UNIVERSIDADE D
COIMBRA

Bernardo Cabrita Figueiredo

Neutron-gamma and pile-up discrimination aiming at plasma real-time control in nuclear fusion experiments

Dissertação no âmbito do Mestrado Integrado em Engenharia Física - Instrumentação, orientada pelo Doutor Nuno Sérgio Castelo Branco da Cruz e co-orientada pelo Doutor João Manuel Rendeiro Cardoso e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Fevereiro de 2023

UNIVERSIDADE D
COIMBRA

Neutron-gamma and pile-up discrimination aiming at plasma real-time control in nuclear fusion experiments

Supervisor:

Dr. Nuno Sérgio Castelo Branco da Cruz

Co-Supervisor:

Prof. Dr. João Manuel Rendeiro Cardoso

Jury:

Prof. Dr. Custódio Francisco Melo Loureiro

Prof. Dr. José Paulo Pires Domingues

Dissertation submitted in partial fulfillment for the degree of Master of Science in Engineering Physics.

Coimbra, March 2023

Acknowledgments

This work was funded by Portuguese National Funds through Fundação para a Ciência e Tecnologia, in the scope of the project UIDB/04559/2020. IPFN activities received financial support from Fundação para a Ciência e Tecnologia through projects UIDB/50010/2020 and UIDP/50010/2020.

Quero agradecer aos amigos que me apoiaram ao longo desta longa jornada, em bons e maus momentos. Ao meu irmão António e a cada um dos meus primos. Aos meus pais por financiarem a minha existência. À minha família. À Dra. Rita Pereira pelos primeiros passos neste projecto. Ao co-orientador Dr. João Cardoso por inspirar a organização que acabou por ser responsável pela escrita desta tese. À Universidade de Coimbra e ao Instituto Superior Técnico pela oportunidade de me provar academicamente.

Ao Nuno, um enorme obrigado por tudo. Por todas as reuniões, orientação e partilha de informação, pela paciência e pela inspiração.

À Rachel, nada disto teria sido possível sem a tua ajuda. Tornaste-me uma pessoa diferente.

À minha avó, que tanto queria ver este fim.

Resumo

A fusão nuclear é um candidato potencial para a produção de energia no próximo século. Um desafio relevante no seu desenvolvimento é simultaneamente científico e técnico: como controlar o plasma em ambientes destrutivos. O controlo do plasma em tempo real depende de enormes quantidades de informação proveniente de ferramentas de diagnóstico. Um destes diagnósticos é o perfil de emissividade de neutrões, que requer o processamento de centenas de milhares de pulsos por segundo. Os processos geradores de neutrões na fusão estão associados à emissão de raios gama, o que torna necessária a discriminação dos neutrões dos raios gama num detector. Num ambiente de alta taxa de contagem, a probabilidade de sobreposição de pulsos, ou de empilhamento, pode ser uma proporção significativa de eventos. A rejeição de empilhamento e a compensação estatística podem mitigar o seu efeito estatístico. Os algoritmos de PSD (*Pulse Shape Discrimination*/Discriminação da forma do pulso) discriminam os neutrões dos gammas, mas os algoritmos precisos são geralmente demasiado lentos ou computacionalmente intensivos para se aplicarem ao controlo de plasma em tempo real, em particular para discriminar os tipos de partículas em sinais empilhados.

A simulação de Monte-Carlo é utilizada para gerar formas de onda personalizadas, com base numa equação semi-empírica conhecida que parametriza formas de pulso de cintilação de acordo com uma taxa de decaimento exponencial e tempo de subida exponencial. Esta forma de onda bem estudada é utilizada num algoritmo de discriminação da forma de onda baseado em TM (*Template Matching*), misturado com uma recente abordagem PSD offline, classificando cada evento de empilhamento contra todas as combinações possíveis de modelos de neutrões-gama. Descobriu-se que o algoritmo pode executar adequadamente o PSD em pulsos de empilhamento, usando dados reais. Com a parte computacionalmente intensiva do algoritmo a ser executada na calibração, tem o potencial para utilização em tempo real.

Abstract

Nuclear fusion is a potential candidate for energy production in the coming century. One relevant challenge in its development is both scientific and technically demanding: how to control plasma in destructive environments. Real-time plasma control relies on massive amounts of information from diagnostics tools. One of these diagnostics is the neutron emissivity profile, which requires the processing of hundreds of thousands of pulses each second. Neutron-generating processes in fusion are associated with the emission of gamma rays, which makes discrimination of neutrons from gammas in a detector a necessity. In a high count-rate environment, the probability of superimposed pulses, or pile-up, can be a significant proportion of events. Pile-up rejection and statistical compensation can mitigate its statistical effect. Pulse-Shape Discrimination (PSD) algorithms discriminate neutrons from gammas, but accurate algorithms are generally too slow or computationally intensive to apply for plasma control in real-time, in particular for discriminating particle types in piled-up signals.

Monte-Carlo simulation is used to generate custom waveforms, based on a known semi-empirical equation that parametrizes scintillation pulse shapes according to an exponential decay rate and exponential rise time. This well-studied waveform is used in a template-matching pile-up separation algorithm, mixed with a recent PSD offline approach, by rating each pile-up event against all possible neutron-gamma template combinations. It is found that the algorithm can adequately perform PSD in piled-up pulses using real data. With the computationally expensive portion of the algorithm being performed in calibration, it has the potential for real-time usage.

"Roma uno die non est condita."

Mr. 9

Contents

List of Acronyms

Euratom	European Atomic Energy Community
ITER	International Thermonuclear Experimental Reactor
JET	Joint European Torus
DEMO	DEMOnstration Power Plant
IPFN	Instituto de Plasmas e Fusão Nuclear
IST	Instituto Superior Técnico
RNC	Radial Neutron Camera
RGRS	Radial Gamma-Ray Spectrometer
DAQ	Data Acquisition System
FPGA	Field-Programmable Gate Array
sCVD Detector	Single-crystalline Chemical Vapor Deposition Detector
PMT	Photomultiplier
PSD	Pulse Shape Discrimination
SDD	Silicon Drift Detector
CCD	Charge Coupled Device
Eg	Energy Gap
APD	Avalanche PhotoDiode
SSPM	Solid-State Photomultiplier
MPPC	Multi-Pixel Photon Counters

FWHM	Full Width at Half-Maximum
SCA	Single Channel Analyser
PHA	Pulse Height Analyser
MCA	Multi-Channel Analyser
ADC	Analog to Digital Converter
RMS	Root-Mean-Square
LSB	Least Significant Bit
DPU	Digital Processing Unit
NIM	Nuclear Instrument Module
CAMAC	Computer Automated Measurement and Control
VME	Virtual Machine Environment
PCI	Peripheral Component Interconnect
ATCA	Advanced Telecommunications Computing Architecture
DSP	Digital Signal Processor
MCU	Microcontroller Unit
ASIC	Application Specific Integrated Circuit
GPU	Graphical Processing Unit
MARTe	Multithreaded Application Real-Time executor
GAM	Generic Application Modules
I/O	Input/Output
FOM	Figure of Merit
CDF	Cumulative Density Function
PDF	Probability Distribution Function
SNR	Signal-to-Noise Ratio
a.u.	arbitrary units

List of Figures

List of Tables

1 Introduction

Climate neutrality is one of the main challenges humankind currently faces. The transition into a green society requires structural change involving the global participation of all economic sectors. In particular, as technology advances, population grows and urbanization spreads, the global electricity demand has steadily been increasing over the past few decades. The only recent year showing a decrease in global energy consumption was 2020, mostly due to the economic slowdown caused by the response to the COVID-19 epidemic justifying the small 1% decrease. When it comes to overall energy consumption, energy demand is outpacing what renewable energy sources can provide. In spite of renewable sources' impressive growth, 40% of 2022 global energy demand growth is expected to be met by increasing fossil fuel-based electricity production, mostly by increased coal firing [1]. The European Union has focused on emancipating economic growth from increased carbon emissions, achieving the self-imposed sustainability goals [2], but the overall energy demand increase highlights the global nature of the issue, in which developed countries have a head-start in economic growth, and are afforded the opportunity to change the future paradigm of energy consumption.

Further investment and improvements in renewable energy production are therefore a necessity. With current technology, fully solving the energy storage problem of current renewable methods seems unlikely [3][4], so what is needed to bridge their daily and seasonal variable output issue is a carbon-free, sustainable, safe energy source that can replace the function coal plants currently fulfill. That is the potential nuclear fusion offers. Depending on future energy costs, it is expected to be a big player in the carbon-free energy source ecosystem for the latter half of the 21st century, especially in countries with limited access to other sources [5].

Uranium-235 is the principal fissile isotope of Uranium, the main nuclear fission fuel. At around 0.7% of natural abundance, it needs isotope separation, or enrichment, to produce usable quantities at higher percentages, up to 5% [6]. The same process is a necessary part of nuclear weapon proliferation, which added to the mediatism of disasters such as Chernobyl and Fukushima, creates a climate of social skepticism. Figure 1.1 shows the progressive net change in nuclear energy production capacity. It shows a phasing out of nuclear power in Europe, with the rest of the world pushing the overall growth into the positive.

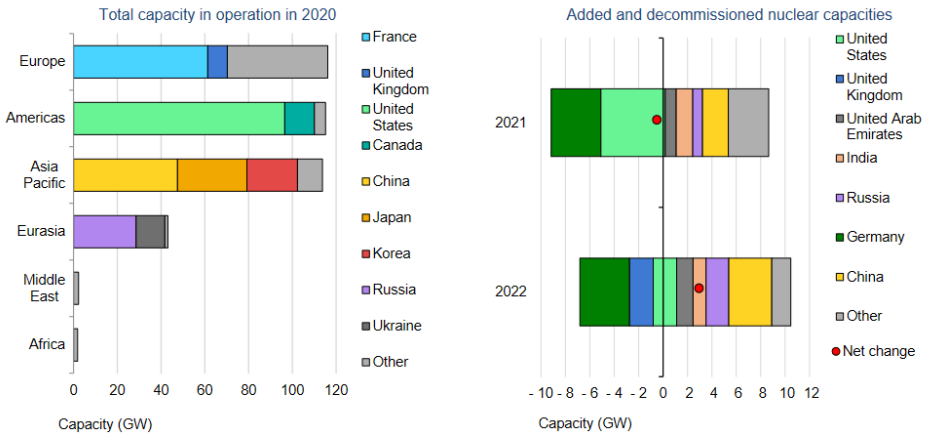


Figure 1.1: Left: Total nuclear energy production capacity in 2020. Right: Added and decommissioned nuclear capacities in the 2021 to 2022(planned) period. Image from [1]

The Levelized Cost of Generating Electricity (LCOE), a metric that evaluates the cost of producing energy, from extraction to the resource expenditure in building, producing, and decommissioning a power plant, not including transmission/distribution, shows Nuclear fission as a very competitive form of carbon-free energy production, see fig. 1.2. Overall, first-of-a-kind projects have inflated costs. At the same time, the long-term operation of nuclear fission facilities is unmatched in LCOE [7]. In terms of normalized energy accident risk, meaning how many fatalities, and how frequent the accidents are expected per TWh of energy produced, nuclear fission is one of the safest energy production methods, even when including the catastrophic meltdowns[8].

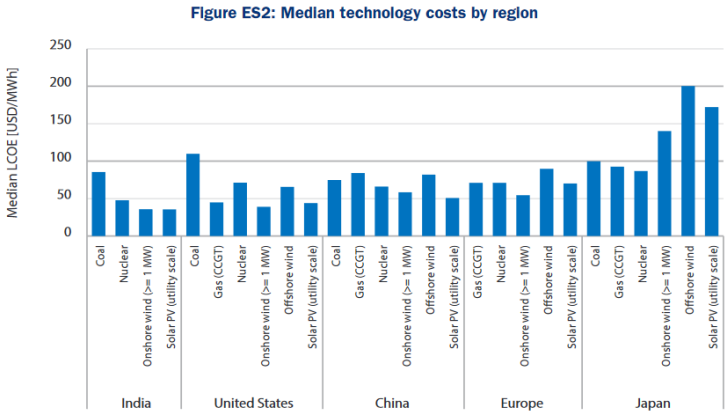


Figure 1.2: Median Levelized Cost of Generating Electricity in several regions. Image from [7]

Nuclear fusion, when compared to nuclear fission, is placed as a first-of-a-kind research project, with the potential to create an abundant source of carbon-free energy, with baseline operation capacity. The fuel is virtually limitless and its extraction doesn't come with the environmental impact of mining operations that coal, gas, and nuclear fission have. Compared to fission, there is no risk of a meltdown and a limited risk of nuclear weapon proliferation. In addition, there are no long-term radioactive waste products [9].

When faced with a similar advantageous profile to nuclear fission, fusion theoretically fixes most of the sources of nuclear skepticism. The ambitious challenge here posed is then the recreation of conditions similar to a star, at a local scale. In other words, developing the technology to support the confinement of burning plasma (magnetic or otherwise) as an energy source.

The creation of a fusion power plant stands furthest on the roadmap of a long-term international project in fusion research, organized by the European Atomic Energy Community (Euratom)[10], one of the founding treaties of the European Union. Euratom's membership is shared with the EU, having additional bilateral agreements with "Argentina, Australia, Brazil, Canada, India, Japan, Kazakhstan, Russia, South Africa, South Korea, Ukraine, the United States, and Uzbekistan on nuclear fission/fusion research or peaceful use of nuclear energy and nuclear safety"[11]. In 2006, Euratom, the United States, the Russian Federation, Japan, China, South Korea, and India signed an agreement to develop and build the International Thermonuclear Experimental Reactor (ITER), an experimental reactor aimed at testing the feasibility of this energy source. While ITER is expected to start operating in 2026, reaching full capacity by 2035, electricity production is not a part of its design. Commercial fusion energy isn't expected to happen before the 2nd half of the 21st century with the development of ITER's successor, the DEMOnstration Power Plant(DEMO).

In the European Union, United Kingdom, Switzerland, and Ukraine, fusion research is organized by EUROfusion, a consortium funded by Euratom. The participation of Portugal in EUROfusion is ensured by Instituto de Plasmas e Fusão Nuclear(IPFN), a research unit of Instituto Superior Técnico (IST). EUROfusion manages the Joint European Torus(JET), a magnetic plasma containment experiment which can be considered the main design predecessor for ITER[12].

1.1 Motivation

Harnessing fusion is no easy feat, as it requires the localized creation and containment of plasma hot enough for fusion to occur.

Plasma behaves much like a fluid made of unbound charged particles and is mathematically described as such, composed of unbound electrons and the ions they were ionized from. Individual particle interactions in plasma number too high for a classical approach to describing its mechanics to succeed. This means that describing its state is done in terms of statistical parameters: charge and density, mean velocity and current, pressure, temperature, and heat flux. Unlike other fluids, the harsh, destructive conditions of burning plasma pose a challenge if outright interacting with it - its temperature is high enough to melt any material used to contain it, destroying the vessel and polluting the plasma with impurities, so an alternative containment method is required; in the same vein, observation requires robust tools to probe the plasma reaction, and actuators to control the plasma. The set of metrology tools and algorithms used to measure and calculate a certain parameter, which can be used to describe the plasma, is called diagnostics. The electromagnetic nature of plasma leads to key diagnostics techniques being centered around acquiring appropriate parameters, such as flux, currents, and electromagnetic fields, to describe the internal reaction, behavior, and properties. However, full information on the internal state has to be inferred by measuring the conditions that create outgoing radiation, employing other techniques for the sake of completeness, redundancy, and calibration. These include "approximately 45 measurement systems (...), broadly broken up into the following categories, Magnetic, Optical, Bolometric, neutron and particle, spectroscopic and Edge systems" [13].

Two of the diagnostics in the ITER project are the Radial Neutron Camera (RNC) and the Radial Gamma-Ray Spectrometer (RGRS), represented in figure 1.3. They are based around detecting gamma rays and neutrons at high count rates. Gamma rays are highly energetic photons - they interact very easily. On the other hand, neutrons are massive uncharged particles, which means they aren't contained by the magnetic confinement field. Because of this, gamma detection is often polluted by neutrons, and vice-versa.

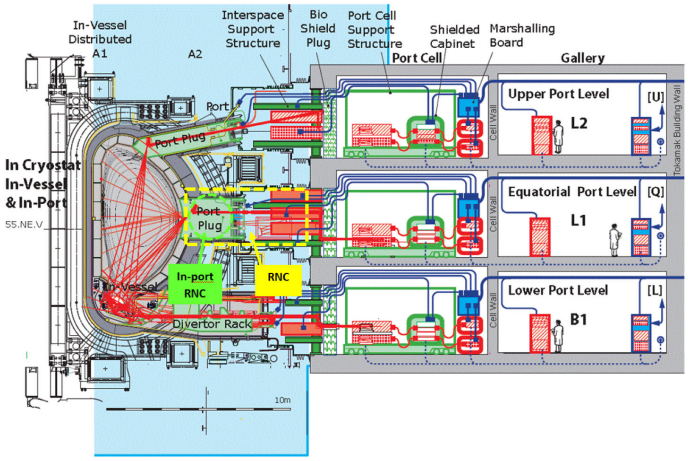


Figure 1.3: Poloidal view of ITER. The RNC is designed to measure the neutron emissivity profile, used in advanced control systems. The RGRS detects runaway electrons and measures the confined alpha profile. Image from [14]

This dissertation is contextualized within IPFN's research, which focuses on Data Acquisition Systems(DAQ) and Data Processing for JET and ITER diagnostics, in particular for the JET EP2 gamma-ray spectrometer and ITER's RNC[15]. The RNC is designed with detectors placed at 26 lines of sight, each acquiring up to a peak rate of 2 million events per second. These produce a massive amount of data, up to 0.5GB/s for each channel, creating additional challenges in data compression and processing [16]. This data is used to calculate a plasma parameter, the plasma emissivity profile, which is used as a part of the control cycle.

Especially at high count rates, it sometimes happens that two detection events occur near-simultaneously, generating overlapping signals. This is called pile-up. Traditionally, piled-up pulses were discarded and statistically compensated, but modern pile-up resolution techniques have been able to retrieve more and more information from these pulses, which can provide a more accurate measurement of the fusion reaction.

The motivation behind resolving the pile-up of neutron and gamma pulses lies in the improvement of gamma and neutron spectra and emissivity profiles for plasma control, in particular considering the high number of particles involved in fusion.

1.1.1 Case Study: Radial Neutron Camera

Cruz et al [15] present a protoype for the control and data aquisition design for ITER's RNC. The Radial Neutron Camera's 26 lines of sight produce up to 52 million events per second in total, to be processed in real-time cycles of 10ms in order to count incoming neutrons at each line of sight. Together with magnetic surface data derived from magnetic diagnostics, the neutron emissivity profile is produced. The processing tasks required to create the neutron emissivity profile are shown in image 1.4. The processing chain that culminates in the measurement of particle counts, shown in chapter 2.6, is summarized in the first block of this diagram. A KC705 Field-Programmable Gate Array (section 2.6.4) is used to process data from each line of sight into a count value, which then feeds forward into the emissivity profile calculation algorithm. Each of these is part of a 10ms control cycle, which means the calculation is redone at that rate. At a peak rate of 2MEvents/s, this sets an average of 20 thousand events processed in each control cycle.

A candidate detector for the RNC in-port subsystem, the closest to the containment, thus subject the harshest conditions, is the CIVIDEC B7 Single-Crystal Diamond Detector (section2.5.2), due to its radiation hardness and tentative thermal fatigue capabilities[14]. From this information, the peak amount of pile-up for these conditions can be estimated. For the pulse length, it's hard to determine exactly the precise start and end point of a pulse (section 2.6.3), but a reasonable estimate comes from the typical Full Width at Half-Maximum (FWHM) for these types of systems (section 2.8.4). For sCVD Diamond detectors coupled to a charge amplifier, the FWHM ranges $8 - 180ns$, according to the CIVIDEC catalogue[17]. For these values, from equation 2.11 the estimated first-order pile-up percentage at peak rates ranges $1.6 - 30.2\%$.

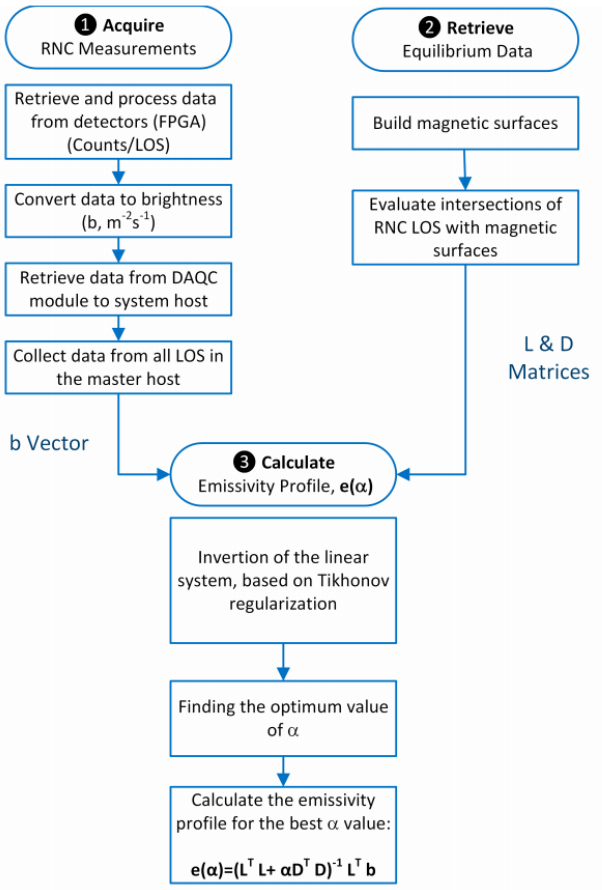


Figure 1.4: RNC Processing tasks for the neutron emissivity profile. From [15].

1.2 Objective: Neutron-Gamma Discrimination

The goal of this thesis is to improve neutron-gamma particle discrimination in environments where both particles are present and pile-up is frequent, namely in ITER/JET-like Tokamaks. The main objectives are: firstly, the creation of a tool for testing, capable of creating simulated synthetic pulses and pulse trains, in accordance with desired event rates, energy spectrum, and particle type, according to the detection system characteristics, including detector type, data acquisition sample rate, and noise; secondly, using this tool to develop and explore existing detection and pile-up separation algorithms, how they can be implemented simultaneously with pulse shape discrimination algorithms, taking into account separation efficiency and energy accuracy; thirdly, completing the acquisition and processing cycle in order to validate the whole process. This cycle starts with a known/input energy spectrum, and generates individual pulses and timings, creating a pulse train. This pulse train is processed with the developed pile-up separation and pulse shape discrimination algorithms, attributing to each detected pulse an energy and particle type. This produces a spectrum that can be compared to the original for valida-

tion. Finally, evaluate the real-time performance of these algorithms and if they can be viably implemented in a real-time fashion, in accordance with the strict ITER RNC requirements. The RNC is composed of 26 lines of sight measurements, each subject to a peak rate of 2MEvents/s. It needs to be able to process these, including pulse shape discrimination and pile-up rejection. This processing needs to be fast enough to fit within a 10ms fast control cycle[16].

1.3 Thesis outline

This dissertation starts by contextualizing fusion reactors and the physics behind fusion, tracing back to the physical source of pile-up in chapter 2.

Throughout chapter 2, it follows the detection and acquisition chain until a digitized pulse is obtained. Then, it provides an introduction to pulse processing, its implementations, and its objectives. Afterwards, it summarizes some of these processing algorithms, and how they can be used to separate overlapping signals and to discriminate neutron from gamma signals.

Afterwards, Chapter 3 entails the development of a simulator capable of synthesizing a pulse train according to desired pile-up amounts, as described in the previous section, and thereafter developing and testing the neutron-gamma discrimination algorithm in chapter 4.

Chapter 5 evaluates the performance of the developed algorithms in terms of accuracy and speed, the possible applications of this work, and how to develop it further.

2 State of the Art

2.1 Basic plasma physics

The principle behind using nuclear reactions to produce energy is the mass-energy difference of a nuclear reaction's products and reagents. Based on Einstein's formula, $E = mc^2$, a small decrease in mass when two light nuclei fuse, or when heavy nuclei split apart and reassemble, releases an amount of energy equivalent to the difference, scaled by a massive factor of $c^2 = 8.9853934 * 10^{16}(m/s)^2$. This is orders of magnitude higher than chemical reactions, and therein lies its energy production potential.

To create fusion plasma, or for hot fusion to happen, there are two main conditions: sufficient temperature and density. Firstly, the temperature needs to be high enough to overcome the binding energies that give matter its stability. As the temperature is raised, a typical material changes state from solid into liquid and then gaseous. Eventually, at high enough temperatures, the electrostatic attraction that binds nuclei to electrons is overcome, ionizing them. The electrons and the nuclei become a neutral gaseous fluid of charged particles called plasma. Despite being charged, the overall charge balances out, so at a macroscopic scale plasma is neutrally charged.

Charged particles interact with each other at a distance as described by Coulomb's Law, $\boldsymbol{F} = k_e \frac{q_1 q_2}{r^2}$, meaning the static electric force between particles of charge q_1 and q_2 is proportional to the charges and inversely proportional to the distance squared, where k_e is Coulomb's constant. This means oppositely charged particles attract each other, binding them electrically, whereas same-charged particles repel each other. This positive-positive electromagnetic repulsion is an obstacle to fusion, as it causes the plasma to behave with a tendency to spread out, where containment is desirable. In addition, a certain density is needed so that individual particle interactions are frequent enough that fusion, a statistically unlikely phenomenon, can happen at a significant rate. Moving charged particles can then interact via the Lorentz interaction, creating a complex web of interactions, waves, and instabilities. Plasma is this ionized, globally

neutral gaseous state of matter that interacts individually as particles and simultaneously as a fluid.

The difficulty in fusion lies in maintaining hot, confined plasma capable of overcoming this electrostatic repulsion of positively charged nuclei, with a density high enough that they fuse at a significant rate, for a long enough period of time. The product of temperature T, density n, and confinement time τ_E is defined as the fusion triple product, a metric to rate the viability of fusion. A triple product of $nT\tau_E = 1 * 10^{21} m^{-3} keV$ is considered the minimum for creating a self-sustainable fusion reaction, where the energy produced is greater than the energy cost to heat and contain the plasma. For energy extraction, the required triple product is estimated to be thrice that value [18].

Not all plasmas are created equally. In a gas at a temperature of T, a particle with mass m has an average thermal velocity, v_T, of $v_T = \sqrt{k_B * T/m}$, where $k_B \cong 1.38 * 10^{-23} J/K$ is Boltzmann's constant, and $k_B * T$ is the particle's thermal energy. The domain that describes a plasma is a function of its temperature and density, seen in image 2.1.

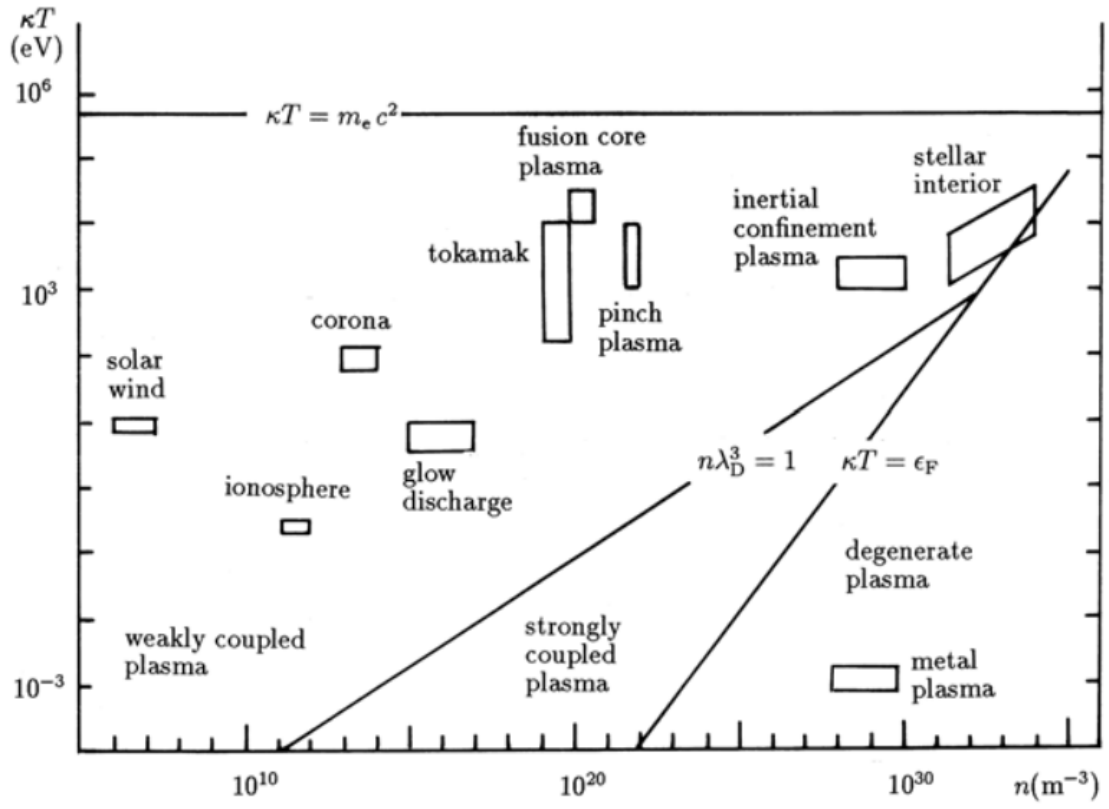


Figure 2.1: Plasma domains in the density-temperature ($n - kT$) diagram. Image from [19]

Stars, namely the Sun, seemingly easily achieve this fusion-capable plasma naturally, via sheer size. Accumulate a high enough quantity of Hydrogen (with 8% the Sun's mass as threshold) and its own gravity will contain it, increasing the temperature to 10 million ºC or higher from its close packing, and spontaneously igniting fusion reactions, releasing energy, releasing light. Harnessing this energy production system on a human scale requires much higher temperatures to offset the missing element of the star's mass and particle density. This can be better understood by comparing the several fusion-related plasma domains seen in figure 2.1, (fusion core, tokamak, pinch, inertial confinement) to the stellar interior domain. An electron of thermal energy $1eV = kT \cong 1.60 * 10^{-}19J$ corresponds to a temperature of $q_{e^-}/k_B = 1.16 * 10^4 K$, which places laboratory fusion temperatures at over 150 million ºC [20].

Unlike gravitational self-containment in the Sun, human-level, controlled fusion requires a different plasma containment method. Plasma is composed of neutral particles, as well as ions and free electrons. The latter two give plasma a charged nature, which can be used to contain it with electromagnetic fields. Alternatively, inertial confinement contains plasma fuel in small pellets or containers, heating them with powerful lasers, causing the Deuterium-Tritium to implode and fuse in a *controlled* fashion. Both these methods are broached in section 2.3.

Goldston [21] provides an introduction to the physics of plasma, how it can be modeled as an ionized charged fluid with magnetohydrodynamics, with the kinetic theory of plasma relating the micro-level interactions and motions to describe the behavior of plasma.

In order to eventually harness fusion, the basic requirements are then the capacity to measure, contain and heat up plasma until it burns. In this chapter, after this first look at what is meant by plasma, comes an explanation of what fusion entails, followed by a summary of current fusion research efforts, and finally explaining the importance of diagnostics.

2.2 Deuterium-Tritium fusion

Any two nuclei that collide at an energy high enough to overcome electrostatic repulsion could be fused into a heavier nucleus. The difference in mass between reagents and products releases or absorbs the equivalent amount of energy. As heavier nuclei are created, the mass difference progressively decreases, and fusion products tend to have more and more energy than the reagents, such that after iron-56 fusion is no longer exothermic. Therefore, when choosing fusion candidates for energy production, lighter nuclei, and more exothermic reactions are preferable. The chain of reactions that produce heavier elements, especially past Iron, is of high interest to astrophysics and nuclear physics. Current models "have not reached a 'mature' enough stage of development" [22], but recent advances in radioactive beam accelerators and improvements in astronomical observations have the potential to provide some answers in this field [23].

The first reagent that comes to mind is the simplest and most abundant: Hydrogen. Proton-Proton fusion almost always produces a diproton. A diproton is simply a Helium nucleus, which is very unstable and usually decays back into protons. However, one out of every 10^-19 Helium nucleus beta-decays into Deuterium, a much stabler particle. Deuterium-based reactions then have the potential to produce stable heavier elements. Even though proton-proton fusion is the starting point for all stellar nucleosynthesis, that process is only viable in stars because of the sheer amount of Hydrogen in a star and the time scale stars live at. This inefficiency makes pure protons unviable as a fusion fuel source. As a star ages, Hydrogen levels dwindle, and other chain-reactions based on Helium also occur, producing heavier elements, up to Oxygen, and as Helium runs out, up to Iron, the last exothermic fusion. With heavier elements, the mass-difference in fusion is smaller, and these reactions produce less energy per event, so fuel reagents should be as light as possible. Heavier-than-iron elements are originated only in high-energy astronomical phenomena such as Supernovae, where the environment allows for endothermic fusion to occur.

The second candidate is Deuterium, a Hydrogen isotope with an extra neutron. D-D reactions occur when Deuterium nuclei fuse, and can occur via two channels of similar probabilities:

$$H_1^2 + H_1^2 \rightarrow He_2^3(0.8MeV) + n_0^1(2.5MeV) + 3.3MeV$$

$$H_1^2 + H_1^2 \rightarrow H_1^3(1MeV) + H_1^1(3MeV) + 4MeV$$

With a much higher (around 100x) cross-reaction than D-D fusion, there is the main candidate for the first fusion reactors: the Deuterium-Tritium (DT) reaction.

In the DT fusion Deuterium D or H_1^2 fuses with Tritium T or H_1^3, producing an excited Helium particle He_2^5, which decays soon after, resulting in an α or He_2^4 particle, a neutron, and a gamma ray. The most common decay releases a 14.1MeV α particle, a 3.5MeV neutron, and a 16.7MeV gamma-ray[24]. Deuterium is a common isotope of Hydrogen and is relatively cheap to extract from water. Tritium is a rarer isotope, but can be bred by bombarding Lithium with energetic neutrons, which are produced in DT fusion, see eq. 2.2, ahead. Due to the availability of reagents, high cross-reaction, and energy release, DT fusion is the choice reaction.

$$H_1^2 + H_1^3 \rightarrow He_2^4(3.5MeV) + n_0^1(14MeV) + 17.6MeV$$

DT fusion is the main reaction of interest, and as such, the main reaction products of interest are neutrons, alpha particles, and gamma rays in these energy ranges. On the other hand, these particles can all interact among themselves, as well as with the components of the plasma containment vessel and measurement instrumentation, resulting in complex secondary reactions.

Deuterium plasma is reasonably understood, greatly thanks to research in JET, with current models for plasma in Tokamak devices generally matching experimental results in linear regimes. Current challenges are presented by non-linear and plasma edges, as well as the transposition of models into other H-isotope plasma, especially DT plasma, the key reaction. Garcia states that "predictability of burning plasmas is a key issue for designing and building credible future fusion devices"[25], which makes accurate readings of transport properties of particles in plasma a priority.

How does one contain and heat plasma?

2.3 Fusion Research

There are two main families of fusion reactor designs: Inertial Confinement Fusion reactors (ICF) and the more successful Magnetic Confinement Fusion reactors (MCF).

Inertial confinement reactors use lasers to fire at DT fuel capsules, heating them and creating burning plasma. Though historically unsuccessful mainly due to limitations in laser and capsule design, recently the National Ignition Facility has achieved a plasma Q-factor of 0.7[26].

This thesis is, however, focused on MCF development, which is based on a different principle.

Lorentz force describes the attraction between parallel conductors. When an electromagnetic field is applied, since plasma behaves as ionized gas, it behaves as a series of conductors. This principle, called Z-pinch[27], has been used in many fields, highlighted in magnetic plasma confinement. Devices such as the Reversed Field Pinch[28], Spheromak[29], Stellarator[30], as well as the toroidal Tokamak seen on figure 2.2, take advantage of this effect to contain plasma.

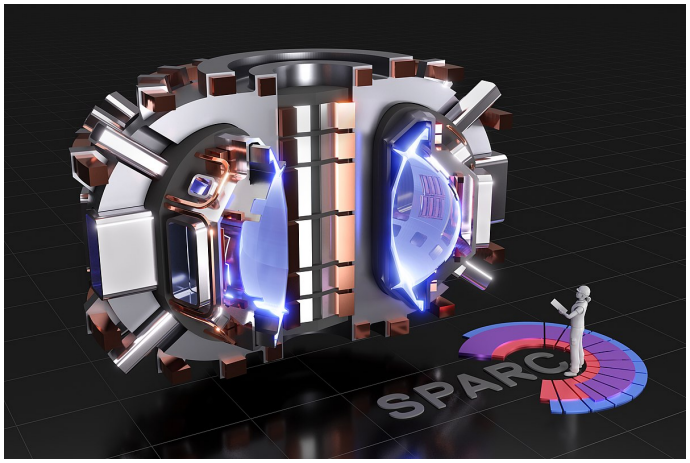


Figure 2.2: Render of the SPARC(Soonest/Smallest Private-funded Affordable Robust Compact) Tokamak, a project by Commonwealth Fusion Systems and the Massachusetts Institute of Technology to use high-temperature superconducting magnets in order to achieve fusion factor Q>2. Image by T. Henderson, CFS/MIT-PSFC, Creative Commons

Bulk heating of the plasma is possible via ohmic heating, using magnetic induction to heat the charged particles inside the containment chamber, but that is not enough to reach fusion temperatures. Further heating can be done via Neutral Beam Injection (NBI), by injecting a beam of neutrally charged, highly energetic particles; Ion Cyclotron Resonance Heating (ICRF), a method which is tunable to the relative amounts of ions in the plasma and helps in containment[31]; and by taking advantage of internal alpha particle heating [32]. As fusion research progresses, alpha radiation from the fusion reaction itself will eventually be enough to make the heating self-sufficient, where fusion products create enough heat that the fusion reac-

tion sustains its own conditions, without requiring auxiliary heating as a primary heat source. This is called burning plasma, a landmark in fusion progress.

Current challenges include increasing the stability of the nuclear reaction - increasing the length of time a reaction can be contained without side effects such as destructive particle/energy leakage, eventually allowing for continuous operation - and increasing the Power Amplification or Q factor, the ratio between energy that is input and output. A full description of research and development efforts for ITER is available[33]. The greatest Q factor achieved to date in a magnetic fusion containment used to be 0.69 by the JET(Joint European Torus) experiment in 1997, producing 21.7MJ of energy, although not managing to achieve the break-even point of Q=1. Note that this is the theoretical Q factor for plasma, accounting only for energy input and output as it relates to the plasma. The total Q factor would include the energy required to run the reactor itself, which makes commercial fusion a distant, and even more defiant objective[20]. More recently, in December 2021, a JET burst achieved a sustained plasma discharge of 59MJ over 5 seconds. This has a lower Q factor than the previous record, but twice the energy, and is a longer, more sustained discharge, in line with the objectives of current fusion experiments.

Of highlight are the international experiments in plasma containment JET and the ITER (International Thermonuclear Experimental Reactor), the latter currently undergoing construction. Both are plasma containment experiments based on the Tokamak design. Complementing the ITER project, the China Fusion Engineering Test Reactor(CFETR) is being designed[34]. CFETR focuses on researching tritium breeding and continuous operation, as well as improved management of fusion byproducts, which reduce the reaction's efficiency and can degrade the reactor's components.

Once net energy can be produced, only then is the door opened for conversion into electrical energy. The method of extraction is unclear as of yet. For this purpose, variations of steam turbines have been proposed as a first choice, as they are readily adaptable from current commercial use in conventional energy production methods[35].

Even though energy production is the ultimate end goal motivating nuclear fusion experimentation, energy extraction is outside the scope of the JET and ITER experiments. That role would be pioneered by a future project: DEMO, the first fusion demonstration reactor, bridging the gap between ITER and commercial nuclear fusion power plants, though hinging on the success of current experiments.

2.4 Gamma and Neutron Diagnostics

Section 2.3 concluded by highlighting the importance of accurate measurements of the fusion reaction. The current section looks at some properties of key reaction products and how their measurement can be useful in diagnostics.

2.4.1 Particles of interest

Radiation from nuclear processes can be categorized in 4 subtypes: Fast beta particles (positrons and electrons) and heavy charged particles (all energetic ions), which are all charged radiation. Neutrons and electromagnetic radiation are uncharged radiation. Knoll[36] provides a comprehensive review of the detection mechanisms for each type of radiation.

Electromagnetic waves, which here result from nuclear interactions, are divided according to their energy. Gamma rays and then X-rays are the two types of radiation with the highest energy, with an overlap in our range of interest, but the term gamma rays refers to their nuclear origin. The lack of charge creates a hurdle if detection is the purpose, so gamma rays are usually detected by inferring their energy from the electrons they transfer it to. Depending on their energy, the dominant interaction changes, from photoelectric effect to Compton effect and pair production. Each source produces a characteristic emission spectrum, so the detection of gamma rays and the production of a spectrum can be used to know what interactions give rise to it. Gamma rays occur from the nuclear interaction between highly energetic ions (protons, deuterons, tritons, helium ions, or alpha particles) and plasma fuel ions or containment materials, which include beryllium, boron, carbon, and oxygen. Additional sources of gamma rays are varied: cyclotron acceleration, neutron beam injection, and fuel reactions, as well as secondary nuclear reactions within the reactor, reactions with plasma impurities and with the walls of the reactor itself. A list of expected reactions in JET's plasma can be found in [37].

In addition, depending on their origin, gamma rays can be divided into prompt or delayed. Prompt rays are created directly as a result of an interaction. On the other hand, if a reaction creates an excited nucleus, its decay can emit radiation. De-excitation via fluorescence can emit characteristic X-rays, and further, nuclear de-excitation, via alpha, beta, or gamma decay, or internal conversion can produce a gamma photon. In either case, this introduces a delay in the emission of a photon which depends on the nucleus' half-life.

Neutrons are massive particles but have no electrical charge. This means that, unlike gamma rays, they don't interact electromagnetically. Instead, the cross-section of neutrons' interactions is a strong function of their energy. At low energies, especially below the Cadmium cutoff energy

of 0.5eV, this can be an obstacle to their detection, due to the unlikelihood of interacting, so often neutron-mediated nuclear reactions are employed to indirectly detect the neutrons by capturing the reaction products. On the other hand, neutrons of interest produced in nuclear fission reactions typically start at 10keV, at which level elastic neutron scattering is indeed an option. When this occurs, the neutron transfers part of its energy in the scattering, onto the target, creating a recoil nucleus. In the case of a Hydrogen nucleus, up to 100% of its energy can be transferred. Since this proton is easily detectable, it can be used as an indirect detection of fast neutrons.

An alpha particle is simply an ionized Helium nucleus. Charged and massive, alpha particles generate a series of secondary reactions inside the reactor and on its edge, which can be measured by their characteristic gamma rays[38]. In DT fusion, eq. 2.2, roughly 20% of the output energy comes in the form of an alpha particle. Assuming that it is all reabsorbed, this means that if the output energy is 5 times the input, the reaction can become self-sustainable, igniting the plasma indefinitely. This is why achieving ITER's goal of $Q > 10$ is so relevant.

Gorelenkov[39] reviews Energetic Particles physics in the context of burning plasma experiments, "going from single particle motion to their interactions with electromagnetic fields, to various collective effects and to the predictive models for burning plasmas (BPs) and ITER in particular", summarizing current and future areas of research.

2.4.2 Diagnostics

Diagnostics are tools designed to measure a specific characteristic in a fusion reactor. They resort to outside measurements or particle probing, because material contact with plasma is generally too destructive, and deduce characteristics about the inside.

Hutchinson[40] summarizes the types of measurements that diagnostics techniques are capable of, and how they can be used to reconstruct the fluid plasma profile. These are divided into the measurement of magnetic fields, plasma particle flux, refractive index, electromagnetic emission and scattering, neutral atom diagnostics, and fast ions and fusion products.

In the context of JET and ITER, it's important to know the role of diagnostics.

Moseev [41] discusses recent progress in the key diagnostics in tokamak and stellarator plasmas for fast ions. These include "neutron and gamma-ray spectroscopy, fast-ion D-α spectroscopy, collective Thomson scattering, neutral particle analyzers, and fast-ion loss detectors", and progress is broadly attributed to contributions from the ASDEX Upgrade and JET Tokamaks.

Kiptily [42] reports how gamma-ray spectrometry in JET can be used to derive the relative

concentration ions in the plasma, and together with the JET neutron profile monitor, reconstruct emissivity profiles inside the plasma. In ITER, Gamma Ray spectrometry is done in neutron cameras, the Vertical Camera, and the Radial Neutron Camera. It can provide information on alpha particle birth, confined particle profiles, including impurities in the plasma. It is also used to detect different plasma instabilities[38].

Besides the physical and mathematical complexity involved in diagnostics, there are additional challenges in designing appropriate acquisition systems. The detection chamber has to be resistant to the harsh environment, with high radiation doses and challenging temperature cycles[14]. In addition, the system needs to be sophisticated enough to process high amounts of data at fast frequencies, in particular real-time diagnostics, which can be used in control sequences to manage the plasma, improving confinement and plasma parameters, to achieve steady state fusion.

These types of acquisitions are challenging all throughout the acquisition chain. A single event occurs when one particle hits a detector, which outputs a small amount of light. The challenge starts with choosing the right detector, usually a scintillator, amplifying and converting the pulse to an electric signal; the pulse is then digitized in an analog-to-digital converter (ADC); pulse processing can then happen, to correct electronic or physical limitations. Information is then extracted from the pulse, finally permitting spectrometry and counting. All of this has to be supported by electronics with sufficient data throughput. Pereira [43] explains some of the challenges in designing the upgraded data acquisition system for the JET EP2 gamma-ray spectrometer, such as the very high count-rates, up to 1GSample/s, neutron background, pile-up, and pulse analysis, and the electronics required to detect and handle this much pulse data. This diagnostic, along with ITER's RNC, is our main focus in particular when it comes to the detection of gamma rays and neutrons.

The RNC is designed to deliver the real-time neutron emissivity profile, equipped with detectors at 26 lines of sight. Each line of sight can be subject to a peak event rate of 2MEvents/s. These measurements build a neutron spectrum which is, together with magnetic flux information from other acquisitions, used to calculate the emissivity profile using an inversion algorithm, inside a time window of 10ms to match the control cycle period. The system needs to be capable of generating spectra with pile-up rejection and neutron/gamma pulse shape discrimination[16].

2.5 Pulse Detection

This section looks at how radiation interacts with matter, such that it produces a signal in some form, called a pulse. Whatever the method of detecting the pulse, the analysis of this output pulse is what characterizes the incoming radiation (section 2.8). The end form of this pulse is electrical, due to convenience, though the pulse might not always begin as such, as is the case with Photomultipliers (section 2.5.3).

2.5.1 Particle detection

Charged Particles

Charged particles going through a material continuously interact via the Coulomb force with the charged particles present in the material. The particles can interact with the material in several ways, such as with Rutherford scattering, Bremsstrahlung or alpha-induced reactions. But taking into account the range of the nucleus compared to orbitals or free electrons, the most likely interactions are elastic collisions with electrons, meaning a detector's electron density is the primary driver behind its stopping power, S. The stopping power is also the rate at which a particle loses energy as it passes through a material, defined by $-dE/dx$. The bulk of energy loss is done progressively, where a single particle interacts with several electrons throughout its path, exciting or even ionizing them and producing secondary electrons, as the incident particle slows down. In each collision, the energy loss is proportional to the ratio in mass between the particle and the electron in the material. The rate of energy loss, more specifically named the specific energy loss, is described by Bethe's formula in eqn. 2.1.

This means that heavy charged particles can take hundreds of interactions to slow down in a roughly linear path, while electrons slow down much faster and in more erratic paths. The $1/v^2$ component can be understood by a slower incident particle spending more time in the vicinity of possible interaction points, which increases the cross-section of interactions at lower speeds. Therefore, as a particle slows down, it deposits increasingly more energy, reaching a maximum called the Bragg peak, shortly after which it stops. This sets a characteristic upper limit for the maximum penetration depth that each type of particle can have, called the particle's range.

Note that slower particles have less momentum and, like electrons with their lower mass, are more deflected by interactions such as Bremsstrahlung. Bethe's formula is not as accurate in describing these conditions.

As for detectors, this means heavier and denser elements absorb radiation at a faster rate. The absorption of a charged particle is done over time, and involves many particles, meaning

this energy loss is a stochastic process. This results in a natural energy distribution centered around the energy of the incident particle, and in a similarly variable stopping range/time.

$$\frac{-dE}{dx} = \frac{4\pi e^4 z^2}{m_0 v^2} NB \tag{2.1}$$

$$where \tag{2.2}$$

$$B = Z \ln \frac{2m_0 v^2}{I} - ln(1 - \frac{v^2}{c^2}) - \frac{v^2}{c^2} \tag{2.3}$$

Bethe formula for the specific energy loss of a heavy charged particle of kinetic energy E through a material, along its path x. z is the particle's charge number, so the product ze can be interpreted as its charge. v is the velocity. m_0 is the electron rest mass and c is the speed of light. N and Z are the number density and atomic number of the absorption material's atoms. I represents the average excitation and ionization energy of the absorbing material atoms, typically an experimental value. Note that at non-relativistic speeds, the logarithmic component is negligible, making B approximately equal to Z. As far as the detector material is concerned, its stopping power is mainly determined by the product NZ, also known as its electron density.

Uncharged Particles

Unlike charged particles, gamma rays do not present electrical charge and interact with matter quite differently. There are 3 main interaction types: Photoelectric absorption, Compton scattering, and pair production, with the dominating one changing with increasing energies, shown in image 2.3.

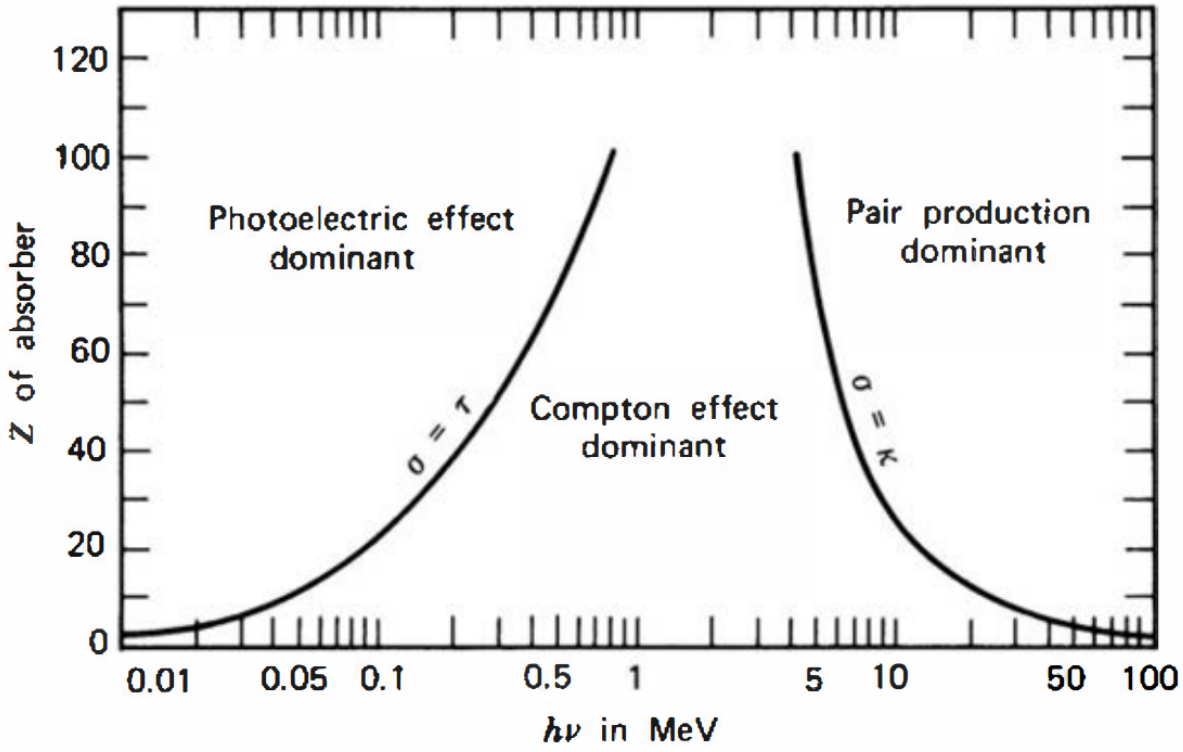


Figure 2.3: Dominant gamma-ray interaction according to energy. Note that fusion reactions range in the keV to MeV energy levels. From [36]

In photoelectric absorption, the absorption of a gamma photon results in the partial or complete transfer of energy to a single electron, where the photon disappears, ejecting a photoelectron from an inner shell of an atom in the material, with energy given by equation 2.4. hv is Plank's constant times the frequency of the photon, and E_b is the binding energy of the ejected electron. The ejected electron creates a vacancy in the atom, which can be filled by absorbing an electron from the medium, or instead from another orbital in the same atom, resulting in a cascade of electron transfers, releasing energy via a photon. In some cases, an electron from the atom's outer shell is emitted instead, called an Auger electron. Thus, each of these transfers or absorption results in the emission of characteristic X-rays or an Auger electron.

$$E_{e^-} = hv - E_b \tag{2.4}$$

Compton scattering occurs when an incoming gamma photon elastically scatters with a weakly-bound electron, partially transferring its energy. The amount of energy transferred depends on the angle of the scattering, with equal probabilities overall and maximum energy corresponding to a dispersion angle of 180º, called back-scattering.

At 1.02MeV, a gamma-ray has the equivalent energy to twice the mass of an electron/-positron. This means that starting from this energy, the production of a pair of electron and a positron can occur, with total kinetic energy equal to the leftover gamma energy. Total absorption of these particles is possible, but the positron very easily interacts with an electron, annihilating both and producing two gamma rays of 511keV. The gamma-ray absorption spectrum is then characterized by a plateau through Compton scattering angles, called the Compton continuum, ending at the Compton edge. If incident radiation is above 1.02MeV, there is then a total absorption peak when pair-production occurs and is fully absorbed via photoelectric effect, at 1.02MeV. If one or both of the annihilation gamma rays escapes the detector without being absorbed, this creates two additional peaks, at 511keV incrementally lower energies, the single and double escape peaks, seen in figure 2.4.

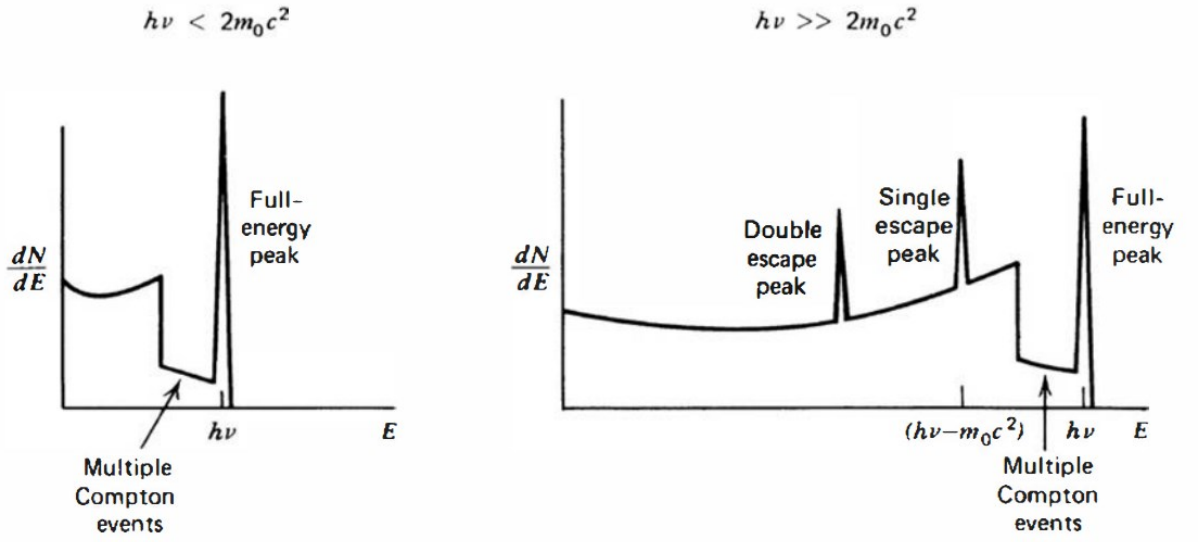


Figure 2.4: Gamma-ray detection spectrum according to incident gamma energy, $h\nu$. $m_0c^2 = 1.02MeV$ is twice the mass of an electron or positron. From [36]

Neutrons are massive uncharged particles and are almost always detected indirectly by taking advantage of an interaction that converts them into charged particles. The energy of the neutron heavily influences the cross-reaction of their interaction with matter, so the detection of neutrons depends on the energy range that one is interested in detecting. Below 0.5eV, the Cadmium

cutoff, these are called slow neutrons, of little interest in nuclear fusion. For fast neutrons, above 10keV, the main reaction of interest is elastic scattering, where a neutron interacts with a nucleus, transferring some of its energy to the nucleus, creating a recoil proton. This energy transfer depends on the angle of the scattering, creating a continuous distribution of energies, up to a maximum that depends on the target nucleus, according to equation 2.5, where A is the mass of the nucleus divided by the neutron mass and Θ is the scattering angle in center-of-mass coordinates.

$$RecoilEnergy = 2A/(1+A)2(1-cos(\Theta)) \tag{2.5}$$

Hydrogen is the only case in which all the kinetic energy of the neutron can be transferred in elastic scattering to a nucleus; other target nuclei have a maximum transfer of $4A/(1+A)^2$, a fraction smaller than 1, which always implies several collisions for full energy transfer unless the target is a proton.

Moderating neutrons is a method that involves surrounding a slow-neutron detector with Hydrogen in order to slow down neutrons with successive collisions until they reach an energy level where they can be efficiently detected, a process called thermalization of neutrons. The thickness of the moderator increases the moderation effect but also decreases the probability that the neutron reaches the detector, so moderation is not a linear conversion. Moderation-based detection involves several collisions, so these are typically too slow for nuclear fusion detectors. Slow neutron capture is usually done using the Boron reaction (equation 2.6), with an energy loss Q equal to 2.792MeV, or 2.310MeV if ${}^{7}_{3}Li*$ is excited.

$${}^{10}_{5}B + {}^{1}_{0}\,n \mapsto {}^{7}_{3}\,Li + {}^{4}_{2}\,\alpha \tag{2.6}$$

$$\tag{2.7}$$

Q = 2.792MeV or 2.310MeV if ${}^{7}_{3}Li*$ is excited (which it is 94% of the time)

The Boron reaction is very popular due to its high cross-section and the availability of Boron in this form.

For fast neutron spectroscopy, ${}^{3}He(n,p)$, equation 2.8 or ${}^{6}Li(n,\alpha)$, equation 2.9, are the main reactions used to convert neutrons to charged particles. The energy loss in these reactions, the Q-value, can be added to the detected recoil particle energy to obtain the incident neutron's original kinetic energy, which wasn't the case in slow neutrons, as their energy is too low compared to the Q value to extract information. Higher Q-values indicate more energy transferred to the products, which is otherwise superior. The incident neutron energy heavily influences the cross-section of the reaction, decreasing inversely proportionally to the neutron's velocity over most

of the energy range, with exceptions being the moderator's resonant frequencies. So special care has to be taken in order to consider the Light Output Function and Pulse Height Distribution of each detector.

$$ {}^{3}_{2}He + {}^{1}_{0}n \mapsto {}^{3}_{1}H + {}^{1}_{1}p \tag{2.8} $$

Q = 764keV

$$ {}^{6}_{3}Li + {}^{1}_{0}n \mapsto {}^{3}_{1}H + {}^{4}_{2}\alpha \tag{2.9} $$

Q= 4.78MeV

In any of these particle interactions, what happens is the accumulation of charge over time in the detector. This charge can be collected in the form of a pulse in several ways, such as directly with proportional counters or solid state detectors, or indirectly in the case of scintillators with photomultipliers (PMTs). In either case, the result is the production of an electric signal with amplitude proportional to the energy deposited by the particle, as will be shown next.

The resolution of a detector can be quantified by the Full Width at Half Maximum at the position of a known peak in the spectrum, usually the photopeak, which is the gamma pair-production double absorption peak.

2.5.2 Scintillators

Scintillation at its core is a property exhibited by some materials that, when excited by ionizing radiation, exhibit luminescence, emitting detectable, longer wavelength light after a characteristic decay time. There's a huge variety in scintillation materials and their properties, so a focus will be put on describing the main mechanisms that have to do with gamma and neutron detection, namely the physics that govern organic and inorganic scintillators.

In designing scintillators for radiation detection, the right scintillator should be chosen for each application. This means choosing the right size to capture the desired radiation and the material with appropriate efficiency, energy and time resolution, cost, radiation resistance, decay time, etc.

In nuclear fusion, scintillators should be able to adapt to the high neutron flux and gamma rays.

Organic Scintillators

Organic scintillators get their name from their molecular structure - aromatic hydrocarbon molecules with Π-electron structure - so their scintillation properties are based on how they absorb radiation as a whole molecule, exciting valence electrons onto higher energy orbitals, which later de-excite, emitting light with energy corresponding to the difference in energy between the orbitals.

Π-electron can be excited from the lowest energy orbital, S_0, S meaning singleton or spin 0, S_0 indicating the lowest energy orbital, S_1 second lowest, etc, onto higher energy orbitals. De-excitation from singlet states back to baseline corresponds to the emission of light via prompt fluorescence, which has a short decay time. Once in higher orbitals, electrons can then cross into triplet, T orbitals of spin 1 via inter-system crossing, interacting with a phonon and changing the spin and, usually slightly, the energy level. De-excitation from triplet states directly to S_0 orbitals or vice-versa requires an additional element in the interaction for the spin inversion and is much less likely to occur. As such, these transitions are characterized by much longer decay times than from S states, and the light emitted in de-excitation is called phosphorescence instead. If an electron lingers onto the harder-to-decay from T state, crosses back onto S_1, and then decays to S_0, a retarded fluorescence photon is emitted, naming the process delayed fluorescence. Thus, you have three types of luminescence in a scintillator, prompt and delayed fluorescence, and phosphorescence. These are characterized by their respective time constants, with fluorescence typically in the nanosecond range, while phosphorescence ranges in the milliseconds and is not so relevant for detection.

The prompt fluorescence intensity can be described simply by $I = I_0 e^{-t/\tau}$, an exponential decay over time with decay constant τ. An overall light pulse can be described by $I = I_0(e^{-t/\tau} - e^{-t/\tau_1})$ where the first exponential, the rising edge, is dominated by the prompt fluorescence, whereas the limited speed of populating optical levels introduces the second exponential. So there is a fast component, related to prompt fluorescence, and a slow component, related to the triplet states with longer decay times. Different incoming particles lose energy through the scintillator at different rates and create different concentrations of triplet states. The time response function of a scintillator to a particle is therefore characterized by how the particle loses energy as it travels throughout the material, as described in equation 2.1. As a consequence of this, different particles at the same energy should have similar rising shapes, but differently shaped tails. The identification of incident particles from the shape of the light pulse is called pulse shape discrimination (PSD), especially relevant in the discrimination of gamma-induced events in neutron detectors, and discussed in section 2.8.

Organic scintillators can come in several types. Pure organic crystals are the simplest, but these suffer the typical limitations of crystals, such as being hard to grow large enough, with enough isotropy and purity, and being more fragile.

An organic material can be dissolved into a solvent, creating a liquid organic solution without having to resort to a crystal. This makes them usually cheaper to produce, and lack of structure means liquid organics are typically more radiation-resistant. However, the solvent can cause quenching, which is unwanted absorption by dissolved particles that reduces scintillation efficiency. If the solvent used can be polymerized, a solid solution can be produced, called a plastic scintillator. These have the cost and regularity advantages of liquids, but the added structure means they lose some radiation resistance.

Organic scintillators can be adapted to each use-case. The size of the scintillator is usually chosen to allow for as much absorption as possible in the name of efficiency, but thin-film plastic scintillators have been developed to allow for the transmission of radiation with longer travel times, creating unique response functions. Typically, organics show a good response to charged particles. Another use-case is fast neutrons detection, which through elastic scattering produces recoil protons and can be indirectly detected. Gamma rays react more easily with heavier elements, which isn't the case with organics, as hydrocarbons are composed of Hydrogen, Carbon, and Oxygen, relatively light elements. This results in a low photoelectric reaction cross-section, meaning the dominating reaction is Compton scattering. But this also means that it can be useful to add heavier elements, usually tin or lead, to increase the cross-section of the gamma-ray interaction, at the cost of lower light output and energy resolution.

Inorganic Scintillators

Inorganic scintillators are typically made from glass and/or ceramics, and their luminescence is based on the crystalline structure of the material. The absorption of a particle in a crystal is accompanied by a transition of an electron from the valence band to somewhere in the conduction band, creating a hole in the valence band. The energy gap between these bands is the minimum energy that an incoming particle needs to transfer in order to excite the electron, and at the same time, the energy difference between transition bands is the energy of a photon emitted at the time of de-excitation. In typical materials, the energy gap is too great to create visible light, so the crystals are doped with impurities that introduce energy levels between the conduction and valence band. Transitions between these levels, so-called recombination or luminescence centers, can then produce visible light, with the emission spectrum of the scintillator determined by the impurity and host crystal.

Inorganic scintillators typically have a higher efficiency and better resolution when compared to organics, and a more linear response as well, making them a very popular choice. However, the recombination centers create more complex emission pulses, and phosphorescence decay times are closer to prompt fluorescence in most inorganics. This makes them seldom be chosen for pulse shape discrimination, although in certain applications more complex emission properties can make the discrimination of other particles, such as Hadrons, possible[44].

The output pulse can be described as a variation of a Gaussian, where the leading edge, the rise time, depends on the incident particle[45].

Semiconductor Detectors and others

Semiconductors display a characteristic bandgap, an energy difference between valence and conduction bands. They don't conduct electricity unless an electron manages to cross onto the higher energy valence band. Incident radiation can excite an electron from the valence to the conduction band, creating an electron-hole pair, effectively creating two charge carriers in the medium. This is similar to gas-chamber detectors, where incident radiation creates an ion pair by exciting a gas, creating charge carriers, but this technology has been all but superseded by semiconductor detectors. The collection of this charge is a way of detecting incoming radiation. If a bias voltage is set at both ends of the material, this will create a uniform electrical field on the inside. Drifting charge carriers now have an induced acceleration, which creates a voltage pulse at the terminals. The bias voltage is created by non-injecting electrodes, as to not create additional charge carriers at the terminals. These are called Semiconductor Diode detectors. Their preponderance to interact is characterized by their atomic number Z and their band structure, in particular the bandgap energy. Traditional intrinsic semiconductors, Germanium and Silicon, are limited in their usage up to around 10keV energies. Larger band gaps and higher atomic numbers mean a higher cross-section with high-energy particles, so heavier and compound semiconductors are an area of interest in high-energy radiation detection.

The shape of the pulse can be described by the Shockley-Ramo Theorem [46], and as far as the incident particle is concerned, it mainly depends on the deposited energy. Thus, most semiconductors aren't apt for discrimination of incident particles.

However, creating 100% pure semiconductor crystals is currently impossible, so their characteristics is measured in terms of their impurities, leveled by design for the desired characteristics, adding donor or acceptor energy levels (and electrons) to the band structure, thus creating n or p-doped semiconductors. Impurities are defects in the semiconductor crystalline structure, where an atom is replaced by a different element. Propagation of charge carriers, electrons or

electron vacancies, throughout semiconductors is dependent on the regularity of their structure, and defects create charge traps. Charge trapping, in addition to leakage current (present at any non-zero temperature), as well as charge-recombination are the main limitations to the usage of semiconductors as detectors.

Doped semiconductors have also been used with increasing success to detect radiation. Silicon Drift Detectors(SDDs) are made of a certain configuration of n-doped and p-doped semiconductors, with blocking electrodes at the endpoints that induce a bias voltage and allow the measurement of output charge. Inside, a neutral region is created by drifting a certain impurity to compensate acceptor or donor charge carriers, creating a neutralized active detection region. The most common semiconductor drift detectors are Lithium-infused Germanium or Silicon, usually referred to as Ge(Li) and Si(Li). Particles that hit the active region and interact produce an electrical pulse at the electrodes. Note that advances in purity in Germanium crystals mean semiconductor diode detectors are more prevalent.

The infusion of Lithium impurities is done progressively, and can similarly drift over time, making semiconductor detectors especially vulnerable to temperature changes and radiation. These properties aren't constant over weeks, and the photon absorption spectrum shows changing efficiencies over time, so semiconductors require frequent calibration.

In the usual planar geometry, an incoming particle interacts with a charge carrier, inducing a charge at the detector electrodes due to the bias voltage. This charge profile depends on where the particle hits the detector, which introduces a variation in pulse shape depending on the incoming particle which also reduces time resolution. In some applications, this can be taken advantage of to create position-sensitive detectors, including common camera sensors, such as Charge Coupled Devices(CCDs).

For lower energies, up to X-ray spectroscopy, SDDs allow for good energy resolution at low shaping times and high count rates[47]. For higher energies, other compound semiconductors, with higher Z values, larger bandgaps can be used, such as Gallium Arsenide, or ZnTe and CdTe blends (CZT).

SDDs are the equivalent of gas-filled ionization chambers, where incident radiation creates charge carriers which are collected at the terminals. If charge multiplication is also possible, these become analogues to proportional counters, called avalanche detectors.

Diamond, with a huge bandgap of 5.6eV can also be used as a semiconductor diode, which creates a detector highly robust to radiation and temperature. These can be single crystals and can also be created by chemical vapor deposition(CVD), called CVD Diamond detectors. These properties, in addition to great timing resolution, charge-carrier mobility, low leakage current,

and capacitance make them currently one of the best candidates for fast neutron spectroscopy. A unique characteristic of Diamond detectors is that the read-out pulse shape depends on the ionization profile, which depends on the incident particle [48]. Based on this, Diamond CVD detectors have recently been demonstrated to be uniquely capable of pulse-shape analysis for pulse-shape discrimination and background rejection, in particular for charged particles [49].

2.5.3 Charge Collection

Whichever method of radiation detection, our goal is to produce an electrical pulse that can be interpreted. Current and Voltage are the two forms of electrical encoding.

Current, when used for signal transmission, since it is measured in series, its SNR is less affected the lower the resistance is, which is advantageous in some cases, in particular for very low power applications and to minimize long-distance voltage drop in transmission, neither of which finds a home in fusion instrumentation. Voltage is measured in parallel, which adds less noise, and benefits from higher resistance. For practical use, it's easier to reach effectively infinite resistance than zero resistance.

In the case of Scintillators, they convert incoming radiation into light pulses, and this still needs to be converted. For that purpose, the main devices used are photomultipliers tubes and semiconductor devices, discussed in the following sections. Their output, as well as a semiconductor when used as a standalone detector, is a current pulse. These tend to be low-current, thus poorly apt for coaxial transmission, so at this point, they are converted into voltage. This works by using a charge preamplifier, presented in image 2.5.

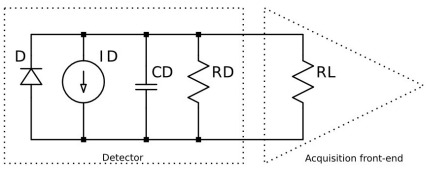


Figure 2.5: Basic charge to voltage conversion circuit, from [50]. A detector creates a signal in current, which acts as a current source with high impedance. By integrating the charges over time in a capacitor assembled in parallel, a signal in voltage is produced at the output load, which can be designed to match the impedance of the subsequent circuits.

Photomultiplier

We arrive at the Photomultiplier Tube with light pulses created by a scintillator. In each pulse's shape is encoded information about the source phenomenon. Both the individual pulse and the

overall statistics provide valuable information on the state of the nuclear reactions inside the Tokamak. The purpose of the PMT is the conversion of weak light pulses into usable current pulses without adding too much random noise. As a black box, a light pulse arrives at the photocathode, and after a 20-50ns delay, an identically shaped current pulse is read at the anode.

The general structure of a PMT, represented in figure 2.6, can be divided structurally into an outer vacuum enclosure containing a photosensitive layer and an electron multiplier which finishes in the anode, where charge is collected as the output. The vacuum's purpose is to allow for the efficient acceleration of electrons by electrical fields.

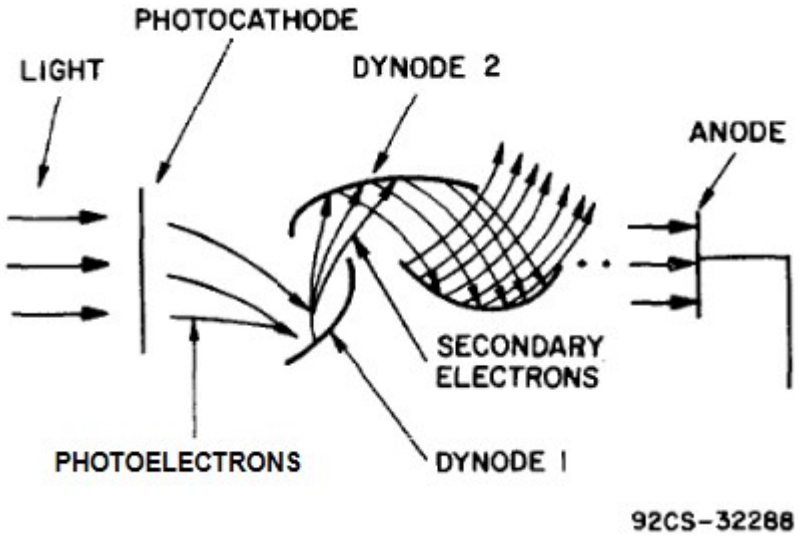


Figure 2.6: Structure of a Photomultiplier Tube, from [51]

The photosensitive layer converts photons into low-energy electrons. Since only a few hundred electrons are produced from each scintillation event, in order to create a usable current this requires an amplification stage, which is the electron multiplier that follows. In this layer, if an arriving photon has energy hv greater than the energy gap (Eg) between the conduction and valence bands of the photosensitive material, it can be absorbed, transferring its energy into an electron onto an excited state with energy hv-Eg. This electron can be created anywhere inside the material, so if it is to escape it will need to migrate to the surface of the layer, where it will face the material-vacuum interface. If it is energetic enough to overcome this potential barrier, it can then escape from the surface. This process is called secondary emission. The vacuum introduces a minimum energy barrier for the photoelectrons, which for the typical PMT means there is a cutoff at the red or near-infrared frequencies. Throughout the migration process, an electron will continually lose some energy via phonon interactions and electron-electron colli-

sions. Upon reaching the bottom of the conduction layer, the conduction electron can recombine with a hole, and its information is lost. This process lasts around 1ps.

This energy loss and time limitation introduce a compromise in the size of the layer: it has to be thick enough that it absorbs incoming photons, but not enough that the escape depth is too long and electrons reach the surface without enough energy to escape. In metals, this translates to an escape depth of a few nm, while in semiconductor materials, because of smaller energy gaps and fewer interactions the layer can be up to 25nm.

When speaking of efficiency of a PMT, you can speak of the overall efficiency in terms of amperes of output current per lumen of incident light, which is the commercial approach. The quantum efficiency is the ratio of output photoelectrons per incident photon, and typically maxes out at 30%. An alternative measurement, common in scintillators, is the average energy loss required to create 1 photoelectron. Each photolayer material has a different absorption spectrum, and this method has the advantage of averaging out the efficiency throughout the absorption spectrum. The average energy loss is often compared to NaI(Tl) crystal as a standard, due to its common usage and relatively broad absorption spectrum.

In normal semiconductors, the electron affinity is such that at the surface of the photosensitive layer, the outside potential is higher than the conduction band. This difference, added to the continuous phonon-interaction energy loss and eventual recombination is limiting in photoelectron emission. If a material with negative electron affinity is introduced, such as p-doping with Zinc, an intermediate energy level is created between the valence and conduction band, which allows for electrons to linger longer in the conduction band before recombining, granting them a greater chance to escape. The immediate consequence is that more electrons are produced at the first stage. Additionally, the greater chance of being in the conduction band when emitted results in a lower average electron energy distribution, which is more uniform. These bring advantages to the PMT's efficiency and resolution.

Some electrons can naturally escape the photosensitive layer, even without being excited by incoming light. This is called spontaneous emission. At any temperature, the thermal energy of the electrons in the material is a distribution, so naturally, it can occur that an individual electron has enough energy to leak. In metals, the high potential barrier results in low leakage, while in semiconductors the lower barrier results in greater photosensitivity with the drawback of higher thermal emission. Thermal electrons reaching the multiplication stage will be amplified as well.

The electron multiplication stage consists of a series of electrodes set at increasingly higher voltages, which creates a field between each of them. Electrons accelerate and hit each electrode,

multiply, and repeat this process several times until reaching the output anode as a readable current pulse.

Upon escaping the layer, electrons from the photocathode are accelerated by an electrical field until they strike the surface of an electrode, called the dynode. If they have sufficient energy, this can result in the re-emission of more than one electron of the same energy, explaining this phenomenon's name - secondary emission. A single electron can cause the re-emission of several secondary electrons, for example, if the first dynode has an electrical potential difference of 270V relative to the escaped electrons, and the dynode bandgap is 2,7eV, an electron can reach the dynode at up to 270eV of energy, and generate up to 100 secondary clones. However, since re-emitted electrons are sent in a random direction, only a fraction is sent into the correct direction onto the next dynode.

Secondary emission is a statistical process, so at each dynode one would expect a Poisson distribution of energies to arise. But the first stage has the fewest electrons, and is a greater factor in the noise distribution at the anode. Practical results do not fully agree with the series of Poisson distributions, but the end result is a broadening of the pulse signal. In addition, since different trajectories of capture at each dynode are possible, instead of a fixed delay, the end pulse will suffer from a spread transit time.

In terms of the electron multiplication layer, there are plenty of ways of building a PMT. Side-on PMTs have the initial photosensitive layer at the side, and often employ a circular series of multiplication layers, which reflect onto each other, creating a cascade of multiplications. This reflection mode allows for compact, fast, and cheap PMTs, but their geometry might limit the design of the acquisition system, and they aren't as flexible in design as head-on PMTs. Head-on PMTs operate in transmission mode, with light entering one side and an electrical signal exiting from the opposite side. The dynode design varies a lot, depending on what the application requires. For instance, linear-focused PMTs exist to minimize the response time; some map the input onto different channels in the output, becoming position-dependent PMTs; More complex approaches also exist, such as replacing the dynodes for microchannel plates, which use several tubes for individual photon paths for a decrease in delay but at a higher cost; other dynode geometries might focus on magnetic interference robustness, uniformity, output current, SNR, etc [52].

The limitations of the material and designs have to be taken into account when choosing a PMT. Besides the characteristics of individual chosen materials, such as luminous sensitivity (overall or cathode), radiant sensitivity, and electronic limitations like rise time and dark current, you have other factors limiting the ideal functioning of a PMT.

The aforementioned non-linearity of electron trajectories, especially relevant at the last dynode, creates a spread in transit time. Non-uniformities in larger photocathodes, even when the light is spread out to average the variation out, have the effect of reducing the energy resolution. Effects such as hysteresis can gradually increase the gain at each dynode, an effect called fatigue [53]. The thermal spontaneous electrons mentioned above can be a significant source of noise, and they can be reduced by cooling the PMT, but there are other possible sources of noise. Unwanted signals can appear from external sources such as cosmic rays, or internal sources, like echoed signals from a dynode further ahead that find their way back onto earlier stages of amplification.

Alternative approaches to PMTs can be found in different technologies, especially those using semiconductor photodiodes, discussed ahead in section 2.5.3.

As was mentioned earlier in this section, it is convenient for a pulse signal to be expressed in terms of voltage, not as a current. Therefore, after the anode in the PMT is a charge collection circuit, responsible for this conversion. This circuit is ideally expressed as a parallel RC circuit. In this circuit, an exponential pulse arrives from a scintillator. As long as the scintillator decay constant λ is significantly greater than the spread in transit time, the input current pulse is an exponential decay pulse $i(t) = i_0 \cdot e^{-\lambda t}$. Integrating the charge over the circuit using $Q = \int_0^{\inf} i(t)dt$ results in $i(t) = \lambda Q e^{-\lambda t}$. The voltage can then be solved as $V(t) = \frac{1}{\lambda - \theta} \cdot \frac{\lambda Q}{C}(e^{-\theta t} - e^{-\lambda t})$ [36], where $\theta = 1/RC$.

Issues with light in PMTs often occur where the geometry is prone to reflections, especially at the interface between materials. The transit time and multiple reflections introduce a spread in the arrival time of photons, impacting the time response of the pulse. Coupling between scintillators and PMTs can use light guides or several types of fibers to channel the light output into the desired direction. Sometimes wavelength shifters are incorporated to absorb the scintillator light and re-emit it at a longer wavelength, in order to match the emission spectrum with the most efficient range of the light output of the PMT.

Semiconductor Photodiode

The typical disadvantages of Semiconductor Photodiodes as detectors is their limited sensitivity to higher energy radiation, especially neutrons and gammas. When coupled with a Scintillator, these find a use-case for collecting the scintillator output light.

Semiconductors junctions exhibit particular electric properties, and a common orientation of semiconductor layers is P-doped, Intrinsic, N-doped, as found in PIN photodiodes, which were discussed in section 2.5.2. If high voltages are set at the terminals, the electrical field is enough to

amplify charge carriers, multiplying them, and the photodiode is called an Avalanche PhotoDiode (APD). PMTs and APDs both provide charge amplification, while PINs don't. The amplification factor is highly dependent on voltage. However, past a certain point, an APD can enter Geiger Mode, where there is no upper limit for amplification and a single photon event can generate a full discharge. These can be arranged in a single cell, called solid-state photomultiplier(SSPM), or more often Silicon photomultiplier due to the commonly used material.

Multi-Pixel Photon Counters

Using the CMOS process it is possible to print a matrix of SiPDs in a chip, where every pixel corresponds to an independent APD in Geiger mode and a quenching resistor, assembled in a 2-dimensional array. These are commonly known for their usage in CCD sensors in ordinary cameras and are also used in Multi-Pixel Photon Counters (MPPCs). They can be associated with scintillators when positional information is of interest. Every pixel is highly sensitive, reacting to a single photon and creating a pulse of the same amplitude. If several photons hit different pixels at the same time, the output pulse will be a superposition of the signals, a single pulse with the sum of their amplitudes. If two photons hit the same pixel, a single pulse is output, so MPPCs need to be designed such that one pixel should only receive one photon at a time. To count the pulses, a charge amplifier circuit is used.

MPPCs are more sensitive to thermal fluctuations and spurious pulses. SiPMs in general are less vulnerable to external magnetic fields than PMTs and are immune to ambient light. Because of their technological overlap with other fields, their cost is lower, and very small sizes are available.

2.5.4 Detectors for neutron gamma discrimination

Any medium with a heavy flux of fast neutrons will have a certain amount of neutron-induced gamma rays. This creates background radiation that will show on the neutron spectrum that needs to be compensated. In nuclear fusion, key reactions (equation 2.2) involve fast neutrons and gamma rays, so neutron or gamma radiation is present in every measurement. One could choose detectors that are transparent to one of the types of radiation, but as was seen in section 2.5.1, mediation of uncharged particles means most good neutron and gamma detectors are sensitive to both of these types of radiation. The approach in these cases is then choosing detectors that show a different response to each particle, identifying the undesired ones as background radiation, and correcting the pulse counts or spectra.

Section 2.5.2 described how organic scintillators show a composite fluorescence response.

Though the majority have most of their energy in the fast component, those that have a good proportion of energy in the slow component can be used for pulse shape discrimination, discussed further in section 2.8. Stilbene in particular is widespread in pulse shape discrimination [54][55], as are liquid organic scintillators[56], most famously NE213[57] and similar ones such as EJ-301 [58] showing some of the best pulse discrimination capabilities.

Inorganic scintillators have found widespread usage, for example, with Lanthanium Bromide, LaBr3, coupled to MPPCs or a PMT being used for the JET gamma-ray chamber and spectrometer [59], but their usage for PSD isn't as prevalent as Liquid Scintillators. Though even the most popular ones, Thallium-doped Sodium iodide, NaI(Tl), and Sodium Doped Cesium Iodide, CsI(Na), show some capabilities[60], their longer decay times and poor separation energy resolution don't tend towards high-energy, high count-rates discrimination of neutrons from gamma rays. On the other hand, they can show a good separation of charged particles from neutrons or from gammas, in particular with Tellurium-doped Cesium Iodide, CsI(Tl) [61].

Diamond CVD is used as a neutron detector at JET[62], but it doesn't currently serve as a neutron-gamma discriminator. However, a recent area of development shows that similarly to inorganics, Diamond detectors can be used to discriminate some particles, notably gammas from charged particles [49]. Since neutron detection usually involves conversion to charged particles, this makes PSD theoretically possible in diamond detectors.

2.6 Pulse Processing

Diagnostics, described in section 2.4, are the motivation to extract information from a pulse. First off, the confirmation of the event is information in itself. Further, it allows for counting, which can then provide event rates. Integrating this signal adds up to the whole charge which, once calibrated to the energy response of the detector, describes the energy of the generating event. With enough counts, this allows for the construction of an energy spectrum. In some cases, the shape of the signal can also inform one of the identity of the generating event, as will be explained in section 2.8. The main goals of analysing pulses are then pulse counting and spectroscopy, and sometimes feature extraction, as well as particle identification and physical event description.

This section shows how a current pulse is processed into useful information, leading with a summary of the acquisition chain and following up with digital pulse processing.

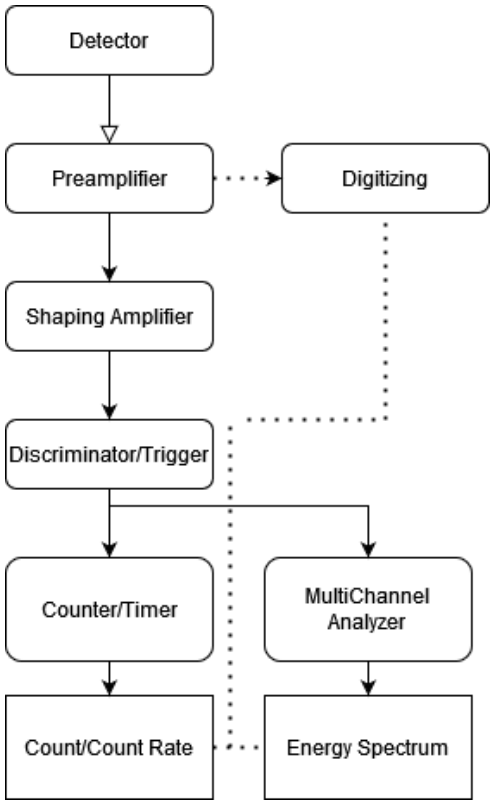


Figure 2.7: Basic Acquisition Chain for current pulses in radiation detectors, adapted from [36]

The basic acquisition chain is represented in figure 2.7. After a particle interacts in the detector, it is assumed that the signal is described by the pulse model. This means that a charge Q has been produced in the detector, as described in section 2.5, and the collection of charge carriers over a period of time produces a pulse, a signal in current that describes the

generating event. The Preamplifier has two roles, as a low-noise interface and as a converter into voltage. This means amplifying the signal as early as possible, so that transmission onto the rest of the electronics doesn't harm the SNR, by integrating the current in order to produce a voltage signal.

In traditional analog acquisition, the output is a voltage step with height proportional to the charge Q, which is then shaped in the Shaping Amplifier into a voltage pulse according to the decay constant of the amplifier, filtering for noise, with added shaping electronics that improve the quality of further processing. At this point, pre-processing logic can be present, rejecting events that can't be processed, such as piled-up pulses, those with heavy distortion or those caught in detector dead-time. Then the pulse is sent to the trigger/discriminator. These filter for pulses with certain characteristics, with the most basic integral discriminator filtering for pulses above a threshold and the differential discriminator, also called Single Channel Analyser(SCA), filtering for pulses between a minimum and maximum value. The output of these is a logic pulse. This logic pulse increments an event count if the height of the pulse is above a certain energy threshold, called a Pulse Height Analyser (PHA). Alternatively, a Multi-Channel Analyser (MCA) can select pulses according to their height/energy, and place them in an appropriate bin. The result is a histogram of pulse energies, or an energy spectrum. Electronic design is a major topic on its own, and outside the scope of this thesis. Kowalski[63] provides a comprehensive review of Nuclear Electronics.

2.6.1 Digitizing

Modern acquisition chains, after the initial pulse is pre-amplified and converted to voltage, then convert it to digital form (section 2.6.1) and use digital logic to replicate these analog processes. The added versatility of using digital allows for more complex and adaptable processing algorithms. In any case, the acquisition chain always includes an analog component for detection and pre-amplification, and sometimes initial shaping and pre-processing to match the electronics ahead. Then follows the digitizing and digital processing.

Digital processing has the major advantage of being able to preserve the shape of the whole pulse, which can be used to conserve its information as long as needed, and paves the way for more complex algorithms, digital filters that are otherwise impossible to create electronically, and more powerful off-line analysis that would take too long to implement in real-time. Redesigning an analog filter requires hardware changes, while a digital redesign requires a code update. Digitizing itself is mainly limited by data throughput and analog to digital conversion, and digital processing is limited by processing power, further discussed in section 2.6.4.

Analog to Digital Conversion

An ideal Analog to Digital Converter(ADC) takes an input analog voltage and linearly and instantaneously converts it into an output digital code. The main properties of the ADC are its frequency, the continuous rate at which it performs conversions, expressed in Hertz or megasamples per second; and its resolution or bit depth, n, the binary range of values it can output. An n bit ADC outputs a range of 2^n channels. The input range is mapped to the 2^n channels, each with a digital resolution of $1/2^n$. Choosing the bit-depth of an ADC should take into account the noise of the source, as it becomes meaningless to introduce more channels if the SNR is too low, as the noise would cause values to fluctuate beyond the correct channel, introducing errors when values are placed onto the wrong bin.

Since an ADC functions by taking a continuum of input values and placing them onto discrete channels, the limits of each bin introduce a baseline non-ideal behaviour, before even considering electrical noise. For instance, an ADC with bins of 1,2,3, and so forth would place 1.1 and 1.9 in the same bin, while 1.9 and 2.1 would sort to different bins, despite the smaller difference. In this sense, the linearity of an ADC refers to how well it matches the range of input voltages to each output channel. The linearity of an ADC can be studied by observing its response to a linearly increasing input voltage, seeing how the conversion rate per channel deviates from the expected - equal conversions per channel. This can be described by the differential non-linearity(DNL) per channel, the Root-Mean-Square(RMS) of all code-width deviations. Another common parameter for linearity is in units of Least Significant Bits (LSBs), the maximum deviation.

A detected pulse is a continuous-time signal, and digital pulse processing converts it into a digital representation, where the function is no longer defined over all points in time, only at the points it was sampled at. This means digital processing is limited by the ADC rate as stated by the Nyquist–Shannon sampling theorem. This theorem establishes the minimum sampling rate to digitally describe a continuous signal to be higher than twice the highest frequency component of the signal. Sampling at a lower rate would mean information loss, where information in the higher frequency components of a signal would be lost, or aliased into lower frequencies. On the other hand, if the sampled signal has a maximum frequency, then it can be perfectly reconstructed from the sampled values as long as the sampling rate is sufficient. This makes the speed of an ADC one of the most important parts of digitizing. The conversion time of the ADC, detector dead-time and memory management are the main factors that establish the time it takes to acquire one pulse, past the duration of the pulse itself.

2.6.2 Pulse Shaping

Between the speed of the detection interaction described in section 2.5 and the time it takes to collect the charge, pulses take a certain time to decay back to baseline. These variable decay times, also present in the preamplifier, introduce several limitations. If the decay constants of the detector and charge-collection are comparable, a pulse will not fully return to baseline before all its charge is captured, a phenomenon called ballistic deficit. The presence of a decay time in a detector means that after an event, while the detector remains in an excited state, there is a period where additional events are lost, called the intrinsic dead-time, or pulses pile-up, meaning two separate pulses overlap to some extent, limiting the information that can be extracted. Pile-up is further discussed in section 2.7.

The non-infinite decay time of a pre-amplifier that band-pass filters the pulse can cause it to undershoot, reaching negative values before returning to baseline. The DC-blocking nature of capacitors used to collect charge means that regular, unipolar pulses, have a baseline value below zero.

These physical limitations can be alleviated by filtering the pulse, adapting it to the decay times of each component. For example, extending the pulse, increasing the time it spends near its maximum value, increases the accuracy of measuring its peak height, which means increased energy resolution. Shortening the pulse can allow for full charge collection, reduce the tail and prevent it from interfering with following pulses, reducing the dead-time of detection and pile-up. A certain combination of these filters, both analog and their digital equivalent, or purely digital filters can be used to reduce the effect of these limitations. For starters, analog filtering is usually present in the pre-amplifier. CR circuits or differentiators, and RC circuits or integrators function as high-pass and low-pass filters, respectively. These remove signals outside the range of frequencies of interest in the pulse.

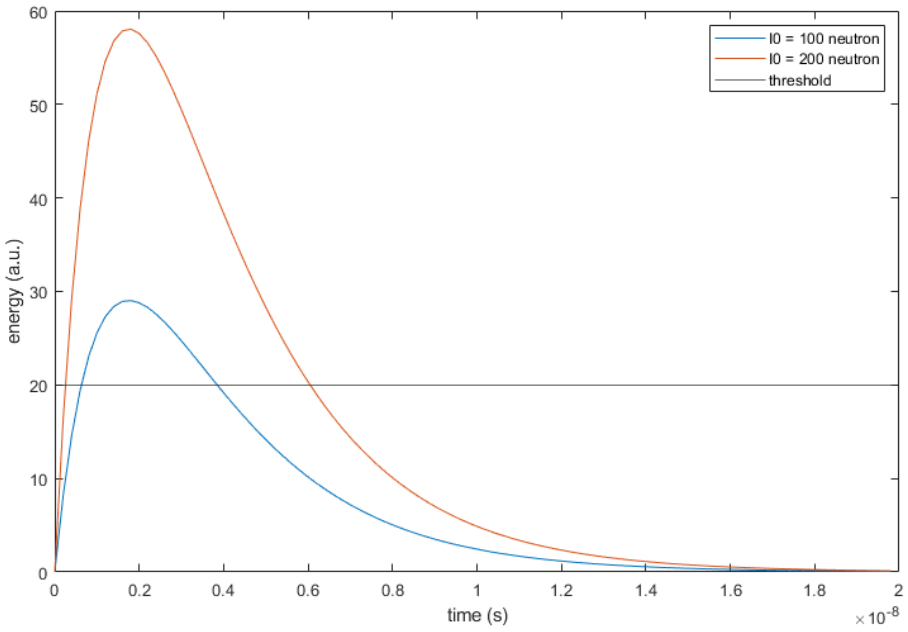
A CR-RC filter in sequence, with appropriate decay constants, can shorten or lengthen the pulse in order to achieve the desired balance of reduced pile-up, ballistic deficit and SNR, described by equation 2.10 as a response to a step impulse [36]. A series RC filter after this, $CR{-}(RC)^n$, shapes a pulse into a Gaussian, whose symmetry provides a faster return to baseline and flat top improves peak resolution. A triangular active filter has similar advantages to the Gaussian, but has a smaller peak and improved SNR. Trapezoidal filtering allows for pulses to have a certain margin of tolerance, where the flat top compensates for variable charge collection time, allowing pulses to have the same height. Otherwise, different collection times, such as in the impact position-dependent semiconductor diodes, would produce slightly different heights, degrading energy resolution.

$$E_{out} = \frac{E\tau_1}{\tau_1 - \tau_2}(e^{-t/\tau_1} - e^{-t/\tau_2}) \tag{2.10}$$

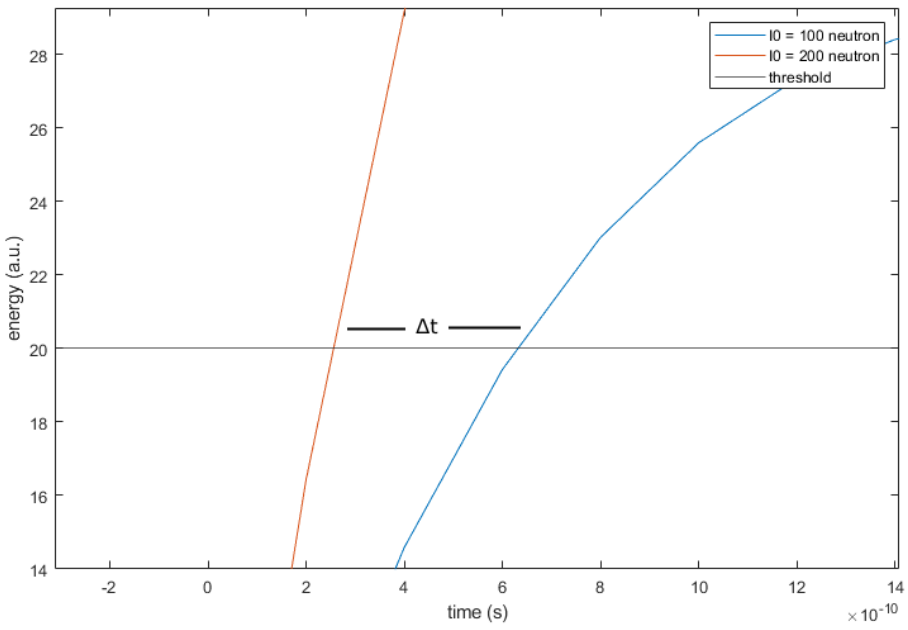
Passive filters, employing inductors and capacitors, and active electronic filters, which add transistors (introducing binary logic and amplification), both still treat the signal analogically. Digital filtering uses only transistors and switched capacitors to modify the signal. There is a mathematical component of digital filtering, consisting of replicating or approximating the transfer function of analog filters in digital logic. Beyond providing an alternative method to filters, and the reliability of having a digital signal, the greatest asset of digital processing is the greater versatility in the types of filters that can be used. Transfer functions are no longer limited by electronic components. For example, a finite duration impulse response, a perfectly square wave, programmable filters, and error detection are possible only through digital processing. An introduction to Digital Signal Processing and Filter Design is provided by Shenoi[64].

2.6.3 Event Detection

Whether an acquisition system is continuously acquiring data, or only registering only data where there is relevant information, there needs to be a way to detect when a particle has interacted with the detector. Because of this, a method to detect an event is a mandatory step for acquisition. The role of creating a logic signal at the moment a pulse is detected is carried out by a time pick-off unit, also called a trigger. Any trigger is vulnerable to fluctuations caused by noise and from a different response to a variety of input pulses, such as time jitter and amplitude walk. The simplest digital trigger is a leading edge threshold, that activates once a pulse crosses above a predetermined value. This is represented in figure 2.8a. There, the effects of amplitude walk can be seen: an arbitrary pulse with twice the amplitude of another will be triggered at a different part of their rising exponential, resulting in inconsistent timing if the amplitudes or pulse shapes vary too much, so this method is best used for more limited dynamic ranges. Time jitter happens due to random fluctuations in the signal, where the threshold can be triggered too early or too late due to these fluctuations, which introduces an interval of uncertainty in the correct trigger time. Because of this, in order to improve timing consistency, some authors use the absolute maximum value in a pulse as a reference, defining the timing based on the time when a signal reaches a relative level, such as 20% of the maximum [65].



(a) Comparison of two neutron pulses, leading edge trigger.



(b) Detail of the threshold, Amplitude walk of Δt .

Figure 2.8: Neutron pulses of identical rise, decay time, with different amplitudes.

If a signal is shaped into a bipolar pulse, what used to be the maximum value is an inflexion point that turns into a crossover point, where the shaped signal turns from positive to negative. This property is used in crossover timing: digitally, interpolating between the time of the last positive and first negative value provides a timing method. In analog circuits this is achieved via RC-circuits, discussed in section 2.6.2, while in digital algorithms this can be achieved at its core with a numerical differential and a linear interpolation, making this algorithm computationally fast. This method is more consistent than leading-edge triggering but introduces more noise and jitter.

Under this umbrella can also be included the Differential Pulse Discriminator(DPD) approach, where the first derivative of a pulse is compared to a threshold value. Values above the threshold signal the presence of an event[66].

In a leading-edge trigger, the best results come from using a trigger level of 10-20% of a pulse's maximum. But the variety in pulse amplitudes means that this isn't constant. Thus, constant fraction timing sets a trigger for each pulse as a fraction of its amplitude. Arc timing also triggers at a constant fraction, but since the leading edge is constant among particle types, as seen in section 2.5.2, this method assumes a regular rise time, linearly regressing to the origin of the pulse.

Extrapolated leading edge timing is similar to this, using two independent discriminators to detect an approximately linear portion of the leading edge of a pulse, and then from the difference in time and amplitude extrapolating (non-linearly) to the origin of the pulse[36]. A similar approach to this, the Lower Level Discriminator(LLD), detects the presence of a pulse by comparing deviations to the predicted value of a signal. For a single event, it scans for sudden spikes compared to the average baseline; for the second event in a 2-fold pile-up event, it uses a deterministic model to determine the expected value of the first pulse, scanning for spikes compared to that expectation[66].

Another caveat of digital triggers is that the acquisition frequency introduces a minimum time uncertainty in the trigger equal to the sampling period, the inverse of the frequency. Again, this means that higher sampling frequencies are advantageous.

2.6.4 Digital Processing Units

DAQ historical overview

The advent of digital processing is tied to the appearance of new technology, which made traditional analog alternatives less attractive in comparison. In this sense, the development of the theoretical basis between analog and digital information in pulse code modulation and later the publishing of A Mathematical Theory of Communication by Claude Shannon[67], and the development of the transistor in 1948 were the first steps into digital processing. As this technology matured, digital progressively overtook analog communication. Printed circuits, the MOSFET in 1974, data compression and fast ADCs were some of the pivotal events that changed the paradigm of communication theory. Nuclear acquisition systems followed the same path and, at least partially, adopted a digital mindset.

The first applications of digital processing rose up as modular replacements for specific functions in an acquisition/processing chain, where digital was superior for the task at hand. Because of this, DPUs(Digital Processing Units) inherit the hardware standards of previous instrumentation. The first of these standards was the Nuclear Instrument Module (NIM) system. The NIM standard in 1969 first defined electronic components as modules such as a preamplifier, integral discriminator or the trigger units that were mentioned in earlier sections. It set their standard electrical/mechanical specifications and of the hardware around them, so that they could be easily powered, interfaced with, and assembled into a rack. This modular construction, as well as the compatibility it afforded, made designing and constructing a whole system from scratch much easier. Later, the Computer Automated Measurement and Control(CAMAC) standard was established, adding a data bus and creating an internal way to access data and communicate between the modules. This facilitated the design of systems which prioritized data processing. For even more data-intensive applications, and those requiring added automation, the VMEbus Standard was a computer bus adapted to fit nuclear instrumentation purposes. The modules have direct memory access to data, while analog interfacing is relegated to the back panel. CAMAC and VME both implement a parallel bus for cross-module communication. This is great for data transmission, but a shared parallel communication bus has disadvantages: it is bandwidth limited by the number of modules used, only two modules can communicate at a time and it is more vulnerable to malfunctions. A serial bus is cheaper and not as limited, but asynchronous communication introduces overhead processing, and it is point-based so only directly connected devices can communicate.

With an increased focus on digital, these systems have evolved to accommodate changes to

the state of the art, such as the parallel/serial communication duality, increased demand for I/O slots, more data throughput and module size requirements. For example, VME was continuously updated with standard extensions, and found a backwards-compatible successor in VPX (VME extension specifications). Modern design philosophy has changed to prioritize some type of interfacing with a general-use PC(Personal Computer), leading to updated standards. Many different implementations coexist in the paradigm of nuclear data acquisition systems, often in the same laboratory. The JET experiment, for example, simultaneously hosts CAMAC[68], VME[69], FPDP(Front Panel Data Port, a VME standard extension), PCI(Peripheral Component Interconnect)[70], and ATCA[43].

The first digital processing modules, due to their novelty, were integrated into a mostly analog system, its digital part enclosed in a black-box, with both input and output being analog. These DSPs(Digital Signal Processors) were processors responsible for accelerating some workload using digital processing. In their most basic form, they required an ADC, a microprocessor, then a DAC tasked with converting the signal back to analog after processing was done. This kept the system identical on the whole, while still being able to take advantage of a specialized circuit with digital algorithms. Modern systems emphasize the importance of early conversion to digital, often placing ADCs and time-stamping early in the acquisition chain, which allows the rest of the system to be digital, with all of its advantages.

When designing a digital processing unit, first comes the theoretical component of digital processing, meaning the mathematical design of filtering and shaping, how it manipulates the data to serve a function. Once that has been decided there comes a decision of what type of implementation to use. Is the system designed for a single experiment? How much will it cost and when is the deadline? What are the Input/Output requirements? How much data will be produced, and does it have to be stored? The choice of implementation, therefore, has to take into account a lot of aspects of the system, such as the instrumentation standard, Input/Output, power requirements, as well as data requirements like processing speed, bandwidth, and memory. The choice can range from a general-purpose consumer PC, with processing done in the CPU and I/O being as simple as a commonplace USB port, to a system that implements expensive and design-intensive hardware to speed up a specific operation.

In the following sections, some of these implementations are discussed. All of these are based on integrated circuit chips, but they each find their specific use-cases, fitting their respective design architecture. The trade-offs can involve several factors. For example, better performance can come from higher clock rates and transistor counts, but these increase power consumption. Lower development costs and flexibility can be achieved with general-use devices, but these can't

match the performance of specific hardware optimized for a task. In addition, it should be taken into account that designing a device for the first time has an incredibly high non-recurring cost. Programming an existing general-purpose device, with well-established production lines, can be advantageous in terms of cost and design time.

DSP

Digital Signal Processors are processors specialized in digital processing tasks. Processing tasks can be divided into data manipulation (saving and loading to memory, logic comparisons, etc) and mathematical operations such as summing and multiplying. The typical workflow in a DSP consists of moving data in, including program data and the target data for processing, performing mathematical operations on it (running the program), and moving the data out. The mathematical operations performed are those required in data processing tasks. These tasks are based on filter implementation and frequency domain operations such as Fast Fourier Transforms and convolutions. Essentially, this involves a high number of somewhat independent multiplications and additions, a decent overhead in data throughput, with little data manipulation. DSPs are designed bottom-up in order to optimize these tasks. The simplest implementation of a finite response filter, for example, requires loading program data and filter coefficients, after which follows a loop of loading data point by point, multiplying it and accumulating it. Another example, fundamental for DSP, is the algorithm for convolution: invert an array, then the core computation loop is multiplying and summing point-by-point, and shifting a register each iteration. Or, if performed in parallel, several multiplications and accumulations performed at the same time.

Consider, for the sake of contrast, a general-purpose task like document processing. The bulk of computing cost lies in data manipulation: opening a file from memory, rendering the words in a certain order, manipulating the words and saving them to a file, with a longer time margin to perform each operation. These tasks have minimal mathematical operations, instead prioritizing data manipulation, so a typical general-purpose CPU, such as the one used to write and develop this project, is not specialized for a specific task, instead being very good at data manipulation. For example, this Intel x86 CPU employs a modified Harvard architecture, which has unified memory for data and code (program data), meaning an excess of throughput in either can limit the other. The serial nature of these general-purpose operations means that the single-core speed is the biggest predictor of a general-purpose CPU's performance (unless the bottleneck comes from running several different programs at the same time). Engineering limitations set a soft cap on single-core frequency, and as developers work around this by adopting features

previously exclusive to a processor implementation, the concept of what defines a DSP versus a microcontroller or CPU is becoming harder and harder to define.

A digital processing task doesn't need to focus on memory size, as once all the required math is performed, the data used to calculate it is no longer needed for the task. Accessing new data (receiving an address from a source or sampling a new value) and outputting an answer both create requirements in data, but unlike general tasks, the requirements lie in how much of it can be input and output, meaning data throughput, not data manipulation. For a DSP, after loading the data, most of the processing time is spent on the mathematical operations part. Even in the case of real-time operation, the processed data often doesn't require storage, only the results do. In these cases, data needs to be loaded fast and continuously as the program runs, again pointing towards data throughput and parallel operation.

The core of a DSP unit refers to the area in the processor responsible for the main tasks. This includes the program sequencer, data registers, address generator, multiplier and arithmetic and logic unit. What separates the core from a complete processor is that the latter also takes into account memory and interfacing with other components. A DSP unit can be a complete processor, or implemented as a part of another system with another CPU managing it, all part of the same integrated chip, either embedded in the CPU or as a separate board[71].

With a general idea of the use-case of DSP units, the hardware commonly featured in them can now be better understood: circular buffers specialize in real-time operation since the processed data is continuously input, unlike offline processing where all the data can be addressed from the start; multiply-add cores can optimize the highly repetitive dot-product operations mentioned earlier, which can be accelerated by performing the main loop in parallel; high-speed I/O and independent memory access focus on data throughput and skipping CPU involvement as much as possible, to minimize data transfer bottlenecks. These types of adaptations are aimed at building a strong hardware basis for faster parallel computing, which is the foundation of digital processing.

Strictly hardware-based solutions are limited to operations that can be performed with logic gates, but having some signal processing functions hardwired results in much better speeds, though only for those tasks. DSP code is usually written in C/C++/Assembly, so it can be changed or optimized, but the hardware is final. If a device implementing a DSP needs to be changed, no more hard-wired functions can be added, which can limit its usefulness.

MCU

Microcontrollers(MCUs) can be seen as a more general purpose version of DSPs, but responsible for controlling a peripheral. MCUs don't have optimized hardware to the level of DSPs, but since most digital algorithms can be adapted to run in any type of processor, the flexibility of MCUs results in a reduction in speed, not in function. This is common to any digital processing. Typically working at lower frequencies, with smaller power consumption, and with a proven legacy, microcontrollers are often used in applications where low-cost, low-power, autonomous operation of a peripheral is required.

ASIC

The hardware optimization present in DSPs, when taken further, culminates in Application Specific Integrated Circuits(ASICs). ASICs are designed from the ground up for a specific application. This means that there is no redundant part in an ASIC, and no component goes unused. Its hardware can be optimized for any criteria, sometimes even implementing other DPUs. This results in the highest speeds, smallest size and lowest power consumption of any DPU, but comes at the cost of having the highest development times and no flexibility.

FPGA

Field-Programmable Gate Arrays (FPGAs) are integrated circuits designed to be configurable after implementation. They include logic blocks which can be altered with a digital input. This makes them capable of the same type of hardware optimization that ASICs and DSP units use, but since the logic is (re-)programmable, they gain the flexibility of being continuously adaptable. They stand in an unique position, merging software and hardware implementations of digital processing.

Unlike a single-core processor running code sequentially, a digital algorithm can be implemented in hardware such that parallel bottlenecks are managed by adding more logic. Therefore, its programmable logic means FPGAs lend themselves to parallel computing.

Modern FPGAs compete for best-in-class in terms of high-speed data processing. Though digital in their nature, some FPGAs can even include analog input ports. These features, when combined with its modular nature grants FPGAs great versatility, interfacing, and memory management. General-purpose processing units have the advantage of a higher user base and reusable code, which further reduces overhead from hardware design.

Some of the limitations of FPGAs come in the form of cost and energy usage. Programmable logic needs redundancy, which requires a greater number of transistors, and leads to bigger chips,

higher energy usage and higher price. It's important to mention that FPGAs require specialized hardware programmers, which can lead to some of the longest development times and high costs.

PC

A personal computer(PC) is an end-point for human interfacing, placed at the end of the acquisition chain. However, digital algorithms aren't specific to any device, and a PC can in some cases take over the functions of a DPU or controller. This is called native processing, and is the cheapest implementation of digital processing, as it only involves sharing the computing load of an already existing device, as long as interfacing is possible. Digital processing in a PC is purely a software endeavour, which limits its optimization options. Additionally, sharing computing loads with processing algorithms, interfacing and control in the same processor can result in bottlenecks. Even with modern parallel computing, PCs can't match the speeds of dedicated processing units, but their minuscule barrier-of-entry grants them a useful niche - a PC should already exist in the environment, they're used in development, human interfacing, data analysis, etc. This thesis was developed and written in its entirety in a PC.

Common PCs include PCI-e (PCI express) interfaces, which can be used to connect to peripherals. Of key notice in digital processing is the usage of Graphical Processing Units(GPUs). These are specialized components originally designed to render digital images, but since there is a high degree of intersection in the type of computing logic required for most types of digital processing, GPUs find themselves used in diverse applications. Offloading digital processing from the CPU of a PC is one of these use-cases, especially in the case of Machine Learning and Neural Networks.

2.6.5 Real-time case-study: MARTe

In the previous sections, several types of hardware implementations for digital processing were presented. The diversity of options available, all capable of the same type of processing, makes the choice depend on several factors, not just the technical requirements such as processing speed or total memory, which gate some of the choices from the start. The development environment; deadlines; available resources, both in budget constraints and human capital, can all be decisive factors in making this decision. An issue arises if e.g. a real-time control system is developed from the ground up each time, aimed at a specific hardware choice: non-portable, non-reusable code is created with each new iteration, increasing development time and costs.

Created with this issue in mind, and based on this relative degree of independence between the environment, the specific hardware, and the software (real-time algorithms), MARTe (Multi-

threaded Application Real-Time executor) is a C++ framework for developing real-time critical applications[72]. MARTe has been especially relevant for fusion control systems, having been used in "the JET Vertical Stabilization system, which uses the Real Time Application Interface (RTAI) operating system on Intel multi-core processors; the COMPASS plasma control system, driven by Linux RT also on Intel multi-core processors; ISTTOK [Instituto Superior Técnico TOKamak] real-time tomography equilibrium reconstruction which shares the same support configuration of COMPASS; JET error field correction coils based on VME, PowerPC and VxWorks; FTU LH reflected power system running on VME, Intel with RTAI." [73].

It's a cross-platform framework, having already been ported to "Wind River VxWorks, Linux, Linux/RTAI, Solaris, and MS Windows"[73]. In MARTe, GAMs (Generic Application Modules) serve as modular software blocks, whose execution is managed by a real-time scheduler. The scheduler defines how several GAMs work together, chained in parallel or in sequence to perform the required work, each being typically responsible for a specific system requirement. GAMs aren't immediately associated with the hardware implementation, instead defining their I/O and data access requirements in their configuration.

The environment in MARTe defines how interfacing with hardware is done, as well as the services that manage memory access and I/O (managed by a special module, IOGAM). The platform is the specific Operating System and processor used, where hardware optimization can be done. MARTe is made with this clear separation into three layers in mind, with the framework taking the role of managing their interactions. This structure lets the development of a system be done in parallel, phased out over time and split into different specialized teams. Furthermore, since the target platform is independent of the software (GAMs), the user code can be reused and cross-tested across platforms, granting it portability and robustness. Since the configuration of GAMs only defines the data input and output requirements, and with the framework being capable of simulating parts of the system, testing is made convenient, with the possibility of testing real-time code in an offline environment.

This section is included for two reasons: first, to showcase a real case of a development environment where this project could be integrated. The simulation part in chapter 3 has a similar philosophy to MARTe's goal of enabling the testing of real-time algorithms. Secondly, the algorithm shown in chapter 4 aims to be capable of real-time operation. So an analysis of how it could be created in a GAM, namely what parts of the pulse processing algorithm can be optimized to process in parallel, is an indicator of how viable it would be implemented. This is discussed in section 5.2.1.

This framework has a successor in MARTe2, with the upgrade focusing on Quality Assurance.

The original framework is already focused on the developer's ability to test, while the new one builds on that by standardizing the C++ version to ISO/IEC 14882:2003[74], making code compliant to MISRA C++:2008[74] and incorporating an agile methodology[75].

2.7 Pulse Pile-up

Radioactive decay is approximated using the statistics of a Poisson distribution, a valid approximation as long as the decay rate is not significantly affected as a material progressively decays. Although not universally valid, for event counting with low-efficiency detectors the approximation holds[76]. In the context of nuclear fusion, radiation sources (section 2.2) usually have short decay times, in the order of nanoseconds, lower or comparable to those of the instrumentation, and orders of magnitudes smaller than the duration of a plasma discharge, in the seconds range, making the approximation valid.

Obeying a Poisson distribution means events occur spaced randomly over time, described by an average rate. Events, defined in section 2.6.3, occur independently of each other, which means that two pulses can occasionally occur simultaneously, interfering with each other. This is called pile-up, a random coincidence of 2 or more events that can affect pulse detection. It can be classified as tail or peak pile-up, depending on where the superposition of pulses is located. Pile-up increases with event rate and pulse length and can be estimated if those factors are known. Pile-up occurs if a second pulse appears before a previously occurring one hasn't fully resolved. From this observation, the amount of first-order pile-up can be estimated from the probability that a pulse happens before a standard pulse length. The probability of observing a time between consecutive pulses in a Poisson distribution is given by equation 2.11, where Δt is the time between pulses and n is the event rate[36].

$$P(\Delta t > PulseLength) = exp(-n * PulseLength) \tag{2.11}$$

Tail pile-up happens when a pulse is detected while the long tail of a preceding pulse is still ongoing. In this case, although the pulses are distinguishable through their peaks, there is still an effect in the spectrum. As was seen before (section 2.6), information on the timing and energy of a pulse is present in the rising edge, so the first pulse is unaffected in that sense. However, the residual tail of the first pulse, or even its undershoot if the baseline energy level hasn't been restored, will affect the height of the second pulse, reducing its energy resolution. The result is the appearance of inaccurately detected events. These false counts have energy levels corresponding to the sum of up to twice the energy of each pulse, depending on how much overlap there is between the piled-up pulses. A single energy peak source would produce

a secondary peak at twice the original energy, a tertiary at thrice the energy, and so forth. This effect is visible in a detection spectrum corrected for pile-up in image 2.9. In addition, a single pile-up event corresponds to at least 2 real events, affecting not only the individual energy resolution and the spectrum, but counting statistics as well.

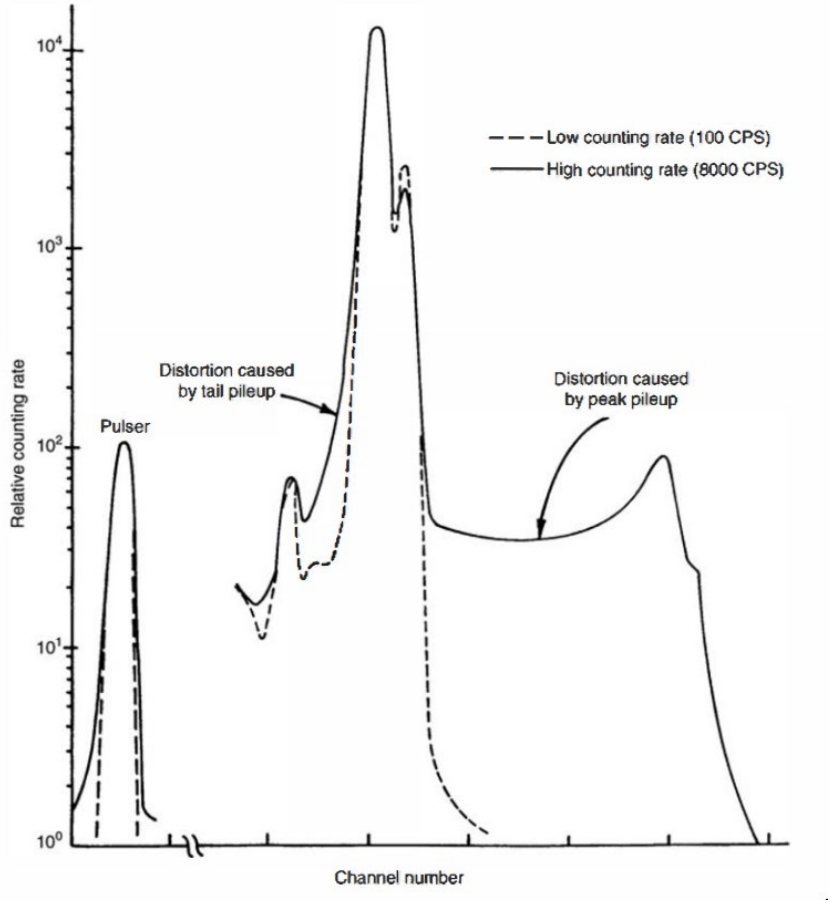


Figure 2.9: Spectral distortion due to pile-up in a low vs. high count rate ^{55}Fe spectrum. Image from [77], adapted from [36]

Pile-up happens as a continuum, not strictly at the same time, so the partial superposition of pulses also has the additional effect of flattening the real peaks, as some pulses will have their energy partially inflated by the tail of previous ones. This can also happen if baseline energy levels aren't fully reset after each event. Pile-up rejection and correction methods are discussed in the following sections 2.7.1.

Longer decay times at any point in the acquisition chain increase the duration of each pulse, therefore increasing the likelihood of pile-up. Shorter pulses are desirable if pile-up minimization is a goal. This can be achieved in some ways, but as mentioned earlier, the decay constant of a pulse originally comes from the detector itself (section 2.5.2), which is then affected by the electronics of the acquisition process, in particular when it comes to charge collection as seen in

sections 2.5.3 and 2.5.3, and pulse shaping for pre-amplification (section 2.6.2). Shaping a pulse with a longer decay time introduces longer tails.

In counting statistics, there is then a certain degree of data loss that is unavoidable, introduced from different sources: pile-up and the inherent dead-time from the detector, circuitry, and ADC conversion and storage [78].

Besides affecting the duration of pulses, detectors and electronics alike can have a dead-time, a period of time after an event where additional events are discarded/affected. In paralyzable systems, if a second event is detected before the resolving time τ, no output is produced and the resolving time is extended. The system is paralyzed until it has been allowed to experience some time without events. In non-paralyzable systems, the resolving time is fixed, not extended if new pulses occur during relaxation. At low event rates, these models are identical, but at high enough event counts, paralyzable models eventually stack enough successive paralysis per event that a higher event rate can actually reduce the total counts. This is not the case for non-paralyzable models, as dead-time is not extended. Hybrid models with more realistic results are an active area of research, recently reviewed in [78].

Pile-up processing can be tackled at different points in the acquisition chain. It can be prevented, rejected or corrected. Since the decay constant of the pulse is the main driver of pulse length, its changes over the acquisition process, especially its shape after the pre-amplifier, are vulnerable to manipulation. In section 2.6.2 it was explained how pulses can be shaped according to a design objective. In pile-up reduction, the objective is preventing pile-up, so they can be shaped into a shorter duration to reduce coincidences, as demonstrated by equation 2.11; or flattening the top of a waveform to reduce the ballistic deficit and improve energy resolution, for example with a triangular or trapezoid filter [79].

Pile-up correction methods take advantage of digital methods, after digitization, and have a common factor of being computational in nature. These have the objective of deconvoluting piled-up events, either figuring out their individual amplitudes (for PHA) or deciphering what single pulses would have generated the acquisition event, and are discussed in section 2.7.2.

2.7.1 Rejection

Pile-up rejection relies on identifying which, or how many pulses contain pile-up and discarding them as a way of removing the spectral distortion they cause. Various methods can be used for this purpose, grouped under the umbrella of pile-up rejection. An alternative lies in preventing it with pulse shaping, or attempting to recover information or reconstruct the individual pulses that originated the piled-up signal, discussed in section 2.7.2.

Any method has limitations, and even correction methods have a minimum time difference between pulses in order to discriminate them. Pulses that occur simultaneously, too close for the time resolution of a system, inevitably produce a secondary peak. These types of effects can't be detected on an individual basis, so some authors have taken to spectral corrections, such as the mathematical approach of equating pile-up to extending dead-time[80] and Monte Carlo Simulation to predict true spectral response[81].

Lindstrom[80] interprets amplifier pile-up as a mathematically equivalent case of extending dead time, estimating the amount of pile-up and arriving at equation 2.12 for a rate correction, where I is the output rate; I_0 is the input rate, derived from the ADC dead-time; and k is a constant. Amplifier dead-time correction is a part of accurate counting statistics, making this a simple statistical approach extended to pile-up correction. Variations of this approach, of statistically compensating pile-up-related counting losses in a similar approach to dead-time correction, are found in nearly every acquisition system, often integrated into the acquisition system itself.

$$I = I_0 * exp(-k * I_0) \tag{2.12}$$

A simple method of pile-up detection involves parallel processing of events with a fast and slow branch. The fast branch is made to process the pulse as fast as possible, only serving to detect the presence of pile-up. If pile-up is detected, it rejects the event. Otherwise, the slow branch normally processes the pulse.

A common hardware solution to this is including a periodic pulser at a low frequency. Since these have a known regularity, they are not susceptible to pile-up, and a correction factor can be determined from the number of missing pulser events[80].

Undetected pile-up can still occur during the processing time of the fast branch, and in any case, rejecting events always negatively affects the spectrum accuracy, since fewer events can be used.[36]. Rejection isn't an ideal solution, but it is a necessity for accurate spectroscopy.

2.7.2 Deconvolution algorithms

Pile-up rejection doesn't add significant information for counting statistics, only how many events have pile-up detected in them. Then they can be discarded, or an attempt can be made at recovering some information from them. Separation or deconvolution algorithms attempt to separate individual events from a pulse containing pile-up. This is possible thanks to the advent of digital methods, as mentioned in section 2.6.4. When approached by a pulse with 2-fold pile-up, an ideal deconvolution algorithm separates it into the two independent events that generated

it, which allows for pulse processing to be done on it. Another approach is interpreting a 2-fold event as a whole.

Methods like Template Matching(TM)[82], Maximum Likelihood Estimation(MLE)[83], High-Yield pile-up event recovery(HYPER)[84] and Single Event Reconstruction(SER)[84], are based on the linearity and time-invariance of the acquisition system. In these conditions, piled-up pulses are an arithmetic sum of two independent events. As a result of this, if the model for the pulse shape is known, as well as an accurate and consistent starting point for each pulse, some deconvolution is possible, obtaining the pulses that would have generated the piled-up signal. Note that TM and MLE can also be used for pulse timing.

These types of methods have a few limitations. They are accurate for the model they are adapted to, and there is no generalized pulse shape, as it is characterized by the convolution of the detector output and the response function of the acquisition system in question. However, in these conditions where the linearity of an acquisition system is assumed, a 2-fold pile-up event can be parameterized with only the first and second pulse heights and the offset between them. This creates a common opportunity for all methods under these conditions: if the pulse shape is known, all the pulse parameters can be fit and optimized, which in turn describes a family of fitting algorithms for the deconvolution of piled-up signals that can range widely in complexity, depending on the computational requirements. The simplest optimization that can be considered is least-square-fitting these 3 parameters, for example in Guo's work with the organic scintillator exponential shape[66]. In this 3-parameter approximation is implicit the calibration of the acquisition system. Otherwise, the decay constant of the system can also be interpreted as another parameter to optimize.

Some authors avoid this pulse shape dilemma by parametrizing by opting for more general models. A set of calibration pulses can be used to determine the empirical pulse shape, such as with spline interpolation, where a pulse shape is decomposed into polynomial segments. [66][85]

Yang Xiao-Feng et al[86] use more complex search methods to find these parameters in a generalized double Gaussian pulse model. Artificial neural networks have been used for this purpose [87][88][89]. An issue with increasing complexity is that real-time processing might become impossible, invalidating some of these algorithms' possible uses.

The bi-exponential decay for organic scintillators (eqn. 2.5.2) is commonly used to parameterize pulses. Diamond detectors, however, have variable charge collection time, depending on the physical location of the particle interaction, so this parameterization loses accuracy. Marrone et al obtain a signal shape from the convolution of the impulse response of the readout system with the scintillator shape[90].

Single Event Reconstruction(SER) is based on reconstructing a single pulse based on a known pulse shape and subtracting it from the waveform, revealing in this process the next event, without the coincidence. This is done over the whole waveform, each pulse in succession, iteratively[91]. Though similar to the following methods, SER actually provides a reconstruction of each individual pulse, not just the heights as is the case with Template Matching methods.

Template Matching(TM) assumes the linear time-invariance to normalize a fixed pulse shape. Each piled-up waveform is therefore a linear combination of these events, scaled in amplitude and time between events and can be fitted accordingly, and the height of each original pulse is determined. The most common approach is least-squares fitting to find the values that best fit the reconstructed values to the original piled-up waveform heights [92]. Weng[93] uses a simplified model that considers the pulse shape as the convolution of an incident gamma-ray with the detector response matrix. Since the latter is linear, it is possible to pre-calculate it, so the deconvolution is a matter of solving for a matrix of parameters such as noise and delay, essentially inverting the convolution.

Other than an imperfect template or a non-naive approach that doesn't presume linear-time invariance, the only room for error here is in the parameters. However, if the matrix gets too big or too sparse[94], it can become computationally problematic to process the inversion algorithm.

Maximum Likelihood Estimation (MLE) is a subtype of Template Matching that characterizes a piled-up waveform as a linear combination of pulses, modelled by a stable template. It can use an external source of timing information to supplement the matrix. Its increased matrix size makes it require more resources to calculate[84]. High-Yield Pile-up Event Recovery (HYPER) assumes the shape of a single event to be a single exponential decay, instead of the common double-exponential. In HYPER the first signal in a piled-up is integrated until a second event is detected; the remainder tail is compensated, and the process continues onto the following event[95].

Mohammadian-Behbahani et al [84] provide a numerical comparison of this family of pileup resolution methods. They conclude that no single method is perfect for each use-case. An informed decision should be made according to the required capacities, whether it be noise resistance, robustness to timing variance or charge collection, throughput, etc.

Kafaee et al[96] introduce genetic algorithm and artificial neural network approaches to pile-up resolution, which show promise in that they don't require taking into account the hardware pulse shaping that characterizes traditional methods. These are more generalized solutions with comparable accuracy to state-of-the-art methods.

2.8 Pulse Shape Discrimination Algorithms

In section 2.5, it was discussed how particles deposit their energy in a detector at a rate that is characteristic of their properties, creating a characteristic curve. From the shape of this pulse, some properties of the incident particle can be determined. The identification of these properties, namely the type of particle that generated an event, is called Pulse Shape Discrimination.

These methods rely on an acquisition setup producing a reliable shape, which is semi-empirically justified in the case of scintillators, as seen in section 2.5.2, and also in Diamond detectors[49]. This provides some crossover with pile-up separation methods: if a pulse can be parametrized or compared with a template, it can be compared to a model or calibration pulse and exploited for classification. In the case of PSD, instead of amplitude and offset, a parameter like the rate of exponential decay might be appropriate. For instance, Chandhran et al[97] demonstrate the use of the cross-correlation with a reference pulse to perform neutron-gamma PSD. Alharbi[98] uses Principal Component Analysis to classify neutron or gamma based on the cross-correlation of individual pulses with a reference covariance matrix, created using a calibration run.

Neural Networks are an area of increasing interest for PSD with great diversity of methods, some even relying on the same type of parametrization[99].

As is a pattern throughout this work, digital methods can replicate analog methods. Sosa et al[100] compare the performance of digital and analog PSD head-to-head, with digital methods coming ahead in discrimination power (measured with eq. 2.14) across the board.

The following sections focus on describing individual pulse shape discrimination methods and then PSD in the context of piled-up signals.

2.8.1 Zero-crossing

Roush et al[101] demonstrated in 1964 that the bi-exponential approximation (eq. 2.5.2) of the output of an organic scintillator and PMT combination meant that the maximum in a pulse is driven by the relative amounts of the slow and fast components. This creates a peaking time that is independent of the pulse amplitude, being driven only by the decay time of the scintillator and PMT, besides the light characteristics. By RC-differentiating this signal, the peaking time corresponds to the moment of change into the negative values, which can become a timing method for this position. Roush used a tunnel diode as a baseline stabilizer to remove undershoot, then integrating the signal while it is still positive. This integration creates a signal proportional to the time during which the signal was positive. Comparing this current

integration to set thresholds allowed for the separation of particle types in electron+proton, gamma+alpha and neutron+gamma signals.

Pulse shaping in an analogue system requires precisely designing the whole acquisition chain to match decay times according to the principal objectives. Digital methods remove most of this burden, and can accurately replicate analogue methods[102].

2.8.2 Charge Integration

This technique is based on the same assumptions as the zero-crossing method. However, in this instance, the slow and fast components of the double-exponential are integrated separately. Again, the slow portion of the exponential tail is related to the type of incident particle, so the proportion of charge accumulated in the slow tail, compared to the total charge, is indicative of the type of particle[88].

A trapezoidal filter has also been used for PSD, using the characteristics of a trapezoid-filtered waveform (shortened pulse, flat top) to improve some of the variation in timing and energy resolution. This method defines the integration lengths differently, but could be seen as an improvement on CI methods. [103]

2.8.3 Pulse Gradient Analysis

Pulse Gradient Analysis (PGA) classifies particle type according to the gradient between the peak amplitude of a pulse and the discrimination amplitude. Discrimination amplitude is defined as the pulse height after a predetermined time period. This method gains speed from only needing to analyze the signal at and near the peak, saving some processing in the long tail. [104]

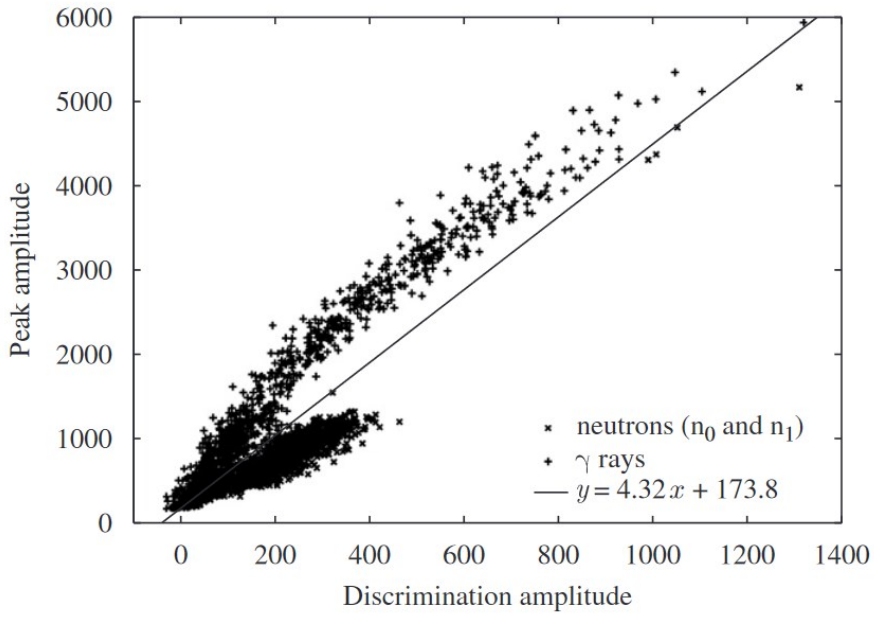


Figure 2.10: Neutron-Gamma Pulse Shape Discrimination using Pulse Gradient Analysis. Peak amplitude vs. Discrimination amplitude (a.u.). From [104].

2.8.4 Validation

In terms of classification, a pulse shape discrimination algorithm is better if it can better distinguish between particle types. This separation becomes increasingly difficult at low energies and is usually a qualitative measure.

In order to measure the actual experimental ability to separate pulses of interest, Roush defined a Figure of Merit (FOM), equation 2.13, where T_x is the zero-crossing time for each type of event of interest, and the slope is calculated with the smallest of these values. This metric, or variations on it[105], is still used to compare the performance of a scintillator's PSD capabilities.

$$FoM \equiv (T_2 - T_1) \cdot Slope_x|_{Tx} \tag{2.13}$$

$$FoM \equiv (T_2 - T_1)/FWHM_2 + FWHM_1 \tag{2.14}$$

A problem with validating PSD experimental results is the common necessity of reference pulses to compare to. All the aforementioned methods require some sort of calibration or training. Gamma-rich environments are typically also neutron-rich, so in the presence of both, a reference pulse has an undetermined type before classification.

Time-of-flight methods exploit the speed difference between massive and non-massive particles to classify them. These can be used to classify particles, using that information to cross-validate the classification of PSD algorithms[105].

Validating a separation algorithm comes down to the definition of a separation parameter, mapping out the different particles and calculating how well they can be distinguished, resulting in the figure of merit. Due to its ease of implementation and good results, the charge integration parameter defined in section 2.8.2 is commonly used, usually compared to the pulse energy or time-of-flight.

2.8.5 PSD with pile-up

PSD is very often used concomitantly with pile-up, though in the sense that pile-up rejection is used to filter out piled-up events before PSD is performed. This is done because PSD performance using strictly traditional methods suffers from high misclassification rates, especially in high count rate environments[105].

Fu et al introduce the use of neural network algorithms to classify pile-up events by training these on synthetic pulses formed by all the permutations in neutron and gamma 2-fold pile-up events[105].

In a similar vein, Luo et al[106] treat piled-up events as a whole, either neutron-neutron, neutron-gamma, gamma-neutron or gamma-gamma. Using calibration pulses as a reference, they create a model for each of these combinations that can generate a pile-up event, then use a goodness-of-fit measurement to identify the most likely type of event.

3 Pulse Simulation

3.1 Pulse generation

The first goal of this project, as stated in section 1.2, *is the creation of a tool for testing, capable of creating simulated synthetic pulses and pulse trains, in accordance with desired event rates, energy spectrum and particle type, according to the detection system characteristics, including detector type, data acquisition sample rate and noise.* This means having the ability to create a pulse train that follows the characteristics of real data such as a data stream from JET or ITER, to be used as a test environment. This chapter describes the approach taken to create this program.

Prototyping and testing pulse processing systems and algorithms usually first involves using synthetic data, generated with proprietary pulse-generating boards, before moving on to tests with real data. Without access to proprietary boards, and lacking the time to learn how to integrate into an environment like MARTe from section 2.6.5, the first choice was to start by developing the capacity to generate pulses according to our specifications.

A possible approach is simulating the particle interactions and transport, as seen in the work of Pozzi et al[107], which uses Monte Carlo N-Particle Transport Code to study PHA in organic scintillators, or Patronis et al using GEANT4 code to simulate and test the response function of a BC501A neutron detector [108]. This dissertation focuses more in the signal chain during and after the signal acquisition, so an accurate simulation of the physics behind the pulse is beyond the scope of this project. In addition, these types of approaches are unfit for real-time processing due to their computational costs.

This work mainly follows up on the Gamma-Ray Diagnostics Systems for the JET experiment[43], as well as the work on Data Processing Algorithms and Electronic Tests for the Radial Neutron Chamber [109][15]. The baseline for the project was adapting an existing Python code for processing FPGA signals. The goal of that project is real-time, low-latency counting of gamma and neutron pulses. These use sample rates ranging from 500MHz up to 1GHz, [43] and a digitizer

frequency of 1.6GHz [15]. The inverse of these frequencies establishes our range of time resolutions. As for energy resolution, the early digitization of detector data means the energy values are relative to the ADC bins, not related to absolute energy values before some calibration is done. Calibration doesn't affect pulse shapes, and values don't need to represent anything for PSD or pile-up separation algorithm testing, therefore when applicable throughout chapters 3 and 4, the original digitized values were interpreted as arbitrary energy units.

Taking into account the variety of detector types and acquisition systems, and their respective response functions (section 2.5.4), a black-box approach was chosen to reduce the overhead of accounting for hardware variations. This can be done because an event can be accurately replicated using a pulse model, representing the acquisition system's transfer function, which was established throughout sections 2.8 and 2.7. The choice of a model is discussed in chapter 3.2. These single pulse events are then converted into a series of pulses matching the chosen time distribution. With these variables, the result is a pulse train with characteristics that can be matched to a real pulse train in terms of the energy spectrum, pulse rate and noise. In order to generate the energy and time distributions, stochastic simulation was used, following the approach of [106], where Monte-Carlo simulation is used to generate individual pulses based on a semi-empirical formula.

3.2 Single pulse descriptors

Techniques that are based on a detector's response are widespread in usage. These rely on the shape of the pulses a detector outputs being consistent for the same particle type and energy. In these cases, whether through physical reasons (section 2.5.2) or mathematical approximations (section 2.8), there is a semi-empirical formula for the shape of a pulse. This is especially applicable for organic scintillators and the double exponential decay (eq.2.5.2).

Koechlin et al[110] compare a few empirical formulas for the shape of a pulse as far back as 1964. Among them, they suggest the use of an exponential rise time, with a secondary decay, introducing the double exponential (eq. 2.5.2).

Marrone et al[90] nearly four decades later, despite much scientific maturity, elaborates on the pulse shape analysis of neutron and gamma single pulses: "An overall reasonable description of the pulse shape can be obtained by using these values for the two exponentials, convoluted with the response function of the system". The response function of the system relates to the acquisition instrumentation used (section 2.6.2). They conclude that the leading edge exponential parts of each pulse are roughly constant between particle types, while the tail decay constant characterizes the particle, as suggested in section 2.5.2.

Different methods have also been used to describe a pulse, see section 2.8, such as spline polynomials [85], Gaussian approximation[36] and in the context of PSD, convolutional neural networks [87]. For our choice, there is a strong argument for creation of a data stream for processing in an FPGA. This disqualifies methods focused on simulating particle transport[108][107].

The physics is briefly considered at the start in order to understand the phenomenon and where each pulse's shape comes from. But up to the creation of the pulse, there's a certain degree of black-box approach after acquisition - the main interest here is in the fast digital processing of a pulse train, for real-time use. The template chosen at the start can be easily changed, which argues in favor of choosing a simple descriptor. Diamond CVD is the main candidate in ITER's RNC for its neutron detection capacity, while organic scintillators' PSD discrimination power is unparalleled.

Furthermore, given the volume of investigation into organic scintillators, the double exponential was chosen as the base model to recreate a pulse shape.

A pulse described with this model can be characterized by 3 variables: The amplitude of the pulse, I_0, characterizes the energy of the incoming particle. I_0 linearly scales with the height of the pulse if the other variables are constant. In the double exponential approximation, the height of the pulse scales proportionally to its initial amplitude I_0, and it can be calculated directly as long as the detector is calibrated.

The actual energy of a pulse is encoded in its charge-integration (section 2.6), but since height, initial amplitude and energy are all proportional, these are mutually interchangeable apart from a multiplication factor, but different for neutrons and gammas.

The exponential decay constants, τ_{rise} and τ_{decay}, are essentially properties of each detector, influenced by the shaping procedures. τ_{decay} encodes the tail of the pulse, and how fast it returns to baseline. Unlike the rise time, since each type of particle introduces a characteristic type of decay, τ_{decay} is a function of the type of event-generating particle, such as T_G for gammas and T_N for neutrons. The difference in pulse shape can be seen in figures 3.1 and 3.2. These values are affected by the pulse shaping happening throughout the acquisition chain, discussed in section 2.6.2, which is why a model is only valid for a specific system.

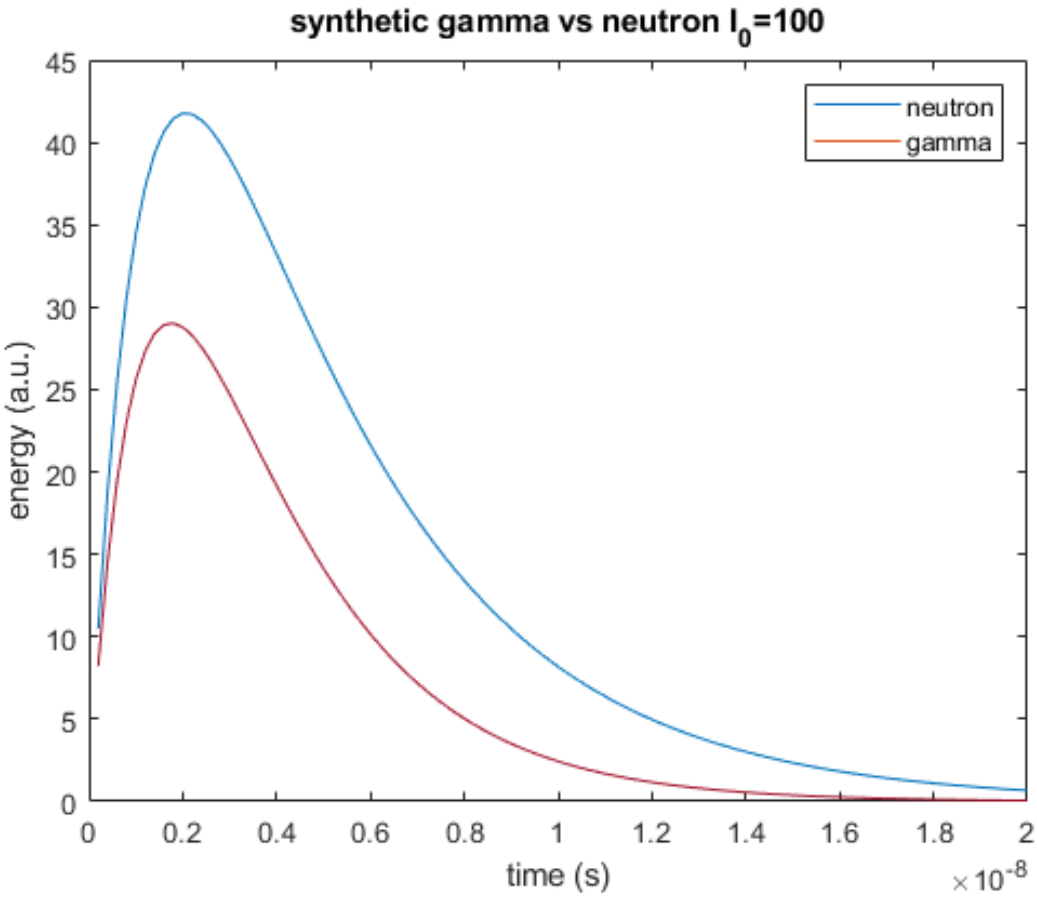


Figure 3.1: A single synthetic gamma pulse and a single neutron pulse. Acquisition frequency = 5GHz means each time step is $\Delta t = \frac{1}{5*10^9} = 0.2$ns; Note that despite having the same I_0, the pulses have very different energies. $\tau_N = 2.7 * 10^{-9}s; \tau_G = 4 * 10^{-9}s; \tau_{rise} = 1.2 * 10^{-9}s$

Figure 3.1 shows two pulses created by sampling a double exponential function at each time step, given by the inverse of the sampling frequency. Each pulse shares most inputs, except particle type, encoded in the different exponential rise times. This produces pulses with

different heights and energy(area). In order to test discrimination algorithms, the ability to match heights and energies is needed, so that the particle type can be isolated as a variable. The relationship that allows this can be derived when looking at equation 2.5.2, where the only maximum is located when $\delta/\delta f = 0$. Solving for this maximum produces eq. 3.1, which allows us to transpose a maximum to I_0 value, as demonstrated in image 3.2.

$$I_{t_{MAX}} = I_0 * ((exp(k * ln(\tau_D/\tau_R)/\tau_D)) - exp(k * ln(\tau_D/\tau_R)/\tau_R)) \\ k = (\tau_D * \tau_R)/(\tau_R - \tau_D) \tag{3.1}$$

Additionally, the neutron pulse in figure 3.1 has some leftover energy by the time the gamma has returned to baseline. This method produces pulses with an infinite impulse response, so to fix this a cutoff point is set based on a threshold value. The fixed pulse width used in FPGA processing is not relevant at this point, only after detection begins, discussed in chapter 4.

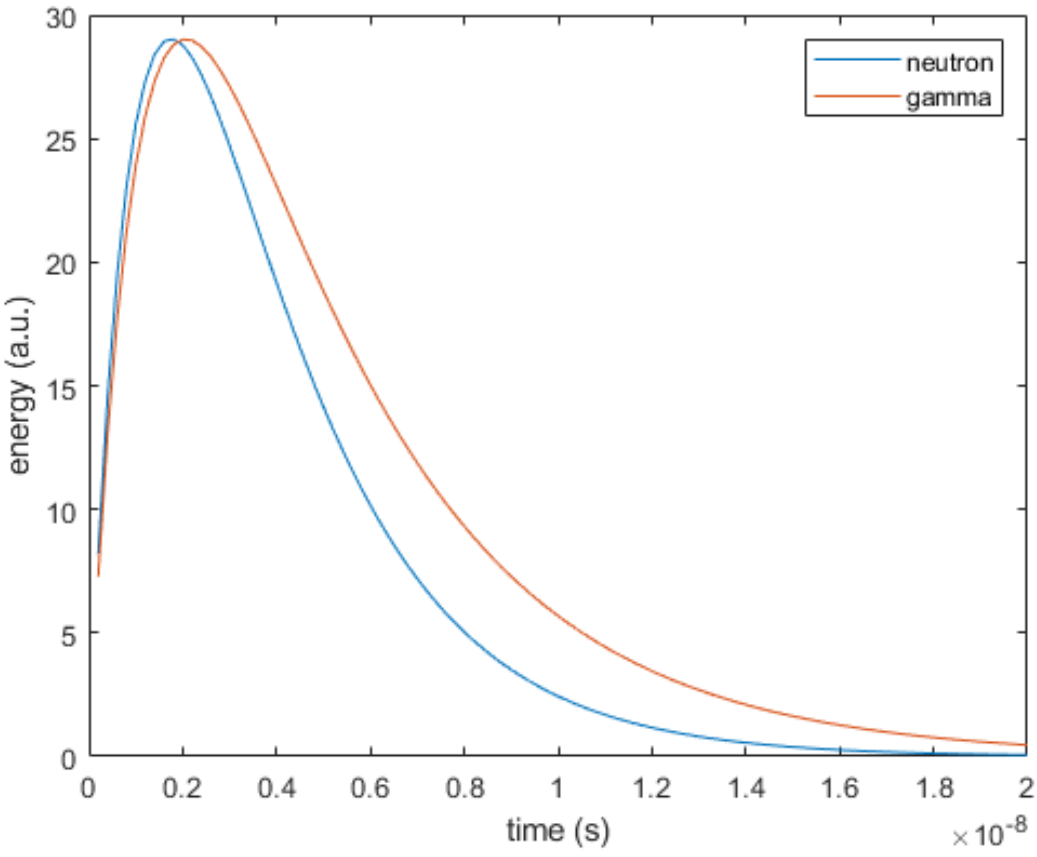


Figure 3.2: A single synthetic gamma pulse and a single neutron pulse of matching heights. Identical conditions as figure 3.1, except I_0 is 100 for the neutron, and 69.5 for the gamma, calculated using equation 3.1

A similar relationship can be derived for the energies, by integrating 3.1 with respect to

time, arriving at eq. 3.2.

$$I = I_0 * (\tau_R * (exp(-x/\tau_R) - \tau * (exp(-x/\tau) \tag{3.2}$$

3.3 Pulse train generation

The previous section explained the decision process behind choosing a single pulse descriptor. This section follows up on how to extend the generation of a single pulse and into a sequence of pulses with timing characteristics that match data acquired in a fusion experiment.

To create a single pulse using the method described in the previous chapter, the input variables are the initial amplitude I_0 and the decay constants for each particle type, as well as the sampling frequency. To take a single neutron pulse and create a congruent sequence of neutron pulses, energy values are sampled from an energy spectrum and plugged into the single neutron pulse generator. This allows for the recreation of pulses matching an arbitrary spectrum, from a simple sequence of mono-energetic neutron pulses to one matching a real material. The utility of this lies in being able to change some conditions of an acquisition experiment while maintaining others. In our case, this can provide some insight into the behaviour of a detection algorithm, for example at differing levels of pile-up, see ch.3.3.2. In order to extend this to other particle types, a probability can be assigned to each type, generating pulses from the same energy spectrum randomly, or independent spectra can be generated for each type, and later merged. Batch-generating single pulses can become computationally taxing as it linearly scales with the number of pulses. This is mostly irrelevant since the whole process only takes a few seconds for tens of thousands of pulses in Matlab and it doesn't concern the speed of the detection algorithms.

In summary, with a method to generate single pulses, a target energy and timing distribution, and a way to sample from those distributions (described in section 3.3.1), a method to simulate a pulse train is achieved.

3.3.1 Stochastic generation

There are two distributions that have to be followed in order to create a pulse train. The first is the spectrum of energies, which comes in the form of a histogram (event counts in each bin), which can be sourced from real events or theoretical. Sampling energies like this creates a batch of independent pulses.

The second distribution is the timing distribution, which should provide the time-points for each pulse - at what point in time does each pulse occur and how to place them.

The method chosen for this step is a Monte-Carlo method called inverse transform sampling [111]. This method involves repeatedly taking a random number from [0,1] and finding the highest value in the Cumulative Density Function(CDF) that would generate it. This requires knowing the CDF of a distribution and creating its inverse. This is equivalent to calculating the quantile of the CDF for that probability value. If you have a uniform number generator for [0,1], this method allows you to generate an arbitrary number of samples from a known distribution. This process is shown in figure 3.3.

The probability distribution function (PDF) is a function of x defined by: for a value sampled randomly from its domain, the probability that it falls at each x point is the value of the function at that point. The CDF is made by continuously summing or integrating the PDF through its domain. For discrete functions, such as a spectrum, CDFs are trivial to calculate. On the other hand, not all continuous functions have an analytical CDF.

This method thus requires the probability distribution function (PDF). As for the energy spectrum, an arbitrary histogram can be input, including a real spectrum. The counts in each bin divided by the total number of events is the probability that an event will fit in that bin, which can be directly interpreted as a PDF when normalized.

As for the time distribution, the approach taken started by noting that detection events are a Poisson process. This means events are discrete and independent and the probability of X events happening in an arbitrary time period is known. A Poisson distribution is modelled after the rate parameter $\lambda = \#events/totaltime$, which can be calculated as the mean time between events.

At any time in a Poisson process, the time until the next event happens is described by a decaying exponential in the form of $P(T > t) = e^{-\lambda * t}$ [112]. In this way, at any point in time, there is a stochastic model for the time it takes until the next event happens, Δt. Rather than populating the time-points in the data stream, pulses can be generated based on the expected time lapsed since the previous pulse.

To generate the time between pulse, note that $P(T > t) = e^{-\lambda * t}$ (eq.3.3.1) is the PDF of the exponential distribution, so the CDF of an exponential distribution of parameter λ can be used to sample values for Δt between pulses, where λ is the average time between events. A similar method is used by Gardner et al [81] in the context of X-ray spectroscopy.

At this point, the following is possible: generating single pulses of gammas or neutrons based on an energy value and pulse shape; generating an array of timing information or time-between-pulses, Δt, that matches a Poisson Process happening at a known rate; generating a sampled table of energies that follows a known spectrum.

By assigning times and energies one-to-one, a pulse is generated from the energy and summed onto a zeroed array after Δt has passed. Repeating this for the number of pulses required generates a full pulse train.

3.3.2 Natural pile-up

Pile-up in pulses occurs when two pulses are so close together that their signal intersects and can't be resolved independently. Chapter 2.7 showed how to estimate the amount of pile-up based on pulse length. In the context of a simulated pulse, this is equivalent to having sampled a Δt below the minimum time difference where consecutive pulses are detected as separate. Since the timing generation method creates an array of the time intervals between pulses, the data needed to estimate the amount of pile-up is readily available. The expected value of Δt is the lambda factor used to generate the exponential CDF. By definition, 50% of the values are below the mean value. The lower this mean-time, the more events will pile up. This makes it possible to investigate how much pile-up happens at each interval, and eventually how much can be resolved at each interval with a given detection algorithm.

A graphical representation of this pile-up estimation is shown in image 3.3. In this approach, first a certain pulse-width is set. This value is a power of 2 chosen to match the pulse width used by the acquisition instrumentation. From this value, an upper estimate for pile-up is derived using the method described in eq. 2.11. This is done by first generating all the Δt values, and ordering them. If a pulse has $\Delta t \geq pulsewidth$, it can only ever incur pile-up if a pulse that follows has a smaller Δt. After ordering the time intervals, counting how many are below the pulse-width value gives us an upper bound for the number of piled-up events. It doesn't mean all the events with Δt smaller than the pulse width are piled up, but if Δt is greater than the pulse width, there can be no pile-up with the previous pulse. There could be pile-up with the next pulse, but that would be counted with its respective Δt. This approach doesn't discount multiple pile-ups, but candidates for two-fold pile-up could be detected similarly by counting a sequence of two or more Δt below the critical value.

This numerical approach is necessary because this type of pile-up can be estimated, but can't be forced. This sampling is a simplified attempt at recreating the randomness of detection numerically, using the stochastic nature of the sampling process. The biggest advantage lies in the information produced: once the pulse train is formed, the energy, type, and time-point of every pulse that has been generated are known, invaluable information for checking the performance of detection algorithms.

Gaussian white noise to match the type of noise in [65] was added, as shown in figure 3.4.

3.3.3 Limitations

An issue occurred with the values generated from the CDF method, which were repeating far too often, both for energies and times. This would happen because the algorithm for sampling from an empirical CDF would search for the first value that's above it, and take that inverse. This had two consequences: First, this means that any and all outcomes would be sampled out of a limited number of possible values - those that are used to create the CDF. If there are 3000 pulses used to create a histogram to sample, the energy of any pulse generated would be one of those 3000 values. Despite matching the originals perfectly, nothing new would be created, essentially shuffling around the known events at different time points. This could create unintentional patterns, especially after 3000 generated pulses, and binds the generated energies to a finite spectrum.

The second problem lies with random sampling the first element of the CDF. Any value sampled below that value would redirect to exactly the same bin. In the context of generating times, this means the smallest possible value of Δt that could be generated would be the first element in the X domain of the CDF. Looking at the process in image 3.3, the smallest value for Δt would be $X(1)$, or $\Delta t/10 = 2.5 * 10^{-9} s$. In most modern pile-up resolution algorithms, the possible improvements lie in either solving pile-up for lower energies or for smaller Δts, so having a lower bound for the interval between pulses was a critical flaw.

In order to resolve this, interpolation was introduced to the Inverse Transform Sampling algorithm. Instead of finding the first value in the CDF that is higher than the generated value between 0 and 1, the remainder after finding a bin is used to interpolate the generated energy/time value proportionally. This method solves the discretization problem from the limited number of bins. However, the values between bins are linearly interpolated, making energy resolution depend on spectrum resolution. This could be improved upon with a more complex interpolation method, more bins (energy resolution) or more counts. Another issue arises in the case where there is a high discrepancy in adjacent energy values in an energy spectrum, visualized on image 3.5. In real cases, with spectra closer to being continuous, and a high number of reference values, this is unlikely to be an issue. In simulating discrete spectra, rather than sampling from a CDF, a list of energies is used to prevent this.

This chapter described a method to create a pulse train according to given specifications. The validation of this method is explained in the following chapter (4), because for validation purposes an event detection algorithm is required, therein described.

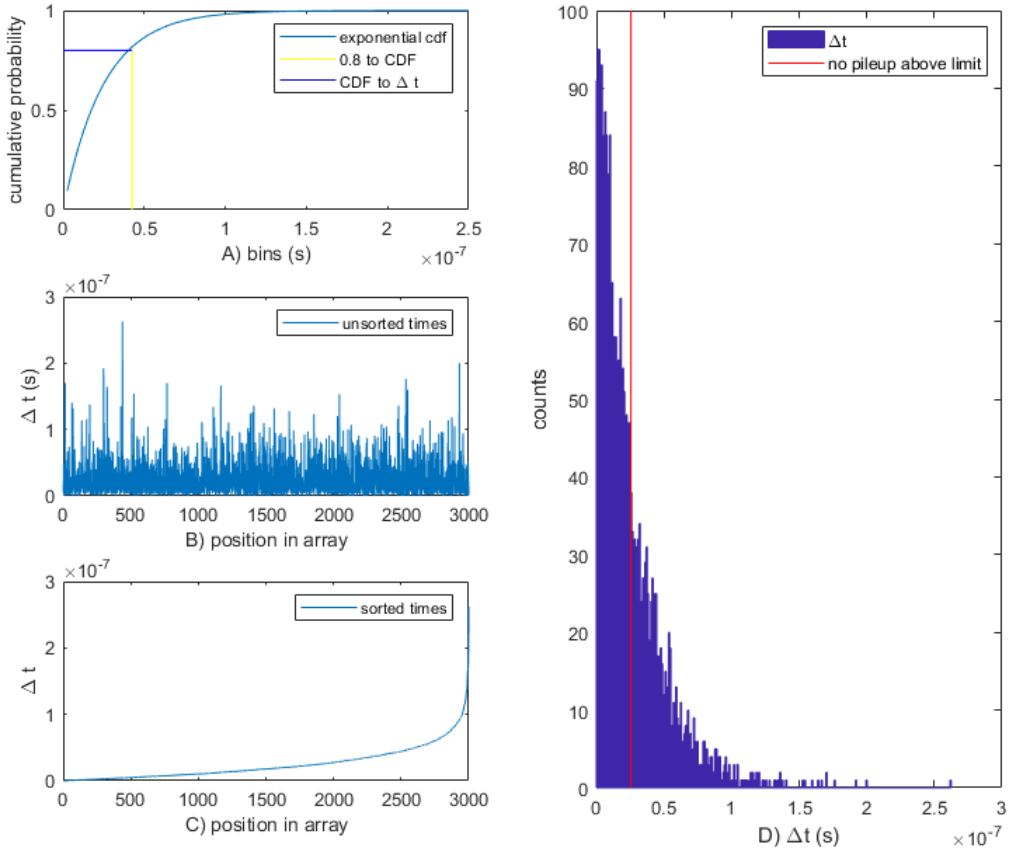


Figure 3.3: Inverse Transform Sampling to generate 3000 Δt between pulses. δt = 2.5e-8(s). Omitted is the first step, choosing the sampling domain X. Here, X is an array of size 100, with step=starting point=$\delta t/10$.

A) Exponential CDF represented in its domain. A random value between 0 and 1 is generated. Illustrated is P=0.8 (blue). At the intersection with the CDF, the first value greater than 0.8<0.817 is the 17th element of the CDF. The 17th element of X is 4.25e-8 (yellow), the sampled value. Statistically, this Δt is a value greater than 80% of all values sampled this way.

B) Repeat A) 3000 times for random values between 0 and 1. The average Δt is 2.452e-8, expected value would be δt = 2.5e-8(s).

C) Sort the Δts. Note that the linear sampling of an exponential CDF produces an exponential graph.

D) Histogram of the Δts. Considering a maximum pulse width of 128*δt, pulses that far separated in time won't pile up. This limit is represented in red. Only pulses to the left of this line are liable to pile up, though not necessarily.

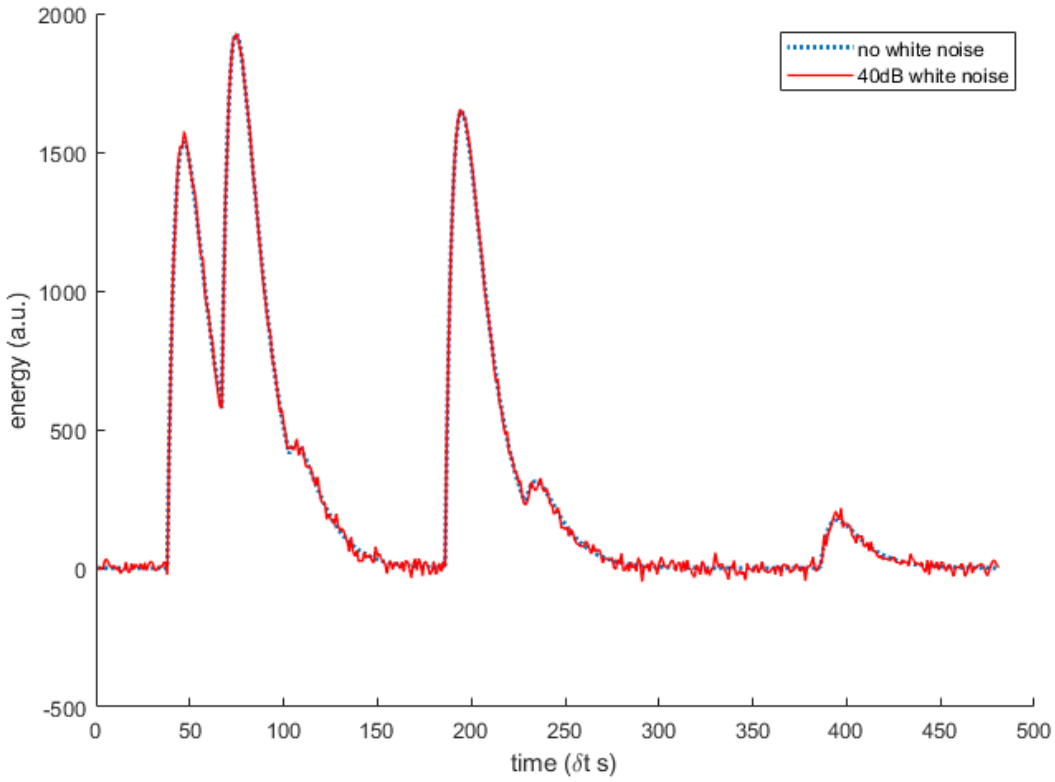


Figure 3.4: Gaussian White Noise can be added to introduce the desired level of Signal-to-Noise ratio (SNR). Pictured is a sample taken from a simulated neutron pulse train containing 3 and 2-fold pile-up and a single pulse, with 40dB SNR.

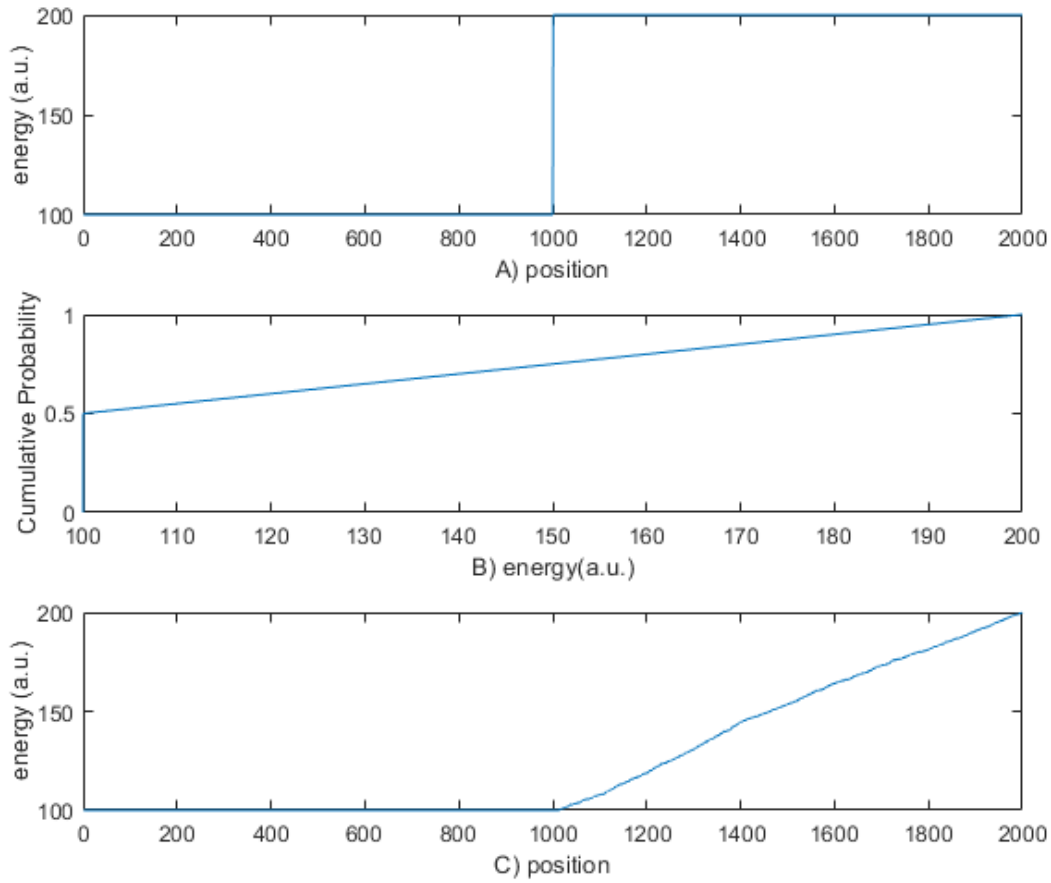


Figure 3.5: Generation of artificial values in Inverse Transform Sampling.

A) Energy array created with 50% 100 (a.u.) and 50% 200 (a.u.), for a total of 2000 values. The expected result would be simulating a roughly similar number of 100 and 200 as an output, around 1000 of each.

B) Empirical CDF formed from the distribution in A). Sampling the values below 0.5 will always result in a 100 being sampled, so 50% should be correct. Though the CDF only contains a point at (200,1) and the slope seen there is a visual artifact, it correctly represents the interpolation that will occur when the random value hits any point between]0.5,1].

C) Energy values generated, sorted. Demonstrated here is the worst-case spectral distortion of the interpolation, for histograms with only 2-bins.

4 Neutron-Gamma discrimination algorithm

4.1 Pulse Separation and classification

This section details the development of an event detection algorithm capable of resolving pile-up and discriminating neutron-gamma pulses. The goals are: validating the simulation algorithm described in chapter 3; validating the algorithm's viability; if valid, testing its limitations.

This project was born as part of JET Tokamak and ITER-RNC research, with the calculation of the emissivity profile for plasma control as the end goal, shown earlier in chapter 1, image 1.4, which is subject to a control cycle of 10ms. As a part of this calculation, a spectrum of neutron counts is required. Continuously acquiring data from detectors would drastically increase data throughput requirements, resulting in data loss. Because of this, the acquisition, which is done in an FPGA, uses a leading-edge trigger to detect and acquire only data above a threshold, data that contain pulses, so-called events. Events are acquired in a window of a fixed pulse width parameter. If at the end of this window, the detected value has not dropped below the threshold, this means there is pile-up in the event, and the acquisition is extended by the pulse width.

The event data can be processed in real-time in the FPGA, or sent to the host computer for processing. The processing tasks are pile-up detection, particle identification to separate neutrons from gammas, pulse height analysis to measure the energy of each pulse, and sorting into bins to create a histogram or neutron spectrum. There ends the scope of this thesis. In order to test the pulse train simulation algorithm created for this project, an event detection algorithm was necessary, which is discussed in section 4.1.1.

For event processing, two existing algorithms are considered. The first algorithm was first developed for the WPJET4 Gamma Spectrometer and adapted for the RNC, and is shown in image 4.1 [103][65]. It implements a trapezoidal filter to flatten the typical exponential pulse

shape into a trapezoid with height proportional to the energy. This acts as a moving average filter, gives some amplification for short pulses and eliminates the need for preamplifier DC offset correction; more critically, the relationship between height and area can be used as a separation parameter to perform PSD. The second algorithm uses a more traditional charge integration method to perform PSD.

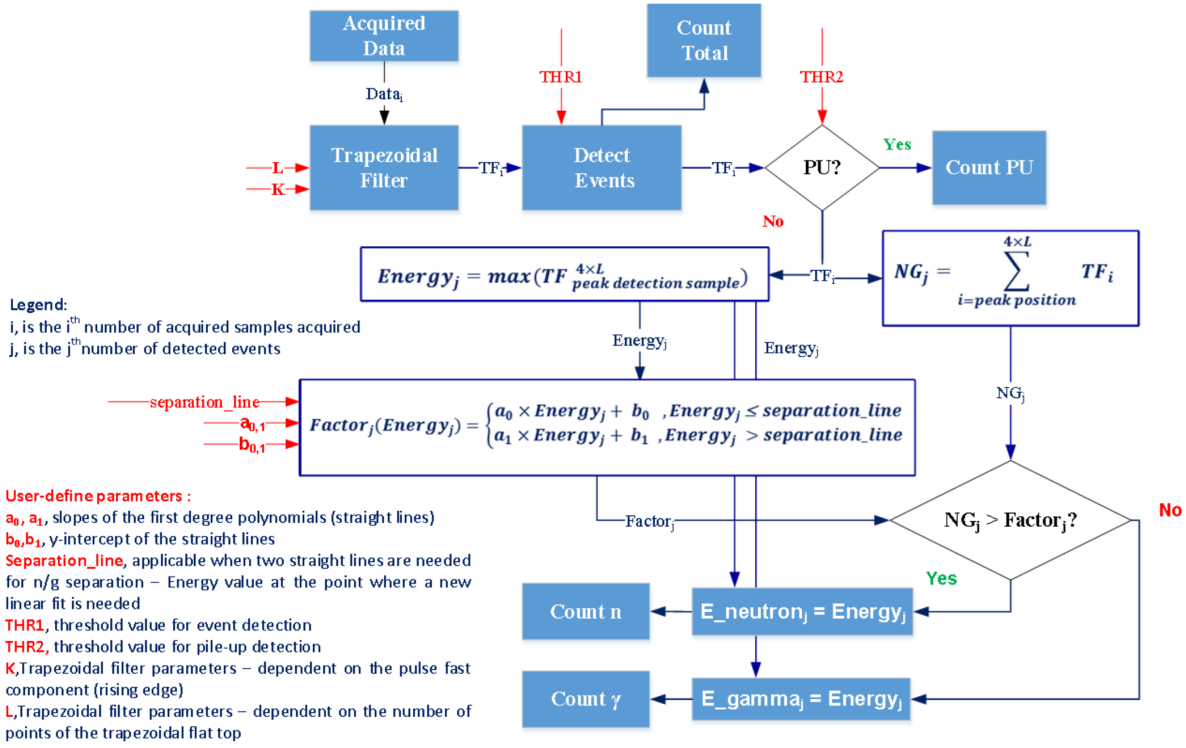


Figure 4.1: Trapezoidal shaper technique for neutron-gamma discrimination. From [65]

The processing algorithm used in this project melds these pile-up separation approaches with the one developed by Luo, adding the capability to perform Pulse Shape Discrimination at the same time as the pile-up separation, applying a version of the Template Matching and Single Event Reconstruction deconvolution algorithms described in chapter 2.7.2[106]. This is explained in section 4.1.3.

The aforementioned algorithm was adapted in a few ways to fit the current project. In order to create a reference signal, low-count rate gamma and neutron acquisitions would be required. Because these weren't readily available, and taking into account the goal of validating the generating algorithm of chapter 3, the double exponential equation was used. This allows for the use of detector datasheets to quickly estimate the decay constants and create reference pulses. Though this might not as accurately describe the response function of the whole acquisition system, that could easily be mitigated when dealing with real data by either using a reference

pulse, or by fitting the decay constants using calibration data.

It should be noted that acquisition systems traditionally perform pile-up rejection and PSD at different points of the acquisition process, while this approach does both simultaneously.

4.1.1 Event detection

The event detection algorithm works by acquiring windows of a fixed size when pulses are above a threshold, which can be expanded up to a certain amount of times if the pulse has not finished by the end of the window. For the sake of compatibility with the data, the same approach was taken, with base acquisition windows ranging 32-128 samples wide, depending on the experiment, which sets the pulse width. At an acquisition rate of 1 Gsamples/s, a 32 sample window lasts for 32ns, a typical value for the duration of pulses.

The baseline correction is based on the pre-trigger, which is the set of n values before hitting the noise threshold which signals the event, usually 4 or 8 values. It works by checking the average energy preceding the event and normalizing it to 0, as to minimize its influence on possible pulse heights.

To estimate multiplicity in each event, the downward crossing method was used. After restoring the baseline, the first order derivative is calculated as the numeric differential, which is 3-point moving-mean filtered. Counting the number of downward zero-crossings in the filtered differential was used as an estimate for the multiplicity. While this method was good enough for as long as noisy, synthetic pulse trains were being processed, when cross-testing with real data, it proved highly inconsistent due to noise and generally non-ideal pulse shape, so the Matlab more powerful peak finding algorithm was temporarily used. This alternative worked great, but at a huge compromise in speed and lack of portability. In the end, returning to the crossover method by removing false positives as suggested by Luo et al proved to be the much faster method, with identical accuracy[106]. The false positives were occurring in points below the threshold in the original array; in points where the derivative change was negligible; in points where the peak is made up of at least two consecutive identical values. Masking these out then allowed for counting the number of negative crossovers, producing the estimated multiplicity. Pulses are then processed according to the multiplicity, as single events, 2-fold pile-up or 3-fold pile-up.

4.1.2 Individual pulses

The principle behind the discrimination algorithm is recreating the pulse according to a known model and seeing what fits best among the possibilities for that event. Generating a single pulse is identical to the process in section 3.2, sampling the double exponential at a known sample rate, but here the input variables need to recreate the target pulse as closely as possible, and the challenge lies in how to create those inputs. The desired height can be achieved by identifying the pulse height and exploiting the relationship between height and energy via equation 3.1. The pulse shape, encoded in the decay times, is a global parameter for each type of pulse, gamma or neutron. This can be sourced from literature[36], which was done for the synthetic testing shown in this chapter, and roughly matches real pulse lengths when acquisition and the sampling frequencies are matched. However, section 2.6.2 discussed how pulse shapes are tied into the whole acquisition system, so these can't be expected to apply to real data acquired with a different system. Sourcing decay times for different particle types acquired with the same setup is not realistic. In these cases, a reference pulse can be built using the average of well-behaved pulses, as mentioned in section 2.8, and applied further ahead in section 4.3.1.

To identify the position of peaks, note that the multiplicity estimation algorithm from section 4.1.1 is already capable of identifying their positions, with limitations discussed in the next section.

The generated pulses are set to end at a predetermined length in the separation algorithm, or after crossing a termination threshold, which was defined based on when the pulse is no longer distinguishable from baseline noise. Discrimination between individual neutron or gamma is done by minimum point-squares difference, comparing it to the original pulse - the recreated pulse with the lowest residual value is considered correct. This can also be done by charge integration, but since that can't be applied to piled-up pulses and single pulses aren't the focus of interest, the choice was based on consistency, with the CI method being used to cross-check the results.

4.1.3 2-fold pile-up

The work of Luo et al[106] is a key reference here. Luo's method starts by creating a reference signal created from averaging a large number of pile-up free waveforms, a standard neutron pulse, and a standard gamma pulse. Pile-up detection is based on the first-order derivative, where the multiplicity is estimated by the zero-crossing method, counting the number of changes from positive to negative in the differential above a certain slope, where the original value is also above a threshold. In an event where 2-fold pile-up is detected, the constituent pulses are recreated

by comparing the pulse to four models, the four possible combinations of the standard neutron and gamma pulses, separated by an offset, in varying amplitudes. Using the normalized first pulse as a basis, a second pulse is introduced at a fixed relative amplitude, changed iteratively until it converges. The recreation that most resembles the original pulse determines the type of pile-up, calculated by the minimum point difference.

The main changes from this algorithm are in the recreation stage. The first peak in a pile-up event preserves most of its rising shape. Taking this into account, the maximum value can be estimated and the first pulse in a pile-up event is reproduced as a neutron and as a gamma. At this point, the type of particle is unknown. This recreated first pulse is subtracted from the original, first as a neutron and then as a gamma, which clears up the 2nd pulse in the pile-up for individual analysis. This is shown in figure 4.2 for the first pulse as a neutron. This takes advantage of the double exponential model by using the height inversion mentioned in the previous section, foregoing the need for normalization and iteratively calculating optimal amplitudes.

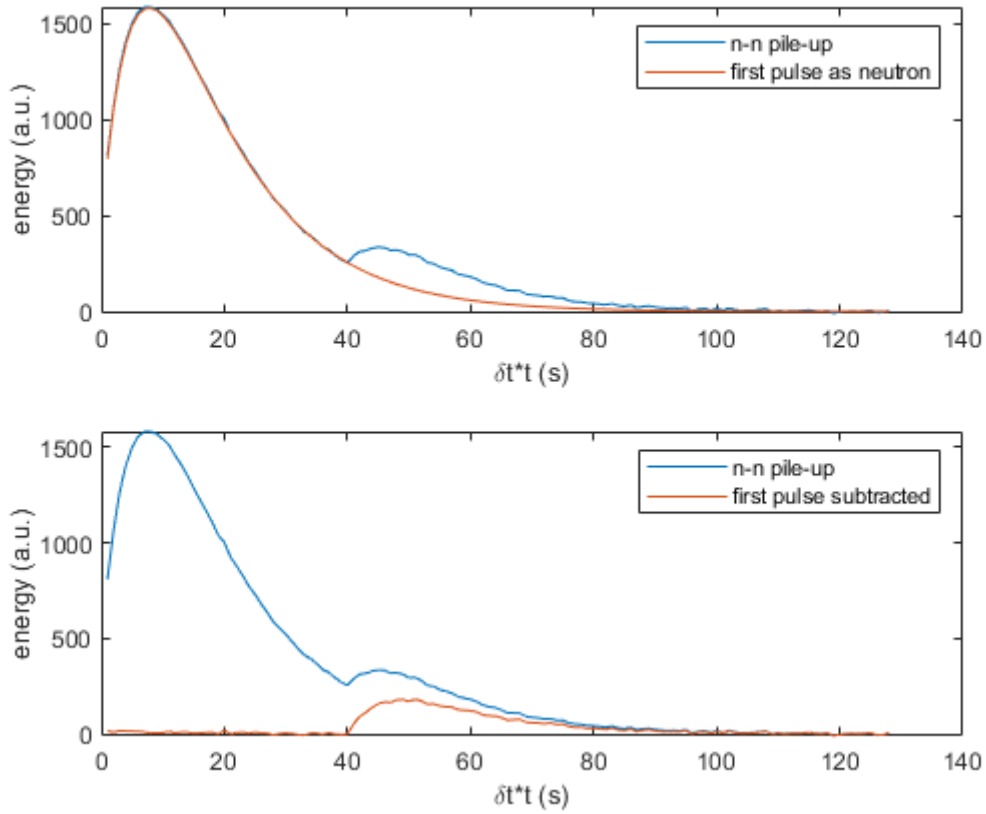


Figure 4.2: A) A 2-fold pile-up is detected; the height of the first pulse is measured and a neutron that would match that height is recreated at that position;
B) The synthesized neutron was subtracted from the original pulse, revealing the 2nd pulse

In this event to which the first pulse was subtracted, as neutron and then as gamma, the height of the second pulse is measured, using the same equation to recreate the second pulse as both neutron and gamma, again adjusting for offset. Thus, a fully synthetic recreation of the 2-fold pile-up is achieved. All 4 combinations to the original point by point are compared, summing the square differences, and the lowest residual is chosen as correct. The residual also serves as an indicator of the quality of the fit, with the possibility of using it to exclude poor fits. The combinations are shown in figure 4.3.

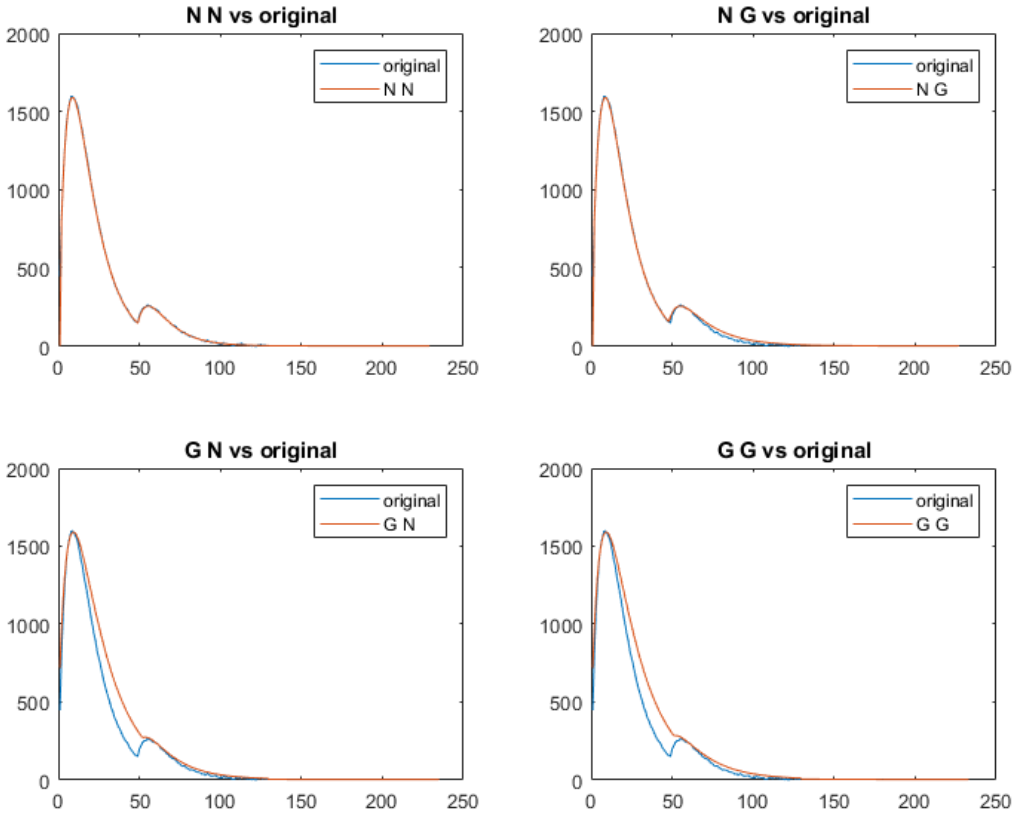


Figure 4.3: 4 possibilities for the 2-fold pile-up algorithm. The original pulse is recreated with the first and second pulse as either neutron or gamma pulse, and the best fit is chosen.

4.1.4 3-fold pile-up

Extending the 2-fold method upwards is possible any number of times, with the computational load doubling each time the multiplicity increases by one. In a 3-fold pile-up, this means 8 cases that have to be considered - the previous four each with a possible leading neutron or gamma. The 3-fold discrimination algorithm is the same as the 2-fold: consider the first pulse, recreate it based on height using the semi-empirical formula, and subtract it from the original. Repeat for the 2nd pile-up, for a total of 8 combinations, as shown in figure 4.4. The one with the smallest residual is chosen.

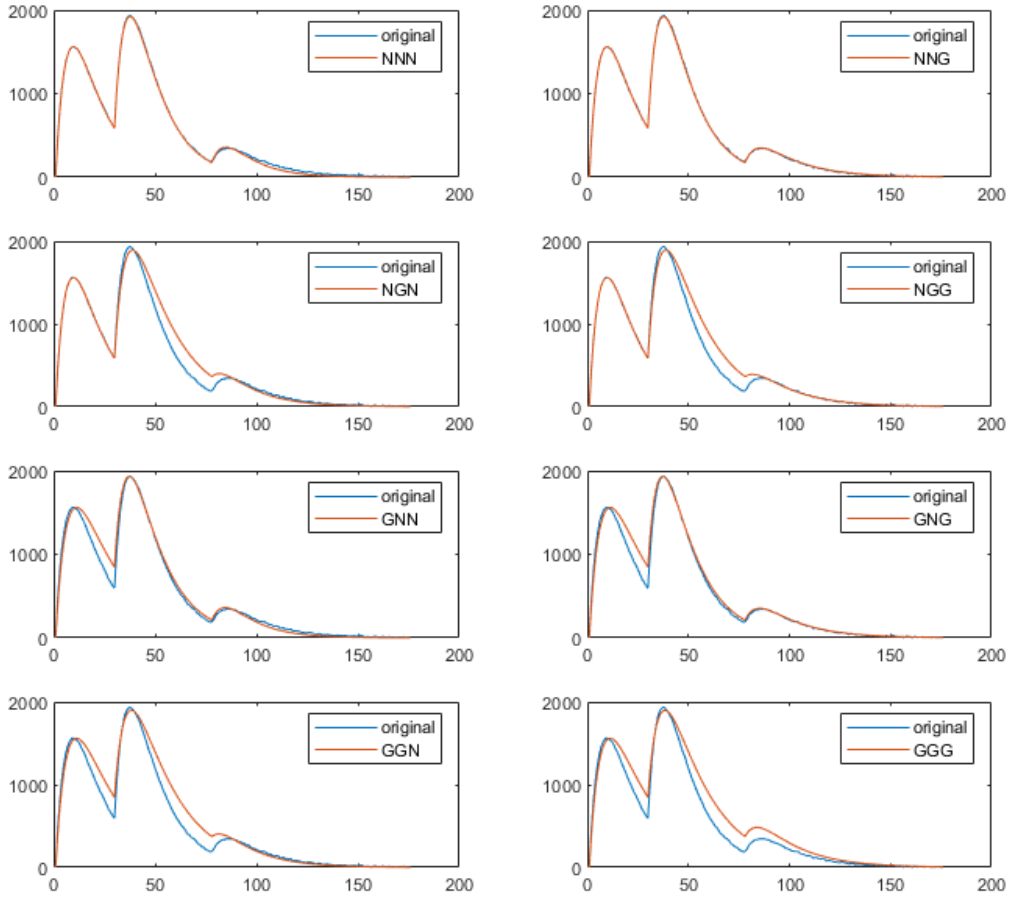


Figure 4.4: 8 possibilities for the 3-fold gamma/neutron pile-up algorithm. Energy(a.u.) vs. time (1 sample = δ t).
3-fold pulse identified using the pulse detection algorithm from a synthetic pulse array, based on a JET test burst.

4.2 Algorithm Validation

4.2.1 Single Energy Neutrons

For the first test case, a simple pulse array was created to validate the simulator and event detection, in close-to-ideal conditions. This was set up with a very low average time, $\lambda = 2.5 * 10^{-7}$ seconds. Acquisition frequency was set at 500MHz and $\Delta t = 2ns$ or $2 * 10^{-9}$s. These values are chosen to match JET EP2 acquisition frequency[43] and create an environment with a minimal amount of pileup, as to specifically test the detection algorithm. Principal decay times were based off a NE213 scintillator, with $T_R = 1.2 * 10^{-9}$s, $T_N = 2.7 * 10^{-9}$s[36]. All 10,000 pulses are neutrons with a height of 500a.u., with a time between pulses stochastically generated by Monte-Carlo inverse sampling an exponential distribution with lambda value equal to the average time between pulses.

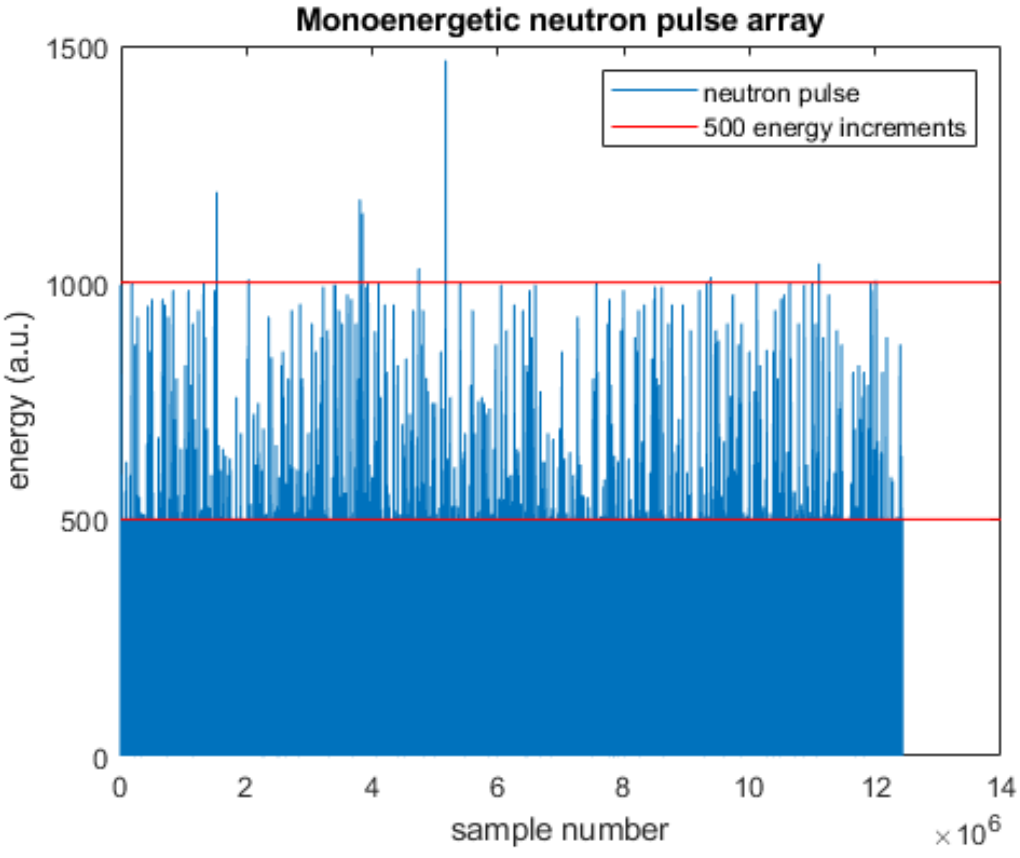


Figure 4.5: Full waveform of 10000 simulated neutrons

The resulting overall waveform is shown in figure 4.5, zoomed out for effect. Every pulse was generated with the same height of 500, meaning every pulse in between the 500 and 1000 threshold has to be a result of a naturally occurring pile-up. In figure 4.6, a zoomed-in segment

of the waveform is shown. A single event is shown next to two 2-fold pile-up events. Of these pile-ups, the one to the right is visually identical to a single event that peaks close to 1000, meaning the single events that generated it were very close in time. The smaller the time between pulses, the less likely to be detected as pile-up, eventually being indistinguishable from a single event. This would mark the limit of the event detection algorithm.

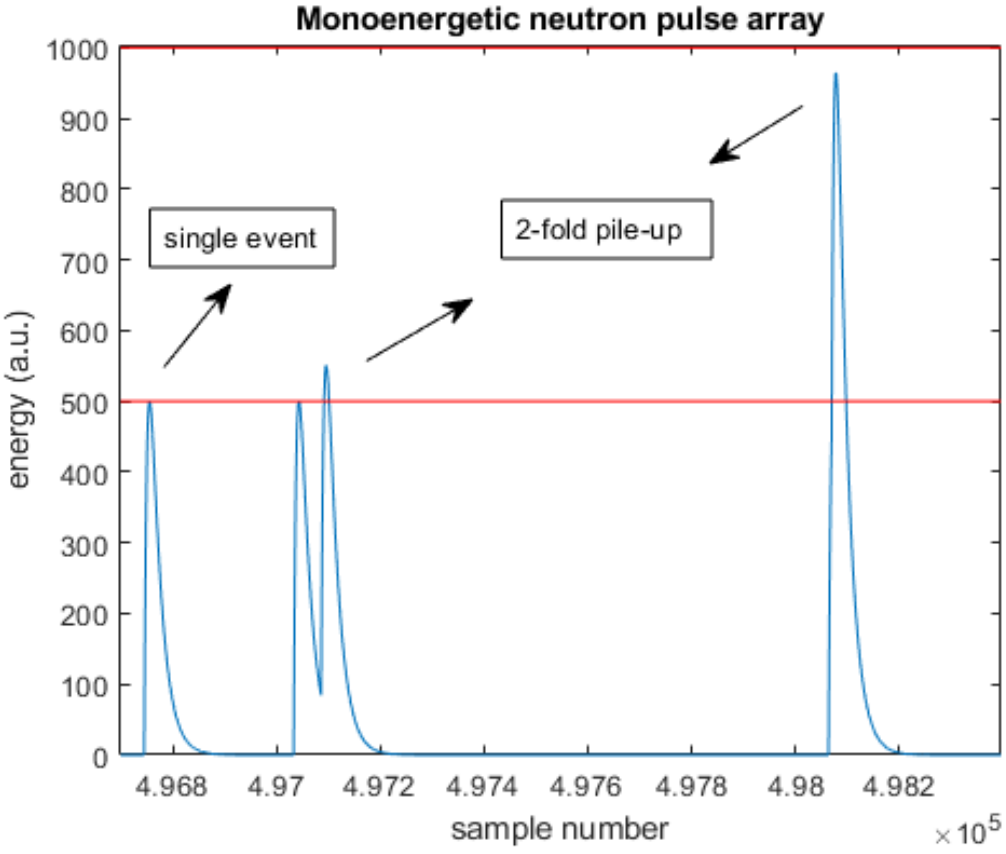


Figure 4.6: Zoomed waveform containing several types of pile-up

In the full waveform, figure 4.5, a few pulses crossing the 1000 threshold can also be seen, marking a triple pile-up event. Out of the 10000 pulses, 324 were detected as being part of a double pile-up and 18 of a triple, with a total of 9977 processed pulses. 23 pulses were then either not detected, or part of a pile-up beyond the detection limit, which led to it being interpreted as a single pulse. All detected pulses were correctly interpreted as neutrons, though on previous similar test runs occasionally a single phantom gamma would be detected.

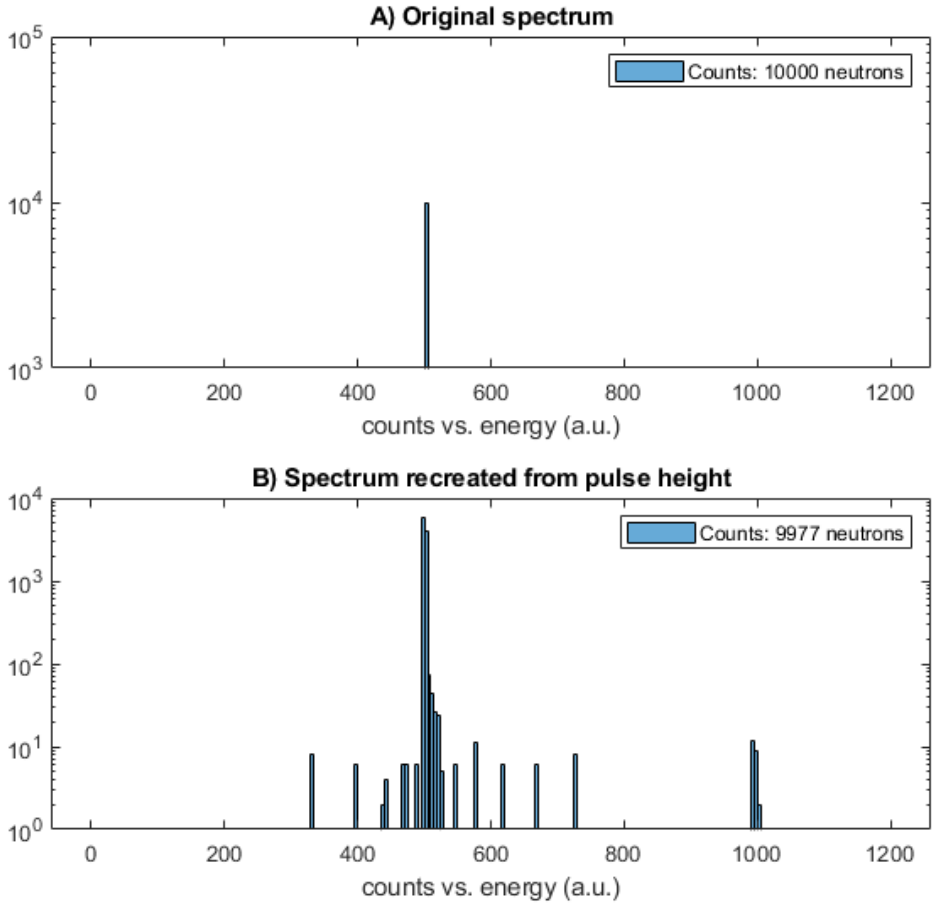


Figure 4.7: Effects on mono-energetic spectrum

Figure 4.7 shows a comparison between the theoretical spectrum and the one obtained by processing with this algorithm. Some peak broadening is expected, as well as the appearance of a secondary peak at twice the height of each pulse, corresponding to 2-fold pile-ups interpreted as single events. Indeed, centered around 10a.u. of 1000, can be found 23 events, the amount missing from the count. Two unexpected artifacts can also be seen. Firstly, the central peak broadening seems to be lopsided towards the right. Sampling a few of these cases show that there is some leftover tail being processed: this could be explained if baseline correction wasn't being performed, pointing towards a poorly defined baseline threshold. Secondly, there were at least 70 pulses spread out in pairs, in a mirrored fashion around 500, throughout the 0-500 and 500-1000 range.

One of these cases is shown in detail in figure 4.8, where a pile-up that is hard to spot by visual inspection is detected based on the change in the derivative. Checking the finite difference differential shows this is a 2-fold peak pile-up, occurring on the rising edge of the first pulse. At the point in time at which the second pulse is detected, the first one hasn't reached its maximum

value. This is leading the naive estimation to assign the wrong height, attributing the remainder energy to the second pulse, thereby explaining the mirrored pulses. This is evident when looking at the energy values for these mirrored piled-up pairs, which add up to 1000. Looking at the extra gamma pulse reveals it to also be a peak pile-up in a 3-fold pile-up, with a similar error in estimating energies. It's likely this is the source of the error in classification. This behavior was unexpected, given that rising edges are typically much shorter than the detection time resolution. On the other hand, the event detection algorithm is more powerful than expected.

In order to obtain an estimation of the peak broadening, the center peak was fitted to a normal distribution. The values outside the [470,530] range were excluded due to the mirroring mentioned earlier, obtaining $\mu = 500.09 \pm 0.02$ and $\sigma = 2.15 \pm 0.02$, which can be used to estimate $FWHM \approx 2.355 * \sigma = 5.05$.

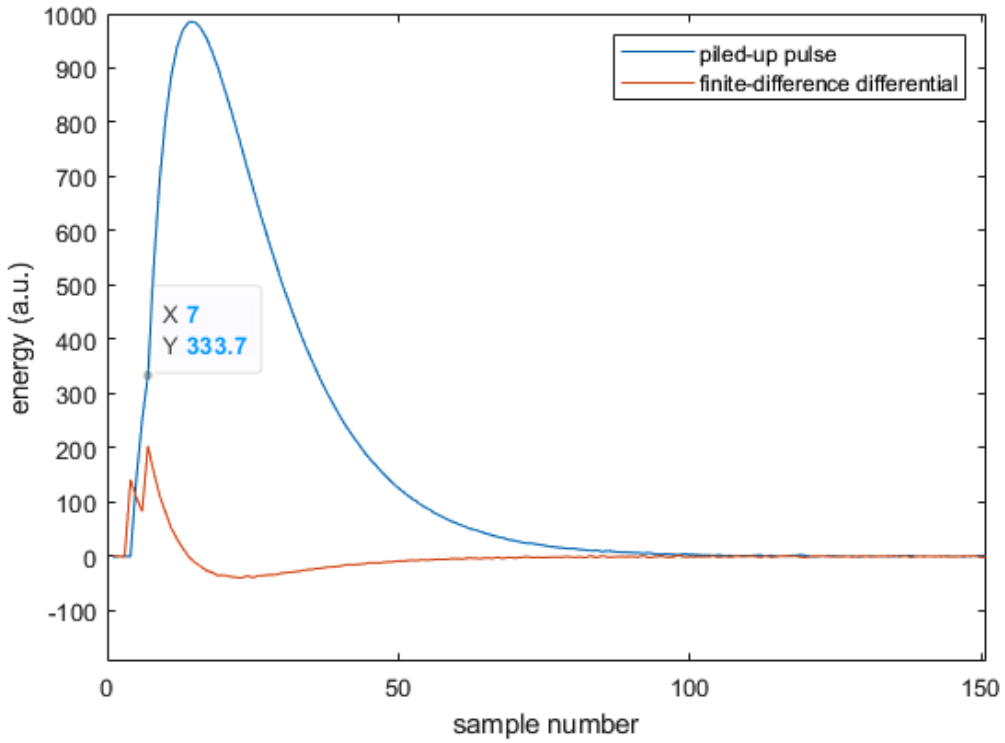


Figure 4.8: Rising edge pile-up

4.2.2 Single energy neutron and gamma

The second test array is shown in figure 4.9. The conditions are identical to the previous test, with an additional 10000 gamma pulses of 200 height each generated in the same algorithm as the neutrons, for a total of 20000 pulses. This time, the purpose is to check the effects on a spectrum composed of a single peak of gamma events and another of neutron events, which should serve as a test for the classification algorithm. The decay times for neutrons were $T_R =$ 1.2ns, $T_N = 2.7$ns, with $T_G = 4$ns to simulate the longer component.

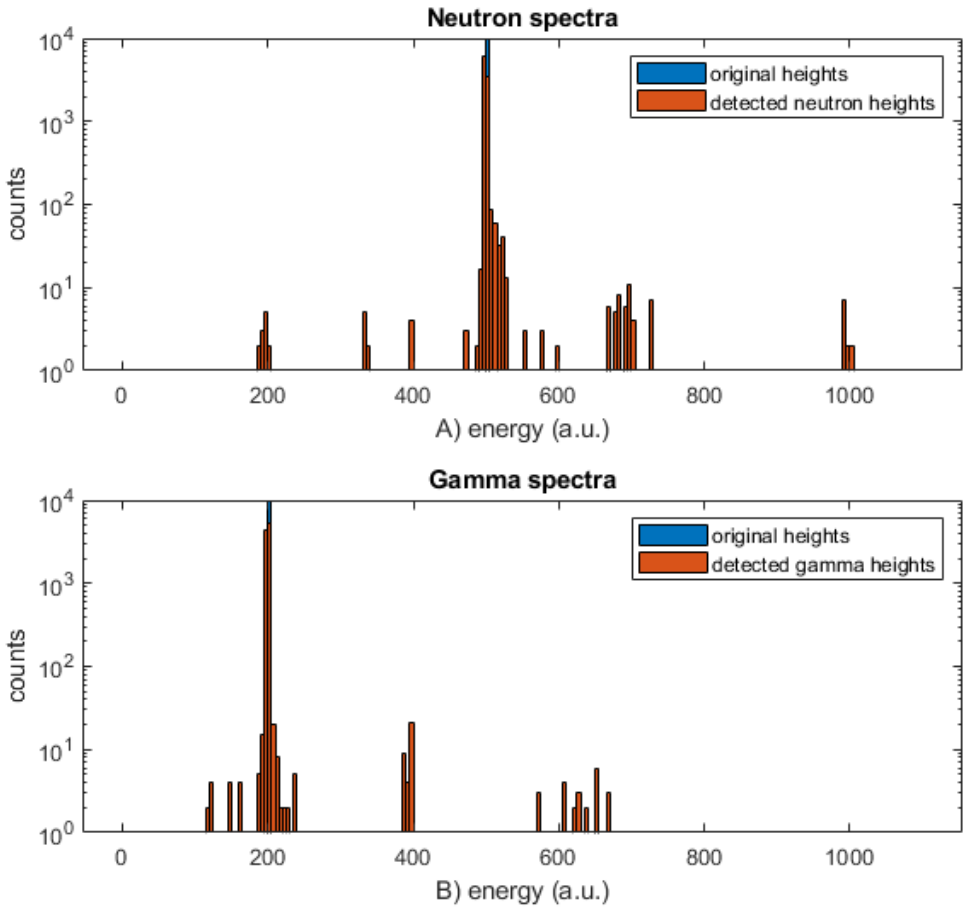


Figure 4.9: Effects on monoenergetic spectrum with neutron and gamma

Out of the 20000 pulses, 19893 were detected, 821 via 2-fold pile-up resolution and 31 via 3-fold, with 4 rejected due to higher than 3 multiplicity. This is a pile-up rate of approximately 4%. 9969 neutrons and 9922 gammas were detected. There is some peak broadening and low energy miscounting. A small number of pulses with combinations of the energies are also revealed, with peaks at the sum of these values, especially 400 and 1000. This is the expected behaviour from fully piled-up pulses. The same errors mentioned in the previous test are also present here.

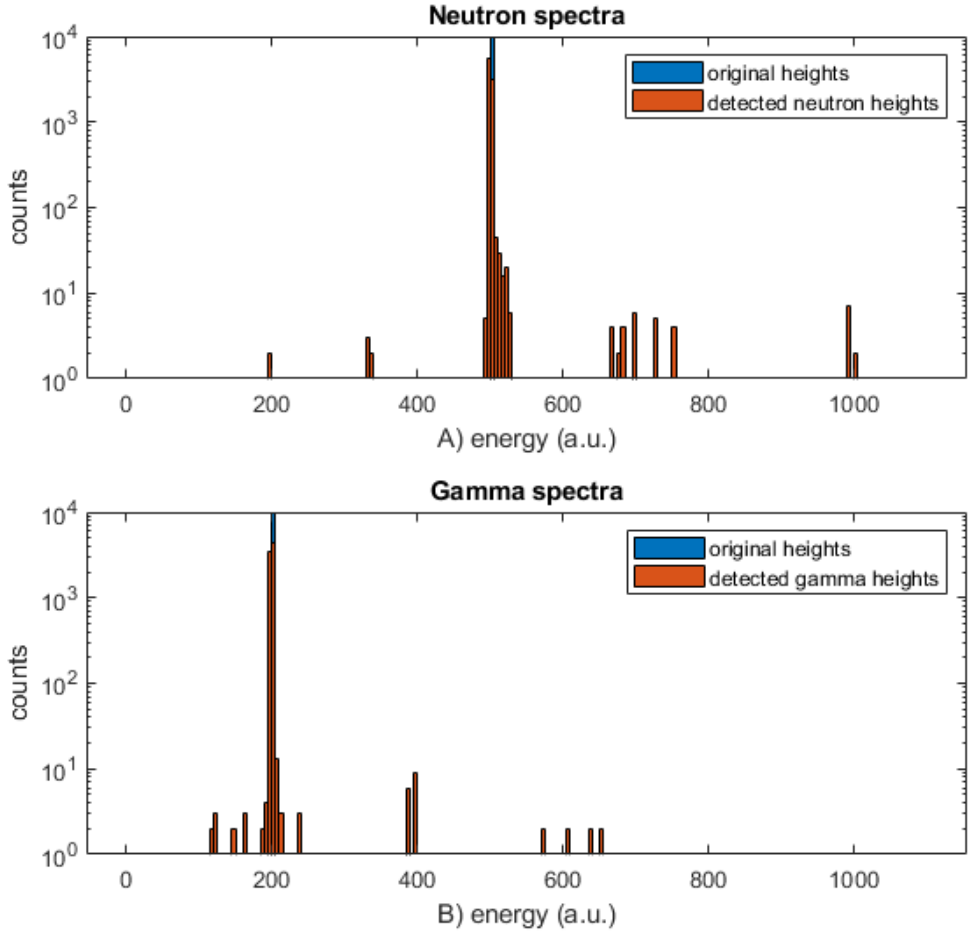


Figure 4.10: Corrected effects on monoenergetic spectrum with neutron and gamma

The residual from the squared difference can be used as a threshold to filter out events where the separation is of poor quality. For this project, with the leeway of saving all the events, this is used ad-hoc to build the spectra and identify limitations. This corrected spectrum is shown in figure 4.10. This solves the two key issues from the previous test, but as 1116 neutrons and 2075 gammas were rejected like this, this results in a decreased detection efficiency decrease of approximately 16%. The threshold used can be more stringent at the cost of decreased efficiency. Fitting for the unfiltered values produces the following values: $\mu = 500.17 \pm 0.02$ and $\sigma = 2.46 \pm 0.02$, $FWHM \approx= 5.89$. This correction results in $\mu = 500.03 \pm 0.02$ and $\sigma = 1.82 \pm 0.01$, $FWHM \approx= 4.29$.

4.2.3 Noise impact on FWHM

The test conditions in section 4.2.2 were repeated at four different noise levels, from a SNR of 40dB to 25dB in 5dB increments, with results shown in table 4.1. Although originally designed for dimensioning the level of noise, this test reveals a drastic drop in the quality of separation at higher noises, with particle type highly influencing the number of rejected events. 40dB of noise match the 48dB SNR of an 8bit ADC (6 times the number of bits). The noise shown at a baseline level, be it electronic noise or secondary low-energy signals being detected, suggests that the precision of the ADC isn't the limiting factor here, and could be above-spec for this type of acquisition.

Table 4.1: Noise level impact on the FHWM of synthetic peaks: 10k 500 energy neutrons and 10k 200 energy gammas, with or without filtering for the residual value from the separation algorithm. 40dB to 25dB. Values outside a 30-energy range of the peaks are rejected in all cases.

	40dB		**35dB**	
particle type				
neutron	count	9671	count	9670
	percentage	51.4%	percentage	51.4%
		500.171 [500.122, 500.219]		500.216 [500.165, 500.268]
		2.449 [2.414, 2.484]		2.587 [2.551, 2.624]
gamma	count	9153	count	9150
	percentage	48.6%	percentage	48.6%
		200.279 [200.241, 200.318]		200.486 [200.446, 200.525]
		1.875 [1.849, 1.903]		1.942 [1.915, 1.971]
neutron	count	9203	count	9170
	percentage	56.9%	percentage	60.7%
		499.863 [499.851, 499.875]		499.902 [499.884, 499.919]
residual filter		0.5822 [0.5739, 0.5907]		0.8730 [0.8605, 0.8858]
gamma	count	6974	count	5925
	percentage	43.1%	percentage	39.3%
		200.114 [200.102, 200.127]		200.322 [200.303, 200.341]
residual filter		0.5146 [0.5062, 0.5233]		0.7415 [0.7284, 0.7551]

	30dB		**25dB**	
particle type				
neutron	count	9664	count	9658
	percentage	51.4%	percentage	51.4%
		500.441 [500.386, 500.497]		501.091 [501.026, 501.157]
		2.780 [2.742, 2.820]		3.301 [3.255, 3.348]
gamma	count	9151	count	9145
	percentage	48.6%	percentage	48.6%
		200.962 [200.919, 201.005]		202.023 [201.971, 202.075]
		2.107 [2.076, 2.137]		2.525 [2.489, 2.562]
neutron	count	9004	count	4868
	percentage	62.9%	percentage	59.8%
		500.113 [500.085, 500.14]		501.266 [501.211, 501.321]
residual filter		1.325 [1.306, 1.345]		1.951 [1.913, 1.991]
gamma	count	5300	count	3277
	percentage	37.1%	percentage	40.2%
		200.828 [200.797, 200.859]		201.900 [201.842, 201.957]
residual filter		1.140 [1.118, 1.162]		1.678 [1.638, 1.719]

4.2.4 Detection limit

This section describes the attempts at quantifying the pile-up separation limit of the event detection algorithm, and how close can pulses occur until they are detected as individuals. Taking another look at the data from figure 4.5, where all the pulses have a theoretical height of 500, the theoretical time between pulses can be checked individually from the data used to generate the pulse train, and matched with the processed data. The pulses detected as a single pulse at around 1000 energy in these conditions must be undetected 2-fold pile-ups. Checking the time-between pulses, these range up to a maximum of $4.8 * 10^{-9}$s, or 2.42 samples at the current acquisition frequency.

The same logic was applied to the separation algorithm: The pulses corresponding to the peak pile-up that is detected but not resolved can be found by checking the tabled data for pulses interpreted as 2-fold pile-up, which also have an interpreted height well outside the vicinity of 500. These signal correct pile-up detection but poor separation, with the shortest time between pulses of this type being $5.0 * 10^{-}9$s. The separation algorithm can't correctly separate these energies, assigning more energy to one pulse and less energy to the other, preserving the total. The individual energy values of each pair are symmetrical around 500, With increasing distance decreasing the distortion in energy values. From $27 * 10^{-}9$s (14 samples) and above, there is consistently less than 1% of error in the pulse energies, compared to the theoretical value.

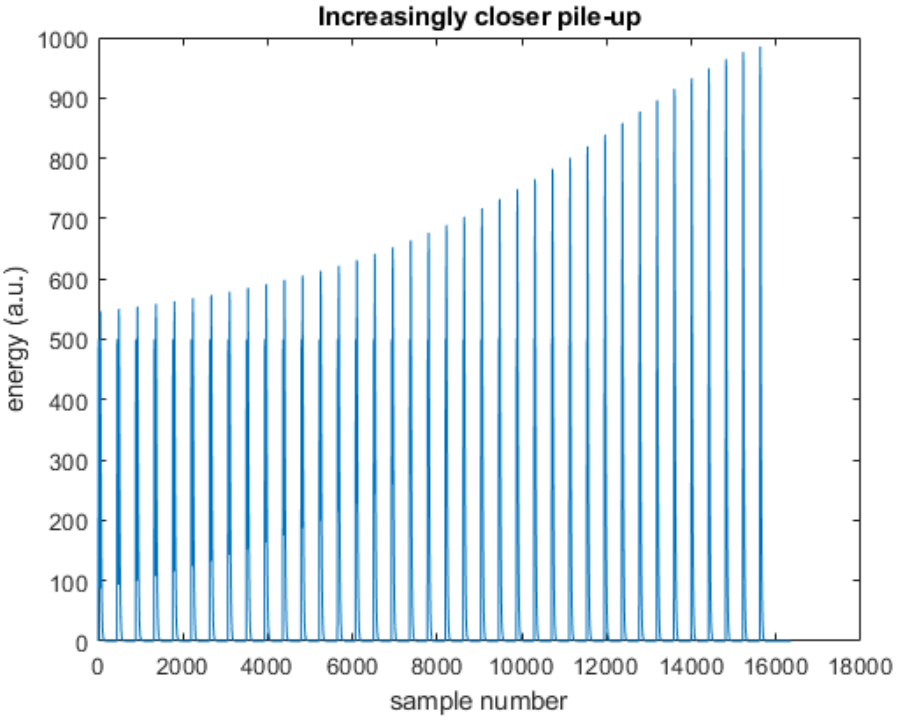


Figure 4.11: Pulse train of increasingly closely piled-up events

In order to further test the time resolution, a pulse array was created, consisting of pairs of increasingly closer pulses. Each of these pairs of pulses is well-distanced from other pairs, but each pair is piled up. This makes it possible to see how the detection and processing algorithm is interpreting each pair and see when it breaks down. Pulses are 500 height neutrons, at 500MHz acquisition frequency, and Gaussian white noise at 40dB was added. The result can be seen in image 4.11. The first pair has a separation of 40*dt (80ns). From what was seen earlier, the 2-14 (4-28ns) samples of distance between pulses should be the region of interest, so this value is chosen to provide some leeway. The closest pair has a separation of dt (2ns). Note that at an acquisition frequency of 500MHz, dt is 2ns.

The pulses at or below 4ns in distance were not detectable as pile-up. They are interpreted as single pulses and would be rejected both from poor quality of fit and at multiplicity estimation. From 5-28ns, although the pile-up is detected, the rising edge pile-up error in attributing the first pulse energy, explained in section 4.2.1, has a significant impact on the processed energy.

This test was repeated at increasing noise levels. The base 40dB had a limit of 28ns. Increasing the noise to 35dB pushed it up to 29ns; 30dB to 32ns. Further increases in noise deteriorate the energy resolution so much that this can't be tested. The lower limit is identical at all tested noise levels, including 25dB and 20dB, which is unexpected. While noise significantly impacts discrimination, crossover-based pile-up detection demonstrates a good degree of robustness to noise.

4.3 Application to Real Data

An important proof of concept for this project would be demonstrating that the event detection and pulse separation and PSD algorithm work on real data, and that the simulator can replicate real data. The following sections describe the process of attempting this, reusing data from an acquisition made in the Frascati Neutron Generator, described in the Final Report on Data Processing Algorithms and Electronic Tests[65].

The setup for the tested acquisition used a NE213 Stilbene organic scintillator coupled to a Hamamatsu model R5505-70 PMT, connected directly to the DAQ system described in full in the report[65]. The results are discussed in section 4.3.2.

4.3.1 Decay time fitting

The decay of a Stilbene scintillator in response to a gamma photon can be described as having a main fluorescence of 5.21ns carrying 95% of the energy, with a secondary component of 21.33ns and a third at 134.77ns carrying the rest. Similarly, for a neutron, a main component is found at 5.01ns, with the others at 27.70ns and 253.19ns[113]. Tests using these values were unsuccessful at both simulating pulses and pile-up separation/PSD. Though there is a dependency of the decay times on the scintillator size, it doesn't fully explain the various different values found in literature [114][36].

Throughout the research done towards this project, when gathering detector decay times, it was very rare to have the response to different particle types published, other than in research specifically concerning the pulse shape. Because of this limitation, added to the dependence of decay times on several components in the acquisition, it became clear that decay times can't be relied upon consistently to model pulses, but rather to choose a detector with a time scale appropriate for the use-case.

In order to achieve a realistic model, a consistent way to get the decay times for any input was needed. With the knowledge that reference pulses were used in this context, from sections 2.8 and 4.1, the same was done here. The approach taken involved gathering a large number of pulses, normalizing them, and using a bounded simplex search to optimize for the parameters in the double exponential function[115].

This was attempted with limited success, starting with the 2 decay components and offset for a single normalized pulse, using the theoretical values to define initial search values and boundaries. This was partly unsuccessful, in large part due to the difficulty of converging this function. Defining search bounds is critical for this search to converge, and dimensioning these

from the literature proved misleading, as mentioned earlier. The rise times output by the search were often of comparable magnitude to the decay times, and when these parameters' ranges intersect in the search, it behaves erratically. This can be verified by looking at a standard scintillator pulse and noting that the rise times last for over a sample, which is in the nanosecond range at typical acquisition frequencies, whereas literature also references values in the tens of picoseconds range[116], strictly for the detector.

The rise time has a negligible effect on the pulse shape beyond its rising edge, so a systematic error in this value would affect all pulses equally, whereas an error in the neutron/gamma decay times would affect the separation rating. So a choice was made to make the decay times a priority. First, a 3-variable search is done to find the most likely value for the rise time, after which the rise time is treated as a constant in a 2-variable search for the decay times.

The resulting decay times were then sorted into a histogram. This was first tested with 2000 synthetic pulses using the setup from section 4.2.2, with $T_R = 1.2$ns, $T_N = 2.7$ns, with $T_G = 4$ns. The obtained histograms, for the fixed rise time, and the decay times, are shown respectively in images 4.12 and 4.13. A single peak in the rise times can be seen, which, when fitted with a normal distribution, comes out to $T_R = 1.13$ns. The Decay times show well-separated peaks, fitted to $T_N = 2.88$ns and $T_G = 4.2$ns.

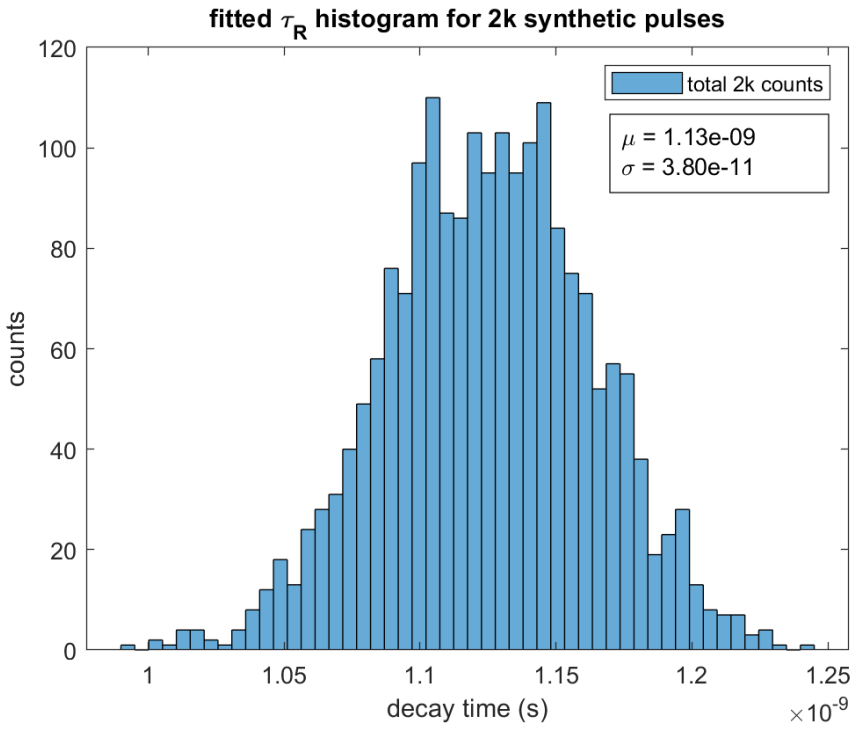


Figure 4.12: Synthetic rise time fitting

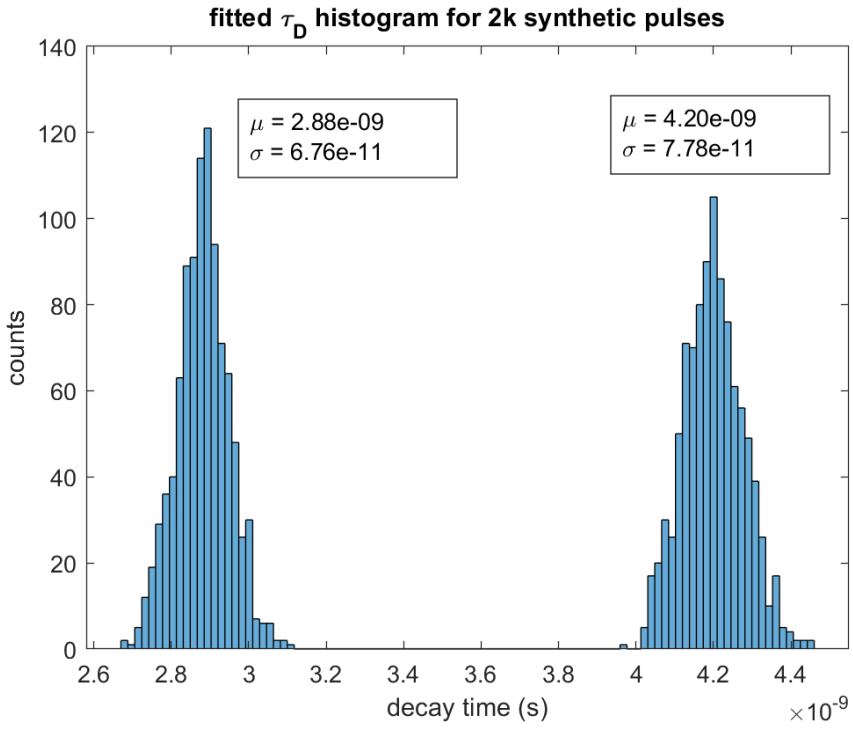


Figure 4.13: Synthetic decay time fitting

Repeating this process with an acquisition (run) from the data referenced in section 4.3, starting with a 3-variable search, a histogram for the rise times was built, shown in image 4.14. Additionally, 3 peaks can be seen. The first one, in the sub-nanosecond range, would be the first candidate for the rise time. However, these are a result of a search that got stuck in a local minimum at the lower boundary. The second graph shows the same data set but filtered for a residual value to remove poor-quality searches. This removes the first peak, but an unexpected second peak is still present. The first peak matches the expected result[65] and has a higher count, while the second peak matches one of the decay times attributed to gammas in this search, shown in image 4.15, likely an artifact of intersecting boundary conditions. Because of this, the first peak was interpreted as representing the rise time, rather than a dual-mode interpretation. The lower boundary convergence can also be seen in the decay time histogram.

By fixing the rise time to the one obtained, the search is repeated, now with 2 degrees of freedom. The resulting decay time histogram is shown in image 4.16.

The peaks in image 4.16 were interpreted as the two particle types: neutron and gamma. Using a normal regression around each peak, values for the decay times were obtained. The obtained neutron and gamma decay times were respectively $\tau_N = 6.434 * 10^{-9}s$ and $\tau_G = 7.465 * 10^{-9}s$, while the rise time was $\tau_R = 2.630 * 10^{-9}s$.

These decay times worked for the algorithms, and they are similar to the expected 5ns[113]

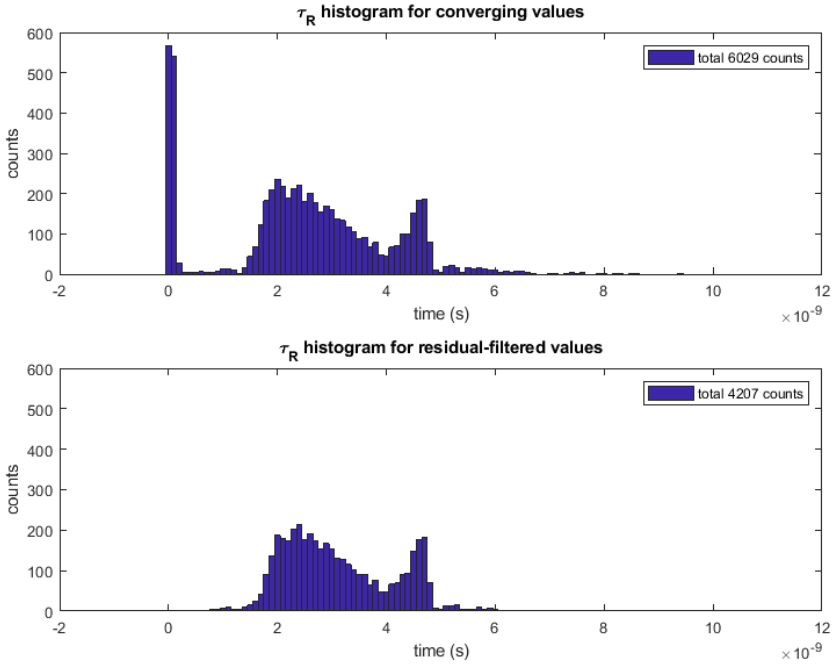


Figure 4.14: Stilbene rise time distribution

and 3.2-7.3ns range depending on thickness[114]. The rise time is also similar to the empirically expected 3.5ns[65].

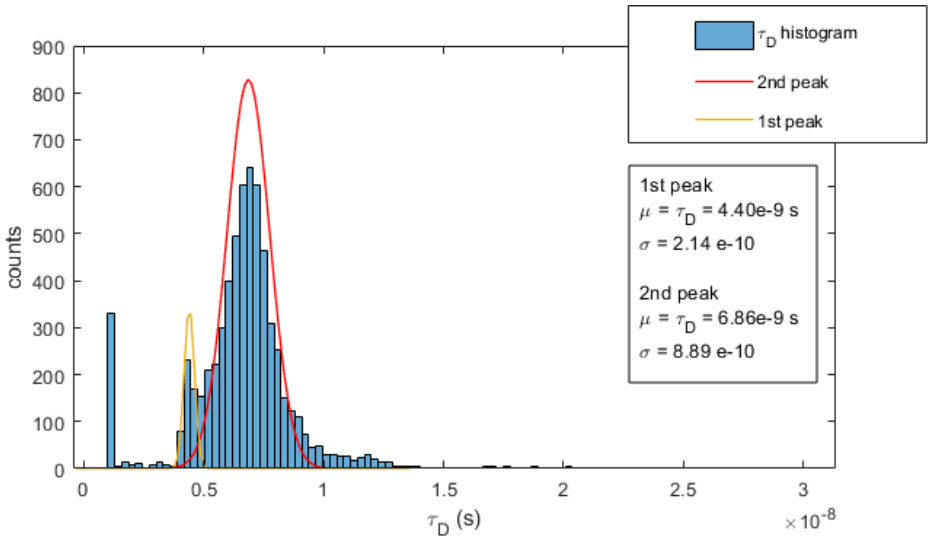


Figure 4.15: Stilbene decay times fitting

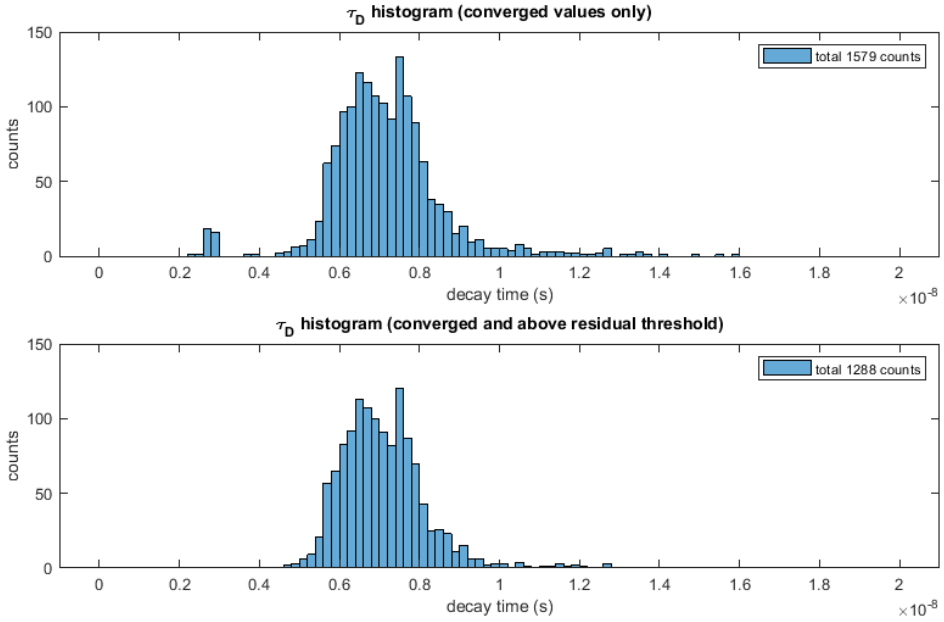


Figure 4.16: Stilbene decay times fitting

4.3.2 Case study: Stilbene detector

The discharge chosen for this case study, run 27, had been processed in the existing off-line processing code, with 714kEvents/s and a pile-up percentage of 6.32%. This run was chosen due to the relatively high levels of pile-up; because there is neutron-gamma pile-up, and the organic scintillator used was seen in section 2.5.2 to be capable of Pulse Shape Discrimination; and due to the availability of a spectrum to compare to, shown in image 4.17.

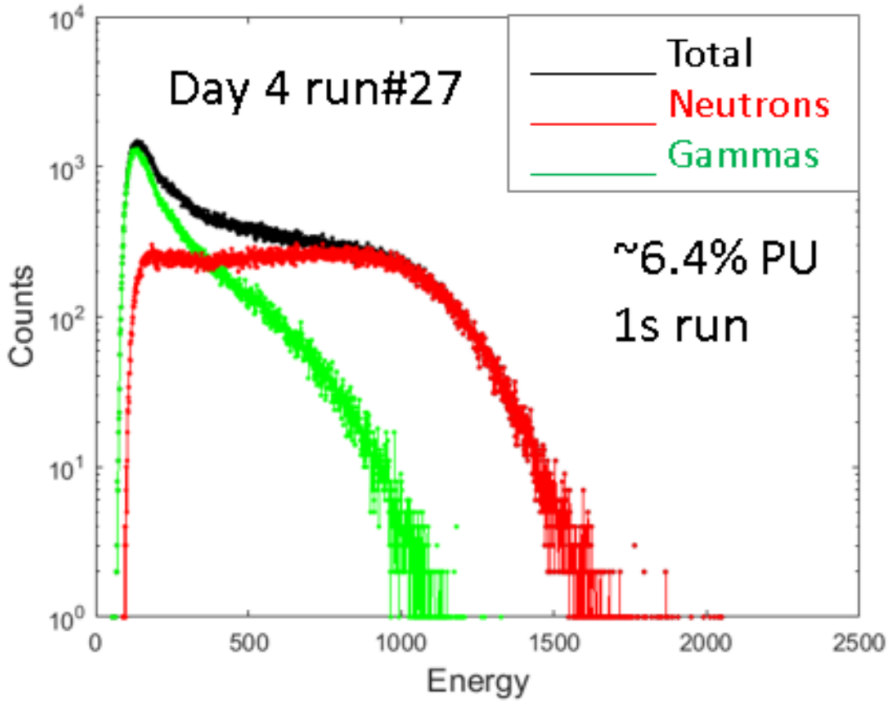


Figure 4.17: Spectrum of run 27, from [65]

The detection algorithm was adapted to interpret this data in event mode, with incoming pulses formatted into windows of 32 samples, extendable up to 128 samples. The first 4 samples encode a timestamp, with the last two encoding a tag for identification and an estimated number of pulses contained in that event.

The separation algorithm was set to run on 50k events. This number is an approximation, as the window size and pulse count per event are variable - the total count was 50730 pulses. Out of these, 39820 were interpreted as neutrons and 9251 as gammas, roughly 19%gammas with a detected pileup of 13,4%. The output spectrum is shown in image 4.18. When calculating the multiplicity of an event, in cases of disagreement with the estimate of the number of pulses from the acquisition, the event was rejected. These rejections totalled 1652 pulses. 7 events were rejected due to multiplicity higher than 3. Although the general shape of this detected spectrum

matches the original, at low energies the erroneous classification of gammas as neutrons is still present.

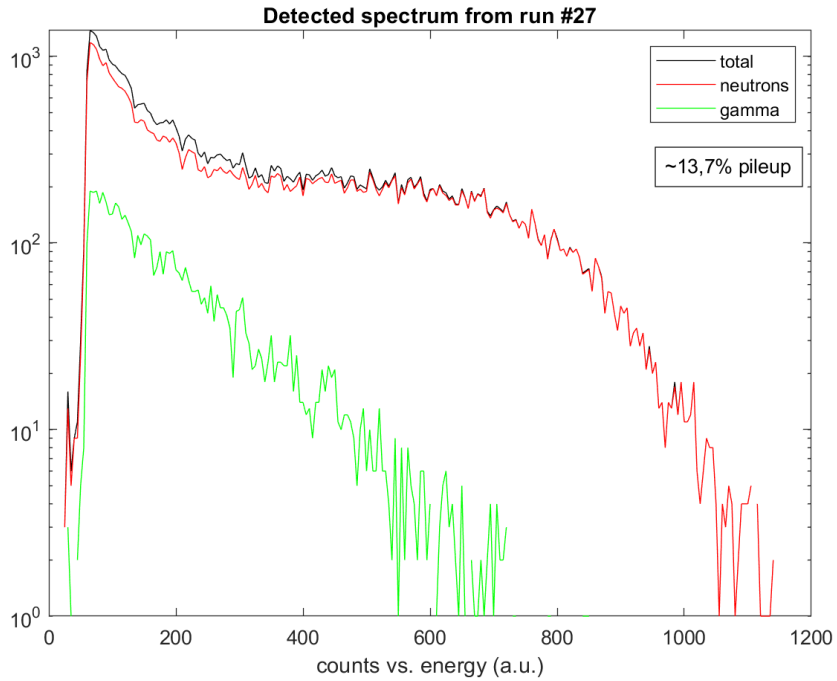


Figure 4.18: Spectrum of run 27, using the separation algorithm

Using this spectrum as a basis, the procedure from chapter 3 was followed: the energies were sampled 10k times according to their particle type, creating single events based on the double exponential and extracted decay times for this scintillator. These were randomly placed over time by sampling an exponential distribution corresponding to an average time-between-pulse of $\lambda = 2.7 * 10^{-7}$, creating a synthetic pulse train of 10k pulses, to which Gaussian white noise at an SNR of 35dB was added. The same algorithm, in continuous mode, was used to detect these events and build a spectrum, shown in image 4.19. The total count was 9586 pulses, with 8793 neutrons and 788 gammas detected, or 8.22% gammas. The λ value was chosen to reach the target 13.4% pile-up, with a detected amount of 13.5% pile-up. This was estimated based on 2.11, by counting the number of pulses distanced less than the pulse width, using a pulse width of 15 samples and a frequency of 400MHz. Each sample lasts for 2.5ns, so even though the window is 32 samples or 80ns, the actual duration of a pulse is around 15 samples or 37.5ns.

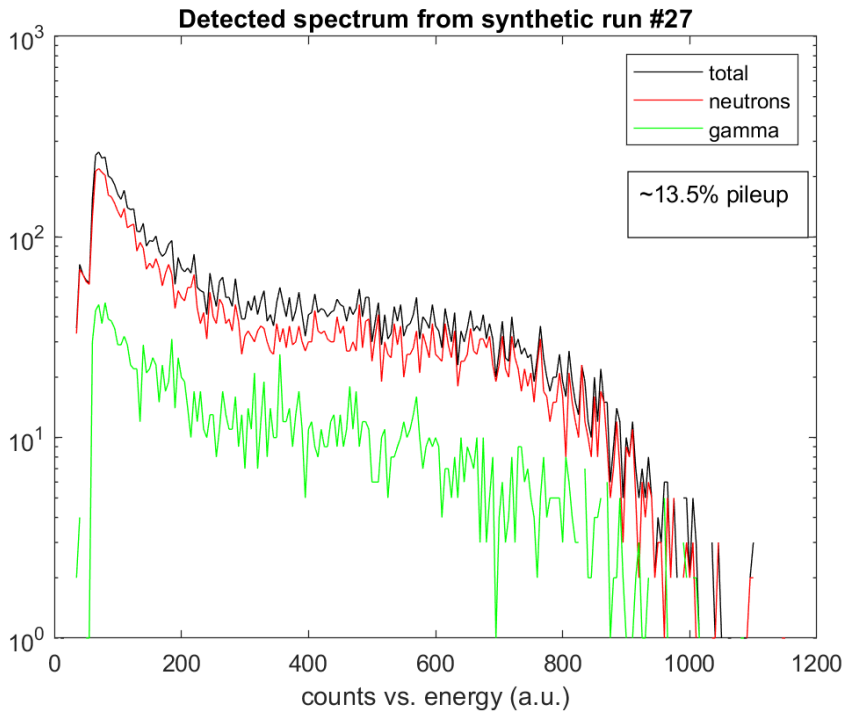


Figure 4.19: Spectrum of recreated run 27, using the separation algorithm

5 Conclusion

Chapter 3 presented the development of a simulation tool, capable of creating synthetic pulse trains in accordance with the desired acquisition characteristics.

Chapter 4 presented the development of an event detection algorithm used to process the pulse train data into events. This chapter also explained the pile-up separation and pulse shape discrimination algorithm, which was tested with custom synthetic pulse trains and with real data, validating it and discussing some of its limitations.

This means the main objectives of this thesis were achieved, and the discussion of results is left for the current chapter. Section 5.1 evaluates the performance of the developed algorithms. Section 5.2 talks about general considerations and limitations observed throughout the development process. Section 5.2.1 provides a short analysis of how the algorithm would perform in a real-time environment. Finally, section 5.3 explains how further development could be done, and how to possibly answer questions raised in this work.

5.1 Results

For synthetic pulses, the event detection algorithm showed great accuracy. Though omitted from this work, the performance was similar using well-behaved calibration runs. However, sections 4.2.4 and 4.3 showed that in noisier environments, or where the time and energy resolution are a limiting factor, there can be a significant degradation of the detection efficiency.

The algorithm developed by Luo et al achieved a disentanglement limit of 20ns, with a sub-6% reconstruction error at a 50ns distance between pulses[106]. The pile-up detection and multiplicity estimation algorithm used in this is identical and was found, surprisingly, to be capable of detecting pile-ups at even shorter intervals, even in the 5-28ns range. However, the pulse height estimation used in the separation algorithm misattributed the energies in this interval, which corresponds to the occurrence of a rise pile-up, which is an unsolved issue in this field. It would have been helpful to get the reconstruction accuracy in real data, but that would require another classification method to cross-check the particle types, such as charge

integrating the reconstructions or time-of-flight analysis.

The separation algorithm wrongly classified gammas at lower energies. Gamma pulses are shorter than neutrons, so a worse energy resolution is expected, making it common to find this type of degradation[65][106]. However, the levels shown in section 4.2.3 were higher than expected, in particular with an SNR of 30dB or smaller. This effect is clearly visible by comparing the various spectra from section 4.3.2. While this only attenuates the issue, the residual value from the fit was shown to be capable of rejecting poor results.

The real-data analysis from the Stilbene acquisitions from section 4.3.2 provided some interesting observations. The multiplicity estimation provided by the FPGA used to acquire the data was frequently in disagreement with the crossover method. Looking at the waveforms where this happened showed that both estimates were wrong on occasion. This was addressed by rejecting pulses where the estimates didn't match, but the source of the discrepancies could be further investigated, especially given that the pile-up percentage was roughly double the expected.

5.2 Discussion

The discussion starts by addressing the multiplicity estimation used as a way to locate the position of each pulse in a piled-up event. The position might not be the true maximum, firstly limited by the time resolution encoded in the acquisition frequency, and being somewhat vulnerable to noise. However, a bigger issue comes from the range of possible values, which are bound to a specific sample in the data. This is where the acquisition frequency has the greatest impact because the duration of each sample is determined by the acquisition frequency's inverse. At the time scales in these acquisitions, the duration of a sample is comparable to the rise time, so an intrinsic error of 1 sample has a massive impact on the reconstruction of the event, as a small offset creates a slight deviation in the waveform that quickly stacks up when comparing arrays point-by-point.

When this was noticed, a simplex search was attempted to find the optimal values for the amplitudes and offset. This method, as explored by Magdowski[117], provided great preliminary results, for both single pulse maximum position and energy, and while the accuracy was impressive when successful, I wasn't able to consistently get the search to converge, so this method couldn't be generalized. Fortunately, this eventually led to the decay time fitting from section 4.3.1, which posed a different challenge, first covered there. For starters, this type of exponential sum is a notoriously difficult function to converge onto, behaving unpredictably with small changes to the decay times, so the search failed frequently or converged onto local minima or boundaries, wasting a lot of pulses and processing time. Some topics were found in mathematics

forums attempting to linearize this specific search, suggesting this might be a possible avenue for improvement, though it seems more pragmatic to *simply* increase sample rates. In either case, the computational costs of a single simplex search disqualify this method from real-time use, much less the 4 searches required for 2-fold pile-up separation and PSD.

A more viable real-time approach would be embracing the fast peak estimation as having an inherent precision of 1 sample, using that value as a starting point to obtain a better sub-sample maximum position estimate. This could be done by interpolating the recreated pulse at smaller intervals in the case of a reference template, or sampling the exponential function at different time points if that proves to be faster. In this way the peaks would be more precisely defined, drastically improving the separation algorithm. This was tested for a while, but the same point-difference error accumulation was found to also come from the amplitudes. In addition, this method of obtaining the maximum position or the offset between piled-up pulses approach starts to resemble Luo's iterative search, which was avoided to reduce computational costs. If this were then extended to amplitude and position, now the iterative search is extended to 2 dimensions, at which point the simplex search starts to look appealing. In the end, a choice was made to correct the offset at each stage of the recreation algorithm in integer samples, but there is no doubt that errors in pile-up classification occur due to this compromise.

Unlike the double exponential model used throughout this project, which approximates the fluorescence to two main decay components, rise and decay times, some authors describe the fluorescence more precisely, with more than one main decay constant per particle, as mentioned earlier[113]. However, the time resolution required to model these accurately exceeds the acquisitions used in section2.8.4, and modelling the pulse with those additional decay constants would increase the degrees of freedom in the search by two. Considering that the main component carries most of the energy (95%), even a large change in the secondary components would impact the pulse shape less than the variation in the main decay, which can be seen in image 4.16. Since there was a minimal benefit to modelling the extra components, at the cost of the search's quality and speed, the 2-decay components model was kept. However, for a system with better time resolution, more precise models might make more sense.

The question that remains to be answered is: Why this specific pulse model? The main initial argument was relatability to a physical concept, as it is a semi-empirical formula. But even this model is an approximation of the real signal, limited mainly because current acquisition frequencies and energy resolutions aren't good enough to model more realistic responses, especially not at these event rates. The other reason comes from having an analytic formula, which was used to great effect to calculate initial amplitudes from the pulse height, shown in

section 3.2, and later applied in the separation algorithm. If the main goal is strictly improving the pulse separation quality, the model that gives the best results should be picked, regardless of physical significance. For example, creating a reference pulse by averaging waveforms (for each detectable particle type) is a solid alternative[106], as is using a spline approximation, which takes advantage of the speed improvement of sampling a polynomial function over an exponential.

Another question is whether the pulse shape is constant throughout the energy spectrum. All of these models imply a linear scaling with amplitude, so a perfectly acquired neutron pulse of height X would have precisely half the height and area of a perfectly acquired neutron pulse of height 2X. But averaging *normalized* pulses does not test for this, so that remains an assumption until further investigation. Sorting the fitted decay times according to the respective pulse height showed signs of non-linearity, but this was done with an earlier version of the search, using calibration (synthetic) data, so not much can be said.

The data used in the Stilbene case study from section 4.3.2 had a very limited energy and time resolution for pulses with a duration of 15 samples, so the decay constants were obtained in a different run with an identical setup, but better conditions. This might explain some of the variance in the two decay times, and contribute to the spectral distortion seen mostly at lower energies.

5.2.1 Real-Time considerations

For real-time implementation, it is helpful to frame this algorithm as part of the processing tasks for the neutron emissivity profile shown in image 1.4. Tasks 2 and 3 are already implemented in MARTe as GAMs (Generic Application Modules), which can be run in under 2ms[65]. As a part of the 10ms control cycle, this leaves up to an 8ms margin for task 1. Most of the processing time in this task is taken by the acquisition and processing algorithms described in this thesis, though without measuring the complete task, it's hard to give an estimate of how much would be occupied by each stage. So instead, this section discusses possible processing bottlenecks.

The real-time performance of the separation and PSD algorithms depends on how fast each individual event is processed, and how this speed scales when events are processed in bulk. Section 2.6.4 establishes that this is dictated by how well the algorithm parallelizes in the digital processing unit used, likely to be an FPGA. If e.g. 128 events are processed in 128 or fewer times the duration it takes to process a single pulse plus the parallel overhead, the algorithm can be profitably parallelized at this stage. So the time it takes to create a spectrum improves with the number of threads available to process events. The separation algorithm is designed to

work on event data, so scalability should be close to linear.

The event detection stage also affects this time, but since it operates on a real-time data stream, there is little point in parallel processing at this stage. Bottlenecks at this point can appear if the event rate is too high for the DPU to process all the events. Therefore, improvements at this stage come mainly from reducing data loss from saturation or detector dead-time. For instance, the pulse-width parameter used to create acquisition windows affects the maximum event rate (MB/s) that can be processed without data loss[65]. The speed of the event detection algorithm itself doesn't seem relevant at a time scale of milliseconds. On the contrary, because there is a good margin to spare within the control cycle, it might even make more sense to improve the detection algorithm in terms of accuracy at the cost of increased processing, making it slower. Whatever changes are made at this stage are independent of the rest of the spectrum building.

5.3 Future work

Most of the limitations in the performance of the separation and PSD algorithm could be answered by testing with more real data from different sources, in order to establish if the source of errors is the algorithm itself, or how it interacts with noisy data. In the same vein, the lack of cross-validating the PSD with competing algorithms limited the performance testing to spectral observations and statistical analysis, which is not ideal. It is also unclear how applying PSD methods to single pulses reconstructed pulses from piled-up events affects the PSD algorithm's performance.

Implementing the separation algorithm in an FPGA would be the next step, but this was regrettably not pursued due to lack of time.

The event detection algorithm proved more promising than expected. If adapted to be used as a real-time triggering algorithm, it could provide some improvements to event detection, applicable to different processing tasks, though it should be noted that the current one is designed with data throughput limitations in mind.

Bibliography

[1] "Electricity market report - july 2021." https://www.iea.org/reports/electricity-market-report-july-2021. Accessed: 2022-4-21.

[2] E. COMMISSION, "2020 report on the state of the energy union pursuant to regulation (eu) 2018/1999 on governance of the energy union and climate action," 2020.

[3] M. Arbabzadeh, R. Sioshansi, J. X. Johnson, and G. A. Keoleian, "The role of energy storage in deep decarbonization of electricity production," *Nature communications*, vol. 10, no. 1, pp. 1–11, 2019.

[4] M. C. McManus, "Environmental consequences of the use of batteries in low carbon systems: The impact of battery production," *Applied Energy*, vol. 93, pp. 288–295, 2012.

[5] K. Gi, F. Sano, K. Akimoto, R. Hiwatari, and K. Tobita, "Potential contribution of fusion power generation to low-carbon development under the paris agreement and associated uncertainties," *Energy Strategy Reviews*, vol. 27, p. 100432, 2020.

[6] B. W. Brook, "Could nuclear fission energy, etc., solve the greenhouse problem? the affirmative case," *Energy Policy*, vol. 42, pp. 4–8, 2012.

[7] S. Lorenczik, S. Kim, B. Wanner, J. M. Bermudez Menendez, U. Remme, T. Hasegawa, J. H. Keppler, L. Mir, G. Sousa, M. Berthelemy, *et al.*, "Projected costs of generating electricity-2020 edition," tech. rep., Organisation for Economic Co-Operation and Development, 2020.

[8] B. K. Sovacool, R. Andersen, S. Sorensen, K. Sorensen, V. Tienda, A. Vainorius, O. M. Schirach, and F. Bjørn-Thygesen, "Balancing safety with sustainability: assessing the risk of accidents for modern low-carbon energy systems," *Journal of Cleaner Production*, vol. 112, pp. 3952–3965, 2016.

[9] "Advantages of fusion." https://www.iter.org/proj/inafewlines.

[10] A. Donné, "The european roadmap towards fusion electricity," *Philosophical Transactions of the Royal Society A*, vol. 377, no. 2141, p. 20170432, 2019.

[11] M. Szczepański, "2017 euratom briefing." `https://www.europarl.europa.eu/RegData/etudes/BRIE/2017/608665/EPRS_BRI(2017)608665_EN.pdf`.

[12] " Translating JET into ITER." `https://www.iter.org/newsline/-/3520/`, 2020. [Online; accessed 14-January-2020].

[13] M. Walsh, P. Andrew, R. Barnsley, L. Bertalot, R. Boivin, D. Bora, R. Bouhamou, S. Ciattaglia, A. Costley, G. Counsell, *et al.*, "Iter diagnostic challenges," in *2011 IEEE/NPSS 24th Symposium on Fusion Engineering*, pp. 1–8, IEEE, 2011.

[14] F. Pompili, B. Esposito, D. Marocco, S. Podda, M. Riva, S. Baccaro, A. Cemmi, I. Di Sarcina, L. Quintieri, D. Bocian, *et al.*, "Radiation and thermal stress test on diamond detectors for the radial neutron camera of iter," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 62–64, 2019.

[15] N. Cruz, B. Santos, A. Fernandes, P. F. Carvalho, J. Sousa, B. Gonçalves, M. Riva, C. Centioli, D. Marocco, B. Esposito, C. M. B. Correia, and R. C. Pereira, "The design and performance of the real-time software architecture for the iter radial neutron camera," *IEEE Transactions on Nuclear Science*, vol. 66, no. 7, pp. 1310–1317, 2019.

[16] N. Cruz, B. Santos, A. Fernandes, P. Carvalho, J. Sousa, B. Goncalves, M. Riva, C. Centioli, D. Marocco, B. Esposito, *et al.*, "The design and performance of the real-time software architecture for the iter radial neutron camera," *IEEE Transactions on Nuclear Science*, vol. 66, no. 7, pp. 1310–1317, 2019.

[17] "cividec instrumentation portfolio," 2017.

[18] A. Costley, "On the fusion triple product and fusion power gain of tokamak pilot plants and reactors," *Nuclear Fusion*, vol. 56, no. 6, p. 066003, 2016.

[19] K. Miyamoto, *Plasma physics and controlled nuclear fusion*, vol. 38. Springer Science & Business Media, 2005.

[20] "What is Iter?." `https://www.iter.org/sci/Fusion`. [Online; accessed 16-December-2022].

[21] R. J. Goldston, *Introduction to plasma physics*. CRC Press, 2020.

[22] M. Arnould and S. Goriely, "The p-process of stellar nucleosynthesis: astrophysics and nuclear physics status," *Physics Reports*, vol. 384, no. 1-2, pp. 1–84, 2003.

[23] T. Kajino, W. Aoki, A. Balantekin, R. Diehl, M. Famiano, and G. Mathews, "Current status of r-process nucleosynthesis," *Progress in Particle and Nuclear Physics*, vol. 107, pp. 109–166, 2019.

[24] F. Ajzenberg-Selove, "Energy levels of light nuclei a= 5- 10," *Nuclear Physics A*, vol. 490, no. 1, pp. 1–225, 1988.

[25] J. Garcia, R. Dumont, J. Joly, J. Morales, L. Garzotti, T. Bache, Y. Baranov, F. Casson, C. Challis, K. Kirov, *et al.*, "First principles and integrated modelling achievements towards trustful fusion power predictions for jet and iter," *Nuclear Fusion*, vol. 59, no. 8, p. 086047, 2019.

[26] D. Clery, "Laser-powered fusion effort nears 'ignition'," 2021.

[27] M. A. Liberman, J. S. De Groot, A. Toor, and R. B. Spielman, "Physics of high-density z-pinch plasmas," 2012.

[28] H. Bodin, "The reversed field pinch," *Nuclear Fusion*, vol. 30, no. 9, p. 1717, 1990.

[29] T. R. Jarboe, "Review of spheromak research," *Plasma Physics and Controlled Fusion*, vol. 36, no. 6, p. 945, 1994.

[30] A. H. Boozer, "What is a stellarator?," *Physics of Plasmas*, vol. 5, no. 5, pp. 1647–1655, 1998.

[31] J. Wilson and P. Bonoli, "Progress on ion cyclotron range of frequencies heating physics and technology in support of the international tokamak experimental reactor," *Physics of Plasmas*, vol. 22, no. 2, p. 021801, 2015.

[32] R. Budny, J. Cordey, T. Team, and J. Contributors, "Core fusion power gain and alpha heating in jet, tftr, and iter," *Nuclear Fusion*, vol. 56, no. 5, p. 056002, 2016.

[33] I. Organization, "Iter research plan within the staged approach," *ITER Technical Report ITR-18-003*, pp. 57–164, 2018.

[34] G. Zhuang, G. Li, J. Li, Y. Wan, Y. Liu, X. Wang, Y. Song, V. Chan, Q. Yang, B. Wan, *et al.*, "Progress of the cfetr design," *Nuclear Fusion*, vol. 59, no. 11, p. 112010, 2019.

[35] S. Ishiyama, Y. Muto, Y. Kato, S. Nishio, T. Hayashi, and Y. Nomoto, "Study of steam, helium and supercritical co2 turbine power generations in prototype fusion power reactor," *Progress in nuclear energy*, vol. 50, no. 2-6, pp. 325–332, 2008.

[36] G. F. Knoll, *Radiation detection and measurement*. John Wiley & Sons, 2010.

[37] D. Makivić, "Production of gamma rays in fusion reactors,"

[38] V. Kiptily, T. Kondoh, V. Naidenov, I. Polunovski, Y. V. Tuboltsev, I. Chugunov, and A. Shevelev, "Gamma ray spectrometry in iter: conceptual design," in *Diagnostics for Experimental Thermonuclear Fusion Reactors 2*, pp. 511–519, Springer, 1998.

[39] N. Gorelenkov, S. Pinches, and K. Toi, "Energetic particle physics in fusion research in preparation for burning plasma experiments," vol. 54, p. 125001, nov 2014.

[40] I. H. Hutchinson, "Principles of plasma diagnostics," *Plasma Physics and Controlled Fusion*, vol. 44, no. 12, p. 2603, 2002.

[41] D. Moseev, M. Salewski, M. Garcia-Muñoz, B. Geiger, and M. Nocente, "Recent progress in fast-ion diagnostics for magnetically confined plasmas," *Reviews of Modern Plasma Physics*, vol. 2, no. 1, pp. 1–68, 2018.

[42] V. Kiptily, F. Cecil, O. Jarvis, M. Mantsinen, S. Sharapov, L. Bertalot, S. Conroy, L. Ingesson, T. Johnson, K. Lawson, *et al.*, "γ-ray diagnostics of energetic ions in jet," *Nuclear Fusion*, vol. 42, no. 8, p. 999, 2002.

[43] R. Pereira, J. Sousa, A. Fernandes, F. Patrício, B. Carvalho, A. Neto, C. Varandas, G. Gorini, M. Tardocchi, D. Gin, *et al.*, "Atca data acquisition system for gamma-ray spectrometry," *Fusion engineering and design*, vol. 83, no. 2-3, pp. 341–345, 2008.

[44] S. Longo and J. M. Roney, "Hadronic vs. electromagnetic pulse shape discrimination in csi (tl) for high energy physics experiments," *Journal of Instrumentation*, vol. 13, no. 03, p. P03018, 2018.

[45] J. BiaŁkowski, Z. Moroz, and M. Moszyński, "Further study of timing properties of scintillation counters," *Nuclear Instruments and Methods*, vol. 117, no. 1, pp. 221–226, 1974.

[46] Z. He, "Review of the shockley–ramo theorem and its application in semiconductor gamma-ray detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 463, no. 1, pp. 250–267, 2001.

[47] P. Lechner, A. Pahlke, and H. Soltau, "Novel high-resolution silicon drift detectors," *X-Ray Spectrometry: An International Journal*, vol. 33, no. 4, pp. 256–261, 2004.

[48] C. Weiss, *A CVD diamond detector for (n, α) cross-section measurements*. PhD thesis, 2014.

[49] M. I. Kobayashi, M. Angelone, S. Yoshihashi, K. Ogawa, M. Isobe, T. Nishitani, S. Sangaroon, S. Kamio, Y. Fujiwara, T. Tsubouchi, A. Uritani, M. Sakama, and M. Osakabe, "Thermal neutron measurement by single crystal cvd diamond detector applied with the pulse shape discrimination during deuterium plasma experiment in lhd," *Fusion Engineering and Design*, vol. 161, p. 112063, 2020.

[50] M. Hamel, ed., *Plastic Scintillators*. Springer International Publishing, 2021.

[51] "Large-area picosecond photo-detectors project." `https://psec.uchicago.edu/library/`.

[52] H. Photonics, "Photomultiplier tubes - basics and applications 3rd edition," 2007.

[53] L. Cathey, "Fatigue in photomultipliers," *IRE Transactions on Nuclear Science*, vol. 5, no. 3, pp. 109–114, 1958.

[54] M. Riva, B. Esposito, D. Marocco, F. Belli, B. Syme, J.-E. Contributors, *et al.*, "The new digital electronics for the jet neutron profile monitor: Performances and first experimental results," *Fusion engineering and design*, vol. 86, no. 6-8, pp. 1191–1195, 2011.

[55] Y. K. Kim, S. K. Lee, B. H. Kang, J. B. Son, and G. D. Kim, "Performance improvement of neutron flux monitor at KSTAR," *Journal of Instrumentation*, vol. 7, pp. C06013–C06013, jun 2012.

[56] M. Ahmed, "A comparative study of n-γ discrimination properties of scintillators ne 213, c6h6, c6d6 and stilbene," *Nuclear Instruments and Methods*, vol. 143, no. 2, pp. 255–257, 1977.

[57] T. Craciunescu, G. Bonheure, V. Kiptily, A. Murari, I. Tiseanu, and V. Zoita, "A comparison of four reconstruction methods for jet neutron and gamma tomography," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 605, no. 3, pp. 374–383, 2009.

[58] C. Kim, J.-Y. Yeom, and G. Kim, "Digital n- Pulse Shape Discrimination in Organic Scintillators with a High-Speed Digitizer," *Journal of Radiation Protection and Research*, vol. 44, pp. 53–63, June 2019.

[59] D. Rigamonti, A. Muraro, M. Nocente, V. Perseo, G. Boltruczyk, A. Fernandes, J. Figueiredo, L. Giacomelli, G. Gorini, M. Gosk, *et al.*, "Performance of the prototype labr3 spectrometer developed for the jet gamma-ray camera upgrade," *Review of Scientific Instruments*, vol. 87, no. 11, p. 11E717, 2016.

[60] J. Malkar and P. Kunte, "Pulse shape discrimination technique in a nai(tl)csi(na) x-ray detector," *Nuclear Instruments and Methods in Physics Research*, vol. 202, no. 3, pp. 465–468, 1982.

[61] S. Rawat, M. Tyagi, P. Netrakanti, V. Kashyap, A. Mitra, A. Singh, D. Desai, G. A. Kumar, and S. Gadkari, "Pulse shape discrimination properties of gd3ga3al2o12:ce,b single crystal in comparison with csi:tl," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 840, pp. 186–191, 2016.

[62] D. Lattanzi, M. Angelone, M. Pillon, S. Almaviva, M. Marinelli, E. Milani, G. Prestopino, A. Tucciarone, C. Verona, G. Verona-Rinati, *et al.*, "Single crystal cvd diamonds as neutron detectors at jet," *Fusion Engineering and Design*, vol. 84, no. 7-11, pp. 1156–1159, 2009.

[63] E. Kowalski, *Nuclear electronics*. Springer-Verlag, 2013.

[64] B. A. Shenoi, *Introduction to digital signal processing and filter design*. John Wiley & Sons, 2005.

[65] R. M, "Rt-f4e fpa327 05-05 final report on data processing algorithms and electronic tests," tech. rep., RNC & RGRS Consortium, 2018.

[66] W. Guo, R. P. Gardner, and C. W. Mayo, "A study of the real-time deconvolution of digitized waveforms with pulse pile up for digital radiation spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 544, no. 3, pp. 668–678, 2005.

[67] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[68] J. Sousa, A. Combo, A. Batista, M. Correia, D. Trotman, J. Waterhouse, and C. Varandas, "A distributed real-time system for event-driven control and dynamic data acquisition on a fusion plasma experiment," *Fusion engineering and design*, vol. 48, no. 1-2, pp. 31–36, 2000.

[69] N. Cruz, J. Sousa, R. Pereira, Á. Combo, C. Varandas, M. Beldishevski, S. Dorling, B. Alper, *et al.*, "A large memory vme data acquisition system for the jet heterodyne radiometer upgrade," *Fusion engineering and design*, vol. 71, no. 1-4, pp. 167–173, 2004.

[70] A. Combo, R. Pereira, J. Sousa, N. Cruz, P. Carvalho, C. Varandas, S. Conroy, J. Källne, and M. Weiszflog, "A pci transient recorder module for the jet magnetic proton recoil neutron spectrometer," *Fusion engineering and design*, vol. 71, no. 1-4, pp. 151–157, 2004.

[71] S. Smith, *Digital signal processing: a practical guide for engineers and scientists*. Elsevier, 2013.

[72] A. C. Neto, F. Sartori, F. Piccolo, R. Vitelli, G. De Tommasi, L. Zabeo, A. Barbalace, H. Fernandes, D. F. Valcarcel, and A. J. N. Batista, "Marte: A multiplatform real-time framework," *IEEE Transactions on Nuclear Science*, vol. 57, no. 2, pp. 479–486, 2010.

[73] A. C. Neto, D. Alves, L. Boncagni, P. J. Carvalho, D. F. Valcarcel, A. Barbalace, G. De Tommasi, H. Fernandes, F. Sartori, E. Vitale, R. Vitelli, and L. Zabeo, "A survey of recent marte based systems," *IEEE Transactions on Nuclear Science*, vol. 58, no. 4, pp. 1482–1489, 2011.

[74] "Marte2 quality assurance - v1.0.0 documentation." `https://vcis.f4e.europa.eu/marte2-docs/master/html/overview.html`, 2018.

[75] A. C. Neto, F. Sartori, R. Vitelli, L. Capellà, G. Ferrò, I. Herrero, and H. Novella, "An agile quality assurance framework for the development of fusion real-time applications," in *2016 IEEE-NPSS Real Time Conference (RT)*, pp. 1–7, 2016.

[76] A. Sitek and A. M. Celler, "Limitations of poisson statistics in describing radioactive decay," *Physica Medica*, vol. 31, no. 8, pp. 1105–1107, 2015.

[77] L. Wielopolski and R. P. Gardner, "Prediction of the pulse-height spectral distortion caused by the peak pile-up effect," *Nuclear Instruments and Methods*, vol. 133, no. 2, pp. 303–309, 1976.

[78] S. Usman and A. Patil, "Radiation detector deadtime and pile up: A review of the status of science," *Nuclear Engineering and Technology*, vol. 50, no. 7, pp. 1006–1016, 2018.

[79] R. C. Pereira, A. M. Fernandes, A. Neto, J. Sousa, C. A. F. Varandas, J. Cardoso, C. M. B. A. Correia, M. Tardocchi, M. Nocente, G. Gorini, V. Kiptily, B. Syme, and M. Jennison, "Pulse analysis for gamma-ray diagnostics atca sub-systems of jet tokamak," *IEEE Transactions on Nuclear Science*, vol. 58, no. 4, pp. 1531–1537, 2011.

[80] R. M. Lindstrom and R. F. Fleming, "Dead time, pileup, and accurate gamma-ray spectrometry," 1995.

[81] R. Gardner and S. Lee, "Monte carlo simulation of pulse pile up," 1999.

[82] X. Wen and H. Yang, "Study on a digital pulse processing algorithm based on template-matching for high-throughput spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 784, pp. 269–273, 2015.

[83] M. Ruiz-Gonzalez, V. Bora, and L. R. Furenlid, "Maximum-likelihood estimation of scintillation pulse timing," *IEEE transactions on radiation and plasma medical sciences*, vol. 2, no. 1, pp. 1–6, 2017.

[84] M.-R. Mohammadian-Behbahani and S. Saramad, "A comparison study of the pile-up correction algorithms," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 951, p. 163013, 2020.

[85] V. Boronoyev and O. Rinchinov, "Methods of spline approximation in the problem of amplitude-time analysis of a pulse wave," *Radiophysics and quantum electronics*, vol. 41, no. 8, pp. 706–715, 1998.

[86] Y. Xiao-Feng, H. Hong-Quan, Z. Guo-Qiang, G. Liang-Quan, J. Kai-ming, G. Min, H. Chuan-Hao, and L. Mao-Lin, "Pulse pile-up correction by particle swarm optimization with double-layer parameter identification model in x-ray spectroscopy," *Journal of Signal Processing Systems*, vol. 94, no. 4, pp. 377–386, 2022.

[87] M. Astrain, M. Ruiz, A. V. Stephen, R. Sarwar, A. Carpeño, S. Esquembri, A. Murari, F. Belli, and M. Riva, "Real-time implementation of the neutron/gamma discrimination in an fpga-based daq mtca platform using a convolutional neural network," *IEEE Transactions on Nuclear Science*, 2021.

[88] E. Ronchi, P.-A. Söderström, J. Nyberg, E. A. Sundén, S. Conroy, G. Ericsson, C. Hellesen, M. G. Johnson, and M. Weiszflog, "An artificial neural network based neutron–gamma discrimination and pile-up rejection framework for the bc-501 liquid scintillation detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 610, no. 2, pp. 534–539, 2009.

[89] G. Fantini, A. Armatol, E. Armengaud, W. Armstrong, C. Augier, F. Avignone, O. Azzolini, A. Barabash, G. Bari, A. Barresi, *et al.*, "Machine learning techniques for pile-up rejection in cryogenic calorimeters," *Journal of Low Temperature Physics*, pp. 1–8, 2022.

[90] S. Marrone, D. Cano-Ott, N. Colonna, C. Domingo, F. Gramegna, E. Gonzalez, F. Gunsing, M. Heil, F. Käppeler, P. Mastinu, *et al.*, "Pulse shape analysis of liquid scintillators for neutron studies," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 490, no. 1-2, pp. 299–307, 2002.

[91] M.-R. Mohammadian-Behbahani and S. Saramad, "Pile-up correction algorithm based on successive integration for high count rate medical imaging and radiation spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 897, pp. 1–7, 2018.

[92] X. Wang, Q. Xie, Y. Chen, M. Niu, and P. Xiao, "Advantages of digitally sampling scintillation pulses in pileup processing in pet," *IEEE Transactions on Nuclear Science*, vol. 59, no. 3, pp. 498–506, 2012.

[93] X. Wen and H. Yang, "Study on a digital pulse processing algorithm based on template-matching for high-throughput spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 784, pp. 269–273, 2015. Symposium on Radiation Measurements and Applications 2014 (SORMA XV).

[94] T. Trigano, I. Gildin, and Y. Sepulcre, "Pileup correction algorithm using an iterated sparse reconstruction method," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1392–1395, 2015.

[95] W.-H. Wong, H. Li, J. Uribe, H. Baghaei, Y. Wang, and S. Yokoyama, "Feasibility of a high-speed gamma-camera design using the high-yield-pileup-event-recovery method," *Journal of Nuclear Medicine*, vol. 42, no. 4, pp. 624–632, 2001.

[96] M. Kafaee and S. Saramad, "Pile-up correction by genetic algorithm and artificial neural network," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 607, no. 3, pp. 652–658, 2009.

[97] P. Chandhran, K. E. Holbert, E. B. Johnson, C. Whitney, and S. M. Vogel, "Neutron and gamma ray discrimination for clyc using normalized cross correlation analysis," in *2014*

IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–8, IEEE, 2014.

[98] T. Alharbi, "Principal component analysis for pulse-shape discrimination of scintillation radiation detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 806, pp. 240–243, 2016.

[99] A. Caldwell, F. Cossavella, B. Majorovits, D. Palioselitis, and O. Volynets, "Signal recognition efficiencies of artificial neural-network pulse-shape discrimination in hpge 0-decay searches," *The European Physical Journal C*, vol. 75, no. 7, pp. 1–9, 2015.

[100] C. Sosa, M. Flaska, and S. Pozzi, "Comparison of analog and digital pulse-shape-discrimination systems," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 826, pp. 72–79, 2016.

[101] M. Roush, M. Wilson, and W. Hornyak, "Pulse shape discrimination," *Nuclear Instruments and Methods*, vol. 31, no. 1, pp. 112–124, 1964.

[102] M. J. Balmer, K. A. Gamage, and G. C. Taylor, "Comparative analysis of pulse shape discrimination methods in a 6li loaded plastic scintillator," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 788, pp. 146–153, 2015.

[103] R. Pereira, A. Fernandes, N. Cruz, J. Sousa, M. Riva, D. Marocco, F. Belli, and B. Gonçalves, "Neutron/gamma discrimination code based on trapezoidal filter," *Fusion Engineering and Design*, vol. 134, pp. 118–122, 2018.

[104] M. Aspinall, B. D'Mellow, R. Mackin, M. Joyce, N. Hawkes, D. Thomas, Z. Jarrah, A. Peyton, P. Nolan, and A. Boston, "Verification of the digital discrimination of neutrons and γ rays using pulse gradient analysis by digital measurement of time of flight," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 583, no. 2-3, pp. 432–438, 2007.

[105] C. Fu, A. Di Fulvio, S. Clarke, D. Wentzloff, S. Pozzi, and H. Kim, "Artificial neural network algorithms for pulse shape discrimination and recovery of piled-up pulses in organic scintillators," *Annals of Nuclear Energy*, vol. 120, pp. 410–421, 2018.

[106] X. Luo, V. Modamio, J. Nyberg, J. Valiente-Dobon, Q. Nishada, G. De Angelis, J. Agramunt, F. Egea, M. N. Erduran, S. Ertürk, *et al.*, "Pulse pile-up identification and reconstruction for liquid scintillator based neutron detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 897, pp. 59–65, 2018.

[107] S. A. Pozzi, M. Flaska, A. Enqvist, and I. Pazsit, "Monte carlo and analytical models of neutron detection with organic scintillation detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 582, no. 2, pp. 629–637, 2007.

[108] N. Patronis, M. Kokkoris, D. Giantsoudi, G. Perdikakis, C. Papadopoulos, and R. Vlastou, "Aspects of geant4 monte-carlo calculations of the bc501a neutron detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 578, no. 1, pp. 351–355, 2007.

[109] R. Pereira, N. Cruz, A. Fernandes, J. Sousa, B. Gonçalves, M. Riva, C. Centioli, D. Marocco, S. Podda, B. Esposito, *et al.*, "Real-time data acquisition and processing system design for the iter radial neutron camera," in *Proc. Sci. 1st EPS Conf. Plasma Diagnostics*, 2015.

[110] Y. Koechlin and A. Raviart, "Analyse par echantillonnage sur photons individuels des liquides fluorescents dans le domaine de la sub-nanoseconde," *Nuclear Instruments and Methods*, vol. 29, no. 1, pp. 45–53, 1964.

[111] K. Sigman, "Professor karl sigman's lecture notes on simulation," 2012.

[112] R. Gallager, "Discrete stochastic processes - chapter 2: Poisson processes," 2011.

[113] H. Kim, G. Cho, and H. Kim, "Characteristics of a stilbene scintillation crystal in a neutron spectrometer," *Radiation Measurements*, vol. 58, pp. 133–137, 2013.

[114] M. Moszyński, "Light pulse shape study from crystal organic scintillators," *Nuclear Instruments and Methods*, vol. 153, no. 2-3, pp. 439–443, 1978.

[115] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder–mead simplex method in low dimensions," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[116] S. E. Derenzo, W.-S. Choong, and W. W. Moses, "Fundamental limits of scintillation detector timing precision," *Physics in Medicine Biology*, vol. 59, p. 3261, may 2014.

[117] M. Magdowski and R. Vick, "Estimation of the mathematical parameters of double-exponential pulses using the nelder–mead algorithm," *IEEE transactions on electromagnetic compatibility*, vol. 52, no. 4, pp. 1060–1062, 2010.