The GREYSTAR prototype
Designing a multidetector for automatic isotope identification

Thèse

Barton Wallace

Doctorat en physique
Philosophiae doctor (Ph. D.)

Québec, Canada

The GREYSTAR prototype
Designing a multidetector for automatic isotope identification

Thèse

Barton Wallace

Sous la direction de :

René Roy, directeur de recherche

Résumé

L'utilisation de la spectroscopie nucléaire comme outil de détection est limitée par le ratio du signal sur bruit. Les méthodes traditionnelles qui permettent d'outrepasser cette difficulté sont d'augmenter l'intensité du signal ou de diminuer le bruit. Pour amplifier le signal, il existe deux méthodes : avoir un émetteur plus intense ou avoir un détecteur de meilleur rapport signal sur bruit . Les techniques typiques de réduction de bruit de fond sont : l'application de filtres numériques et l'ajout de systèmes, tels des circuits de coïncidence, pour favoriser l'étude du signal. Dans le cadre de notre recherche, nous cherchons à détecter des traces d'éléments en utilisant l'activation par neutron. Nous sommes donc limités à avoir des émetteurs d'intensité faible. Beaucoup de chercheurs se penchent déjà sur la création et l'amélioration de détecteurs ayant de meilleures résolutions en énergie. L'approche que nous proposons utilise des détecteurs typiques ayant des résolutions standards et aucun système de suppression de bruit de fond. Nous avons développé un montage expérimental astucieux qui se charge de récolter un maximum d'information. Nous le couplons par la suite à un système d'analyse statistique robuste basé sur l'estimation du maximum de vraisemblance d'un modèle de mixture pour extraire les paramètres d'intérêt. Cette méthode mène à une amélioration flagrante du pouvoir de détection d'un tel système. Les résultats de cette recherche sont par la suite très utiles à tous les systèmes de spectroscopie semblables. Les méthodes d'acquisition et d'analyse développées sont en parfaite synergie avec les travaux de recherche qui se font ailleurs sur des systèmes de détecteurs plus exotiques.

Abstract

The use of nuclear spectroscopy as a detection tool is limited by signal to noise ratio. The most common ways of overcoming this difficulty are either boosting the signal or dampening the noise. Boosting the signal can either be done by having a more powerful emitter or by having a better signal to noise ratio. Applying digital filters and having special experimental setups are the typical ways of enhancing the signal. In our case, the detection of trace elements with neutron activation analysis, we are limited by the nature of the problem to having small emitter signals. A good amount of research is already being done to develop new and exotic detectors to improve their energy resolution. The approach we propose uses typical detectors with average resolution and no background suppression. We propose an intelligent experimental design intended to extract a maximum of information from the experiment and an analysis algorithm based on a rigorous statistical model using a maximum likelihood approach to solving a mixture model in order to drastically improve detection efficiency. This method can be combined with more complex setups that use higher resolution detectors. As such the results of this project synergize well with what is being done elsewhere in the fields of spectroscopy and detector physics.

Table of contents

Résumé iii
Abstract iv
Table of contents v
List of tables vii
List of figures viii
List of acronyms x
Acknowledgements xi
Introduction 1
1 Physical foundations 4
1.1 Introduction 4
1.2 Neutron source 4
1.3 Neutron activation analysis 8
1.4 Scintillation detectors 13
1.5 Gamma spectroscopy 18
1.6 Conclusion 21
2 Data acquisition system 22
2.1 Introduction 22
2.2 Detectors 22
2.3 Analog to digital converter (ADC) 27
2.4 Field programmable gate array (FPGA) 29
2.5 The control FSM 30
2.6 C# data logger and Graphical user interface (GUI) 33
2.7 Results 34
2.8 Conclusion 39
3 State of the art in peak identification 41
3.1 Introduction 41
3.2 The non parametric side 41
3.3 The parametric side 42
3.4 Asking the right question 43
3.5 Shortcomings of traditional methods 43
3.6 Our proposed solution 46
3.7 Conclusion 47
4 Analysis system 48
4.1 Introduction 48
4.2 Statistical model 48
4.3 Metaheuristic optimization 58
4.4 Kernel functions (reference spectra) 60
4.5 C# implementation 63
4.6 Results 63

4.7 Conclusion 65
5 Prototype characterization and first results 66
5.1 Introduction 66
5.2 Isotope detection and poor fit discrimination with the GREYSTAR Analysis Tool 66
5.3 Time-energy event recording 69
5.4 Deep analysis with 2D fitting results 71
5.5 Benchmark results for known isotopes and activated materials 72
5.6 Conclusion 74
6 Result analysis and discussion 75
6.1 Introduction 75
6.2 Analysis of first results 75
6.3 Limitations of the system 76
6.4 Discussion and further work 78
6.5 Conclusion 79
Conclusion 80
Bibliography 82
Annexes 87
Bootstrap 88
Random tree classification on field programmable gate arrays 89
On/Off Am-Be neutron source 91
User guide for the GREYSTAR Analysis Tool (GAT). 92
Properties of the materials used in the benchmark tests 97
The Au sample: 97
The Ag sample: 97
The Cu sample: 97
The Al sample: 98
The ^{137}Cs source : 98
The ^{22}Na source: 98
The ^{60}Co source : 98
The ^{54}Mn source : 98
Screen captures of the benchmark tests. 99
Block diagram of the acquisition system. 104

List of tables

Table 1. Classification of neutrons by their kinetic energy. 10
Table 2. Sensitivity (in picograms) of different elements for a thermal neutron flux of 10^{13} n cm^2 s^{-1} [15] 12

List of figures

Figure 1. Energy distribution of a typical Am Be neutron source [47]. 5
Figure 2. Shielding for the neutron source. 6
Figure 3. Drawing of the main steps in neutron activation. 9
Figure 4. Typical graph of neutron capture cross section as a function of kinetic energy [39]. 11
Figure 5. Gamma ray absorption coefficient as a function of energy for natural Al [41]. 13
Figure 6. Relative contribution of each process: Photoelectric, Compton and Pair production as a function of gamma energy [42]. 14
Figure 7. Drawing of the various paths a photon can take when interacting with a scintillation detector [17]. 15
Figure 8. Energy diagram of the electrons in the crystal lattice. 17
Figure 9. Schematic drawing of a spectrum for a monochromatic isotope. 19
Figure 10. Real spectrum of a monochromatic isotope (^{137}Cs). 20
Figure 11. Volumetric prices of different detector material. 23
Figure 12. Canberra model 802 NaI(Tl) scintillation detector with PMT. 25
Figure 13. NIM electronics crate with two high voltage sources and two model 571 Ortec research amplifiers. 25
Figure 14. The acquisition and analysis experimental setup. 26
Figure 15. Cyclone V FPGA starter kit from Terasic, EVAL-AD7352/AD7356/AD7357 ADC from Analog and EVAL-CED1Z power source from Analog. 28
Figure 16. Timing diagram for communication with the AD7356 ADC from Analog. 29
Figure 17. Drawing of a 8-bit UART communication protocol with no parity bit [3]. 32
Figure 18. Screen shot of the GREYSTAR Acquisition Tool graphical user interface. 34
Figure 19. Background spectrum obtained with a Canberra model 802 scintillator coupled with the GREYSTAR system. 35
Figure 20. Background spectrum obtained with a Canberra model 802 scintillator coupled with the EASY-MCA system. 35
Figure 21. Time-Energy plot obtained from event recording with the GREYSTAR system. 36
Figure 22. Kernel density estimation[9] of the time-energy plot with a bandwidth of (0.5, 5). 37
Figure 23. Kernel density estimation of the time-energy plot with a bandwidth of (1, 10). 38
Figure 24. Kernel density estimation of the time-energy plot with a bandwidth of (3, 19). 39
Figure 25. When no conditions are imposed, the fit will trivially pass through each point. 50
Figure 26. The trivial MLE fit with no constraints on pi. 51
Figure 27. Linear regression obtained by estimating parameters a and b. 52
Figure 28. Linear regression (red) that maximizes the combined product of each channel's likelihood function. 53
Figure 29. One dimensional (energy spectrum) kernel for ^{137}Cs. 54
Figure 30. Two dimensional (time-energy spectrum) kernel for ^{137}Cs. 55
Figure 31. Examples of density kernels for various isotopes. 56
Figure 32. Schematic representation of the metaheuristic algorithm. 59

Figure 33. Kernel density estimations of a ^{137}Cs spectrum for different bandwidths. 61
Figure 34. Time dependent (exponential decay) spectrum of ^{137}Cs and time independent (constant) spectrum of background radiation. 62
Figure 35. Spectrum of a low activity ^{137}Cs source. 64
Figure 36. Screenshot of the GUI. 67
Figure 37. Screenshot of the GUI. 68
Figure 38. Screenshot of the GUI. 69
Figure 39. Time-energy spectrum of activated aluminum. 70
Figure 40. The time-energy KDE of activated Al data. 71
Figure 41. Spectrum morphing for different brick depths. 78
Figure 42. Drawing of the proposed geometry of the Am-Be source. 91
Figure 43. The start, stop and clear buttons on the GUI. 92
Figure 44. The log check box and the range selection fields. 93
Figure 45. File name fields on the GUI. 94
Figure 46. System communication window 1. 95
Figure 47. Communication windows 2, 3 and 4. 96
Figure 48. Au 3 minute spectrum and fit. 99
Figure 49. Ag 6 minute spectrum and fit. 100
Figure 50. Cu 6 minute spectrum and fit. 101
Figure 51. Al 3 minute spectrum and fit. 102
Figure 52. Al and Cu 3 minute spectrum and fit. 103

List of acronyms

Most acronyms are introduced in the text. However, if need be, this table sums up most of the acronyms used in this work.

Acronym:	Meaning:	Provenance:
ADC	Analog to Digital Converter	Electrical Engineering
ASIC	Application Specific Integrated Circuit	Electrical Engineering
CPU	Central Processing Unit	Electrical Engineering
FPGA	Field Programmable Gate Array	Electrical Engineering
FSM	Finite State Machine	Computer Science
FWHM	Full Width at Half Maximum	Nuclear Physics
GUI	Graphical User Interface	Computer Science
IC	Integrated Circuit	Electrical Engineering
MLE	Maximum Likelihood Estimation/Estimator/Estimate	Statistics
NAA	Neutron Activation Analysis	Nuclear Physics
OLS	Ordinary Least Square	Statistics
PDF	Probability Density Function	Statistics
PE	Photoelectric Effect	Nuclear Physics
PGAA	Prompt gamma Activation Analysis	Nuclear Physics
PLL	Phase Locked Loop	Electrical Engineering
PMT	Photo Multiplier Tube	Nuclear Physics
UART	Universal Asynchronous Receiver Transmitter	Electrical Engineering
VHDL	VHSIC Hardware Description Language	Electrical Engineering
VHSIC	Very High Speed Integrated Circuit	Electrical Engineering

Acknowledgements

In no particular order here are the people and entities I would like to thank:

My parents, Catherine Morin, Patrick St-Onge, René Roy, Gabriel Bédard, Pierre Amiot, the Céline & Jacques Lamarre Foundation, Lise Dubé, Denise Dugas, the support staff at Laval, the people at TRIUMF, Anne-Sophie Charest, Amine Miled and many others for their support and interest in this project.

Introduction

In the world of elemental spectroscopy, we have two main types of analysis: destructive and non-destructive. Destructive methods include chemical based analysis, mass spectroscopy, and other traditional methods. Non-destructive spectroscopy approaches include neutron activation analysis (NAA) and X-ray fluorescence analysis (XRF). These methods, especially X-ray fluorescence, are fairly limited in penetration depth. As a result, both NAA and XRF are often used on powdered samples, making them destructive methods. NAA having a much better penetration than XRF is a candidate for the spectroscopy of unaltered volumes. It is used for a variety of static analysis systems, such as in nuclear reactors or in-field monitoring of key elements. What we propose is a mobile NAA system capable of in-field measurements. The drawback of such a method is the loss of high neutron fluxes from static nuclear reactors. As such, in this project, an effort is made to improve both the physical setup whose job is to acquire relevant data and the analysis algorithm used to extract useful information about the probed system. Born from necessity for our project, these improvements will find useful applications in traditional spectroscopy systems, medical physics and fundamental research.

The goal of this doctoral project was the elaboration of a detector array prototype following the work done in [1]. Our main objective was to develop a better statistical analysis approach and then demonstrate its functionality by building a working prototype. To test our hypothesis we needed an acquisition system capable of delivering data in an event-recorded format. This type of system was commercially unavailable to us, as such; the design of an adequate acquisition system became necessary. The GREYSTAR prototype can be broken down into two equally important parts. The first part is the acquisition system. It is responsible for transforming physical interactions into digital data. The second part is the analysis system. Its job is to use the digital data provided and extract physical interpretations.

The system we have built is able to detect gamma radiation produced by neutron activation of unknown specimens. It then compares the experimental data to a library of known isotopes and presents a report on the composition of the irradiated sample. The novelty of this project resides in the acquisition of data in the form of events, through the use of field programmable gates arrays for

multi-detector array control, and the application of a robust statistical model combined with a metaheuristic algorithm for data analysis. The use of FPGAs is commonplace in nuclear physics for their flexibility in design. It is this flexibility that enabled us to produce a functional prototype for event driven acquisition. The statistical model used considers detection event as statistical objects that can be modeled by a mixture model. This type of analysis is, to our knowledge, novel in the field of neutron activation analysis. Although many analysis methods have overlapping concepts with our method, we found that building the statistical model from the ground up enabled us to go beyond the application of typical heuristics and avoid pitfalls associated with those heuristics.

Our main focus was integrating: a smart experimental design, computing power and robust statistical analysis to permit the use of traditional scintillator detectors instead of more modern and costly semiconductor detectors for gamma ray spectroscopy. We may add that most if not all of the advances made in this project can be directly applicable to more modern or exotic types of detectors.

An enormous amount of technical work has gone into designing this prototype. I designed a VHDL code that was implanted in a field programmable gate array (FPGA). This code controls and communicates with the analog to digital converters coupled to the detectors. It also does rudimentary pulse shape analysis to extract the pulse height and it assures communication to a laptop via a USB emulated UART. I also wrote a C++/C# graphical user interface (GUI) to control the FPGA and to log the incoming data. A statistical analysis code is coupled to the GUI. I developed the statistical model and coded a metaheuristic algorithm for the analysis code. The result is a high performance acquisition system capable of automatic isotope identification with confidence interval estimation.

A detailed explication of these systems follows a review of the underlying physics used in neutron activation analysis. A list of acronyms can be found at page ix. Technical information, which was not necessary to the text but relevant nonetheless, can also be found in the appendix.

We start the thesis with a review of the physical considerations needed to understand the rudiments of neutron activation analysis. This includes a review of: neutron activation, radioactive decay, and interaction of gamma radiation in matter, scintillator detector physics and gamma spectroscopy.

We then present the acquisition system. In this section we demonstrate that our system has similar capabilities to commercial systems with the added benefits of being cheaper and producing the data

format we are interested in. We also discuss our choice of detectors and electronics. We present the GUI we built to help in data acquisition. We show with a few results that our system is capable of delivering the data in the desired format.

In the third section, we discuss a brief review of typical spectrum analysis techniques. This review is a high level approach that tries to inform the reader as to where our novel method falls on the spectrum of existing methods. Direct comparison of these methods is either outside the scope of this thesis or flat out uninformative. We propose a simple metric to which we will compare our method.

In the fourth section, we detail how our statistical analysis method is constructed. We start from a statistical description of the data and the statistical properties of a histogram. We then construct what is essentially a mixture model of isotope spectrums. A local search algorithm, used in the optimisation process is discussed and the code implementation is presented. We finish by showing the result of a simple experiment that demonstrates the good working of the implemented code.

The fourth section is comprised of a series of simple experimental tests whose purpose is to show the good function of the whole prototype. This section goes deeper in the methodology used for the tests than previous sections. The purpose of this section is to demonstrate that we have succeeded in creating a functional prototype.

The final section is an analysis and discussion about the limits of both the prototype and the method. We review some of the results presented in the fourth section and describe how the initial results tie in with the larger goals of the GREYSTAR project. We outline the future work needed to go from our prototype to a more advanced system. We also discuss the implications of our novel analysis method for other spectroscopy applications.

1 Physical foundations

1.1 Introduction

In this section we review the physical concepts necessary to the understanding of the project. We review neutron capture, beta and gamma decay, interaction of radiation in matter and detector physics. The last part is of key importance for the creation of our statistical model. The rest of the chapter is simply there to help the reader understand the context of neutron activation analysis.

1.2 Neutron source

The choice of neutron sources is discussed at length in [7]. We will recall the important points of that discussion here. Two options were possible for our needs: Isotopic sources and discharge tube sources.

Discharge tubes use deuterium-deuterium or tritium-deuterium reactions to produce high fluxes of fast neutrons. This option is relatively expensive and so was discarded from the present project for monetary reasons. Given the resources, this type of neutron source is by far the most advantageous. The two main reasons are the high neutron flux and the possibility to shut off the source for transport.

Isotopic sources combine an alpha emitter with neutron rich target. The target absorbs the alpha and emits a neutron. The source we chose for the project uses an americium emitter with a beryllium coating. It was the most intense neutron emitter available in the Physics Department at Laval University. It has an alpha activity of 74 GBq that translates to roughly a 5.1 MBq neutron activity at 70 neutrons per million alphas for Am 241 [12]. The neutrons emitted have a distribution in energy as seen in figure 1.

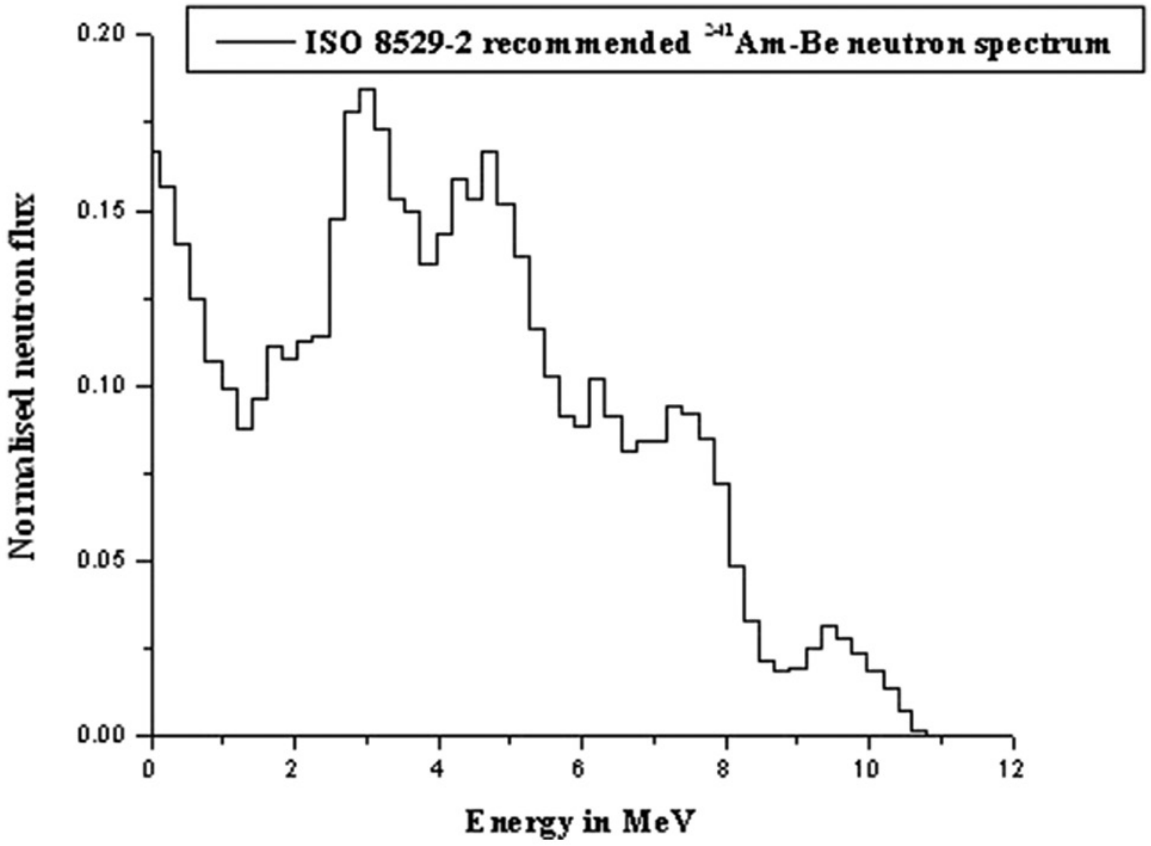

Figure 1. Energy distribution of a typical Am Be neutron source [47].

The source is placed in a lead (Pb) shield to attenuate the gamma radiation from the americium. GEANT4[1] simulations show that this does not affect neutron flux in any significant way [23]. The neutrons are thermalized by 2 cm of paraffin[2]. A mixture of borax and paraffin is used as an absorber to stop neutrons from emitting in all directions. A small solid angle window (~0.2 sr) is not blocked by absorbing material and lets thermal neutrons through for activation purposes. Finally a second lead shielding is placed around the absorber to stop the gamma radiation produced from the boron

[1] Montecarlo simulation toolkit developped by CERN.

[2] This value was chosen as a functional heuristic. It is equivalent to the thermal diffusion length [40] for neutrons in paraffin. We want to minimise the capture of thermal neutrons by the thermalizing material. In the case of partial thermalization, we expect the epithermal neutrons to thermalize in the irradiated sample. We also guide our choice by the remarks found in [46] where for an Am-Be source, 2.5 cm of paraffin were found to be optimal for thermalization.

absorption of the thermal neutrons. Once the neutrons have been thermalized, we expect them to have a Maxwell Boltzmann distribution around the thermal temperature of 0.025 eV.

Figure 2. Shielding for the neutron source.

Figure 2 shows one of the shielding setup for our system, comprised of a lead brick with a hole in the middle where the neutron source can be placed. Around this lead core are large blocks of paraffin. Notice that in this particular setup, the paraffin blocks are not mixed with borax. On the bottom side of the lead core is a thinner 2 cm paraffin block. This thin window leads directly to the area where materials are placed for activation. Around the rest of the wax layer is a second layer of thick lead bricks. We take the time to note that for tests in a controlled environment, such a setup was not crucially necessary. We also note that improvements to our shielding would have to be made in order to improve the mobility of such a heavy structure.

Work was done on proposing a new novel geometry for an Am-Be source. This work is presented in the annex. At the time of writing this thesis, work on this geometry has been put on hold. The purpose of shielding the neutron source is mainly to make its use safer. In reality, for the experimental tests that are done in the context of this thesis, this safeguard is not absolutely necessary given the possibility of doing the activation in a shielded room.

1.3 Neutron activation analysis

Neutron activation is a nuclear reaction where a target nucleus absorbs a neutron. The target nucleus becomes a different isotope of the same element:

$$ {}^{A}_{Z}X + {}^{1}_{0}n \rightarrow {}^{A+1}_{Z}X. \quad (1) $$

Often, the product of the reaction will not be in its ground state. This excited state will decay by gamma emission to find its ground state very quickly (~10^{-12} to 10^{-9} seconds) [7]. The difference between an excited nucleus and an unstable nucleus is crucial. A simple distinction is that the excited nucleus, as stated above, is a nucleus with an excess of energy. This energy can decay to the ground state by emitting electromagnetic radiation. There is no change in the composition of the isotope. Nuclear decay on the other hand changes the proton and neutron composition of the nucleus and produces a different isotope.

If the new isotope is stable we talk only of neutron absorption. In the other more usual case, the neutron surplus in the isotope makes it unstable. This type of instability usually leads to β^- decay:

$$ {}^{A+1}_{Z}X \rightarrow {}^{A+1}_{Z+1}Y + \beta^- + \bar{\nu}_e. \quad (2) $$

This is an example of transmutation. The initial element X has changed to Y by emitting an electron and an antineutrino. This happens because one of the extra neutrons undergoes transformation (to proton) so that the nucleus comes closer to a more stable neutron to proton ratio. This type of reaction is typically much slower than that of gamma decay. In general we can talk of half-lives of the order of microseconds to tens of years. These half-lives are characteristic of the isotopes and can be used to identify the parent nucleus. Once the β^- decay has occurred, our product nucleus finds itself in the same situation as it did right after the neutron absorption. It will usually have an excess of energy that can be radiated through gamma decay. The decay energy of this reaction is well defined, as it is a two-body problem. The energy of the gamma ray is then characteristic of the reaction and can be used to identify the parent and daughter isotopes.

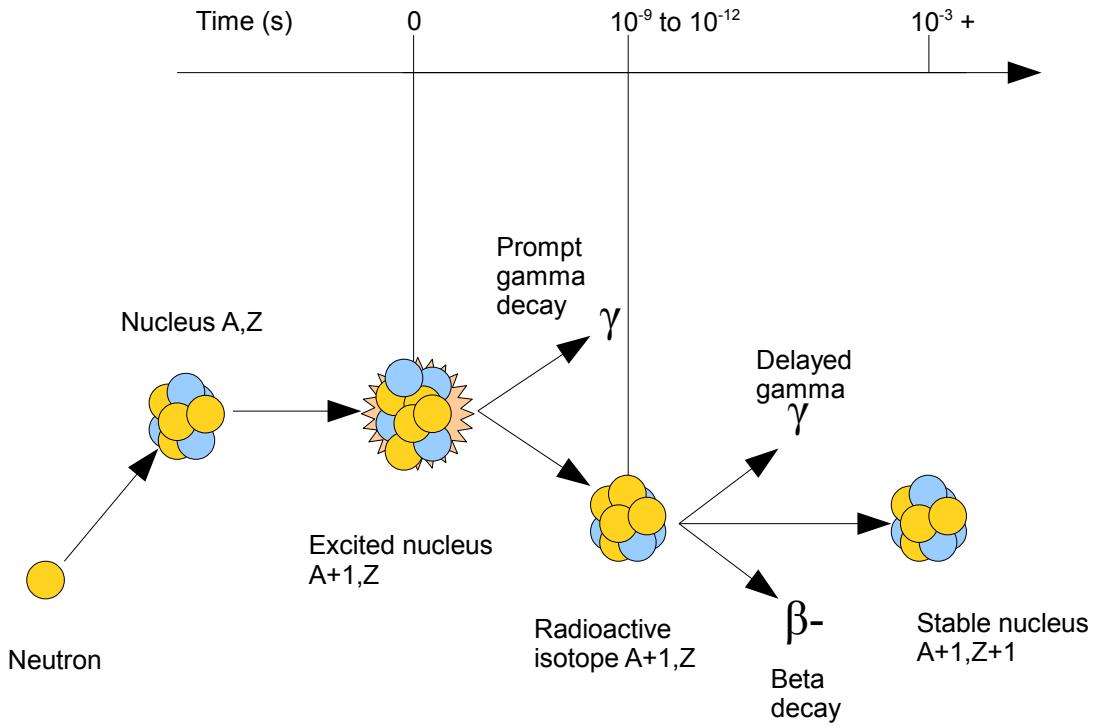

Figure 3. Drawing of the main steps in neutron activation.

There are two main types of analysis done from neutron activation. The first concerns itself with the analysis of prompt gamma decays and the second with the analysis of delayed gamma decays. The choice of which decay is best suited for a particular application depends heavily on the elements one needs to detect. As a general rule, decay spectrums for prompt gamma decays are more complex (more spectral lines) and have very short half-lives. To detect these short half-lives, it is necessary to have a neutron source that can pulsate at high frequencies. Delayed gamma activation is somewhat simpler and more widely known. The decay chains are simpler and the half-lives easier to measure and use for identification. Given the elements that are of interest and the type of neutron source we work with, we have decided to go with delayed gamma analysis (NAA or neutron activation analysis) for this project. It is important to note that much of what has been developed for this project is still directly applicable to prompt gamma neutron activation analysis (PGAA).

The probability of neutron capture depends on the absorption cross section of the target nucleus. This cross section is dependent on the kinetic energy of the neutron. We can divide the neutrons into three classes based on their energy:

Neutron classification:	Kinetic Energy:
Slow or thermal neutrons	<0.5 eV
Epithermal neutrons	0.5 eV to 0.5 MeV
Fast neutrons	>0.5 MeV

Table 1. Classification of neutrons by their kinetic energy.

Neutrons that are emitted by typical sources generally fall in the fast neutron category. As they are moderated (interact elastically with matter), neutrons lose kinetic energy. Once they have reached thermal equilibrium with the environing matter, we say that the neutrons have thermalized. The cross section for absorption is dependent on the overlap of the wave functions of the neutron and the target nucleus. As neutrons slow down, the absorption cross section grows. In addition, when there is particularly good overlap of the wave functions, there is the added effect of resonance, which boosts the cross section.

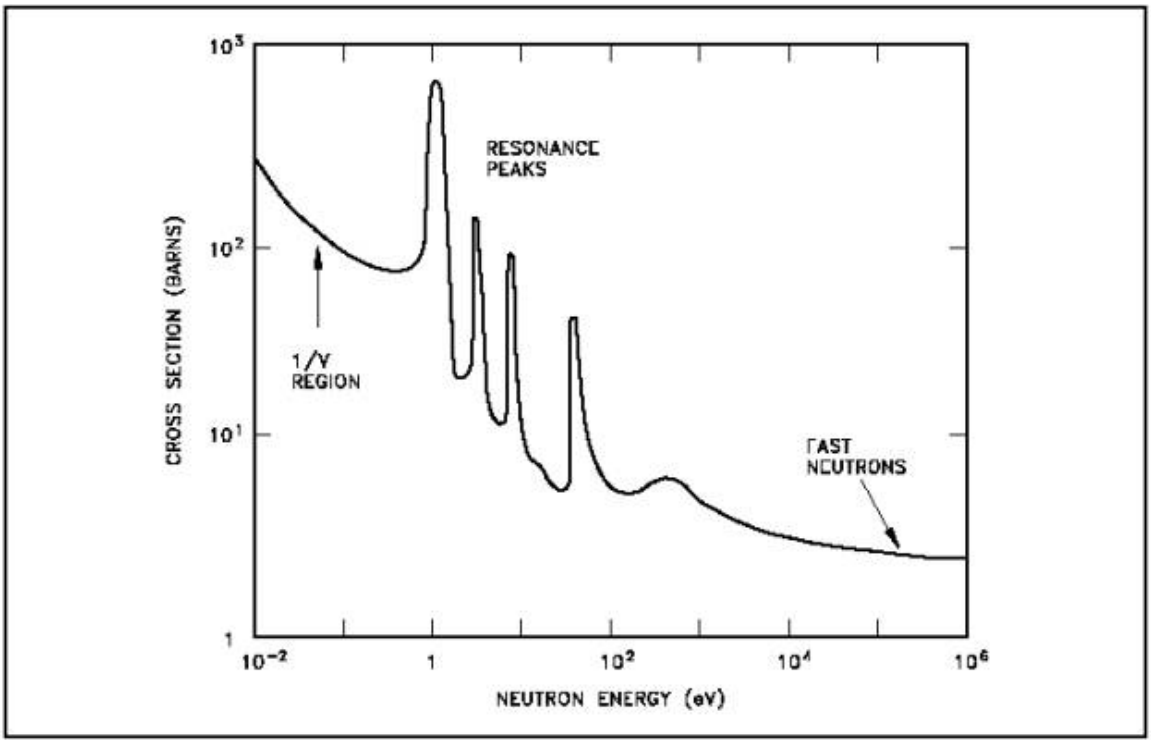

Figure 4. Typical graph of neutron capture cross section as a function of kinetic energy [39].

For these reasons, we have selected to work with thermal neutrons. For fast neutrons the elastic collision cross section dominates the absorption cross section. To thermalize neutrons from a neutron source, we can simply place some material with a high hydrogen concentration such as water or paraffin in the path of the neutrons. Hydrogen is the most efficient at thermalizing neutrons given the similarity of its mass with that of the neutron. Mass symmetric collisions maximize the transfer of momentum and kinetic energy. A few centimeters of paraffin are enough to thermalize neutrons from an Am-Be source [46].

Combining the probability of absorption with the fact that certain elements lead to a reaction chain that we can observe (activation and gamma decay), we can compute the sensitivity of NAA for certain elements at a given neutron flux. This list gives the limits of detection for a certain mass with a thermal neutron flux of 10^{13} n cm^{-2} s^{-1}:

Sensitivity (picograms)	Element
1	Dy, Eu
1–10	In, Lu, Mn
10–100	Au, Ho, Ir, Re, Sm, W
100–1000	Ag, Ar, As, Br, Cl, Co, Cs, Cu, Er, Ga, Hf, I, La, Sb, Sc, Se, Ta, Tb, Th, Tm, U, V, Yb
1000–10^4	Al, Ba, Cd, Ce, Cr, Hg, Kr, Gd, Ge, Mo, Na, Nd, Ni, Os, Pd, Rb, Rh, Ru, Sr, Te, Zn, Zr
10^4–10^5	Bi, Ca, K, Mg, P, Pt, Si, Sn, Ti, Tl, Xe, Y
10^5–10^6	F, Fe, Nb, Ne
10^7	Pb, S

Table 2. Sensitivity (in picograms) of different elements for a thermal neutron flux of 10^{13} n cm^2 s^{-1} [15]

Given the limitations on our neutron flux and detector area, we estimate that the elements in the first three or four rows will be possible to detect with our system. With a 10^6 n cm^{-2} s^{-1} source[3], the sensitivity for the fourth line (100-1000 picograms) would become 1-10 milligrams. This means that masses exceeding this limit will be detectable.

In summary, the important parameters needed to detect elements are the element ability to capture thermal neutrons and its ability to produce gamma decays once activated so that the element can be identified. The two features of this chain reaction that interest us are the energy of the decay and the

[3] The order of magnitude of the maximal estimated neutron flux (on the surface of the source) for our Am-Be neutron source. It also corresponds to the flux of a typical Am-Be source at 10 cm.

half-life of the transmutation product. This project is preoccupied with detecting these two important features and analyzing them.

1.4 Scintillation detectors

The ability to analyze samples with NAA depends heavily on the ability to detect gamma rays and their energy. The choice of detector will be discussed later in this thesis. Nonetheless, we take the time here to review the basis of gamma radiation interaction in matter. We also present the basics of scintillator detector physics.

To understand scintillation detectors or any proportional detectors for that matter, we must first understand the interaction between radiation and matter. Gamma rays and X-rays are high-energy electromagnetic radiation. Although they are photons, they behave very differently that their optically visible brethren. The range of interest is the keV to MeV region. Three interactions dominate at these energies: photoelectric absorption, Compton scattering and pair production. The cross section of each effect changes as a function of energy (figure 5):

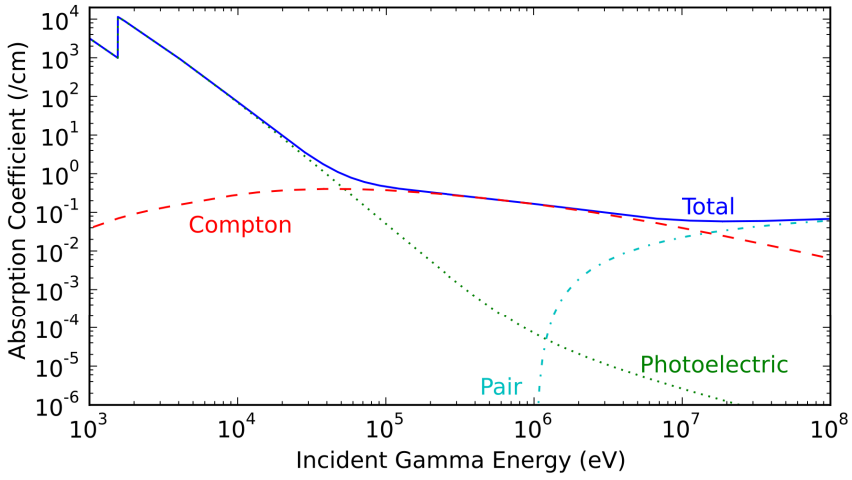

Figure 5. Gamma ray absorption coefficient as a function of energy for natural Al [41].

We notice the three distinct processes: Photoelectric, Compton and Pair production. The relative contribution of each effect gives insight to what we mean by photoelectric dominant, Compton dominant and pair production dominant regions of energy in Al (figure 6).

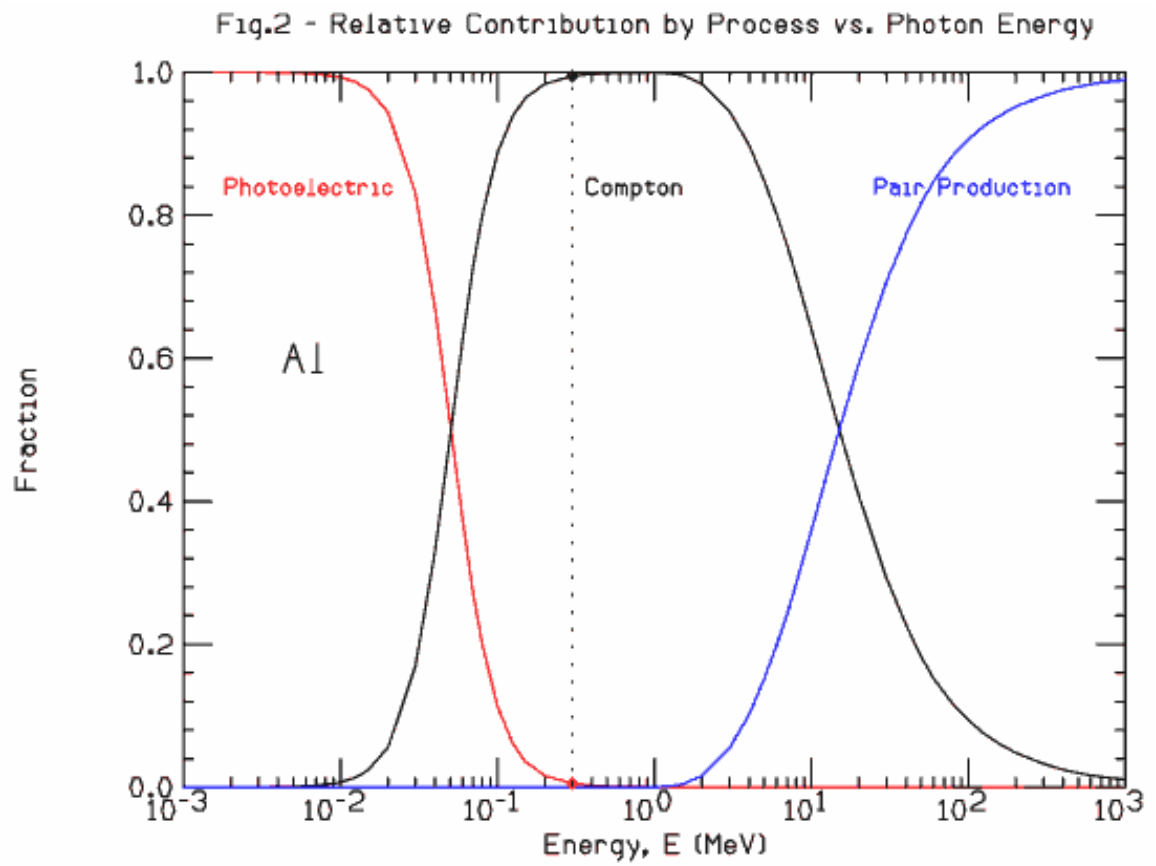

Figure 6. Relative contribution of each process: Photoelectric, Compton and Pair production as a function of gamma energy [42].

Pair production occurs when a high-energy gamma (>1.022 MeV) passes close to a nucleus and converts part (1.022 MeV) its energy to matter by creating an electron-positron pair. The remainder of its energy is conserved as the kinetic energy of the created pair. Compton scattering occurs when photons collide elastically on electrons in the electron cloud. The manner in which these collisions occur is described by the Klein-Nishina formula [17]. It suffices to conceptualize Compton scattering as a quantum billiard where the photon transfers a fraction of its energy to the electron it comes in contact with. This typically occurs repeatedly until the gamma's energy is small enough so that

photoelectric absorption occurs. The photoelectric effect (PE) is simply the absorption of the photon (and all its energy) by an electron.

The energy range of gammas we want to detect is below 3 MeV. For this range the two dominant interactions will be Compton and photoelectric. For typical scintillation detectors, a gamma entering a detector will usually go through multiple Compton scatterings and end up being absorbed via the photoelectric effect. In this scenario, the photon has transferred all its energy inside the detector volume. This type of event is what we refer to as full energy loss. It is possible for the incident gamma to undergo some scattering but to escape before losing all of its energy; this gives rise to the Compton continuum. Paths 1 and 2 in figure 7 represent these two narratives:

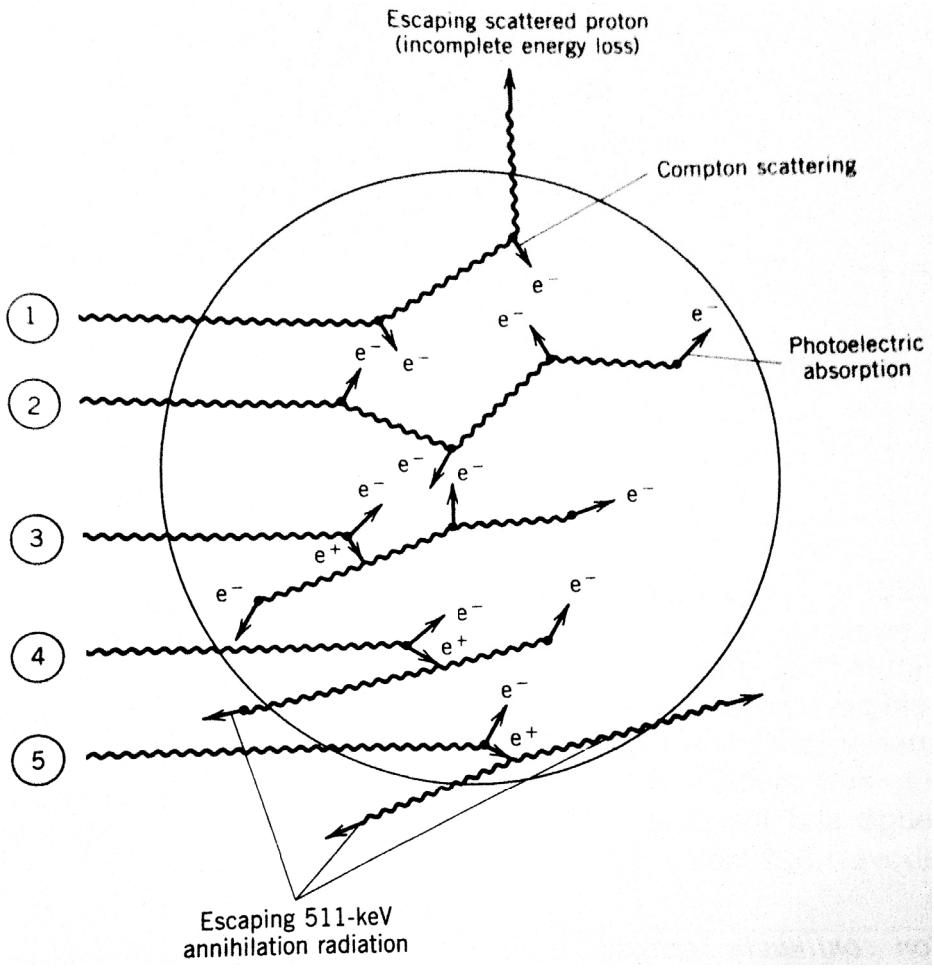

Figure 7. Drawing of the various paths a photon can take when interacting with a scintillation detector [17].

Energy transferred to electrons can be detected by the system. Any escaping photon will reduce the amount of energy deposited and will lead to partial detection of the full energy of the original gamma. Paths 3, 4 and 5 are the narratives that result from pair production where: all the energy is lost, one of the annihilation 511 keV gammas escapes and both 511 keV gammas escape respectively. Path 4 is known as a single escape and path 5 is known as a double escape. We will mainly be concerned with the two first narratives of full energy loss and partial energy loss. Some experimental setups detect escaping gammas and use an anticoincidence circuit to reduce the Compton continuum background. We believe that there is precious information in the Compton continuum so we have decided to keep it and use it in our analysis.

Incident gammas transfer their energy to the electrons of the detector material. When this material is a scintillating crystal, the electrons travel through the material transferring their newly acquired kinetic energy to other electrons via Coulomb excitation. These excited molecules then decay back to their ground state by emitting visible photons. In a pure scintillating crystal such as NaI, the photons produced are of the exact wavelength to excite other electrons in the lattice. This means that the medium is not transparent to that wavelength. Given the probability of recapture, photons travel in a random walk towards the exterior of the crystal. Energy transportation is thus slower than if the material were transparent, i.e. no recapture process. This is why NaI crystals are doped with Tl. This introduces new energy levels. A decaying electron will cascade through these additional levels creating multiple photons: one photon with slightly lower energy than the excitation energy and multiple other photons of much smaller energies.

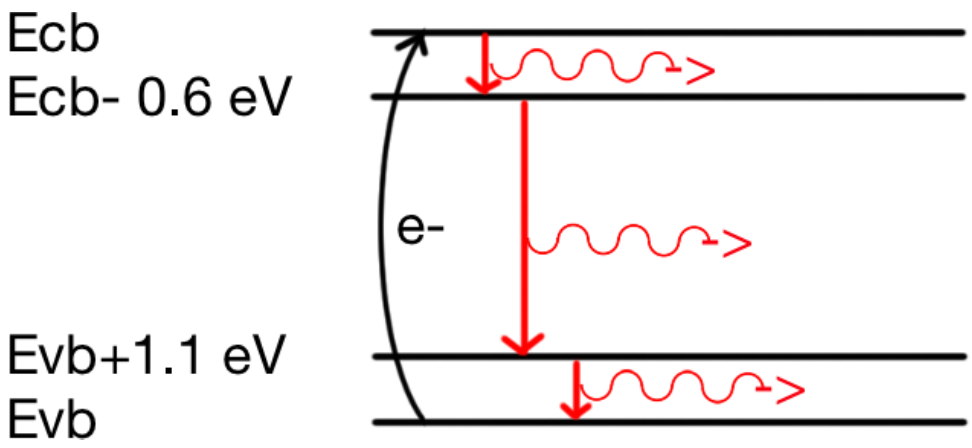

Figure 8. Energy diagram of the electrons in the crystal lattice.

The electron is initially at the valence band energy (Evb) (see figure 8). It is Coulomb excited to the conduction band energy level (Ecb). It then chain-decays through the intermediary energy levels produced by impurities (Ecb -0.6 eV and Evb + 1.1 eV). The main photon no longer has enough energy to recombine with electrons. The medium thus becomes transparent to the photons. Mirror deposits on all but one of the surfaces ensure that the light produced in the crystal finds its way to a photodetector such as a photomultiplier tube (PMT). The PMT is made of multiple layers of dynodes. A vacuum exists between each dynode and applying an electrical potential to them produces an electrical field. Incident visible photons eject electrons from the photocathode through PE. These electrons are accelerated towards the next dynode, where they eject more electrons. After each dynode, the electron flux grows. The output signal of the PMT is then sent to a pre-amplifier or an amplifier.

These electrons are proportional to the number of visible photons entering the PMT, which are in turn proportional to the amount of energy lost in the crystal by gamma rays. The decay time of NaI(Tl) and time resolution of the PMT are good enough to detect individual gamma ray events[4]. The signal pulse that exits the PMT is of the order of half a microsecond. The amplitude of the signal pulse is proportional to the transferred energy. For two identical gamma rays, the pulse height might be

[4] We discuss this when we present the model 802 detectors used in our prototype.

slightly different. The reason is that many levels of the processes are stochastic. The conversion rate of gamma energy to photon is on average 40 photons/keV for NaI(Tl) crystals. The quantum efficiency of the PMT is in the range of 30%. This means that the number of electrons produced by the PMT follows a Poisson distribution. To be clear, we are not referencing the shape of the time dependent signal form the PMT but the sum of all electrons produced for a given gamma detection. One gamma detection event will of course give a single number of electrons but if the experiment were repeated, this number would follow a Poisson distribution because of the underlying stochastic processes. This variation is directly related to the resolution of the detector, which is given in percentage at a given energy. The percentage comes from the ratio of the full width at half maximum (FWHM) divided by the energy. For a normal distribution the FWHM is given by:

$$FWHM \approx 2.35\,\sigma.$$

Typical resolutions for NaI(Tl) detectors are 8.5% at 662 keV [44]. Resolutions can also be given directly in keV at a given energy peak.

1.5 Gamma spectroscopy

Gamma spectroscopy is simply the study of the energy histogram of a given radioactive material. As described previously, one can determine the energy of individual gamma rays. A histogram is simply a representation of the frequency of detecting gammas of certain energies. A typical monochromatic gamma emitter will present a spectrum that looks like the one in figures 9 and 10:

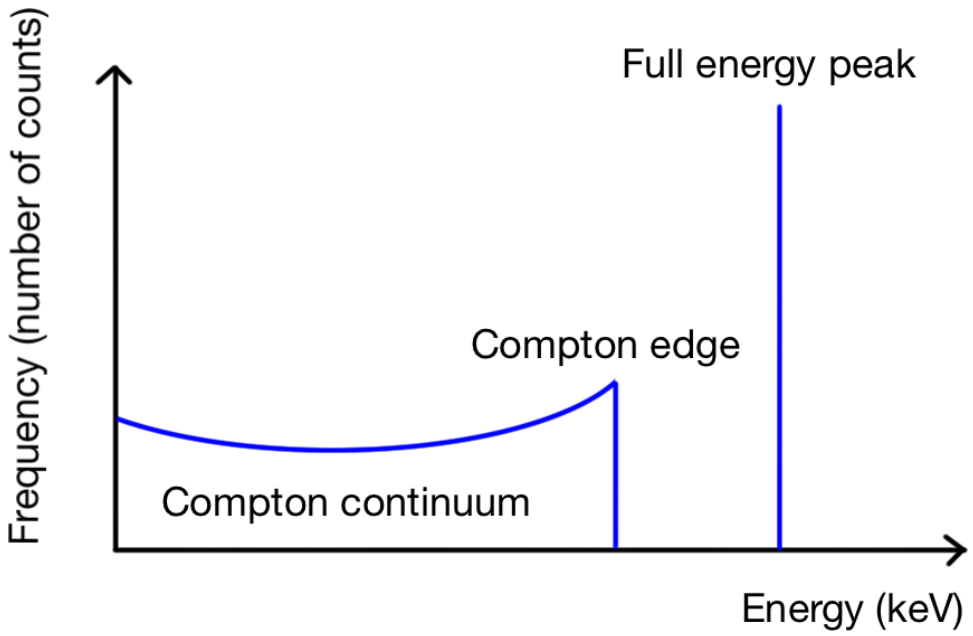

Figure 9. Schematic drawing of a spectrum for a monochromatic isotope.

The full energy peak comes from a complete loss of energy of the incident gamma to the crystal. The Compton continuum is present because of incomplete energy loss. The Compton edge corresponds to the maximum energy the incident gamma can transfer in a single head on collision with an electron. In reality, typical spectrums look like this for NaI(Tl) detectors:

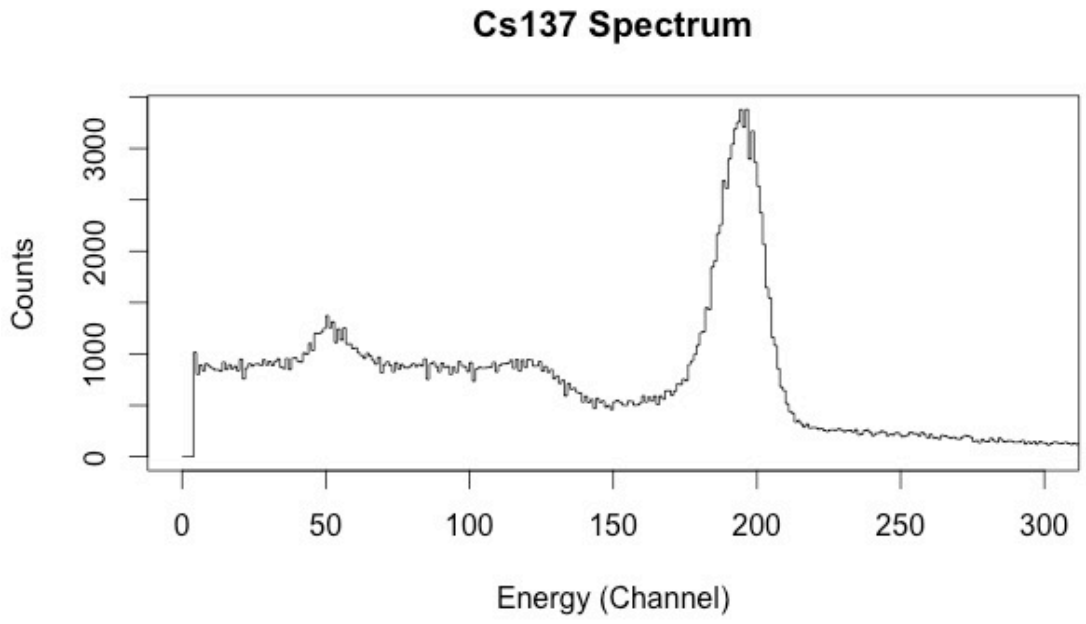

Figure 10. Real spectrum of a monochromatic isotope (^{137}Cs).

Notice the broadening of the full energy peak in figure 10, due to the resolution of the detector. Other dissimilarities between the schematic spectrum of figure 9 and the real spectrum of figure 10 are discussed at length in [1], [7] and [17]. Lastly, an important point to note is that the probability of detection decreases with energy. Low energy incident gammas have a higher photoelectric cross-section and will be easier to stop. This must be taken into account when comparing sources with different energies, especially when attempting to evaluate the relative activity of a given radioactive isotope. Depending on the method of identification, some form of calibration is necessary.

Typical spectrum analysis consists of detecting the full energy peaks and comparing their energies to a known list of emitters [7]. A rule of thumb in the field is to have a peak with at least twice the amplitude of the noise or at a certain number of times the background[5] [7]. Various fitting, filtering or smoothing methods are often used. We discuss this in our review of the state of the art in peak detection. We propose a much more robust isotope detection algorithm that utilizes both the full

[5] It can be shown that the statistical noise is proportionnal to the square root of the number of counts. Depending on the complexity of the spectrum, different methods are used to estimate statistical significance.

energy peak and the Compton continuum. In the results sections of the analysis system chapter we will show that we are able to detect seemingly invisible peaks.

Recalling how neutron activation analysis works, we consider another parameter useful to identification: the half-life. Typical setups for half-live estimation will look at the activity of the full energy peak as a function of time. Most experimental setups can only look at one parameter or the other for a given experiment [7][10]. Either they integrate over time to get sufficient data for a standard histogram, or they integrate over a window of energy and obtain a histogram of the activity as a function of time. Our system is more agile, doing both at the same time by recording events in time and energy. In addition, our analysis system will statistically infer the presence of isotopes in a two dimensional space (energy and time) using all available information, including incomplete energy loss events. In the next few chapters we will describe how our system is designed to perform such a feat.

1.6 Conclusion

We have given an overview of what neutron activation analysis is and we have introduced important notions about detector physics and gamma spectroscopy. We have seen that the properties of the spectrum are a function of the incident radiation as well as the detector. This realisation is the foundation for our research efforts into a statistical model capable of adequately combining the deterministic and random parts of the detection process as a whole.

2 Data acquisition system

2.1 Introduction

In this section, we will describe how we designed our data acquisition system. We present here the final incarnation of our prototype, which has gone through different phases of development. The approach to designing the system has been modular to enable the switching out of components for better ones. The focus of our efforts was to obtain a functional prototype, capable of delivering the data needed for our analysis system. We had to build our own system to enable us to do event recording because no commercial systems available to us could produce this type of data. We describe the choice of detectors, our electronics chain, the FPGA implementation of our control system and the graphical user interface designed to help for the acquisition.

2.2 Detectors

For gamma ray spectroscopy, there are two types of detectors that dominate the market. The first type is scintillation detectors such as NaI(Tl) and LSO, the second is semiconductor detectors such as Ge(LI) and CdTe. Various detectors were considered. The major deciding factors were: the availability of the detectors, the technical properties of these detectors and the marginal gain given the context of prototyping.

In terms of availability, we considered the cost and ease of purchase of the detectors. By far the NaI(Tl) was the cheapest as will be discussed in this section. Omitting a budget constraint, most detectors were easily acquirable.

The determining factor was the marginal gain of using one detector over another. Given our prior expertise and the availability of NaI(Tl) detectors in our lab, the gain in time for us was enormous. In the context of prototyping, choosing a standard scintillation detector will help demonstrate the gains

of the analysis method better than by using less common detectors. This is, after all, the object of this prototype.

To convince ourselves that our choice is not completely detrimental to our project we, nevertheless, investigate the technical properties of other detectors.

Both CdTe (Cadmium Telluride) and Ge(Li) (Germanium) crystals have a much better energy resolution than NaI(Tl) (Sodium Iodide). Ge(Li) has a typical resolution of 500 eV at 122 keV [48], NaI(Tl) has a resolution of 18 keV at 122 keV [44] and CdTe has a resolution of <1.5 keV at 122 keV [49]. Given the type of statistical analysis we will do, the gain from going with a semiconductor detector is not as great as for analysis by the human eye. The computer does not suffer as much from the broadening of the energy peak.

The stopping power for all three materials is similar [18]. This means that to have greater detection efficiency, we will have to optimize the volume. We also note that for CdTe and Ge(Li) detectors, there are strict limits to the size we can grow these crystals. If need be we would prioritize the use of crystals that can be grown to large volumes to maximize detection power.

The determining factor for our application is then the price per volume of detector material. In figure 11, we show a chart of the prices per mm^3 for the three detectors of interest.

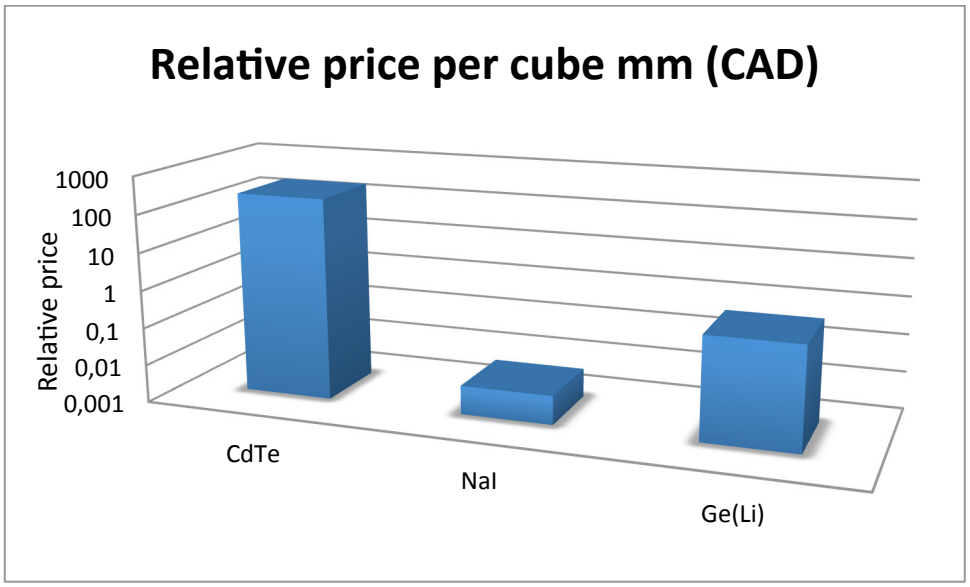

Figure 11. Volumetric prices of different detector material.

Prices include support material (Nitrogen cooling, PMT, etc.) necessary to function but not software. Prices were averages over different companies for single detectors. We notice that by far the NaI(Tl) is the cheapest solution. The Ge(Li) could have also been considered but it needs liquid nitrogen cooling to operate properly due to thermal noise being the same order of magnitude as its band gap. An interesting alternative to the NaI(Tl) would be the use of LSO. These detectors have a greater stopping power than NaI(Tl) but a slightly lower energy resolution. Given our theory that poorer resolution is not as detrimental to our analysis method as for traditional methods this could be an interesting alternative. We see that the choice of detector is very application specific. Our purpose being a functional prototype that shows the advantage of our analysis method, we somewhat arbitrarily choose the most typical detector. We are conscious about the possibility of modifying the setup to improve either resolution or stopping power should need be in a continuation of the GREYSTAR project.

The detectors we bought were model 802 scintillation detectors from Canberra. They are composed of a 2” by 2” NaI(Tl) crystal set atop a photomultiplier tube (PMT) (figure 12). They are connected to model 2007p tube base voltage dividers from the same company. The signal is sent through a 50 ohm BNC connector cable to a research amplifier (Ortec 571). The PMTs are driven at about 1.2 kV by a high voltage source (figure 13). The NaI(Tl) crystals have a typical decay constant of 230 ns [12]. The PMTs have a rise time of less than 20 ns and a fall time of 50 μs [45] The shaping constant of the amplifier is set to have a pulse width of approximately 50 microseconds (10% height). This corresponds to a shaping time constant of 10 microseconds on the amplifier. The choice of 50 microseconds was informed by our sampling frequency. It was the longest pulse width we could get with the pulse-shaping amplifier. Because of the low count rate (<100 Hz), using such a long pulse width was not likely to cause pileup. Choosing a long pulse width reduces the necessary sampling frequency. Future work should address the optimal choice of both the sampling frequency and the pulse width rather than simply choose a combination that works based on available material.

Figure 12. Canberra model 802 NaI(Tl) scintillation detector with PMT.

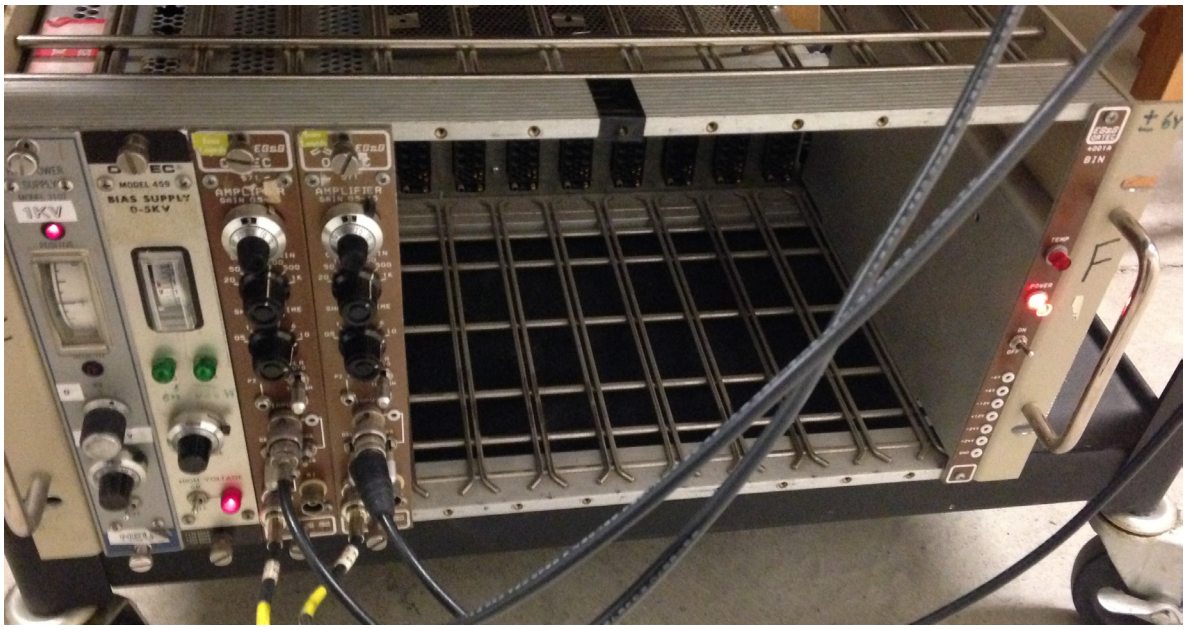

Figure 13. NIM electronics crate with two high voltage sources and two model 571 Ortec research amplifiers.

Figure 14. The acquisition and analysis experimental setup.

Our setup (figure 14) incudes two NaI(Tl) detectors with PMTs, the electronics chain, two ADCs, one FPGA and one laptop computer. We describe in further detail the ADCs and other electronics in this

chapter. The size of the chassis could be reduced and made waterproof for eventual experiments in the field.

2.3 Analog to digital converter (ADC)

Analog to digital converters are the gateway between the analog world and the digital systems we use to analyze the world. There are a number of different types of analog to digital converters. All of them are essentially voltmeters that output a binary value proportional to the input voltage. In our case we selected a successive approximation ADC, the AD7356 from Analog [20]. This model was chosen because it was a relatively cheap option that came with a test bench. We could therefore test the ADC on a functional setup before having to program our own control finite state machine (FSM). Essentially, this type of ADC compares the input voltage to a midpoint value of the input range and stores the result of comparison as Boolean value in a register. Given the outcome, the ADC compares the input voltage to successively smaller ranges of voltage. The register of Boolean bits is then interpreted as a binary number proportional to the input voltage. The resolution (in bits) of the ADC depends on the number of successive approximations made by the converter. The larger the number of approximations is, the higher the conversion time will be. There is a tradeoff between speed and resolution. In our case, 12 bits was sufficient as we are planning to use a maximum of 4096 channels[6]. We also needed a throughput rate (the time it takes to convert a voltage to a digital value and communicate that value to an integrated circuit) of more that 3 MSPS (Mega samples per second). We have around 150 samples of any given pulse. Shortening the pulse width up to half of its maximal value did not seem to affect the resolution of the system given our high sampling rate. We could have optimised the shaping constant but, again, given the low count rate, there was no great strain on the system.

[6] In the end the useful range of channels could have been lower but the analysis system was designed to work on a 4096 channel input.

For our setup, we used the AD7356 on the EVAL-AD7352/AD7356/AD7357 board [19][20]. The EVAL-CED1Z from Analog [21] was used to test the ADC for defects and then used for as a power supply for the EVAL-AD7352/AD7356/AD735 board. Control and communication with the ADC was given to a Cyclone V FPGA [5] by diverting the appropriate pins from the 96 way connector to input and output pins on the FPGA (figure 15).

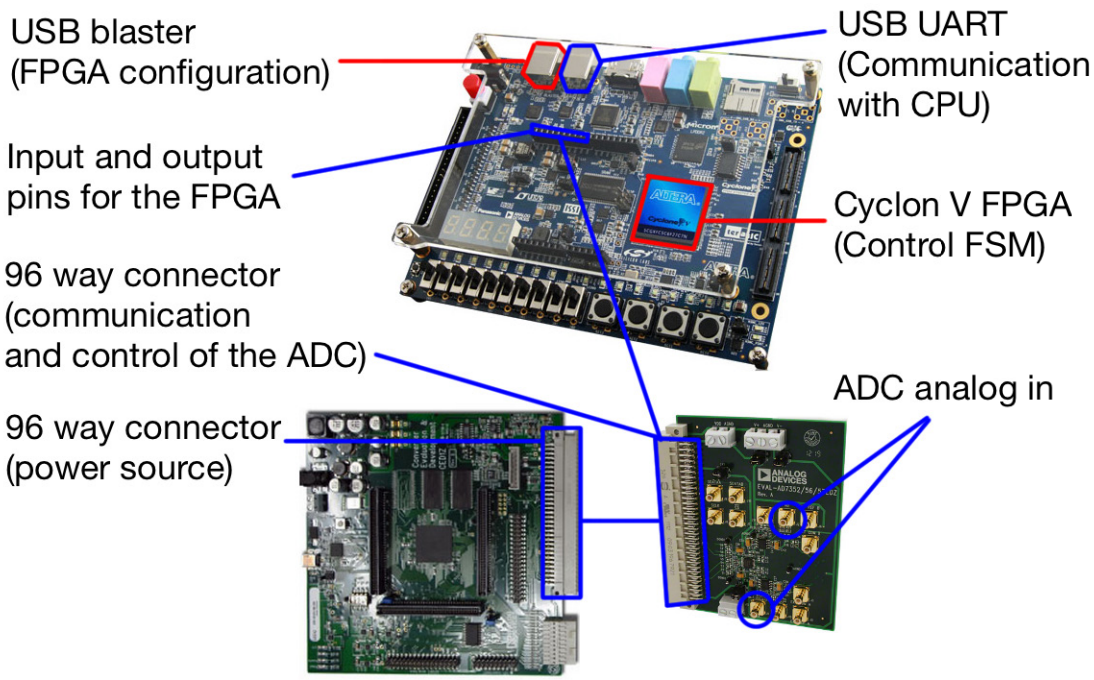

Figure 15. Cyclone V FPGA starter kit from Terasic, EVAL-AD7352/AD7356/AD7357 ADC from Analog and EVAL-CED1Z power source from Analog.

The AD7356 uses a serial communication protocol. Serial communication sends a message of a certain bit length over a single pin. The throughput rate of the ADC is limited by the message length, in our case 12 bits and the clock frequency. To transmit 12 bits of data, you need at least 12 clock cycles. In reality you need 2 extra bits for the start and stop protocol. In the case of parallel communication, each bit is transmitted over its own pin. The main advantage is that it takes one clock cycle to transmit the whole message but the drawback is each ADC would need 12 pins. In the conception of FPGAs, the real estate of communication pins is a precious commodity. Given that the

AD7356 can give us a 5 MSPS throughput rate with serial communication, we opted against using parallel communication.

The communication protocol for the ADC is fairly simple and can be found in the data sheet [19]. Simply put, the control FSM (Finite State Machine) drops the $\overline{CS}$ pin to logic zero to start communication, the SData line then changes to logic 0/1 with each rising clock edge from the SCLK pin (figure 16). The SData takes 14 clock cycles to communicate 12 bits of data. After the 14 clock cycles, the $\overline{CS}$ pin comes up for a few clock cycles to let the SData pin return to a high impedance (High-Z) state.

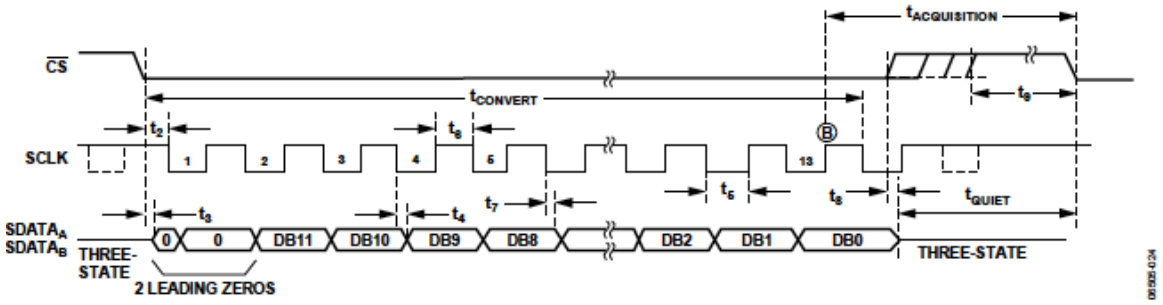

Figure 16. Timing diagram for communication with the AD7356 ADC from Analog.

It is important to remember that this ADC only evaluates a voltage at a given time, any pulse shape analysis, even simply finding the maximum, must be done on an IC (integrated circuit) or with software.

2.4 Field programmable gate array (FPGA)

Field programmable gate arrays are integrated circuits that can be reprogrammed. They are composed of logic cells that can be connected to each other to create simple to complex logic circuits. FPGAs can be programmed with a circuit diagram or a hardware description language

(HDL). Because of the complexity of our integrated circuit, we opted for the VHDL language. VHDL stands for VHSIC (Very High Speed Integrated Circuit) Hardware Description Language.

The main purpose of using a FPGA for our prototype is the low upfront cost for prototyping. One of the most important points is the ability to test IC designs without having to wait for the time it takes for a normal ASIC (Application Specific Integrated Circuit) design cycle. The other advantages that weighed in our decision are the inherent parallelism of FPGAs, their high speed computing power and their multiple input and output capacities.

For our project we selected the Cyclone V GX Starter Kit from Terasic [21]. The main reasons are the low cost of the board, the high performance of the Cyclone V FPGA from Altera, the free design suite (Quartus II web edition) and useful peripherals such as a USB-UART emulator chip (Universal Asynchronous Receiver/Transmitter), USB-Blaster, phase locked loops (PLL) to deliver a 50 MHz clock to the FPGA and memory blocks to store the configuration file or bit stream.

2.5 The control FSM

Using VHDL we designed a control FSM and components to be implemented on the Cyclone V FPGA. The block diagram for our code can be found in the annex. The purpose of this circuit is to control and communicate with the ADC, extract useful information from the high frequency sampling of the scintillation crystal signal. The system then distils the information into useful low frequency data that is packaged and communicated to a CPU (Laptop computer) via a UART communication protocol emulated over a USB cable.

The first component is the ADC controller. It ensures synchronous communication and control with the ADC. It is a small state machine that sends a read signal to the ADC by dropping the $\overline{CS}$ line and then records the serial data into a register. It removes leading and trailing zeroes from the data. Once the data is clean, the controller updates a second register with the 12 bits of data[7]. The purpose of

[7] Each ADC has his own secondary register. Both ADCs are read in parallel and synchronously.

this controller is to prevent the reading of partial data. To the outside world this controller updates a full 12 bits of good data every 14 or so clock cycles.

The second component is the peak estimation FSM. Again this is a small state machine that looks at the ADC controller register periodically (at the same frequency as the register is updated). The FSM compares the 12 bit binary value of the register to a threshold digital value. When the threshold is crossed, a register records the highest value read from that point on until the threshold is no longer crossed. Once this happens, the maximum value is passed on to another register and a flag is set to one. As long as the flag is not reset to zero, the peak estimation is frozen in that state. For our needs, we found, through testing, that using the maximum value was sufficient. We are of course conscious of various improvements that could be made to this module such as pulse shape analysis. Given the scope of the project, we elected to leave this as it is and revisit it at a later date.

The control FSM incorporates one ADC controller and two peak estimation machines (PEM). The control FSM tells the ADC controller to sample at a given frequency and links the PEMs to the two data lines. Each PEM separately waits to catch the maximum value of an event coming from its own detector. Once the PEM finishes and flags itself, the control FSM copies the data from the PEM register to one of its own registers. It then resets the PEM's flag to zero. If the UART is not busy, it packages the 12 bits of data into three 8-bit packets and sends it to the transmitter of the UART component. The first two bits are to send the data from the register and the third bit is used to indicate from which detector it comes. This part of the control FSM can be started and stopped by sending an 8-bit value through the receiver part of the UART. The FSM continually checks the receiver and updates an internal register. As long as this register has the correct 8-bit value or key, the control FSM continues working.

The last component is the UART. The FPGA board is equipped with a USB-UART emulator chip. We communicate with the laptop through this port using the UART protocol. The UART is composed of two state machines: the transmitter and the receiver. The UART protocol is simple: the idle state of the pin is held at logic high (1). To start a serial transmission, the communication pin is taken to logic low (0) for one slow clock cycle (start bit). For the following 8 cycles, the pin is forced to the value of the bit (0 or 1). At the end of these 8 cycles, the pin is taken to logic high for the stop bit (figure 17).

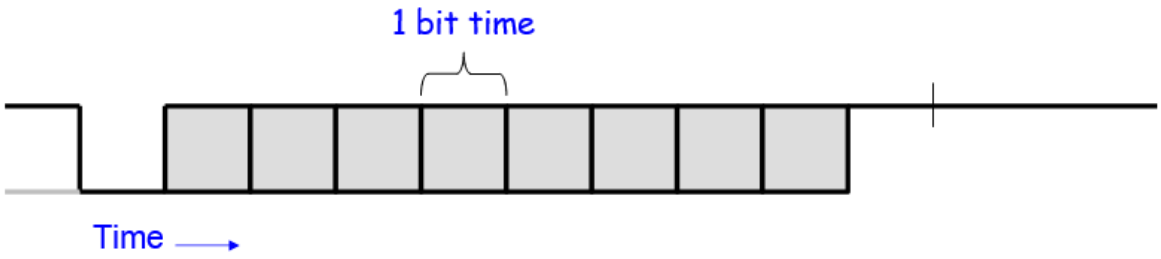

Figure 17. Drawing of a 8-bit UART communication protocol with no parity bit [3].

The baud rate (speed at which data is transmitted per second) is dependent on the slow clock period. Our system uses a baud rate of 115 200 bits per second. Which means that our slow clock is about 54 times slower than our 50 MHz internal clock.

The receiver FSM samples the receiver pin at 50 MHz until it goes to low. The FSM waits for half a slow clock cycle to check if the value is still low then proceeds to sample the middle of each bit at the frequency of the slow clock. Once it has received all the data, it updates a register for the control FSM to see.

The transmitter FSM packages the data by adding a zero to the front of the 8 bits and a one at the end. It then forces the transmitter pin to high or low for the appropriate amount of time to transmit the data with the correct baud rate.

In the VHDL source code, there exists an additional module that is not used. This module's job was to analyze the pulse shape to detect pileup events. This module was going to use random tree classification to discriminate between good and bad signals. It is further discussed in the annex. It was removed from the final code due to the insignificant amount of pileup our system will encounter. The whole purpose of this system is to extrapolate the maximum information from low activity spectrums. Because of the low count rate, the probability of pileup events is negligible. For a count rate of 100 Hz, a pulse width of 50 microseconds, a back of the envelope calculation gives the probability of two events being piled up as being approximately 0.5%. We were able to witness this visually with the EVAL-CED1Z board, which records the pulse shape over a given period of time.

2.6 C# data logger and Graphical user interface (GUI)

The C# language was chosen for two reasons: it is a compiled language similar to C++ and it has a plethora of integrated functionalities that help for communication and GUI design. It is a pseudo parallel language, which means that it is easier than ordinary C or C++ to have multiple event handling and method or function calling.

The data logging part of the program is straightforward, it reads the serial port buffer and reconstructs a 12-bit value and the detector it comes from. It opens a .txt file and writes the 12-bit value in decimal, the detector number and the current time (in microseconds) since the beginning of the acquisition. It also updates a histogram vector, which record the number of times an event of certain amplitude has occurred since the beginning of the acquisition.

The code has an error recognition algorithm to prevent corrupted data from being recorded. The system reads the 3 bytes of data knowing what is expected to be in each. Knowing that the first byte contains a 1 or a 2 (detector number) and that the third byte is limited to the 4 bits remaining from the 12 bit code sprit into 8 bits in the second byte and 4 bits in the third byte, the system checks if the values in each byte is in the expected range. If this is not the case, the buffer is cleared, the triggering data is not recorded and an error message is sent to a text box of the graphical user interface to let the user know a problem has occurred. The logger then resumes recording clean data as it arrives.

The GUI enables the user to start and stop the control FSM on the FPGA by sending the correct 8-bit key through the USB-UART. The user can also clear the histograms and .txt files from the GUI (figure 18).

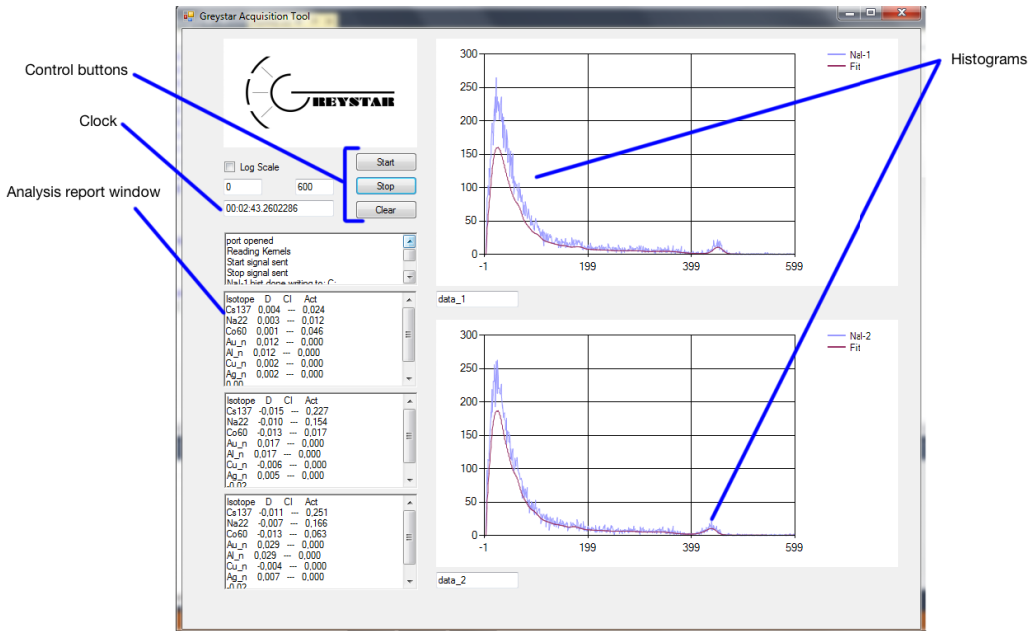

Figure 18. Screen shot of the GREYSTAR Acquisition Tool graphical user interface.

As the data comes in, the histogram vector is updated. Every 500 ms, the histogram plots are updated. Every 2 minutes a call is made to the analysis method, which will be discussed in the next section. Once this call returns its fitted values, the histogram is updated with the regression line superimposed on the histogram data. Results from the statistical analysis are sent to the appropriate text boxes for the user to interpret.

2.7 Results

The user interface code is stable over extended periods of use (weeks), thanks in great part to the choice of the language. The control of the FPGA is seamless and reliable. The spectrums exhibit similar count rates to spectrums from commercial systems, at least in a low count regime. Here are some spectrums obtained from our system compared to traditional systems (figures 19 and 20).

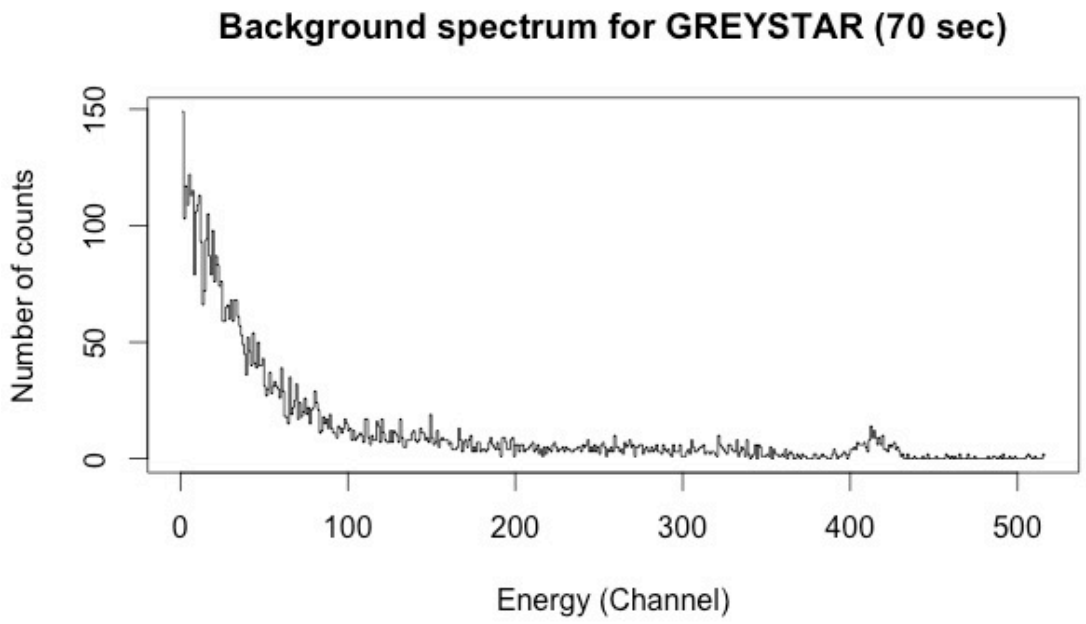

Figure 19. Background spectrum obtained with a Canberra model 802 scintillator coupled with the GREYSTAR system.

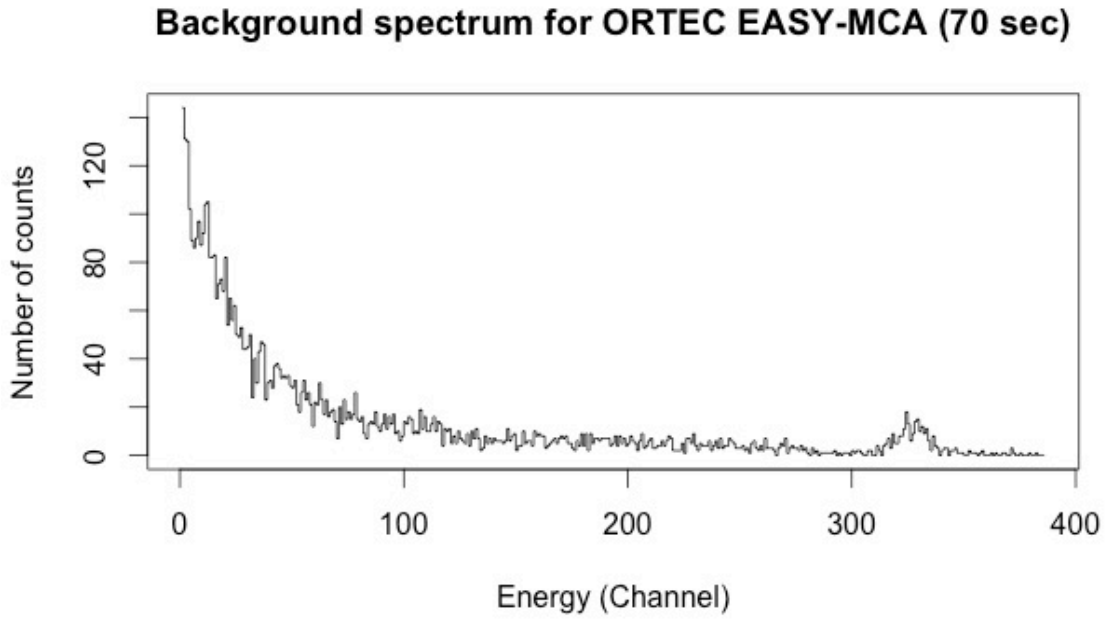

Figure 20. Background spectrum obtained with a Canberra model 802 scintillator coupled with the EASY-MCA system.

Both the GREYSTAR system and the ORTEC EASY-MCA system have comparable count rates for background spectrums. This indicates that when it comes to count rates, at least for low count rates, our system is at least on par with commercially available hardware and software. Both acquisitions were done with the same detector and electronics chain and for the same period of time. We remark that the EASY-MCA does have additional features such as gain selection from the GUI and a more complete software suite. We had no need of testing at high-count rates, given the desired application of our prototype. The purpose of this test was to evaluate if the latency of our system was affecting our count rate in low activity settings. We take these results as an indication that our system is no worse than the commercially available system in the context of our desired application.

The following figures (21, 22, 23 and 24) show the same data in a scatter plot and two dimensional density spectrums. We show different smoothing parameters so that the reader can better visualize the data. The fading line around channel 200 was obtained by varying the distance of a ^{137}Cs source from the detector as time changes.

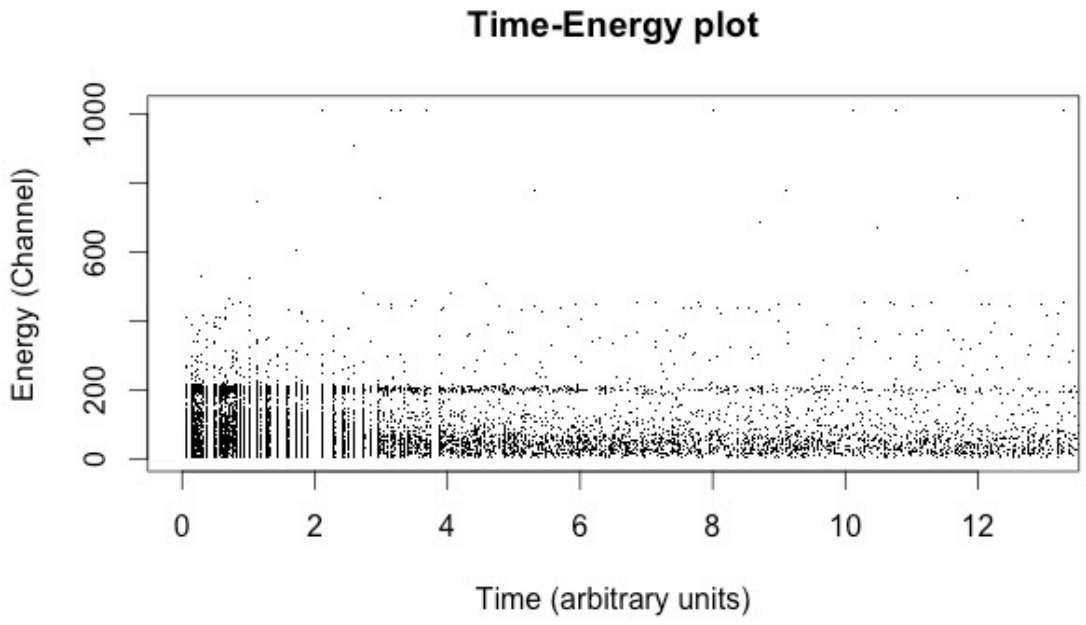

Figure 21. Time-Energy plot obtained from event recording with the GREYSTAR system.

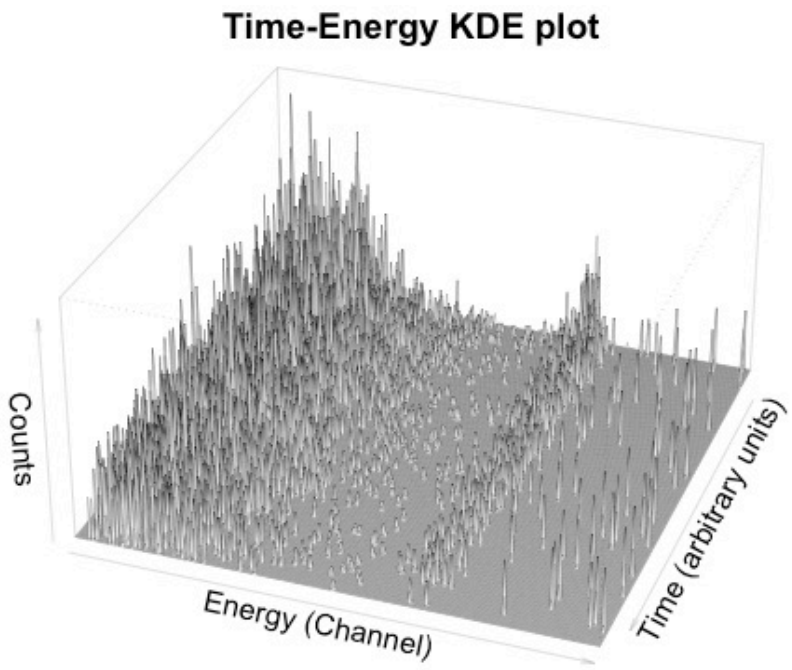

Figure 22. Kernel density estimation[9] of the time-energy plot with a bandwidth of (0.5, 5).

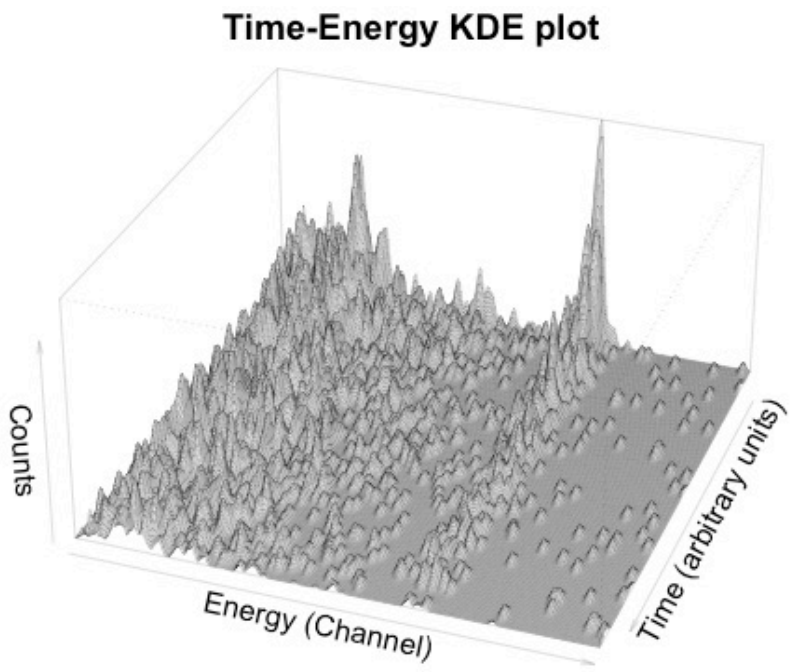

Figure 23. Kernel density estimation of the time-energy plot with a bandwidth of (1, 10).

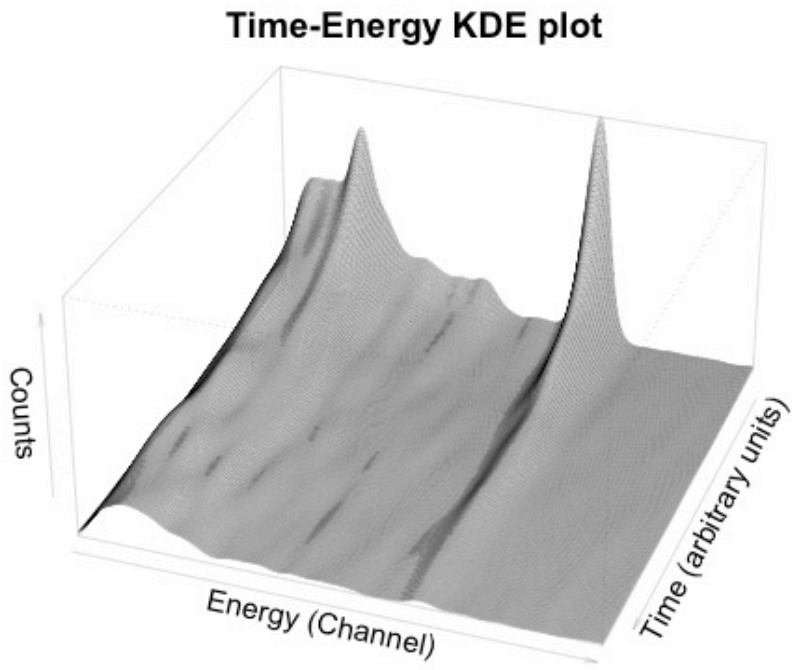

Figure 24. Kernel density estimation of the time-energy plot with a bandwidth of (3, 19).

The time-energy plots show that our acquisition system works for event recording. A plot for raw data is given (figure 21) and a 3D plot of the density is given in figures 22, 23 and 24. We have performed 2D kernel density estimations (KDE) with various kernel widths to help the reader see the evolution of the spectrum in time. KDE is presented in more detail in the next section. We use the CRAN library that comes standard with the programming language R for our implementation of the KDE algorithm. The kernel bandwidths are given in parenthesis.

2.8 Conclusion

We have demonstrated that our experimental setup is capable of acquiring useful data and present the data in the format required by our statistical analysis method. Using available modern electronics, we have constructed an experimental setup, which is more flexible than standard neutron activation spectroscopy setups. The flexibility of our setup comes from the fact that we can do both types of

analysis (energy and half-life) a posteriori. The type of analysis does not impact the experimental setup as it might on other experiments. We have made an effort to use a modular approach for the conception of this part of the prototype. It is thus very simple to improve or modify the setup by, for example, trading out the NaI(Tl) detectors for either higher resolution semiconductor detectors or higher stopping power LSO detectors.

3 State of the art in peak identification

3.1 Introduction

This chapter will review from a conceptual standpoint the past and present efforts done in peak identification and automated isotope detection. Reviews of commercially available software have been done and continue to be done [50]. We believe our proposed method differs substantially in its approach as to warrant explaining where it cuts off from traditional spectrum analysis.

The first thing to do is to define the question that any method tries to answer: For a given spectrum of acquired data, what isotopes can we say are present. This question might seem trivial but it needs to be stated to avoid falling into certain caveats. The second thing we invite the reader to do is imagine a line that goes from completely non-parametric, i.e. almost no assumptions about the underlying process to completely parametric, i.e. all is known or parameterized. We will explore from a high level various existing approaches and state where they fall on the parameterization line and how they try and answer the initial question.

3.2 The non parametric side

Starting on the non parametric side, we note the use of classification techniques for isotope identification [32][51]. In [51], the authors explore using both support vector machines and artificial neural nets [22] before testing the Mahalanobis distance [52] and variance in angle spectrum [53] techniques. The first two classification techniques are almost fully non parametric but the third and fourth make some assumptions about the distribution of the spectrums. The authors end up using one of the more parametric techniques as it yields better results.

Coming back to the initial question of determining the presence of isotopes, this method answers the question: is there anything else than background? Once this question has been answered, the

spectrum is then passed on to a human to do the spectrum analysis. Our belief is that, although very interesting from a machine learning point of view, using fully non-parametric or adaptive methods is sub-optimal as we are not using known properties analysed data. In addition, we believe that the inclusion of parameters can be done intelligently given careful considerations to the underlying physics and statistics.

3.3 The parametric side

Next we consider the family of traditional peak identification. At this end of the spectrum, parameterization and some times over parameterization is king. This approach is probably the most familiar to physicists and engineers. There is a plethora of different techniques which add small twists and turns to a common narrative [54][56][57][58][59]. The general idea is to detect a peak using a peak identification method, estimate the energy of said peak and compare that to a database of known isotopes.

This mindset comes from the knowledge that certain isotopes decay with certain properties. One of these properties is the energy of the emitted gamma. Knowing this, physicists and engineers developed detectors to measure this value. Unfortunately, the detectors have limitations so we try to estimate the best we can where the photopeak is. Looking at this process in terms of answering the primary question, what we are doing is splitting the question in two. The first part is concentrated on finding where the peaks are and the second part tries to tie the position of these peaks to the presence of isotopes. The second part of the method is seldom discussed in papers tackling the first part of the problem.

3.4 Asking the right question

We will bring attention to an important point about the conditionality of spectrum analysis: given the presence of an isotope emitting gammas of a certain energy, the probability of detecting a peak at said energy is strictly a function of the experimental setup, inversely, the probability of the existence of an isotope given the detection of a peak at that energy is also a function of all possible isotopes and the rest of the spectrum. Because we are answering the question what is the probability of an isotope being present given the acquired spectrum, we believe that some information is lost with the traditional peak identification paradigm. The probability we are estimating is dependent on the entire spectrum and this spectrum is, in some form, reduced before it is handed to the analysis method. Nevertheless, we will go through the typical approaches shortcoming and discuss what lead us to search for a better solution for our own system.

3.5 Shortcomings of traditional methods

The first problem that comes up with peak identification is the channel-to-channel variation known a statistical noise. This prevents us from simply finding local maxima and using their position as an estimate of the energy of the peak. Filtering is the most common tool to overcome this problematic. Filters range form simple moving averages, weighted averages, non-parametric kernel density estimation, wavelet smoothing, orthogonal function smoothing, etc. [9][58] For all their apparent complexity, they all essentially do the same thing: they reduce channel-to-channel variation while leaving the underlying trend relatively unchanged.

We note that when we decide to smooth a spectrum we are making a decent but imperfect assumption: the noise from channel to channel is high frequency compared to the underlying features and it is something we want to get rid of. Filtering techniques work because generally our signal is spread out over many channels (low frequency) and the statistical noise is channel to channel (high frequency). This might not hold for different detector resolutions, for example Ge(Li) detectors will

suffer more from over smoothing because of the proximity, in the frequency space, of the features and the noise.

Furthermore, no explicit effort is made to distinguish between multiple channels with a large number of counts and a single channel with a very large number of counts. Simply put: smoothing is averaging and averaging increases informational entropy. Still, in this paradigm, the handoff between loosing a bit of information and identifying the peak is worthwhile. Once we have an estimate of the position of the peak, all that is left is to estimate its relative strength. One can either simply take the height of the smoothed spectrum at a given point or, usually, one will fit a function (generally Gaussian) on that spot and estimate the area under the curve. Sometimes, more exotic functions are used [24] but the general idea never changes.

The second problem that comes along with peak identification is peak resolution. What happens when two peaks overlap? One approach taken by the community is to develop detectors with better resolutions. A different approach, the one that mathematically resembles the most our proposed approach, is peak deconvolution [55]. It seeks to artificially enhance the resolution of detectors by essentially fitting a large number of Gaussian functions and estimating their amplitude. The spectrum is then analyzed in the parameter space where peak resolutions are better. There are a few drawbacks to this method. The first is that if one uses OLS (Ordinary Least Square) to estimate the amplitude space, one generally needs a well-defined function such as the normal distribution. For low counts, one must abandon OLS for MLE techniques because of a low signal to noise ratio and the fact that some of the assumptions for OLS do not hold [9][60]. The interpretability of the parameter space in terms of signal to noise is not straightforward and good care must be taken to avoid type 2 errors (false positives). In addition, most detectors do not hold a constant resolution as energy varies. This means that the function chosen for deconvolution might not hold over the whole spectrum. No assumptions are made on the position of possible peaks, which means that we end up searching impossible combinations of peaks. This can make the analysis prohibitively long. In fact this applies to all peak search paradigms. Because of the split between peak identification and analysis, the identification process is blind to known constraints such as the universe of possible peak locations and peak widths.

The third problem to face these methods is background attribution. When trying to determine the amplitude of a signal, these methods must be able to parse between background contributions and

isotope contributions. The typical approach is background subtraction, where the background is known and a smoothed version of it is subtracted from the spectrum prior to smoothing, peak search or deconvolution. This can lead to negative counts in some channels. This is not all that problematic for high intensity peaks but can lead to difficult situations when the number of counts in the photopeak is small.

The fourth problem to face both deconvolution and simple peak search is the detection of spectrum artefacts such as the Compton edge, backscatter, single escape peak, etc. This means that an automated detection scheme might misinterpret these artefacts as photopeaks. The use of more complex fitted functions or stripping [24] can be useful but lead to the same problems as background subtraction. We note that there exist Compton background suppression setups to combat this problem. We believe that this problem is only a problem within the peak-searching paradigm. If we take a step back and think at why we are looking for photopeaks, we realize that we have unwittingly made a decent yet imperfect approximation. We presume that all the useful information is found in the photopeak because it is easily comparable to our second step of analysis. In reality, the Compton background and other artefacts are dependent on the incident gamma energy. Their detection holds some information even if it is not as valuable as that of the peak. Discarding non photopeak features only makes sense in our limited view of the problem. Machine learning approaches operate on the same instinct. The algorithms look for features in all of the data, not only the photopeak. We simply state that we know how these features interact to produce the observed spectrum, from our understanding of the underlying physics therefore we constrain the problem to one where no training is needed.

The fifth hurdle for peak searching methods, and perhaps the most damning is how does peak searching work when no peaks are present. We know that as gammas travel through matter, they will scatter and the resulting spectrum will morph. At one point, virtually none of the photopeaks will be detectable but some radiation will still get to the detector. We need a method that can be adapted to peak-less spectrum identification. By definition traditional peak identification techniques are inadequate here.

3.6 Our proposed solution

Our technique, which will be described in the next section, will aim to answer the question of which isotopes are present, given the spectrum, directly. We essentially construct a typical statistical test but let our physical knowledge of the problem at hand guide our intuition as to what approximations to make and how to parameterize the problem. In the end, our method can be interpreted as: a mixture model problem, a highly parameterized smoothing method, a special case of deconvolution, a multiple hypothesis test or a multivariate linear regression problem. It is different from a peak searching algorithm or an ordinary classification method. The hypotheses and approximations we make are physically and statistically sound given the problem we are trying to solve.

The inclusion of time dependent data in our analysis is also considered. We are not the first to suggest using an additional dimension to help in isotope identification [7]. Smoothing and deconvolution methods exist for two-dimensional data but the curse of dimensionality has the effect of reducing the density of counts. This is equivalent to having a low signal to noise ratio and puts an additional strain on the smoothing methods necessary for peak search algorithms to work properly. Usually, the spectrum is integrated over time to have a sufficient statistic for peak identification. A window around that peak is then selected and integration over these channels is done while the time dependency is kept to evaluate the half-life. One or more exponential decay functions are fitted to the data to detect the presence of multiple overlapping photopeaks. Again, it is assumed that most of the useful information is found in the photopeak. No analysis of the decay of Compton background is generally done and no combinatory analysis with other peaks is done at the identification level. For low count rates, background subtraction and smoothing hinder exponential curve fitting. What we believe is novel in our approach is that by design we avoid many of the unnecessary approximations of other methods and avoid many caveats that affect these methods.

From a high level perspective, our method starts out by determining the universe of possible spectra that can be combined to make the observed spectrum. We then approach our observed spectrum as a mixture model. Estimating the amplitude of each base spectrum is simply a form of multivariate linear regression. Using the likelihood as the fitness function of our regression forces us to determine a parametric form for the statistical noise from one channel to another but also makes our estimates

more robust to large differences in count rates. Carrying the likelihood throughout the process, in the end we can statistically test different hypotheses, such as the presence of a given isotope. We believe that analyzing directly all of the raw data avoids unnecessary filtering and voluntary blindness to inter-spectrum dependencies.

Finally, to compare our method to existing methods, rather than tryout all available systems, we will test out system on a monochromatic spectrum. We will calculate the time our system takes to identify an isotope to a given confidence interval. We will then calculate the time it takes for the area under the peak (or amplitude of the peak), excluding background, to be n times the standard deviation of the surrounding statistical noise, where n is equal to the number of standard deviation needed to for a p-value equal to the confidence interval. We have a good method of estimating the standard deviation of the noise, knowing that the number of counts in a channel follows a Poisson distribution; we also know that the standard deviation is simply the square root of the average number of counts. We can thus take an average of the background and estimate the standard deviation. Although some existing methods might outperform this method in their own right, we believe that it is a standard measure of statistical significance in the field. This approach to comparing is generous to traditional peak search algorithms, as it does not require finding the peak, which for some algorithms would be impossible given the low number of counts. We believe this is an adequate and conservative method of comparing the benefits of our proposed approach.

3.7 Conclusion

We have demonstrated that our approach is structurally different from traditional approaches used in spectrum analysis. What makes our proposed method different from others is not necessarily any major breakthrough in statistical estimation but simply an adequate description of the problem from a physical and statistical point of view, combined with a judicious choice of statistical models and methods. In the next section we will describe how we arrived at this model.

4 Analysis system

4.1 Introduction

In this section, we will build our statistical model from the ground up. We start by reviewing the concepts of likelihood and curve fitting. Through our effort to understand the statistical properties of the experimental data, we discover that an adequate method for extracting physical observables from the data ends up being a type of mixture model. We then construct what we call kernels, which are reference spectrums for known isotopes. We use the name kernel because we renormalize the integral of the spectrum to one. We then use our new model and construct an analysis algorithm to extract estimates of physical observables such as the relative activity of a given isotope.

4.2 Statistical model

To analyze the data we developed our own statistical model. The premise is to have a set of basis functions that correspond to the spectrum of different radioactive isotopes. We then find the best linear combinations of these basis functions. To evaluate the goodness of fit, we compute the combined log likelihood of the fitted function with respect to the experimental data.

To compute the likelihood of a fit with respect to the histogram we begin by noting that the distribution of counts in a single channel follows a binomial distribution. In our case, this is interpreted as: every time a gamma ray is detected there is a probability p that its energy will be inside a given range. Knowing p and the total number of events n, the number of events inside said range k will follow the binomial distribution:

$$\binom{n}{k} p^k (1-p)^{n-k}. \tag{3}$$

Unfortunately, only nature knows the true value of p. Our job is to estimate p with $\hat{p}$. We start by presuming that p can be any value in the interval [0,1]. We can then compute the likelihood of p as a function of k and n. The likelihood is defined as:

$$L(\theta|x) = P(x|\theta), \tag{4}$$

where $L(\theta|x)$ is the likelihood of parameter values θ given the outcome x. In the case of a simple binomial distribution, p is the only parameter in θ and x is the number of counts k. The likelihood is then:

$$L(p|k) = P(k|p) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{5}$$

Notice that it is the same function as the binomial distribution. The trick here is that in one case the free parameter is k and the fixed parameter is p. This gives us the PDF (probability density function) of k. In the other case, the one that interests us, we have k as the fixed parameter and p as the free variable. This gives us a kind of PDF for p. We can interpret this function as the probability that the observed outcome (k counts given n trials) comes from a situation where p was the probability of success. If we chose an estimator $\hat{p}$ that maximizes the likelihood function we get the MLE or maximum likelihood estimator. In our simple case this would trivially be $\hat{p} = \frac{k}{n}$. For the sake of rigorousness we underline that we have taken a shortcut in giving the likelihood function. The complete form of the likelihood function is:

$$L(p|k) = \frac{P(k|p)\, L(p)}{P(k)}. \tag{6}$$

Knowing that $P(k) = 1$ by definition and that the prior has been postulated as $L(p) = 1$, we get Eq. (4).

The previous example gives the model for a single channel. In the case of spectroscopy, we have a histogram of multiple channels. Thankfully, we can get the combined likelihood of multiple channels by multiplying the likelihood functions and renormalizing. This is equivalent to combining i independent experiments together:

$$L_{total}(\theta|x) \sim \prod L_i(\theta_i|x_i). \tag{7}$$

If no conditions are set on the estimated function (regression function) $\hat{p}(i)$ where i is the channel number then:

$$\hat{p}(i) = \frac{k_i}{n_i}. \tag{8}$$

This obviously leads to over fitting and is quite useless.

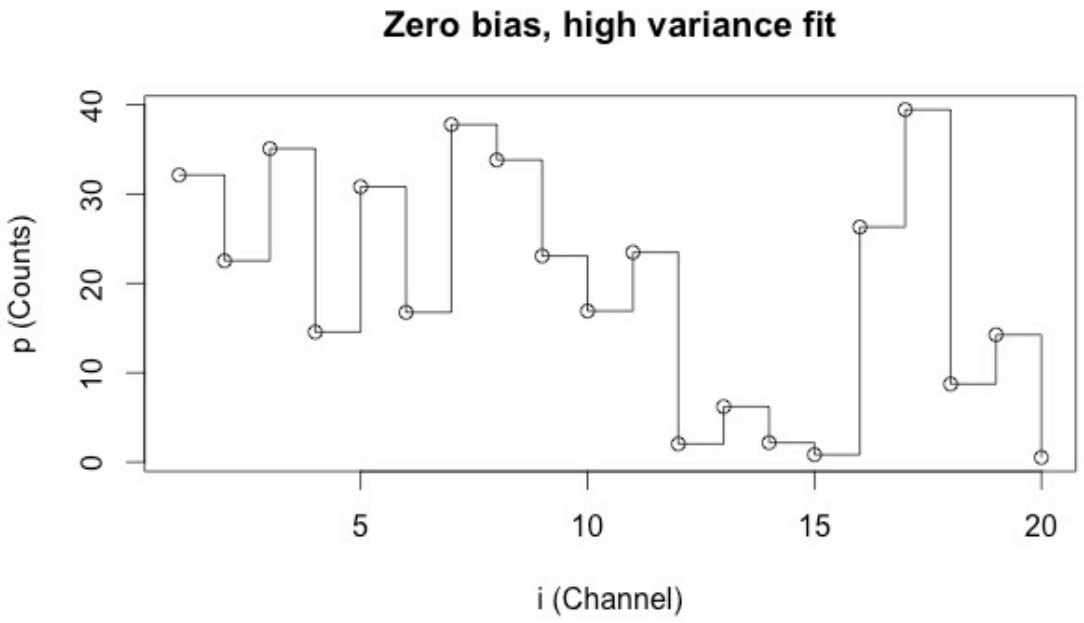

Figure 25. When no conditions are imposed, the fit will trivially pass through each point.

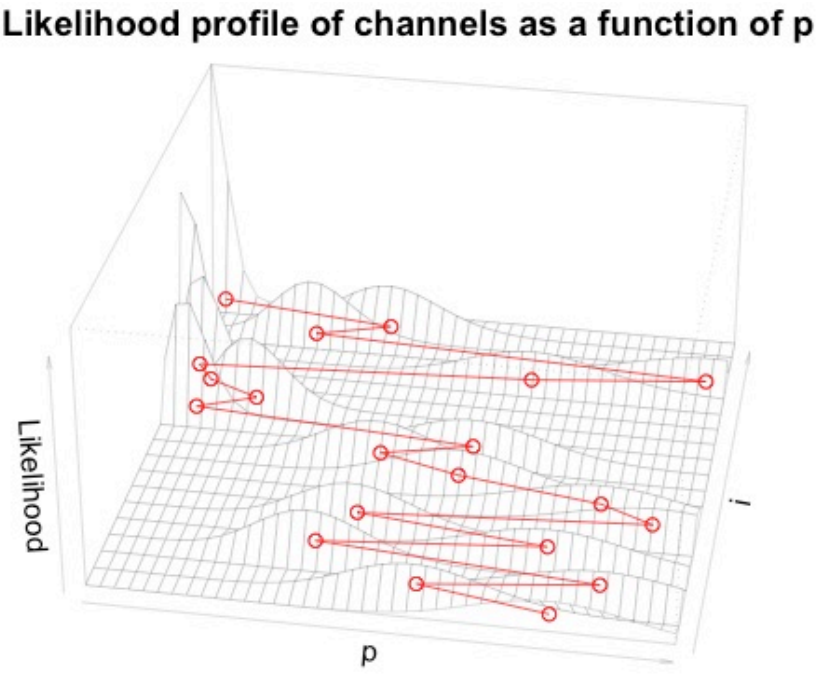

Figure 26. The trivial MLE fit with no constraints on $\hat{\boldsymbol{p}}(\boldsymbol{i})$.

In the case of figures 25 and 26, the regression line passes through the MLE of each channel. A better idea is to impose a model on $\hat{p}(i)$. It can be any number of analytical functions such as: linear, exponential, Gaussian, piecewise defined, etc. It can also be a linear combination of any of these. Coming back to our definition of the likelihood function:

$$L(\theta|x) = P(x|\theta), \tag{9}$$

where θ is now the ensemble of parameters that define the function. For example, in the linear case:

$$p(i) = ai + b \rightarrow \theta = \{a, b\}. \tag{10}$$

Maximizing $L(\theta|x)$ will yield the maximum likelihood estimator:

$$\hat{\theta}_{MLE} = \{\hat{a}_{MLE}, \hat{b}_{MLE}\}. \tag{11}$$

Using the MLE as the plugin estimators [9], our estimated linear regression becomes:

$$\hat{p}(i) = \hat{a}_{MLE} i + \hat{b}_{MLE}. \tag{12}$$

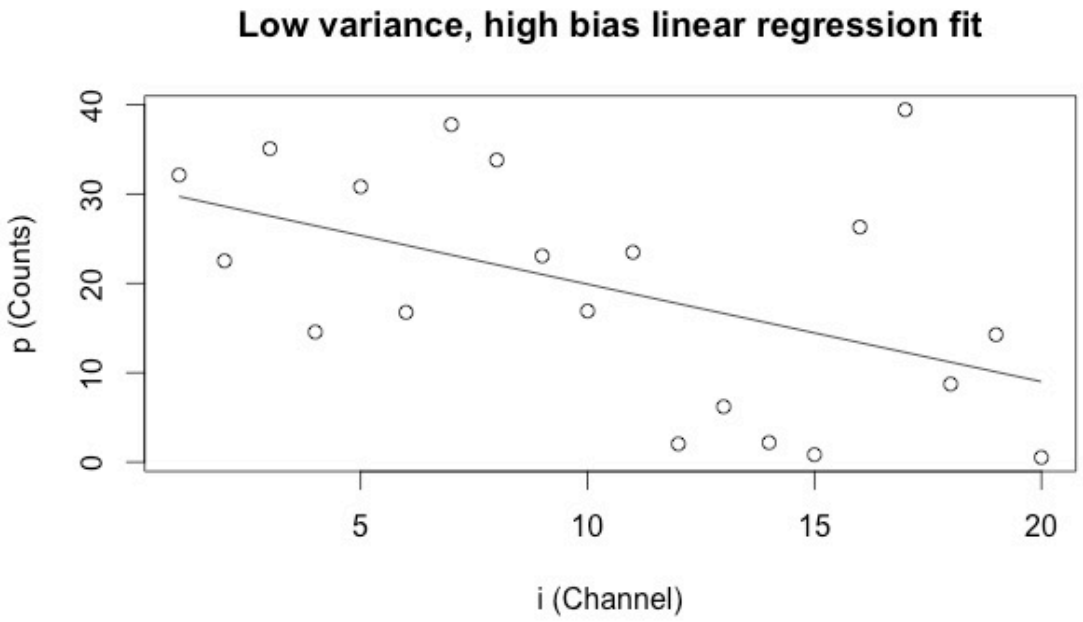

Figure 27. Linear regression obtained by estimating parameters a and b.

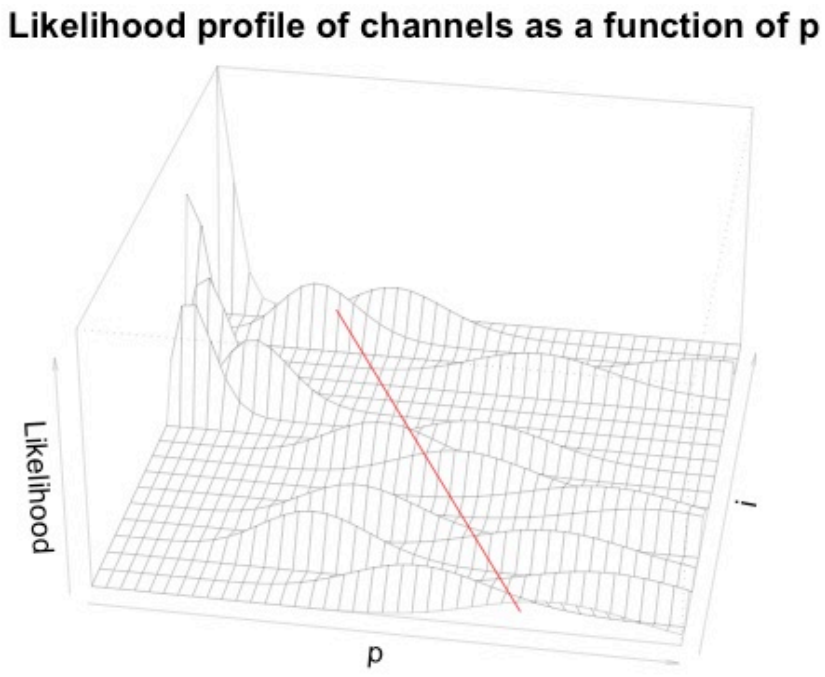

Figure 28. Linear regression (red) that maximizes the combined product of each channel's likelihood function.

In figures 27 and 28, the conditions are so strict that we have no variance but a high bias. The restrictions on the regression line are either imposed by a model in the parametric case or by a risk function in the non-parametric case [9]. A risk function is simply a way of optimizing two criteria at the same time. Generally, the problem will be minimizing both bias and variance.

In the problem at hand, we have information about the underlying physics so we will take a parametric approach. In fact, our understanding of the problem is such that we not only know how to parameterize the channel to channel variation of the underlying probability density but we also know how to model the intra channel likelihood function.

Before going further, we address the question of why not use the ordinary least square regression as a penalizing function. The short answer is that the OLS works well with large statistics and normally distributed error terms. In our case, we know how to model the error terms and we know that they are not constant or always normally distributed. In the case of very few counts, the variance of the estimated parameters is larger with the OLS than with the MLE. Using the likelihood function enables us to have a more robust curve fitting method that does not fail for large differences in counts per

channel. The effects of this are not always visible for simple regressions given that for certain regression functions, the OLS tends towards the MLE. A simple test example where we fit an exponential regression on low count channels can gave us up to 10% bigger variance for the OLS than the MLE estimators. We note here that using the MLE, when feasible, is generally always the best idea [9][60]. The predominance of OLS in many scientific fields stems from the fact that it is easy to compute and in some cases, such as simple linear regression where the error terms are normally distributed and constant, the OLS is equivalent to the MLE. Again, in our case, we know how the error term is distributed and we know that it is not constant nor is it always normal. The use of the maximum likelihood estimator is feasible in our context so we might as well use it.

Applying this model to our problem we take two approaches: the first is energy histogram fitting and the second is time-energy histogram fitting. The energy histogram fitting is simply taking the time integrated histogram (traditional histogram) and fitting an additive model of piecewise defined kernel functions. The time-energy histogram fitting is similar but the data set and kernel functions are two-dimensional (figures 29 and 30).

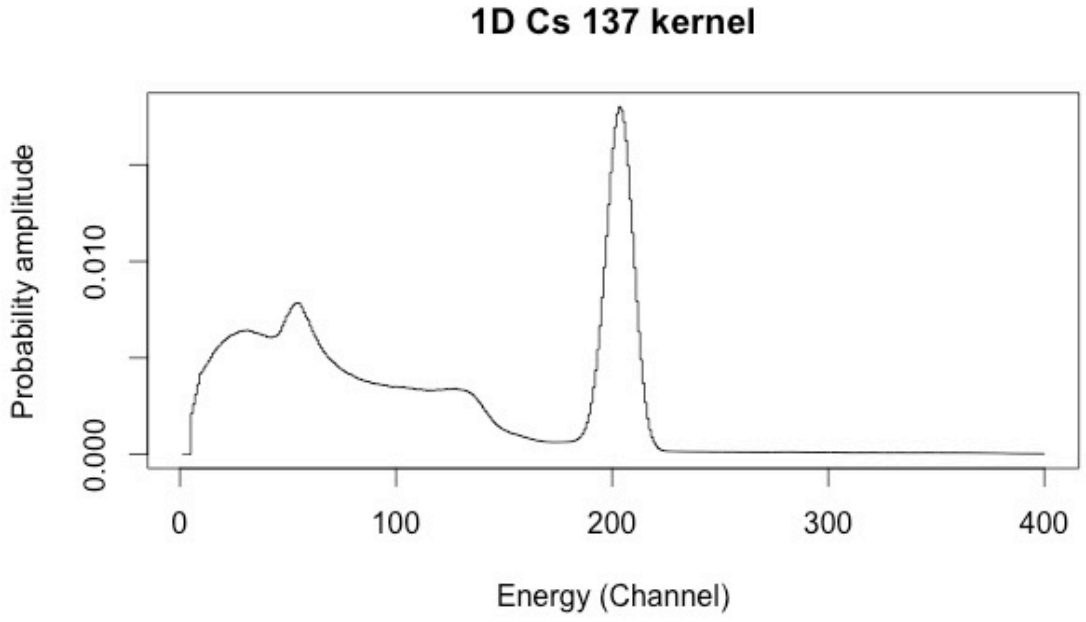

Figure 29. One dimensional (energy spectrum) kernel for ^{137}Cs.

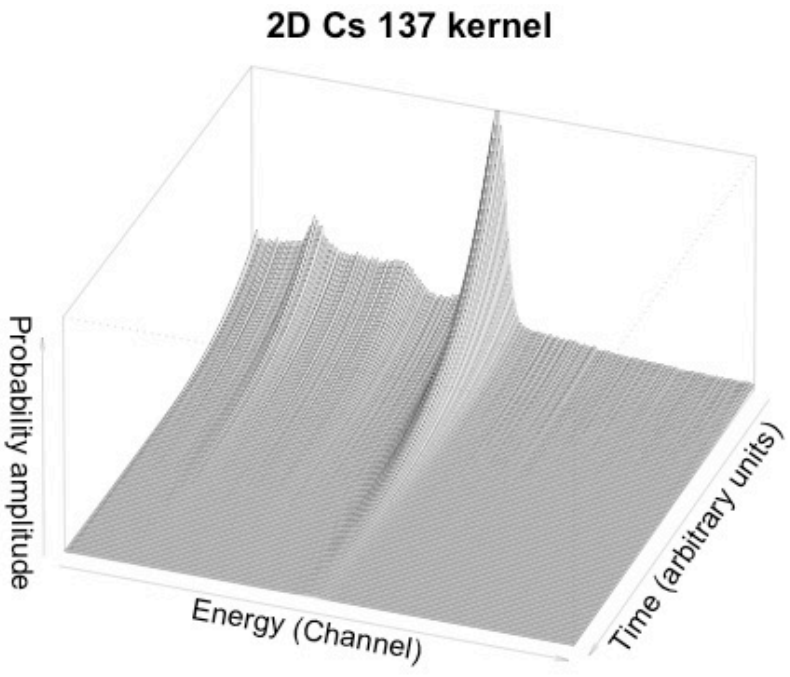

Figure 30. Two dimensional (time-energy spectrum) kernel for ^{137}Cs.

Both methods are similar; we first construct kernel functions by acquiring the background spectrum and the spectrums of isotopes of interest. We can smooth these spectrums or wait until we have a very high statistic. We then renormalize the spectrum, i.e. we divide by the total number of detected events. Figure 31 shows examples of these isotopic one dimensional kernels.

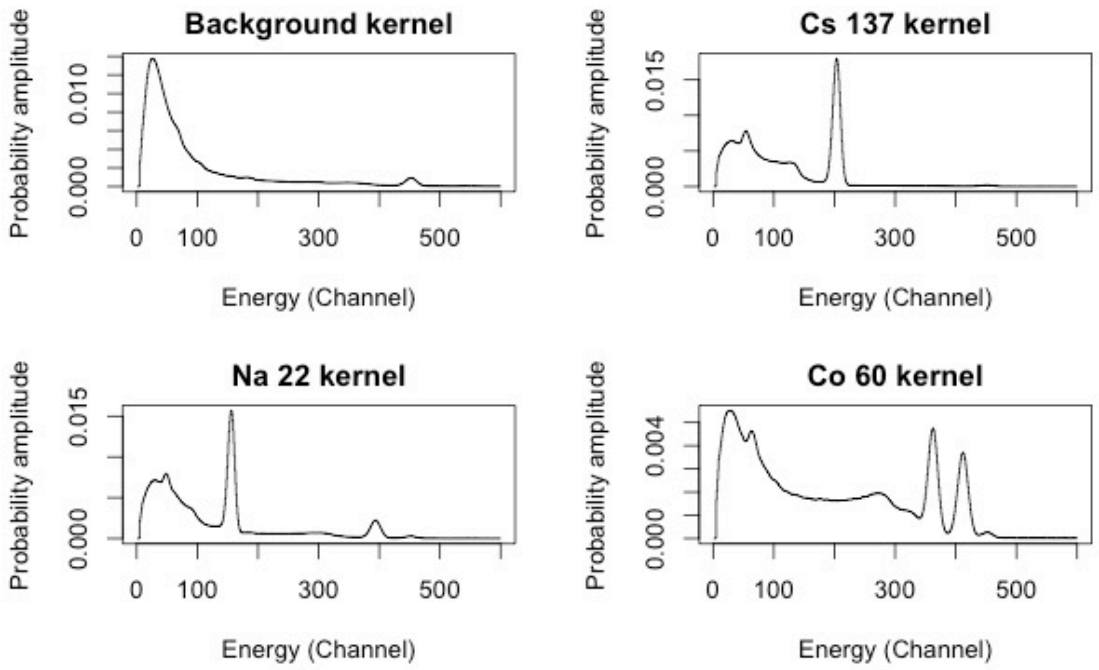

Figure 31. Examples of density kernels for various isotopes.

We construct a model for our regression, which is simply the linear combination of these kernel functions $f(i)_{iso}$. For example:

$$Fit = a_{Bk}f(i)_{Bk} + a_{Cs137}f(i)_{Cs137} + a_{Na22}f(i)_{Na22} + \cdots. \quad (13)$$

A metaheuristic algorithm then maximizes the total likelihood and yields the maximum likelihood estimates of the amplitudes $\widehat{a_{iso}}$. With these estimates we can compute the activity of each isotope and the activity ratio with respect to the background. We expect the amplitudes to belong to the positive real numbers. In addition, the sum of all the amplitudes should be similar to the number of counts detected in the spectrum.

The reason for carrying the total likelihood until this point will now become apparent. Given that we can compute the likelihood of our fit, we can use the likelihood ratio test to test against a null hypothesis. In simple terms, we can say to what degree of confidence we are detecting a given isotope. For example, if we are interested in determining if ^{137}Cs is present or not, we construct a null hypothesis $H0$ and an alternative hypothesis $H1$:

$$Fit\ H0 = a_{Bk}f(i)_{Bk} + a_{Na22}f(i)_{Na22} + \cdots \quad (14)$$

and

$$Fit\ H1 = a_{Bk}f(i)_{Bk} + a_{Cs137}f(i)_{Cs137} + a_{Na22}f(i)_{Na22} + \cdots. \quad (15)$$

Notice that the difference between both fits is only removing the kernel function responsible for fitting the ^{137}Cs spectrum. H1 has n degrees of freedom and H0 has n-1. The likelihood ratio test statistic is:

$$D = -2\ln L(H0) + 2\ln L(H1). \quad (16)$$

We also know that D follows a chi-squared distribution with one degree of freedom for our case:

$$df = df(H1) - df(H0) = n - (n-1) = 1. \quad (17)$$

Our metaheuristic algorithm returns the log-likelihood of our fit, which we use directly in the computation of D. If D is greater than 3.84, we can reject the null hypothesis of no ^{137}Cs with a significance level of 0.95.

This method of detecting signals in the spectrum yields very impressive results as will be shown in the results section. A few caveats to this method must be noted. This method presumes knowledge of all possible spectrums likely to be present in the experiment. The easiest way to ensure this is to setup the experiment as to not have unknown isotopic spectrums. Another, more robust way would be to know (code) all possible isotopic spectrums. This was not done for obvious reasons of time constraint during the elaboration of the project. Only a total of 8 kernel functions were coded: the background, ^{137}Cs, ^{22}Na, ^{60}Co, and the spectrums of activated Au, Cu, Al and Ag (some are shown in figure 31). To prevent errors due to the incomplete library of isotopes, the analysis algorithm looks at extreme lacks of fit between the data and the regression spectrum. In the case of no unknown isotopic spectrums, the distance from the data to the fit follows a statistical distribution. We count the number of times where an extreme lack of fit occurs $\Delta > 5\sigma$. Over 1024 channels the average number of times such lack of fit would happen is zero. In the case of unknown isotopes, the fit will have a hard time of sticking to the data. When this happens, we get numerous (~90) extreme lack of fit events. Above a certain threshold (~4 extreme lack of fit events), the analysis algorithm sends a

message to the user warning that there might be a problem with the fit and to dismiss the results any statistical inference from the experiment. The original idea was to set a lower threshold (~3 σ) so that each channel had around 1% chance of being an extreme event. Then knowing that each channel becomes a Bernoulli trial with p=0.01, repeated n=1024 times, we could create a statistical test to see if we were outside of what would be a typical binomial distribution for those parameters. By testing different values heuristically we ended up choosing 5 σ and the threshold described above because it lead to a more stable and systematic exclusion of pathological cases.

In the case of the two-dimensional fitting, our time-energy space is divided into bins and the amplitude of the kernel is compared to each bin much like it is with the standard histograms. The analysis of this space is slower because the computation time now scales as the square of the binning. Instead of having quasi real time fitting, as is the case for the histograms, the analysis of the time-energy spectrums is done once the acquisition is complete. Not having the constraint of live fitting could enable us to make more computer intensive algorithms such as including the use of bootstrap to evaluate the error of our fitted values. The bootstrap in general and how we propose to use it in our system is discussed in the annex.

4.3 Metaheuristic optimization

The optimization of the likelihood function is not trivial. Our attempts at finding an analytical solution failed. We then turned to metaheuristic approaches, which did the trick. The algorithm is very simple. It is a local search algorithm [43] that adjusts the size of its neighbourhood automatically. The schematic of the algorithm is presented in figure 32.

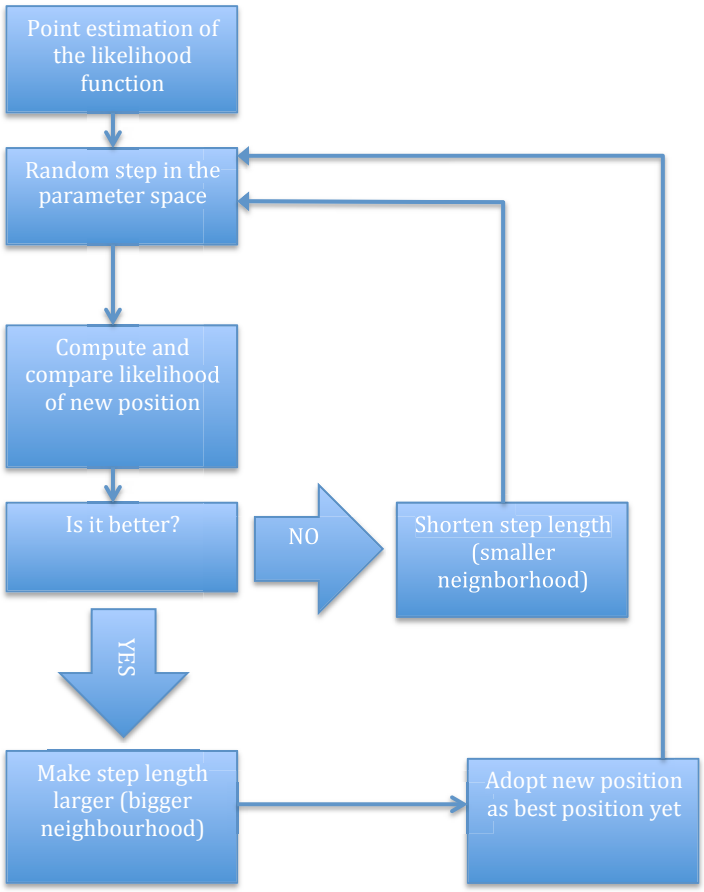

Figure 32. Schematic representation of the metaheuristic algorithm.

We set a limit on the size of the step. When an inferior threshold is reached, the algorithm stops and returns the best-found parameters and the fitness function. On a typical CPU core, the optimization takes around 10 seconds to converge. The cut off value for the smallest step is selected so that the change in the estimated value by such a step would be insignificant from a physical point of view. The reason why we chose to have an adaptable step is that, early on in the optimisation process, we want the algorithm to explore many distant points in the hyper plane to avoid being stuck in a local maximum. We note here that our intuition tells us that if local maxima exist they should be relatively small compared to the absolute maximum. This intuition is guided by the fact that in general the combined likelihood is not poly-modal. In addition to this safeguard, we want the algorithm to converge faster. We let the step get larger when it seems as if finding higher points on the hyperspace is easy. We can visualise is as an explorer at the foot of a mountain. Any step in the direction of the mountain will be beneficial. As the likelihood of finding higher ground drops, we

reduce the step size to help the algorithm not jump over the top of the mountain. Continuing with our metaphor, sometimes the algorithm encounters ridges. In this environment the algorithm reduces the step size much like a mountain climber would walk across said ridge rather than run in all directions. Once on the other side, if the terrain becomes easier, the step size can then get larger again.

4.4 Kernel functions (reference spectra)

As described previously, the construction of the kernel functions is done by taking an experimental spectrum with good statistic, smoothing it and renormalizing it. Two methods are available to us for smoothing: the moving average and the kernel density estimation (KDE). Only adaptive methods are useful here. We have no underlying model for the curve we want to smooth; we are essentially creating this model at the moment. We do know some things about the data such as the approximate width of photopeaks, the nature of the statistical noise but we have no complete view of the spectrum. Other filtering methods exist but KDE is a staple of non-parametric smoothing and using a Gaussian kernel seems intuitive given the nature of photopeaks in our data.

The moving average is simply creating a histogram, where the value of a given channel is the average of itself and its neighbours. This acts as a low pass filter, reducing channel-to-channel variance but keeping the shape of the spectrum. The advantage of this method is its ease of computation. The drawback is that for sharp photopeaks, it attenuates the maxima. A second more robust approach is to use KDE. The premise of KDE is to place a Gaussian (or other) kernel on top of each count. This has a very similar effect to having a weighted moving average where the weight function is Gaussian (or other). In both cases the goal is to adjust the width parameter: the number of neighbours considered for the moving average or the sigma of the Gaussian kernel for the KDE. Choosing a larger width will lead to more smoothing (high bias, low variance) but can eventually lead to over smoothing: attenuation of real signal, as in the case of the photopeak. Choosing a smaller width will lead to a better fit of the smoothed data with the original data but might be unable to smooth properly the noise, i.e. over fitting (low bias, high variance). By minimizing the risk function we can find an appropriate width for our smoothing. A common approach is to take the mean squared error (MSE) as the risk function. The MSE is the sum of the variance and the square of the bias of the

fitted function. Minimizing this function yields a good estimation of the width parameter and prevents over and under fitting. Other risk functions exist for different applications. In our case nothing indicates that the standard MSE function is invalid a priori.

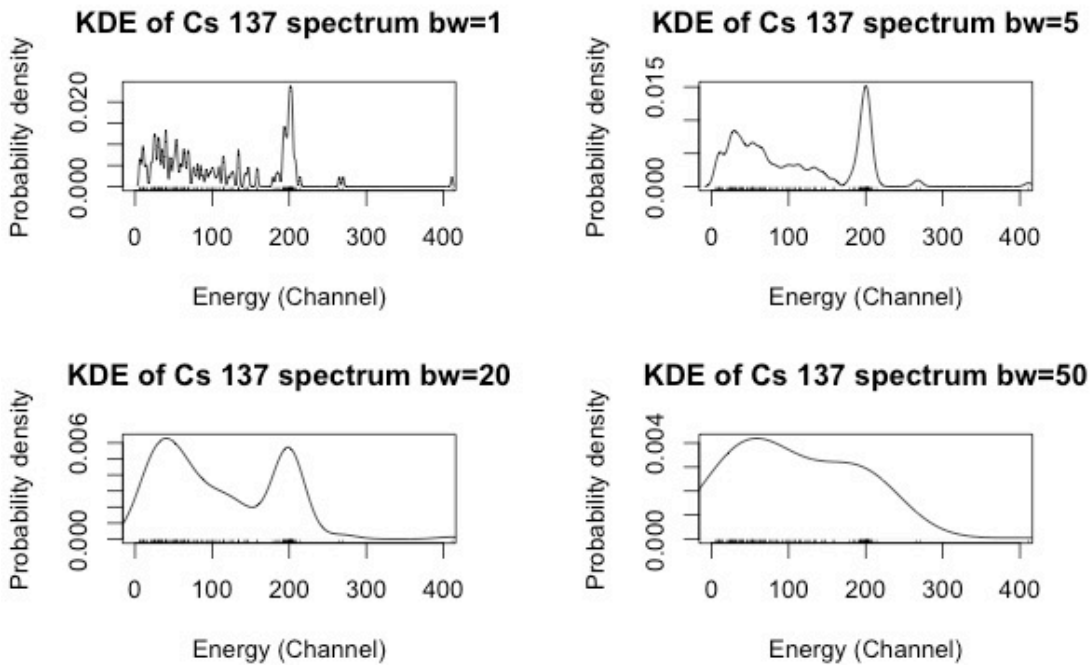

Figure 33. Kernel density estimations of a ^{137}Cs spectrum for different bandwidths.

Minimizing the risk function in this case (figure 33) yields a bandwidth of approximately 20. In a non-parametric setting this would probably be the ideal choice of bandwidth. It would indicate that there are two regions of interest: the full energy peak and the background lobe. This choice of bandwidth also leads to a broadening of the peak. Using non-parametric methods usually leads to a loss of information that could be prevented by the use of a parametric method [9]. In our case we could use a bandwidth (~5) similar to the detector resolution to avoid unnecessary peak broadening. A bandwidth of 1 leads to under smoothing and the bandwidth of 50 leads to over smoothing. A posteriori, minimising the risk function was sub optimal given our knowledge of the approximate photopeak width.

The approach is similar for the two dimensional time-energy spectrums. The only difference is that for time varying spectrums i.e. spectrums from activated isotopes; we sum multiple spectrums together to get a better statistic. For time constant spectrums such as the background or longed lived isotopes (half-life > 100 h), we get the time-integrated profile and keep it constant through time.

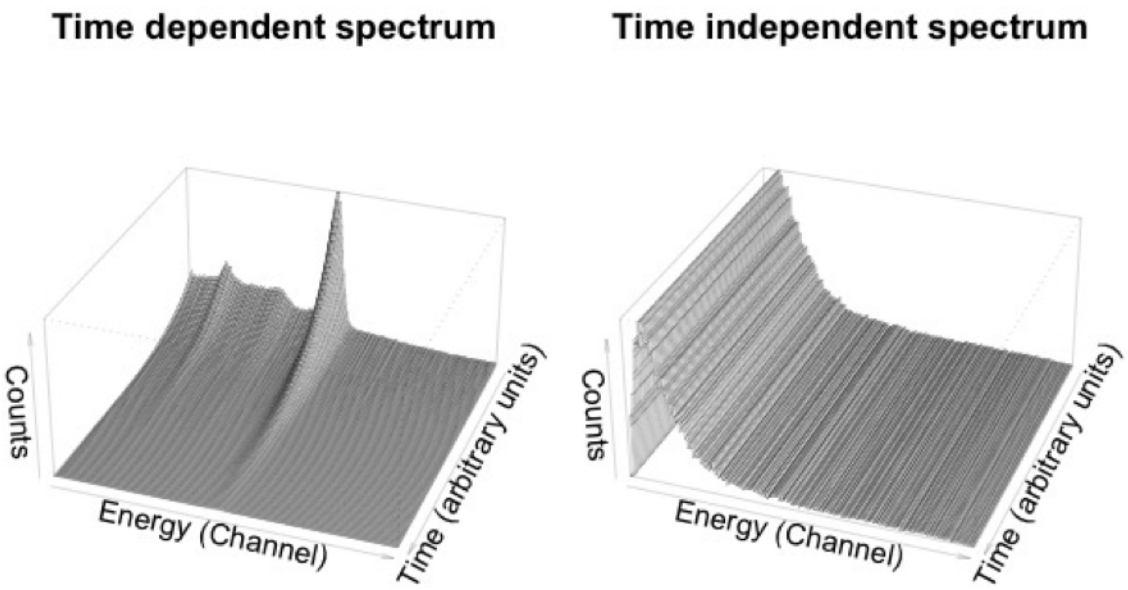

Figure 34. Time dependent (exponential decay) spectrum of ^{137}Cs and time independent (constant) spectrum of background radiation.

4.5 C# implementation

The C++/C# code that evaluates the fitness function actually computes a function proportional to the log likelihood function times a constant. Neither the constant nor the log transform has incidence on the calculations of the MLE or the likelihood ratio test. The constant is the renormalizing factor of the binomial distribution, which depends on the counts per channel k and the total number of counts n. Because these values are the same and are constant for both hypotheses, they cancel out for the D-test. The reason why the constant multiplication and the log transform do not affect the position of the maximum is that they are monotonous functions.

Computation time and memory limitation for double-precision floating-point numbers are the main reasons for the modification of the fitness function. It is simpler to record the logarithm of a very small number ($\sim 10^{-5000}$) than to record the value itself. The smallest floating-point 64-bit unsigned variable that can be stored is of the order of 10^{-324}. These extremely small numbers come from the fact that the PDF of the likelihood function evaluated at any given point is already quite small ($\sim 10^{-8}$ to 10^{-2}). The total likelihood estimation is then the product of 1024 small amplitudes. In addition, by using the log likelihood, all multiplications of likelihoods become additions of log likelihood and exponents (from the binomial distribution) become multiplication. This is where the discarded normalisation constant comes from.

4.6 Results

We see here that our analysis system can identify isotopes with much looser criteria than traditional methods. In this case (figure 35), we can positively identify the presence of ^{137}Cs with >99% confidence. The red part of the histogram is where the 662 keV peak is. Notice that it is invisible to the naked eye. In fact the amplitude of the peak is orders of magnitude smaller that either the statistical noise or the average background in that region.

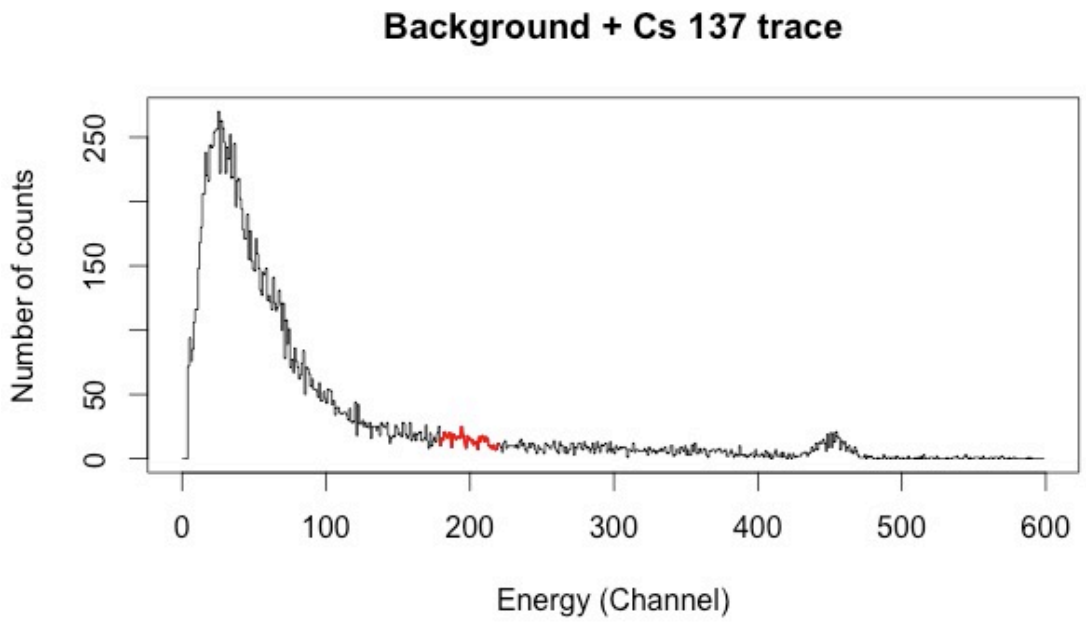

Figure 35. Spectrum of a low activity ^{137}Cs source.

In figure 35, the background radiation dominates the ^{137}Cs activity. The region of interest (red) indicates the area of the full energy peak for ^{137}Cs. It is highlighted to help the reader identify the region. Because traditional methods have stricter criteria for peak identification, they are much slower (10 to 100 times) and generally do not offer a credible estimation of the confidence interval. We estimate the traditional method as needing a signal that is 2.97 sigma out of the noise. To estimate the time, we let the acquisition continue until either the center of the peak or the integral of the peak (minus an estimated contribution of the background) is at 2.97 times the average variance of the background. This experiment was realised 10 times with a ^{137}Cs source and 30 times with no source. The typical time for the experiment was 6 minutes[8]. In every case where the isotope was present, the system was capable of identifying it on both detectors separately. The system did not identify any other isotope that the one present. When no isotopes were present, the system did not identify any isotopes. The confusion matrix indicates no type 1 or type 2 errors. We indicate here that for acquisition times of more than 30 minutes with no isotopes present, the system would sometimes

[8] This was the time needed for our system to detect the presence of an isotope. The experiment was left to run longer (up to an hour) when we were trying to determine the time it took to get a visible peak.

identify radioactive sources. It is our belief that this limit comes from the fact that the reference spectrums, event after smoothing, have higher channel-to-channel variance. By taking longer reference spectrums, this problem was pushed back to longer acquisition times. When the acquired spectrum becomes smoother than the reference spectrums, it is possible that adding very small amplitudes of the isotope spectrums helps smooth the background locally. As stated, we can avoid this problem by having better reference spectrums or limiting our acquisition times. Also, we must remember that statistical significance, by no means, indicates physical significance.

4.7 Conclusion

In this chapter we demonstrated how we built our statistical model from the ground up. Guided by our knowledge of the physical nature of the data, we arrived at a statistically robust model. We showed with a simple experiment the ability of the method in discovering very small signals in the spectrum. We note that the approach for simple histogram analysis and time dependant histogram analysis is virtually the same form a computational point of view. This is, in our view, an example of the strength of the current model. Whereas traditional approaches would have to be thoroughly reviewed to include time dependent analysis, all we have to do is modify the kernel functions and we are good to go. In the next section we present a series of simple experimental test designed to display the good function of the prototype.

5 Prototype characterization and first results

5.1 Introduction

In this section we will show examples of the GREYSTAR system at work and show off the different elements introduced in the earlier chapters. The first results simply demonstrate the ability of the GUI to inform the user of detected isotopes past a statistical threshold. We then present a series of experimental tests designed to showcase the functionality and the limits of the current prototype.

5.2 Isotope detection and poor fit discrimination with the GREYSTAR Analysis Tool

First of all, we show a screenshot of the GAT (GREYSTAR Analysis Tool) GUI having correctly identified a high activity ^{137}Cs source (figure 36). The source was placed in close proximity to the detector without any blocking material for 8 minutes. We note the absence of false detections for other isotopes. We suggest the reader take the time to read the annex describing the GUI. We repeat here that the first two windows represent the analysis results of each detector. The third window is the combination of the statistical tests. The statistical tests are combined on a likelihood level rather than on a spectrum combination level. This ensures the best way to mix these sets of experimental results.

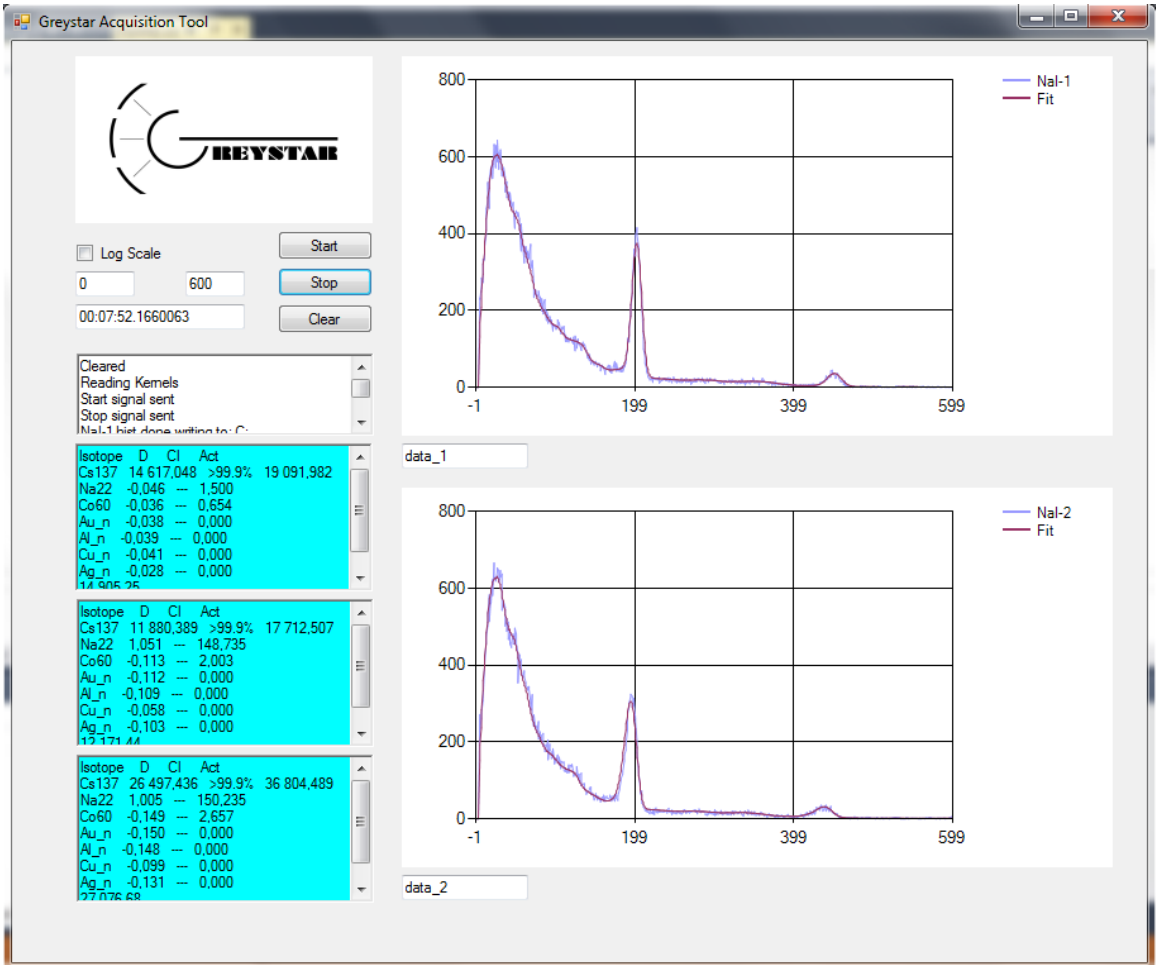

Figure 36. Screenshot of the GUI.

The blue background in figures 36 and 37 on the communication windows indicate that an isotope has been detected beyond 99.9% certainty.

We also have a screenshot of the GAT fitting multiple isotopes (figure 37). In this case we want to test the ability to detect multiple isotopes concurrently while not identifying absent isotopes. We placed ^{22}Na, ^{137}Cs and ^{60}Co in proximity to the detector and made a 3-minute acquisition.

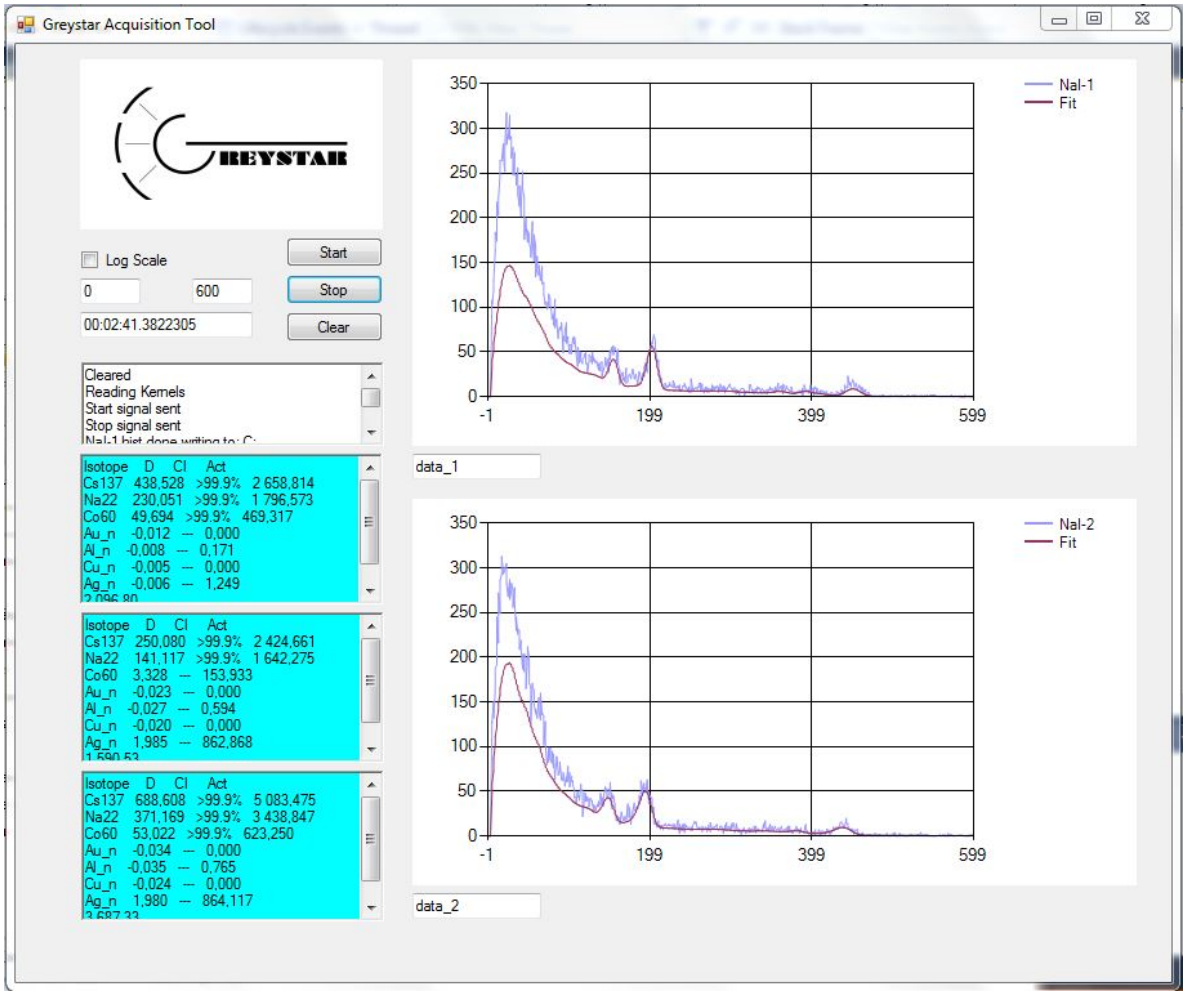

Figure 37. Screenshot of the GUI.

Here is an example (figure 38) of the analysis algorithm rejecting the fit because of too many extreme lack-of-fit events. To obtain this result, we placed a source of ^{54}Mn in proximity to the detector knowing full well that the reference spectrum for this isotope is absent from the library of known isotopes. For this reason the system has a hard time fitting a good model. The lack of fit algorithm kicks in to inform the operator that the estimated amplitudes and statistical significances are probably wrong. Notice the orange background in figure 38 on the communication windows indicating a poor fit.

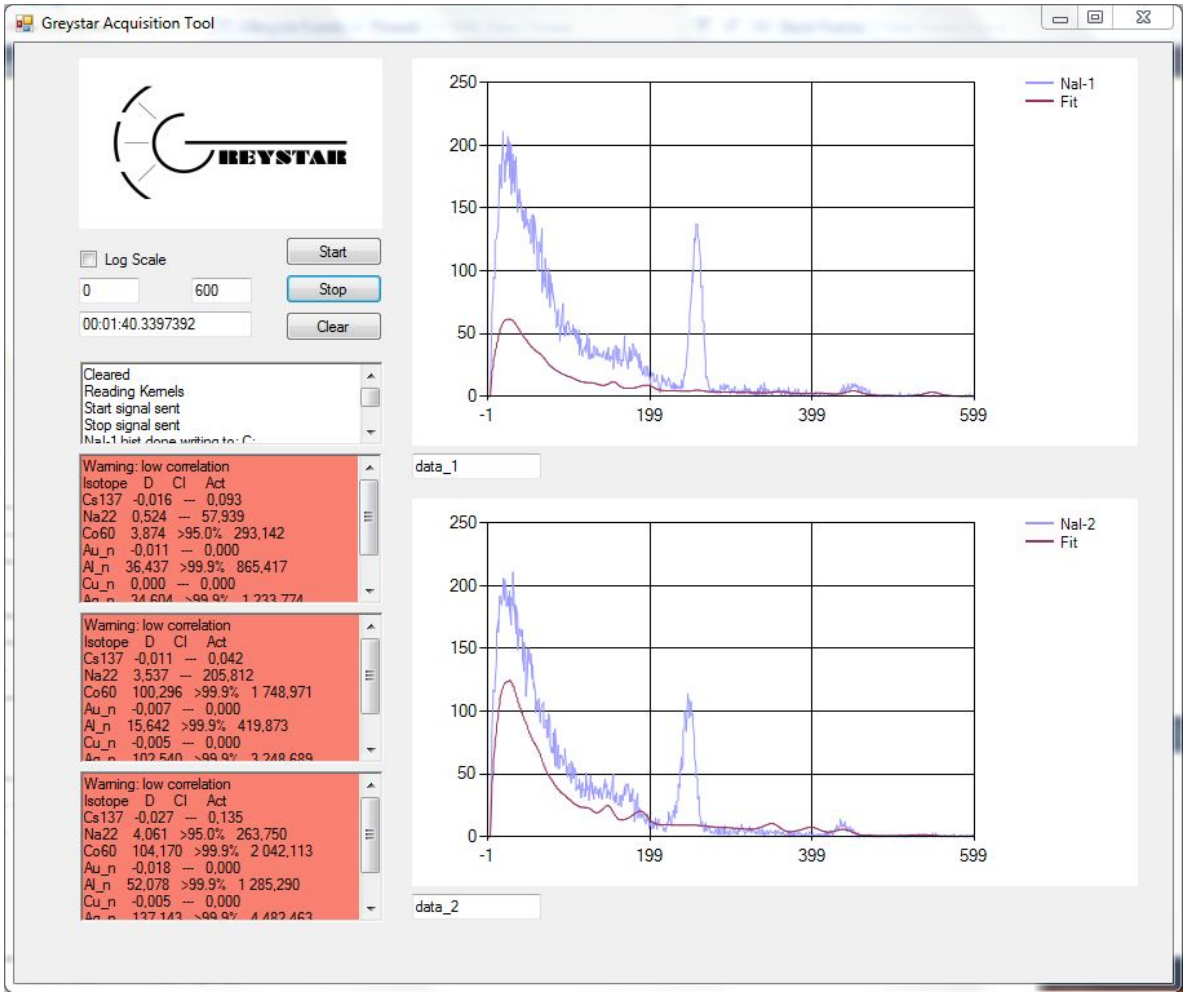

Figure 38. Screenshot of the GUI.

5.3 Time-energy event recording

The system is able to record detection events with enough time and energy resolution to observe both the energy and the time decay of activated elements. Here is an example of the data obtained from the activation of an aluminum rod (figure 39).

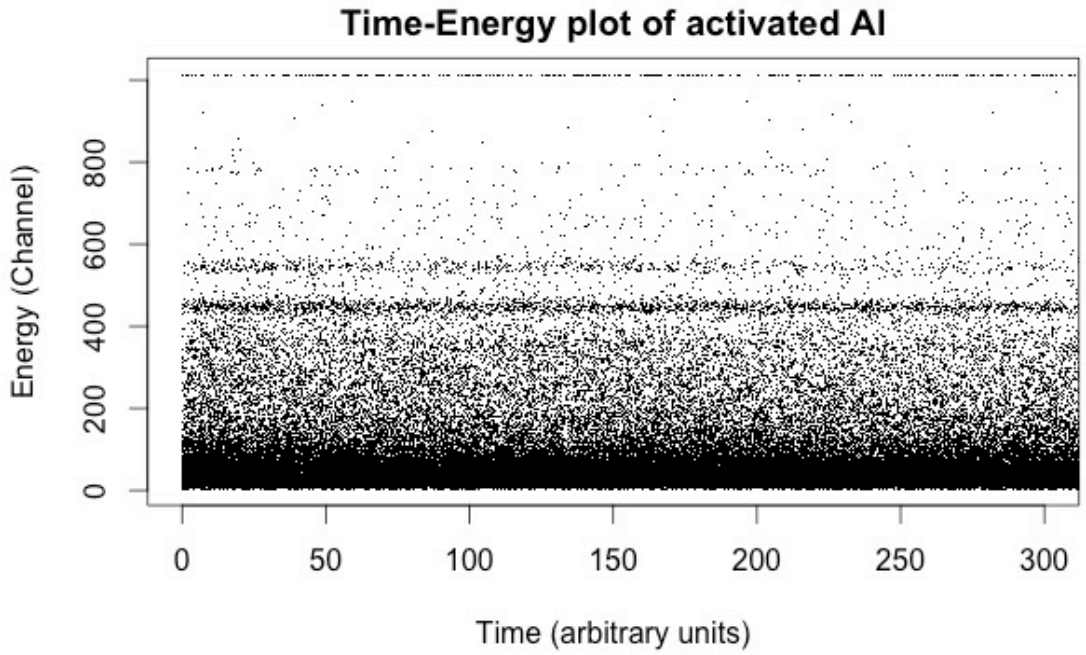

Figure 39. Time-energy spectrum of activated aluminum.

Notice the horizontal lines around channel 450 and channel 550. These lines are the background ^{40}K peak and the activated Al peak respectively. The background stays constant while the activity of the aluminum decays over time. Figure 40 shows the same data in a density spectrum.

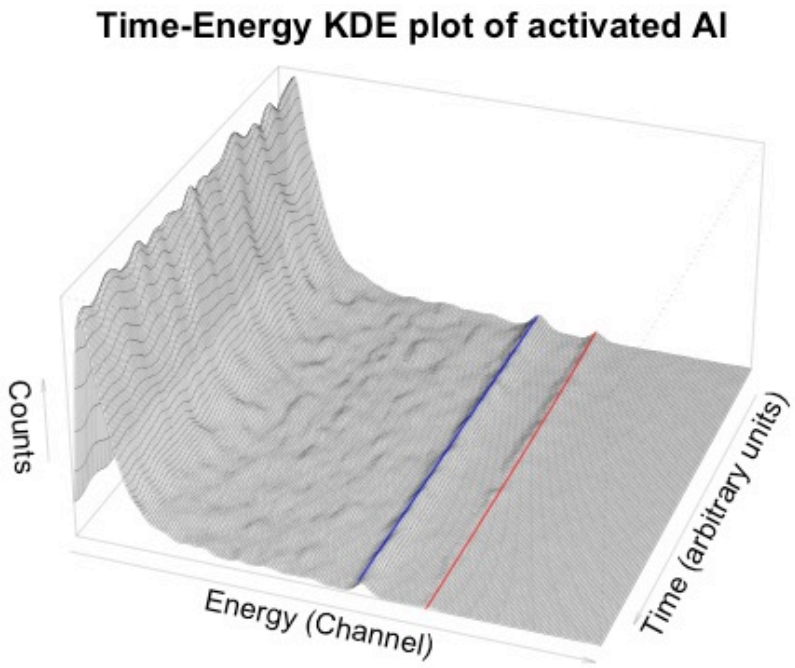

Figure 40. The time-energy KDE of activated Al data.

The blue line indicates time independent ^{40}K peak. The red line indicates the time decreasing activated Al peak. The KDE bandwidths were chosen only to help visualization. No detection algorithm was used to produce these plots. They are presented here to showcase the ability of recording data with the system. The recorded data, at least visually, exhibits features that we would expect this kind of measure to have. In the next section we discuss how this recorded data is treated to extract observables such as the relative activity amplitudes and confidence intervals.

5.4 Deep analysis with 2D fitting results

The deep fit analysis tool was tested using a combined source composed of activated aluminum and a ^{137}Cs source. The purpose was to see if the system was in fact able to fit multiple reference spectrums, in this case, three (Al, ^{137}Cs and Background). Computation times are orders of magnitude longer than for 1D fitting (around 20 to 30 minutes). As such, we did not implement a

bootstrap for the tests, which would require thousands of analyses. It would be advisable to use computer clusters to parallelize the computation. Our test was conclusive giving very high D-values (likelihood ratio test statistic). The setup was comprised of activated Al in close proximity to the detector and a 5 μCi ^{137}Cs source placed at 1 m from the detectors. The acquisition time was 3.0 minutes. This experiment shows the ability of the prototype to analyse two-dimensional data. Further test should be scheduled to rigorously evaluate the detection power gained from this type of analysis. We have limited ourselves to demonstrating the ability of the prototype of doing this type of analysis. From an experimental point of view, we have shown that we can analyse radioactive isotope of largely different half-lives in the same experiment.

Given the difficulty in generating time decreasing reference spectrums for poly-isotopic materials, only Al was tested. Further work is being done to simulate, using GEANT4, the appropriate mono energetic and mono-isotopic spectra. These spectra can then be combined to create adequate reference spectrums. As of the redaction of this thesis, these efforts have not yet been completed. The main reason is the need to develop a credible filter that emulates the effect of the acquisition system in producing comparable spectrums. Two experimental methods exist for the creation of poly-isotopic materials. The first is to record a time dependent spectrum a great number of times, but the curse of dimensionality makes this prohibitively long. The alternative is to integrate over different time periods the acquired data and split the spectrums of both isotopes. This technique was used for the creation of the one-dimensional spectrums of copper and silver.

A side note: when taking time dependent measurements, the experimenter must be careful about not moving the radioactive sources. The incident radiation on the detector is dependent on the distance; a change in this distance could give the impression of activity fluctuation to the analysis tool. It would then become unwise to try and consider activity decrease due to the half-life.

5.5 Benchmark results for known isotopes and activated materials

The purpose of this section is to give the reader an idea of the systems' efficiency and to help guide further work. These are preliminary results and should not be taken as a complete characterisation of

the system. Experimental setups for the measures are given as well as any notes or warning about the system's behaviour and performance. Information about the reference sources used and the mass of the activated materials is given in the annex. Screen captures of the relevant experiments can also be found in the annex as a visual proof of the system working.

The first series of tests were done to prove the ability of the system to detect all 4 activated elements: Al, Au, Ag and Cu. All samples were irradiated by the Am-Be source for 10 minutes then placed in close proximity (<10 cm) of the two detectors. These tests were conducted by activating a single element at a time and seeing if the system was able to identify only that element.

The results for Al and Au were conclusive within one round of detection (10 min activation and 3 min detection). They returned very high D-values (>100), no other elements were detected and no warning messages were generated for lack of fit.

Cu required 3 rounds of detection before both detectors indicated a D-value greater than 10.83 (99.9% confidence interval). No other elements or isotopes were detected by the system and no warnings were generated.

Ag proved more confusing for the system. After one round of activation, the system had discovered Ag with D-values around 23. It also reported ^{137}Cs with a D-value around 6 (>95%). The reason for this is that there is a significant overlap, given the detector resolution and similar peak energy, between the 661 keV gammas from ^{137}Cs and the 657 keV gammas from activated ^{110}Ag. A second round of activation gives us D-values around 100 for Ag and still around 6 for ^{137}Cs. Setting a strict hypothesis test (low p-value, high D-value) enables us to discriminate between the two and reject the presence of ^{137}Cs. Another important note is that when comparing the activity amplitudes, we discover that the fitted ^{137}Cs activity is ten times lower than the Ag activity. From a physical significance point of view, we could also have rejected the hypothesis.

It is important to remind the reader that the confidence interval calculated should not be misinterpreted as the probability or degree of certainty that an element is present. It is a hypothesis test about the inclusion of an element. Once we have selected an arbitrary cut-off D-value, if the test falls under said value, we must accept that we cannot reject the null hypothesis. In this case the null hypothesis is that there is no statistical indication that the addition of the element's spectra makes the

fit better. This does not mean that we accept the absence of this element either, just that we cannot say if it is there.

A poly-elemental test was done by activating Al and Cu together for 10 minutes then detecting for 3 minutes. Both detectors returned an identification (>99.9%) of both elements without returning any warning message or detecting other known isotopes. The purpose of this test was to show the ability of the system to detect multiple activated materials with a common experimental protocol.

Similar tests were done on long-lived radioactive isotopes sources: ^{137}Cs, ^{22}Na and ^{60}Co. The mono-isotopic tests were passed by the system detecting (>99.9%) the appropriate isotope and no other. No lack-of-fit warnings were generated. Poly-isotopic tests were also conclusive. The system was able to fit different combinations of these three isotopes. When confronted with an unknown isotope such as ^{54}Mn, the system returned a lack of fit warning, which can be interpreted as the presence of an unknown spectrum.

Reliable detection of a 5 µCi ^{137}Cs source was possible up to 5 m in air. Detection of a 10 µCi ^{137}Cs through more than 4 cm of brick was impossible without triggering the lack of fit warning and false identification. As will be discussed in the next chapter, spectrum morphing was a great hindrance for reliable detection with the current reference spectrums. Reducing the library to only ^{137}Cs and ^{60}Co and ignoring lack of fit warnings, we were able to detect up to 26 cm. This measure should be interpreted as the possible detection limit of a more sophisticated system, one with integrated spectrum morphing.

5.6 Conclusion

In this section we presented initial experimental results that demonstrate the function of the prototype in the context of radioactive isotopes and activated material. We also show the limits incurred in this line of testing. In the next chapter we will review the results presented here and discuss how they relate to both the objectives set forth in the introduction and how they tie into the larger GREYSTAR project as well as discuss the limits of the current prototype.

6 Result analysis and discussion

6.1 Introduction

In this section we will first review our objectives and discuss how we have met those objectives. We then discuss the limits of the current prototype. Finally, we will review the larger objectives or the GREYSTAR project and how our preliminary results tie into the larger picture. We will also outline the improvements and detail what further work should be done.

6.2 Analysis of first results

The main objective of this project was to demonstrate the efficacy of our novel statistical analysis method. To do so we had to develop our own acquisition system. In the last chapter we presented a series of experimental tests that showcased the ability of our prototype to detect activated elements in a controlled environment. We were able to detect isotopes with very small activities. We also demonstrated the limits of our system when having to detect unknown isotopes. Time-energy recording and detection was also shown to work.

Tests with interposed material were only partially successful. We will discuss improvements needed in the following section. Although the larger objectives of the GREYSTAR project would require this issue to be solved, in the scope of the current prototype the description of these limits is sufficient. We believe that we have shown a promising avenue towards solving this problem.

Preliminary test were done for remote detection, up to several meters. These tests are only marginally interesting, as we know that the detection power is directly a function of the detector area, the distance and the activity of the probed material. Nevertheless, these tests were done and no unexpected results came up.

6.3 Limitations of the system

We have seen that the analysis code beats traditional peak detection algorithms in terms of our general signal to noise metric. It is important to understand what we are asking of the analysis in order to avoid making false interpretations. The statistical significance test tells us if including the spectrum of a given isotope improves significantly the fit between the data and the regression. If the regression has a poor fit to begin with, we may end up having false positives. This is generally controlled by the lack of fit warning.

The system cannot detect an isotope for which it does not have a reference spectrum in its library. It is simple to add more known isotopes to the library by modifying the source code. Although it might seem like a good idea to add every known isotope as references, in reality a smart selection of certain isotopes is better. One should make a list of likely isotopes and include only these references. The reason for this is that, as we add more references, the computation time grows. There is a linear growth from having to test the existence of an additional spectrum and there is a non-linear growth that comes from needing more time to converge in a higher dimension parameter space. Resampling methods in conjuncture with a committee type voting system such as bagging (bootstrap aggregating) [22] could be used to robustly determine the presence of isotopes but these methods are outside the scope of this thesis. Nevertheless, we describe the bootstrap, in the annex.

When we have kernel functions with significant overlap of peaks, more time should also be allocated to the convergence of the metaheuristic algorithm. It would also be wise to consider using the time-energy spectrum analysis discussed previously, especially if the overlapping spectrums have different half-lives.

To insure good results, we should make sure that there is good agreement between the experimental data and the kernel functions. If there is a significant change in the gain or offset[9] of the electronics chain, this can distort the incoming data. In this case, the system will not be able to perform optimally. Make sure that the temperatures of the PMTs are fairly stable to avoid gain changes. In the case of gain drift, we used a simple calibration method where we would adjust the gain and offset of the

[9] The analysis system could be programmed to calibrate itself, but as of the current state of the prototype, the gain and offset must be adjusted experimentally.

amplifier so that the ^{40}K peak from the background and the ^{137}Cs peak fell on their respective reference spectra. This is akin to energy calibration.

Statistical significance does not infer physical significance. When interpreting the confidence interval given by the analysis software, one should be careful before claiming any physical interpretation. It is important to check the relative activity of the detected isotope. Being 99.9% sure that there is 0 Bq activity of something does not mean that it is present. In fact, in this particular case it means it is not present.

The system was engineered to detect trace amounts of activated matter. As such, it does not expect high count-rates. To use the system in a very radioactive environment, one should enable pile up discrimination on the FPGA. This can be achieved by editing the VHDL source code. An empty module is already present for this type of discrimination. Once modified, the code must be recompiled and implanted on the FPGA.

The sensitivity, and by extension the probing depth, of the system can be improved directly by having a higher neutron flux, and by having a greater detection area. As described previously, other portable neutron sources exist and can readily be used, as discussed in the earlier chapters. To add detector area, all that is needed is to add detectors to the system. If this is to be done, additional high-voltage sources, amplifiers and ADCs must be added. The VHDL code has to be edited to add new ADC control modules and MEMs. This is relatively easy. The FPGA board used at the moment could support another ~50 detectors before running out of I/O pins. As we add detectors, the USB-UART communication might become a bottleneck for data transmission. Using another communication method (PCI), having a greater baud rate or changing the communication protocol could solve this problem.

It is important to note that, as the depth of the probed element grows, the spectrum is deformed. First there is a broadening and attenuation of the photopeak then an increase in the Compton background.

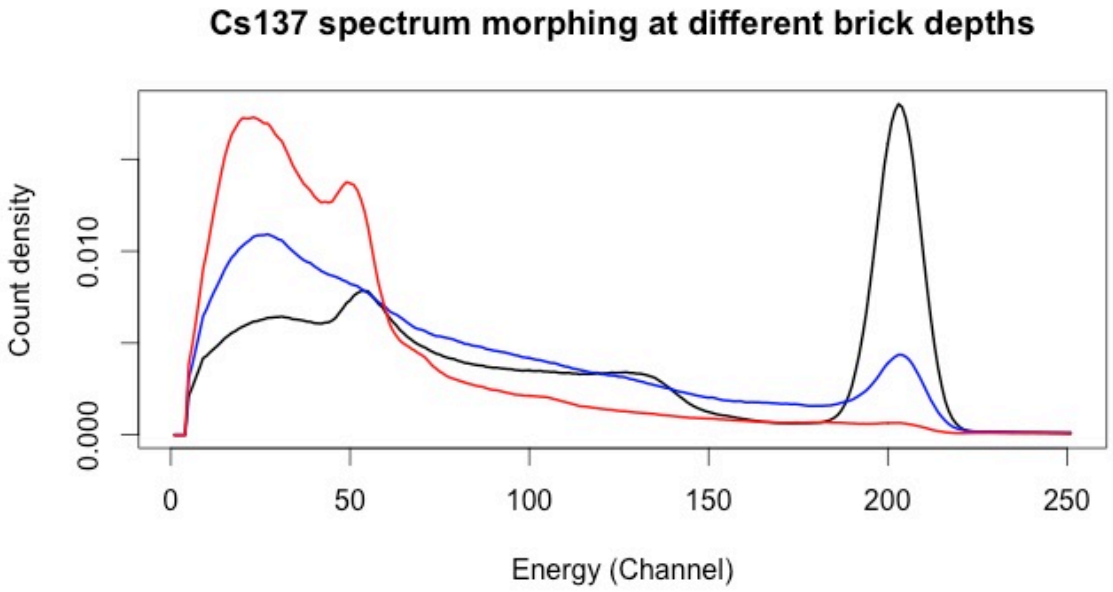

Figure 41. Spectrum morphing for different brick depths.

In figure 41, the black line is the standard ^{137}Cs kernel with no brick. The blue line is the morphed spectrum from the ^{137}Cs source behind 13 cm of brick. The red curve is the source behind 32 cm of brick. Because of this spectrum morphing, the algorithm has an increasingly difficult time fitting its reference spectrums. This results in the lack of fit warning triggering. In some cases, the broadening of the peak will trick the algorithm into thinking that there are two peaks in close proximity and will try to fit an absent element. The workaround we propose for future improvements is to add attenuated spectrums to the library of known isotopes. The depth could then be determined by the user or by maximum likelihood estimation.

6.4 Discussion and further work

The larger objectives of the project are to develop a system able to detect trace elements at a distance in the field. We have been able to demonstrate our ability to activate samples of material with high element concentrations. The first step would now be to revisit estimates done in [1] and

adjust them to the prototype at hand. In the field tests for surface detection could be carried out with no modifications. For deep detection, modifications to the isotope reference spectrums should be made as discussed previously. As already stated, we believe our method is flexible enough to enable peak-less isotope identification, especially with the time-energy detection method. We already know that we can detect isotope sources up to several meters in the air. The next step would be to estimate the neutron flux needed to enable surface detection of different concentrations of various elements for a given distance.

A study of the gain in detection power as a function of detector resolution could also be carried out. We expect to observe diminishing returns on the gain our method provides as the detector resolution increases. Nevertheless, it would be interesting to determine the relationship between the two.

Barring a few minor improvements to the analysis system, we believe that it has been demonstrated as being useful and in many respects better than traditional approaches, at least in the confines of our project. Much more work on parameterizing this prototype can be done but in order to reach the larger goals set forth in [1], certain parts of the prototype will have to be reviewed. The modular nature of the prototype should make this work easily implementable. We believe that the first improvements would have to do with the neutron source and the addition of detector volume.

6.5 Conclusion

We have discussed how our previous experimental results fit into the lager scope of the project and have given indication as to where the project should go next. In the last section we will recap the implications of what has been done in this doctoral project.

Conclusion

We have shown the capability of the prototype to automatically detect activated isotopes. The system is easily scalable on the hardware end and the limits have been discussed as to the inclusion of more and more reference spectrums. As a prototype for automated neutron activation analysis the project is a success. The main limiting factor for probing depth is the morphing of the spectrum as the radiation is attenuated. Further work could be done to model how the spectrum evolves as a function of depth but this is outside the scope of the current project. In addition, the development of a faster method for the construction of reference spectrums will be necessary. As discussed previously, first steps were made to investigate the use of simulations. In addition to this, the use of readily available libraries of spectrums [7] should prove interesting. The limiting factor for both of these approaches is, as stated in the thesis, the need for a filter that emulated the electronics chain of the setup.

As a novel acquisition and analysis method, the project is event more successful. We have built an acquisition system that rivals commercially available products that can be tenfold the cost[10]. Our system not only does simple histogram recording but event recording as well. Our open source and modular approach gives us greater freedom in experiment design. Our analysis method can be exported to any field of research where spectroscopy is used and we have shown that detection sensitivity can artificially be improved by applying the correct statistical model.

Our system was built from the ground up to capture and analyse a maximum of the available information. A multidisciplinary tour de force: using modern electronics, strong statistical tools, exotic algorithmic paradigms and a clear understanding of the underlying physics, made the design of this system possible.

[10] We compare the price of the parts used for our prototype to the price of a similar commercially available system the EASY-MCA. The development time was not accounted for. The price of common materials such as the detectors, amplifiers and computers is omitted, as they are independent of the commercial system. We also note that rebuilding our prototype now would be even cheaper as we would not need to buy a development kit. In the end, our acquisition system costs in the order of 600$ to develop but we could rebuild it for under 250$ given we do not need the EVAL-CED1Z board for anything else than a power supply. The EASY-MCA price is in the order of 2500$.

Many more experimental tests can be done with the prototype. In addition, many components can be improved. Our modular approach should make testing and implementing improvements simple. We believe that the strength of the analysis method resides in the consideration of the prototype as a whole. Modification should not however incur great modifications to the analysis system. Of course any changes that affect the acquired spectrum such as changing the detectors or improving the electronics chain will result in the need to update the library of reference spectrums. An automated approach to building this library and calibrations would probably be one of the better improvements to make.

Bibliography

[1] B. Wallace, *Theoretical and experimental foundations for the GREYSTAR project, Master's thesis* (Université Laval, Québec, 2011)

[2] B. Wallace, *EPJ Web of Conferences* ***66***, 10018 (2014).

[3] A. Miled, *Conceptions de systèmes VLSI, notes de cours* (Université Laval, Québec, 2014).

[4] Terasic Technologies Inc., *DE0-Nano User Manual* (2003-2013).

[5] Terasic Technologies Inc., *Cyclone V GX starter kit User Manual* (2014).

[6] P. J. Ashenden, *The Designer's Guide to VHDL* (Morgan Kaufmann, San Francisco, 2008).

[7] G. L. Molnar, *Handbook of Prompt Gamma Activation Analysis* (Kluwer Academic, London, 2003).

[8] H. Spieler, *Pulse Processing and Analysis* (Lawrence Berkley National Laboratory, Berkeley, 2002).

[9] L. Wasserman, *All of statistics: a concise course in statistical inference* (Springer, New York, 2004).

[10] E. L. Hoffman, *Journal of Geochemical Exploration* ***44***, 297-319 (1992).

[11] N. A. Macmillan, C. Douglas Creelman, *Detection Theory A User's Guide* (2005).

[12] W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments*, Second ed. (Sprigner-Verlag, Berlin, 1987,1994).

[13] S. Arlinghaus, *Practical Handbook of Curve Fitting*, First ed. (CRC Press, 1994).

[14] National Institute of Standards and Technology, *Neutron scattering lengths and cross sections*, http://www.ncnr.nist.gov/resources/n-lengths (Web. 2011).

[15] "The Archaeometry Laboratory at the University of Missouri Research Reactor." *The Archaeometry Laboratory at the University of Missouri Research Reactor* (Web. 2014).

[16] "Nuclear Methods in the Carbon Cycle." *Nuclear Methods In The Carbon Cycle* (Web. 10 Oct. 2014).

[17] K. S. Krane, *Introductory Nuclear Physics*, Second ed. (John Wiley & Sons, Inc., USA, 1987).

[18] National Institute of Standards and Technology, *Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients.* (Web. 2015).

[19] Analog Devices, *EVAL-AD7352/AD7356/AD7357 data sheet.* (2010).

[20] Analog Devices, *AD 7356 data sheet*. (2008-2011).

[21] Analog Devices, *EVAL-CED1Z data sheet.* (2007).

[22] T. Hastie et al., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second ed. (Springer, New York, 2011).

[23] D. Leveque, *Utilisation d'un programme de simulation pour l'optimisation d'un flux neutronique issu d'une source d'Américium-Béryllium*, *rapport de stage* (Université Laval, Quebec, Canada, 2012).

[24] Trombka and Schmadebeck, *A numerical least-square method for resolving complex pulse height spectra*. (NASA, 1968)

[25] Nadkarni, R.A., and W.D. Ehmann. *NEUTRON ACTIVATION ANALYSIS OF WHEAT FLOUR SAMPLES*. *Radiochem. Radioanal. Lett. 6: 89-96* (1971)

[26] Debertin, K., & Helmer, R. G. *Gamma-and X-ray spectrometry with semiconductor detectors*. *North-Holland* (1988).

[27] Reilly, D., Ensslin, N., Smith Jr, H., & Kreiner, S. *Passive nondestructive assay of nuclear materials* (No. NUREG/CR-5550; LA-UR--90-732). Nuclear Regulatory Commission, Washington, DC (United States). Office of Nuclear Regulatory Research; Los Alamos National Lab., NM (United States). (1991).

[28] Eberth, J., & Simpson, J. *From Ge (Li) detectors to gamma-ray tracking arrays–50 years of gamma spectroscopy with germanium detectors*. *Progress in Particle and Nuclear Physics*, *60*(2), 283-337. (2008).

[29] Grau, J. A., Schweitzer, J. S., Ellis, D. V., & Hertzog, R. C. *A geological model for gamma-ray spectroscopy logging measurements*. *Nuclear Geophysics* (1989).

[30] Szentmiklósi, L., Belgya, T., Révay, Z., & Kis, Z. *Upgrade of the prompt gamma activation analysis and the neutron-induced prompt gamma spectroscopy facilities at the Budapest research reactor. Journal of radioanalytical and nuclear chemistry*, *286*(2), 501-505. (2010).

[31] Shyu, C. (1991). *Development of the Monte Carlo-library least-squares method of analysis for neutron capture prompt gamma ray analyzers*. North Carolina State Univ., Raleigh, NC (United States).

[32] Olmos, P., Diaz, J. C., Perez, J. M., Garcia-Belmonte, G., Gomez, P., & Rodellar, V. *Application of neural network techniques in gamma spectroscopy*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *312*(1), 167-173. (1992).

[33] Koskelo, M. J., & Mercier, M. T. *Verification of gamma spectroscopy programs: a standardized approach.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 299(1), 318-321. (1990).

[34] Bronson, F., Atrashkevich, V., Geurkov, G., & Young, B. *Probabilistic uncertainty estimator for gamma-spectroscopy measurements*. *Journal of Radioanalytical and Nuclear Chemistry*, *276*(3), 589-594. (2008).

[35] De Soete, D. *Neutron activation analysis*. (John Wiley & Sons, Inc., USA, 1972).

[36] Gordon, G. E., Randle, K., Goles, G. G., Corliss, J. B., Beeson, M. H., & Oxley, S. S. *Instrumental activation analysis of standard rocks with high-resolution γ-ray detectors*. *Geochimica et Cosmochimica Acta*, *32*(4), 369-396. (1968).

[37] Laul, J. C. *Neutron activation analysis of geological materials*. (1979).

[38] Ward, N. I., & Mason, J. A. *Neutron activation analysis techniques for identifying elemental status in Alzheimer's disease*. *Journal of Radioanalytical and Nuclear Chemistry*, *113*(2), 515-526. (1987).

[39] US department of energy, *DOE Fundamentals Handbook, Nuclear Physics and Reactor Theory, Volume 1*. (1996).

[40] Kayelaby.npl.co.uk,. *Attenuation Of Fast Neutrons: Neutron Moderation And Diffusion 4.7.3.* (Web. 2015)

[41] Joshua Hykes , *Al-gamma-xs, Own work*. (Web. 2011)

[42] Durpdg.dur.ac.uk,. *Interactions Of 0.3 Mev Photons In Aluminum And Lead*. (Web. 2015).

[43] Mladenovic, N & Hansen, P. *Variable neighbourhood search, Computer Ops Res. Vol. 24, No. 1 I. pp. 1097-1100*. (1997)

[44] Canberra, *Model 802 Scintillation Detectors, data sheet.* (2015).

[45] Canberra, *Model 2007/2007P Photomultiplier Tube Base/Preamplifier, data sheet.* (2015).

[46] Alfassi, Z. *Activation Analysis, Volume 2*. (CRC *Press*, Boca Raton, FL 1990)

[47] International Organization for Standardization. *Reference Neutron Radiations — Part 2: Calibration fundamentals of radiation protection devices related to the basic quantities characterizing the radiation field, ISO 8529-2,* (Geneva, 2000)

[48] Canberra, *Low Energy Germanium Detector (LEGe), data sheet.* (2015).

[49] Amptek, *XR-100T-CdTe X-Ray and Gamma Ray Detector, data sheet.* (2015)

[50] Guilherme S. Zahn, Frederico A. Genezini, Maurício Moralles, *Evaluation of Peak-Fitting Software for Gamma Spectrum Analysis, International Nuclear Atlantic Conference (Santos, Brazil).* (2009)

[51] Sharma, S., Bellinger, C. and Japkowicz , N., *Anomaly Detection in Gamma Ray Spectra: A Machine Learning Perspective, 2012 IEEE Symposium on Computational Intelligence for Security and Defence Applications (CISDA),* (Ottawa, 2012)

[52] Gnanadesikan, Ramanathan; and Kettenring, John R., *Robust estimates, residuals, and outlier detection with multiresponse data, Biometrics 28:81–124,* (1972)

[53] H. P. Kriegel, M. Schubert, and Z. Arthur, *Angle-based outlier detection in high-dimensional data*, *in Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ser.* (New York, 2008)

[54] Bader, S. *Identification and Quantificationof Peaks in Spectrometric Data, Dissertation.* (Dortmund, 2008)

[55] Meng, L.J. and Ramsden, D., *An inter-comparison of three spectral-deconvolution algorithms for gamma-ray spectroscopy*, (2000)

[56] M.A. Mariscotti, *A method for automatic identification of peaks in the presence of background and its application to spectrum analysis,* (1967)

[57] Miroslav Morháč, *Multidimensional peak searching algorithm for low-statistics nuclear spectra*, (2007)

[58] Sullivan, C.J., Lu, J., *Automated photopeak detection and analysis in low resolution gamma-ray spectra for isotope identification*. (2013)

[59] M. Morháč, J. Kliman, M. Jandel, L. Krupa, and V. Matoušek, *Study of Fitting Algorithms Applied to Simultaneous Analysis of Large Numbers of Peaks in γ-ray Spectra*. (2003)

[60] Kirkpatrick, J.M., Canberra Ind., Inc., Meriden, CT, Young, B.M., *Poisson Statistical Methods for the Analysis of Low-Count Gamma Spectra.* (2009)

[61] Wallace, B. *Using MLE and metaheuristics for event based neutron activation spectroscopy, CAP conference.* (Montreal, 2013).

Annexes

Bootstrap

The bootstrap is a statistical tool that can be used to evaluate confidence intervals and reduce the effects of outlying data points. The idea is simple; you can approximate the probability density function with the data set. In the case of the non-parametric bootstrap, all you have to do is to resample the data set to create a new data set. One usually resamples the same number of times, as there are data points in the original set. Obviously we resample with replacement as to avoid getting the same distribution all the time. Once a new data set is created, we can apply any function or estimator to the data and extract a quantity of interest, such as the mean, variance, MLE, etc. By repeating this process multiple times (~1000), we get a distribution of the quantity of interest. Analyzing this distribution gives us estimates on the variance of the estimator. We can also construct confidence intervals on the estimated parameters.

Aggregating bootstraps is a way of reducing the variance, and at the same time the risk, of a given estimator. A simple example would be to bootstrap an estimator and then use the mean or mode of the distribution function of estimators as our new estimator.

We will be using the non-parametric bootstrap in our analysis system. Given that the events that are detected are independent and identically distributed, we can use the bootstrap to produce new data sets. Given the computational stress this method produces, it is only used for the time-energy spectrum analysis. It could very well be used for the simple energy spectrum analysis but it kill the on the fly curve fitting. We will describe the methodology so that anyone wanting to reproduce it either on bi-dimensional data sets or one-dimensional datasets will be able to do so.

The first step is to create 1000 new data sets. We simply resample the events from the experimental data set. An important point is that we do not presume to know the form of de PDF. In reality we could have a good idea from our MLE fit of all known isotopes. Using a parametric bootstrap could lead to better estimates of the confidence interval but since we have not coded all known spectrums, we do not assume to know the PDF.

On each of these new data sets, we apply our analysis system and extract the amplitudes of the fitted kernel functions and their test statistic from the likelihood ratio test. We now have a distribution for these amplitudes and their corresponding test statistic. Constructing confidence intervals gives us an idea of the error on our estimators and their significance levels. The idea is to end up with an output that looks like this: Au activity = 100 +/- 10 Bq with 99.5 +/- 0.3 % confidence 99% of the time.

Random tree classification on field programmable gate arrays

Although not used in the final incarnation of the prototype, we thought important to leave the traces of the work that was done and planned to implement this technique. It was first thought of to discriminate between good and bad signals coming from the detector circuit. Our main concern was to throw out pile-up events. This occurs when two pulses overlap and the electronic circuit cannot discriminate between the two. For low count-rates, this is extremely rare and for this reason, this method was scrapped for our system. It would nonetheless be interesting to apply it to systems with high count-rates. It could also turn out that other gains could be made from this type of discrimination. For example: complex pulse shape discrimination for partial energy loss, Compton suppression, etc.

Classification is a sub-field of machine learning. Its goal is to predict a response function Y witch can take a value of 0 or 1 by looking at explanatory variables X. The algorithm that takes in X and spits out Y is called the classifier. For example we could have Y taking the value of cat or dog and X being the values of: pet weight, pet length, pet tail length, pet whisker length, etc. The classifier then takes all these parameters into consideration and predicts if the pet is a cat or a dog. To construct the classifier one generally needs a training set. This is a data set where the outcome is known for given predictor values. The classifier usually tried to see how similar an unknown case is to known cases and then uses it's training to decide how to classify it.

Random trees are a type of classifier. Their training is fairly straight forward: for each predictor covariant, we try to split the training set in two, where one part contains more 1s and the other more 0s. Using the pet example, we might split pet weight as greater than Z kg or lower than Z kg. We would choose Z as to maximize the splitting of cats and dogs. We then would do the same for the other predictor variables: pet length, tail length, etc. The predictor variable that bests splits the response variable Y is kept. The data is split into to subsets following this rule. We sub divide each subset using the same methodology until we get a good enough grouping of cats and dogs. The final result is a decision tree where each termination of the tree is a leaf. In each leaf there should be a population of mostly one or the other of the response variable. For a more detailed description of random trees and other classification methods we refer the reader to [22].

The main advantage of the random tree classifier is that it is easily implementable with logic gates or if/else statements. For FPGA implementation this is perfect; if/else statements can be executed within a single clock cycle. It is important to note that in our proposition, the classifier would be hard coded into a module that the top level could use to discriminate good events from bad events on the FPGA. The training of the classifier would have to have been done prior to the acquisition process.

How we proposed to do this is to have a large set of simulated spectrums either from a home brew simulation code or something like GEANT4. Simulation is essential because to train our data we must know if our pulse is from a good or bad event. It is either very difficult or impossible to know this from only observing nature. Once we have the simulated pulse, we sample it at the same frequency as the ADC would sample the real pulse. These points form a vector of digital values. We must extract quantities of interest from this vector. For example: the maximum value, the length of the pulse, the integral of the pulse, etc. In choosing these quantities, one must remember that any mathematical

operation will be done using logic gates and registers. Addition and subtraction are fairly simple and if/else statements are also fairly easy. Multiplication on the other hand can take up much more space on the FPGA. Trigonometric functions, exponential function and resampling methods are probably too costly. Any computation that requires many steps, such as searching through the vector is much slower and requires a clock cycle per step. Some computations, such as finding the maximum or integrating can be done as the pulse is being digitalized. An effort must be made to consider all tricks available from parallel computation and pipelining.

The first step would be to compute a large number of predictors and assign a computability score that takes into account the time it takes to compute and the space required on the FPGA to compute it. Using the predictors we could train a random tree. The R programing languages has multiple packages that can help for training classifiers. Once we have a trained tree, we only keep the predictor that it actually uses. All we have to do is to implement modules that calculate these predictors from the digitalized pulse vector and implement the if/else tree structure for our classifier.

An easy improvement would be to implement multiple (different) trees and aggregate their outputs. This is known as bagging or bootstrap aggregating. The if/else tree structure should be relatively small on the FPGA and the predictors would already be calculated. This means that we can take advantage of the inherent parallelization of the FPGA to compute multiple trees as fast as a single tree (at the expense of space).

On/Off Am-Be neutron source

The main drawback of an americium-beryllium neutron source is that it is constantly emitting neutrons. Storage of this type of neutrons is voluminous and heavy. We have proposed a new geometry that would enable turning on and off the emission of neutrons.

Given that alpha particles can travel several cm in air, we propose that we could separate the americium and the beryllium and insert a slicer between them. When the slicer is lifted, the alpha particles reach the beryllium target unobstructed. When the slicer is dropped, the alpha particles are stopped before neutrons are produced.

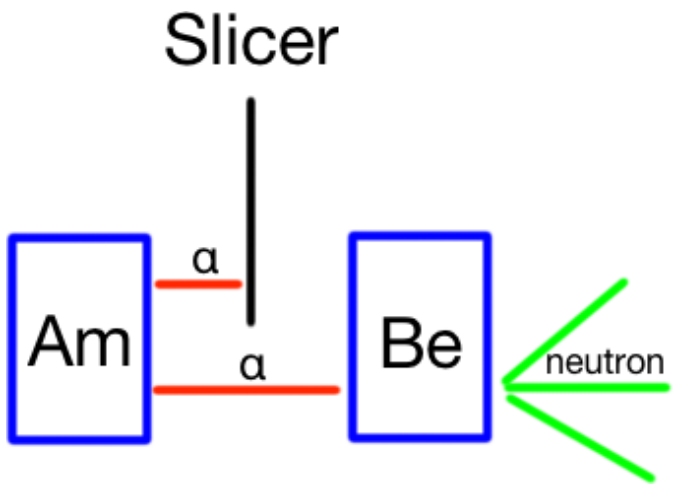

Figure 42. Drawing of the proposed geometry of the Am-Be source.

A mobile slicer is used to stop alpha particles before they reach the Beryllium target.

No evaluation was done on how this would affect the neutron yield for a given alpha activity. A GEANT4 simulation would be a good place to start to evaluate the efficiency of such a system.

User guide for the GREYSTAR Analysis Tool (GAT).

When booting up the system, give a few seconds for the bitmap to load onto the FPGA. Make sure the ADCs have power. It is necessary to start the EVAL-AD7352/AD7356/AD7357 software from Analog and to turn on both ADCs. You can close this software once you have turned on the analog to digital converters.

Launch the GREYSTAR Analysis Tool executable or compile it from the source code. To start acquisition, press the start button. This sends a signal to the FPGA to start communication with the computer. Press the stop button to stop the acquisition. Pressing the stop button resets the clock and writes the histogram to the corresponding file. The clear button erases the contents of both the time-energy data file and the histogram file for both detectors.

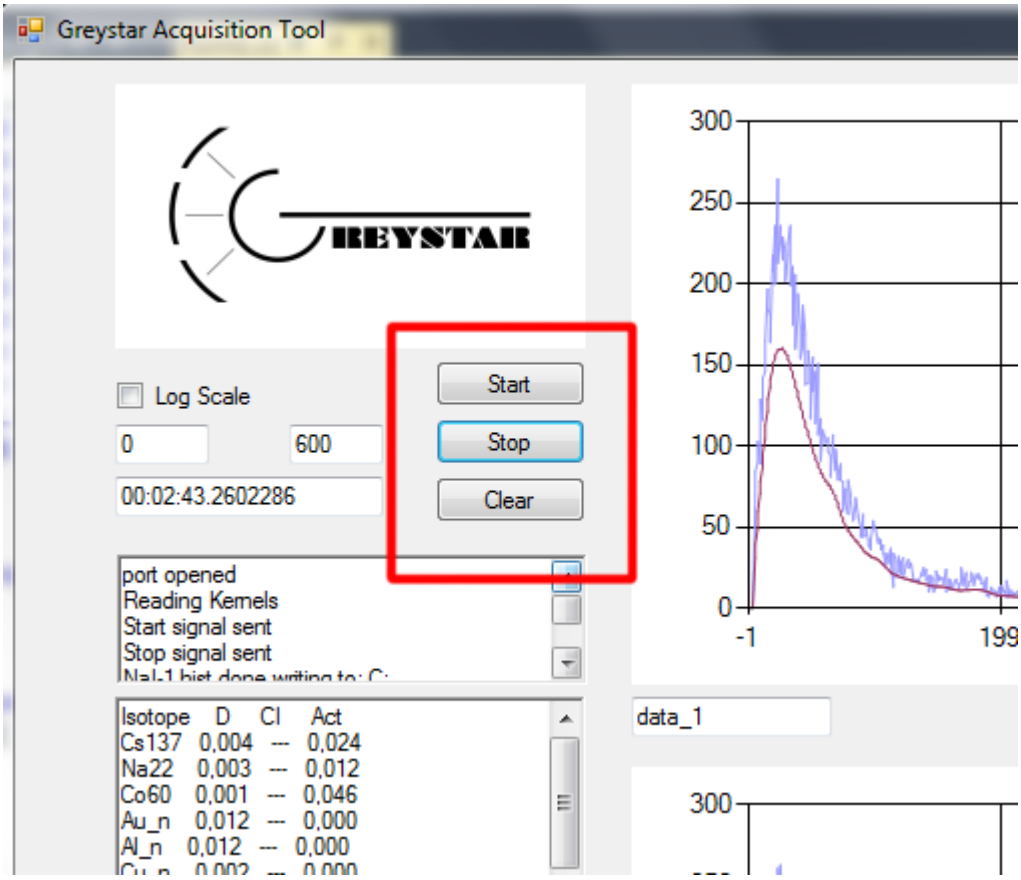

Figure 43. The start, stop and clear buttons on the GUI.

If you want to acquire multiple runs on a single file, start the acquisition then stop it. Don't clear the contents and start the acquisition again. The clock will be reset to zero. The histogram will continue from where it stopped. The time-energy data will continue recording but the timestamp will be reset to zero.

The histogram is updated every 500 milliseconds. The analysis function is called every 2 minutes. During this time the system still records incoming data but a slight lag will be observed on the GUI. By checking the “log” box, the histogram is plotted on a log scale. The user can define the channel range for the histogram display. Input the minimum and maximum of the desired range in the appropriate field.

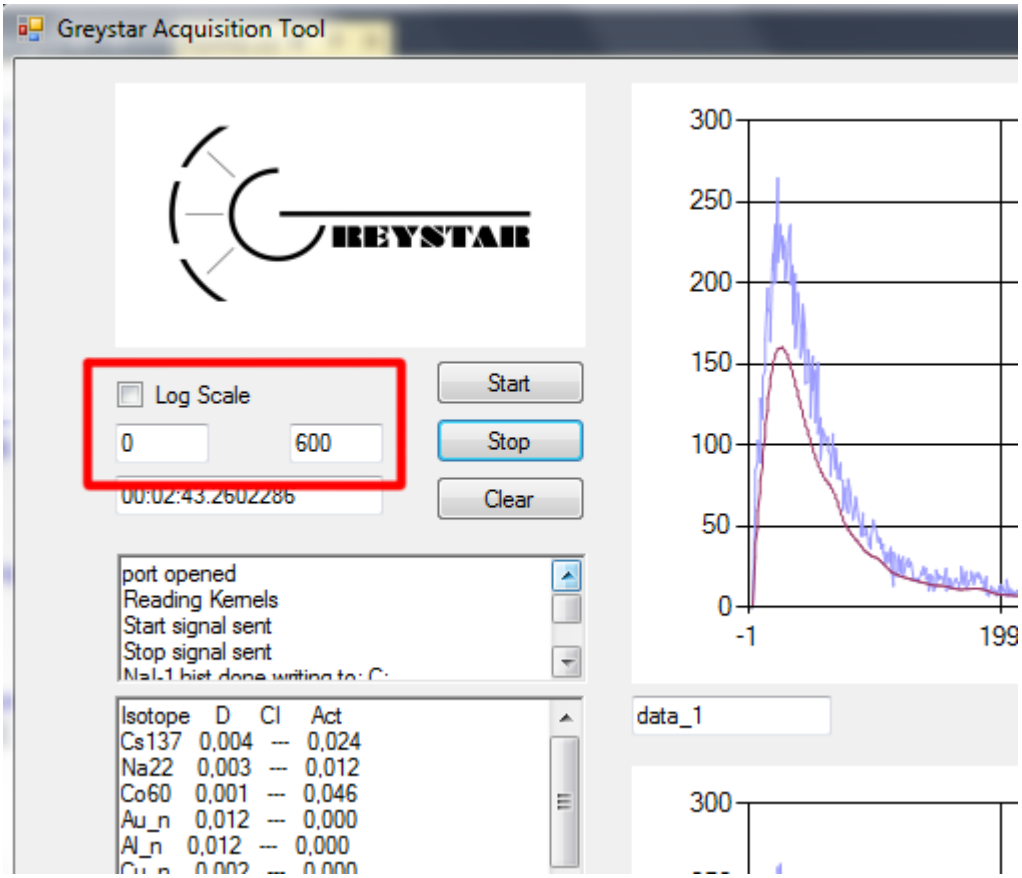

Figure 44. The log check box and the range selection fields.

You can edit the file name where the time-energy data is recorded. Edit the appropriate field.

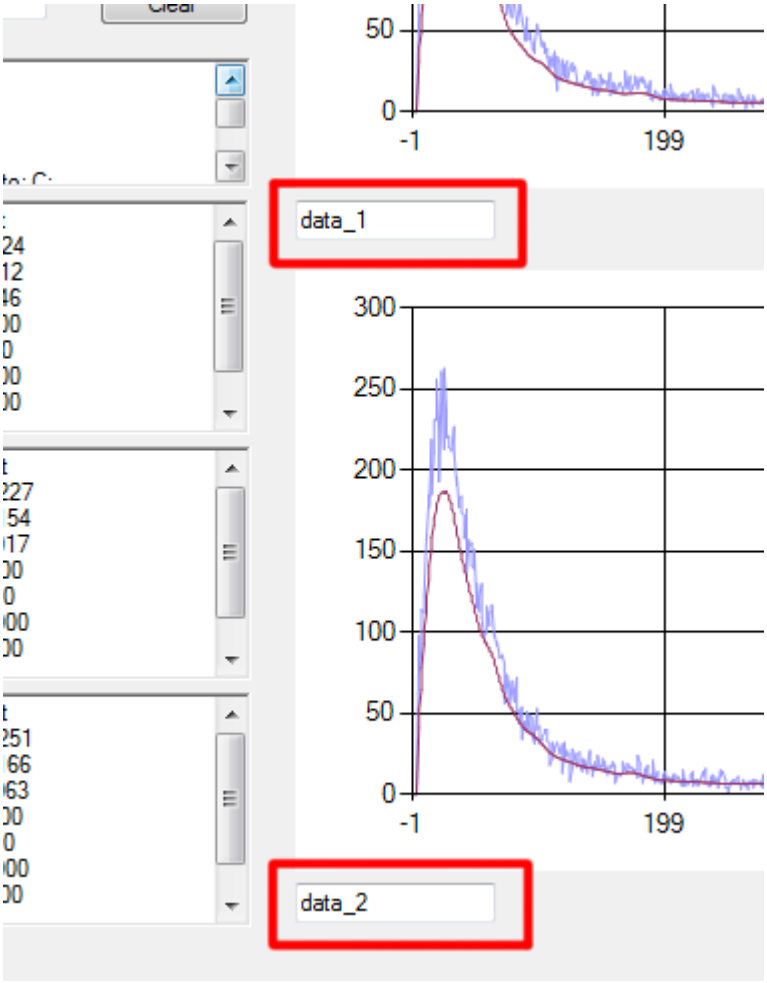

Figure 45. File name fields on the GUI.

There are four windows that communicate information to the user. The first window communicated information about the communication with the FPGA. Messages that are displayed on this window warn the user when the COM port has been opened, when events are dropped due to corrupted packets, when start and stop signals, when kernel functions are loaded and when histograms are written to file.

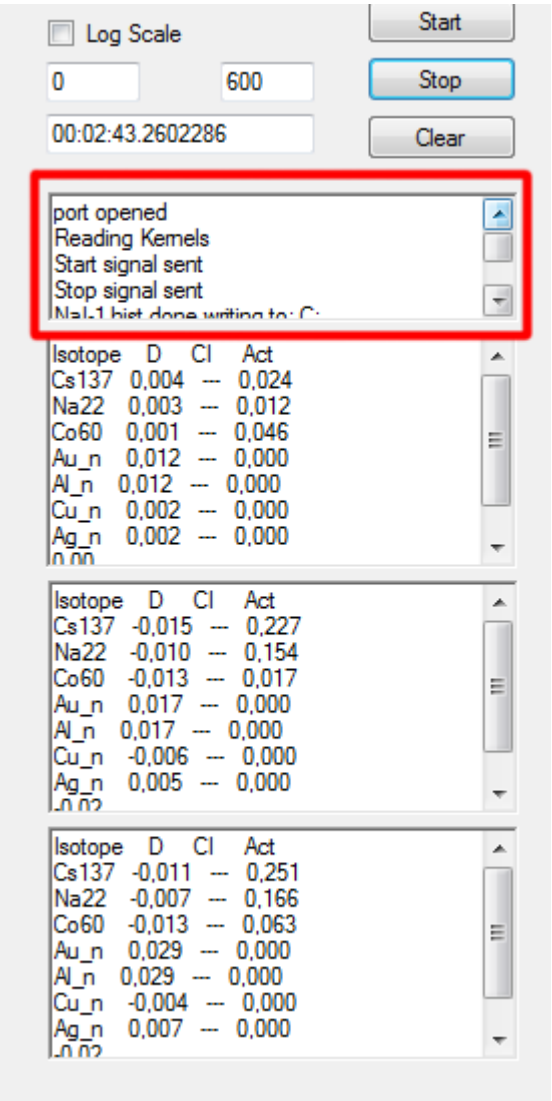

Figure 46. System communication window 1.

The three other communication windows are used to inform the user about the statistical analysis done in real time. The information displayed here is the chi-squared test statistic D, the confidence interval and the estimated activity of each isotope in the library. Windows 2 and 3 are associated to detectors 1 and 2 respectively. Window number 4 is the statistical combination of both detectors. It is possible to have statistical significance sooner by integrating both analyses. When an isotope is detected with greater than 99.9% confidence, the background of the window turns light blue. When a lack of fit is detected, an error message is produced in the window and the background turns orange. The number of channels with and extreme lack of fit is indicated at the bottom of each message. When this number is greater than 5, we consider it is a bad fit. This value can be modified in the source code.

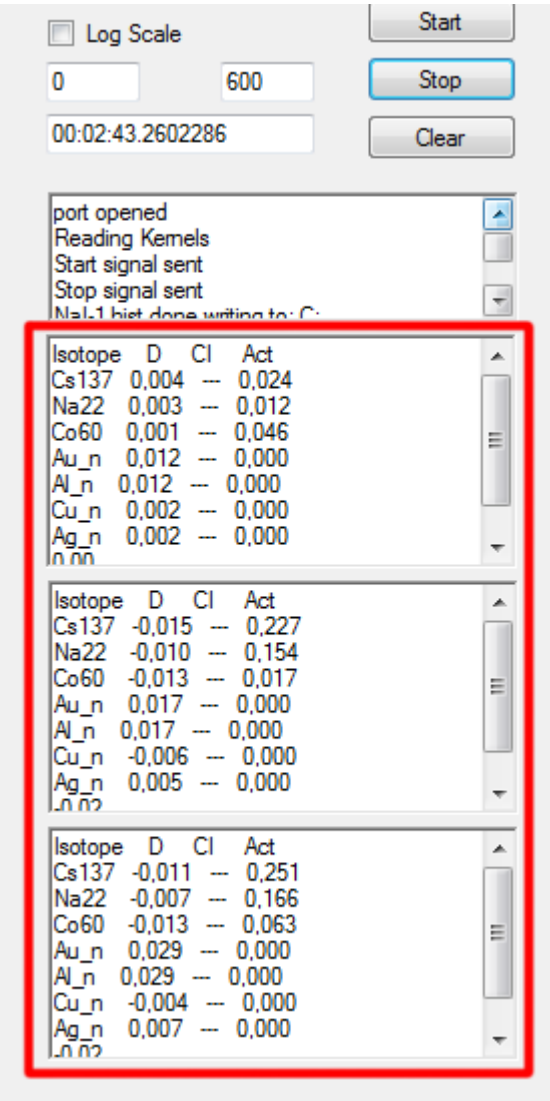

Figure 47. Communication windows 2, 3 and 4.

Properties of the materials used in the benchmark tests

The Au sample:

Mass: 7.08 g.

Composition: 99.99% ^{197}Au.

Activation reaction: ^{197}Au + n -> ^{197m}Au.

Half-life: 7.73 sec.

Main gamma energy: 279 keV.

Capture cross-section: 98.6 b.

The Ag sample:

Mass: 29.15 g.

Composition: 25.9% ^{107}Ag, 24.1% ^{109}Ag, 34.5% ^{63}Cu, 15.5% ^{65}Cu.

Activation reaction: ^{107}Ag + n –> ^{108}Ag, ^{109}Ag + n -> ^{110}Ag.

Half-life: 2.37 min (^{108}Ag), 24.6 sec (^{110}Ag).

Main gamma energy: 633 keV (^{108}Ag), 657 keV (^{110}Ag).

Capture cross-section: 19.5 b (^{107}Ag), 43.8 b (^{109}Ag).

Note: although this reference material contains a significant amount of Cu, the large difference in neutron capture cross-sections makes this an acceptable proxy for pure Ag.

The Cu sample:

Mass: 2.4 kg.

Composition: 69.1% ^{63}Cu, 30.8% ^{65}Cu.

Activation reaction: ^{63}Cu + n -> ^{64}Cu, ^{65}Cu + n -> ^{66}Cu.

Half-life: 12.7 h (^{64}Cu), 5.12 min (^{66}Cu).

Main gamma energy: 1346 keV (^{64}Cu), 511 keV* (^{64}Cu), 1039 keV (^{66}Cu).

Note: ^{64}Cu can decay through β^+ decay, this leads to a secondary emission of two 511 keV gammas when the positron is annihilated. The branching probability that leads to this is 37 times greater than for the 1346 keV gamma (17% vs 0.47%). This leads to a prominent 511 keV peak in the ^{64}Cu spectrum (two gammas are emitted at 37 time the probability, the activity is 74 time greater).

Capture cross-section: 3.12 b (^{63}Cu), 0.66 b (^{65}Cu).

The Al sample:

Mass: 700 g.

Composition: 100% ^{27}Al

Activation reaction: ^{27}Al + n ->^{28}Al.

Half-life: 2.24 min.

Main gamma energy: 1778.9 keV.

Capture cross-section: 231 mb.

The ^{137}Cs source :

Activity: 2x 5 µCi.

Gamma energy: 661 keV.

The ^{22}Na source:

Activity: 1 µCi.

Gamma energy: 511 keV, 1274 keV.

The ^{60}Co source :

Activity : 0.5 µCi.

Gamma energy: 1.17 MeV, 1.33 MeV.

The ^{54}Mn source :

Activity : 1 µCi.

Gamma energy: 834 keV.

Screen captures of the benchmark tests.

The screen captures were taken after the system returned the fitted values, as such we should not concern ourselves with the time indicated on the clock as it may or may not include the fitting time or additional acquisition time that is not considered for the fit. We also note that the coloured background feature was not implanted for every isotope. As such, we recommend that the reader rely on the numerical output of the user interface instead. Both detector histograms and report windows are present (see user guide in Annex).

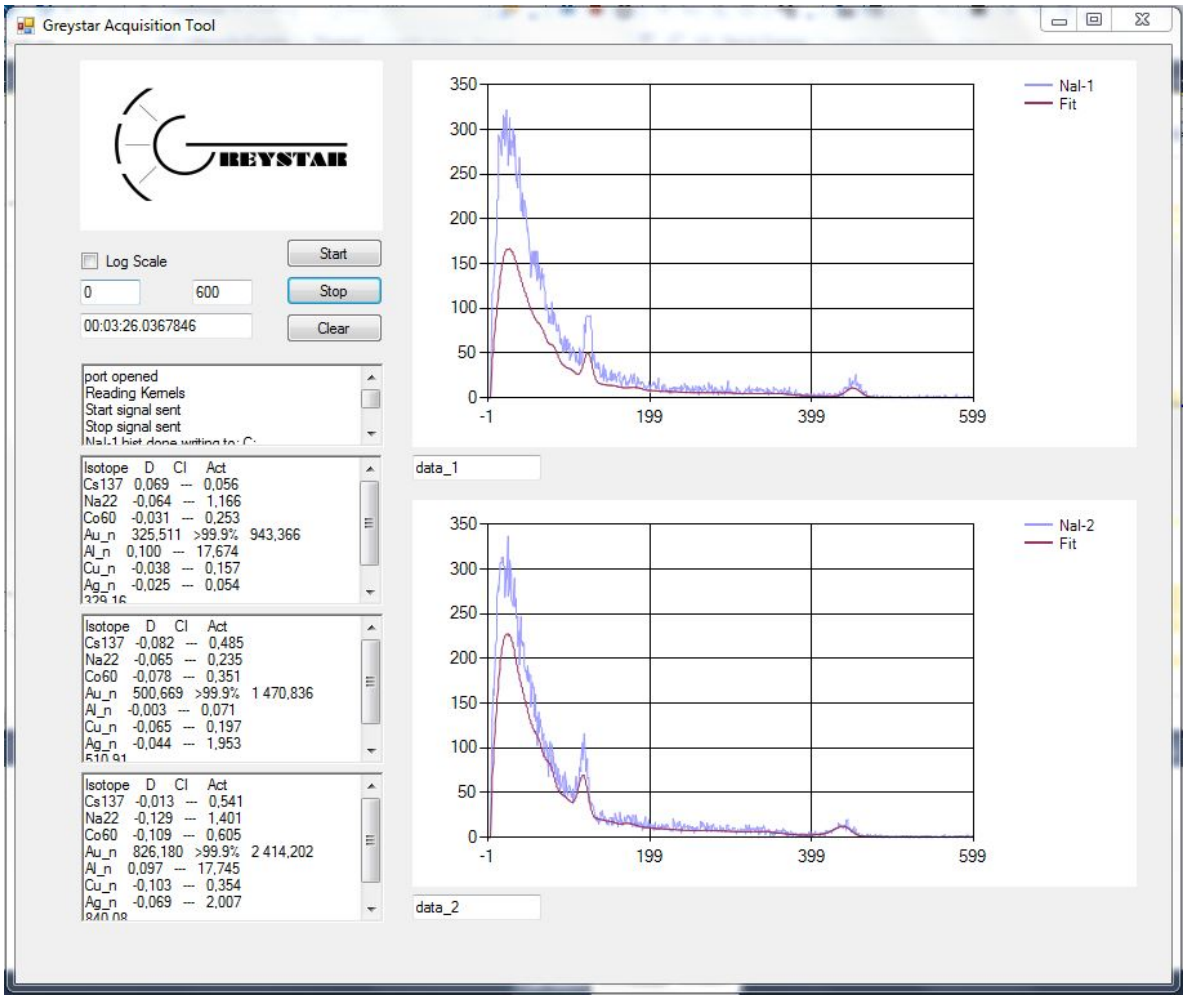

Figure 48. Au 3 minute spectrum and fit.

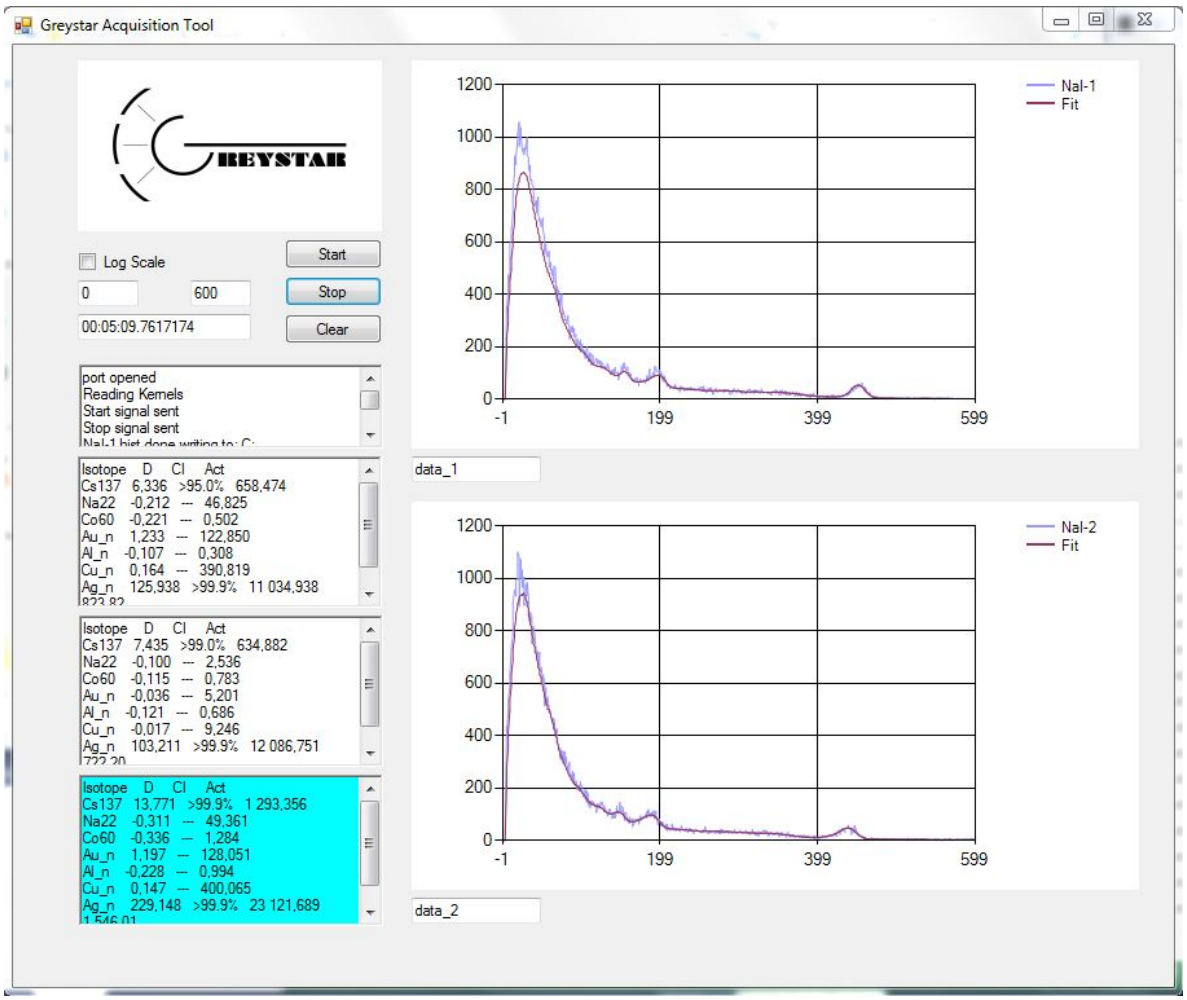

Figure 49. Ag 6 minute spectrum and fit.

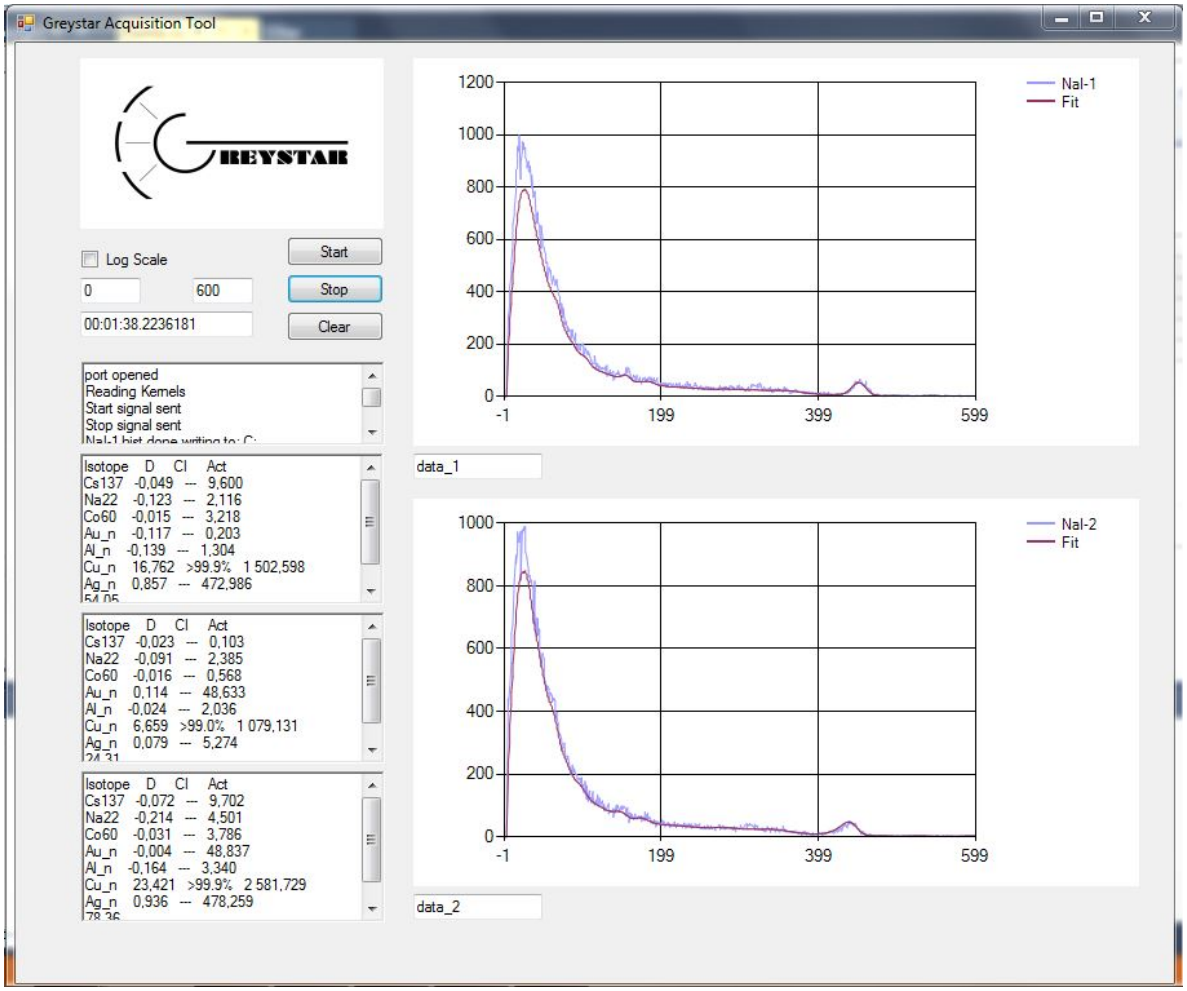

Figure 50. Cu 6 minute spectrum and fit.

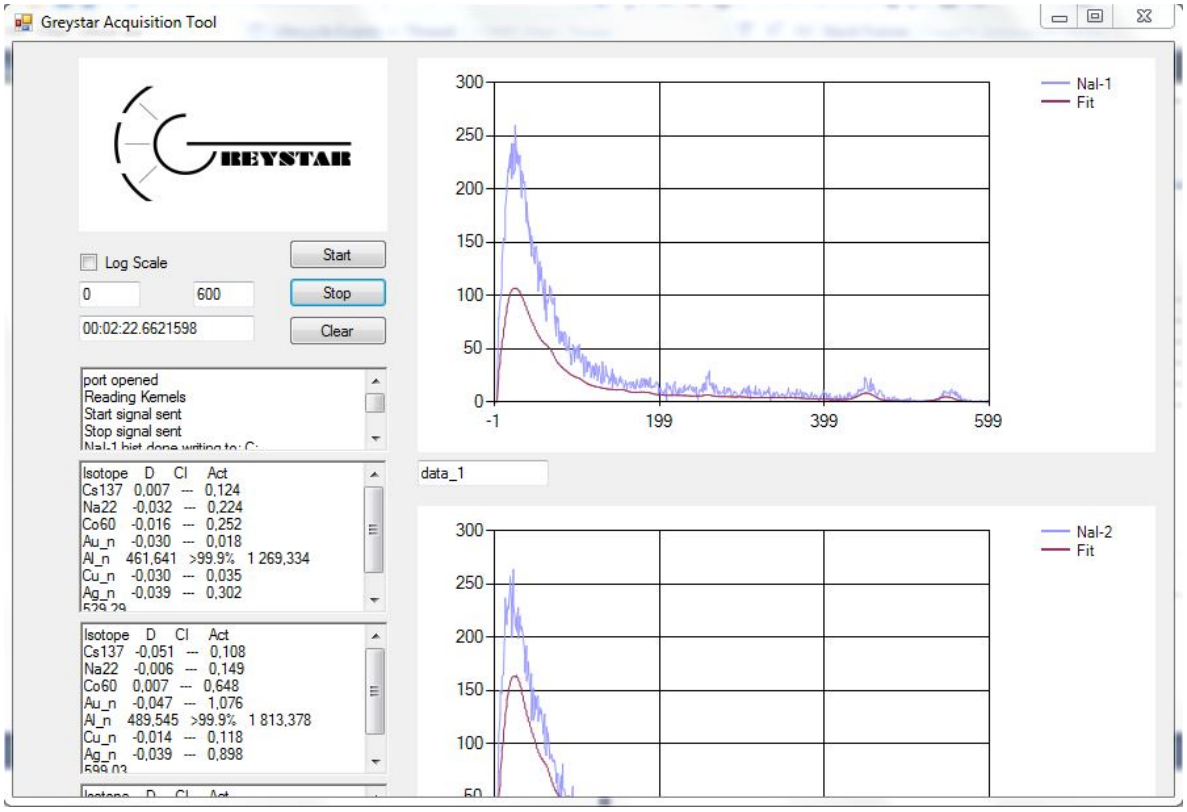

Figure 51. Al 3 minute spectrum and fit.

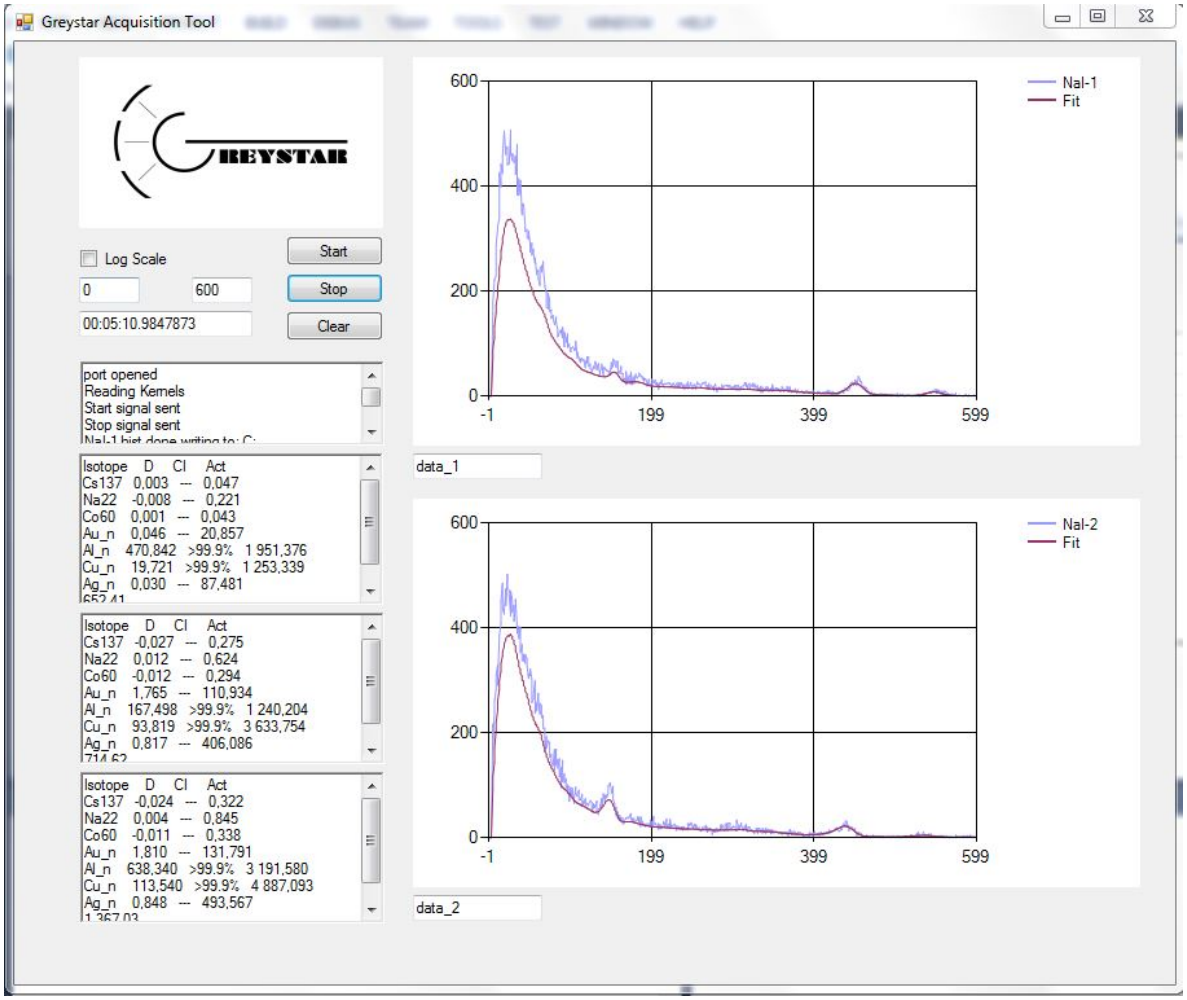

Figure 52. Al and Cu 3 minute spectrum and fit.

Block diagram of the acquisition system.