

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI" Corso di Laurea Triennale in Fisica

TESI DI LAUREA TRIENNALE

Background discrimination techniques using Artificial Neural Networks for the GERDA experiment

Relatore: Dott. Alberto Garfagnini

Laureando: Mirko Pittaluga Matricola: 1008693

Anno Accademico 2012 - 2013

Contents

1	Intr	troduction 1				
	1.1	GERDA	1			
		1.1.1 GERDA suppression of external background	2			
		1.1.2 GERDA detectors	3			
	1.2	Artificial Neural Network (ANN)	3			
		1.2.1 ANN Learning	5			
		1.2.2 ANN Topology	5			
2	Dat	a analysis framework	7			
	2.1	Signals and background characteristics	7			
	2.2	Energy Spectrum	8			
	2.3	Software Environment	10			
		2.3.1 BOOT	10			
		2.3.2 Gelatio	10			
		2.3.2 TMVA	11			
3	Wor	rking method	13			
	3.1	Analysed signals	13			
	3.2	Parameter choice	13			
	3.3	Used ANN	15			
	3.4	Efficiency evaluation	15			
4	Res	ults	17			
	4.1	Simulated Pulses	17			
	4.2	ANN testing efficiency	18			
	4.3	ANN DEP/Bi peak evaluation	19			
	4.4	Compton edge cut	19			
5	Con	clusions	23			
Bi	Bibliography 25					

Introduction

The aim of this thesis is to perform an analysis of Signal recognition and Background rejection in order to outline a framework for data analysis based on Artificial Neural Network (ANN) suitable to the discrimination of samples from the GERDA experiment.

To be more precise, this work will focus on setting up an ANN that, through Pulse Shape Analysis techniques, is able to recognize a possible $0\nu\beta\beta$ decay from background events. Before dealing with this issue, we will introduce the main arguments this thesis harks back; we will briefly describe the GERDA experiment and the ANN theory.

1.1 GERDA

The GERmanium Detector Array (GERDA [1]) is a low background experiment at the Laboratori Nazionali del Gran Sasso (LNGS) designed to search for the neutrinoless double beta decay ($0\nu\beta\beta$) of ⁷⁶Ge. Double beta decay is the simultaneous beta decay of two neutrons in a nucleus. It is a second order weak process, predicted by the Standard Model:

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

The process $(2\nu\beta\beta)$ has been experimentally observed in even-even nuclei and can be detected only when the single beta decay is energetically forbidden. Typical half-lives of $2\nu\beta\beta$ are very large (between 10¹⁸ yr and 10²⁴ yr). Several extensions of the Standard Model predict that also neutrinoless double beta decay $(0\nu\beta\beta)$

$$(A,Z) \rightarrow (A,Z+2) + 2e^{-1}$$

could occur. Its observation implies that the lepton number is violated by two units; it would prove that neutrinos have a Majorana mass component and it would have far-reaching consequences (in other words, it would be the signature of new physics). For experimental results and a recent review on $0\nu\beta\beta$ see [2].

The GERDA experiment in Europe, and MAJORANA [3, 4] in the USA are the current state of the art experiments to search for $0\nu\beta\beta$ in ⁷⁶Ge using germanium detectors. The process has a clear signature, with a mono-energetic line in the observed energy spectrum at $Q_{\beta\beta} = 2039.061 \pm 0.007$ keV [5], corresponding to the sum of the two electrons energies. Two previous experiments, Heidelberg-Moscow (HdM) and IGEX, have studied double beta decay in germanium and set limits on the $0\nu\beta\beta$ half-life: $T_{1/2}^{0\nu} > 1.910^{25}$ yr [6] and $T_{1/2}^{0\nu} > 1.610^{25}$ [7, 8, 9], respectively.

In 2004 part of the HdM collaboration claimed evidence for a $0\nu\beta\beta$ peak at $Q_{\beta\beta}$ which corresponds to a half-life central value of $T_{1/2}^{0\nu} = 1.19_{-0.23}^{+0.37}10^{25}yr$ [10].

Recently, the GERDA experiment has published a first measurement on $0\nu\beta\beta$ decay in ⁷⁶Ge, partly excluding the previous claim. The experiment will continue the measurement with more detector mass and improved background suppression [11].

The very long half-life characteristic of a $0\nu\beta\beta$ involves considerable experimental difficulties for any purpose to highlight the decay. Background suppression techniques are crucial to improve the experiment sensibility. In GERDA two kinds of background reduction are applied:

- Suppression of all the possible external background causes;
- Offline background discrimination using Pulse Shape analysis.

The latter has been investigated in this work.

1.1.1 GERDA suppression of external background

The experimental challenge of GERDA is to have nearly background free conditions in the region of interest around $Q_{\beta\beta}$. The main feature of the GERDA design is to operate bare Ge detectors made out of material enriched in ⁷⁶Ge (^{enr}Ge), immersed in 64 m³ of Liquid Argon (LAr). This design allows a significant reduction in the cladding material around the diodes and the accompanying radiation sources. Furthermore, the background produced by interactions of cosmic rays is lower than for the traditional concepts of HdM, IGEX or MAJORANA due to the lower Z of the shielding material and to the place where GERDA takes place (3800 m under the Earth's surface).



Figure 1.1: Scheme of the GERDA structure (taken from [1])

Background sources that have to be taken into account include neutrons and gammas from the decays in the rock of the underground laboratory, radioactivity in support materials, radioactive elements in the cryogenic liquid as well as internal backgrounds in the Ge diodes. These backgrounds were considered in the design and construction of GERDA and resulted in specific design choices, selection of materials used and also in how detectors were handled. Natural Ge (^{nat}Ge) contains about 7.8% ^{76}Ge , and could in principle be used directly for a

 $0\nu\beta\beta$ decay experiment. Indeed, the first searches for $0\nu\beta\beta$ decay did use natural *Ge* detectors [12, 13, 14]. Enriched detectors allow a better signal-to-background ratio and also yield reduced costs for a fixed mass of ⁷⁶*Ge* in the experiment [1].

Figure 1.1 shows a model of GERDA design: the core of the experiment is an array of germanium diodes suspended in strings into a cryostat filled with LAr. The latter serves both as cooling medium and shield. The cryostat is a steel vessel with a copper lining used primarily to reduce the gamma radiation from the steel vessel. The cryostat is placed in a large water tank (590 m³) that fulfills the functions of shielding the inner volumes from radiation sources within the hall, such as neutrons, as well as providing a sensitive medium for a muon veto system. These installations are supported by a steel superstructure.

1.1.2 GERDA detectors

In GERDA, two kind of detectors are used: *Coaxial Detectors* and *Broad Energy Germa*nium Detectors (BEGe) [15].

This work will only focus on data taken from BEGe detectors. These, while operating, are characterised by a low-strength electric field that allows to collect (through charges drifting) all the pulses of events within the detector. Figure 1.2 shows the electric field potential in a BEGe detector.



Figure 1.2: BEGe detector electric field profile (taken from [15])

1.2 Artificial Neural Network (ANN)

Artificial neural networks are mathematical models inspired by animal central nervous systems that are capable of machine learning and pattern recognition. For a review on the subject see [16, 17, 18]

The building unit of a neural network is a simplified model of what is assumed to be the functional behaviour of an organic neuron. For almost all organic neurons roughly three parts can be distinguished: a set of incoming fibers (the dendrites), a cell body (the soma) and one outgoing fiber (the axon). See Figure 1.3.

The axons divide up into different endings, each making contact with other neurons. The bulb-like structures where the fibers make contact are called synapses. Electrical pulses can be generated by neurons (the so-called firing neurons) and are transmitted along the axon to the synapses. When the electrical activity is transferred by the synapses to another neuron, it may contribute to the excitation or inhibition of that neuron.

The synapses play an important role because their transmission efficiency for electrical pulses

from an axon to the dendrites of other neurons can be changed depending on the 'profitability' of that alteration. The learning ability of human beings is probably connected to the facility of changing the transmission efficiency of those synapses. The change of the synaptic transmission efficiency acts as memory for past experiences [16]. We can consider an ANN as a highly simplified model of the structure of a biological neural network. In our artificial model of the



Figure 1.3: Example of biological neuron (taken from [16])

neuron, the synaptic transmission efficiency is replaced by a real number w_i that weights the dependence of the behaviour of the receiving neuron from the state of the firing neuron. The emission of pulses by the neurons is modelled by the variable x_i , which represents the state of the neuron. If x_i assumes only $\{0, 1\}$ values, we talk about a binary artificial neuron, if x_i assumes all the values between [0, 1], then we talk about a continuous artificial neuron. In almost all artificial models of neurons, all inputs x_i are weighted by the synaptic transmission efficiency and summed in the so-called *transfer function*:

$$S = \sum x_i w_i$$

The S function determines through the activation function (ϕ) the output value o_j of the artificial neuron, see Figure 1.4.

In a binary neuron, for example, the output o_j will be 1 (neuron firing) if the weighted input exceeds a threshold θ_j , and will be 0 (the neuron is silent) if the weighted input is below the threshold.

In a continuous artificial neuron, the *activation function* may be some monotone increasing function of the weighted input. An ANN is a set of artificial neurons structured according to



Figure 1.4: Example of artificial neuron

precise topology classes. The network receives external signals on an input layer of neurons,

each of which is connected with a number of internal nodes organized in several levels. Each node processes the received signals and transmits the result to subsequent nodes.

It has been shown that an ANN built with this kind of artificial neurons behave in the same way as a computer, more precisely, it is possible to implement networks reproducing all the basic logical operations.

For an ANN all the information about its behaviour is stored in w_i and θ_j values; it is therefore clear that any change of w_i and θ_j values involves modifications in the ANN behaviour.

1.2.1 ANN Learning

The most interesting property of ANN is their capability to learn, i.e.: Given a specific task to solve (T), and a class of functions (F) that solves the problem, learning means find the optimal $f^* \in F$ that solve the problem on the basis of ANN experience.

In this work we will focus on a classification problem using the *Supervised Learning* as learning paradigm.

This means that first of all we will need to provide a set of data with known characteristics (i.e. of known classification) for training our ANN. The fundamental aspect of the training phase is the application of a back-propagation algorithm; an algorithm that by using the known set of input data, modifies the ANN w_i and θ_j values to better accomplish that classification [17].

1.2.2 ANN Topology

There is an incredible variety of types of neural networks, each of which is optimal for a specific application. In this work we will use only two different configurations of MultiLayer Perceptron (MLP), a class of networks characterized by the organization in subsequent layers of the nodes. Each neuron in a layer receives input signals from the previous layer, and its output signals are directed to neurons of the following layer. See Figure 1.5.

An important parameter to define a MLP is the number of hidden layers among the input



Figure 1.5: Example of single hidden layer MLP

and output layers. In our work two types of MLP will be used:

• Single Hidden Layer MLP;

• Double Hidden Layers MLP.

MLP occupies a special role in ANN theory; a mathematical theorem, the Universal approximation theorem, proves that a single hidden layer MLP can approximate arbitrarily closely every continuous function that maps compact subsets of \mathbb{R}^n [18].

This means that using a MLP for a classification problem, after the training with enough data, data can be separated with an arbitrary precision.

Data analysis framework

The main aspects of our work are here described:

- the fundamental properties and characteristics of the signals;
- a qualitative description of the energy spectrum analysed;
- the main software environments used for the elaboration;
- the organization of the data used for the analysis.

2.1 Signals and background characteristics

As explained in the introduction, the aim of this work is to create a methodology (or even a tool') to recognise a possible $0\nu\beta\beta$ decay.

Since the only products of the considered reaction are two electrons, all the reaction energy is released in a precise position of the detector. Events of this type are called *Single Site Events* (SSE).

On the contrary, events corresponding to the simultaneous release of energy in several places of the detector (e.g. multiple Compton scattering) are classified as *Multiple Site Events* (MSE). Examples of SSE or MSE in the detector are shown in Figure 2.1.



Figure 2.1: Scheme of SSE and MSE events in the detector (taken from [19])

The exact energy released in the $0\nu\beta\beta$ decay is equal to the Q-value of the reaction $(Q_{\beta\beta})$; the energy spread is given by the detector resolution.

The difficulty of the analysis (as stated in the introduction) comes from the very long halflife time $(T_{1/2})$ expected for $0\nu\beta\beta$ in ⁷⁶Ge, which leads to a very small number of expected decays (considering a detector used in GERDA and supposing a year-running experiment, it is expected to observe 4 decays).

It is known from the physics of detectors that any measuring apparatus is affected (in spite of how many precautions are applied) by the inevitable presence of background. Considering the smallness of the awaited sample, it is clear that even the slightest background signal can greatly affect the measurement.

Since a priori techniques cannot remove background, we will need to develop off line techniques to distinguish the good signal from the background. These techniques are called *Pulse Shape Analysis* (PSA) [19].

Each event within the detector corresponds to an electrical pulse of precise shape (and energy) recorded by the same. Two examples of signal detected by Germanium detectors are displayed in Figure 2.2.



Figure 2.2: SSE and MSE examples (taken from [15])

Applying PSA means developing techniques that allow the recognition of $0\nu\beta\beta$ from the recorded impulses. The first parameter that will be used is clearly the energy of the signal; the second one is the signal shape.

In the GERDA apparatus of measurement the only 'physical' event expected at $Q_{\beta\beta}$ energy is the $0\nu\beta\beta$ decay, so events recorded at $Q_{\beta\beta}$ energy would be due mainly to two causes: $0\nu\beta\beta$ decay or multiple background events superimposing (characterized by $Q_{\beta\beta}$ summed energy).

The first target of this work is to determine the parameters that from the shape of the signal allow to recognize whether it is a SSE or a MSE.

2.2 Energy Spectrum

As just introduced, the main features of the signal used to distinguish $0\nu\beta\beta$ from background events are pulses energy and shape; we are interested, in particular, in distinguish a SSE or MSE by the signal shape analysis.

Since our goal is to use neural networks for discrimination, we need to find a sample of data to train our network. This means we have to retrieve a set of events that are almost exclusively Single Site (to teach the networks to recognize good events), and a set of events that are almost exclusively Multiple Site (to teach the networks to recognize bad events).

We retrieve such a set of events from the test measurements carried out on the detectors used in GERDA. In these measurements the detectors are exposed to high-energy (γ) radiation by known sources (in particular ²²⁸Th).

We expect that the detected γ radiation spectrum will be constituted by all the possible photonmatter interactions: *photoelectric effect*, *Compton scattering*, *Pair Production*. The Pairs Production spectrum is particularly interesting for our purposes [20].

Pair production is a process in which the energy of the photon is converted to a positronelectron pair. The photon energy has therefore to be twice grater than the energy corresponding to the rest mass of the electron (i.e. $2m_0c^2 = 1022$ keV). The excess energy, $E_{\gamma} - 2m_0c^2$, is shared between the two particles as kinetic energy. Both the electron and the positron produced will be slowed down in the adjacent material and the positron will finally react with another electron and annihilate. If this process occurs after the positron has lost all of its kinetic energy, two photons with energy of about $m_0c^2 = 511.0$ keV are generated.

Although it is energetically possible from 1022 keV, pair production starts to play a role for energies above 1.5 MeV. In small detectors, pair production mechanism is associated to multiple peaks in the energy spectrum. If the two photons generated from the positron annihilation escape from the detector, a $E_{\gamma} - 2m_0c^2$ energy will be detected and a "double-escape" energy peak (DEP) at this value can be observed in the spectrum. If one of the positron annihilation generated photon interacts in the crystal, a "single-escape" peak (SEP) at $E_{\gamma} - m_0c^2$ can be observed in the spectrum. The interaction of both positron annihilation photons within the detector leads to a "full-energy" peak (FEP) in the spectrum at the energy E_{γ} [20]. Look at Figure 2.3 for a graphic representation of pair production interactions with the detector.



Figure 2.3: Example of FEP, SEP, DEP production mechanism (taken from [19])

FEP, DEP and SEP peaks, due to their production mechanism, are easily classifiable as SSE or MSE. The SEP and FEP events are clearly MSE, while the DEP events are SSE. Considering DEP peak events is therefore equivalent to considering a set of almost completely SSE, therefore we obtained a set of "good" signals to train networks. To train the network in recognizing MSE we will use signals belonging to a bismuth (Bi) emission peak at an energy very close to that of DEP events; the characteristic of this peak is that it is composed mostly by MSE. See Figure 2.4 for a graphical representation of used peaks.

We have identified at this juncture the two data sets required to train our neural networks. We now need to build neural networks and proceed with the analysis.



Figure 2.4: Example of FEP, SEP, DEP peaks extracted from Bhima Detector

2.3 Software Environment

Our analysis is conducted through existing software environments, which are described below.

2.3.1 ROOT

ROOT[21] is an object oriented framework for large scale data analysis officially developed by CERN since 2002.

ROOT, as well as being a powerful tool for the analysis, also provides several libraries for data organization and manipulation. Particularly useful for our work is the organization of files in structures called Trees and substructures called Branches, with which we will organize all the data we are dealing with. In general, all the work done in this thesis (including other software environments) use ROOT (*ver.* 5.34/09) to organize and process the data.

2.3.2 Gelatio

Gelatio[22] is the framework for digital signal analysis specifically developed for GERDA. It is based on MGDO[23] library, and it is designed to support multichannel and modular analysis. Each module used by Gelatio handles a precise and self-consistent task of the signal processing. We are now going to shortly describe the characteristic Gelatio signal processing flow and the data structure.

GERDA data model

The charge pulses from the high purity Ge detectors operated in GERDA are digitized by 14-bit flash-ADCs (FADC) running at 100 MHz sampling rate. For each event, the FADC computes in run-time two traces that are eventually written to disk. The first trace is sampled at 100 MHz and is 4 μs long (high-frequency-short trace). It includes the signal leading edge and it is used to identify background events through pulse shape discrimination techniques. The second trace has a sampling frequency of 25 MHz and is 160 μs long (low-frequency-long trace). It is used for those operations, as energy reconstruction, which involve the integration of the signal. The two traces are processed along different chains of Gelatio modules (see Figure 2.5) [22]. Just to name some examples of information extractable from the signal, using specific



Figure 2.5: Gerda signal processing flow (taken from [22])

modules we can deduce direct information as the signal's amplitude, energy and rise-time, or, with other modules, we can obtain more complex information as the reconstruction of the current associated to the event.

Every module is composed by smaller units, the MGDO transformations. Using MGDO transformations is possible to implement custom modules in order to derive from electrical impulse the desired information. In our analysis, we will implement a module to extract information on the way the signal rises.

Data structure

The data analysis with Gelatio is organized in several consecutive steps. Data organizations varies according to phases; data are organized in a hierarchical *tier* structure [24].

From the raw data, in which each event is associated to the signal as detected by the FADC $(tier\theta)$, is obtained by a software conversion a .root file in which all events are organized in trees (tier1). Files from tier1 to tier2 are obtained by applying the modules implemented in Gelatio (in tier2 modules value are associated to each event). If data are further manipulated (e.g. quality cut or energy calibration) we than obtain the tier3.

2.3.3 TMVA

The Toolkit for MultiVariate Analysis (TMVA) [26] provides a ROOT-integrated environment for the processing, parallel evaluation and application of multivariate classification. All



Figure 2.6: Gelatio data organization (taken from [25])

multivariate techniques in TMVA belong to the family of "supervised learning" algorithms. They make use of training events, for which the desired output is known, to determine the mapping function that describes a decision boundary (classification). TMVA package includes Artificial neural networks MVA.

TMVA usage can be divided in two phases:

• Training and testing phase. In this phase the supervised learning is applied. Using the good-signals data set, and bad-signals data set, TMVA teaches the selected MVA methods to recognize the different events.

ANNs associate to every event a number in the real axe. In principle, good events are associated to number 1 and bad events to 0; during this phase ANNs are trained (through w_i values modification) to evaluate the signals of the good data set with a number close to 1, and the signals of the bad data set with a number close to 0. At the end of this phase a weight file that store all the final w_i and θ_j values estimated by the chosen learning algorithm is created. The ANN training and testing phase also provide a cut-value, that is the optimal value in the real axe that allows to separate good from bad events. Once the cut-value has been estimated and the weight file has been created, is possible to proceed with the second phase.

• Application. In this phase the MVA methods previously trained are applied to an unknown classification data set. The MVA parameters are loaded from the weight file. In this phase (in ANN case) every event is evaluated by the network and associated to the network response. If the response is higher than the cut-value, the event is considered "good", otherwise the event is considered "bad". At the end of this phase, the results are shown and stored in TMVApp.root file.

Working method

3.1 Analysed signals

First of all, we have to choose the signals to analyse. We will use the data recorded by BEGe detectors *Bhima* and *Agamennone* during a calibration measurement with ^{228}Th . The calibration conditions are listed below:

Detector name	Voltage (V)	Acquisition duration (s)	Acquisition date
Bhima	4000	28200	26/09/2012
Agamennone	4000	25200	09/03/2012

The source data $(tier\theta)$ contain the raw FADC signal recorded for the entire acquisition duration. The first step is about generating a *tier1* file from the raw data. In *tier1* data are stored in a .root file in the form of individual events; the corresponding recorded electrical signal is associated at each event. Such passage from *tier0* to *tier1* is accomplished by a Gelatio script (*Raw2MGDO*).

Once the *tier1* has been obtained, it is than possible to extract all the information related to the signals using Gelatio modules. We now have to choose which signal parameters to use for the classification.

3.2 Parameter choice

In Figure 3.1 two examples of signal taken from *Bhima* are shown. The upper graphs show the events pulse (i.e. the electric signals as collected by the detector), the lower graphs show the derived of the pulse (i.e. the current signal associated to the events). Looking at the lower graphs it is easy to recognize which of the two events is a Single Site (SS) and which is a Multiple Site (MS); the two signals have been chosen at two comparable energies. The recognition was possible through the observation of the current graphs. While the current graphs of SSE show a single, well shaped, peak, the current graphs of MSE show multiple peaks. This is connected to the non-simultaneous collection of signals associated to multiple site event.

If we observe with sufficient sensibility the pulse rising edge, we can notice that the pulse edge of SS signals grow rapidly and all at once, while the edge of the MS signals have multiple growth points (in this case two) due to the superimposition of different site events.

For training and application with ANNs we have to provide real number parameters that efficiently describe the nature of the event (SSE or MSE). As we have just pointed out, the



Figure 3.1: Pulse examples taken from Bhima

analysis of the signal rising edge is an effective method to classify the signals. In this work we choose to deduce the classifying parameters from the signal edge profile.

We will extract 50 values form the signal edge, corresponding to the normalized rise time when the pulse reaches 1, 3, 5, ..., 99% of the full height; the time when the pulse height reaches 50% serves as reference. In other words, for each pulse we calculate the time taken by the signal to pass from 1% to 50% of its maximum amplitude, and we then divide it by the time taken by the signal to pass from 1% to 99%. We repeat the same procedure for the intervals 3%-50%, 5%-50%, \ldots , 47%-50%, 49%-50%, 50%-51%, 50%-53%, \ldots , 50%-97%, 50%-99%.

See Figure 3.2 for a graphical representation of the procedure.

Since there is not any Gelatio module that derives these signal parameters, we have imple-



Figure 3.2: Pulse sampling example

mented one using MGDO libraries.

3.3 Used ANN

Once the parameters for the pulse description have been chosen, we just have to choose the ANN to use for the classification. In parallel, we will conduct the analysis with three different ANN, the networks differ in topology and in the learning algorithm. Using TMVA, the implementation of each neural network is very simple; to build an ANN it is sufficient use a TMVA Factory object method specifying in the options the network type and topology. We use the following ANN:

- MLP using TMVA learning algorithm with a single 54 nodes hidden layer. For an example of a simple single layer structure see Figure 1.5;
- MLP using TMVA learning algorithm with double hidden layer (51 nodes in the first layer, 50 nodes in the second layer). For an example of a double single layer structure see Figure 3.3;
- MLP using root learning algorithm with double hidden layer (51 nodes in the first layer, 50 nodes in the second layer);



Figure 3.3: Double hidden layer ANN example

3.4 Efficiency evaluation

Having concluded the training of ANNs, we now need methods to evaluate the efficiency of built networks. For this purpose, we will apply the criteria presented below.

TMVA testing efficency

At the end of every ANN training and testing, TMVA provides an ANNs ranking based on efficiency. We will use the testing results as a first efficiency parameter.

Simulated pulse classification

Using the information on *Bhima* detector (shape and electrodes position), it is possible to simulate pulse associated to SSE or MSE occurring in different detector positions. The advantage of simulations is that the background-free pulses so obtained constitutes a good test case for ANNs. Simulations are particularly effective to verify if the chosen signal parameters are good to identify SSE and MSE.

When using simulated pulses for parameters check only training/testing phase is executed.

DEP/Bi peak evaluation

After training and testing the networks with DEP and Bi peaks a good efficiency parameter for the ANN is the ratio of accepted pulses over total pulses. This ratio quantifies how the built networks recognize/reject events certainly MSE or SSE. We expect the most of DEP events to be accepted and the most of Bi events to be rejected.

Compton edge cut

After training and testing the networks with DEP and Bi peaks, we can check how the Compton edge events are classified. Physically the most of these pulses are MSE, and therefore should be rejected by the network.

Results

The analysis results are presented below.

4.1 Simulated Pulses

Table 4.1 shows the results of training and testing phase on the Bhima simulated events. For each ANN type its efficiency and optimal cut-value are shown; the separation value, which estimate the degree of separation of the two samples and so its classification efficiency, is particularly interesting. The listing order represents the efficiency ranking of ANNs. In Figure 4.1 a representation of the classification efficiency for two ANNs is given.



Figure 4.1: Bhima simulation testing distribution

ANN Type	Separation value	Optimal cut-value
root MLP Double Layer	0.580	0.3884
MLP Double Layer	0.572	0.3163
MLP Single Layer	0.558	0.2467

Table 4.1: ANN evaluation and cut-value for simulations

4.2 ANN testing efficiency

Table 4.2 presents the TMVA testing results on Bhima and Agamennone detectors.

	ANN Type	Separation value	Optimal cut-value
	MLP Single Layer	0.541	0.6158
Bhima	MLP Double Layer	0.536	0.5671
	root MLP Double Layer	0.531	0.5842
	MLP Single Layer	0.563	0.6375
Agamennone	MLP Double Layer	0.562	0.5427
	root MLP Double Layer	0.552	0.6354

 Table 4.2: ANN evaluation and cut-value

Figure 4.2, 4.3, 4.4 and 4.5 show the networks evaluation for the test sample (the biggest figures are associated to the more efficient ANN). These graphs show the distribution of network evaluation of signal and background data set. The more the signal sample is close to 1 and the background sample is close to 0, the better the classification is accomplished.





(b) Bhima root double hidden layer

Figure 4.2: Bhima testing distribution (double hidden layer and root double hidden layer ANN)



Figure 4.3: Bhima testing distribution (single hidden layer ANN)



Figure 4.4: Agamennone testing distribution (double hidden layer and root double hidden layer ANN)



Figure 4.5: Agamennone testing distribution (single hidden layer ANN)

4.3 ANN DEP/Bi peak evaluation

After the ANNs training on DEP/Bi events, we have evaluated how the peaks were classified by the ANNs. To do this, in Figure 4.7 and 4.6, we have superimposed a histogram with only accepted events (by the ANN) to a histogram with all detected events, both for *Bhima* and for *Agamennone*. Fitting the two peaks with Gaussian functions (red for all events, blue for accepted events) we have estimated the peaks area before and after the ANN application; in Table 4.3 the estimated areas and their ratio are displayed.

4.4 Compton edge cut

Figure 4.8 and 4.9 show the application of double hidden layer ANN on *Bhima* spectrum and single hidden layer ANN on *Agamennone* spectrum respectively. The figures show both the cuts on the complete spectrum and a magnification on the Compton edge (red histogram for the Total Spectrum, blue histogram for the Accepted Spectrum).

Table 4.4 shows the $\frac{Accepted Events}{Total Events}$ parameter for the Compton edge per each ANN and detector.

Total number of events						
		DEP	Bi			
	Bhima	20284	111(02		
	Agamennone	10478	954	.8		
	ANN Type	DEP A	Acc.	Bi Acc.	$\frac{DEPAcc.}{DEPTot.}$	$\frac{BiAcc.}{BiTot.}$
	MLP Single Layer	1829)4	2130	0.902 ± 0.002	$0.192{\pm}0.004$
Bhima	MLP Double Layer	1861	.8	2384	$0.918{\pm}0.002$	$0.215 {\pm} 0.004$
	root MLP Double Layer	1860)8	2365	$0.917{\pm}0.002$	$0.213 {\pm} 0.004$
	MLP Single Layer	889	9	1110	$0.849 {\pm} 0.003$	$0.116 {\pm} 0.003$
Agamennone	MLP Double Layer	911	5	1270	$0.870 {\pm} 0.003$	$0.133{\pm}0.003$
	root MLP Double Layer	896	6	1165	$0.856 {\pm} 0.003$	$0.122{\pm}0.003$

Table 4.3:	DEP/Bi	peack	acceptance	ratio
------------	--------	-------	------------	-------



Figure 4.6: Agamennone root Double Layer ANN peaks comparison (DEP peak on the left, Bi peak on the right)



Figure 4.7: Bhima Single Layer ANN peaks comparison (DEP peak on the left, Bi peak on the right)

	Total	Total Compton edge events			
	Bhima	318003			
	Agamennone	236410			
	ANN Type	Compton edge ev. Accepted	$\frac{ComptonAcc.}{ComptonTot.}$		
	MLP Single Layer	5097	0.0160 ± 0.0002		
Bhima	MLP Double Layer	6849	$0.0215{\pm}0.0003$		
	root MLP Double Layer	12664	$0.0398{\pm}0.0003$		
	MLP Single Layer	1603	$0.0068 {\pm} 0.0002$		
Agamennone	MLP Double Layer	9294	$0.0393{\pm}0.0004$		
	root MLP Double Layer	8799	$0.0372 {\pm} 0.0004$		

 Table 4.4:
 Compton edge acceptance ratio



Figure 4.8: Bhima Double hidden layer ANN application on the spectrum



Figure 4.9: Agamennone Single hidden Layer ANN application on the spectrum

Conclusions

From the analysis of simulated events, emerges a good separation between SSE and MSE after training and testing phase. The separation values obtained from simulations are good and very close to those obtained from the testing phase with real events (although the best are clearly those related to simulations). This first observation proves that the chosen classification parameters allow to distinguish SSE from MSE, thus giving a first confirmation of the used method efficiency.

We point out that during the analysis of both simulated and real events, the training and testing sample was composed by 20000-30000 pulses. Due to the used classification method, the ANNs topology is characterized by a lot of nodes and consequently a lot of nodes connection (about 2500 $w_{i,j}$ for single layer MLP, and 5000 $w_{i,j}$ for double layer MLP, to estimate).

Real sample testing phase highlights the following ANN efficiency order:

- 1. Single hidden layer MLP
- 2. Double hidden layer MLP
- 3. Double hidden layer MLP with root learning algorithm

The single hidden layer MLP is the most efficient network both for *Bhima* and *Agamennone* even though its the most simple ANN. This is probably due to the just outlined training sample number of events. Considering the high number of $w_{i,j}$ to estimate, we should have a lot of training events to reach an optimal training. Probably 20000-30000 pulses are not enough for the training of our networks, especially for the more complicated ones. A possible cause of the higher efficiency of single layer ANN is that being the network with less connections, its the network that comes more close to the optimal training condition using a small training sample.

From DEP and Bi peaks acceptance ratio results that 80%-90% of DEP events are accepted, while only 10%-20% of Bi evens are accepted. It can be noticed that ANNs trained with *Bhima* events have a higher DEP acceptance ratio, while *Agamennone* trained ones have a higher Bi rejection ratio. This is probably due to the different events number of DEP and Bi training samples in the two cases. While for *Bhima* the DEP training sample is sized twice than the Bi training sample (with a consequent higher DEP acceptance), for *Agamennone* DEP and Bi training sample have the same size (with a consequent higher Bi rejection).

Overall, the acceptance/rejection ratios are very good, and therefore ANNs trained with this method are a good instrument for background discrimination.

All ANNs execute a good Compton cut, just about 1%-4% of Compton edge events (that are principally MSE) are accepted. This result (the cuts are even too good) made us suspicious, bringing us to analyse the entire detectors spectrum. From Figure 4.8 and 4.9 can be noticed how very few events with an higher energy than the Bi peak are accepted, highlighting the possibility that the ANN are executing an energy cut rather than a SSE, MSE selection.

We conclude our analysis stating the effectiveness of ANNs classifications, that have very high performances in the DEP/Bi peak recognition and in Compton cut. We report, however, some critical aspects of our analysis:

- an ANN with the same topology of those we used, needs a big-size training sample to reach an optimal training;
- considering the impossibility of completely understand the working mechanism of an ANN (that acts in a sense like a black-box), there is the need of support the ANN analysis with another, more transparent, method with which confirm ANN effectiveness.

For further developments of the work we would suggest to:

- given the complexity of the used networks, increase the size of the training samples;
- obtain a training sample with the same number of acceptable and unacceptable events;
- verify if the used ANNs execute an energy cut, using alongside ANNs analysis an already tested method (e.g. A/E cut).

Bibliography

- [1] K.H. Ackermann et al (GERDA Collaboration). The GERDA experiment for the search of $0\nu\beta\beta$ decay in ⁷⁶Ge. Eur. Phys. J. C, (73):2330, 2013.
- [2] B. Schwingenheuer. Ann. Phys, (525):269, 2013.
- [3] D. G. Philips II et al. The MAJORANA Experiment: and ultra-low background search for neutrinoless double-beta decay.
- [4] P. Finnerty et al. The MAJORANA demonstrator: Progress towards showing the feasibility of a tonne-scale ⁷⁶Ge neutrinoless double-beta decay experiment. Mexico, June 2012. XVIII International Symposium on Particles, Strings and Cosmology.
- [5] B. J. Mount, M. Redshaw, and E. G. Myers. *Phys. Rev. C*, (81):032501, 2010.
- [6] H. V. Klapdor-Kleingrothaus et al (HdM Collaboration). Eur. Phys. J. A, (12):147, 2001.
- [7] C. E. Aalseth et al (Igex Collaboration). Phys. Rev. D, (65):092007, 2002.
- [8] A. Morales. Nucl. Phys. B, (77):335, 1999.
- [9] J. Morales and A. Morales. Nucl. Phys. B, (114):141, 2003.
- [10] H. V. Klapdor-Kleingrothaus et al. Phys. Lett. B, (586):198, 2004.
- [11] M. Agostini, M. Allardt, and E. Andreotti et al (GERDA Collaboration). Results on neutrinoless double beta decay of ⁷⁶Ge from Gerda Phase I. arXiv:1307.4720v1, 2013.
- [12] A. Morales et al. J. Phys. G, (17):S211, 1991.
- [13] J. Busto et al. Nucl. Phys. A, (513):291, 1990.
- [14] A.A. Vasenko et al. Mod. Phys. Lett. A, (5):1299, 1990.
- [15] M. Agostini et al (GERDA Collagoration). Pulse shape discrimination for GERDA Phase I data. arXve:1307.2610v1, July 2013.
- [16] L. P. J. Veelenturf. Analysis and Applications of Artificial Neural Networks. Prentice Hall, Hemel Hempstead, Hertfordshire, 1995.
- [17] B. Yegnanarayana. Artificial Neural Networks. Prentice Hall of India, New Delhi, 2006.
- [18] S. Haykin. Neural Networks: A Comprehensive Foundation, volume 2. Prentice Hall, 1998.
- [19] R. Brugnera. L'esperimento gerda. Conference presentation, May 2013.
- [20] K. Debertin and R. G. Helmer. Gamma and X-ray spectrometry with semiconductor detectors. Elsevier Science Publisher B.V., Amsterdam, The Netherlands, 1988.

- [21] K. Debertin and R. G. Helmer. Root users guide, May 2013. Official CERN guide.
- [22] M Agostini, L Pandola, and P. Zavarise. Off-line data processing and analysis for the GERDA experiment. Technical report, November 2011.
- [23] M. Agostini, J. A. Detwiler, P. Finnerty, et al. The MGDO software library for data analysis in Ge neutrinoless double-beta decay experiments. arXiv:1111.7260, November 2011.
- [24] M. Agostini, L. Pandola, and P. Zavarise. GELATIO The GERDA framework for digital signal analysis. ACAT 2011 Conference presentation, September 2011.
- [25] M. Agostini, L. Pandola, and P. Zavarise. Analysis workshop: Introduction and layout. Tubingen University Conference presentation, October 2011.
- [26] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA 4 Toolkit for Multivariate Data Analysis with ROOT, November 2009. Official TMVA user guide.