Neutron Calibration and Characterisation of the DEAP-3600 Experiment Using a 74MBq AmBe Neutron Source

Franco La Zia

Department of Physics
Royal Holloway, University of London

A thesis submitted to the University of London for the degree of Doctor of Philosophy

December 2018

Declaration

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the document.

Signed:

Date: 02/01/2019

For my parents Francesco and Maristella, and my sister Angela.

Abstract

The existence of dark matter was first proposed by Fritz Zwicky in the 1930's to explain the motions of galaxies in the Coma cluster. Since then the evidence for dark matter has mounted, however its nature has remained illusive. One of the leading dark matter candidates is the Weakly Interacting Massive Particle (WIMP). The DEAP-3600 experiment is a single-phase liquid argon detector located 2 km underground at SNOLAB, Sudbury, Ontario, Canada which aims to measure WIMP interactions with a sensitivity of 10^{-46} cm^2 for a 100 GeV/c^2 WIMP mass. To detect WIMPs liquid argon detectors look for elastic scatters, using pulse shape discrimination (PSD) to distinguish between nuclear (WIMP-like) and electronic (background-like) recoils. Characterisation of PSD is therefore of paramount importance to achieve a low energy WIMP search threshold resulting in increased sensitivity. DEAP-3600 makes use of a 74 MBq AmBe neutron source to populate the detector with nuclear recoils, allowing for calibration of the detector to WIMP-like nuclear recoils. The goal of this thesis is to characterise DEAP-3600 response to nuclear recoils and to determine the WIMP nuclear recoil acceptance using the AmBe source. This is non-trivial due both to the higher cross-section of neutron scatters on liquid argon compared with WIMP interactions as well as the geometry of the detector relative to the source deployment position, as the neutrons will multiple scatters as they traverse the detector. This work presents a method to isolate single scatters (WIMP-like) in the AmBe data by using machine learning algorithms trained on AmBe Monte Carlo simulations. Due to the complexity of neutron interactions a method for simulation to achieve the statistics required for this analysis was developed. The result of this thesis is a determination of the nuclear recoil acceptance in DEAP-3600.

Acknowledgements

First I would like to thank Jocelyn Monroe, my advisor, for having supported me in the last four years. You will always be a source of inspiration for me and I am very grateful for having had the privilege of working with you.

I would also like to thank all my friends at Royal Holloway. Thank you Alistair Butcher, for making me feel welcomed in the Dark Matter group even since the day of my interview and for being an invaluable source of help and advices as a senior PhD student first and Postdoc researcher later. Thank you to Navin Seeburn, for being the first roommate ever not complaining about my snoring. I will not forget your kindness and willingness to help me out debugging a piece of code even at 4 AM in the morning. A special thanks to Nasim Fatemighomi, I have learned a lot from working closely with you. Your commitment and determination are inspiring and I'm glad to have found a friend and a caring listener in you. Ashlea Kemp, I would like to sincerely thank you for having brought the joy and enthusiasm that only a very smart and talented young physicist as you are could have brought in our group and for being a gorgeous friend. Thank you Asher Kaboth for your always helpful comments in our endless group meetings. Thank you Alberto, Gabriela, Emily, Zac, Will P. and S., Andrey, Adriana, Flavia and all the wonderful people that I have had the pleasure of working with or simply sharing good time with in the last four year at Royal Holloway.

I would also like to thank all the collaborators and friends at SNOLAB. Thank you in particular to Mark Ward, for having taught me how to keep calm when things go wrong and having driven me around Sudbury for months, to Mark Kos for being a good friend and a perfect night shift companion, and to Tony Flower and the beloved Tracy McGinn for the great time under and above surface when in Sudbury.

Further along, I want to thank everyone that has made it possible for me to feel at home and has made me feel like belonging to an enlarged and happy family in Egham over these past years. You will forgive me if I am just listing your names, but a book would not be enough to express my gratitude to you all and mention why I should thank each one of you. Seb, Jorge, Ana, Stef, Marianne, Humarrah, Flo, Eugene I and II, Ciara, Roberto, Giulia, Andrea, Mateo, Hannah, Alessandra, Valeria, Lavinia and many others that I am shamefully not mentioning, my deepest gratitude for having been so nice to me. Each one of you has given me something that I will never forget.

Special mention to Lorenzo and Teresa, and you both know why.

I would also like to thank Giuseppe, Salvatore and Giovanni for having been and being the great friends that they are for more than two decades now.

Thanks to my parents, my gratitude to you for what you have done for me is beyond words and I love you. To my sister Angela for always having been there for me and Piero for being the wise big brother I never had, but above all thank you little Vittoria: you are the greatest joy of my life and I hope I will be a good uncle to you.

Last but not least, I would like to thank Joseph Walding, my supervisor. You have been an incredible source of guidance to me and I will always be grateful to you. Thanks for your patience, for having been there at any hour of the day or day of the week when I needed support, for having bought me more whiskey than I deserved after an intense week underground and for having introduced me to the weirdest pizza ever: Cheesburger Za... no pickles, no onion etc...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A consistent number of astronomical and cosmological observations over the last 80 years has suggested that our universe is mainly made of an unknown substance referred to as dark matter. The idea of dark matter was first introduced by Fritz Zwicky in 1933 [1] [2]. A confirmation to Zwicky's hypothesis came in 1970 from Vera Rubin's studies on the rotational velocity of spiral galaxies. By studying the Doppler shift as a function of galactic radius in the Andromeda Nebula, contrary to what one would expect from Newtonian dynamics, she observed that stars further away from the center were moving as fast as those closer in [3]. The explanation to this observation is the existence of a dark matter halo which dominates the gravitational field in the outer regions of the galaxy. Today we know that dark matter constitutes approximately 26% of the total matter in the universe but its nature is still to be determined [4].

1.1 Indirect and Observational Evidence for Dark Matter

Studies of rotational curves in galaxies, gravitational lensing and investigation of the Cosmic Microwave Background (CMB) have provided over the last century very strong indirect evidence for Dark Matter. This section provides a review of the studies and experiments that motivate the search for dark matter.

1.1.1 Rotational Curves of Galaxies

In 1933, Fritz Zwicky observed some anomalies in the velocity dispersion in the Coma cluster to what was expected given the known amount of luminous mass of those galaxies. In his paper *On the masses of nebulae and of clusters of nebulae* published in 1937 [2], which

is a refined and extended analysis of the Coma cluster analysis presented four years earlier, Zwicky states that masses of nebulae estimation either from the luminosities of nebulae or from their internal rotations are unreliable. The mass-luminosity relation, requires a knowledge of the conversion factor from absolute luminosity to mass for all different types of stars in the nebulae. Zwicky, since his first work, instead applied the virial theorem to the Coma cluster in order to estimate its mass. The theorem relates the mean kinetic energy $< T >$ in the centre of mass frame and the mean potential energy $< U >$ according to:

$$< T >= -\frac{1}{2} < U > \tag{1.1}$$

Zwicky's observations using the virial theorem showed an excess in the measured mass compared to the mass calculated from the light emitted from the stars constituting the cluster. In particular, he found that the average density in the cluster would have to be 400 times greater than what is derived from luminous matter to get the large observed average velocity dispersion along the line-of-sight of approximately 1000 km/s [1]. This led him to conclude that most of the missing matter in the cluster was some form of non-luminous (dark) matter. A confirmation to Zwicky's hypothesis came in 1970 from Vera Rubin's studies on the rotational velocity of spiral galaxies. By studying the Doppler shift as a function of galactic radius in the Andromeda Nebula, contrary to what one would expect from Newtonian dynamics, she observed that stars further away from the centre were moving as fast as those closer in [3]. In fact, if we assume Newtonian dynamics is correct , the velocity distribution should follow:

$$(\frac{m \cdot v^2(r)}{r}) = (\frac{m \cdot M(< r) \cdot G}{r^2}), \tag{1.2}$$

where m is the mass of the star, $v(r)$ is the rotational velocity, r the radial distance of the star from the galactic centre, $M(< r)$ is the total mass within the radius r and G the gravitational constant, which gives for $v(r)$:

$$v(r) = \sqrt{\frac{M(r) \cdot G}{r}}. \tag{1.3}$$

For stars inside the galactic disk, $M(< r) \propto r^3$, therefore from 1.3 $v \propto r$. For distances beyond the galactic disk ($r > \mathrm{R_{disk}}$), M $\sim$ constant (Gauss' Law) therefore again from equation 1.3 $v(r) \propto r^{-1/2}$ as predicted by Netwon's laws of gravity. Hence, the velocity

of stars at small r should increase whilst the velocity of far away from the galactic centre (large r) should decrease. However, what observations show instead, is that the rotational velocity flattens out at distances $r \gtrsim \mathrm{R_{disk}}$ implying that $M(r)$ is proportional to r as shown in Fig.1.1.

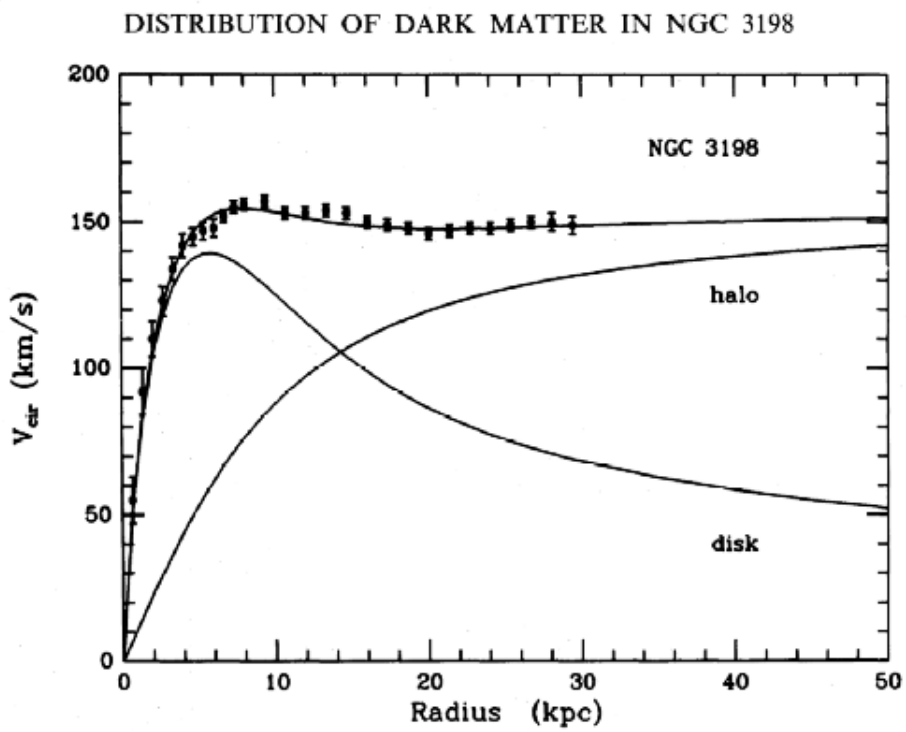


Figure 1.1: Rotation curve for the NGC 3198 galaxy. The plot includes the actual velocities of the outer stars (points) together with the expected velocity given the mass of visible matter in the disk and the contribution of the dark matter halo. [5].

1.1.2 Gravitational Lensing

Another piece of evidence for the existence of dark matter comes from gravitational lensing studies of the mass distribution in galaxy clusters [6], an example of which is the study of a pair of merging clusters known as the Bullet Cluster. Gravitational lensing occurs when light gets deflected by a massive object that stands between the light source emitter and the observer. Predicted by Albert Einstein's theory of general relativity, gravitational lensing can provide important information about the amount and location of dark matter. If M is the mass of the lensing massive object, the deflection factor α is given by [6]:

$$\alpha = -\frac{4GM}{c^2 b}, \tag{1.4}$$

where b is the distance of closest approach, G the gravitational constant and c the speed of light.

This allows an indirect measurement of the mass and location of the object via observation of the light "distortion". By combining the optical hot X-ray emission gas - which makes up most of the baryonic mass of the cluster - and gravitational lensing observations, it was revealed that the hot gas (coloured map in Fig.1.3) that gets slowed down by interactions in the collision stays predominantly around the centre of the collision whereas it was observed that most of the mass passes through unruffled during the collision and is located at the edges of the cluster (green contours Fig.1.3).

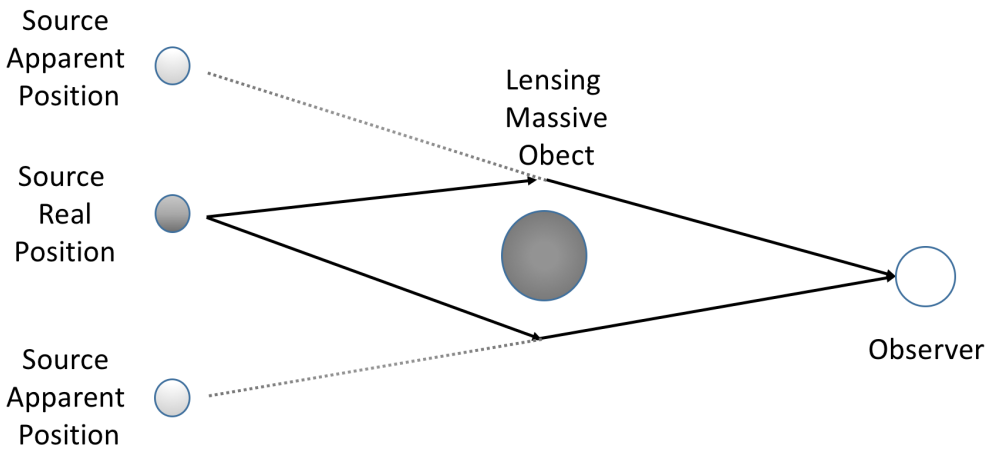


Figure 1.2: Schematic representation of gravitational lensing. The gravitational field of the massive object between the source and observer position causes the bending of light rays from the source. The resultant source image appears distorted into two or more distinct images, sweeping arcs or a complete ring depending on the alignment between the objects involved and the structure of the lensing object.

This is a clear evidence that most of the dark matter should be non-baryonic in nature and represents the first example of an observed difference between dark matter and baryonic matter [8].

1.1.3 Cosmic Microwave Background Cosmic

Observations of the Cosmic Microwave Background (CMB) is another strong piece of evidence for the existence of dark matter on cosmological scales. The CMB is a relic radiation from an early stage of the universe. The CMB emission spectrum agrees precisely with a black body spectrum and is uniform to one part in 10^5. Many space missions have

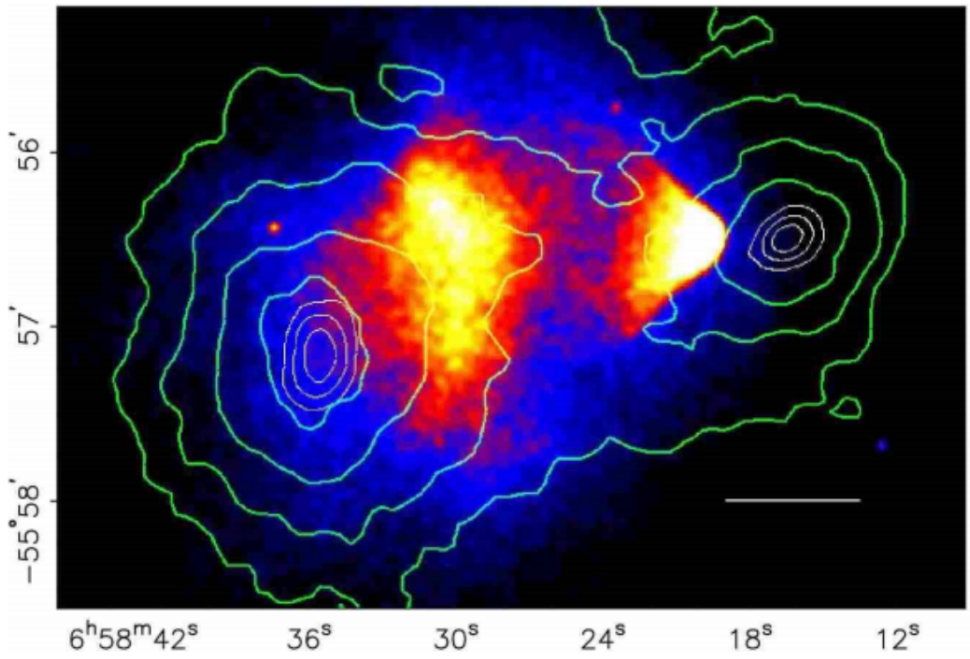


Figure 1.3: Image of the bullet cluster. The Chandra X-ray image of the cluster collision is shown by the coloured map. Gravitational lensing observation from Hubble infers the green contours that are proportional to the projected mass in the system [7].

measured this spectrum with very high precision, revealing that small anisotropies in the photon temperature reveal very fundamental informations about the genesis of the universe's evolution and structure. These anisotropies can be studied by looking at the power spectrum of the temperature fluctuations which can be expanded in terms of spherical harmonics. In Fig.1.5 the magnitude of these fluctuations is shown as a function of the spherical harmonics coefficients I, where I is inversely proportional to the angular scale.

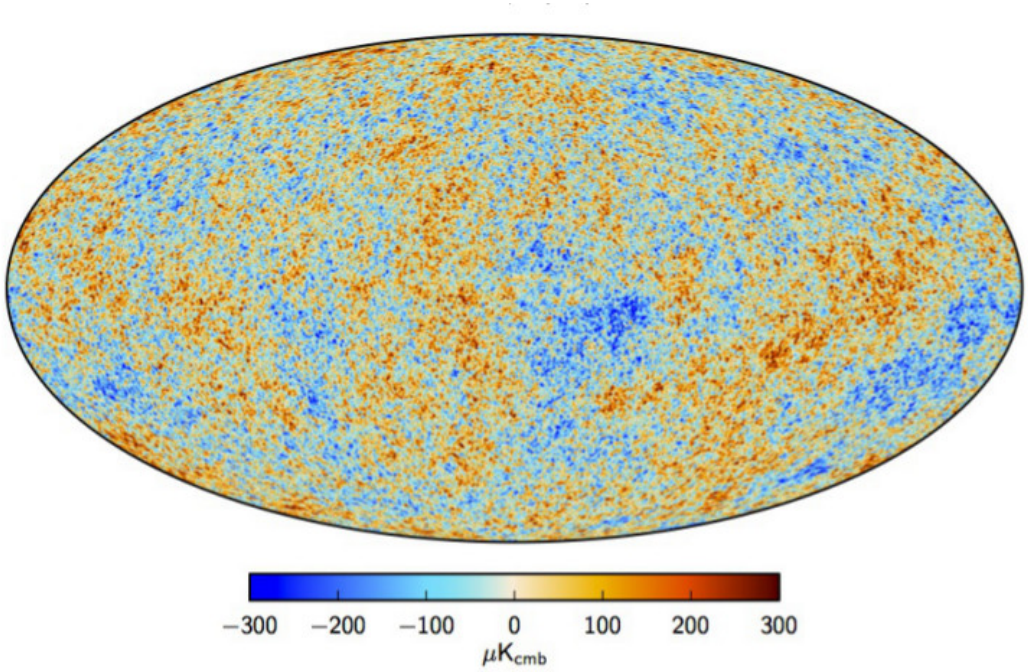


Figure 1.4: Map of the temperature fluctuations of the CMB measured by Planck [4].

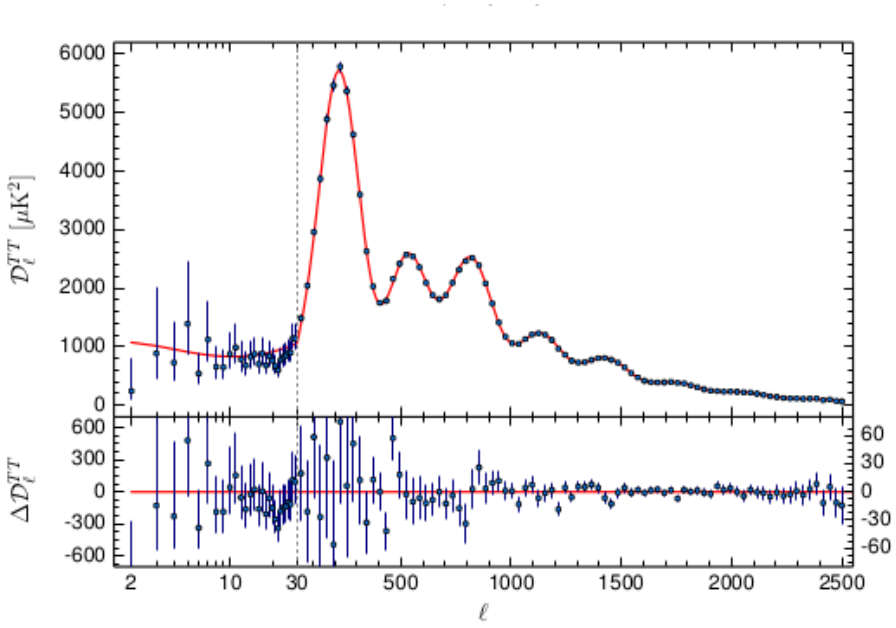


Figure 1.5: Measured CMB temperature fluctuations as a function of multipoles from the Planck 2015 results and residual between the power spectrum and the ΛCDM model [4].

The power spectrum is fitted with the standard model of cosmology ΛCDM (flat universe with dark energy and cold dark matter) and residuals between the model and data from the Planck 2015 experiment are also shown in the plot. The latest results from the Planck collaboration report the contribution of baryonic matter to be

$\Omega_b h^2 = 0.02226 \pm 0.00023$ and $\Omega_{nbm} h^2 = 0.1186 \pm 0.0020$ for the non-baryonic matter (dark matter contribution), with h being the Hubble constant in units of 100 km/s/Mpc. Ω, the density parameter, is defined as the ratio between the mass density ρ_x and the critical density ρ_c for a flat universe ($\Omega_x = \frac{\rho_x}{\rho_c}$). Another constraint on the value of Ω_b comes from Big Bang nucleosynthesis (BBN) theory. According to BBN, light elements like ^{4}He, ^{4}H, ^{3}He and ^{7}Li, started forming as the universe expanded and it's temperature dropped below the relevant nuclear binding energies. BBN took place in the first $t \sim 100\ s$ of the universe when the temperature went from $T \sim 10\ MeV$ to $T \sim 0.1\ MeV$. When the temperature was $T > 1\ MeV$, the universe was made of a plasma containing relativistic particles such as $e^{\pm}$, ν_i, $\bar{\nu_i}$ and γ, and non-relativistic like n and p that were at equilibrium. The ratio of neutrons to protons, governed by the Boltzmann law was then:

$$\frac{n_n}{n_p} \simeq (\frac{m_n}{m_p})^{\frac{3}{2}} e^{-\frac{Q}{k_b T}} = 1, \tag{1.5}$$

but as the temperature dropped, the weak interaction rate $\Gamma_{n \leftrightarrow p} \sim G_F^2 T^5$ fell faster than the expansion rate of the universe $H \sim \sqrt{g_* G_N} T^2$ until the temperature $T \sim 1\ MeV$ was reached and $\Gamma_{n \leftrightarrow p}$ became slower than H. This moment is known as "neutron freeze-out" and fixed the n/p ratio at $\frac{n_n}{n_p} = \frac{1}{6}$. After freeze-out neutrons can still β-decay, so this fraction drops to $\frac{n_n}{n_p} = \frac{1}{7}$ by the time nuclear reactions began. Neutrons that did not decay initiated the nuclear reaction chain that brought about the synthesis of the light elements. Their abundance depends on the ratio of baryons to photons $\eta = \frac{n_b}{n_\gamma}$. Estimating these abundances results in determining the allowed range of the parameter η which predicts the value of Ω_b.

The elemental abundance is shown in Fig.1.6 as a function of η calculated by R.H. Cyburt et al. [10], together with the observational inferred primordial abundances represented with the boxes. CMB measurements of $\Omega_b h^2$ are consistent with these abundances.

This and the previous piece of indirect evidence presented in this section, point towards a non-baryonic particle as a candidate for dark matter.

1.2 Dark Matter Candidates

Some properties of dark matter can be inferred from the evidence presented in section 1.1 . For example, dark matter particle candidates must be electrically neutral and should be weakly interacting with ordinary matter. Non-baryonic dark matter candidates must be

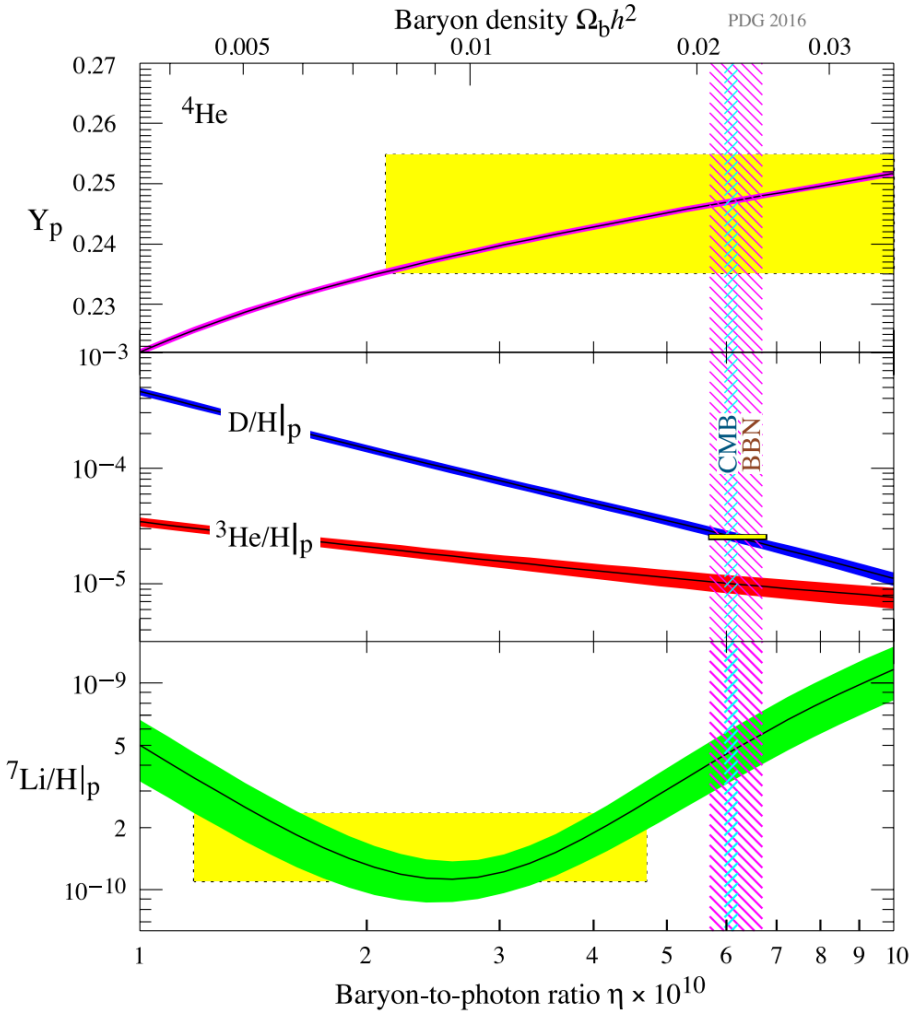


Figure 1.6: Aboundances of ^{4}He, D, ^{3}He and ^{7}Li measured (boxes) and calculated as functions of baryon-to-photon ratio η. The narrow light blue vertical band indicates the CMB measure of the cosmic baryon density, the wider magenta band indicates the BBN concordance range (both at 95% CL). [9]

stable on cosmological time scales, have very weak interactions with electromagnetic radiation and have the right relic density. Many theoretical models, and particles, have been proposed over the last few decades as candidates. Some of the most common candidates are axions, sterile neutrinos, primordial black holes, and weakly interacting massive particles (WIMPs). In what follows, a brief description of those models and particles towards which experimentalists are mainly focusing their attention for discovery will be given.

1.2.1 Axions

Axions were originally postulated in 1977 by Peccei and Quinn as a solution to the strong CP problem in quantum chromodynamics (QCD). If axions exist, they are a dark matter particle candidate. The strong CP problem arises from the experimental observation that the electric dipole moment (EDM) of the neutron is extremely small ($|d_n| < 2.9 \times 10^{-26}$ $e \cdot cm$) compared to the QCD prediction ($|d_n| \sim 10^{-16}\bar{\theta}\ e \cdot cm$). $\bar{\theta}$ is the effective strong CP violating term which is a SM input not predicted by the model. The limit on the EDM measurement sets this term to be $\bar{\theta} < 10^{-9}$. An elegant solution to this problem was the one introduced by Peccei and Quinn. Their ideas was to add a new global symmetry ($U(1)_{PQ}$) to QCD, that becomes spontaneously broken introducing a Nambu-Goldstone boson of the $U(1)_{PQ}$ symmetry, the axion. This particle would not only solve the strong CP problem, but it would have the right properties to be a dark matter particle candidate. In fact it is a non-baryonic, very weakly interacting and stable particle, that could have been produced as a Bose-Einstein condensate in the early stage of the universe, and if it exists in a low mass range between μeV and meV, can account for the correct relic density.

1.2.2 MACHO

Baryonic matter candidates such as MACHOs have been proposed as components of dark matter as well. MACHO stands for MAssive Compact Halo Objects. These objects could be brown dwarfs, black holes and/or neutron stars, each of which emit little or no light. Their masses can be more than 70 orders of magnitude greater than axions highlighting the large mass range for dark matter candidates. MACHOs constitute the main baryonic candidates and their presence can be "observed" though the microlensing effect. The MACHO, EROS and OGLE collaborations are the main groups that have set limits on the total dark matter component due to MACHOS [11]. EROS-2 excluded MACHOs as the dominant form of dark matter in our galaxy for MACHOs with masses in the

range 0.6×10^{-7} to $15M_{\odot}$. The MACHO collaboration concluded that MACHOs cannot contribute more than 40% to the mass of the galactic halo in our galaxy. These results confirm that there must be a non-baryonic component contributing.

1.2.3 WIMPs

Weakly Interactive Massive Particles (WIMPs) χ, are the favoured hypothesis for dark matter particles. Their mass ranges from 10 GeV to a few TeV with weak scale cross-section. The hypothesis that they could be massive neutrinos, has been conclusively disproved by experiments at the LEP electron-positron collider at CERN. However, Supersymmetry (SUSY), an extension to the Standard Model theory, postulates the existence of supersymmetric particles that could be WIMP candidates. SUSY particles are created in pairs wth opposite values of R-parity (R = $\pm$1), a conserved quantum number. In SUSY, the lightest supersymmetric particles (LSP) result from heavy SUSY particles decaying into the lightest stable particles conserving R-parity. Thus they survived from the primordial era of the universe. In many SUSY models the LSP is called the *neutralino*, which is a neutral fermion linear combination of the photino, zino and two higgsino, these are the SUSY partners of the photon, Z^0 and Higgs boson respectively. The main reason why WIMP models are attractive is that it can be shown that they produce the correct dark matter abundance observed today for WIMP masses of the order of 100 GeV and with interaction cross sections at the weak scale. Assuming that a WIMP particle χ exists, the number density of these particles in thermal equilibrium is [12]:

$$n_{\chi}^{eq} = \frac{g}{(2\pi)^3} \int f(\mathbf{p}) d^3\mathbf{p}, \tag{1.6}$$

with g number of internal degrees of freedom of the particle and ($\mathbf{p}$) the Bose-Einstein or the Fermi-Dirac distribution. When the temperature T of the universe was high ($T \gg m_{\chi}$), dark matter particles were relativistic and $n_{eq} \propto T^3$. As the universe cooled down ($T \ll m_{\chi}$), dark matter particles became non relativistic and their density was Boltzmann suppressed by a factor $e^{(-m_{\chi}/T)}$ ($n_{eq} \simeq g(m_{\chi}T/2\pi)^{3/2}e^{(-m_{\chi}/T)}$). If the universe was static and in thermal equilibrium all the time, this exponential suppression would basically lead to a universe with no WIMPs today. The reason why a relic abundance remains today is because at some point the rate of $\chi\bar{\chi}$ annihilation $\Gamma =< \sigma_A v > n_{\chi}$, which was very high when $T \gg m_{\chi}$, dropped below the expansion rate H of the Universe when T dropped

below m_χ, resulting in the freeze out of the WIMPs. The time evolution of the number density of WIMPs n_χ is quantitatively described by the Boltzmann rate equation

$$\frac{dn_\chi}{dt} + 3Hn_\chi = - < \sigma_A v > [(n_\chi)^2 - (n_{eq}^2)^2], \tag{1.7}$$

where H is the Hubble constant and $\sigma_A v$ is the thermally averaged annihilation cross section of $\chi\bar{\chi}$ times the relative velocity v.

This shows how the relic density of a WIMP model ($\Omega_\chi h^2$), depends on the proposed $\sigma_A v$ for that model. An energy independent approximation of the relic density can be derived as shown in [12] as:

$$\Omega_\chi h^2 = \frac{3 \times 10^{-27} cm^3 s^{-1}}{< \sigma_A v >}, \tag{1.8}$$

which is independent from the WIMP mass but inversely proportional to its total annihilation cross section. At this point the question why WIMP models are the favoured hypothesis can be answered. Having found the relic density abundance as in Eq. 1.8, if a stable new particle χ with weak-scale interactions exists, its annihilation cross section can be estimated to be $< \sigma_A v > \sim \alpha^2 (100\ GeV)^2 \sim 10^{-25}\ cm^3 s^{-1}$. This leads to a value for the relic abundance of dark matter which is incredibly close to the observed one.

1.3 Dark Matter Direct Detection Search

Physicists have tried to detect dark matter particles for many decays using different techniques and technologies. Given the nature of the DEAP-3600 experiment, the subject of this thesis, a detailed discussion of direct detection dark matter searches will be given in this section. However, it is worth mentioning searches are also undertaken at large particle colliders such as the LHC as well as indirect detection searches with satellite based detectors (PAMELA [14], AMS [15]) or terrestrial large scale Cherenkov detectors (like IceCube [16]). In the case of searches at accelerators, the idea is that colliding high energy particles (TeV scale), leads to the production of dark matter particles that can be directly observed in the detector as missing energy. Indirect detection relies on the possibility of detecting standard model particles that have been generated as a consequence of dark matter annihilation or decay. Finally, in the direct detection scenario, a signal is expected to be generated by the scattering of dark matter candidate particles with nuclei of the target

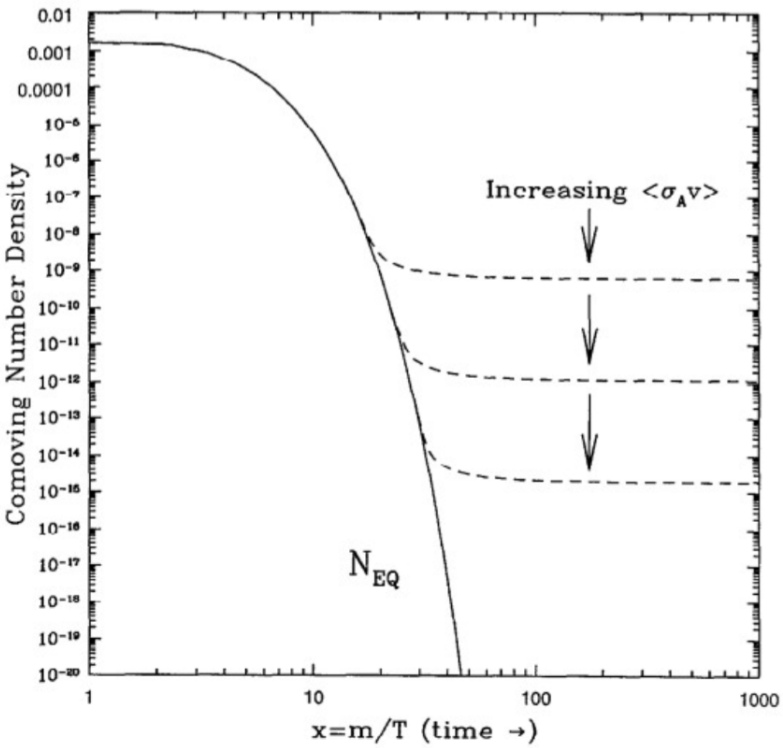


Figure 1.7: Co-moving number density of a non relativistic WIMP particle in the early universe. The solid curve is the equilibrium abundance while the dashed curves are the actual abundance [13].

material. For WIMP-nucleus scattering experiments like DEAP-3600, a nuclear recoil is expected and the differential scattering rate can be expressed as [17]:

$$\frac{dR}{dE}(E,t) = \frac{\rho_0}{m_\chi m_A} \int v f(\mathbf{v},t) \frac{d\sigma}{dE}(E,v) d^3v, \tag{1.9}$$

where v is the dark matter velocity defined in the rest frame of the detector, m_A is the nucleus mass of the experiment target material, m_χ the dark matter particle mass and $\frac{d\sigma}{dE}(E,v)$ its differential cross section, ρ_0 is the local dark matter density and $f(\mathbf{v},t)$ is the WIMP velocity distribution in the detector reference frame, which is time dependent because of the revolution of the Earth around the Sun. The two parameters extracted of a direct detection dark matter experiment are the mass m_χ and the WIMP cross section σ.

Exploiting the energy dependence of Eq. 1.9 and following [17], the rate equation can be approximated as:

$$\frac{dR}{dE}(E) \approx (\frac{dR}{dE})_0 F^2(E) exp(-\frac{E}{E_c}), \tag{1.10}$$

where E_c is a constant that parameterise the energy scale which depends on the target nucleus and WIMP mass, $(\frac{dR}{dE})_0$ indicates the rate at zero momentum transfer. $F^2(E)$ is the nuclear form factor correction, in fact, it needs to be taken into account that the scatter occurs on the whole nucleus. WIMP interactions can be spin-dependent (SD) or spin-independent (SI), with the former only possible if the target nuclei have a net spin. Mediums like argon are not sensitive to SD interactions as their number of protons and neutrons is even, thus DEAP-3600 is only sensitive to SI interactions. $\frac{d\sigma}{dE}(E,v)$ from Eq. 1.9 can then be written as:

$$\frac{d\sigma}{dE} = \frac{m_A}{2\mu_A^2 v^2} \cdot (\sigma_0^{SI} \cdot F_{SI}^2(E) \cdot + \sigma_0^{SD} \cdot F_{SD}^2(E)), \tag{1.11}$$

where $\mu_a = \frac{m_\chi m_A}{m_\chi + m_A}$ is the WIMP-nucleus reduced mass.

Focusing on the SI part only given the nature of this thesis (SD interactions can be found here [17]), and assuming a similar strength of WIMPs coupling to neutrons and protons, the SI cross section can be written as:

$$\sigma_0^{SI} = \sigma_{SI} \frac{\mu_A^2}{\mu_n^2} A^2, \tag{1.12}$$

where μ_n is the reduced mass of the WIMP-nucleon system and A is the atomic mass of

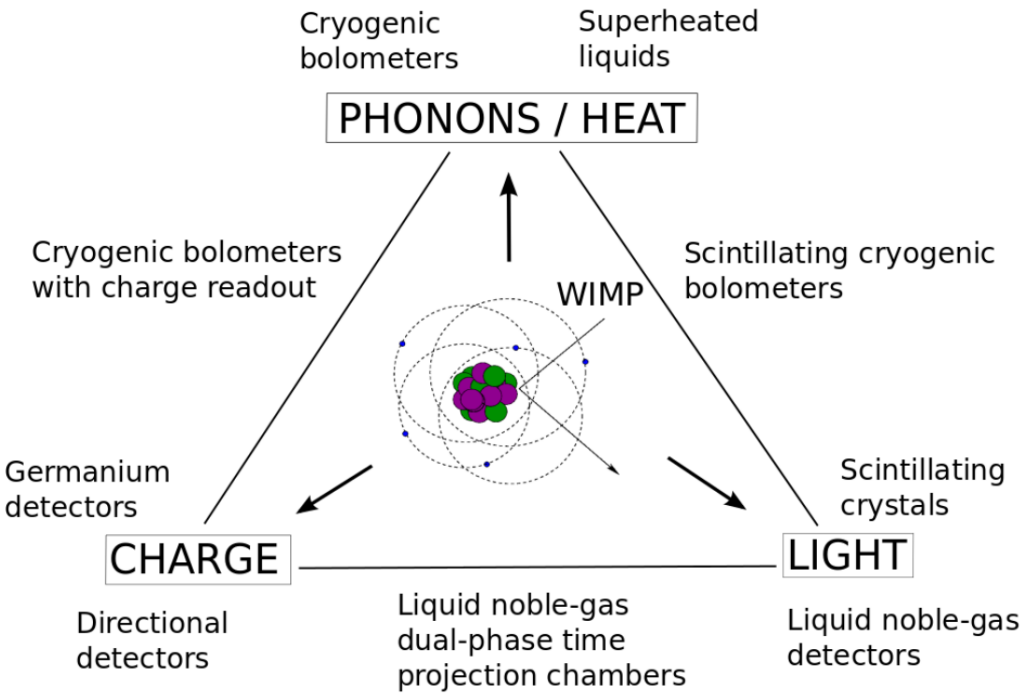


Figure 1.8: Direct detection experiments schematic of possible signals that can be measured independent on the technology used for detection [17].

the target nucleus. σ_{SI} can then be used to compare experimental results from different experiments or to theory. In the rate calculation one usually assumes a local WIMP density of $\rho_\chi = 0.3\ GeVcm^{-3}$, mean WIMP velocity of 230 kms^{-1}, galactic escape velocity of 600 kms^{-1}, and the WIMP spectrum, in the case of zero momentum transfer, follows from elastic scattering calculations assuming that the velocity distribution of WIMPs is Maxwellian.

1.4 Direct Search technologies

Detecting dark matter particles is a very challenging task. A huge effort has been made over the last few decades in developing technologies to pursue this goal, many different experiments are currently competing around the world and very exciting new ones have been proposed for the decades ahead. In this section, an overview of the current state-of-the-art of direct search experimental strategies and techniques will be given and some of the future plans in the field will be briefly discussed.

As discussed in the previous section, a WIMP direct search experiment looks for

"some signal" from an elastic scatter of a WIMP on a target material atom. A detailed discussion of liquid argon scintillation physics will be given in the next chapter. Scintillation light collection is one of a few ways to detect the occurrence of a scatter. One can also read out the heat or charge generated by a recoiling nucleus. Depending on what the experiment is aiming to read, different strategies, material and techniques are used to build a direct search experiment. Fig.1.8 offers a schematic representation of the different possibilities. Most direct WIMP search experiments utilise either one of these three signals or a combination of them. Below, a brief review of the working principles of these different categories is given:

- **Acoustic detection:** A development of the bubble chamber technology using superheated fluids is used in this category with the PICO experiment [18] at SNOLAB being an example of this technique. PICO uses C_3F_8 as a target material and captures bubbles with fast cameras and piezo sensors. It is equipped with acoustic transducers that can read the acoustic emissions from bubble nucleations. The acoustic power released in an event can then be measured and it can be used to place background rejection cuts together with analysis from bubble information such as their number and size.

- **Phonons detection:** Heat (phonons) is produced as a results of particle interacting with atoms in a crystal absorber. A device that collects a phonon signal produced in a crystal is known as a "cryogenic bolometer" and a schematic of such a detector for dark matter detection purposes is shown in Fig.1.9. Energy deposited by a nuclear recoil that creates both thermal and non-thermal phonons in the crystal absorber, determines a temperature rise ΔT which can be measured by a thermal sensor.

- **Phonons/Charge detection:** Experiments in this category collect both phonons and charge generated by recoiling nucleii to allow for particle discrimination. CDMS [19] and EDELWEISS [20] are examples of experiments that read out both these signals and are both Germanium detectors. A characteristic of using this detection strategy is that the energy threshold can be set at very low values ((3-14) keV_{nr} in CDMS II, [17], [21]) allowing the exploration of the low mass WIMP region.

- **Phonons/light detection:** Another possibility is to combine phonons and light collections. The second phase of the CRESST experiment [22] is an example of this

strategy. Detectors using this technique are very competitive in the low WIMP mass search.

- **Charge detection:** Detectors that collect only charges are generally germanium detectors or directional gas time projection chambers (TPCs). Germanium detectors operated in ionisation mode are not able to discriminate between particle types, but they can discriminate between surface events and events that occur in the centre of the medium by measuring the rise time of the charge readout. Their threshold is very low allowing light WIMP mass searches (order 1 GeV). The CoGENT experiment, for example, uses p-type point contact germanium detectors reaching an energy threshold of 500 eV_{ee} [23]. An example of a direction gas TPC can be seen in Fig.1.11 where a schematic of the DRIFT-II experiment TPC is shown [24]. The figure shows how a NR track is reconstructed using this technology: charges are drifted through an electric field and read-out in the anode wires plane.

- **Charge/light detection:** This strategy is used by experiments that are currently leading the dark matter limit set (at mass higher than 10 GeV) such as Xenon1T [25], PandaX [26], LUX [27].dua Experiments that collect both charge and light are referred to as dual-phase detectors. These detectors are TPCs where charges are collected through an electric field in a similar fashion as described for DRIFT-II above, and are equipped with layers of photomultipliers to also collect light from events. This technique ensures an efficient discrimination between electronic and nuclear recoils and is one of the reasons why it has been adopted by a large number of collaborations searching for dark matter.

- **Light detection:** These are detectors referred to as single-phase detectors and they only collect scintillation light generated by recoiling nucleii. This light is generally read by photomultipliers tubes and the discrimination between electrons and nuclear recoils relies on properties of the target material. DEAP-3600 is a single-phase detector and will be discussed in more detail in chapter 2.

1.5 Current Experimental Outlook

The current leading exclusion limits are set by dual-phase TPC Xenon experiments such as Xenon1T [25], LUX [28] and PandaX [29] as can be seen from Fig.1.12 where the spin

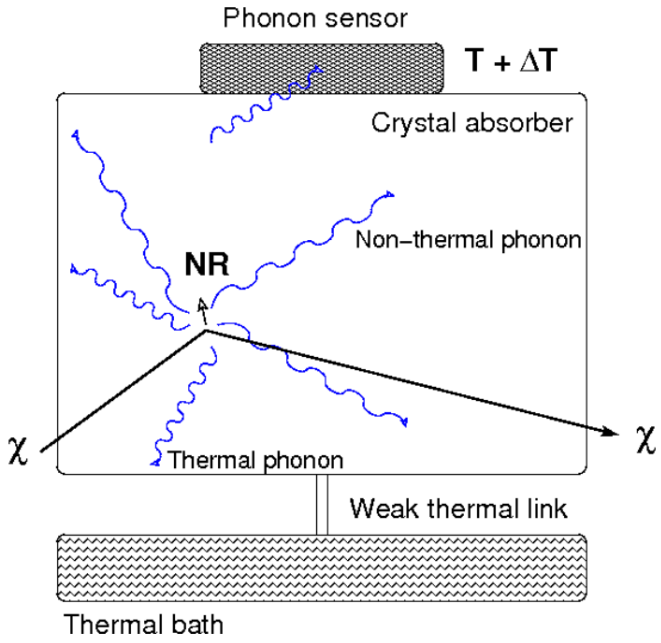


Figure 1.9: Schematic of a cryogenic phonon detector: A nuclear recoil (NR) deposits an energy E in an absorber of capacity C(T). This produces a temperature rise ΔT which is measured by a thermal sensor [17].

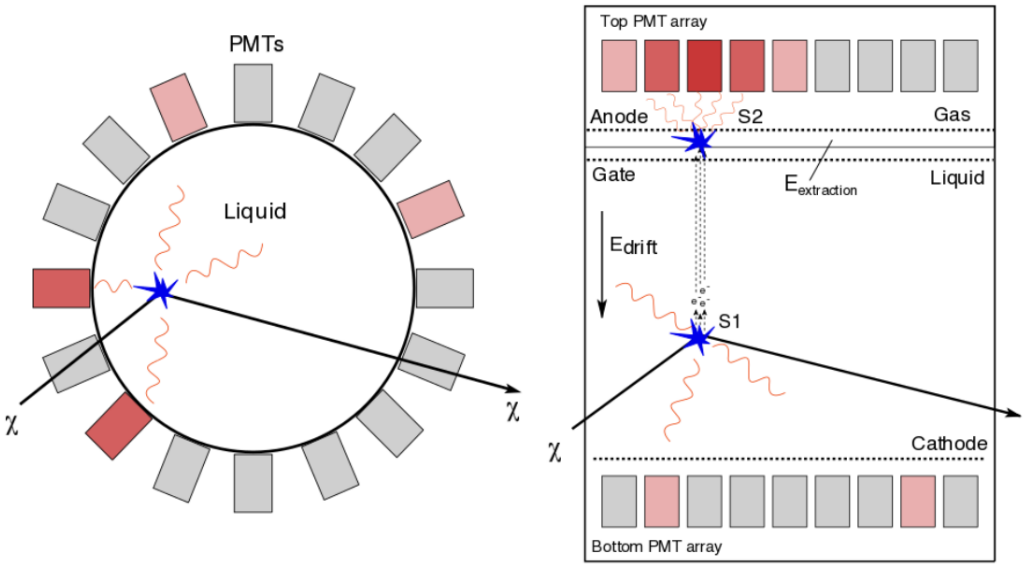


Figure 1.10: Schematic of single-phase (left) and dual phase (right) liquid noble-gas detectors. A single phase detector (like DEAP-3600) cannot collect charge signal from ionisation produced by an energy deposition as a dual phase detector, and uses pulse shape as the main particle discrimination parameter [17].

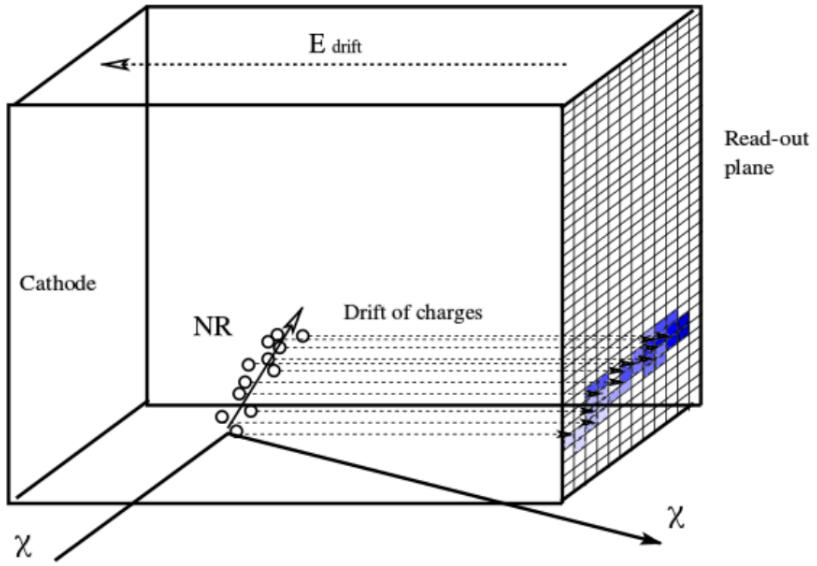


Figure 1.11: Schematic of a low-pressure gaseous time-projection chamber (TPC). The ionisation charge produced by a NR is drifted by a homogeneous electric field to the read-out plane allowing the track to be reconstructed. [17]

independent WIMP-nucleon cross section is plotted against the WIMP mass. The curves in the plot represents the limit that exclude the parameter space above them. The figure also includes signal indications represented with closed contours. Those are areas where experiments have claimed that a dark matter signal has been observed. It should be noted that the WIMP mass range covers four order of magnitude in this plot and many orders of magnitude in cross section. The best limit for spin dependent couplings comes from the PICO experiment [18]. For the spin independent case, liquid noble gas experiment have achieved huge progress over the last decade without seeing any signal as yet. Experiments that use xenon as a target material have currently the best sensitivity for WIMP detection. However, significant progress in the field have been made with liquid argon technology with DEAP-3600 currently having the best limit for argon detectors with a short dataset of 4.2 days [30]. To reach the irreducible neutrino-induced background at sensitivities $\sigma_{\chi p} \sim 10^{-49}\ cm^2$, for all the parameter space for WIMP masses greater than 10 GeV, will require multi-tonne argon and/or xenon detectors with basically perfect background discrimination. On the other side, for WIMP masses below 10 GeV, the challenge will be to decrease a detectors energy threshold as much as possible without requiring the building of large scale detectors. The DAMA/LIBRA claim remain an outstanding issue as even though they observed a signal with significance at 9.3 σ in their annual modulation

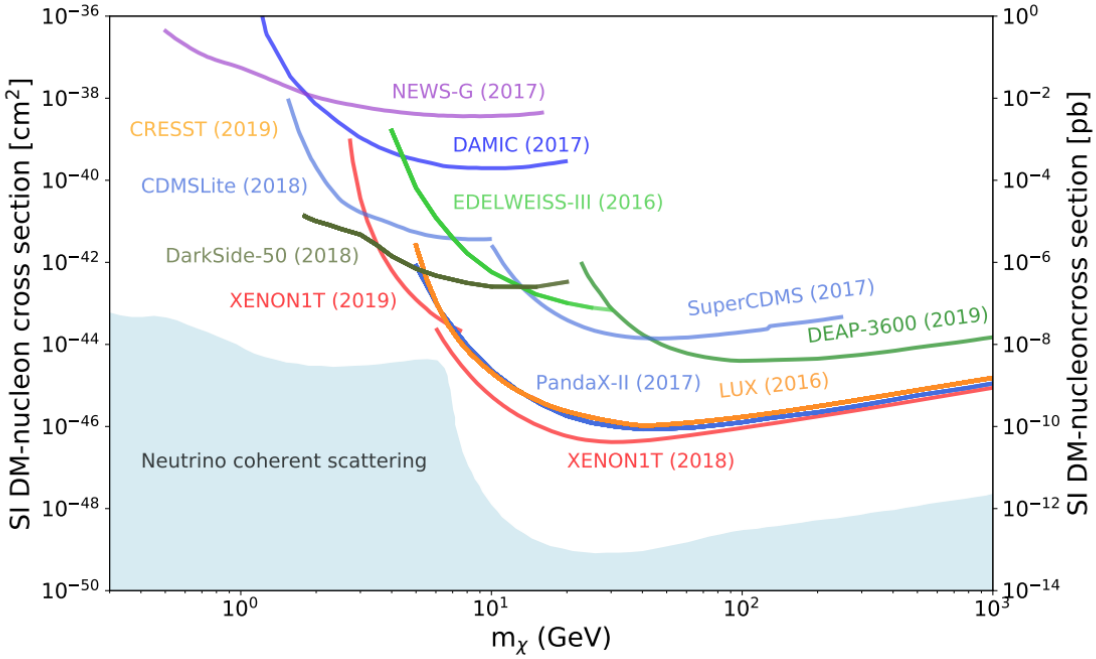


Figure 1.12: WIMP-nucleus scattering cross section limits from multiple direct detection experiments for spin-independent coupling. Xenon detectors have the leading limit at high mass. [35].

analysis, many other experiments have excluded that parameter space region. A number of experiments using similar technologies as DAMA/LIBRA in various underground facilities around the world, are currently or soon start taking to further investigate the nature of the DAMA/LIBRA signal observation (SABRE at LNGS in Italy [31], Anais at LSC in Spain [32], DM-Ice at the South Pole and PICO-LON at the Kamioka mine in Japan [33] [34]).

Chapter 2

Direct detection with the DEAP-3600 detector

The DEAP-3600 experiment, located 2 km underground at SNOLAB in Sudbury, Canada, is a single phase liquid argon direct detection experiment with a sensitivity to the spin-independent WIMP-nucleon cross section of 10^{-46} cm^2 for a 100 GeV/c^2 WIMP mass and can hold up to 3600 kg liquid argon.

2.1 Nobel liquids detection technique

The detector medium used in DEAP-3600 is liquid argon. In the past decade there has been increased development of liquid noble gas detectors within the field of particle physics for both precision neutrino and dark matter physics experiment. Argon, xenon and neon in particular, show excellent properties for detection of rare events. When radiation hits these mediums, both scintillation and ionisation is generated. This leads to two possible detection techniques: single or dual phase detection. DEAP-3600 is a single phase detector, meaning that it only uses scintillation light produced in the target material. A dual phase experiment measure both the charge and the light signal. Examples of some dual phase experiments are given in Sections 1.4 and 1.5.

Dual phase detectors consist of a cylindrical time projection chamber (TPC) filled with a noble gas. The target mass is in the liquid state with a layer of gas above it. Light is collected by layers of PMTs above and below the target mass and a uniform electric field is applied across the target volume parallel to the cylindrical axis. A detailed discussion of this detection technique, which is beyond the scope of this work, can be found in [36].

Briefly, the working principle of a dual phase detector, shown in Fig.2.1 in the case of a xenon detector, is as follows: the incoming particle produces scintillation light and ionises the noble gas. The electric field drifts the ionisation electrons from the interaction site up to the gaseous region above the liquid where a stronger electric field excites the xenon atoms generating a secondary scintillation light signal (S2) proportional to the charge liberated by the interaction. Both the S2 signal and the primary scintillation light S1 signal are recorded by the two arrays of PMTs. The ratio of S2/S1 is different for nuclear and electronic recoils allowing for background rejection. Detection of the two signals allows also for a three-dimensional reconstruction of the event vertex inside the TPC volume.

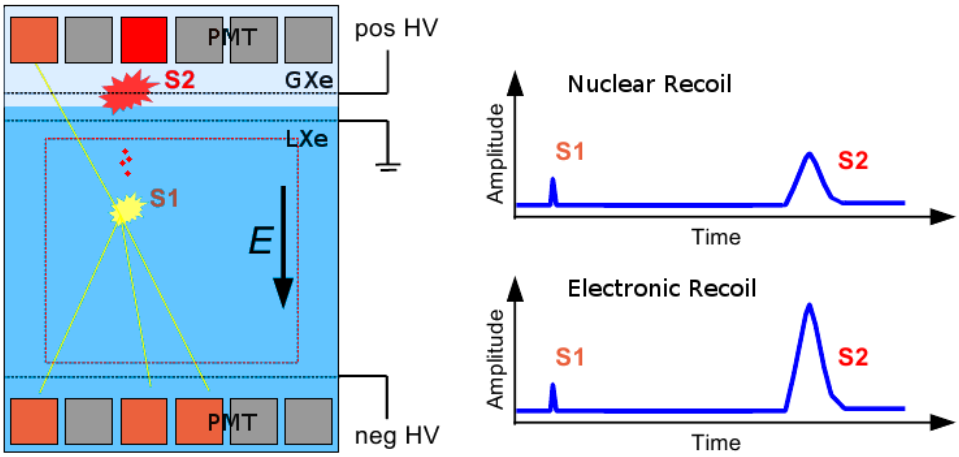


Figure 2.1: Schematic of a dual phase xenon time projection chamber. Both scintillation light (S1) and the ionisation charge signal are measured. The latter is converted to a proportional scintillation signal (S2) in the xenon gas region. [37].

Single phase detectors instead rely on the scintillation light observed by the PMTs only for both particle discrimination and position reconstruction. This technology will be described in the next sections as DEAP-3600 makes use of this detection technique.

Table 2.1: Some physical properties of noble elements used for neutrino and dark matter detection [38].

Property	He	Ne	Ar	Kr	Xe
Atomic number	2	10	18	36	54
Atomic mass	4.0	20.18	39.95	83.80	131.29
Boiling point (T_b) at 1 atm [K]	4.22	27.1	87.3	119.74	165.0
Melting point (T_m) at 1 atm [K]	-	24.6	83.8	115.8	161.4
Gas density at 1 atm and 298 K [gl^{-1}]	0.16	0.82	1.63	3.43	5.40
Liquid density at 1 atm and (T_b) [gcm^{-3}]	0.12	1.21	1.40	2.41	2.94
Average cost [$/100 g]	5.2	33	0.5	33	120

2.2 Scintillation in noble liquids

Scintillation light in noble liquids is produced when an interacting particle transfers part of its energy to the medium. When this happens the medium atoms are ionised and excited dimer states (excimer) forms and decay releasing scintillation photons. Excimers are produced in a singlet ($^1\Sigma_u^+$) or triplet ($^3\Sigma_u^+$) state. A transition between either the singlet or triplet first excited electronic state to the ground state ($^3\Sigma_g^+$) results in the emission of a scintillation photon. The excitation mechanism, assuming liquid argon being the medium for example, is as follows:

$$\chi + \mathrm{Ar} \rightarrow \mathrm{Ar}^* \tag{2.1}$$

$$\mathrm{Ar}^* + \mathrm{Ar} \rightarrow \mathrm{Ar}_2^* \tag{2.2}$$

$$\mathrm{Ar}_2^* \rightarrow 2\mathrm{Ar} + h\nu \tag{2.3}$$

where Eq. 2.1 describes the argon atom excitation due to the incoming particle χ, eq. 2.2 the recombination with a ground state argon atom to form the excimer Ar_2^* which decays to two non-excited argon atoms and the emission of the scintillation photon $h\nu$ emitted in the vacuum-ultraviolet (VUV) range (eq. 2.3). When argon atoms undergo ionisation, the scintillation photon is produced through the following mechanism:

$$\chi + \mathrm{Ar} \rightarrow \mathrm{Ar}^+ + e^- \tag{2.4}$$

$$\mathrm{Ar} + \mathrm{Ar}^+ + e^- \rightarrow \mathrm{Ar}_2^*, \tag{2.5}$$

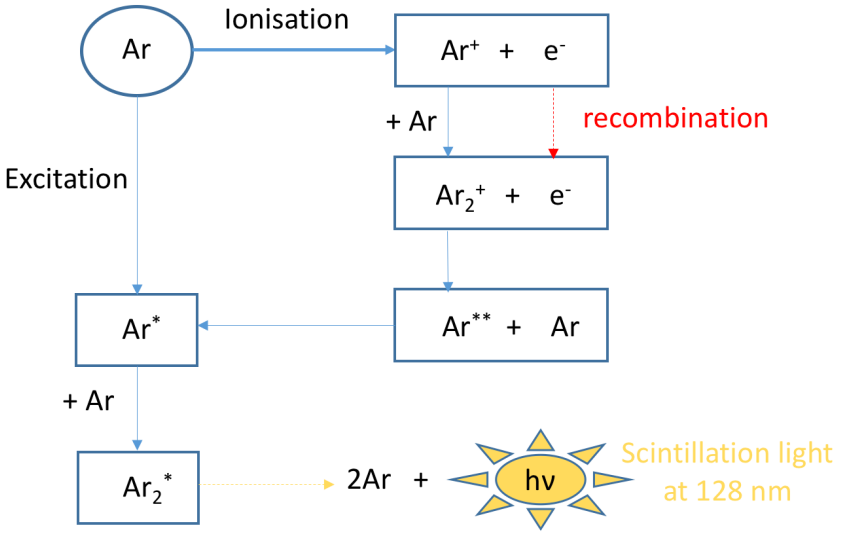


Figure 2.2: Schematic of the scintillation light production through excitation and ionisation in argon.

Where an orbital electron from an argon atom gets released due to the incoming particle χ creating a positively charged argon atom Ar^+ (eq. 2.4), which then leads to the formation of excimers when it recombines with a neutral argon atom and a nearby free electron (eq. 2.5). The excimers follow the same decay as described in Eq. 2.3 to create the VUV photon.

Relaxation process from excited states

An important property of noble liquids, which turns out to be very useful for rare event detections, is that the singlet and triplet decay constants are not the same. This is due to the fact that a single state decay is an allowed transition, whereas a triplet state decay is a forbidden transition that requires a spin flip, thus a longer lifetime, and is made possible due to spin orbit coupling [39]. The triplet decay lifetime very much depends on the atomic number of the noble gas. The singlet state lifetime (τ_s) in liquid argon is 7 $\pm$ 0.1 ns and the triplet state lifetime is much higher, measured to be (τ_t) 1.6 $\pm$ 0.1 μs [40]. Noble liquids with a higher atomic number than argon such as liquid krypton and liquid xenon have a much shorter triplet decay lifetime (85 ns and 27 ns respectively [41]) due to their stronger spin orbit coupling. Singlet and triplet decay lifetimes for common noble gases used in dark matter detection are summarised in Table 3.1 together with other important properties of these medium materials. Noble gases lighter than argon like neon and helium

have even longer decay constants.

Table 2.2: Properties of common noble gases used in dark matter detection [42].

Property	He	Ne	Ar	Kr	Xe
Electron yield (e^-/keV)	39	46	42	49	64
Photon yield (γ/keV)	22	32	40	25	42
Singlet decay time (ns)	10	10	7	7	5
Triplet decay time (ns)	1.3×10^{10}	1.5×10^4	1.5×10^3	85	27
Scintillation wavelength (nm)	80	78	128	148	175

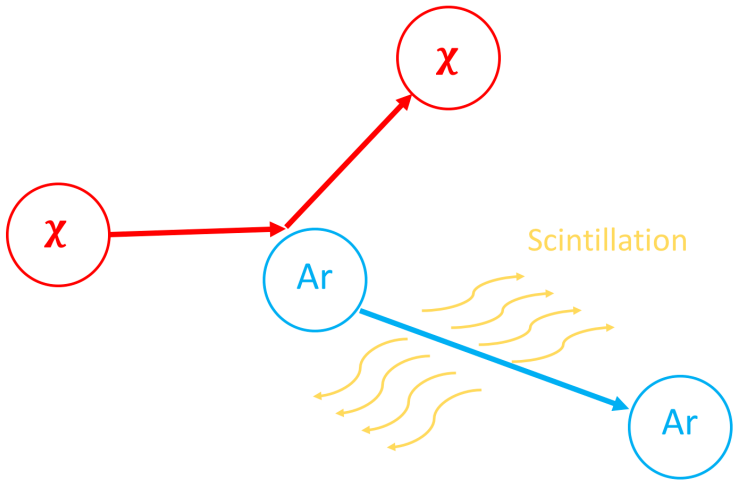


Figure 2.3: WIMP interaction cartoon.

Recombination

The relative populations of singlet and triplet states depends on various factors such as the type of particle that interacts with the medium and recombination. Recombination also affects the scintillation light yield. Experiments on the effect of electric fields on the LAr luminescence excited by energetic electrons conducted by Kubota [43] have shown that 67% of the scintillation light is due to recombination. [44] found that an electric field of 1 keV/cm, reduces the light yield for nuclear recoils in liquid argon up to 32%. [39] measured the singlet to triplet ratio ($\frac{\tau_s A_s}{\tau_t A_t}$) with and without an electric field applied finding an enhancement of the singlet state without the field applied ($\frac{\tau_s A_s}{\tau_t A_t} = 0.045$ with the field applied, $\frac{\tau_s A_s}{\tau_t A_t} = 0.083$ without the field and hence having recombination). The linear energy transfer (LET) that an ionising particle transfers to the material traversed depends on the nature of the radiation for the same material. The higher the LET, the

more ionisation is produced, the more enhanced the single state channel is. Thus, nuclear recoils and alpha particle interactions are more likely to produce excimers in a single state given their LET as opposed to electromagnetic interactions which will preferentially excite the argon excimer triplet state. This property is the basis for background suppression used in liquid noble gas experiments through pulse shape discrimination.

Nuclear recoil light yield

The scintillation light yield is defined as the number of photons N_{ph} emitted per unit energy E. Nuclear and electronic recoils of the same energy show different light yields [45]. The number of photons produced depends on the LET of the interaction. Namely, interactions producing nuclear recoils (high LET) generate for the same total energy transfer fewer photons than interactions with lower LET (electronic recoils). The ratio of total light produced by electronic recoils over that produce by nuclear recoils is know as nuclear recoil efficiency, $\mathcal{L}_{\text{eff}}$, and is expressed in reference to the scintillation light yield for 122 keV γ rays from a cobalt source:

$$\mathcal{L}_{\text{eff}}(E) = \frac{W_{scint,e}(122\text{keV})}{W_{scint,nr}(E)} \tag{2.6}$$

Where the quantity $W_{scint,e/nr}$ is the energy required to produce one scintillation photon. $\mathcal{L}_{\text{eff}}$ in liquid argon was found to be constant by [46] and equal to 0.25 ± 0.02 above 20 keV. SCENE [47] reported $0.235 < \mathcal{L}_{\text{eff}} < 0.296$ for nuclear recoils with energies between 10.3 and 57.3 keV. The discrepancy in light yield discussed so far leads to the use of two energy scales keVee and keVr, where keVee (keV electron equivalent) refers to the energy deposited by an electronic recoil and keVr (keV recoil) refers to the energy deposited by a nuclear recoil. The conversion factor between the two is $\mathcal{L}_{eff}$ according to:

$$E[\text{keVee}] = E[\text{keVr}] \times \mathcal{L}_{\text{eff}}(E) \tag{2.7}$$

For example a 20 keVee electronic recoil, assuming $\mathcal{L}_{\text{eff}} = 0.25$, produces in liquid argon the same number of scintillation photons as a nuclear recoil of 80 keVr.

Quenching

Several processes can contribute to the lower scintillation efficiency of nuclear recoils with respect to electronic recoils. This difference is known as quenching and the mechanisms

through which it occurs are:

- **Heat (Lindhard quenching):** For nuclear recoils some of the energy deposited goes into heat, instead of creating ionisation or excitation.

- **Biexcitonic quenching (or Hitachi quenching):** It proceeds according to:

$$\mathrm{Ar}^* + \mathrm{Ar}^* \rightarrow \mathrm{Ar} + \mathrm{Ar}^+ + e^- \tag{2.8}$$

  where the scattering of two excited atoms (argon in this example) produces relaxation and ionisation resulting in only only one photon generated though recombination rather than the two that would normally produce two excitons [48].

- **Electron escape:** Some fraction of the ionised electrons may escape recombination.

- **Impurities:** These can significantly decrease the light yield by quenching the liquid excimers through energy transfer from them to the contaminant or absorbing emitted UV photons.

Lindhard-Birks model

The two main processes responsible for the lower scintillation efficiency of nuclear recoils are the Lindhard and the Hitachi quenching. A phenomenological model to account for these two effects has been developed by Mei *et al.* [49] and applied to liquid argon experiments by Lippincott [50]. The Lindhard component for a nucleus with atomic number Z and atomic mass A, is parameterised by a factor f_n [51]:

$$f_n = \frac{kg(\epsilon)}{1 + kg(\epsilon)} \tag{2.9}$$

where $\epsilon = 11.5 \times Z^{-7/3} E_R(keV)$, $k = 0.133 \times Z^{2/3} A^{-1/2}$ and $g(\epsilon) = 3\epsilon^{0.15} + 0.7\epsilon^{0.6} + \epsilon$.

The second quenching component, f_l, that accounts for the Hitachi biexcitonic quenching bit, can be modeled using Birk's saturation law that relates the relative scintillation response of scintillators to dE/dx, and the amount of energy deposited per unit length [52].

$$f_l = \frac{1}{1 + k_B \frac{dE}{dx}} \tag{2.10}$$

where k_B is the Birk's constant that can be determined experimentally and the stopping power $\frac{dE}{dx}$ is calculated by the SRIM software package for argon ions in liquid argon [53].

2.3 Pulse shape discrimination

Pulse shape discrimination refers to the technique used to discriminate between potential signal and background in a single phase detector. Liquid argon scintillation properties described in the previous section make it an excellent medium for this purpose. In particular, as already mentioned, the ratio of singlet to triplet states highly depends on the interacting particles. A light particle, like an electron, is more likely to produce excimers in a singlet state (electronic recoils) as opposed to heavier particles such as neutrons that will more likely produce triplet states (nulcear recoils). Given the very different lifetime of these states, a quantity referred to as $\mathrm{F_{prompt}}$ has been defined for particle identification. $\mathrm{F_{prompt}}$ is defined as:

$$F_{prompt} = \frac{\Sigma_{\{i|t_i \in (-28ns, 150ns)\}} Q_i}{\Sigma_{\{i|t_i \in (-28ns, 10\mu s)\}} Q_i} \tag{2.11}$$

where the numerator in eq. 2.11 is the amount of light recorded over over a prompt timing window (150 ns) and the denominator is the total amount amount of light recorded in the entire timing window. The size of this time window is selected such that the separation between the nuclear and electronic recoil bands in the region of interest for the dark matter search is maximased. An example of two waveforms for nuclear recoil and electron recoil events respectively is shown in Fig.2.4. The pulse charge collected within the prompt window (yellow box in the plot) is higher for the nuclear recoil event.

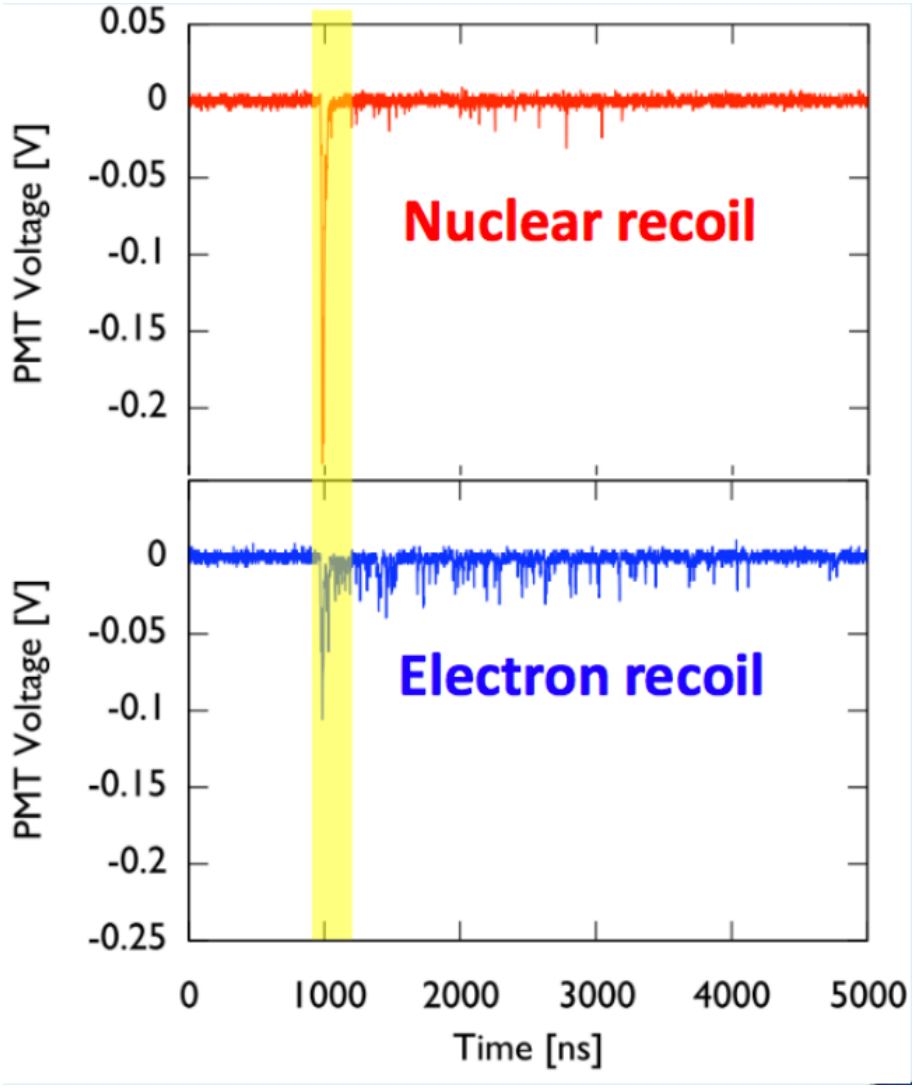


Figure 2.4: Example of waveforms for a nuclear (red) and electron (blue) recoil in DEAP-1. The yellow band defines the prompt time window. Nuclear recoils show more prompt light with respect to electron recoils [54].

The separation power of F_{prompt} can be appreciated in Fig.2.5 where the F_{prompt} distributions for electronic recoils from tagged γ events from a ^{22}Na source run and nuclear recoils from an AmBe neutron source are plotted for the DEAP-1 prototype detector.

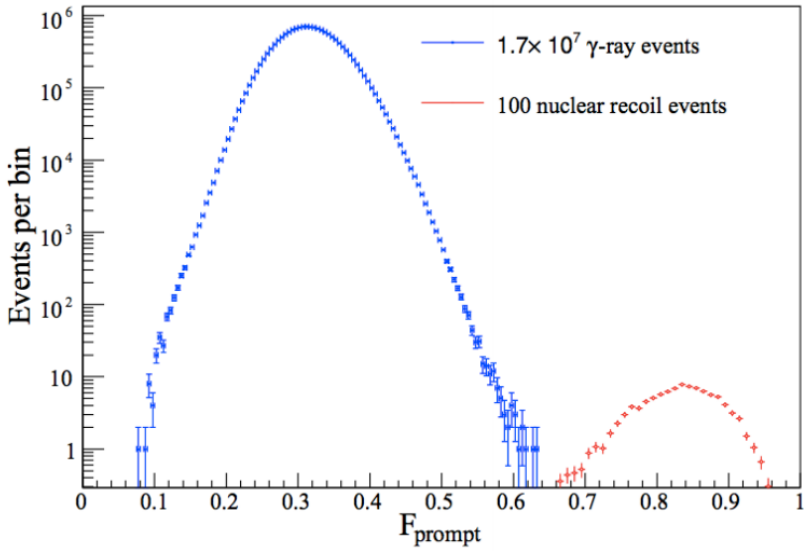


Figure 2.5: Fprompt distribution of 16.7 million tagged γ events from an ^{22}Na calibration source run and 100 tagged nuclear recoil events from an AmBe calibration source in the DEAP-1 prototype detector. Clear separation between the electronic and nuclear recoil distributions can be seen [55].

Finally, Fig.2.6 shows the F_{prompt} as a function of energy for nuclear and electronic recoils where the nuclear recoil and electronic recoil bands are plotted in different colours. Given the 'promptness' of their signal, nuclear recoils will sit in the high F_{prompt} region of the parameter space whilst electronic recoils occupy the low F_{prompt} region. This can be seen in DEAP-3600 with neutron calibration data taken with an AmBe source in Fig.2.8.

2.3.1 Region of Interest

The region of interest (ROI) for the DEAP-3600 dark matter search is a region in the energy - F_{prompt} parameter space where nuclear recoils from WIMP interactions are expected. An important input to defining the ROI is the expected rate of background events leaking into the ROI. The "nominal" energy region of interest in DEAP-3600 is in a window from 120 (15 keVee) photolectrons (PE) to 240 (30 keVee) PE. This range is chosen by calculating the level of PSD by extrapolation of the DEAP-1 model that gives an expected ^{39}Ar background rate of < 0.2 events over 3 years leaking into the 50% WIMP acceptance region. The DEAP-1 projection model, however, was outperformed by the DEAP-3600 effective model (see Fig.2.7) presented in the first result published [30], for a 4.44 live days dataset analysed, allowing for a lower the ROI energy threshold of 80 PE (10 keVee) in the analysis. The ROI contours used for the first result can be seen in Fig.2.8 drawn in black.

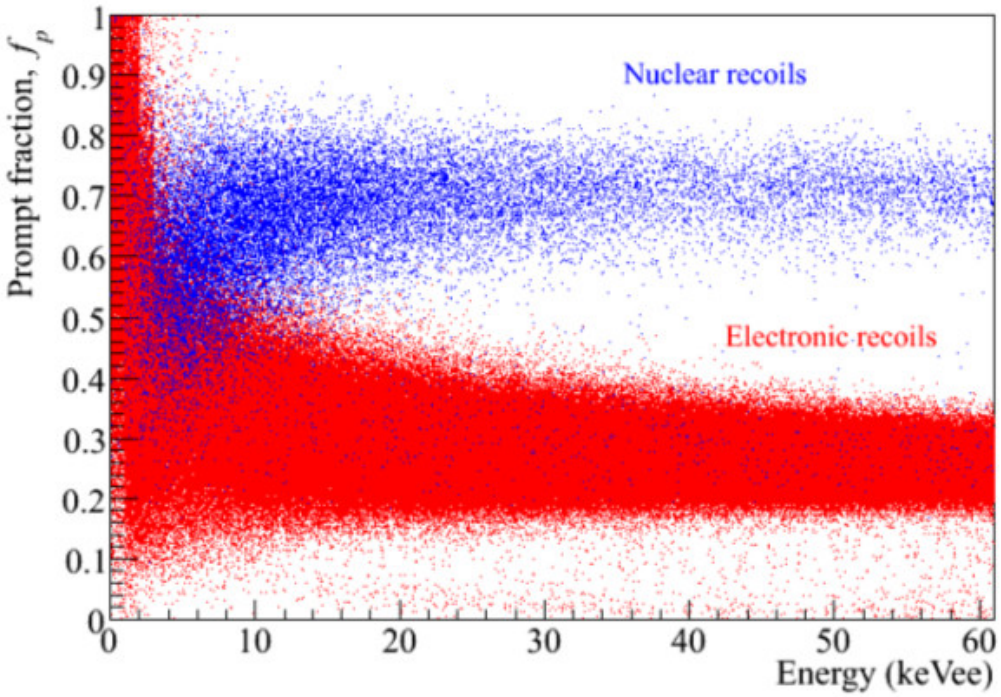


Figure 2.6: Fraction of light arriving in the prompt time window as a function of energy for nuclear and electronic recoils. The separation is what enables the PSD.

The F_{prompt} lower bound of the ROI above 150 PE was chosen to remove 5% of nuclear recoils in each bin with the upper bound of the ROI removing 1% of nuclear recoils in each bin. The maximum energy of 240 PE was kept to its nominal value and was chosen to reduce possible backgrounds from surface alphas.

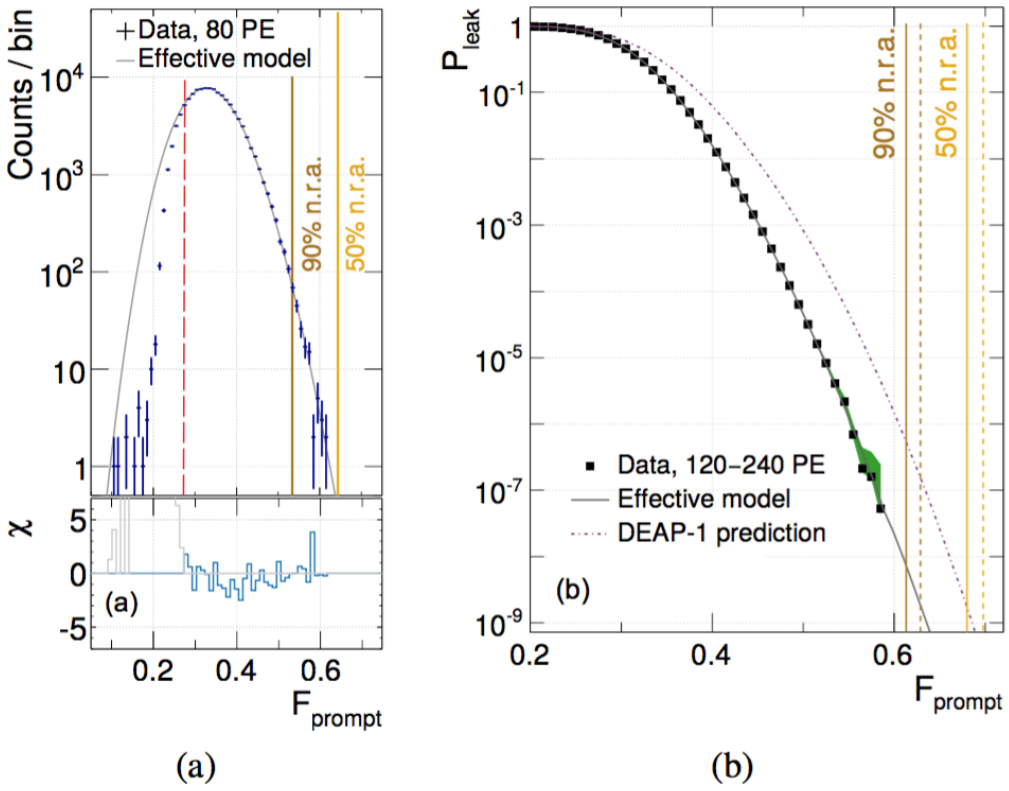


Figure 2.7: (a) Projection of F_{prompt} distribution for 80 PE events, plotted with the effective model fit, as labelled. Red dashed line indicates the lower limit of the fit range. The brown and orange lines represent the 90% and 50% nuclear recoil acceptance boundaries. (b) Comparison of data to model for 120-240 PE range, with 90% and 50% nuclear recoil acceptance indicated. The dashed line represents the DEAP-1 projection to be compared to the improved effective model from DEAP-3600 (solid line) [30].

2.4 The DEAP-3600 detector

The DEAP-3600 detector is a single phase liquid argon detector capable of holding 3600 kg of liquid argon. It consists of a transparent acrylic vessel (AV) with an inner radius of 85 cm and a minimum thickness of 5 cm, surrounded by 255 8-inch diameter R5912 high quantum efficiency photomultipliers (PMTs). The inner AV surface is covered with a thin layer of wavelength shifter of tetraphenyl butadiene (TPB), which converts the 128 nm argon scintillation light into visible blue light. To thermally insulate the PMTs from the argon volume and to shield against external neutrons and neutrons produced in the PMT glass, cylindrical acrylic light guides (LGs) 50 cm long and 9 cm in radius are bonded onto the AV. Further shielding and insulator elements sit between the light guides, referred to as filler blocks, which are made of alternating layers of high density polyethylene and

styrofoam. These structures are housed in a large stainless-steel spherical shell, immersed in an 8 meter diameter ultra pure water tank instrumented with 48 PMTs, which provide shielding against radiation from the surrounding rock as well as acting as a Cherenkov veto for cosmogenic muons. At the top of the AV, the outer acrylic neck with a diameter of 25 cm and height of 50 cm allows access to the inner AV and circulation of argon. A longer steel and acrylic neck is attached to it and contains a cooling coil that cools down the liquid argon using liquid nitrogen. Insertion or extraction of equipment is done through a radon-free glove box at the top of this neck. A rendering of the detector and its main components can be seen in Figure 2.9.

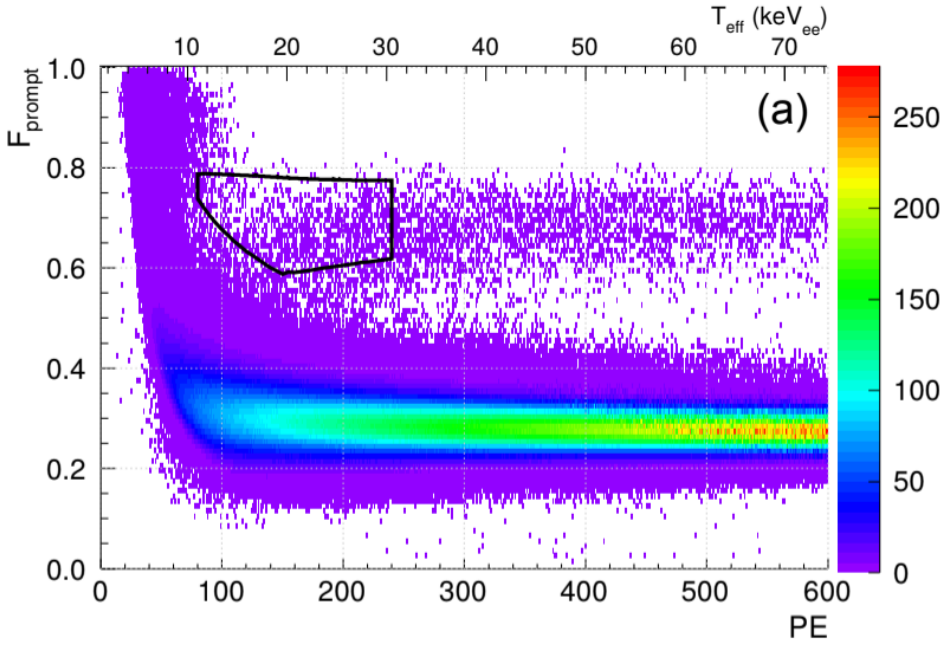


Figure 2.8: F_{prompt} versus photoelectrons (PE) for an AmBe calibration run in DEAP-3600. Nuclear recoil band (high F_{prompt}) is produced by neutron scattering, the electronic recoil band (low F_{prompt}) by β decays from ^{39}Ar and 4.4 MeV γ rays from the AmBe source. The black box shows the WIMP search ROI.

2.4.1 Acrylic Vessel, Light Guides and Material Assay

The AV hold the liquid argon. It is made of a polymer of methyl methacrylate (MMA) produced by the Thai MMA Co. This material was chosen to satisfy the very strict requirements of low radioactivity required for the success of the experiment and its optical properties. Both the inner surface and the bulk of the AV can produce backgrounds that can mimic WIMP events. In particular, alpha decays from the inner surface can produce

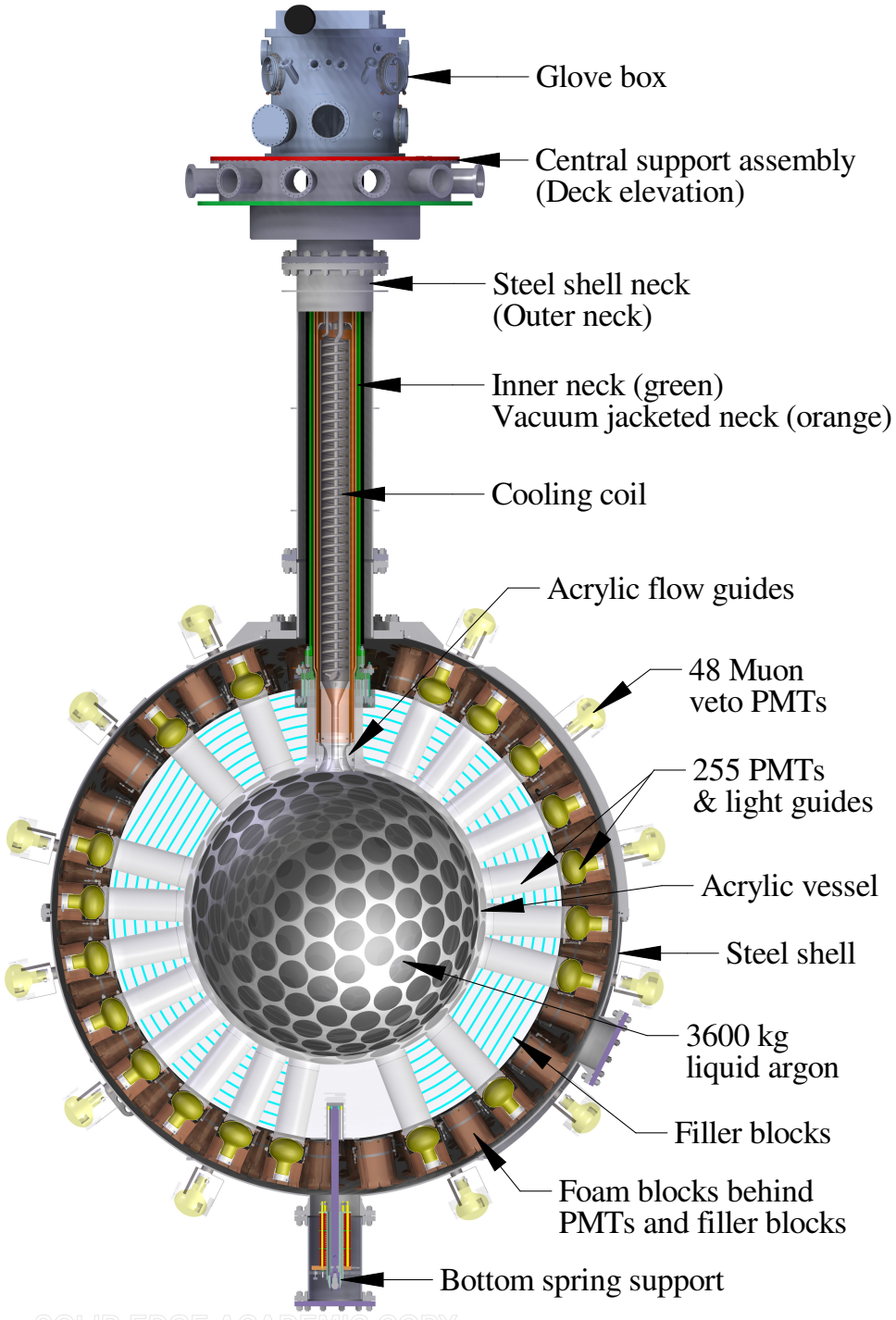


Figure 2.9: Schematic of the DEAP-3600 detector. The acrylic vessel has an inner radius of 85 cm and it is surrounded by 255 8-inch high quantum efficiency photomultipliers. The rendering also shows the 50 cm long light guides, the surrounding filler blocks and the coiling coil [56].

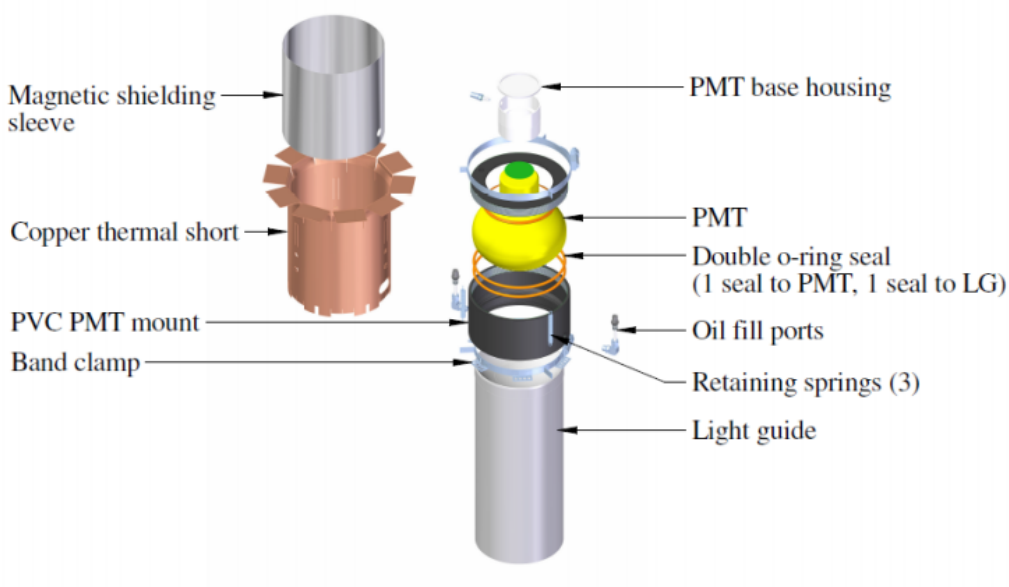


Figure 2.10: Schematic of a PMT coupled to a light guide including the magnetic shield and copper thermal short [56].

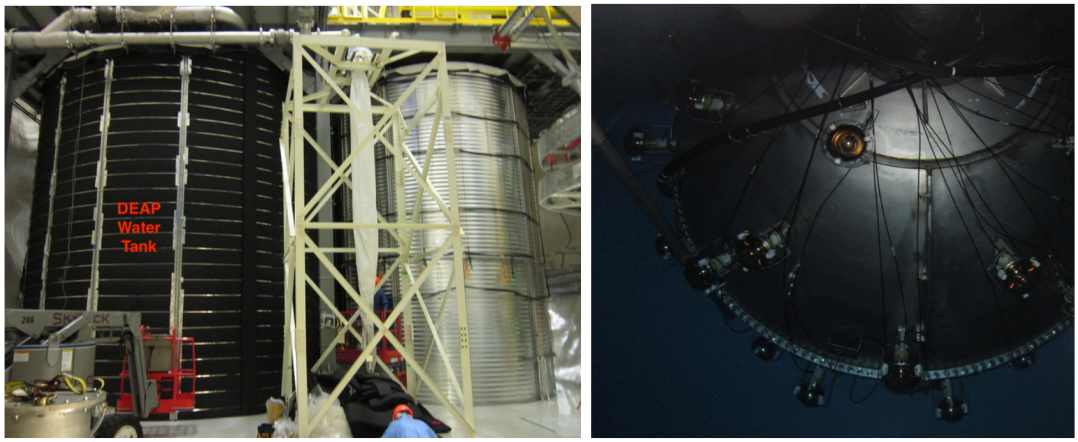


Figure 2.11: Left: Photograph of the Cherenkov water tank (labelled DEAP Water Tank). Right: Image of the outer vessel hanging in the water tank and the veto PMTs.

nuclear recoils while alpha decays in the acrylic bulk can produce neutrons through the $^{12}C(\alpha, n)^{16}O$ reaction.

2.4.2 Tetraphenyl Butadiene deposition

Scintillation light generated in the liquid argon is emitted at 128 nm in the VUV range. However, the R5912-HQE PMTs are more sensitive to the visible spectrum. To maximise light detection across the detector, the inner surface of the AV was coated with a layer of Tetraphenyl Butadiene (TPB) that acts as a wavelength shifter. Photons absorbed by the TPB coating are re-emitted near the peak quantum efficiency of the PMTs in the visible range (∼420 nm) as can be seen in Fig.2.12 for various wavelengths of UV light.

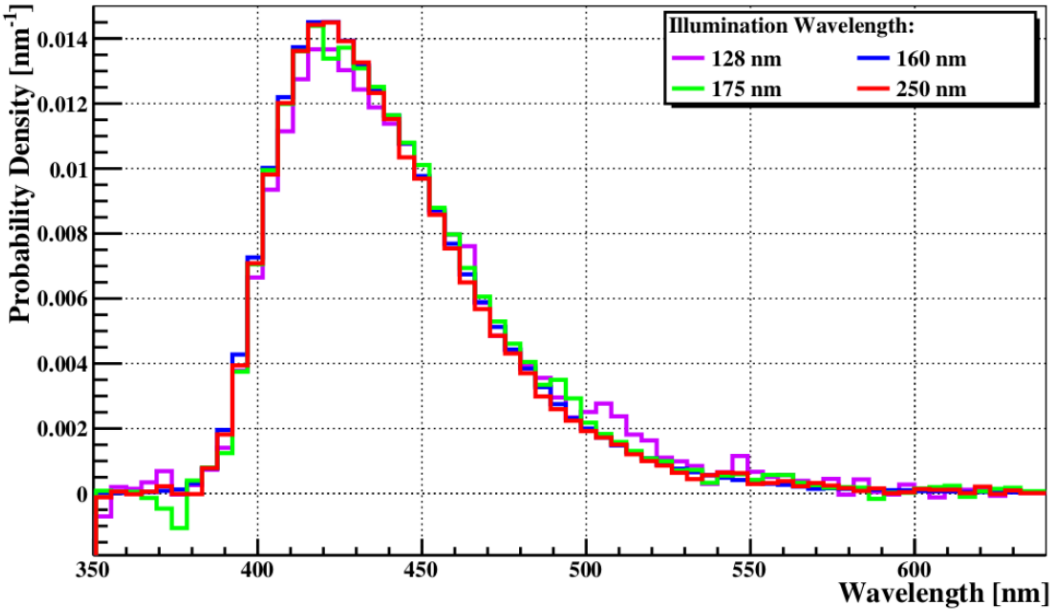


Figure 2.12: TPB fluorescent re-emission spectrum for various wavelengths of UV light. Re-emission wavelength under 128 nm illumination is at 420 nm, close to the peak efficiency of R5912-HQE [57].

Fig.2.13 shows the fluorescence efficiency as a function of incident wavelength for a 1.5 μm layer of TPB. At 128 nm the fluorescence efficiency is 1.2.

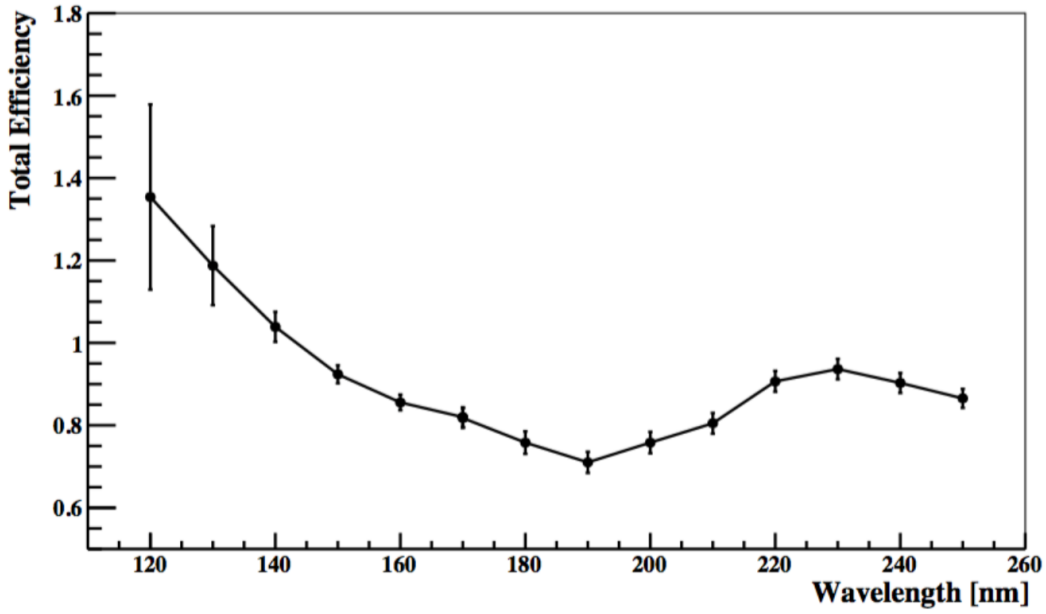


Figure 2.13: Fluorescence efficiency as a function of incident wavelength for a 1.5 μm layer of TPB [57].

The TPB film was deposited inside the AV using physical vapour deposition through the spherical evaporation source shown in Fig.2.14. This source, made of stainless steel, has an 11 cm diameter and hosts an inner copper crucible that holds the TPB powder. A flexible Watlow 125CH93A1X coil heater, wound around the stainless steel surface, radiatively heats the copper crucible housed inside to 208°C above its sublimation temperature. A uniform outgoing flux of TPB is generated through evaporation that reaches the inner surface of the detector through the 14 mm diameter holes. During the deposition process, the AV was under 10^{-6} mbar vacuum and the source was deployed at its centre. To reach this level of vacuum needed for deposition, the inner surface of the AV was brought to 50°C to outgas absorbed water. This was done by using the empty TPB source as a heating element while the acrylic vessel was under vacuum. A total mass of 294 $\pm$ 0.2 g TPB powder was deposited in two evaporations for final uniform thickness of 3.00 $\pm$ 0.02 μm [58].

2.4.3 Filler Blocks and Temperature Sensors

The space between light guides is filled with alternating layer of high-density polyethylene and styrofoam referred to as filler blocks. There are 486 of them in total and they act both as thermal insulators and shielding against neutrons. Figure 2.15 shows a filler block with resistance temperature detectors (RTDs) installed on it for temperature monitoring.

Figure 2.14: Image of the TPB evaporation source. The black filament is a coil heater that by heating the crucible inside the sphere allow the evaporation of the TPB powder for deposition [56].

The inner detector is equipped with 124 RTDs to measure the temperature close to the liquid argon bulk and the gradient temperature across the light guides (acrylic RTDs) and to monitor the PMTs temperatures (PMT RTDs). The 31 filler blocks homogeneously distributed across the AV have three temperature sensors bonded to the bottom, middle and top block layers, at a distance of 0.9 m, 1.1 m, and 1.3 m, respectively, from the centre of the AV. PMT temperatures are monitored by another 31 sensors bonded to the copper thermal shorts on a light guide adjacent to the filler blocks. A gap of 5 mm was left between each block and the light guide to take into account the contraction of both the AV and filler blocks during cool down. Moreover, retaining springs were installed at the warm end of the light guides in order to push the blocks against the AV. This also ensured that they would stay centred during the AV shrinkage during cool down.

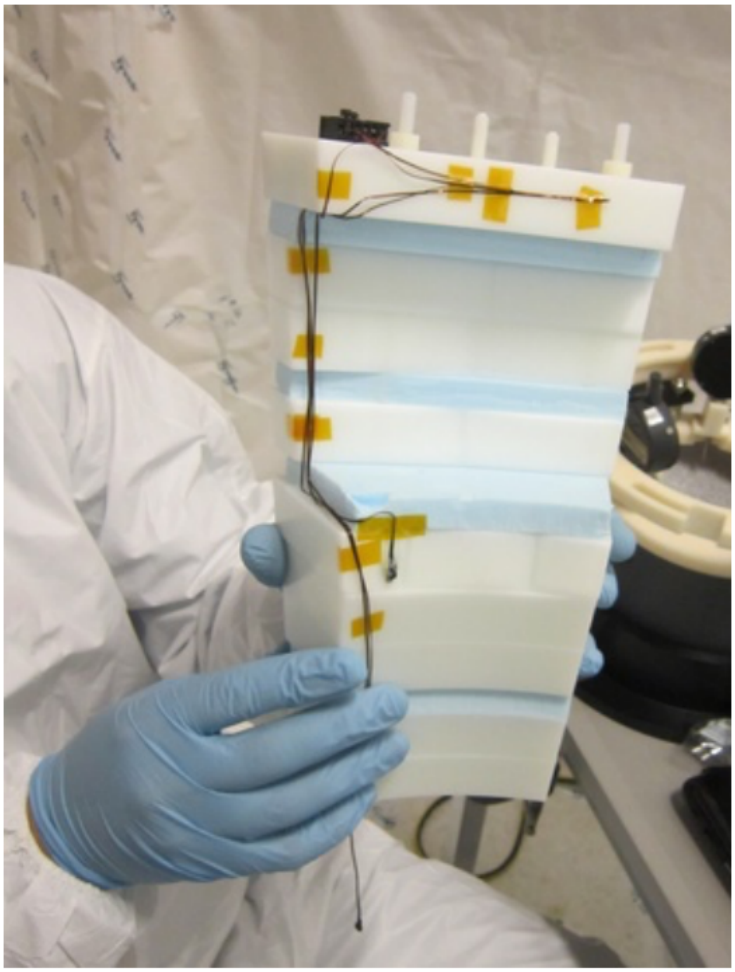


Figure 2.15: Image of a filler block with RTDs installed for temperature monitoring.

2.4.3.1 Photomultiplier tubes

The Hamamatsu R5912 8-inch-diameter HQE PMTs chosen for the DEAP-3600 light detection system, ensure high photon detection efficiency, good timing characteristics and low dark noise. The nominal quantum efficiency of these tubes is 32% and their operating temperature is in the range from -60 to 50°C with a recommended range between -20 to 5°C.

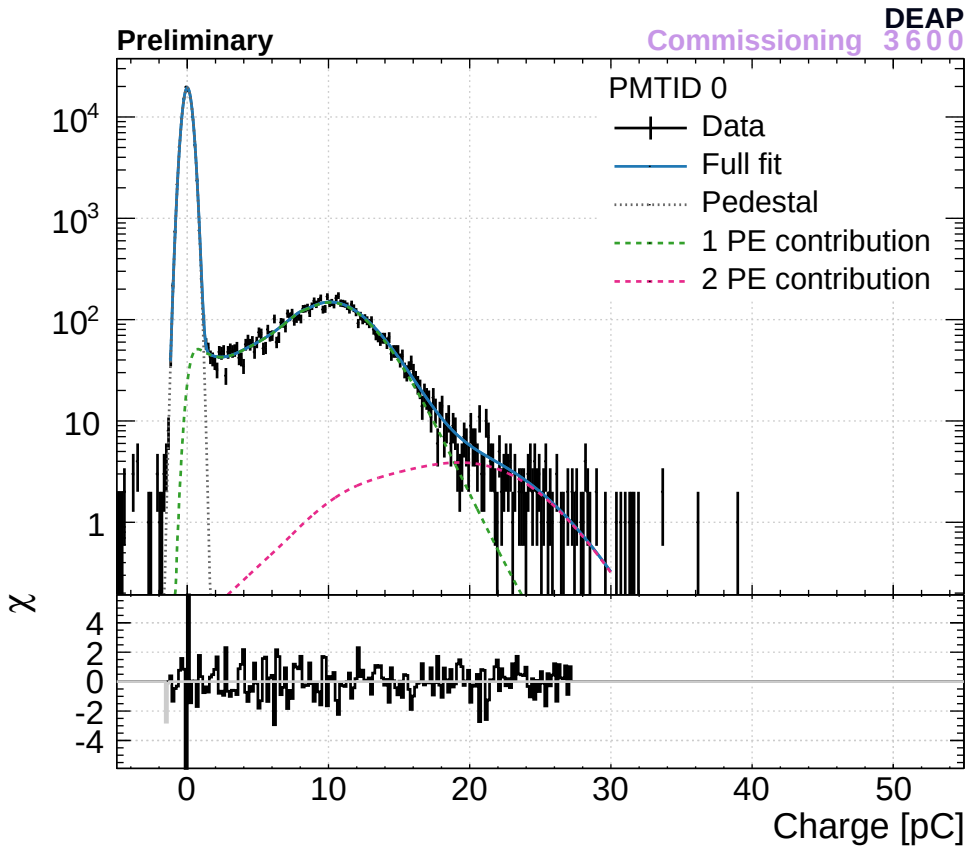


Figure 2.16: Single photoelectron (SPE) spectra for one generic PMT made with an aluminium acrylic reflectors (AARF, section 2.6) optical calibration run. The mean of the single PE charge is around 10 pC. Plot courtesy T. Pollmann

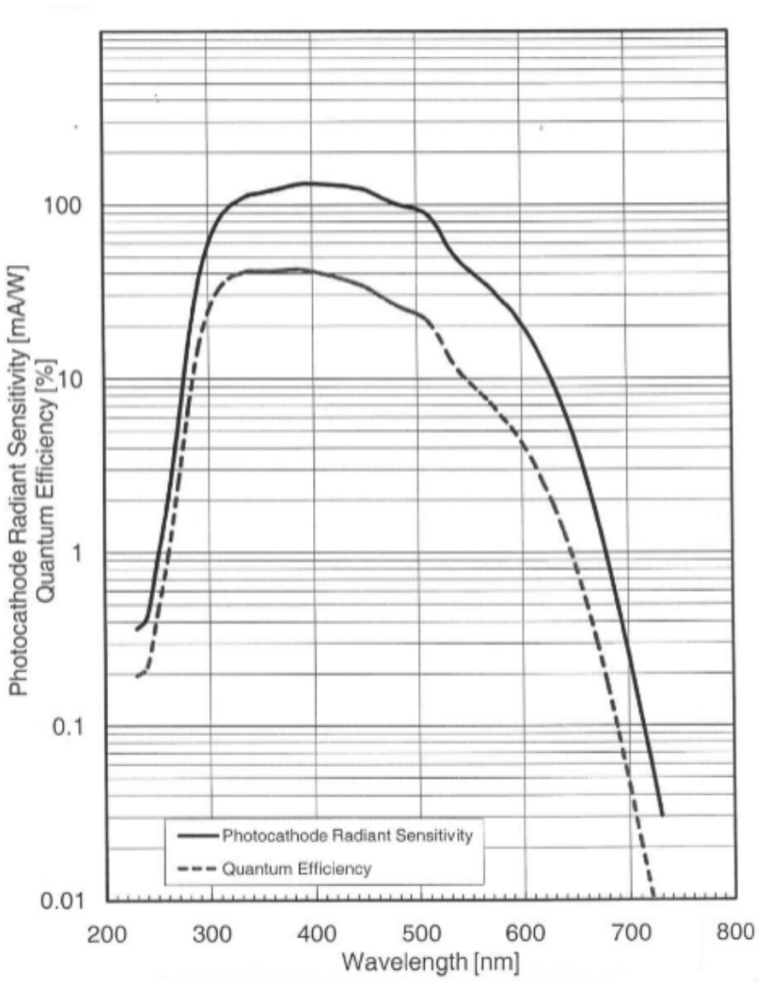


Figure 2.17: Spectral response characteristics measured at Hamamatsu for an R5912-HQE PMT. The maximum quantum efficiency is 42.2% at 390 nm wavelength incident light [57].

The quantum efficiency and radiant sensitivity as a function of wavelength for an R5912-HQE tube can be seen in Figure 2.17. They operate at a bias voltage between 1500 V and 1900 V and the mean single photoelectron (SPE) charge for all PMTs has been found to be 9.39 pC with an RMS of 0.16 pC from in-situ measurements as shown in Fig.2.18. The mean SPE charges $\bar{q}$ - monitored on an ongoing basis - depend on the applied bias voltage V through:

$$\bar{q} = AV^{\gamma} \tag{2.12}$$

where A is a normalisation parameter and γ is a parameter that is necessary to quantify the effect of fluctuations in the bias voltage and to adjust the mean SPE charge such that the gains across the entire PMT array can be matched. The γ parameter was determined to be equals to 6.97 ± 0.01 through LED calibration data taken with the PMTs at different

bias voltage [56].

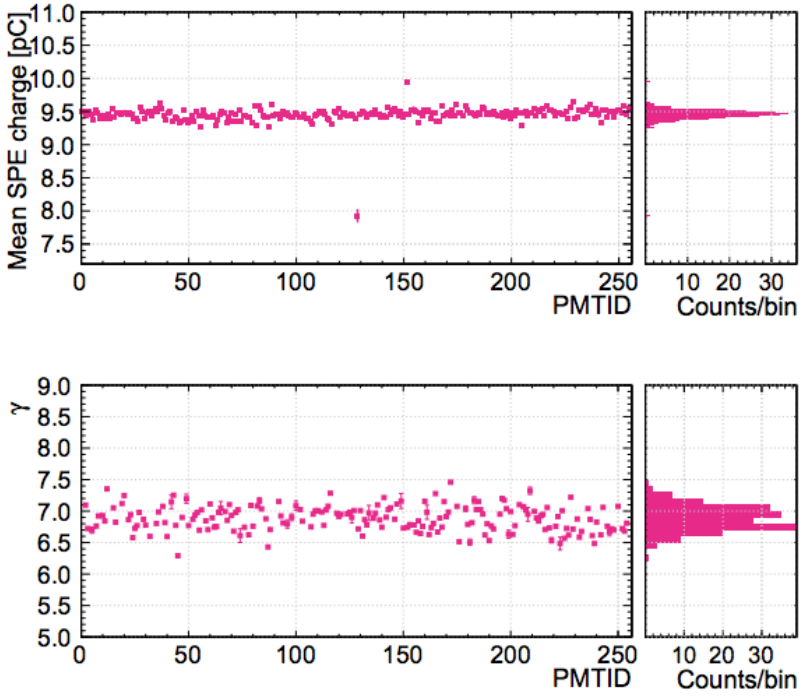


Figure 2.18: Top: Mean single photoelectron charge vs. PMTID. Mean is 9.39 pC with an RMS of 0.16 pC. Bottom: γ parameter vs. PMTID, with a mean of 6.9 and RMS of 0.2 [56].

Figure 2.19 shows the dark noise rates, which are strongly temperature dependent, versus PMT ID while the detector was at room temperature and after cool down with the detector filled with cold argon gas. In this configuration the dark noise shows a mean of 0.24 kHz and an RMS of 0.06 kHz. Measuring the dark noise rate with the AV filled with LAr is impossible due to the high rate of ^{39}Ar events. Figure 2.19 also shows the full-width at half-maximum (FWHM) transit time spread for all PMTs measured ex-situ with a tagged ^{90}Sr source. The distribution has a mean of 2.60 ns and an RMS of 0.12 ns.

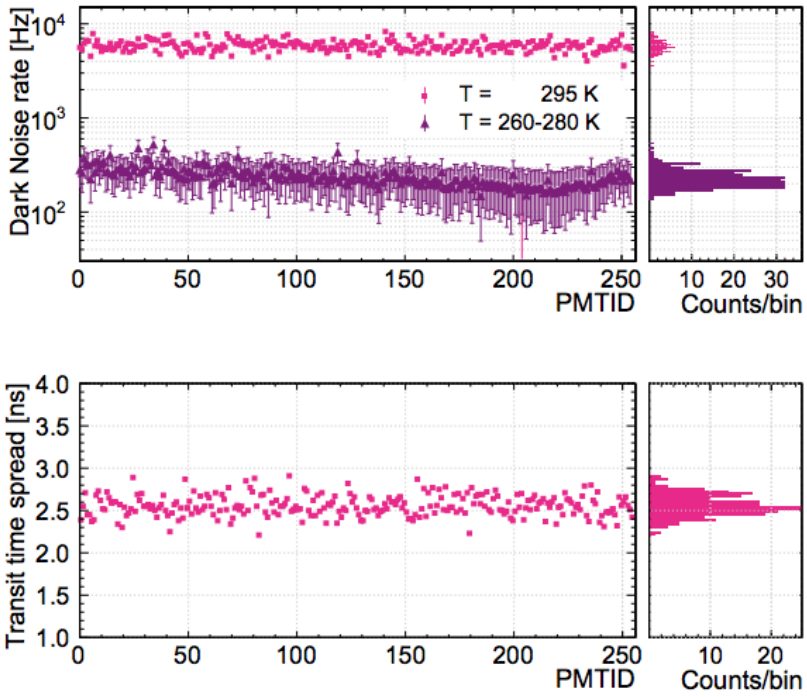


Figure 2.19: Top: Dark noise vs. PMTID for room-temperature (295 K, pink) and after filling the detector filling with argon (260 K for PMT near the bottom of the detector, large PMTID, and 280 K for PMT near the top, small PMTID). Bottom: FWHM transit time spread vs. PMTID, with a mean of 2.6 ns and an RMS of 0.12 ns [56].

One of the main unwanted background sources when working with photomultipliers comes from afterpulsing. An afterpulse occurs late in time after the initial photoelectron signal and is generated by ionisation of the residual gases present inside the tube by the accelerated photoelectrons. The R5912-HQE tubes show afterpulsing in time regions between 100 ns and 10 μs. Figure 2.20 shows the total afterpulsing probability for each PMT.

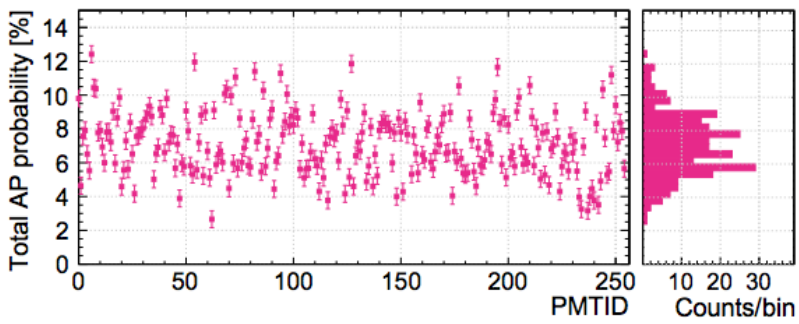


Figure 2.20: Total afterpulsing probabilities vs. PMTID [56].

2.4.4 Magnetic Field Suppression

The presence of external magnetic fields can affect the light collection of the PMTs and hence reduce collection efficiency. The effect of a 0.5 G magnetic field would result in a 25% loss in collection efficiency and 20% gain loss for the DEAP-3600 PMTs. A system to minimise this effect is in place in the form of two detector components: four field-compensating coils around the detector and individual FINEMETs shield around each PMT. The compensation coils are attached to the inner wall of the water tank at ± 0.75 m and ± 2.25 m from the equator and compensate for the vertical component of the Earth's magnetic field. The residual horizontal component is reduced by the FINEMET shielding wrapped around the PMTs (see Fig.2.10). Outside the steel shell six 3-axis fluxgate magnetometers are installed that monitor the magnetic field near the detector.

2.4.5 Steel Shell

A stainless steel shell houses the detector components and is suspended by its neck from the Cube Hall deck. It has a diameter of 3.4 m and is both water and light tight. It was fabricated from 304 stainless steel (All-Weld, Toronto, Canada) with its inner surface electropolished. It can resist a maximum pressure of 30 psig which is the pressure that could be reached as a consequence of a fracture in the AV.

2.4.6 Muon Veto System

Cosmic muons are a serious potential source of background for experiments like DEAP-3600 as they can induce a neutron that could recoil in the liquid argon volume. The easiest way for low background experiments to mitigate this source of background is to install their detector deep underground. The Cube Hall at SNOLAB, where the detector is hosted, sits at a depth of 6800 feet (2 km) in the Creighton active mine in Lively, Ontario, Canada. This corresponds to an overburden of 6010 m water equivalent [59] which results in a substantial reduction of cosmic rays flux. SNOLAB's muon flux is measured to be 0.27 $\mu/\mathrm{m}^2/\mathrm{day}$ which translates to an expected rate of < 0.042 in three years of operation without any veto system in place. To reduce this rate even further, the steel shell is housed in a 8 m diameter ultra pure water tank and instrumented with 48 veto PMTs that can tag muons by detecting Cherenkov light from their interactions in water.

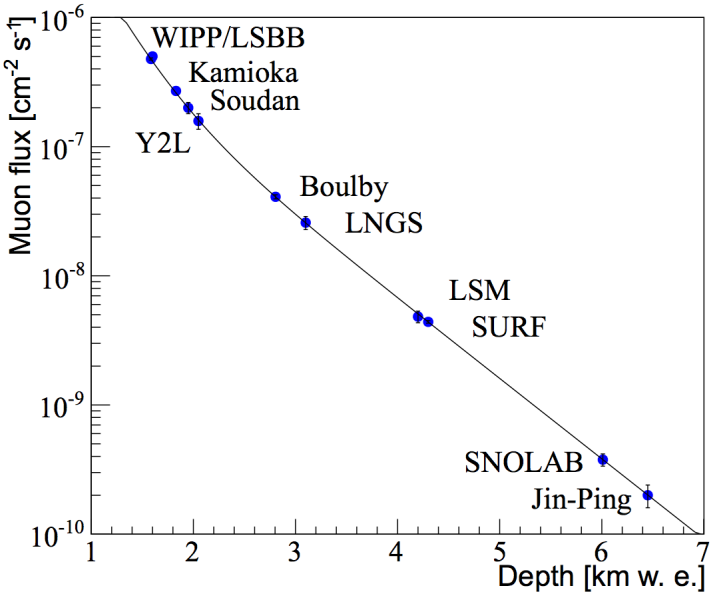


Figure 2.21: Muon flux for various underground laboratories as function of depth in kilometres water equivalent. [17].

2.4.7 Purification and Cryogenic System

Having a clean and pure target material is of crucial importance for the experiment. Impurities in liquid argon can lead to poor light yield. One wants to avoid the presence of electronegative chemical contaminants in argon as well as radioactive contamination (from radon and its progeny). The former would reduce the argon triplet lifetime, the latter would cause background events from alpha decay. To achieve this goal, a complex argon purification system is in place in DEAP-3600. It consists of five main components as shown in the schematic in Fig.2.22:

- **Process pump:** argon gas is injected into the system through this pump at room temperature (300 K) from a large storage dewar.
- **SAES getter:** A chemical argon gas purifier designed to remove chemical impurities to $<$ 1ppb.
- **Radon trap:** a custom built charcoal trap that removes radon and radioactive impurities. It takes gas at 300 K, pre-cools it at 100 K and passes it through a charcoal column. From here the gas is passed to the condenser column via a 5 μm filter.
- **Condenser column:** Liquifies the argon gas coming from the radon trap before it goes into the detector.

• **Boiler:** Vaporises the liquid argon before it gets feed back to the purification loop.

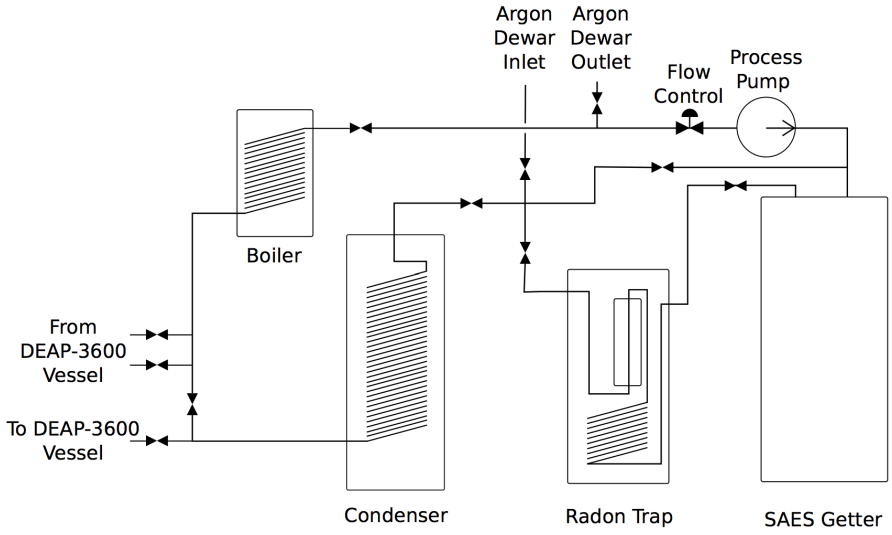


Figure 2.22: Flow diagram of the DEAP-3600 purification system. Argon gas is injected into the loop ahead of a flow controller and process pump which circulates it through a getter, radon trap, and condenser before entering the DEAP-3600 AV. The loop is closed by a boiler unit that allows argon from the AV to be fed back through the purification loop [56].

Argon manifests its liquid state at a temperature of 84-87 K and a pressure of 13-15 psia. The cryogenic system that allows to have liquid argon inside the AV consists of both a liquid nitrogen cooling system and a liquid argon purification loop as described in the diagram above. The LN_2 is stored in a dewar at the top of the Cube Hall, above the detector. As previously mentioned in Section 2.4, inside the neck there is a cooling coil (see Fig.2.9) through which the LN_2 is gravity fed from its dewar to provide cooling for the argon within the condenser column and the cooling coil itself. The convection that keeps argon in its liquid phase is regulated by flow guides that sit at the base of the neck. A computational fluid dynamics analysis was performed to optimise their design in such a way that they can efficiently guide warm argon up the neck where it gets cooled by the coils, and back into the AV once cooled.

2.5 DAQ Overview

The main challenge for the DEAP-3600 data acquisition system (DAQ) is to handle the very high rate of background events coming from ^{39}Ar beta decays while keeping useful information in a large range of energies for the dark matter search, from a few keV where

the WIMP signal is expected to be up to many MeV. Given the high rate of intrinsic backgrounds mentioned above, in order to make sure that all possible WIMP candidate events are not missed, the data acquisition rate needs to be kept below 5 MB/s. In this section a brief overview of the DAQ system will be given with a focus on the electronics and trigger system. An overall schematic of the DEAP-3600 DAQ electronics can be seen in Fig.2.23.

2.5.1 FrontEnd

The front end system includes the signal conditioning boards (SCBs) and the high voltage power supply for the inner detector and veto PMTs (MPOD). Signal from the PMTs is sent to the SCBs before being digitised. They provide the HV and decouple the signal to be sent to the digitisers, and are also coupled with a pulse pattern generator (PPG) for monitoring. There are 22 SCBs for the inner detector PMTs, one for the neck veto PMTs, and 4 for the 48 muon veto PMTs. Each SCB can handle up to 12 PMTs. Each of these 12 channels have three outputs: high gain, low gain, and a summing channel that are sent to two different types of commercial digitisers. The high gain outputs are sent to 250 MS/s CAEN V1720 digitisers that can store either full waveforms or data in n Zero Length Encoding (ZLE) mode. The low gain outputs are sent to 62.5 MS/s CAEN V1740 digitisers which can only record full waveforms. They digitise pulse that have saturated the high gain channel as well as the muon veto PMTs outputs. For each of the 22 SCBs, the 12 channels are added to create an analog sum (ASUM) resulting in 22 ASUMs from the inner detector SCBs.

2.5.2 Trigger System

The core of the trigger system is the Digitiser and Trigger Module (DTM) which makes the trigger decision. It also provides the master clock to synchronise all the SCBs, to the external calibration system and veto system and throttles data collection if the DAQ is busy. The trigger decision in made based on the 22 analogue sum of the signal coming from the inner detector PMTs. Another important component of the trigger system is the PPG which generates pulse signals fed to the PMTs through the SCBs and used for calibration and test purposes. The system is based on a set of trigger sources. These can be internal (periodic triggers, analysis of PMT signals) or external. Once the decision to trigger is made by these trigger sources, these are mapped to logical trigger outputs that

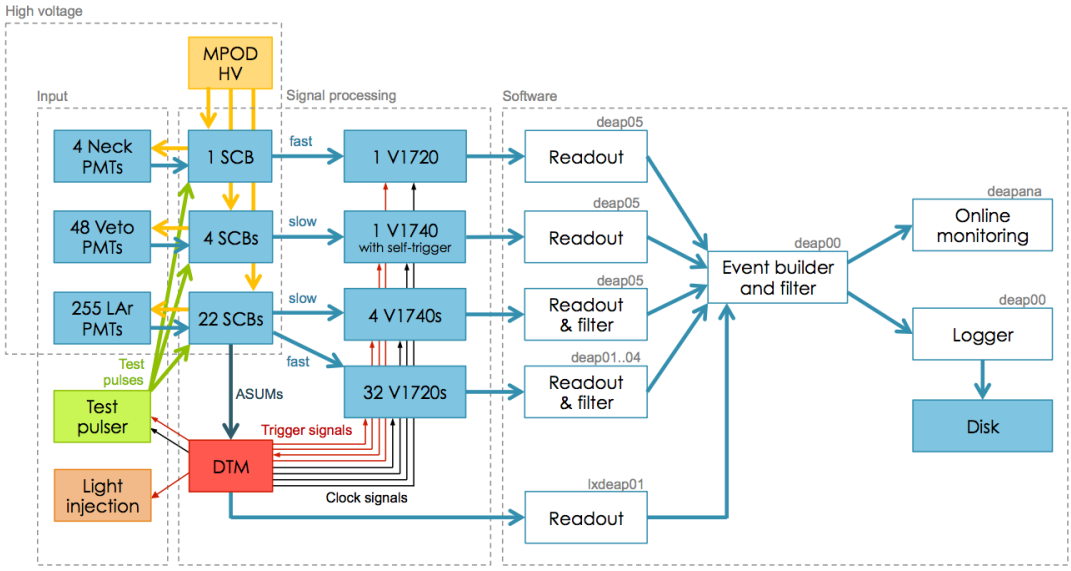


Figure 2.23: Schematic of the DEAP-3600 electronics components. The PMTs signal is first sent through signal conditioning boards to broaden the pulses in time, and then read by commercial CAEN digitisers, a fast, V1720, and a slow V1740 . The digitised signal is then read by front end PCs [56].

decide which hardware the signal should be sent to and whether or not to "pre-scale" the triggered event. Pre-scaling is a method that keeps the rate down by ignoring a percentage of events (typically background events). DEAP-3600 uses many trigger algorithms for different purposes the most important and commonly used are listed below:

- **Physics trigger:** The detector is operated mainly in this trigger mode as it constitutes the normal trigger algorithm for WIMP search runs. As such, it is designed to maximise event acceptance in the WIMP energy region of interest and PSD region of interest. The "physics trigger" adds the 22 ASUMs from the inner detector PMTs together and selects events based on their values in the two parameter space E_{prompt} - F_{prompt} that are calculated through rolling integrals over that ASUMs in two different windows: 177 ns and 3100 ns, alligned to the same start time. E_{prompt} is the total charge in the prompt window and F_{prompt} is as defined in Section 2.3. From Fig.2.24 can be seen that this E_{prompt} - F_{prompt} space is divided into 6 regions, each of them counting as a separate trigger source except for the region labelled as "X" which is discarded. To suppress the ^{39}Ar high rate, events in region "C" are pre-scaled, meaning that the digitisers read only 1% of the events falling in that region. Events falling in the other four regions are all read out.

- **Periodic trigger:** Runs at 40 Hz with test pulses injected at 1 Hz.
- **External trigger:** Is a trigger source coming from either the muon veto system or external calibration sources.

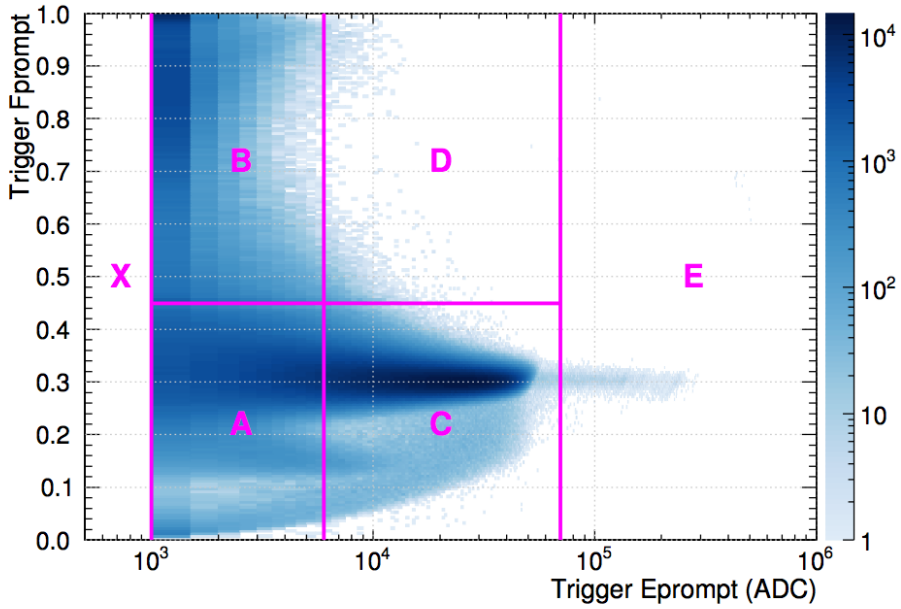


Figure 2.24: Trigger prompt energy $E_{prompt}$and trigger F_{prompt} for example data from the physics trigger. No cuts are applied and darker colours indicate more events. The six trigger regions are labelled in magenta [56].

2.6 Optical Calibration Systems

DEAP-3600 is calibrated (for different purposes) using both optical and radioactive sources. In the next two sections an overview of the hardware for these systems will be given, and a more detailed discussion of calibration results will be presented in Chapters 3, 5 and 6.

Aluminium and Acrylic Reflectors (AARF)

The aluminium acrylic reflectors (AARFs) is an optical calibration system that ensure calibration and detector stability monitoring throughout the lifetime of the experiment. A schematic representation of this system can be seen in Fig.2.26. It consists of optical fibres attached to acrylic studs, coated in aluminium, bonded to the outer surface of a light guide. When light generated by 445 nm LEDs is injected through the optical fibre, the AARF unit guide reflects it at a 90 degree angle and guides it towards the face of the

PMT. The majority of this light is collected by the PMT where the AARF is installed, while a fraction gets diffused into the acrylic vessel. A total number of 22 AARF units are installed on the detector, 20 uniformly distributed across the PMT array and 2 in the neck (see Fig.2.27.) This calibration system allows an accurate timing calibration of the 255 PMTs as well as monitoring the detector response and its stability over time. It represents the primary optical calibration during operation and is used to calculate and monitor the PMT single photo-electron (SPE) charges and their efficiencies.

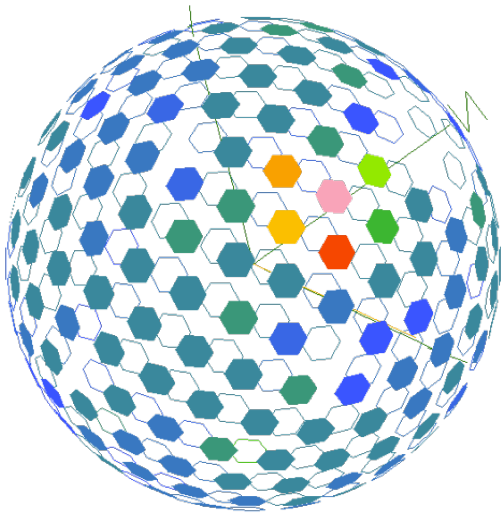


Figure 2.25: Average calibrated charge for the first 1000 events in an AARF run. Plot made with the DEAP-3600 deapdisplay software.

Optical Diffuser Flask

The Optical diffuser flask, also referred to as the "laserball", is the only calibration source deployed inside the AV prior to filling with liquid argon. It was designed to provide an isotropic source of light of different wavelengths from various positions. Lasers with wavelengths of 375 nm, 405 nm and 445 nm were used to illuminate the detector. The 375 nm wavelength was chosen because it activates the TPB and provided useful data to study its properties and thickness. The laserball was deployed before filling but after TPB deposition. For timing characterisation of the PMTs (discussed in the next chapter) the 445 nm laser was used as at this wavelength the TPB does not get activated and so the effect of the TPB on the timing offsets of the PMTs could be decoupled. Fig.2.28 shows the laserball flask. PMT occupancies for a run with the laserball at the centre of the detector and at four different phi angles can be seen in Fig.2.29. These figures display a

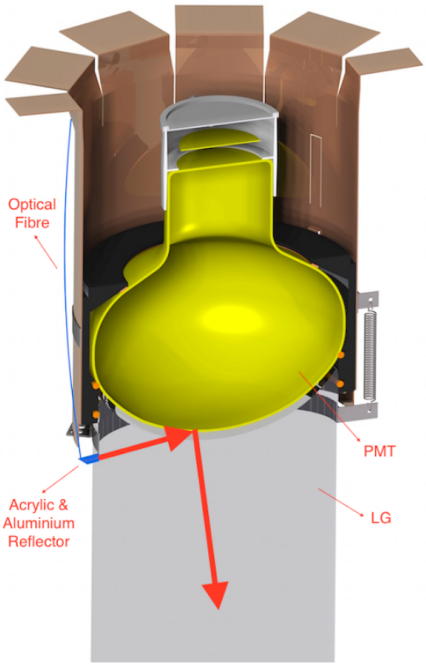


Figure 2.26: Pictorial representation of the AARF system with its optical fibre, the Acrylic and Aluminium Reflector hosted by the PMT, the PMT itself and the lightguide.

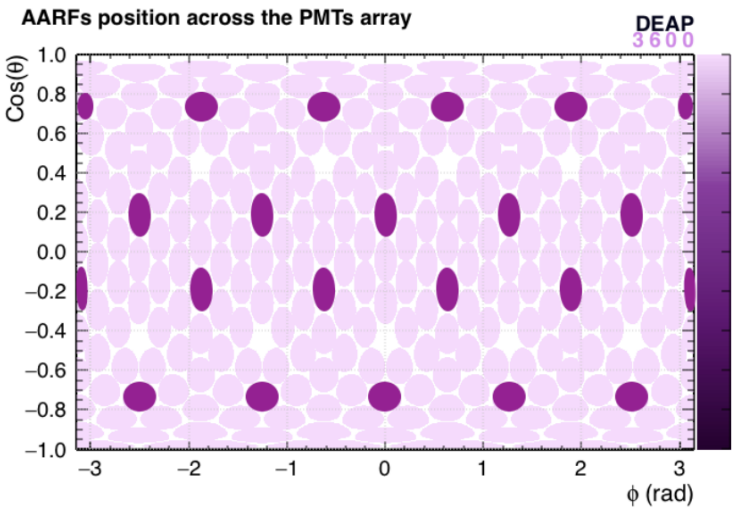


Figure 2.27: Colour map plot showing the position of the 20 detector AARFs (in purple) on the PMTs array, in the $\phi - \theta$ plane. Plot courtesy P. Giampa.

bottom-top asymmetry that can be explained by the material properties of the flask and its geometry. PMTs at the bottom of the detector received more light than the ones at the top. The flask was made from perfluoroalkoxy alkane (PFA) and filled with 40 μm glass beads suspended in silicone gel, with the density of glass beads near the top higher than the bottom. This causes the θ non-uniformity observed in Fig.2.29. To decouple the effect of this non-uniformity the source was deployed in three positions along the z vertical axis that coincides with the centre of the detector neck: $z = 0$ and $z = \pm 55$ cm. Possible rotational ϕ non-uniformity due to imperfections in the PFA were also taken into account and thus data were taken in four position in ϕ 90 degree apart from each other.

Figure 2.28: Image of the PFA laserball flask.

2.7 External radioactive sources

The DEAP-3600 experiment makes use of two external radioactive calibration sources. A tagged 0.999 MBq ^{22}Na gamma source (created April 2012) is used to monitor the energy scale and resolution of the detector. To characterise the response to neutron-induced nu-

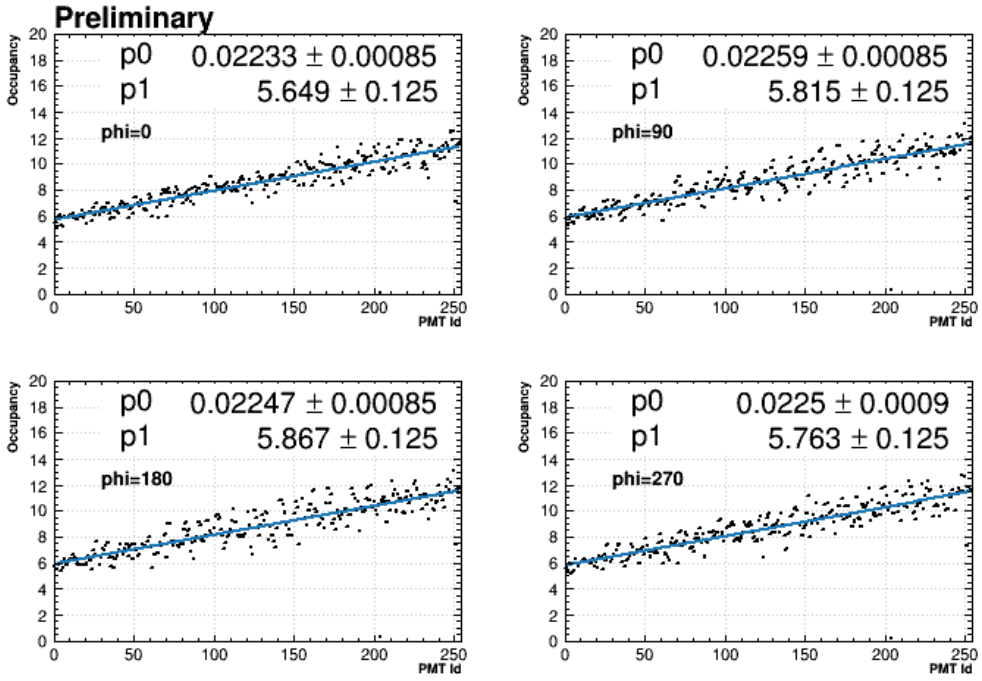


Figure 2.29: Laserball PMT Hit occupancies in the four laserball azimuthal rotations/orientations at z=0 position. Plot courtesy R. Mehdiyev.

clear recoils a tagged 74 MBq americium-beryllium (AmBe) source is periodically deployed externally to the steel shell [60].

Gamma Calibration Systems

The ^{22}Na source is a positron emitter. It emits a positron followed by a 1.27 MeV γ ray. A pair of 511 keV photons is also emitted as a consequence of the positron annihilation and these photons are used to tag the source. As such, the source is see sandwiched between two 20 mm diameter, 20 mm long LYSO (Cerium-doped Lutetium Yttrium Orthosilicate) crystal scintillators coupled to a PMT (Hamamatsu R9880U). Figure 2.33 shows the gamma canister and its components. The source is deployed through the circular high density polyethylene tube wrapped around the detector (Cal F in Fig.2.30). This allows a uniform calibration of the detector and precise position reconstruction studies. The canister is deployed using an automated pulley and carriage system driven by a Mclennan 34HSX-208E stepper motor controlled by a 818 Mclennan SimStep Controller (see Fig.2.32). Ex-situ measurements have shown that the source can be deployed with an uncertainty of approximately 1 cm.

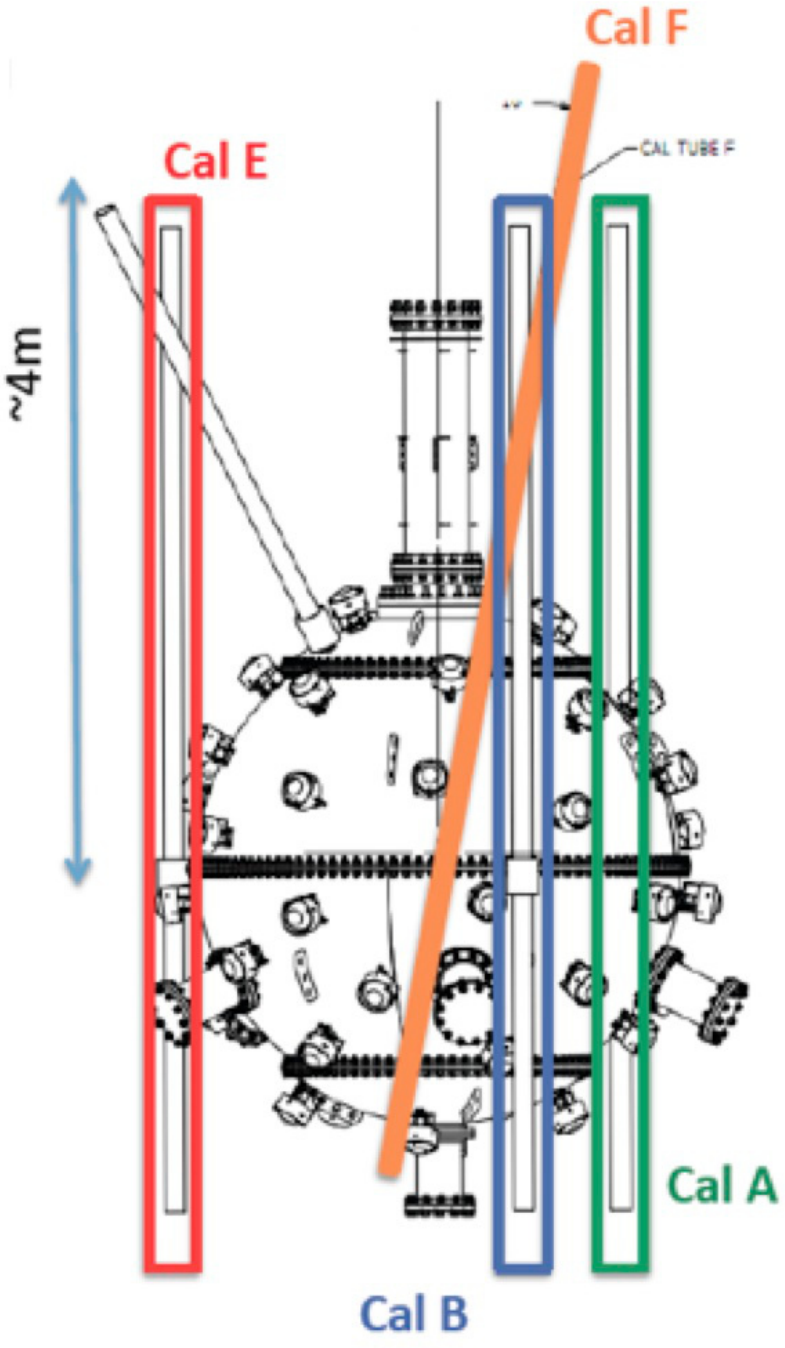


Figure 2.30: Calibration tube A, B, E, and F. The 3 vertical tubes are mainly used for AmBe calibration runs. The circular F is used for ^{22}Na calibration runs.

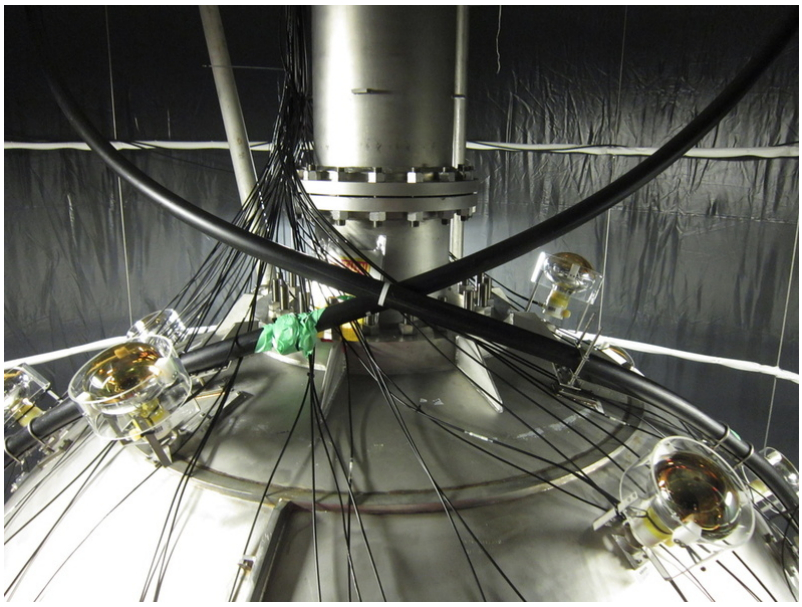


Figure 2.31: Close-up picture of calibration tube F near the detector neck.

Figure 2.32: Picture of the gamma calibration deployment system for Cal F tube (two racks on the left) and neutron calibration deployment system (taller rack on the right).

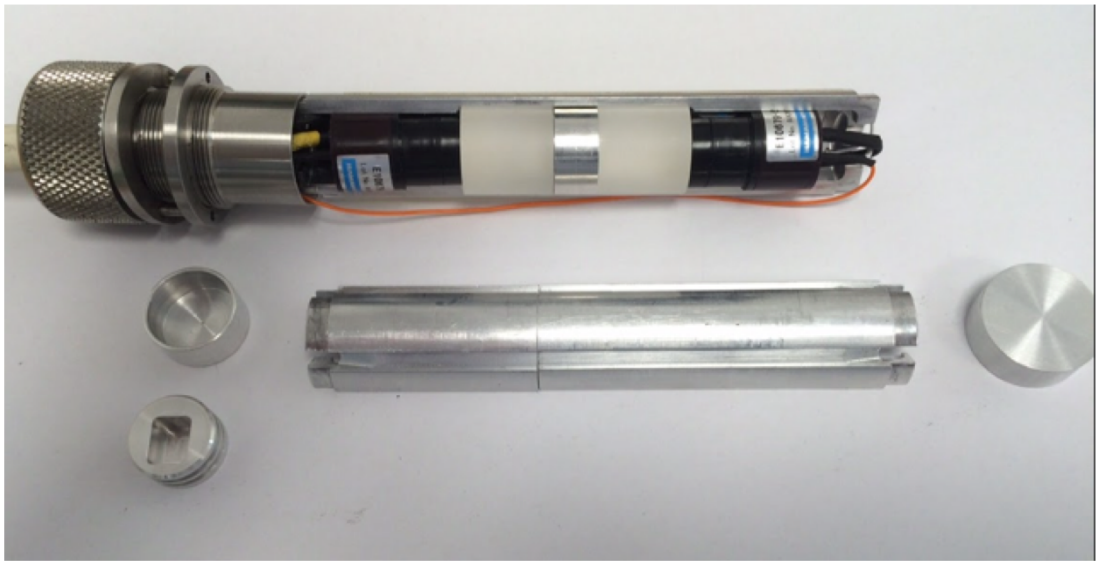


Figure 2.33: Image of the ^{22}Na canister consisting of a two-part outer casing. Inside the outer casing sits two Hamamatsu PMTs and bases, two 3.5kBq LYSO crystals and one 1 MBq ^{22}Na Eckert and Ziegler source contained in two factory encapsulations and two additional soldered copper encapsulations.

Neutron Calibration Systems

The AmBe neutron source is contained in a canister in the same fashion as the ^{22}Na (see Figure 2.35). Two 40 mm diameter, 51 mm long Ametek-packaged NaI crystal are used as scintillators and each of them is coupled to a 38 mm ETL 9102 PMT. During calibration runs the source is deployed in one of the three vertical calibration tubes (A,B and E in Fig.2.30) through a pulley and carriage system similar to the one used for the gamma source. The tag PMTs are powered by a Cockcroft-Walton high voltage generator allowing the PMT voltage to be driven from a 5 V source.

2.8 Backgrounds

Everything that can mimic a WIMP-like event in DEAP-3600 is a source of background that needs to be addressed and reduced either intervening at the hardware level or with software tools in the analysis chain. In this section the most conspicuous sources of background and the work done to minimise them will be discussed. The expected rate of residual backgrounds in three years of operation for these sources is summarised in Table 2.3.

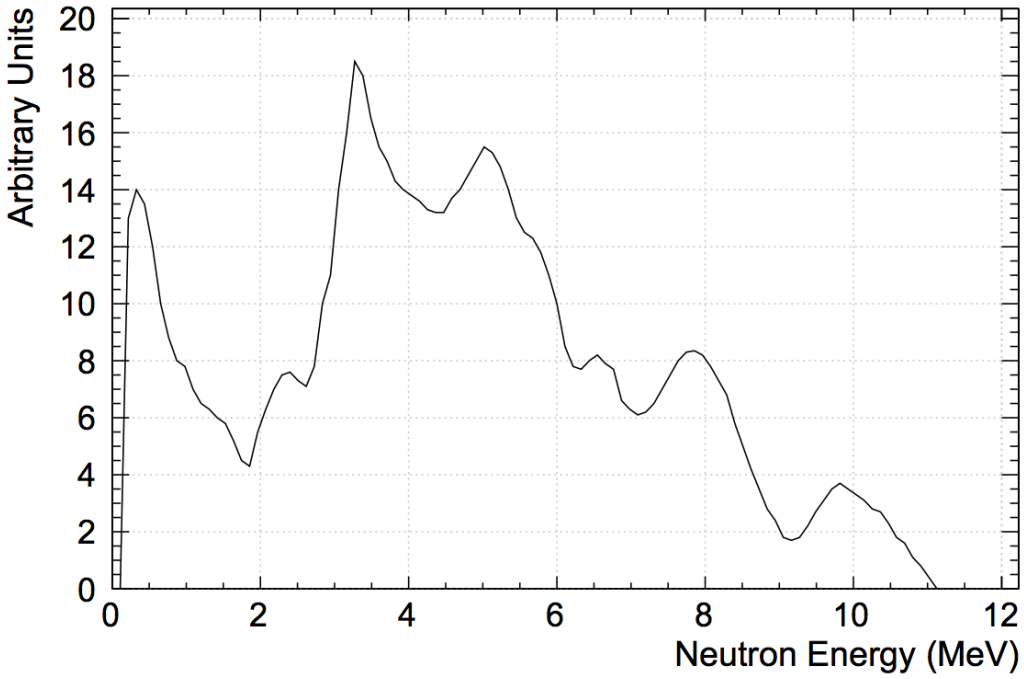


Figure 2.34: AmBe spectrum used in MC generation plotted in arbitrary units [61].

Background source	Events in ROI	Fiducial events in ROI
^{39}Ar	1.6 x 10^9	<0.2
Neutrons	30	<0.2
Surface alphas	150	<0.2
Total		<0.6

Table 2.3: Background budget for the DEAP-3600 experiment, for 3 years exposure and 1 tonne fiducial mass.

2.8.1 ^{39}Ar

The most prominent source of backgrounds comes from an isotope of argon itself: ^{39}Ar. Produced by interactions of cosmic rays with ^{40}Ar, it is an unstable β emitter isotope with a half-life of 269 years and endpoint energy of 565 keV [62]. It produces β's through:

$$^{39}\mathrm{Ar} \rightarrow {}^{39}\mathrm{K} + e^- + \nu_e \tag{2.13}$$

with a 1 Bq/kg activity in natural argon. This means that in DEAP-3600 the expected number of electronic recoils event from this source is about 3.4×10^{11} [63] of which $\sim 10^9$ would leak into the dark matter search region of interest in three years of operation without

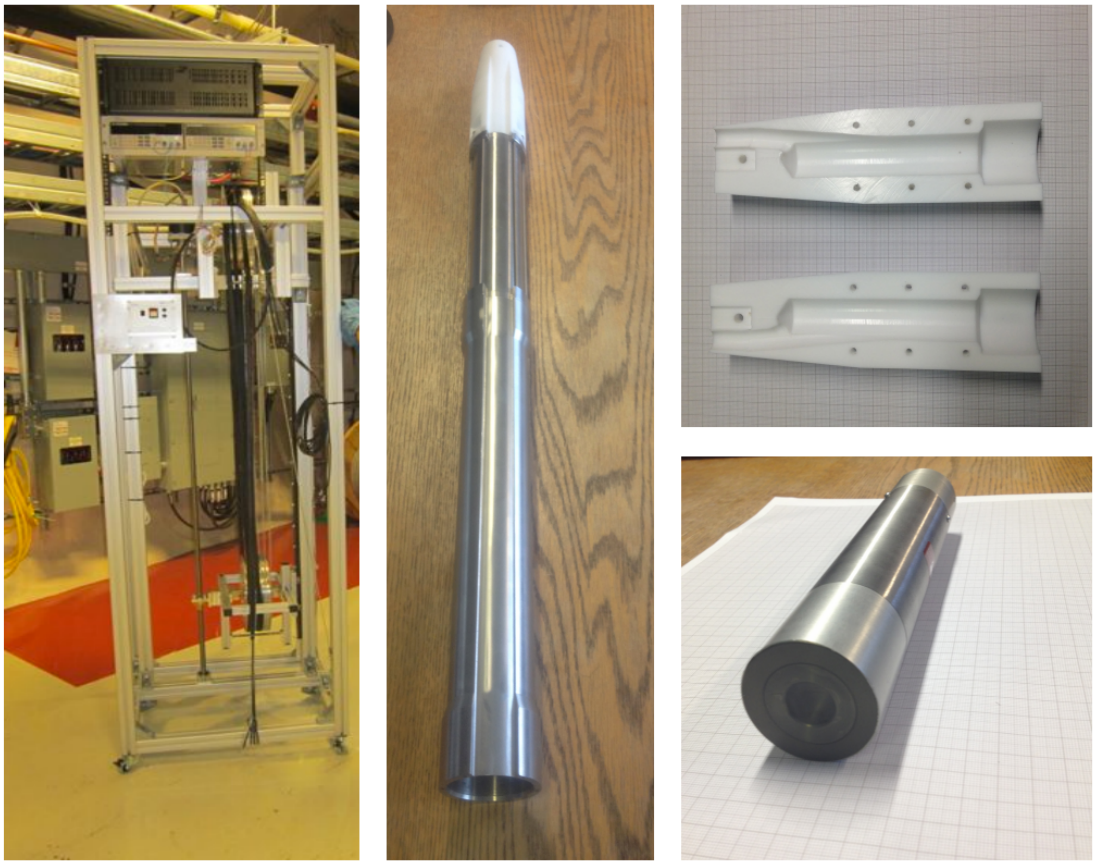


Figure 2.35: Left: neutron calibration deployment system. Middle: neutron source canister. Right: Canister support bracket (top) and PMT-NaI crystal assembly (bottom). The two PMT-NaI crystal assemblies sit back-to-back housed within the neutron source canister.

using the mitigation power of the pulse shape discrimination of the scintillation light. The projected expected number of events in the fiducial ROI is < 0.2 after PSD.

2.8.2 Neutrons

Neutrons are a very problematic source of backgrounds for dark matter search experiments as they can mimic exactly the signal expected by a WIMP interaction. Neutrons will scatter multiple times in liquid argon but a single elastic scatter would create a nuclear recoil event which would be indistinguishable from a WIMP interaction. The main sources of neutrons in DEAP-3600 come from cosmogenics, lab rocks and detector components. To mitigate this backgrounds, a lot of effort has been invested in shielding and detector materials selection. Neutrons from cosmogenic activity (induced by high energy muons) are handled by placing the detector underground as already discussed in Section 2.4.6. Neutrons coming from rocks activity surrounding the detector are mitigated through the water veto system. The most prominent and tough to minimise source of neutrons come from detector itself such us the PMTs, the neck, the steel shell etc, due to (α, n) reactions from uranium, thorium and radon contamination. Simulations have shown that the PMT components such as glass and ceramic are main source of neutrons, with the PMTs internal glass being responsible for generating $\sim 2 \times 9^4$ neutrons/year [64]. These neutrons are mitigated through shielding with the AV, light guides and filler block giving a 10^{-6} level of attenuation by thermalising them.

2.8.3 Surface Alphas

Another potential source of background that can mimic WIMP-like signals come from decays of radon and its progeny on the inner surface of the AV. Alphas with energies between 5 and 8 MeV are emitted in the decay chains of these long-live isotopes. Even though these alphas are well above the energy region for the WIMP search, they can generate events reconstructed at lower energies (near the dark matter search ROI) if they only deposit a fraction of this energy in the liquid argon. Surface alphas background can be generated by decays in the liquid argon bulk, TPB decays and acrylic decays. In the first case both the alpha particle emitted and the recoiling nucleus daughter deposit their full energy ending up being identified in the same F_{prompt} region as WIMPs but at higher energy. Alpha decays can occur in the TPB due presence of the same contaminants as the acrylic vessel. To mitigate this source of background a resurfacing device was run for

198 hours inside the acrylic vessel prior to filling and TPB deposition. The goal of this custom built sanding robot was to remove the inner layer from the AV in order to reduce surface contaminations. A schematic of the device can be seen in Fig.2.36. Equipped with two rotating sanding arms, it was 5.5 meters tall and featured a system to flush water during sanding to remove sediments. It was deployed through the AV neck and it removed an estimated 500 $\pm$ 50 μm of acrylic from the inner surface of the AV. By removing this thickness the ^{210}Pb surface backgrounds were reduced down to near the AV assay upper limit for the bulk radiogenic contamination of 2.2 $\times 10^{-10}$ g/g as shown in Fig.2.37 [56].

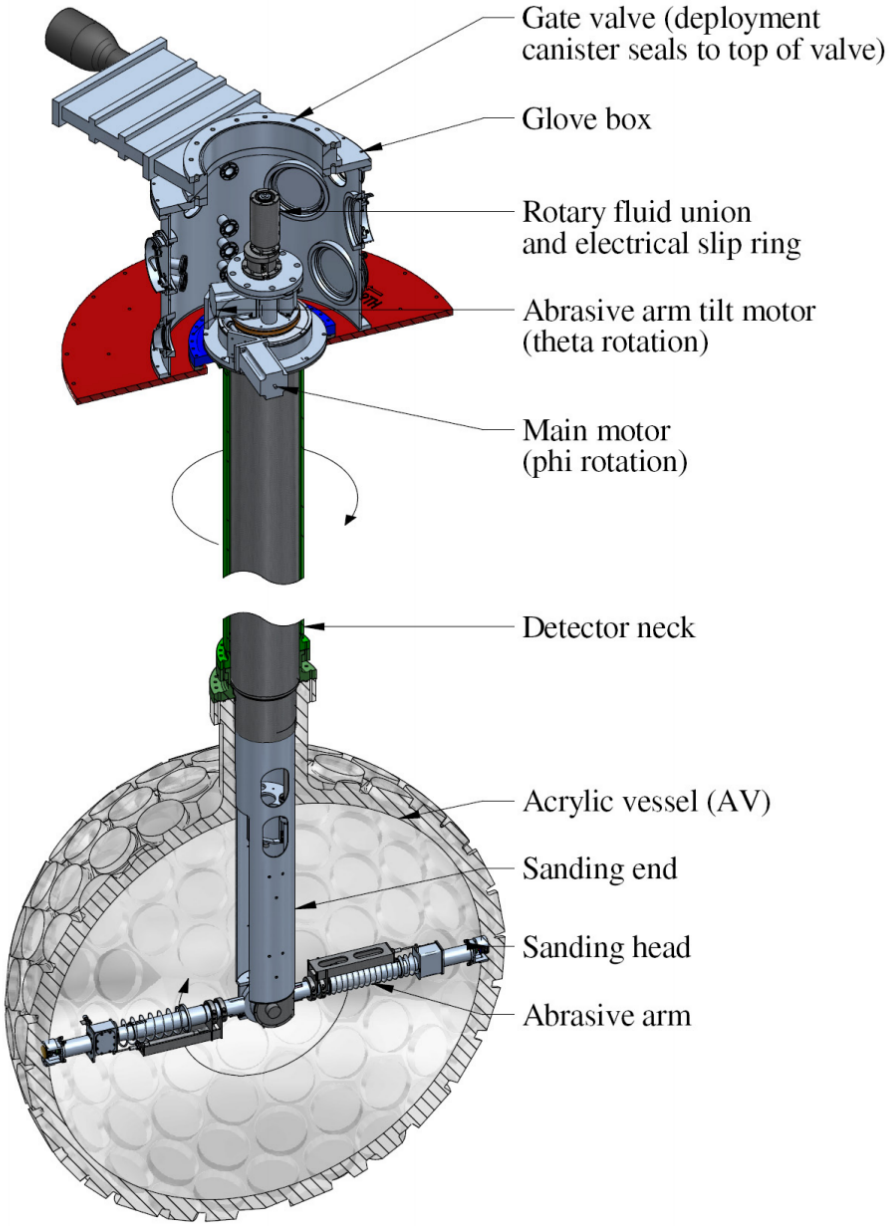


Figure 2.36: Overview of the resurfacer device and its main components when deployed inside the acrylic vessel [56].

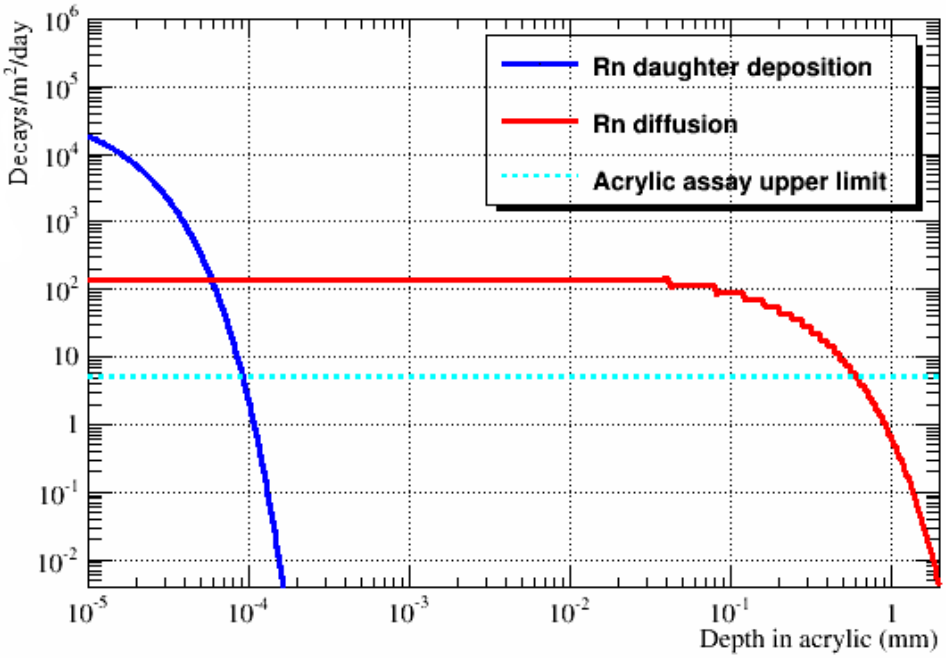


Figure 2.37: Calculated ^{210}Pb alpha activity in the AV before resurfacing. The blue curve shows the activity after radon-laden air exposure whilst the activity due to radon diffusion into acrylic is shown in red. The 500 ± 50 μm of acrylic removed by the resurfacer reduces the activity of ^{210}Pb down to near the assay upper of $2.2 \times 10^{-19} g/g$ limit represented by the cyan line [56].

e

Chapter 3

Optical and timing characterisation with the laserball source

As discussed in Sect. 2.6, an Optical Diffuser Flask (Laserball) was deployed inside the AV prior to argon fill. This calibration source was mainly used for PMT characterisation and study of theTPB properties. Results of a study using laserball data, aiming to remove the PMTs channel-to-channel time offsets and extract T_0 is presented in this chapter.

3.1 Introduction

Timing is of crucial importance to achieve accurate event reconstruction and to reduce the leakage of background events into the fiducial volume. The laser sources deployed within the AV prior to filling provided very fast and narrow pulses allowing a jitter analysis at the sub-nanosecond level. Fig.3.3 shows laserball data for a run where the source was deployed at the centre of the detector with a 445 nm laser head installed. In particular, the plot shows the distribution of light across the detector in term of a quantity, occupancy, defined as the fraction of the total light flashes seen by a given PMT, versus the PMT ID. Low IDs correspond to PMTs near the neck of the detector, high IDs to PMTs at the bottom of the AV. It should be noticed that due to multiple reflections and the laserball geometry itself (the flask neck acting as a shadowing component), there is an asymmetry between the top and bottom PMTs, with the bottom ones seeing more light. Two PMTs were off when these data were taken. In this chapter we will discuss the analysis done to determine the channel-to-channel timing variations due to cable length and electronics channel differences using the laserball data at the centre of the detector. These calibration allows DEAP-3600

Figure 3.1: Picture of the 375 nm laser system [65].

Figure 3.2: Picture of the flask with the laser on.

to achieve sub-nanosecond timing resolution. A timing calibration to correct for timing offsets between channels is done on a run by run basis injecting test pulses through a Pulse Pattern Generator (PPG). PPG timing calibration allows for the correction of the three following offsets:

- A 16 ns jitter caused by some boards that receive the trigger signal too late. It is the same for all 8 channels on a board.
- Channel-to-channel offsets on a board, including 8 ns jumps when the DTM/V1720 clocks re-synchronise. These offsets are different for each channel on the board.
- Fixed offsets caused by the time taken by the test input signals to propagate through the SCB.

An example of pulses alignment across the detector prior to any PPG or laserball

Wavelength(nm)	Wavelength tolerance(nm)	Spectral half-width(nm)	Pulse width(ps)	
			max.	typ.
375	$< \pm 10$	<10	80	50
405	$< \pm 10$	<10	100	60
455	$< \pm 10$	<10	100	70

Table 3.1: Specifications of the PLP-10 laser diode head used for the laserball calibration campaign [65].

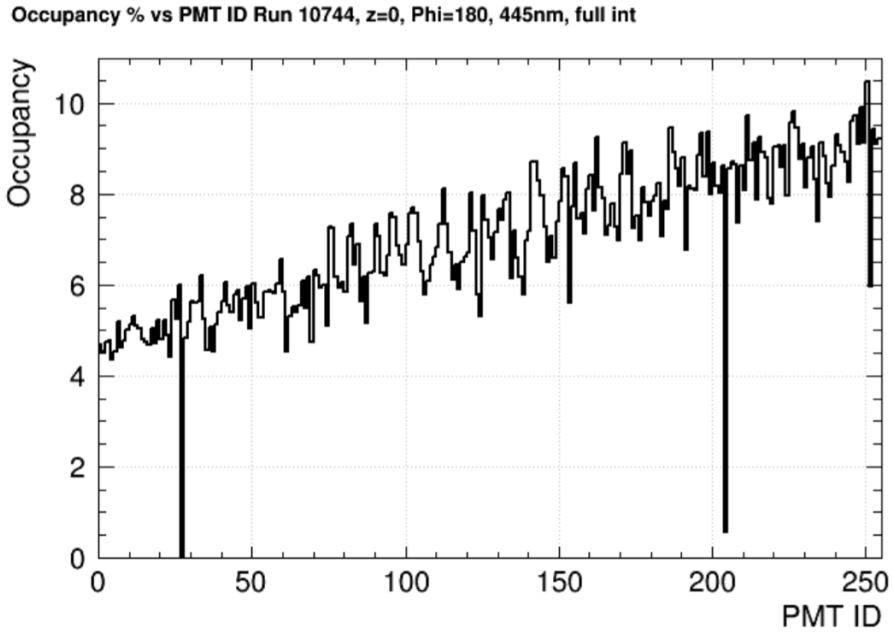


Figure 3.3: Occupancy for a 445 nm laserball run. The channel with 0 occupancy is a dead PMT and PMT ID 204 was off during this run.

calibration are performed can be seen in Fig.3.4 where the time of all the pulses in a run with the 445 nm wavelength laserball at the centre of the detector is shown versus PMT ID. If all channels were aligned one would expect to see a straight line at around 6400 ns, the event trigger time.

In the next sections, the methodology used to extract residual offsets for PPG corrected pulses will be described. Offsets extracted from this study are stored as correction parameters in the database and applied when raw data are calibrated for each run.

3.2 Offsets and Timing definitions

Fig.3.9 shows the time distribution for a 445 nm laserball run at z = 0 cm. Although the laser generates very narrow pulses, light in the detector can scatter several times before being collected by the PMTs. This effect can be clearly seen in the tail of the

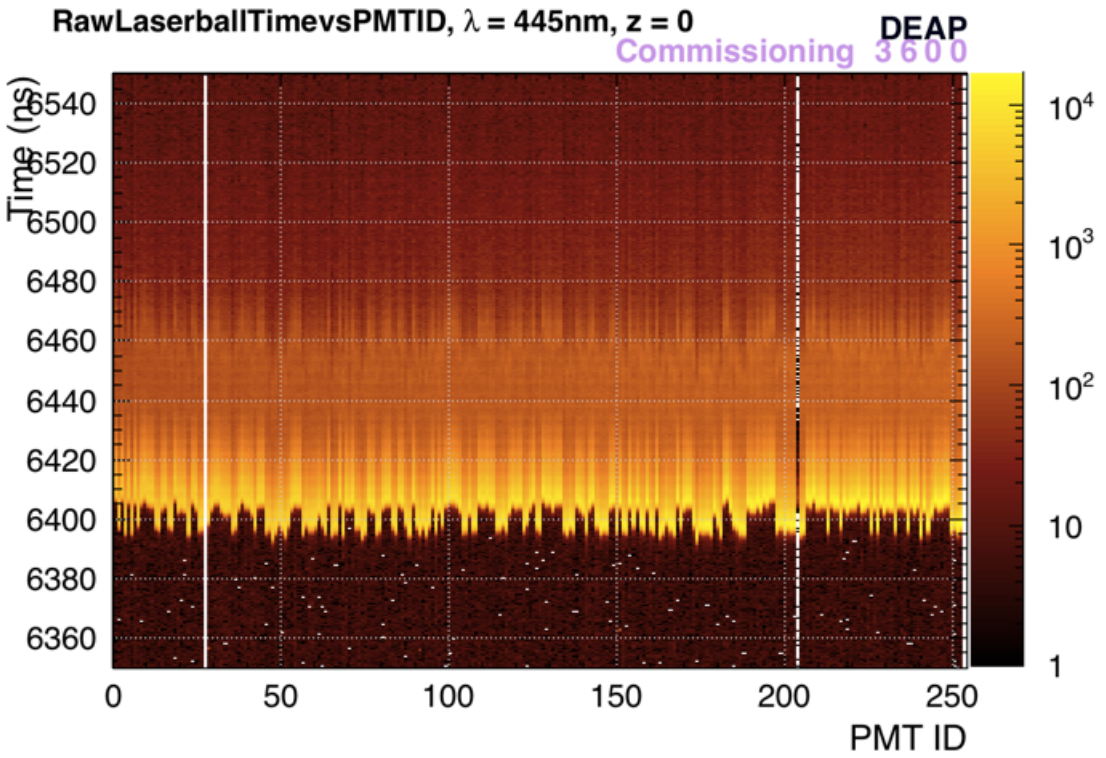


Figure 3.4: 'Raw' time versus PMT ID for a laserball run at z = 0 cm before correcting for any offsets. Pulses are highly misaligned due to PMTs drift differences, cable length differences and time jitter between the front ends of the V1720 and the final digitisation. Two channels were off (white straight lines) during these runs.

time distribution in Fig.3.9 for late times after the main peak. Time t_i for each PMT is calculated as [66]:

$$t_i = t_0 + \frac{\vec{x_i} - \vec{x_0}}{c} + t_{0,i} + t_{V1720,i}, \tag{3.1}$$

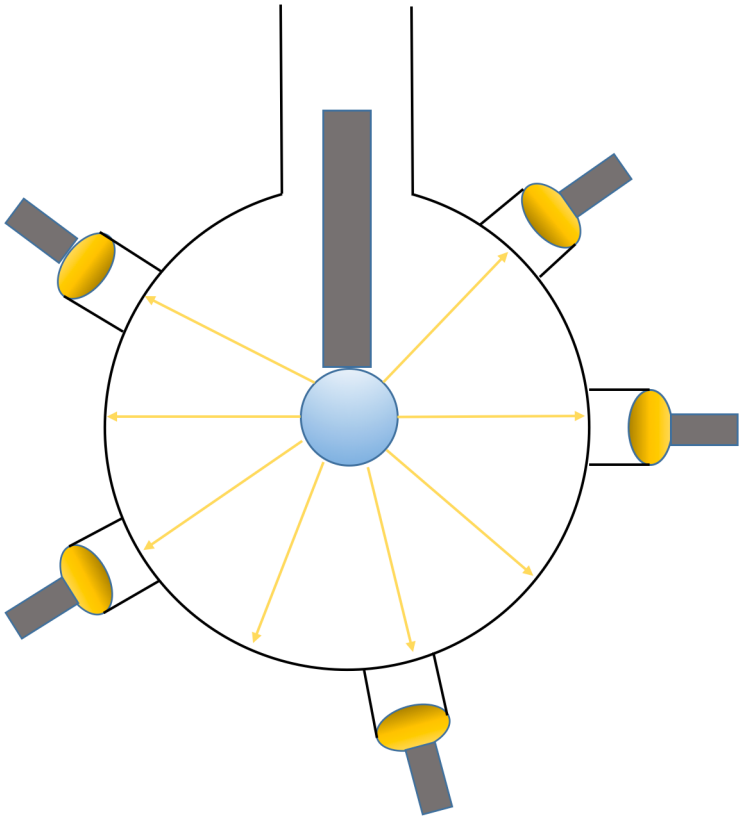


Figure 3.5: Pictorial representation of a laserball data collection run in DEAP-3600 with the laser placed at the centre of the detector. In this configuration the PMTs are hit by the light all simultaneously.

where $\vec{x_0}$ is the position of the event (a calibration source for example), $\vec{x_i}$ the position of the PMT i, t_0 the time of the flash of light and $t_{0,i} + t_{V1720,i}$ is a time offset for each particular channel. In particular, $t_{0,i}$ accounts for channel-to-channel PMTs drift differences and cable length differences and $t_{V1720,i}$ contains time jitter between the front ends of the signal conditioning boards, V1720 and the final digitisation. As already mentioned, the trigger system includes a separate Pulse Pattern Generator (PPG) board, which provides synchronised pulses to all the SCBs boards when it receives a signal from the DTM. The SCBs distributes this signal to all channels. For a PPG run, the time of a pulse for a PMT i is given by:

$$t_i = t_0^{'} + t_{V1720,i}, \tag{3.2}$$

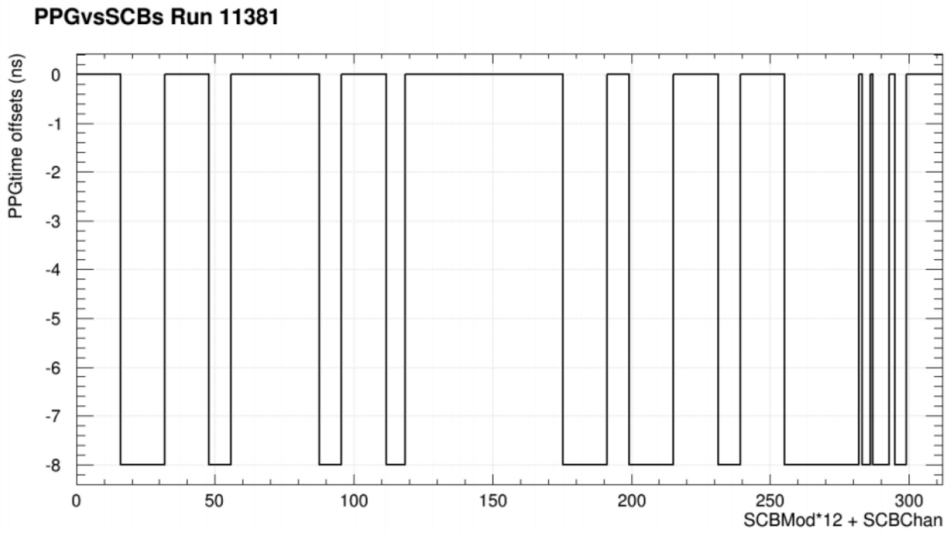


Figure 3.6: The trigger system includes a Pulse Pattern Generator (PPG) board that allows to send synchronised pulses to all the SCBs when it receives a signal from the DTM. This plot shows the offsets correction extracted by using PPG pulses versus SCBs channel for a laserball run at z = 0 cm.

where $t_0^{'}$ is a common offset and $t_{V1720,i}$ includes propagation delays between SCBs and V1720 channels and channel to channel digitiser time shifts as already mentioned. By using PPG runs one can measure $t_{V1720,i}$ (see Fig.3.6).

Once corrected for timing offsets discussed in the introduction and described by Eq. 3.2 by using dedicated PPG runs, in the hypothesis of a central light source ($\vec{x_i} = \vec{x_0}$) the only unknown quantity to extract t_0 is $t_{0,i}$, i.e. residual channel-to-channel offsets due to difference in the cables length.

3.3 T_0 extraction

The problem to solve and the goal of this study are summarised in the pictorial schematic in Fig.3.7. Here on the left side raw pulses are represented and reflect the scenario showed in Fig.3.4 where there are consistent misalignment between all channels. On the right side is the ideal scenario where all pulses are perfectly aligned after correcting for PPG offsets and for the residual offsets extracted with the laserball from this study. It will be shown that indeed a combination of these two methods allows to get an alignment at the sub-nanosecond level.

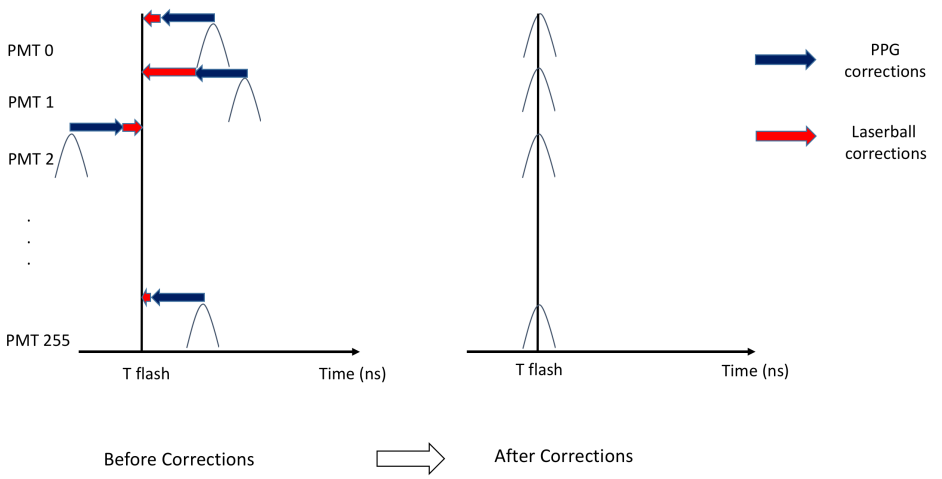


Figure 3.7: Pictorial representation of the study and its goal. On the left side four generic 'raw' pulses are represented and they are few ns off T_{flash}. The goal of the study it to align the pulses from all the PMTs to the time of flash (T_0) by correcting for offsets extracted through the PPG (blue arrow) and residual offsets extracted through the laserball at the centre on the AV (red arrow).

Methodology

In order to extract these residual offsets, raw laserball data (position z = 0) were first corrected for PPG offsets, resulting in the overall time profile shown in Fig.3.8. It is evident that even though PPG correction have removed part of the misalignment, pulses still show jitter that can be removed by using the central position laserball data. In order to do this, five laserball runs at z = 0 cm and different angles ϕ have been taken into consideration.

For each of these runs and for each PMT_i, the quantity ΔT_i, defined as:

$$\Delta T_i = \bar{t_i} - \frac{1}{N} \times \sum_{i=1}^{N} t_i \tag{3.3}$$

has been evaluated. In eq. 3.3 $\bar{t_i}$ is the difference between the mean of the peak pulse time distribution for each PMT and the second term is the mean time for all PMTs, with N being the number of PMTs (255). As already mentioned and seen from the time distribution shown in Fig.3.9, the light can bounce around many times inside the detector before being recorded by the PMTs generating a "late" contribution to that distribution. In order to avoid including late light, a decision to select only prompt pulses in a fixed time window of 20 ns was made for this study. In particular, only pulses with 6400 ns

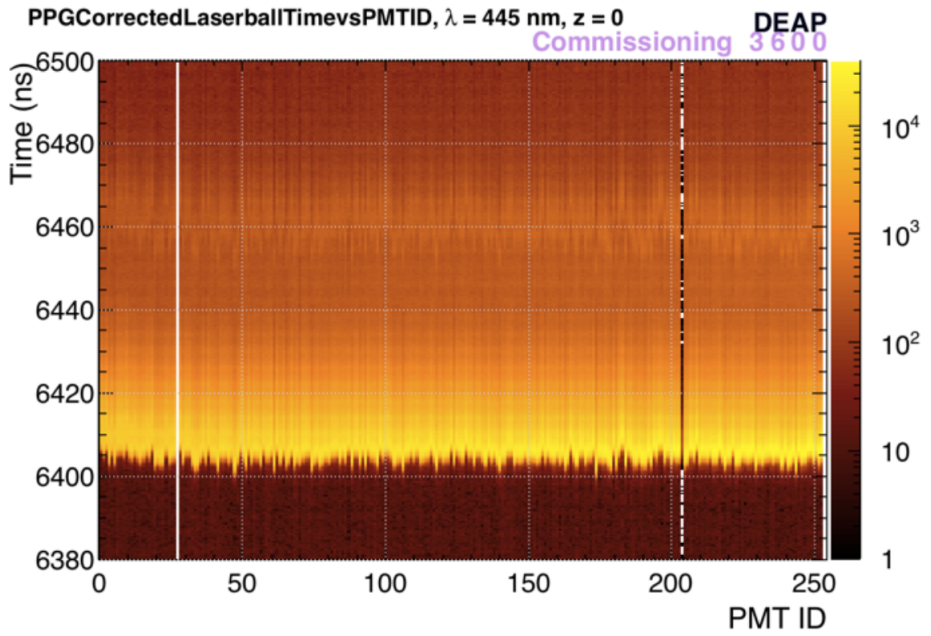


Figure 3.8: Time versus PMT ID after correcting for PPG offsets. The jitter is reduced compared to what seen in Fig.3.4, but residual offsets (due to differences in the cable length) are still present and can be diminished using the laserball data.

$< t_i <$ 6420 ns were taken into account. This is shown in Fig.3.11, where the prompt light for a laserball run already corrected for PPG offsets is plotted. By computing for each of the five runs the ΔT_i as defined in Eq. 3.3 it is possible to determine the residual offsets for each run. This is shown for a run in Fig.3.10. If all the pulses were aligned, one would expect to see a flat distribution centred around $\Delta T_i = 0$ for this variable when the laser is placed at the centre of the detector. Hence, the deviations from zero of ΔT_i across the detector are the residual offsets one needs to apply in order to get aligned pulses.

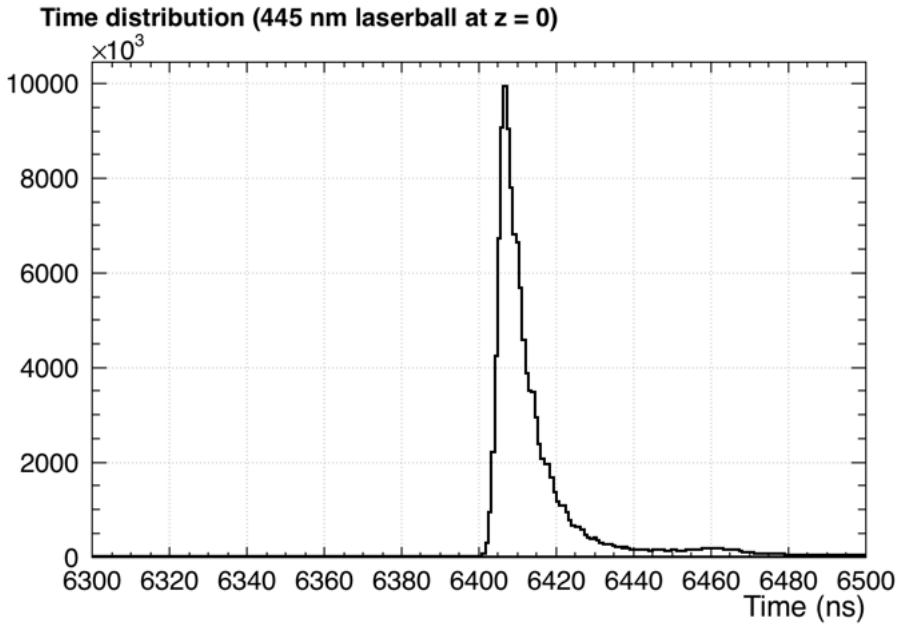


Figure 3.9: Time distribution for a $\lambda = 445$ nm and $z = 0$ cm laserball run. The broadening of the distribution and the visible tail are due to late light that reach the PMT after having being reflected multiple times in the AV.

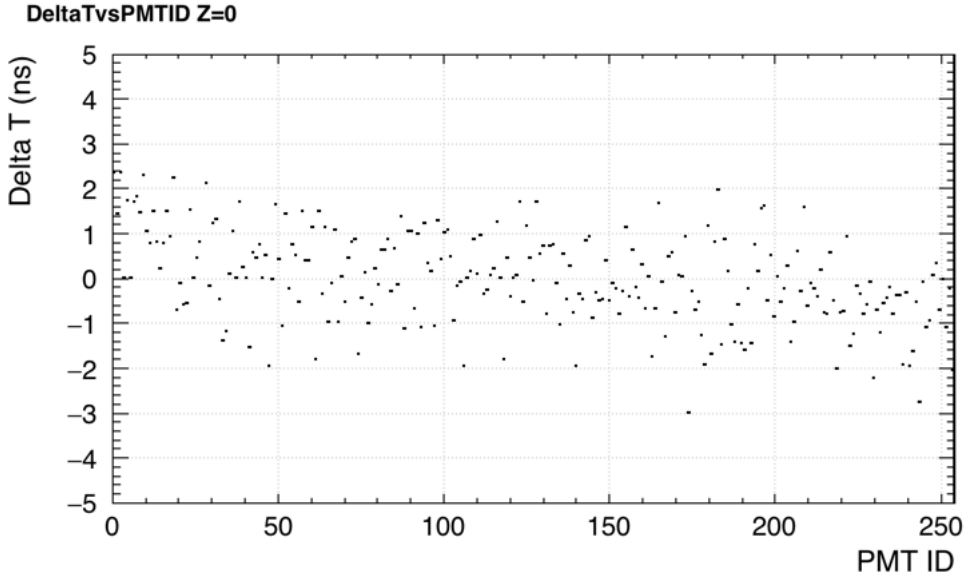


Figure 3.10: Residual offsets extracted by using Eq. 3.3 for a single laserball run.

As already mentioned, five runs at different rotation angles ϕ were taken at z = 0 cm. Fig.3.12 shows the offsets (ΔT from eq. 3.3) for 2 generic PMTs versus run number. The residual offsets applied to raw data were extracted by calculating the weighted mean value for each PMT over the five runs. An offset table for each of the 255 PMTs is obtained in this way. This table is stored in the DEAP-3600 database and is applied when calibrating

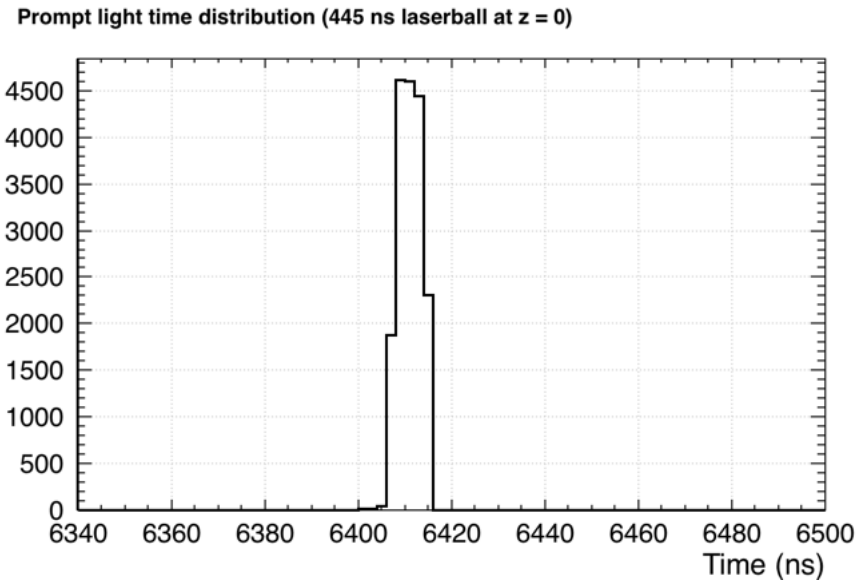
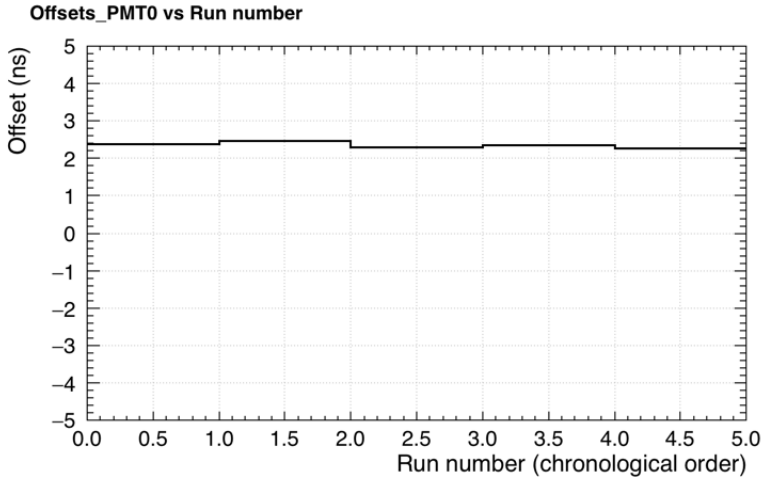


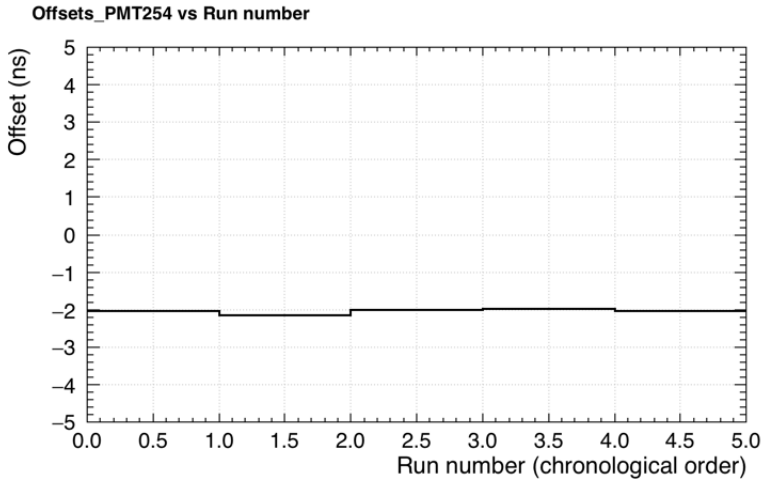
Figure 3.11: Time distribution for a $\lambda = 445$ nm and $z = 0$ cm laserball run including 'prompt' light only.

raw data.

Finally, the overall timing distribution is plotted for 255 PMTs in the two dimensional plot of Fig.3.13 showing the overall effect of the offset corrections across the detector. A projection within $\pm 1\sigma$ of the gaussian peak of each channel after calibration is shown in Fig.3.14 for all the PMTs in black and for a generic PMT (ID 0) in red. The RMS of all PMTs is 0.73 ns.



(a)



(b)

Figure 3.12: Residual offsets extracted by using Eq. 3.3 for a two generic PMTs (PMTID 0 and PMTID 254) for the five laserball runs used in this study. The final table with offsets for data corrections were extracted by getting the weighted mean value for each PMT from the other 223 similar plots.

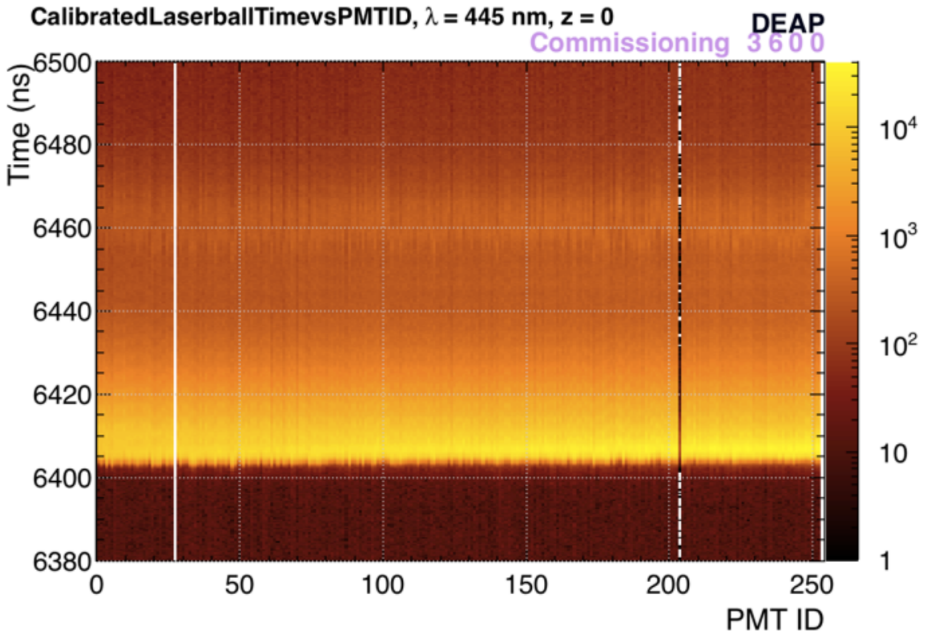


Figure 3.13: Calibrated laserball time versus PMT ID. This plot shows the overall result of correcting for the weigthed mean offsets extracted as described above.

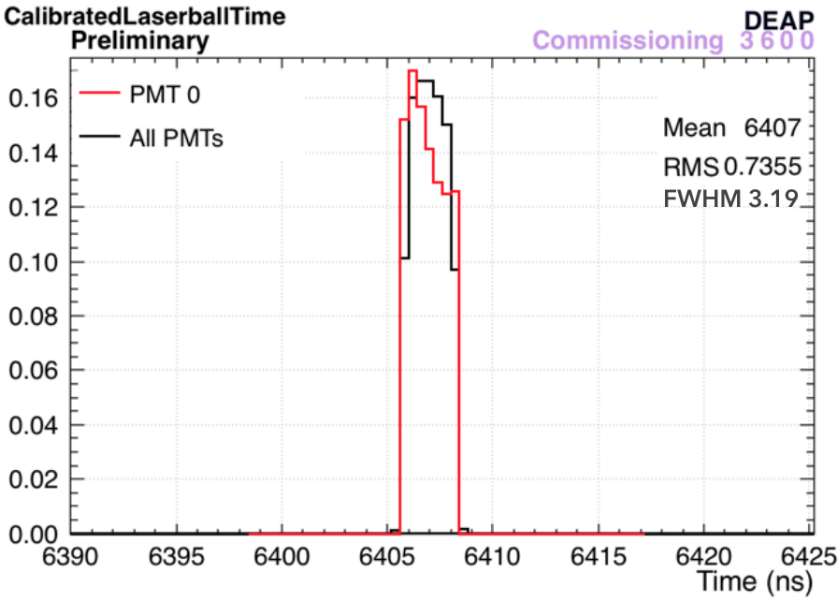


Figure 3.14: Projection of the calibrated laserball time for all PMTs (black) with its mean, RMS and FWHM and a generic PMT (PMT 0 in red).

3.4 Summary

In conclusion, the laserball source has allowed for a complete characterisation of the photomultipliers, giving an independent method from the AARFs to measure their relative

efficiencies and their time response. The data collected with the source placed at the centre of the detector, have allowed for the correction of channel-to.channel timing misalignments due to the detector electronics with a sub-nanosecond channel-to-channel alignment resolution achieved.

Chapter 4

Physics of neutron scattering, the DEAP-3600 74 MBq AmBe source and calibration data collection

4.1 Introduction

The physics of neutron interactions with liquid argon, the characteristic of the AmBe neutron source used in DEAP-3600 for calibration purposes, and a simulation study of this source at SNOLAB will be presented in this chapter. Neutrons are neutral particles, therefore they can travel many centimetres in a straight line through solid materials before they interact in some way with the material. When they interact (collide) with a nucleus of the medium they can be either be scattered or absorbed.

4.2 Neutron scattering in liquid argon

Neutrons in liquid argon undergo three types of interactions: elastic scattering, inelastic scattering and neutron capture.

Elastic scattering produces nuclear recoils, inelastic scattering leads to nuclear recoils and γ emission, neutron capture is followed by emission of a γ and Auger electrons.

The recoil nucleus kinetic energy in elastic scattering can be expressed as:

$$E_R = \frac{2m_n E_n}{(m_n + A)^2}(m_n + A - m_n cos^2\theta - cos\theta\sqrt{A^2 + m_n^2 cos^2\theta - m_n^2}) \tag{4.1}$$

where m_n is the mass of the neutron, A is the atomic mass of the target nucleus, θ the scattering angle of the outgoing neutron in the laboratory frame and E_n the incoming neutron kinetic energy. Eq. 4.1 can be approximated as [46]:

$$E_R \sim \frac{2A}{(1+A)^2}(1 - cos\theta)E_n \tag{4.2}$$

for incoming non relativistic neutrons ($E_n << m_n c$) interacting with heavy target nuclei ($A >> m_n$) as in the case of argon.

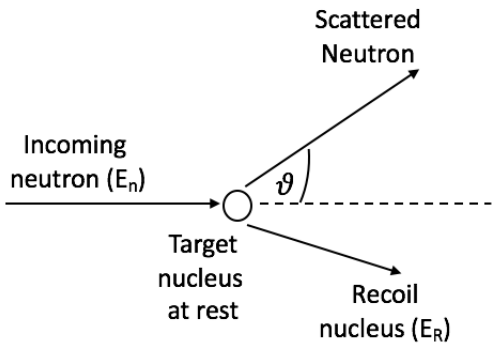


Figure 4.1: Neutron elastic scattering diagram in the laboratory coordinate system.

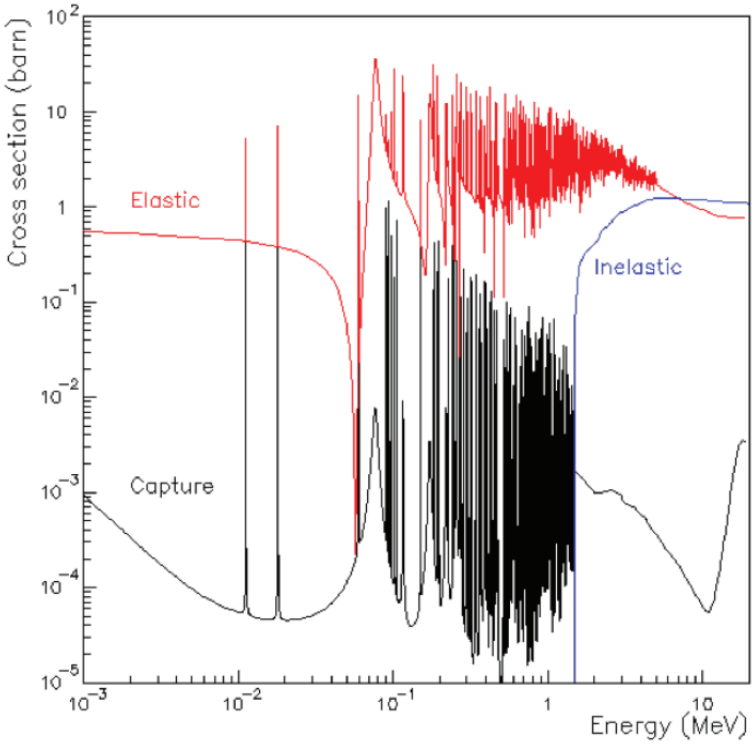


Figure 4.2: Cross-section for different neutron interactions on ^{40}Ar nuclei [67].

The total neutron cross section as a function of energy on ^{40}Ar nuclei for the three type of interactions is shown in Fig.4.2 whereas a diagram of an elastic neutron scattering in represented in Fig.4.1 in the laboratory coordinate system. Fig.4.2 shows that the inelastic contribution becomes relevant at energies above a couple of MeV and neutron capture dominate for thermal energies (0.025 eV). A common way to classify neutrons according to their energies is as slow or fast neutrons. Neutrons whose kinetic energy is below about 1 keV are conventionally referred to as slow neutrons whereas neutrons whose kinetic energy are above 1 MeV are commonly referred to as fast neutrons.

Elastic scattering

In elastic scattering reactions between a neutron and a target nucleus, no energy is transferred into nuclear excitation and both momentum and kinetic energy are conserved before and after the collision. This type of scattering is the most likely to happen for almost all nuclides and neutron energies. In an elastic collision the energy lost by the neutron is transferred to the recoiling nucleus as shown in the pictorial scattering reaction of a 1 MeV neutron with a nucleus in Fig.4.3.

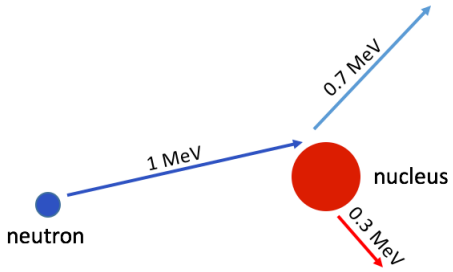


Figure 4.3: Schematic of a neutron elastic scattering for a 1 MeV neutron. Total kinetic energy is conserved and no energy is transferred into nuclear excitation.

The kinetic energy of the recoiled nucleus depends on its recoil angle, with the maximum energy transfer occurring for head-on collision. Elastic scattering between a neutron and an argon nucleus are the type of collisions expected between WIMPs and argon nucleii thus being able to recognise these interactions is of paramount importance.

Inelastic scattering

In the inelastic scattering reactions the kinetic energy is not conserved before and after the collision. Namely, the total kinetic energy is greater in the initial state rather than in the final. Inelastic scattering leads to excitation of the target nucleus that takes the remaining energy after the collision. It will then de-excite and emits one or more gamma-rays. Inelastic scattering is less common than elastic scattering and when it occurs is more likely to involve high energy (fast) neutrons and high Z target nuclei. In inelastic scattering between neutrons and argon nucleii additional or different particles can also often be produced, such as in the inelastic collision ${}^{40}_{18}\mathrm{Ar} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{1}_{1}\mathrm{p} + {}^{40}_{17}\mathrm{Cl}$ between a neutron and an ${}^{40}\mathrm{Ar}$ nucleus.

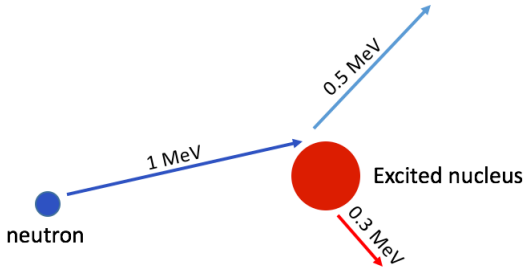


Figure 4.4: Schematic of a neutron inelastic scattering for a 1 MeV neutron. The kinetic energy in the final state is smaller than in the initial and the target nucleus is left in an excited state.

Capture

Capture occurs when the target nucleus absorbs the neutron, subsequently leaving it in an excited state. De-excitation from this state results in the emission of one or more gamma-rays commonly referred to as capture gammas. In a capture reaction the incident neutron is completely absorbed and a compound nucleus is formed as a result. A neutron that does not escape from the detector will be very likely captured as its energy decreases after scattering multiple times in the detector material, and low energy neutrons have the highest neutron capture cross section. A measurement of the $^{40}\mathrm{Ar}(n,\gamma)^{41}\mathrm{Ar}$ capture cross section in the energy range relevant to DEAP-3600 from 0.4 up to 14.8 MeV, can be found in [68]

4.3 The 74 MBq AmBe source

To calibrate the DEAP-3600 detector a 74MBq double-encapsulated source was purchased from Eckert and Ziegler [69]. AmBe sources are a mix of $^{241}\mathrm{Am}$ and $^{9}\mathrm{Be}$ and their neutron yield is ca. 2.0 to 2.4 $\times 10^6$ neutrons/sec/Ci with energies up to around 12 MeV. $^{241}\mathrm{Am}$ has a half-life of 432.2 years and decays via α emission to $^{237}\mathrm{Np}$. Neutrons are generated through the following nuclear reaction:

$$^{9}\mathrm{Be} + \alpha \rightarrow n + {}^{12}\mathrm{C}^{*} + Q(5.704MeV), \tag{4.3}$$

where the α particle comes from the $^{241}\mathrm{Am}$ decay. Also, a 4.44 MeV gamma-ray is emitted simultaneously from the the excited nucleus $^{12}\mathrm{C}^{*}$ 64% of the time. Weaker gamma rays

are also emitted during the ^{241}Am decay with energies of 59.5 keV and 13.9 keV. All alpha particles emitted by ^{241}Am are absorbed by the source encapsulation resulting in an alpha dose rate that is zero at 7 mm. As for the neutron rate it was measured by the manufacturer to be 4.6 kHz. The source received two further encapsulations to reduced to a minimum the risk of contaminating the laboratory: a soldered copper foil and a stainless steel enclosure.

4.3.1 Assessing background rate to experiments for storage

SNOLAB is a facility where multiple experiments are taking data simultaneously in different or even in the same experimental halls. Most of them are looking for rare events for which a high neutron rate can be a problematic background. For this reason it is important to assess what the implication for the laboratory of introducing a neutron source can be and how to mitigate its impact if needed. Monte Carlo simulations were carried out to assess this risk. The SNOLAB geometry was simulated as shown in Fig. 4.5 and includes the two main experimental halls that could be affected by the AmBe source: The Cube Hall (where DEAP-3600 and MiniCLEAN are hosted), the Cryopit, the Bottom Access Drift (BAD) and the Top Access Drift (TAD).

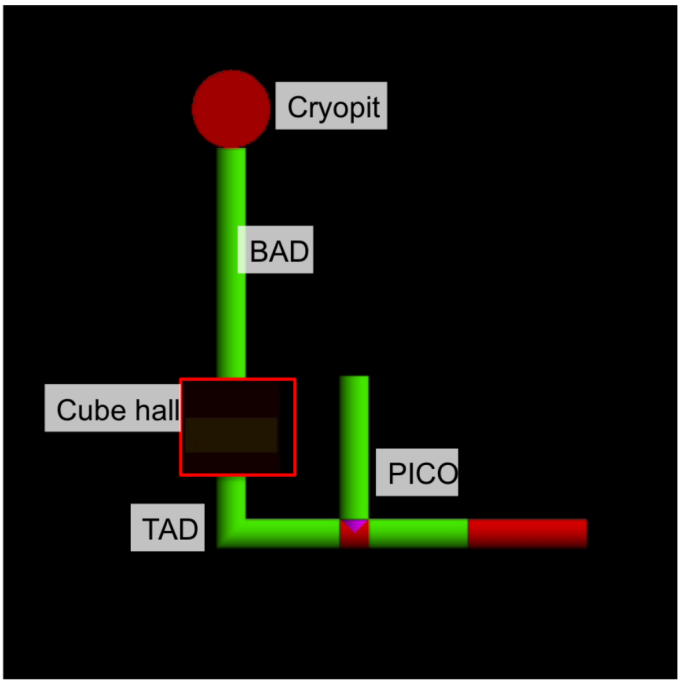


Figure 4.5: SNOLAB simulated geometry with cubehall, cryopit, PICO hall and relevant drifts.

A simulation of 2 million neutrons (about 6 minutes of running) from the bare neutron source on the DEAP-3600 deck in the cube hall found the neutron rate in the cryopit and in the TAD area where the HALO experiment is installed to be 10% and 1% of the fast neutron background rate in the lab. Another simulation showed that the neutron rate at PICO for a bare source simulated on the DEAP-3600 deck is 1.8 neutrons/m^2/hour which corresponds to 1% of the fast neutron background which is of 166 neutrons/m^2/hour.

The neutron energy distributions at PICO and in the TAD for a source placed on the deck are shown in Figures 4.6 and 4.7 respectively.

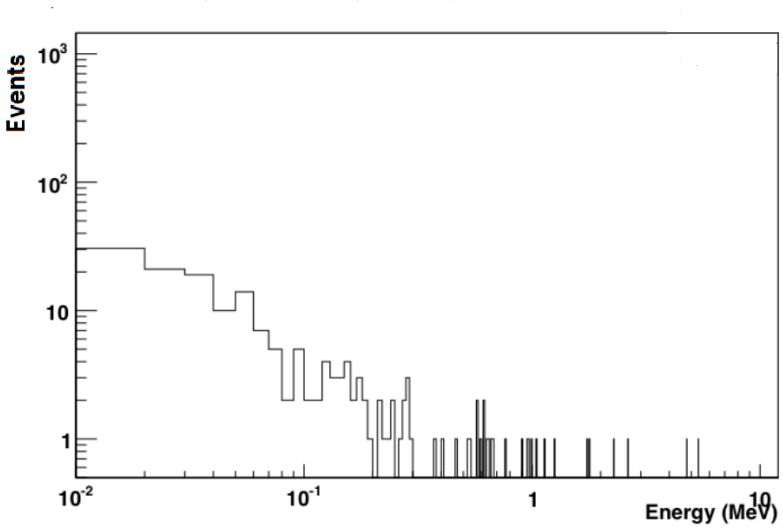


Figure 4.6: Neutron energy distribution at PICO hall with source placed on deck.

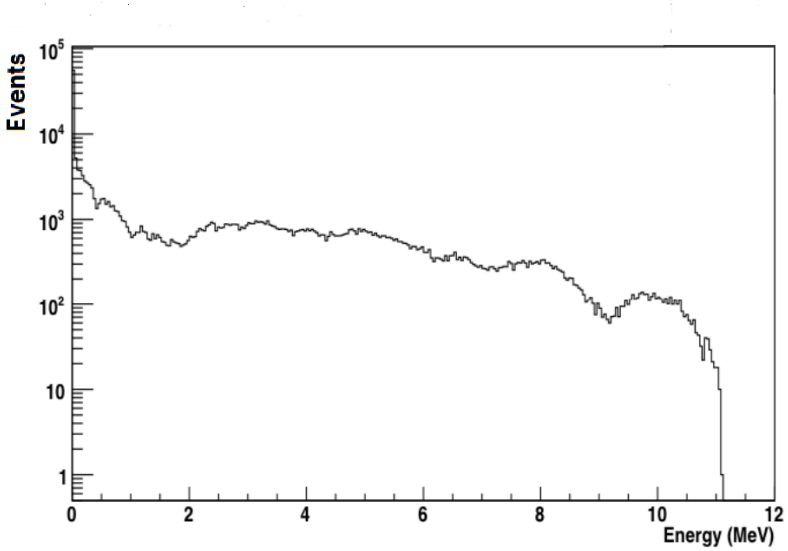


Figure 4.7: Neutron energy distribution at TAD whit source placed on deck.

The viability of storing the AmBe source in the Bottom Drift Access within a borated polyethylene box was also assessed in simulation. In a separate simulation, the bare AmBe source was simulated within the BAD. Fig. 4.8 shows the energy spectrum at the Top Access Drift for a shielded and unshielded source in the BAD.

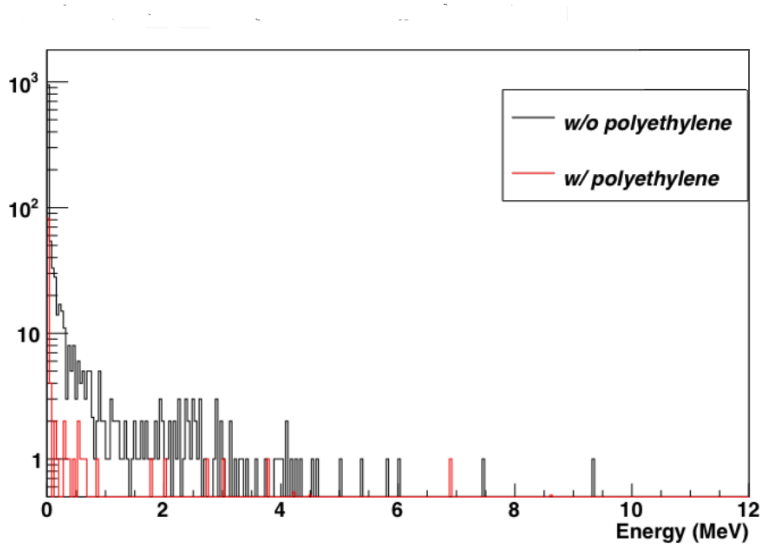


Figure 4.8: Neutron energy distribution at TAD for a shielded and unshielded source placed at BAD.

The neutron rate at the TAD and PICO when the source was within shielding box was found to be 230 neutrons/m^2/day and 0.023 neutrons/m^2/day respectively, equivalent to

5% and 0.001% of the fast neutron rate in the laboratory. In conclusion, the study showed that the impact of having the AmBe source in the Cube Hall and stored within a borated polyethylene box in the BAD does not affect other experiments running at SNOLAB.

4.4 AmBe Calibration Data

The first AmBe calibration campaign was undertaken when the detector was in its gaseous argon phase during cool-down prior filling for a total live-time of 9.41 days. Since then, many other data have been taken with liquid argon and in different positions. These AmBe data collection campaigns, their live-times and the position were the source was deployed are summarised in the table below.

Table 4.1: Breakdown of AmBe calibration campaigns from 2016 until July 2018.

AmBe calibration campaigns livetime in hours						
Year and Detector Phase	CAL A	CAL B	CAL E	Glove box	Neck	Total
2016 (GAr)	0	0	225.62	0	0	225.62
2016 (LAr)	0	0	147.03	0	0	147.03
2017 (LAr)	109.02	72.67	94.57	23.47	21.05	320.77
2018 (until July, LAr)	115.62	109.35	133.47	0	0	358.43
Total (days)	9.36	7.58	25.02	0.97	0.87	43.83

The same information is displayed in Fig. 4.9 where the run live-times are plotted as a function of time and different colours represent different source position.This plot shows how the calibration campaign strategy has changed and has been refined over time.

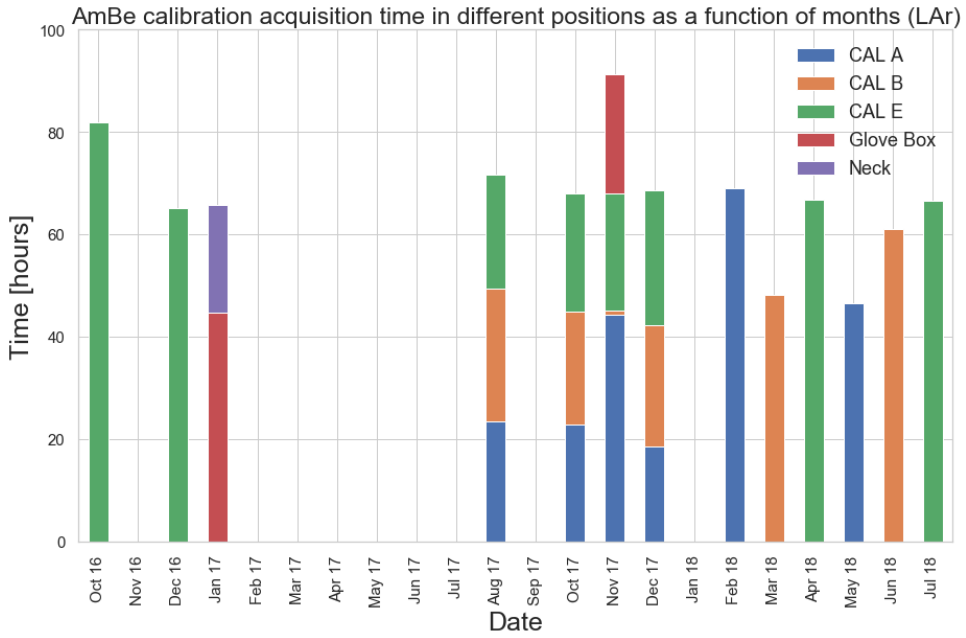


Figure 4.9: AmBe calibration runs up to July 2018. Data have been collected with the source placed in 5 different locations.

In particular, an analysis of the 2017 dataset presented below drove the decision to undertake monthly AmBe campaign, in one of the three calibration tubes per time.

The study aimed to answer the question of how much AmBe data was required to collect in order to obtain a statistically significant sample of single scatters in the threshold PE

bin (both the 80 and 120 PE bins were tconsidered) with low uncertainty (1%), and to measure stability versus time of the nuclear recoil mean F_{prompt} in the 80 PE bin at the 1% level. To answer the first question the mean F_{prompt} at the 80 and 120 PE bin was investigated for ten PE bins. Fig. 4.10 shows the 2017 data for each source deployment (left hand side) alongside the same data with a linear fit applied (right side). Each bin represents one run or the sum of multiple runs taken in the same calibration tube for a total live-time of roughly 23 hours (± 2 hours). It turned out that at least 30 days of AmBe data per year were required to get ∼ 5000 events in 3 years (4% statistical uncertainty) in the 80-90 PE bin.

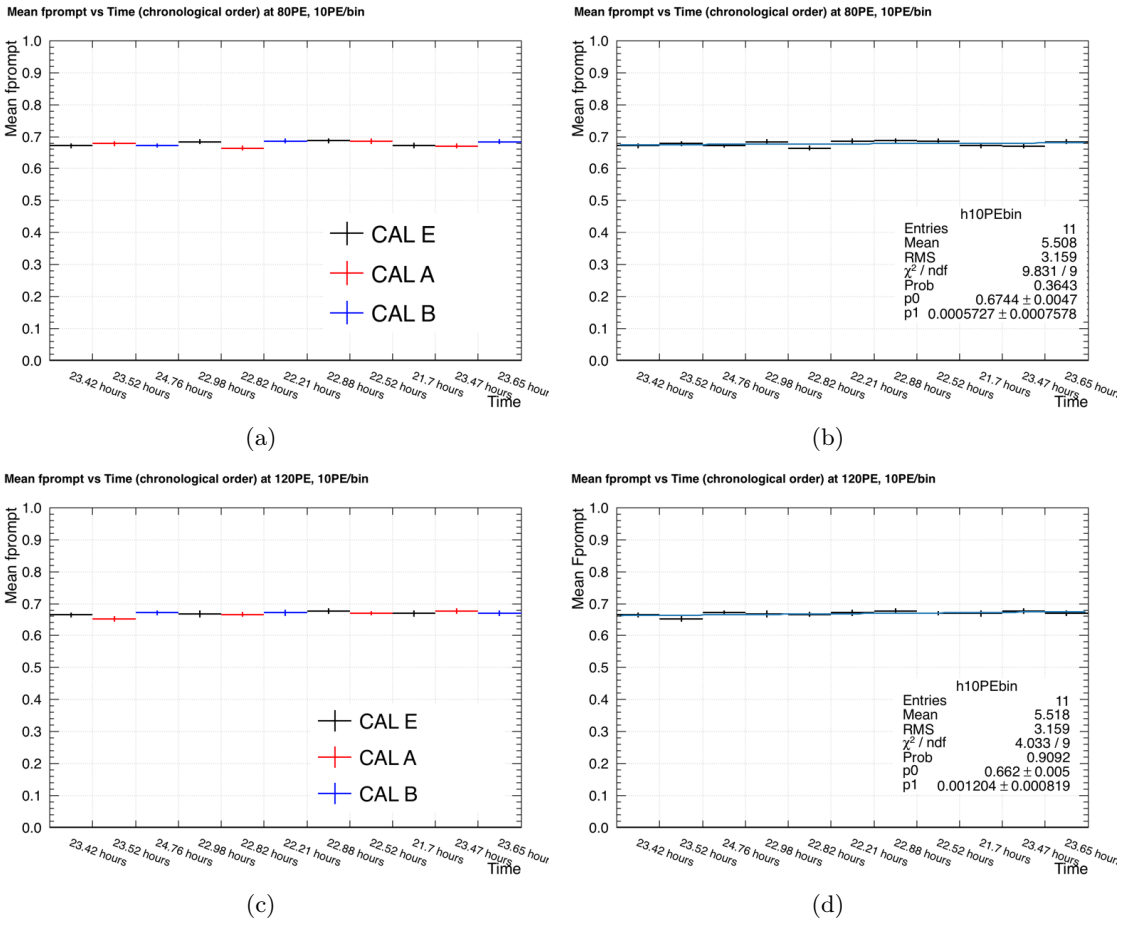


Figure 4.10: Mean F_{prompt} versus time for the 2017 AmBe calibration runs collected with the source in one of the three calibration tubes. (a) Mean F_{prompt} vs time in 80-90 PE bin; (b) linear fit to (a); (c) Mean F_{prompt} vs time in 120-130 PE bin; (d) linear fit to (c).

A decision was then made to collect ∼ 2.5 days of AmBe calibration data every month to reach this goal. Also, a single calibration tube run at each month was suggested such that any dead time to move the source would be minimised as no phi-angle dependence was

observed by looking at the energy distributions in the three calibration tubes at different months as shown in Fig. 4.11 (the difference in the blue distribution in Fig. 4.11(c) is a statistical effect due to a very short run in CAL B in November 2017).

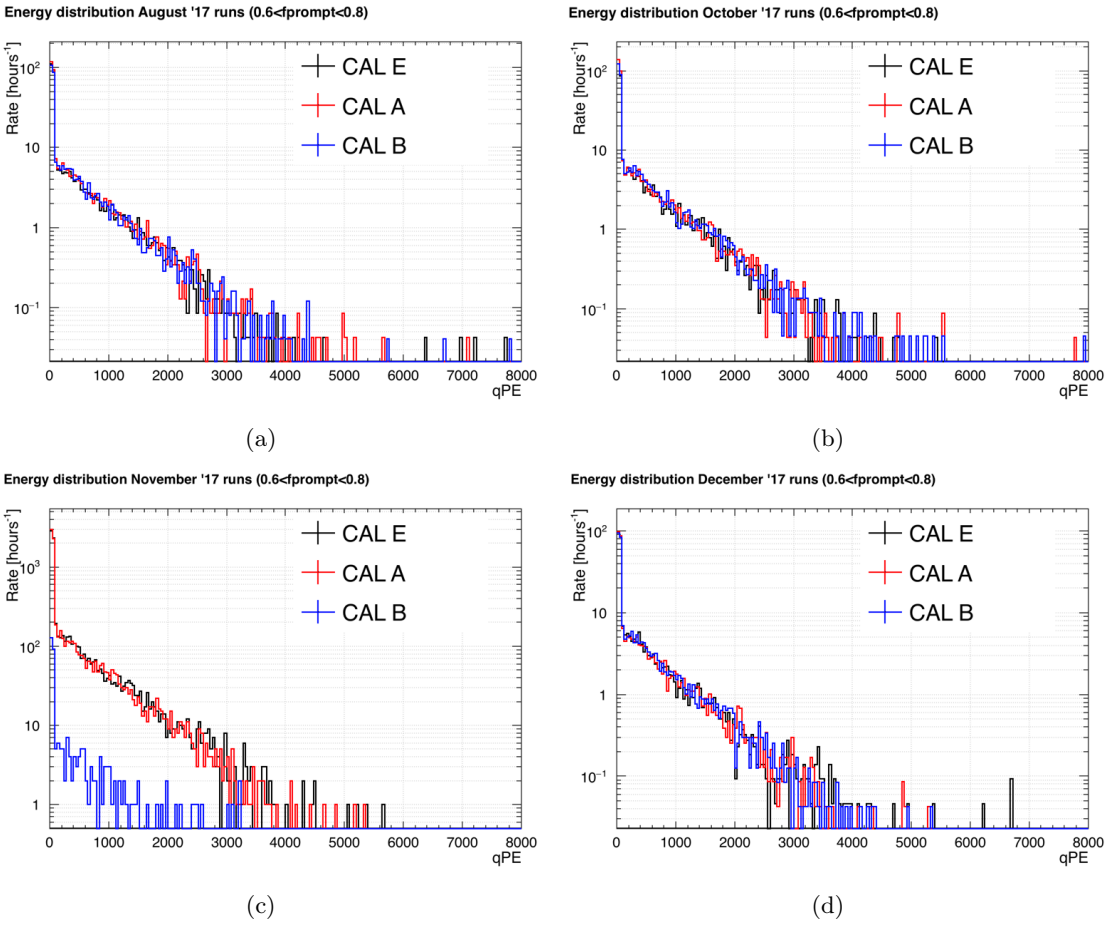


Figure 4.11: Angular dependence for the 2017 AmBe calibration runs in the nuclear recoil F_{prompt} region. No difference is seen for the three tube positions and over time. CAL B data are statistically limited in (c) as the data collection run was only few minutes long. (a) August runs, (b) October runs, (c) November runs and (d) December runs.

The nuclear recoil band rate of the 2017 AmBe runs also appears to be pretty stable over time as shown in Fig. 4.12 where the rate from 80 up to 8000 PE and in the $0.6 < F_{prompt} < 0.8$ region is plotted against run number in chronological order.

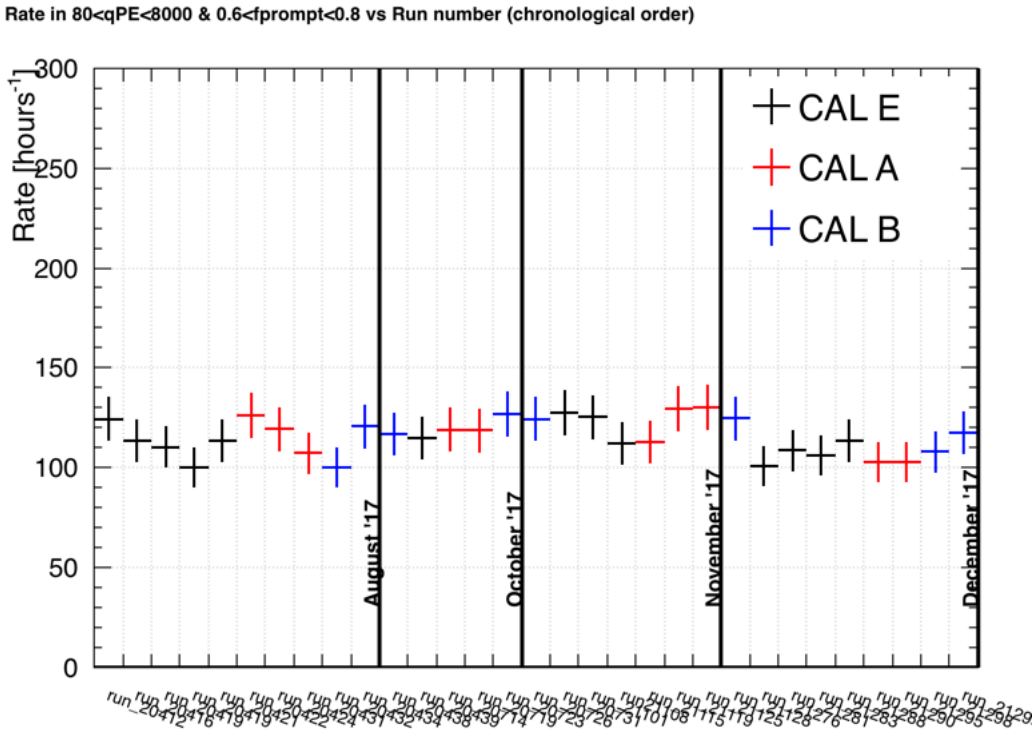
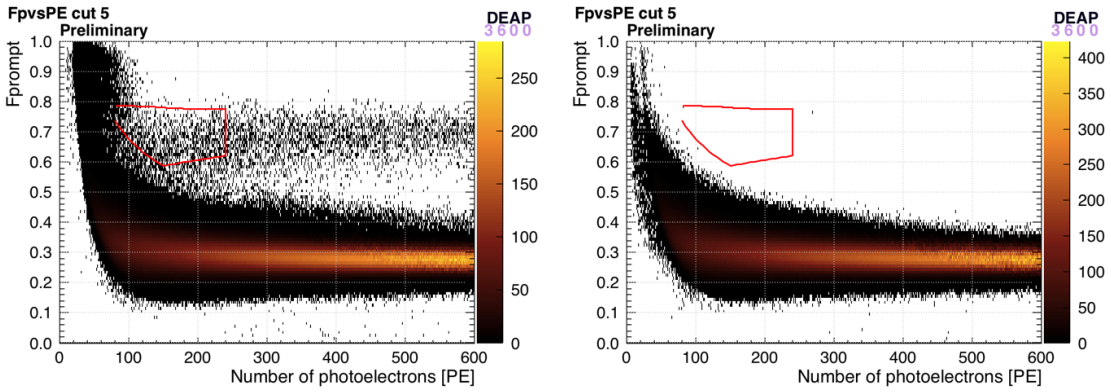


Figure 4.12: Nuclear recoil band rate versus time for the 2017 AmBe calbration data. Runs in the x-axis are displayed in chronological order and rates are calculated in the energy region between 80 and 8000 PE, and 0.6 < F_{prompt} <0.8.

Finally, an example of an AmBe calibration run compared to a normal physics run is shown in Fig. 4.13 in the F_{prompt} - qPE space after a series of selection cuts to remove bad events have been applied, these cuts are described in more detail in the following chapter. The region of interest box used for the first result publication [30] is also shown. When the source is deployed, many events appear in the F_{prompt} region above 0.6 in 4.13(a) populating the nuclear recoil band in contrast to a physics run (4.13(b)) where this region is empty if no WIMPs or background events have survived the cuts. Cherenkov events induced by the AmBe source are also visible in 4.13(a) at F_{prompt} values greater than 0.8.



(a) (b)

Figure 4.13: F_{prompt} versus PE for an AmBe calibration run (a) and a physics run (b) after all selection cuts have been applied. The nuclear recoil band can be seen in (a) at high F_{prompt}.

Chapter 5

Neutron Calibration

The detector's response to nuclear recoils is determined with long calibration runs where the AmBe source is deployed in one of three vertical calibration tubes. The neutron calibration studies presented in this chapter, which involve analyses of both real and simulated data, have been developed alongside the collection of AmBe calibration runs. As discussed in Chapter 4, having adequate statistics is of paramount importance for calibration purposes. Preliminary studies, that will be presented in the first part of this chapter, were undertaken when DEAP-3600 had collected four neutrons calibration campaigns with the detector in different configurations (see Fig.4.9) for approximately 220 hours livetime. A methodology to generate a very large, two stage, AmBe Monte Carlo data set, a cut flow to isolate nuclear recoil events and a simplified way to identify single scatters were developed. In the second part of this chapter, an enlarged AmBe data statistics (data collected up until July 2018), and improved two stage Monte Carlo generation technique and makes use of machine learning algorithms (as discussed in Chapter 6) to isolate a clean sample of single scatters among the nuclear recoil band is discussed.

5.1 Simulation

DEAP-3600 uses the Reactor Analysis Tool (RAT) [70] as a framework for simulation and analysis. RAT is a software package originally designed for the Braidwood neutrino experiment - a spherical liquid scintillator detector instrumented with PMTs - that has been adapted for DEAP-3600. RAT provides a single and coherent framework where both Monte Carlo simulation of the detector and data analysis can be performed. It builds on existing packages such as GEANT4.

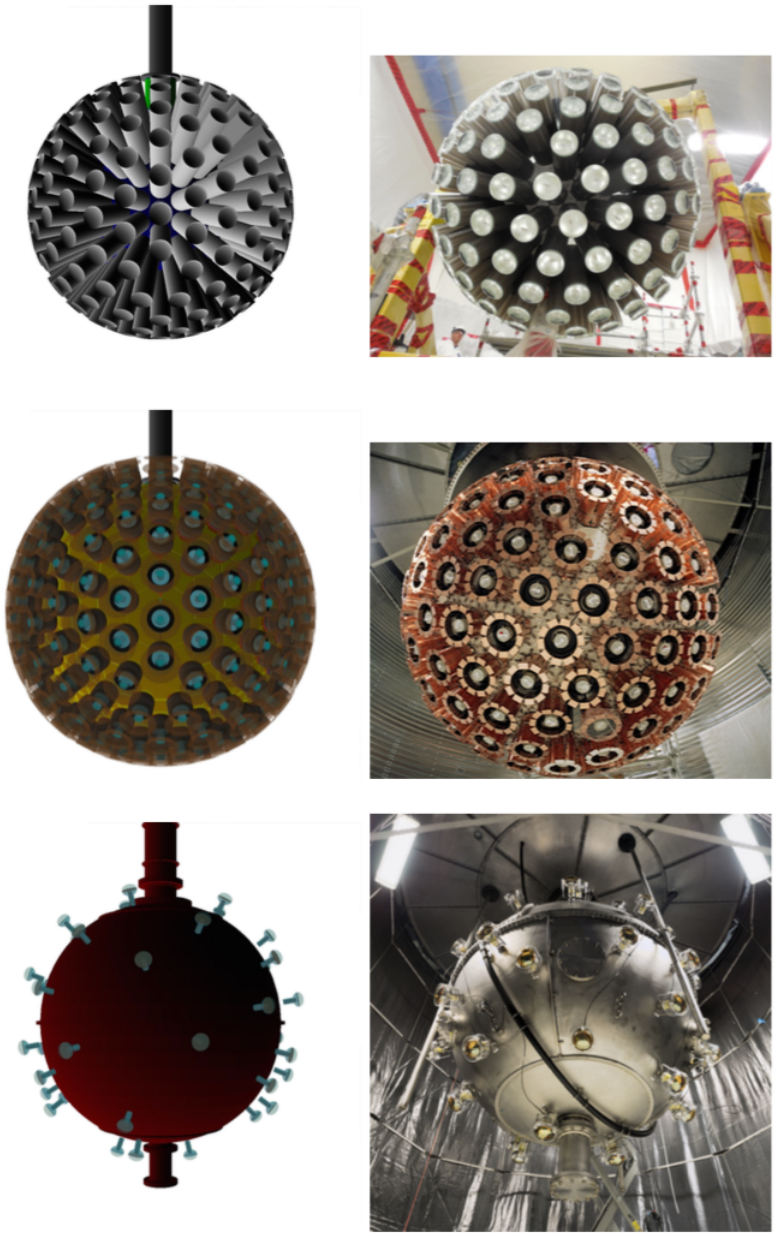


Figure 5.1: Ray traced simulated detector components compared to real DEAP-3600 detector components. Top: Acrylic vessel with light guides attached. Middle: copper heat shorts. Bottom: Steel shell and veto PMTs [61].

Detector geometries in RAT are configured by RATDB tables. RATDB is the database of constants that RAT uses for all the adjustable parameters. The DEAP-3600 geometry is built using plain text macros of the GEANT4 geometry classes. As its starts, RAT loads the RATDB tables and the geometries specified in the macro, loops over all the geometry tables in memory and builds the volume from the outside in. The advantage of using database controlled geometries is that new volumes of predefined shapes can be added very easily. Fig.5.1 shows some parts of the DEAP-3600 GEANT4 detector geometry built in RAT alongside the same components in the real detector. The light guides attached to the acrylic vessel are shown in the top picture, the copper heat shorts in the middle and the veto PMTs attached to the steel vessel in the bottom picture. Not shown in this picture, but included in the simulation, are the external calibration tubes, the water tank and the surrounding rock of the SNOLAB Cube Hall cavern where the experiment is hosted. As mentioned above, all the adjustable parameters such as the physical properties of the materials for the simulation are stored as RATDB files. These include both literature and test bench values. Scintillation in liquid argon is handled by a separate class in RAT that is called at the end of each GEANT4 step and is based on the model from [49]. Nuclear recoil quenching factors and PSD distributions are calculated from the SCENE measurements [47], and are also implemented in RAT.

5.2 Analysis chain

One of the features that makes RAT a very good framework for analysis is that when performing data analysis the software does not differentiate between real data and MC generated data. RAT analyses consist of a set of "processors" that analyse every single event in a specific ordered hierarchy. The order in which these processors are applied to data is important because each processor can potentially utilise variables created by previous processors. The raw data processing starts with running low-level processors such as calibration and pulse identification, all the way up to higher-level processors, such as position reconstruction. For calibration data RAT is run over the data twice: the first time to create and apply calibration constants and then for final processing.

5.2.1 Data quality and cleaning

High data quality must always be paramount at the creation of any dataset that aims to achieve important scientific results. A first-pass analysis of the DEAP-3600 data produces raw event files containing the uncalibrated physics events. From these, calibrated files are created with calibration constants and corrections - such as PMT channel timing offsets as discussed in Chapter 3 - applied

Further data quality control is given by diagnostic plots generated online at the end of each run that allow shifters or analysers to promptly spot anomalies in the data. The process discussed in Section 5.2 then generates calibrated data containing high level informations such as the position and the number of photoelectrons detected for each event. At this level in the analysis, several cut constants have been generated and stored in the dataframe. The cuts applied are:

1. **Low-level cuts**:

   (a) **Charge** This cut discards events that have no charge information.

   (b) **Charge cut** Some example of events cut by this low-level cut are PPG events or when the DAQ is running busy and may suppresse readout of digitisers.

   (c) **Trigger Sources** These are low level cuts that remove specific types of triggers according to flag values. For example periodic and muon veto triggers are removed by these cuts.

   (d) **Sub-event** This is a pile up cut that removes coincidence events. The variable sub-event is an integer indicating the number of events occurring in coincidence. Since WIMP events are expected to be stand-alone events, coincidence events are a background that needs to be removed. Events with only one sub-event are selected by this cut.

   (e) **Early pulses** A variable referred to as *numEarlyPulses* denotes the number of pulses counted in the first 1600 ns of the waveform (900 ns before the event peak). A cut selecting events where no more than 2 of these early pulses is used to remove pile up from previous events.

   (f) **Previous trigger time** The last cut in place to handle pile up is a cut on the start of two consecutive event windows. This cut requires a difference in time of at least 20 μs between the 2 event triggers such that light generated from an event will not leak into the next event.

2. **Quality cuts**:

   (a) **Max charge fraction per PMT** This cut aims to clean the dataset from Cherenkov light generated events produced by scatters of γ/β particles from decays of contaminant nuclei inside the acrylic of the AV or the light guides. A cut variable known as `fmaxPE` was defined as the ratio of the highest PE count observed in a single PMT divided by the total PE charge observed in that event by all the PMTs. The optimal value for this cut was found to be 0.2, with `fmaxPE` > 0.2 removing Cherenkov events from the dataset.

   (b) **Trigger time** Another pile up cut that catches pile-up events not typically identified by the sub-event cut. The variable used for this cut is referred to as calibrated trigger time (CTT), which is the peak of the pulse charges in an event summed over all the PMTs. CTT occurs 2500 ns after the start of the event window on average, so an event with an early trigger time has a previous event piling up in the new time window and an event with a late trigger time is because of pile-up later in the window. Events with CTT $<$ 2350 ns and CTT $>$ 2650 ns are rejected by this cut.

   (c) **Neck veto PMTs** Four neck veto PMTs are located around the outside of the acrylic neck and coupled with plastic scintillating fibres. A variable referred to as `neckVetoN` can take values of 0, 1, 2, 3 or 4 depending on how many veto PMTs fire. A cut requiring `neckVetoN` $= 0$ prevent events like alpha scintillation and Ar decay in gaseous argon that originate in the proximity of/or inside the AV neck to be part of the data stream.

3. **Fiducial cuts**:

   (a) **Max scintillation PE fraction per PMT** This cut is defined similarly to the max charge fraction per PMT (`fmaxPE`) cut. The variable is referred to as `fmaxnSC` defined as the fraction of total event charge found in the highest charge PMT after correcting for afterpulsing.

   (b) **Top two PMT rows** This is a fiducial cut that cuts high charge events in the z-axis. The top two PMT rows refer to two rings of PMTs that are above the liquid level. The detector is not fully filled due to an incident that occurred during the first liquid argon fill that resulted in the decision to not fully fill the detector [56]. These PMTs see events that occur in the gaseous region of the

detector. Events with a fraction of the total PE in the event in these top two rings greater than 0.05 are cut.

5.3 SCENE measurements and the Quenching model in DEAP-3600

The SCintillation Efficiency of Noble Elements (SCENE) experiment is a liquid argon TPC that uses a mono-energetic neutron beam to characterise the S1 scintillation and S2 ionisation signals produced by nuclear recoils between 10.3 and 57.2 keV with and without an applied electric field. A schematic of the experimental setup is shown in Fig.5.2. Protons from a Tandem accelerator were sent to a thick LiF target backing generating a neutron beam. The TPC was located at different distances from the LiF target in different runs (from 73.1 cm to 82.4 cm) and three EJ301 12.7 × 12.7 cm cylindrical liquid scintillator neutron detectors were used to detect the scattered neutrons placed 71 cm from the TPC with the angle between the beam, target and detector adjustable.

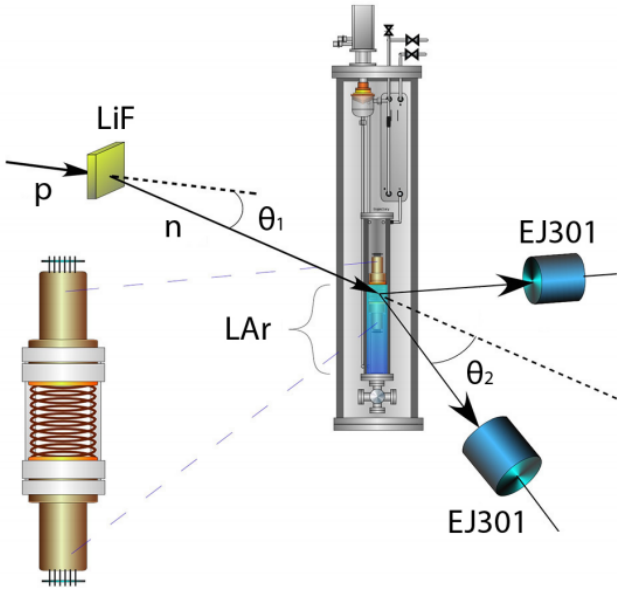


Figure 5.2: Schematic of the experiment setup not to scale. θ_1 is the neutron production angle and θ_2 the scattering angle [44].

The nuclear recoil quenching of ^{40}Ar is described in DEAP-3600 with the Lindhard-Birks model (see Section 2.2) was measured at zero electric field by the SCENE experiment. This model takes into account light loss due to heat generation and biexcitonic quenching.

However, ionised argon atoms not recombining at a relevant timescale (electron escape) is another mechanism that leads to light suppression and needs to be taken into account. Figure 5.3 shows the scintillation efficiency L_{eff} as a function of the nuclear recoil energy from [71] as well as measurements from the SCENE, ARIS and DarkSide-50 experiments.

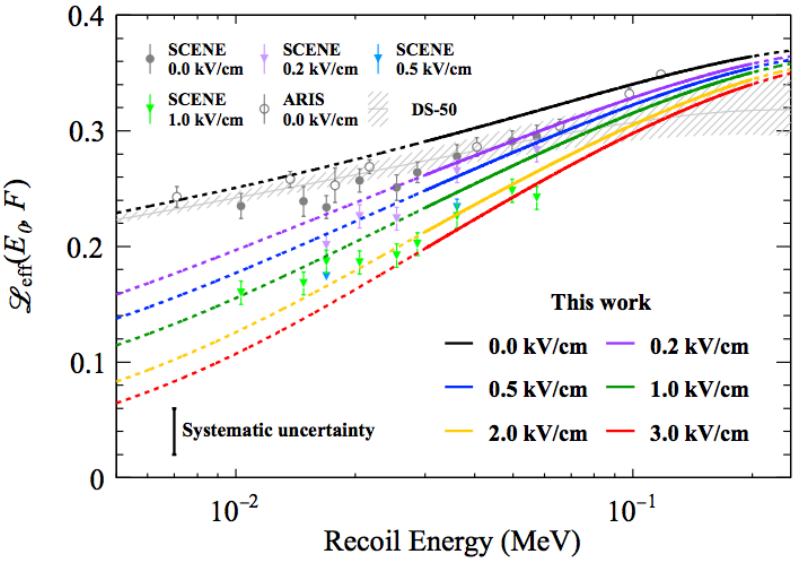


Figure 5.3: Scintillation efficiency L_{eff} as a function of the NR energy measured by Kimura et all. Measurements from SCENE, ARIS, and DarkSide-50 (DS-50) are also shown in the plot [71].

5.4 AmBe Monte Carlo Production

DEAP-3600 is designed to shield against neutrons as they are one of the trickiest sources of background. A neutron scattering once in the liquid argon volume can mimic a WIMP-like signal. Thus, generation of full simulations of neutrons is quite challenging as it requires a very large amount of CPU time and a huge amount of disk space to create a significant Monte Carlo dataset. Given a source strength of 4.6 kHz we would need to generate more than 1.5 billion neutrons to have an equivalent calibration run of 100 hours.

In order to speed up this generation process - that the nominal DEAP-3600 GEANT4 configuration would take many months to generate - and save disk space, a decision was made to split up the MC generation into two separate stages and edit the geometry and processes involved in the generation as discussed in the following section.

5.4.1 Stage 1 Monte Carlo

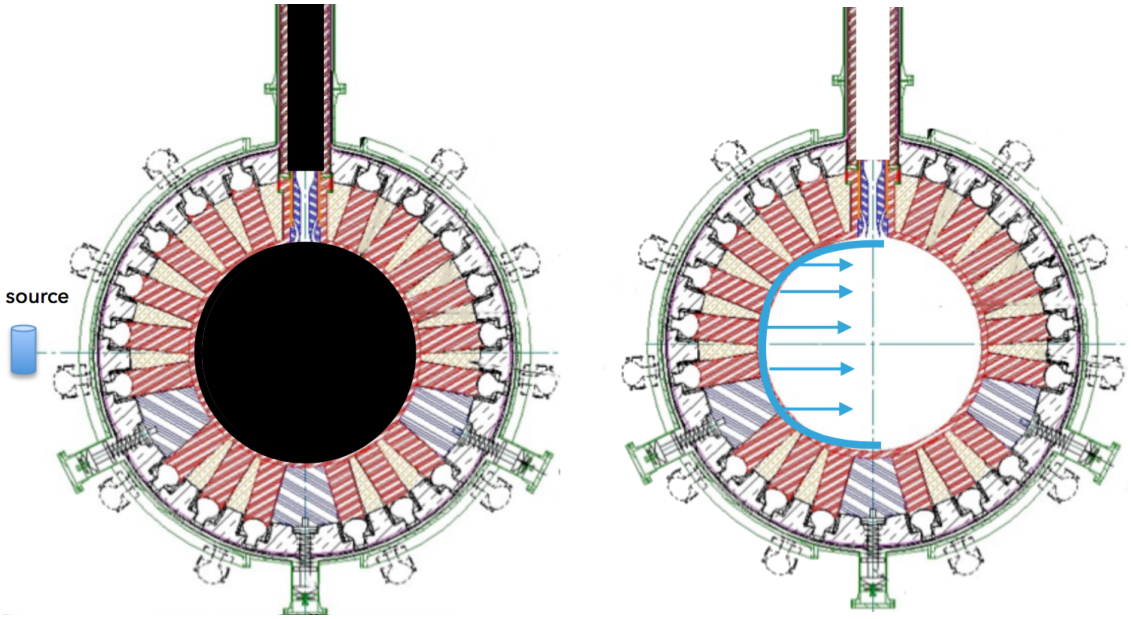
Stage one is a simplified simulation of the detector where some of the GEANT4 volumes are removed to reduce the interaction steps. In particular, the liquid argon and the TPB layer were removed from the GEANT4 simulated detector geometry in this stage, simplifying the otherwise intricate geometry where simulated neutrons can spend long time interacting. However, the neutron absorption rate remains unchanged as the overall mass remains unchanged. No electronics/DAQ response were simulated at this stage either. This makes it possible to generating a very large number of neutrons (order of billions) from the AmBe spectrum in a short period of time, because nearly all neutrons simulated fail to make it to the LAr volume. Also, a processor called "online trigger" was developed such that when running this simplified simulation, only events where a neutron makes it into the acrylic vessel volume (AV) are stored and all the others discarded. This allows to keep the output size of the MC as small as possible and make it faster to spin over this MC set when it comes to generate stage two. In what is presented in this work, $\sim$3 billions neutrons from an AmBe spectrum were generated with this method for an equivalent livetime of 181.2 hrs.

5.4.2 Stage 2 Monte Carlo

Stage two makes use of neutrons generated in stage one and performs a full simulation. Namely, these neutrons are now propagated though the detector with its full geometry in place, i.e. volumes removed during stage one are now replaced, and the DAQ and electronics response are simulated. The liquid argon optical properties in RAT are modified to reflect the SCENE measurements [47] and a fit to the SCENE data was extrapolated to the alpha singlet-to-triplet ratio to better describe the rise in F_{prompt} for increasing PE (number of photoelectrons). Additionally the quenching factor Q_f and the singlet-to-triplet production ratio for nuclear scatters are changed in RAT. In particular, Q_f is taken from table IV and the singlet-to-triplet ratios are calculated using the zero field data in table V in reference [47].

Neutrons from stage one were re-simulated with their momentum as recorded in the last GEANT4 step before entering the detector and their position is extrapolated to the edge of the detector volume using their first step in the detector volume and their last step outside the detector volume (Fig.5.4). Also, the most energetic inelastic gammas coming from neutron interactions in the detector materials were simulated at this stage but not capture

gammas. This was to speed up the simulation but simulation described later include these gammas within their events.



(a) (b)

Figure 5.4: Schematic of the 2 stages Monte Carlo generation. (a) Stage 1: simplified geometry with the neck geometry, the LAr volume and the TPB surface removed (black) and the AmBe source in one of the calibration tubes. (b) Stage 2: Surviving neutrons and associated inelastic gammas re-simulated near the edge of the AV.

5.5 AmBe Monte Carlo - Data Comparison

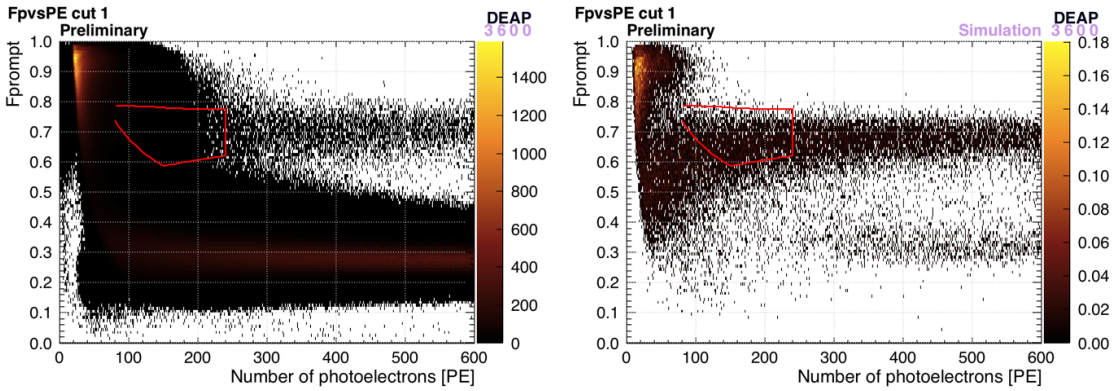
The approach taken by the DEAP-3600 collaboration to define the acceptance to WIMPs for the dark matter search analysis was to use a simulated single scatter nuclear recoil spectrum as opposed to extracting it from the AmBe calibration data. This is because it is not trivial to identify pure single scatter events in the AmBe data due to the complexity of the neutron interaction physics. The strategy of using simulated recoils requires a good match between AmBe data and AmBe Monte Carlo in the dark matter search region of interest (ROI), as if the simulation models the data accurately, we can determine the WIMP acceptance as a function of F_{prompt} and PE by comparing the true single scatter spectrum for the AmBe MC with the simulated single scatter spectrum. In this section, the cut flow applied to data and Monte Carlo to clean the data as well as representative distributions will be presented. It should be noticed that the cut flow presented in this section differs from the one listed in Section 5.2.1 as it reflects the early stages of the analyses as opposed

to the final results that will be presented later in this chapter. The definition of ROI has changed many times throughout the time this study was being conducted as the PE threshold was lowered given the excellent pulse shape discrimination of the detector (power to discriminate between signal and background). We will refer to two ROIs in this work: **basic ROI** - ROI used when this study began, and **ROI** - region of interest used for the first analysis published by the collaboration (red box in Figure 6.14).

The AmBe data used for this analysis were taken with the detector in stable condition and with its full LAr mass of 3322 $\pm$ 110 kg with a livetime of 65.2 hours. Stage two AmBe Monte Carlo simulates a detector with the same amount of LAr and has an equivalent livetime of 181.2 hours. However, it should be noted that the AmBe MC simulation does not include ^{39}Ar pile-up and, as will be discussed in detail later, we believe it under estimates the Cherenkov light production seen in real data. The 4.4 MeV gammas associated with the AmBe source are not included either in the results presented in this section but the effect of adding them with a separated set of Monte Carlo will be shown in the next sections. The standard cuts applied to this analysis are summarised in Table 5.1 and include cuts to remove events from DAQ calibration, Cherenkov events in the light guides and PMTs and pile-up. Cuts 4 and 5 are fiducial volume cuts that rely on position reconstruction, 6 and 7 are equivalent to them but make uses of charge or scintillation light seen from the PMTs instead. Results after cut 5 and 7 will be often compared in this section as they represents the cleanest set from which to extract and study the nuclear recoil band induced by neutrons from the AmBe source. Figures 5.5 to 5.11 show the F_{prompt} versus PE for AmBe data before background subtraction (left-hand side) and MC (right-hand side) after each cut and normalised to their respective livetimes. Some differences that can immediately be inferred by looking at these distributions are the absence of the large band at low F_{prompt} and the population at high F_{prompt} and low PE in MC. The former is not surprising as we know that the low F_{prompt} band in data comes mainly from ^{39}Ar which is not included in the simulation. The latter is a bit more compelling, as we believe it is due to Cherenkov light that is produced in real data but not well modelled in the simulation. The number of events surviving each cut in data in the basic ROI and nominal ROI are summarised in Table 5.2.

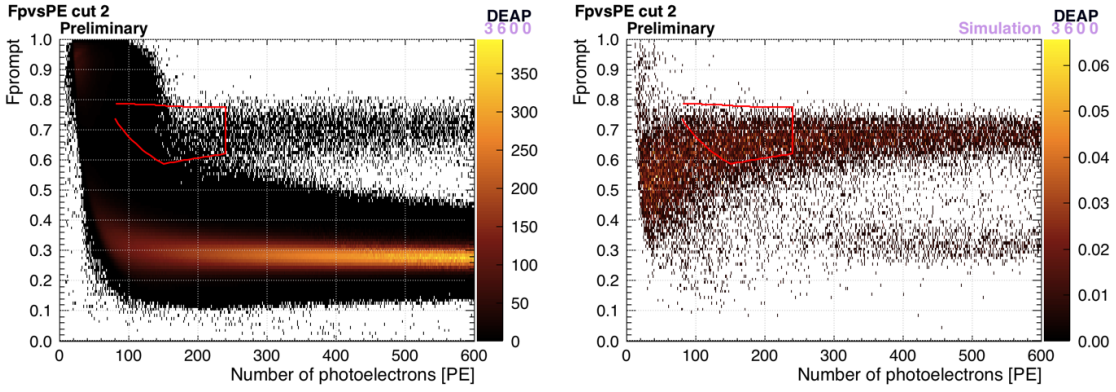
Cut Number	Cut Definition
Cut 1	DAQ calibration, Pile-up, Event asymmetry, Neck veto
Cut 2	Cut 1 + Max charge fraction per PMT
Cut 3	Cut 2 + Event time
Cut 4	Cut 3 + Fill level cut
Cut 5	Cut 4 + Radial cut
Cut 6	Cut 5 + Charge fraction in the top 2 PMT rings
Cut 7	Cut 7 + Max scintillation PE fraction per PMT

Table 5.1: Cut flow applied to both AmBe data and MC analysis definition.



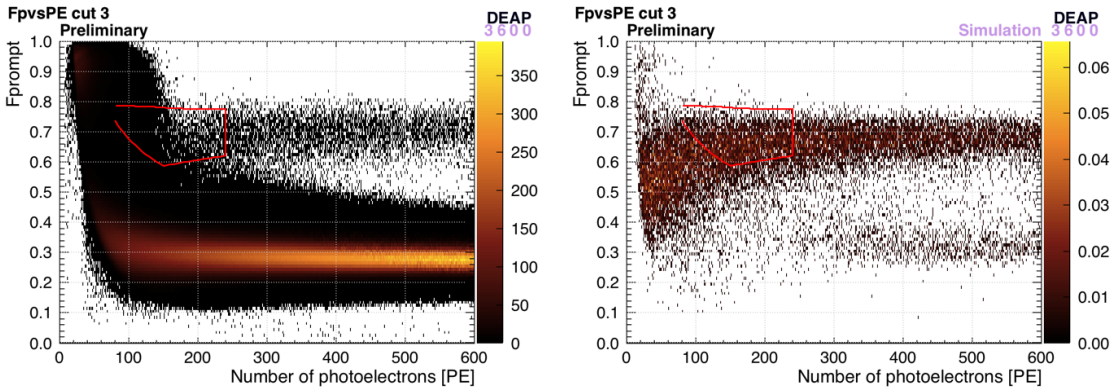
(a) (b)

Figure 5.5: F_{prompt} vs PE for AmBe data (a) and MC (b) after cut 1.



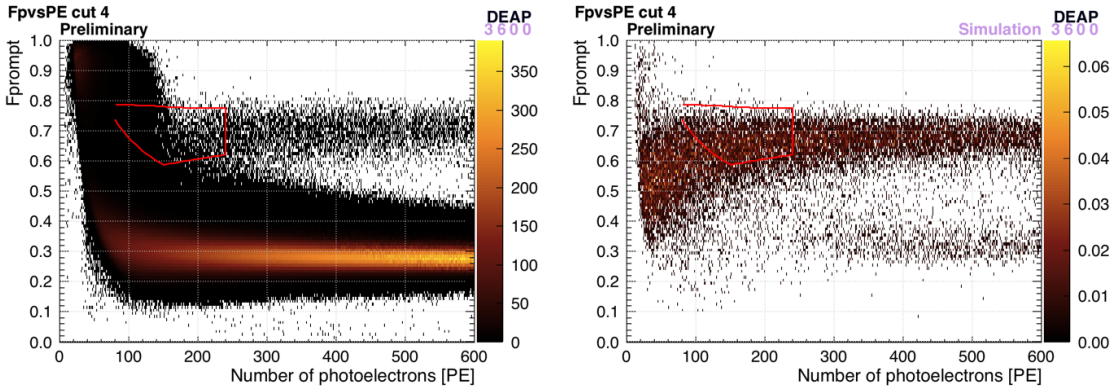
(a) (b)

Figure 5.6: F_{prompt} vs PE for AmBe data (a) and MC (b) after cut 2.



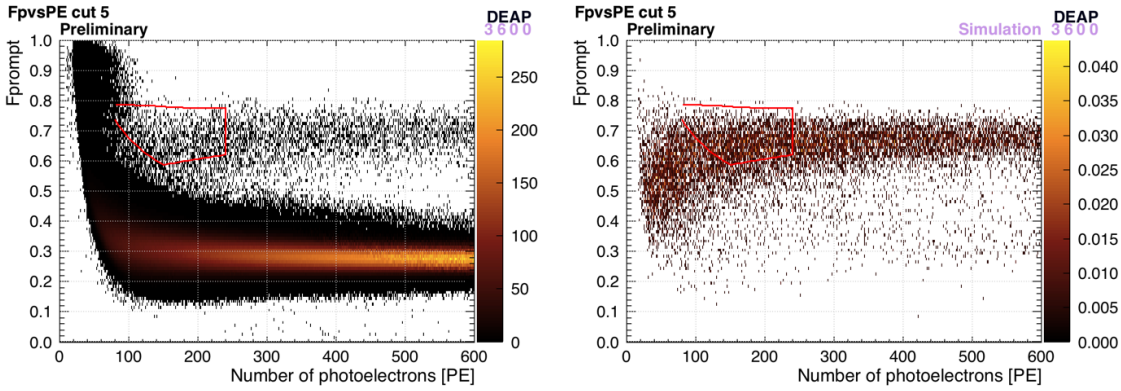
(a) (b)

Figure 5.7: F_{prompt} vs PE for AmBe data (a) and MC (b) after cut 3.



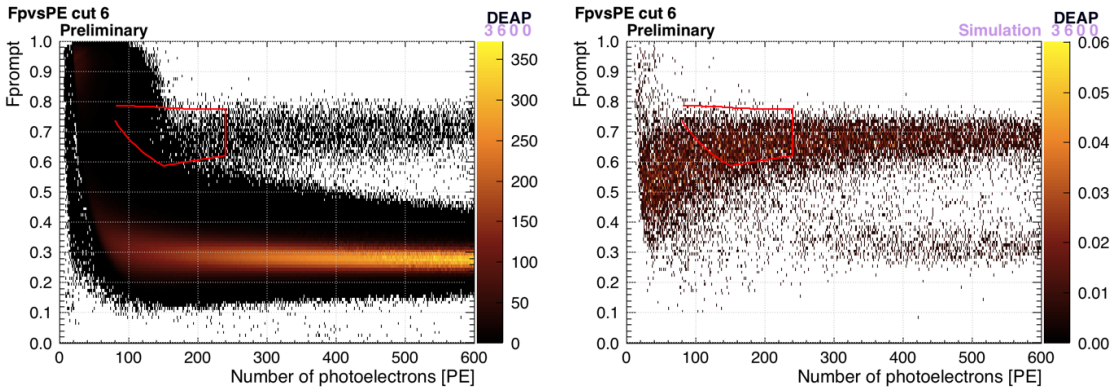
(a) (b)

Figure 5.8: F_{prompt} vs PE for AmBe data (a) and MC (b) after cut 4.



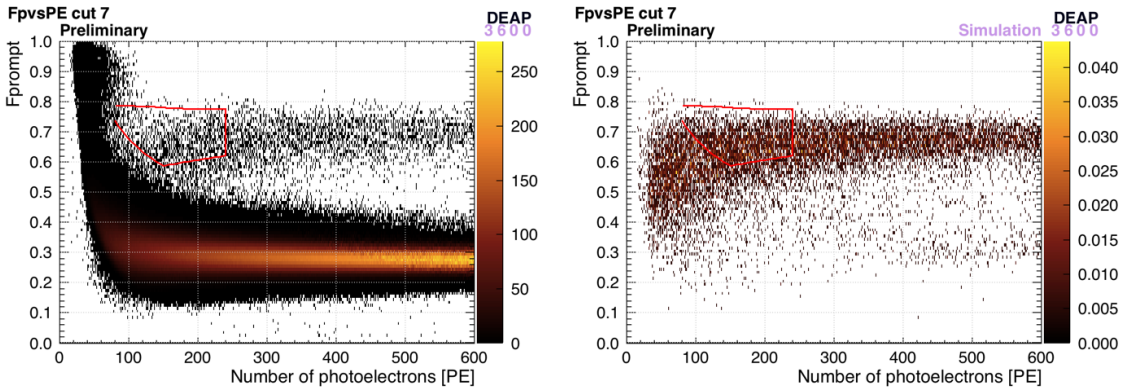
(a) (b)

Figure 5.9: F_{prompt} vs PE for AmBe data (a) and MC (b) after cut 5.



(a) (b)

Figure 5.10: F_{prompt} vs PE for AmBe data (a) and MC (b) after cut 6.



(a) (b)

Figure 5.11: F_{prompt} vs PE for AmBe data (a) and MC (b) after cut 7.

It can be seen how the cuts affect data more substantially than Monte Carlo. This is not surprising as they are mainly designed to remove pile-up and Cherenkov events that happen way less often in Monte Carlo "by construction". In order to make a more realistic comparison to Monte Carlo, one can subtract a physics background run when the AmBe source was not deployed from the AmBe data. A run with a livetime of 18.7 hours and collected right after the source was removed in order to ensure that the detector was in the same status has been used for this purpose. It was noted that after background subtraction and normalisation, the background run had more ^{39}Ar events than the AmBe run (see Figure 5.12(a)). A natural culprit for this behaviour could be a difference in the trigger rate between calibration and physics runs. Investigations in this direction have been conducted and are summarised in Figure 5.14, which shows the difference in time (Δ T) between triggers for the AmBe and background runs. If the higher ^{39}Ar "activity" after

normalisation in background was due to the trigger rate being too high in the calibration run, one would expect to see some difference for the two distributions in Figure 5.14. The fact that no difference was seen and that the same behaviour was observed for the partial fill AmBe calibration runs, suggest that this is not related to the trigger. Given that the ^{39}Ar rate should be the same in AmBe and background runs (full LAr level for both runs), a decision was made to scale down the ^{39}Ar background rate to match the AmBe data as can be seen in Figure 5.12(b). Figure 5.12 shows the F_{prompt} versus PE distribution after background subtraction for cut 5 (fiducial volume based cut). The same procedure was followed to perform the subtraction for cut 7 (charge based cut) as illustrated in Figure 5.13. A comparison of events surviving these two cuts in the region of interest between background subtracted AmBe data and AmBe MC is reported in Table 5.3.

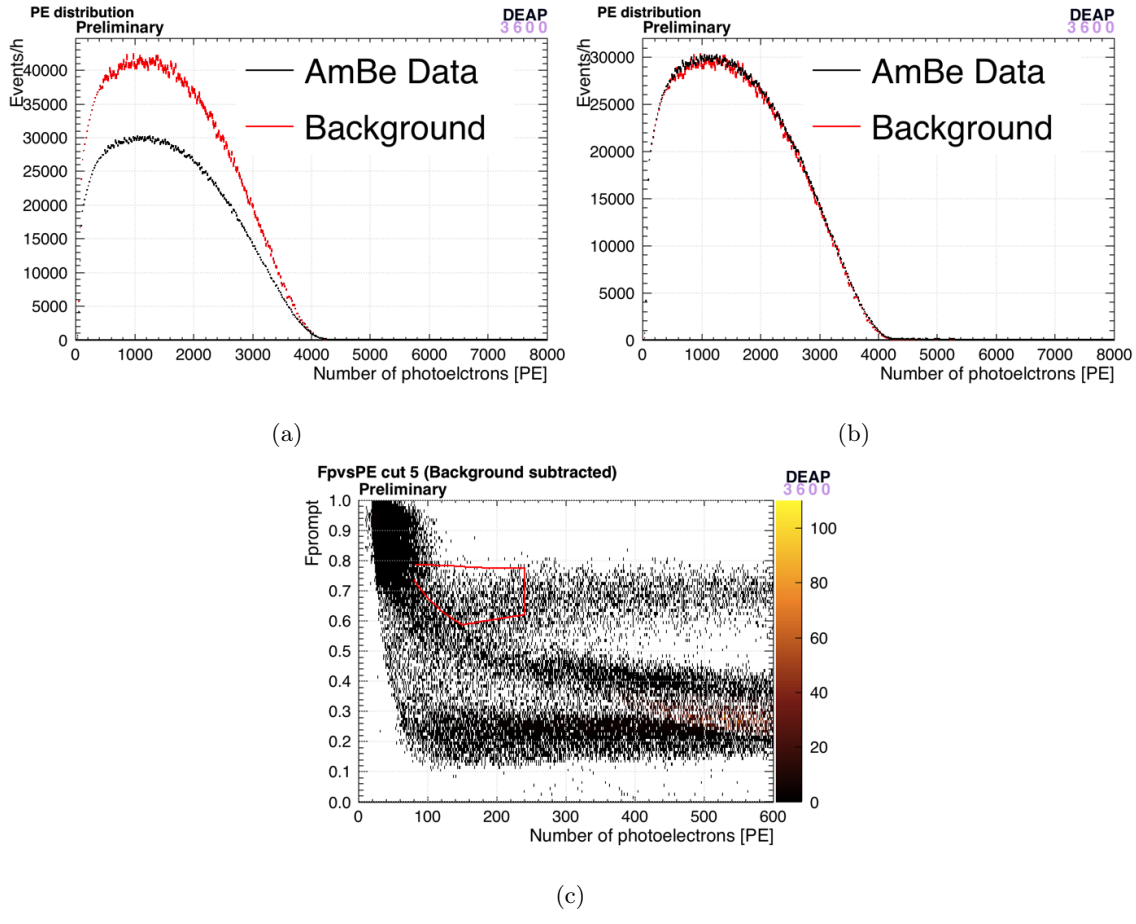


Figure 5.12: PE distribution comparison between AmBe data and physic data (background) before (a) and after rescaling (b) for cut 5 and F_{prompt} versus PE background subtracted after cut 5 (c).

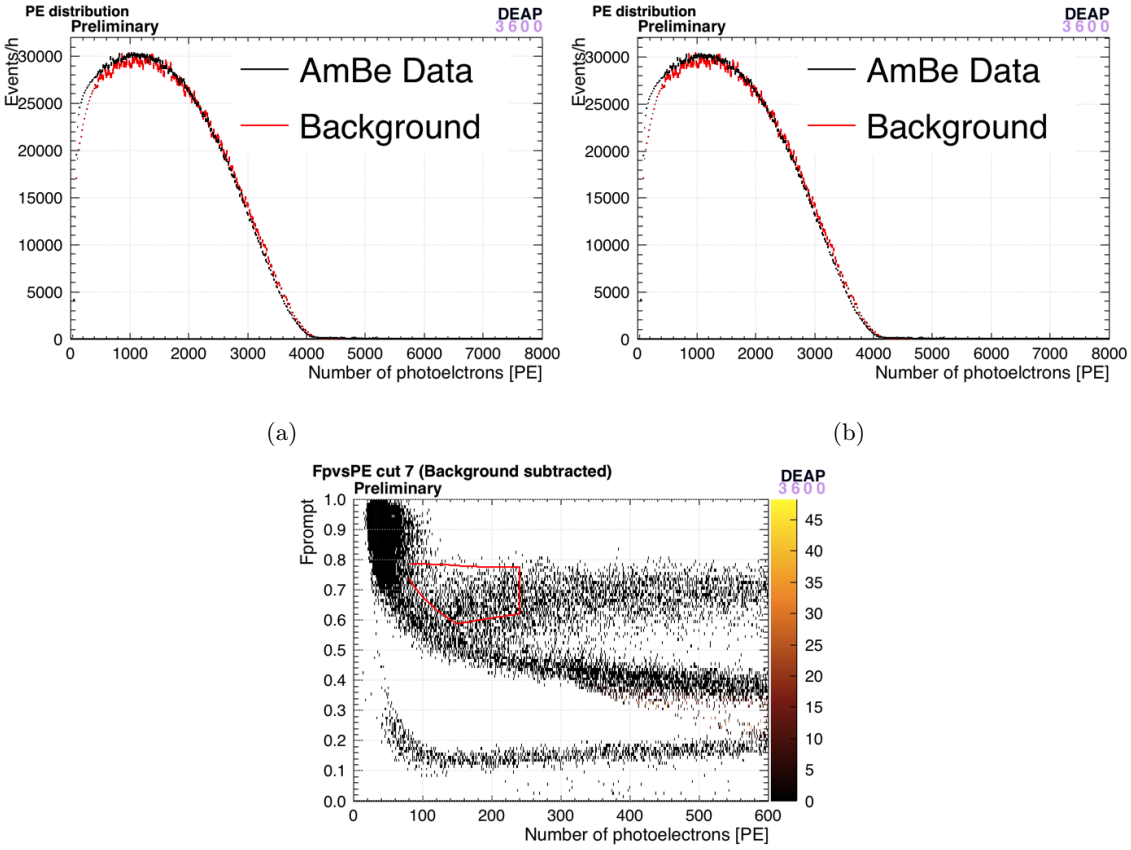


Figure 5.13: PE distribution comparison between AmBe data and physic data (background) before (a) and after rescaling (b) for cut c7 and F_{prompt} versus PE background subtracted after cut c7 (c).

Cut Number	Basic ROI	ROI
Cut 1	1,923	6,877
Cut 2	101.5	309.3
Cut 3	101.4	308.8
Cut 4	96.7	293.5
Cut 6	85.2	256.7
Cut 5	16.6	20.2
Cut 7	16.8	17.4

Table 5.2: Number of events per hour surviving different cuts for the two ROIs.

Table 5.3 shows how lowering the PE threshold results in a higher number of events leaking into the ROI in data as opposed to a very small increase in MC. These events are likely to be Cherenkov events that have not been properly cut in data. Figure 5.15 shows the x, y and z reconstructed distributions for AmBe data and MC in the basic ROI, showing a good agreement for all three coordinates.

Cut Number	Basic ROI data : MC	ROI data : MC
Cut 5	16.6 : 14.1	20.2 : 14.8
Cut 7	16.8 : 14.6	17.4 : 15.2

Table 5.3: Number of events surviving different cuts per hour in data and MC. Data are background subtracted.

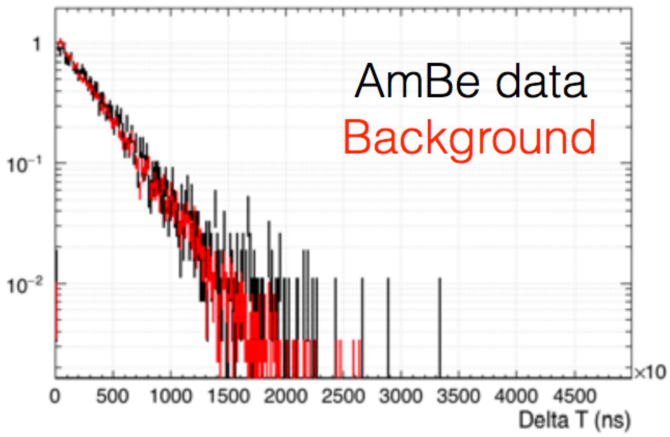
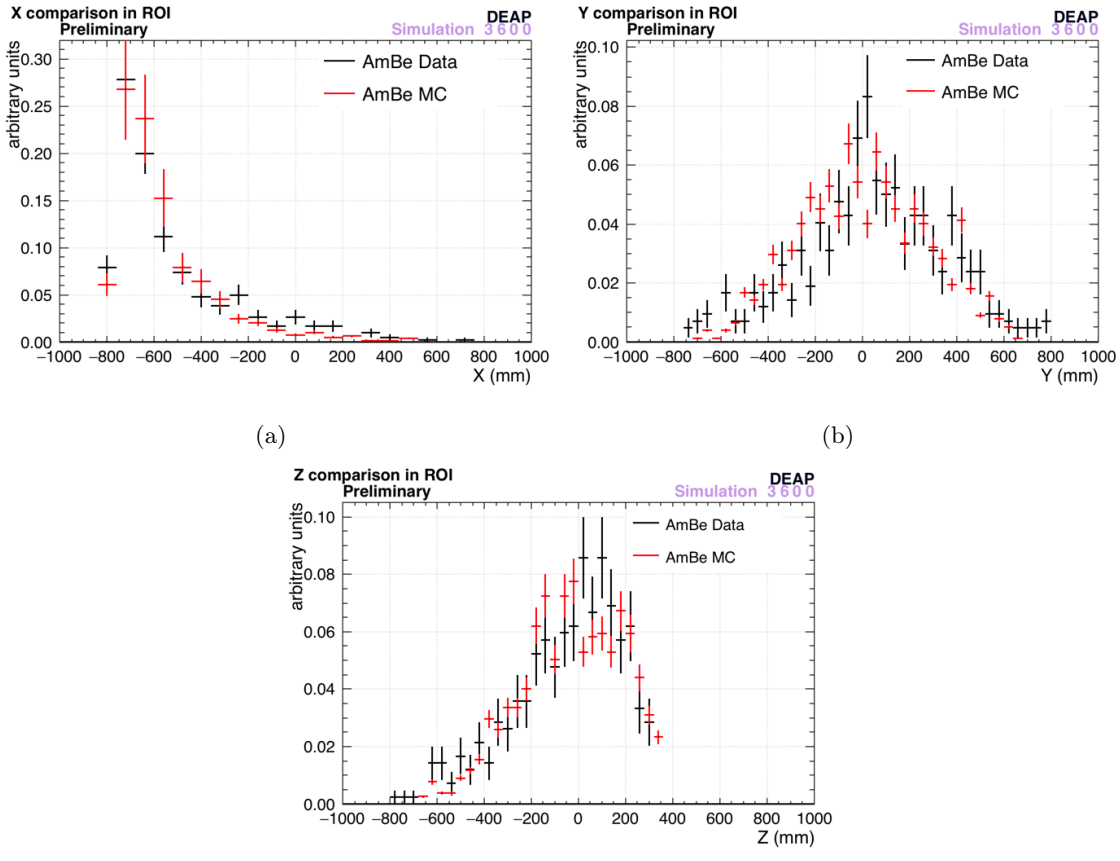


Figure 5.14: Δ T (time) between triggers for the AmBe and background runs.

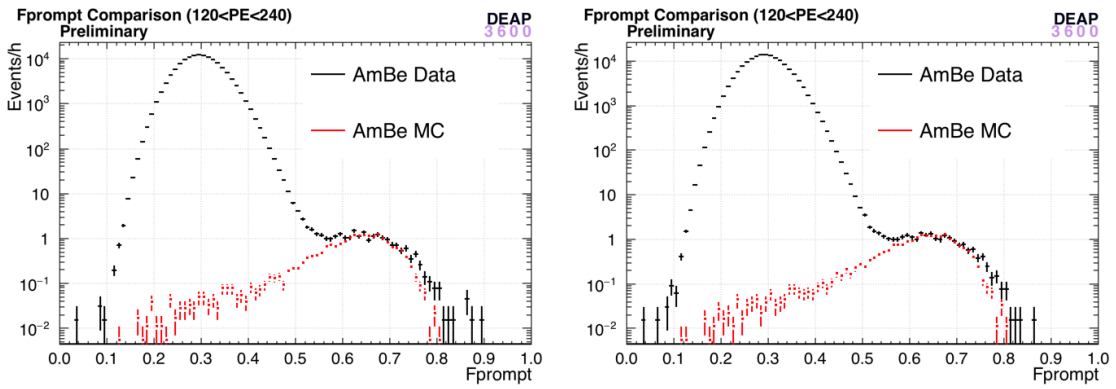


(a) (b)

(c)

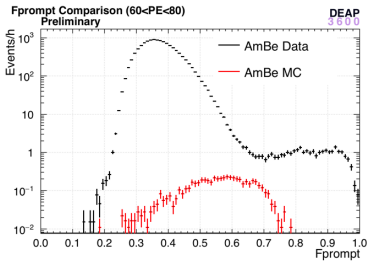
Figure 5.15: Position reconstruction comparison between AmBe data and MC: X (a), Y (b) and Z (c).

A comparison of the F_{prompt} distributions for cut 5 and cut 7 can be seen in Figure 5.16 for a PE region between 120 and 240 PE (basic ROI) and in Figure 5.18 in the 240 < PE < 3000 PE region again for both cuts. As can be seen from Figure 5.16 there is good agreement for $F_{prompt} > 0.6$ (ROI). However, a discrepancy can be seen in the high PE region (Fig.5.18).

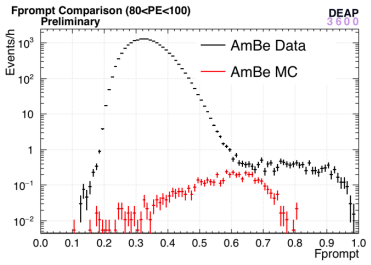


(a) (b)

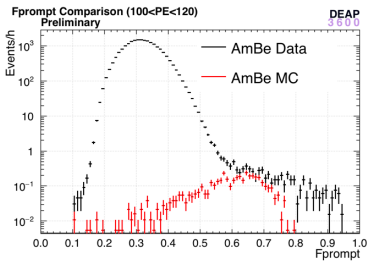
Figure 5.16: F_{prompt} distributions comparison in the basic ROI PE range ($120 < PE < 240$) between AmBe data and MC for cut 5 (a) and 7 (b).



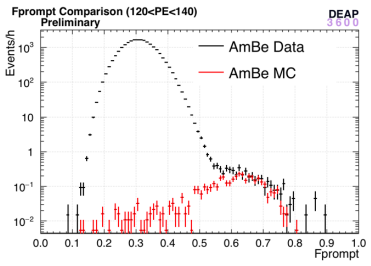
(a)



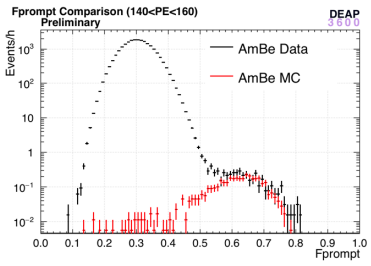
(b)



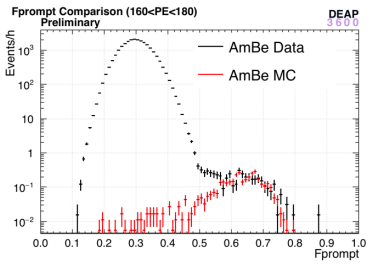
(c)



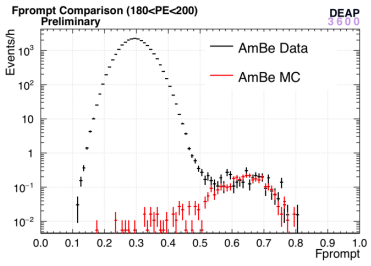
(d)



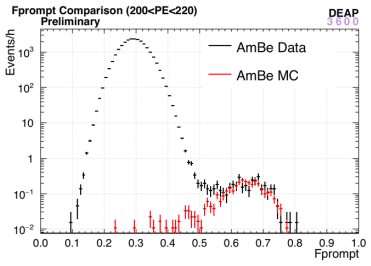
(e)



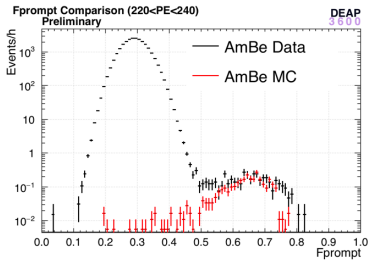
(f)



(g)



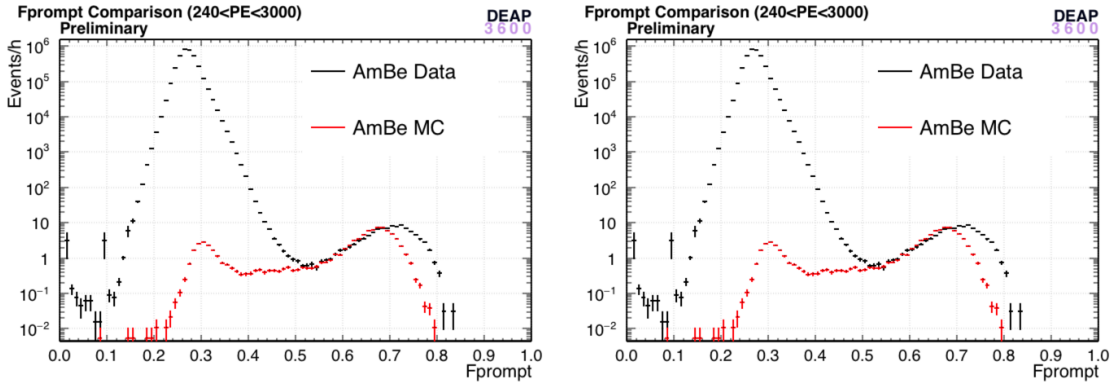
(h)



(i)

Figure 5.17: F_{prompt} distributions comparison between AmBe data and MC in 20 PE slices from 60 PE (a) to 240 PE (i).

The disagreement for all these distributions at low F_{prompt} is well expected as the simulation does not include ^{39}Ar events which are responsible for populating that area at low F_{prompt} (electronic recoils). A comparison of the F_{prompt} distributions in 20 PE slices across the region of interest can be seen in Figure 5.17.



(a) (b)

Figure 5.18: F_{prompt} distributions comparison in the $240 < PE < 3000$ region between AmBe data and MC for cut 5 (a) and 7 (b).

The agreement at high F_{prompt} improves as PE is increased, with there being poor agrement for PE values below 100. We believe this is again due to underestimation in the Cherenkov light production in the AmBe simulation. Figure 5.19 shows the PE comparison for events with F_{prompt} greater than 0.6. AmBe data are slightly above MC in the whole range and work in tuning the simulation optics is on going and should result in a better agreement across the whole spectrum. The large discrepancy at PE below 100 is well understood to be due to Cherenkov events that are not simulated. As Figure 5.16 shows, although the simplified AmBe MC does a good job in matching the F_{prompt} distribution in the basic ROI, when the PE threshold is lowered down to 80 PE, many events that are not simulated start playing a role resulting in a poor agreement. A first attempt to improve this agreement was to study the effect of including 4.4 MeV gammas to the simulation. These events were added to the stage two simulation, livetime normalised and assuming a 64% production rate from the neutron source. This distribution can be seen in Fig.5.21 in light green. Adding these events results in a better agreement in the region $0.5 < F_{prompt} < 0.6$. The other, even larger discrepancy observed when going down to 80 PE, comes in the high F_{prompt} region ($F_{prompt} > 0.8$). As already mentioned, this comes from Cherenkov events leaking in data that are not well simulated in MC. In particular, the hypothesis is that for the AmBe source this Cherenkov light is produced in many light guides from

the many gammas that a neutron produces when entering the detector. That is why we believe the cut on the max charge fraction per PMT is not enough to remove these events from data. To prove this assumption, a study using thorium source data was performed to see if this excess in data could be explained by a Cherenkov spectrum. A population of Cherenkov events was selected from a data set taken with a detector in the same status as for the AmBe runs (LAr phase and same fill level) and background events were subtracted from this sample to make sure that the gammas producing Cherenkov light are unique to the thorium source. This can be seen in Figure 5.20 where the top two plots show the background population for cut 5 and cut 7 and the bottom plots the Cherenkov sample after background subtraction. It should be noted from Fig.5.20(a) and Fig.5.20(b) that no events are seen in the ROI for these background runs. The effect of adding these thorium Cherenkov distribution can be seen in Figure 5.21 (grey curve). These plots show the F_{prompt} distribution in the $80 < PE < 240$ range (ROI) for both cut 5 and cut 7. The other curves present in these plots are: a PSD (pulse shape discrimination) fit applied to the ^{39}Ar spectrum (in blue) and normalised to the peak of the spectrum; the AmBe data from calibration runs (black), the 4.4 MeV gamma simulation discussed above (light green) normalised assuming a 4.6 kHz AmBe source and a 64% gamma production rate from this source; the stage two AmBe MC (in red) normalised to livetime again assuming a 4.6 kHz source and finally the green curve which is the sum of the PSD fit, AmBe MC, 4.4 MeV gammas MC, a flat nuclear recoil spectrum (in magenta) and thorium Cherenkov distribution. This Cherenkov distribution is shape normalised such that it matches data in the $0.5 < F_{prompt} < 0.8$ range. As can be seen from these plots, the 4.4 MeV gammas MC helps in getting a better agreement in the $0.5 < F_{prompt} < 0.6$ region whereas adding the Cherenkov distribution from the thorium source results in a good match in the high F_{prompt} region (above 0.8). This seems to suggest that the reason for discrepancy at high F_{prompt} is related to Cherenkov events surviving the cuts and leaking in the ROI at low PE that are not produced by the simulation.

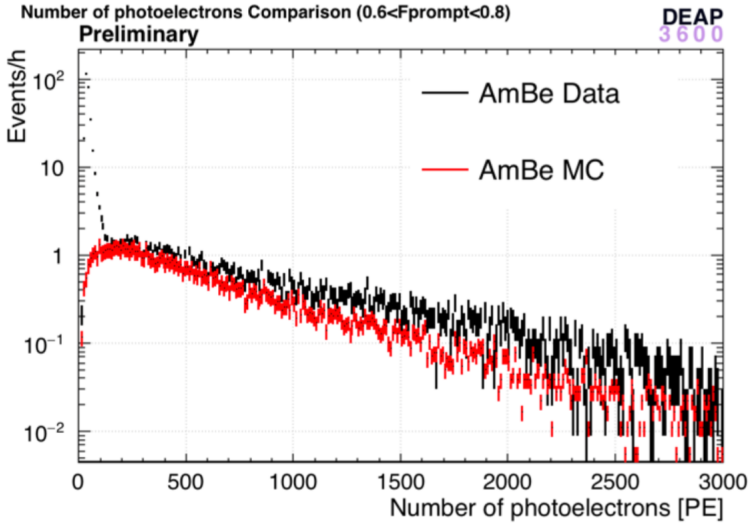


Figure 5.19: PE distributions comparison between AmBe data and MC ($0.6 < F_{prompt} < 0.8$).

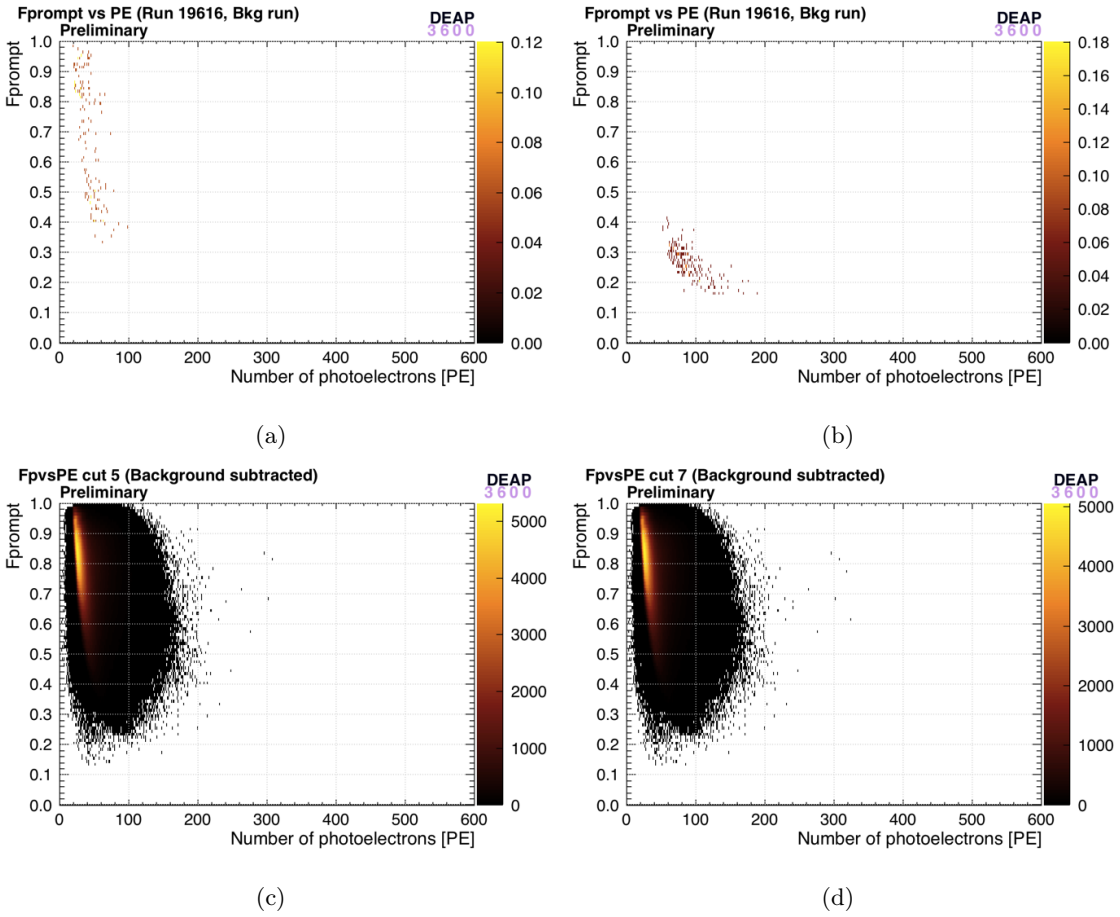
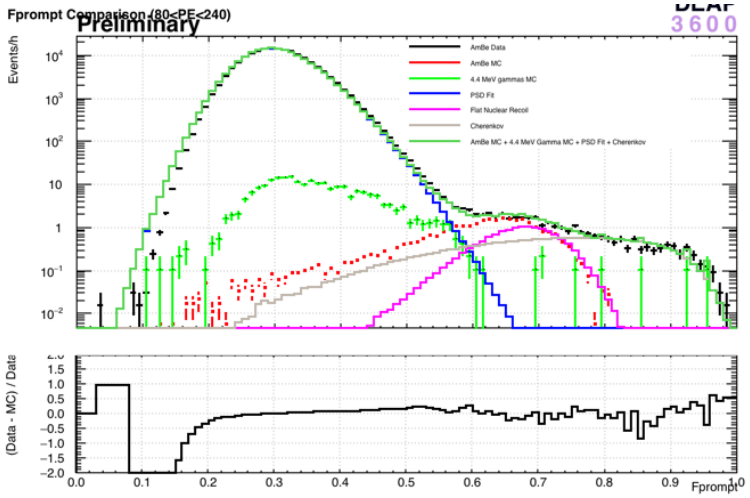
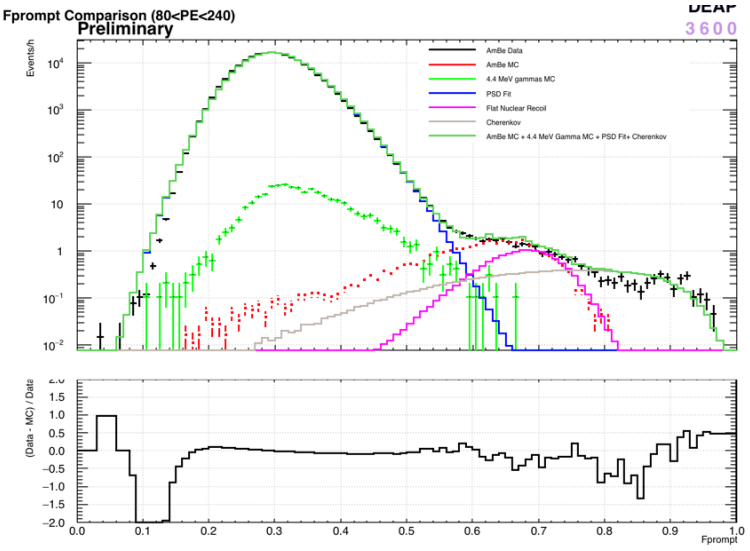


Figure 5.20: Cherenkov selected events for a background run for cut 5 (a) and 7 (b) and the background subtracted thorium run for 5 (c) and 7 (d).



(a)



(b)

Figure 5.21: F_{prompt} distributions for cut 5 and residuals between summed PSD fit model, MC, Cherenkov spectrum and AmBe data (a) and cut 7 (b).

5.6 AmBe Monte Carlo - Data Discrepancy Studies

In order to determine the source of discrepancy seen at high F_{prompt} and high PE some investigations have been undertaken and will be discussed in this section. At first, the possibility that an incorrect AmBe spectrum had been used in the simulation was taken into account. In fact, a variation in the spectrum could result in a different number of inelastic scatters in the target volume that could affect the final F_{prompt} distribution. To

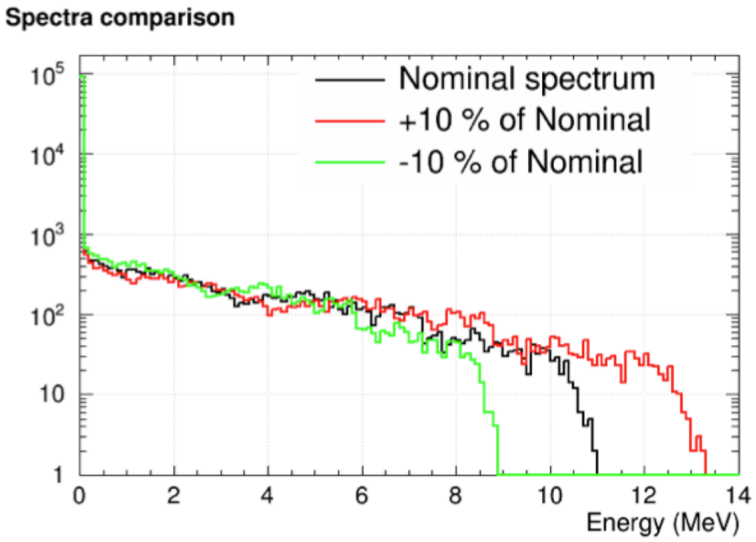


Figure 5.22: Nominal, softer and harder AmBe MC spectrum. Black curve is the nominal spectrum used in stage two. Two additional spectra with 10% increase (red) and 10% decrease in energy were used for simulation to check how this would have affected the F_{prompt} distribution.

disprove this possibility, two simulations with a softer and harder neutron spectra from the AmBe source were generated. Fig.5.22 shows the nominal AmBe spectrum used for the Monte Carlo generation in black, a spectrum where its energy was increased by 10% in red and one with energy decreased by 10%. However, as one can see in Fig.5.23 no changes in the F_{prompt} distribution shape or band peaks were observed.

Another possible explanation investigated was the the presence of detector specific event pathologies (i.e. data/MC disagreement localised to specific regions within the detector). In order to look for these type of events, the high F_{prompt} band was divided into three PE regions which we will refer to as ROI, region 1 and region 2, defined as follows:

- **ROI:** $0.6 < F_{prompt} < 0.8$, $120 < PE < 240$
- **Region 1:** $0.6 < F_{prompt} < 0.8$, $240 < PE < 1000$
- **Region 2:** $0.6 < F_{prompt} < 0.8$, $1000 < PE < 2000$

As can be seen from Figure 5.24, that shows the Z distributions for data and MC for the basic ROI, region 1 and 2, there is no evidence of areas in the Z space with an excess of events for any of the three regions. This was found to be true for X and Y distributions

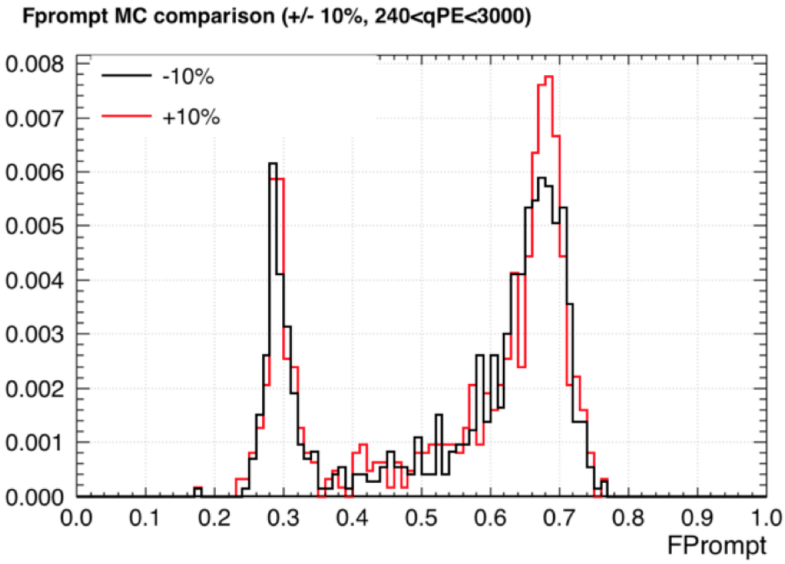


Figure 5.23: F_{prompt} distributions in the $240 < F_{prompt} < 3000$ for the softer and harder AmBe MC spectrum. No difference can be seen in the peak of the high F_{prompt} band.

as well. By comparing the radial distributions instead, a clear excess of events in data at high radius in region 2 which is not present for region 1 can be seen (Figure 5.25). This may be due the fact that the simulation is underestimating the acrylic fluorescence. In particular, it could be possible that the Cherenkov light coming from inelastic gammas in the acrylic is under estimated in the MC. If this is the case, this could explain the excess in data at high radii as this extra light would pull the event's reconstructed position towards the outside of the detector.

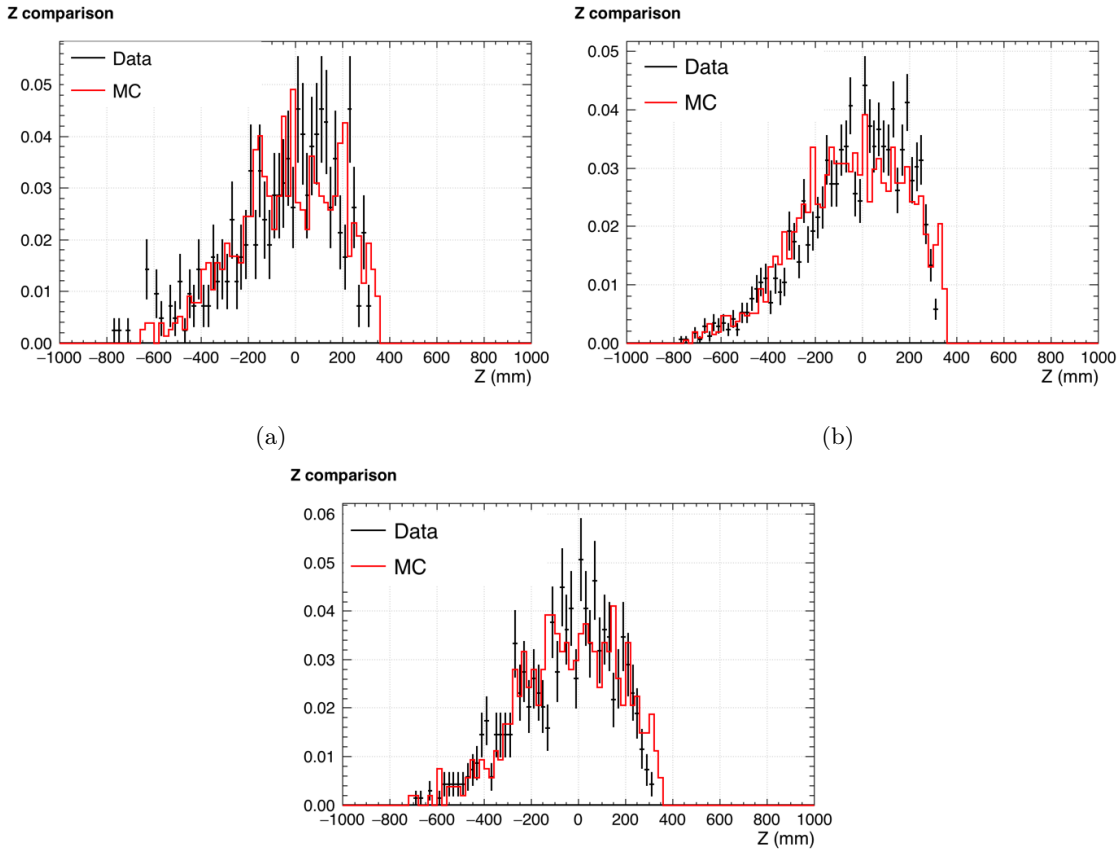


Figure 5.24: Z distributions data-MC comparison for the basic ROI (a), region 1 (b) and region 2 (c).

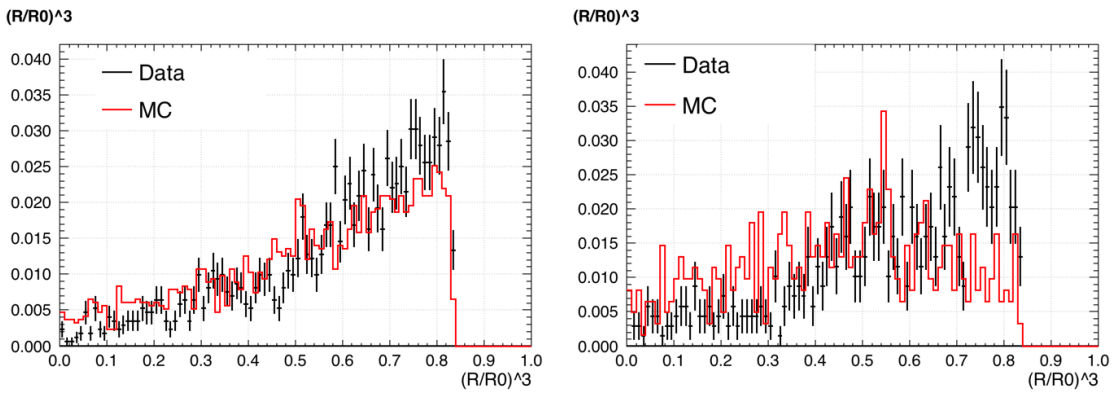
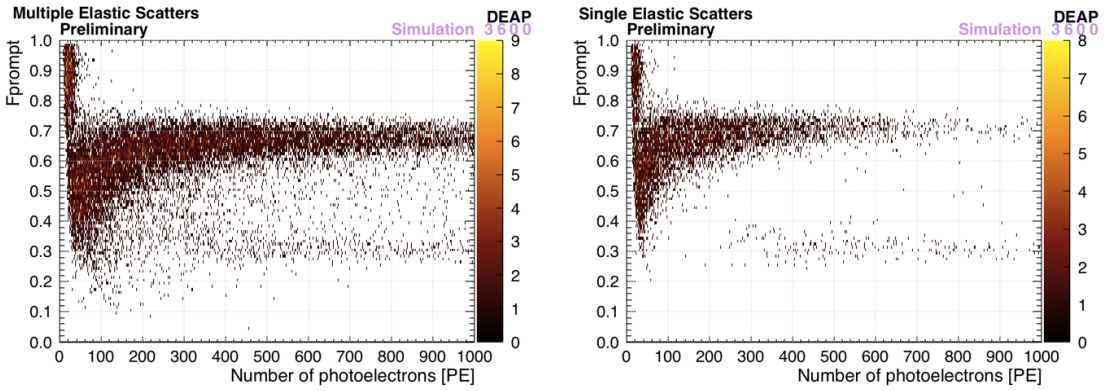


Figure 5.25: $(\frac{R}{R_0})^3$ distributions for region 1 (a) and region 2 (b).

5.7 Single Scatters extraction from Monte Carlo and Data

The ideal way to define the WIMP acceptance would be by using the AmBe calibration data, extracting it from the high F_{prompt} band seen in Fig.4.13. This would be easily feasible if the nuclear recoil band was made of single scatters only. In reality, when neutrons from the AmBe source interact with the detector, most of them will scatter multiple times. This leads to an F_{prompt} band whose shape and position does not match what it would be if only single scatters occur. Although a first attempt to extract a single scatters sample from data was done with the datasets that have been presented in this chapter, a more dedicated study involving machine learning techniques to reach this goal will be the subject of the next chapter. Having shown a good level of agreement between data and Monte Carlo in the region of interest, one can think of extracting the single scatters distribution from data by using simulations by using the truth information from Monte Carlo to separate the true single scatter events from the multiple scatter events. Single and multiple scatters were identified in Monte Carlo and Fig.5.26 shows the F_{prompt} versus PE distribution for ∼3 billion neutrons simulated in stage one, equivalent to 181.2 hours. In particular, Fig.5.26(a) shows multiple scatter events and Fig.5.26(b) the single scatter events.



(a) (b)

Figure 5.26: F_{prompt} versus PE for multiple elastic scatters (a) and single elastic scatters (b).

What can be easily seen from Fig.5.26 is a clear difference between the single and the multiple elastic scatters distributions. It should be noticed that in extracting those single and multiple scatter events, no minimum scatter energy nor spatial or temporal separation between scatters is considered. A study to infer the effect of cutting on these variables was undertaken and concluded that no substantial changes are seen in the distributions

of Fig.5.26. This can be seen for example in Fig.5.27 where the mean F_{prompt} versus the PE distributions are showed for both single and multiple elastic scatter events when a minimum scatter energy cuts are applied. The effect of applying cuts based on spacial separation between scatter locations or on the delta T (time) between two consecutive scatters are shown in Fig.5.28. As one can see from 5.28(a) introducing a spacial cut has no consequences at all whereas the F_{prompt} mean for a Δ T $>$ 10 ns is sligthly pulled down (Fig.5.28(b)). However, this is not a big effect and in what follows none of these cuts have been applied. Fig.5.29 and Fig.5.30 also show the effect of applying these cuts on the energy distributions and total number of scatters respectively. Table 5.4 shows that elastic scatters dominate in DEAP-3600 and that both in the basic and nominal ROI more than 55% of all events are multiple scatters. Given this, we use the ratio multiple-to-single from the simulation to "extract" a single scatter spectrum from the AmBe data. If we assume that what we are left with in the nuclear recoil band in data after all cuts is a mixture of single and multiple scatters event types only, we can write the F_{prompt} for AmBe data as:

$$F_p^{data} = F_p^S \times (1 - R) + F_p^M \times R \tag{5.1}$$

where F_p^S is the "single" scatter spectrum we want to extract from data, F_p^M are the mean F_{prompt} values for multiple scatters from MC and R is the ratio of multiple scatters to total number of scatters. By re-arranging Eq. 5.1 we get:

$$F_p^S = F_p^{data} - \frac{F_p^M \times R}{(1 - R)} \tag{5.2}$$

If the AmBe MC was perfectly modelling real data, this extracted single scatter spectrum F_p^S would of course match the single scatter distribution in the simulation. We will treat any discrepancies between the "extracted" single scatter spectrum and the simulated one as systematic errors due to the Monte Carlo being incomplete. Fig.5.31 shows the results of this approach. In particular, Fig.5.31(a) and Fig.5.31(b) show the extracted "single" scatter mean F_{prompt} distribution, the mean F_{prompt} distributions for data and MC, and the ration of multiple scatters to total scatters for cut 5 and 7 respectively. A residual plot between F_p^S and the MC single scatter spectrum as a function of number of photoelectrons can be seen in Fig.5.31(c) and Fig.5.31(d) again for the two different cuts. These two plots, together with the Gaussian fits shown in Fig.5.31(e) and Fig.5.31(f) show that the extracted single AmBe data is about 0.03 higher in F_{prompt} than the single elastic from

MC for both cut 5 and 7. This value can be set as a systematic error on the F_{prompt} distribution.

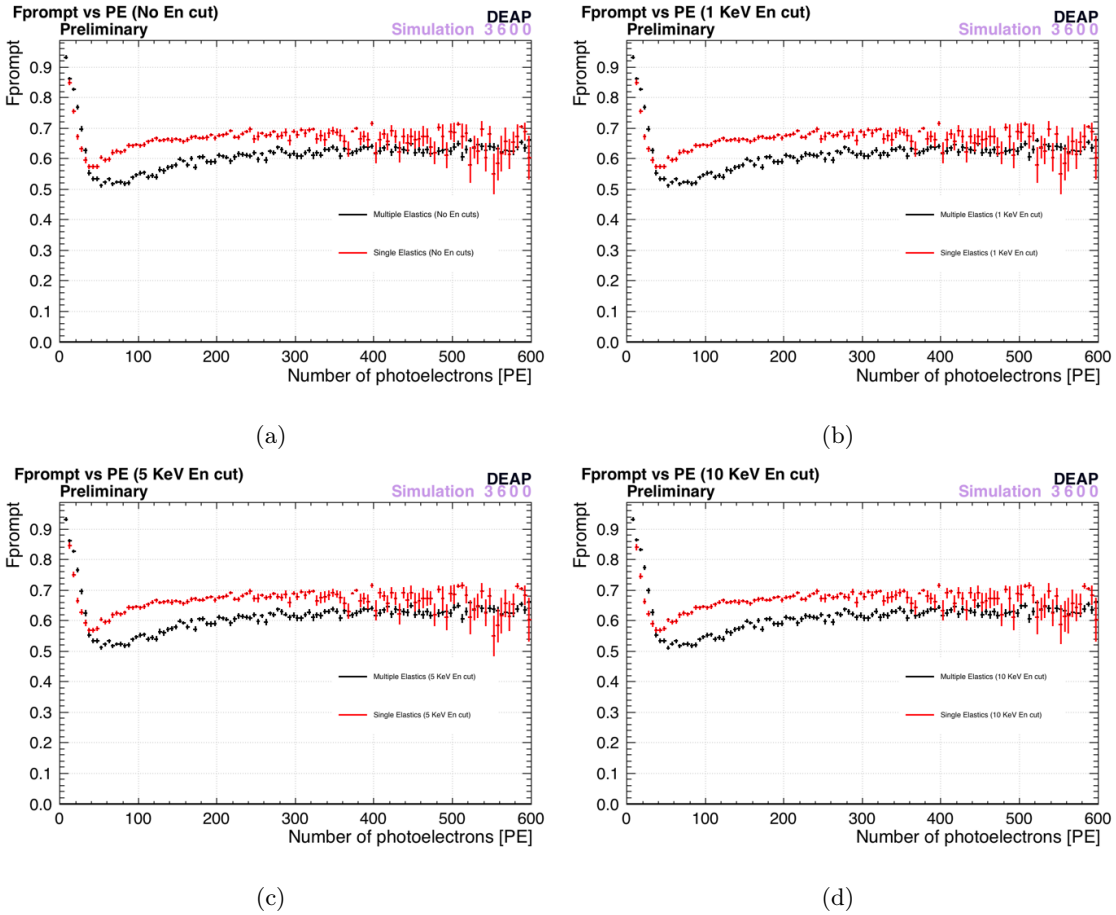


Figure 5.27: Mean F_{prompt} versus PE distributions for different minimum energy required: (a) no energy cut, (b) $E_n > 1$ keV, (c) $E_n > 5$ keV and $E_n > 10$ keV.

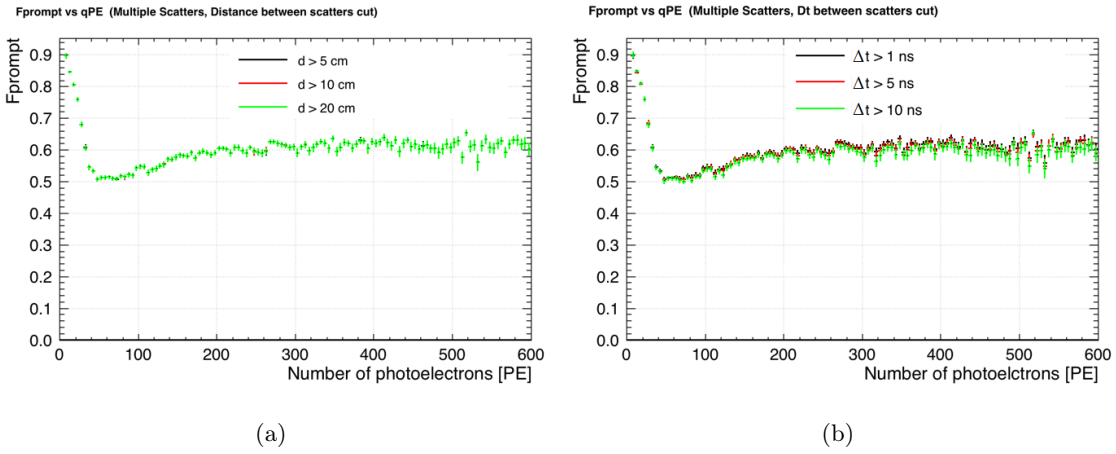
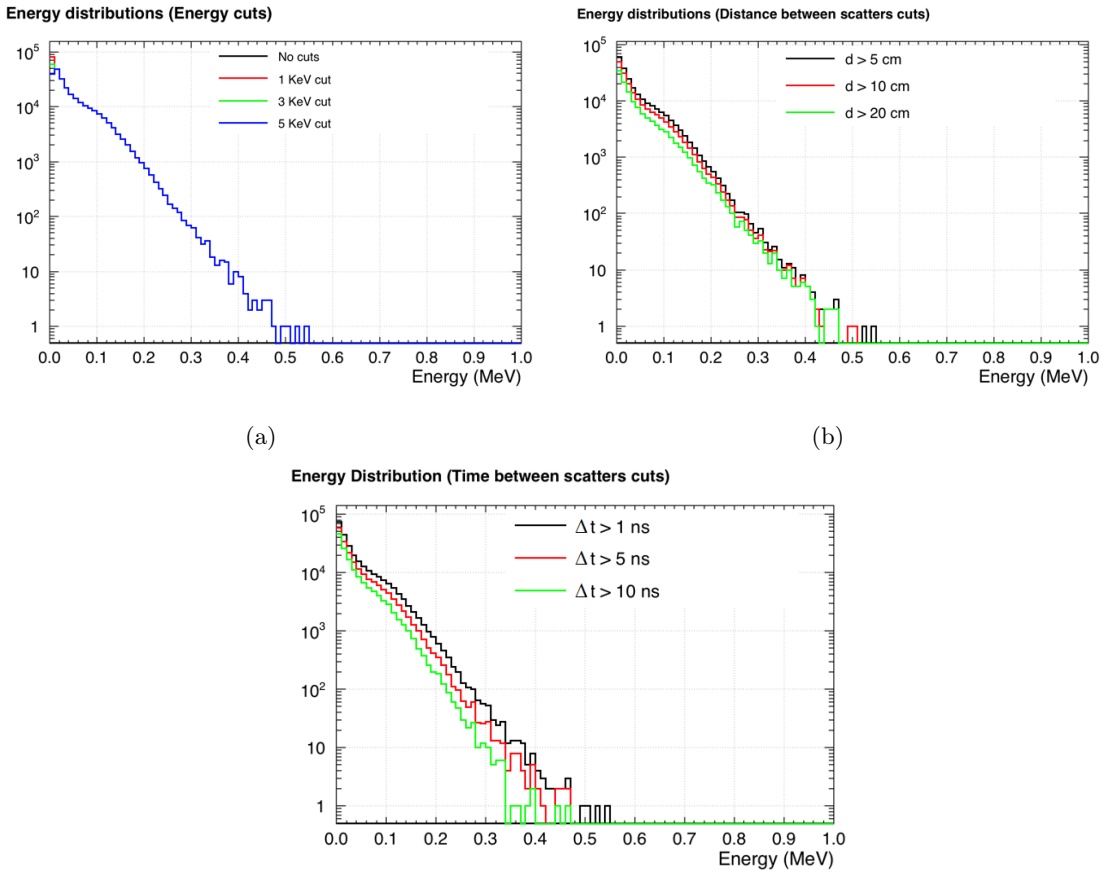


Figure 5.28: Mean F_{prompt} versus PE distributions: (a) distance between scatters cut, (b) difference in time between scatters.



(a) (b)

(c)

Figure 5.29: Energy distributions as a function of minimum energy (a), distance between scatters (b) and difference in time between scatters (c).

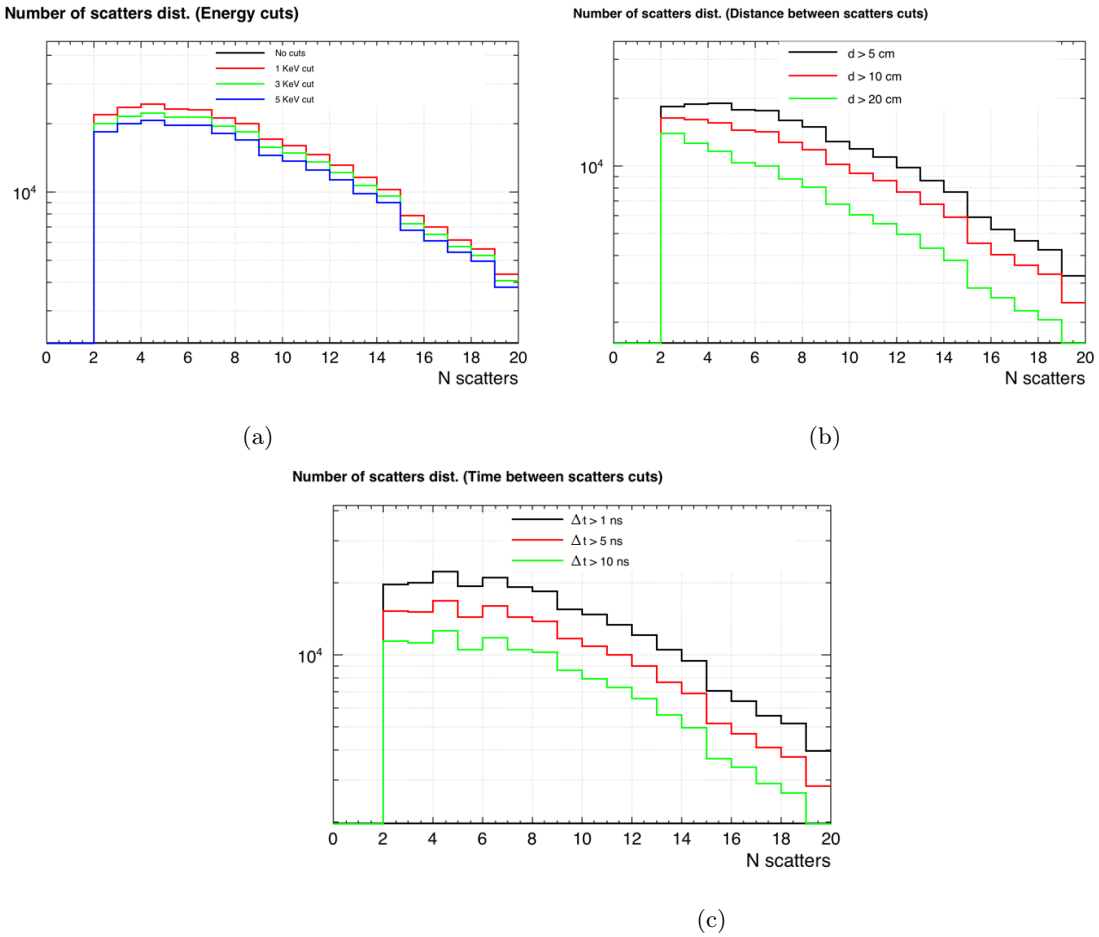


Figure 5.30: Distribution of number of total scatters as a function of minimum energy (a), distance between scatters (b) and difference in time between scatters (c).

Event type	**Events in basic ROI**	%	**Events in ROI**	%
Single elestic	1400±37	41.0	1686±41	42.5
Multiple elastic	2053±45	58.0	2245±47	56.5
Other	-	1.0	-	1.0

Table 5.4: Classification of neutron interactions in basic and nominal ROI.

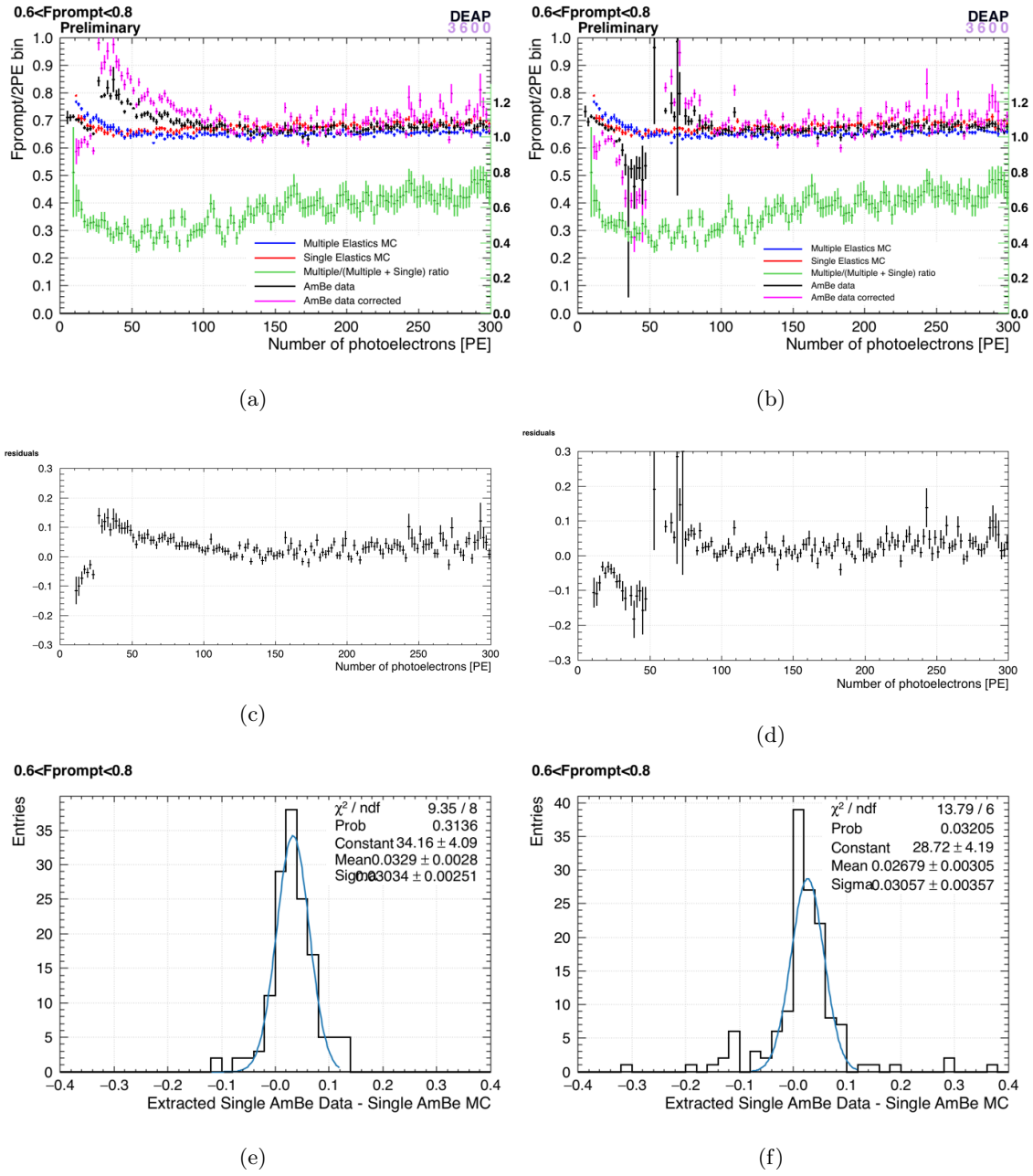


Figure 5.31: (a) Extracted "single" scatter spectrum for cuts 5 and 7 (b) in the $0 < PE < 300$ range and after background subtraction. The right axis shows the multiple scatter fraction (green) in both plots. In black is shown the mean F_{prompt} value for each 2 PE bin for data (black), single scatter MC (red), multiple scatter MC (blue) and the extracted "single" scatter data (magenta). Subfigure (c) and (d) show the residual between the magenta and red curves for cuts 5 and 7 respectively. (e) and (f) show how the systematic can be extracted comparing magenta and red curves over the ROI.

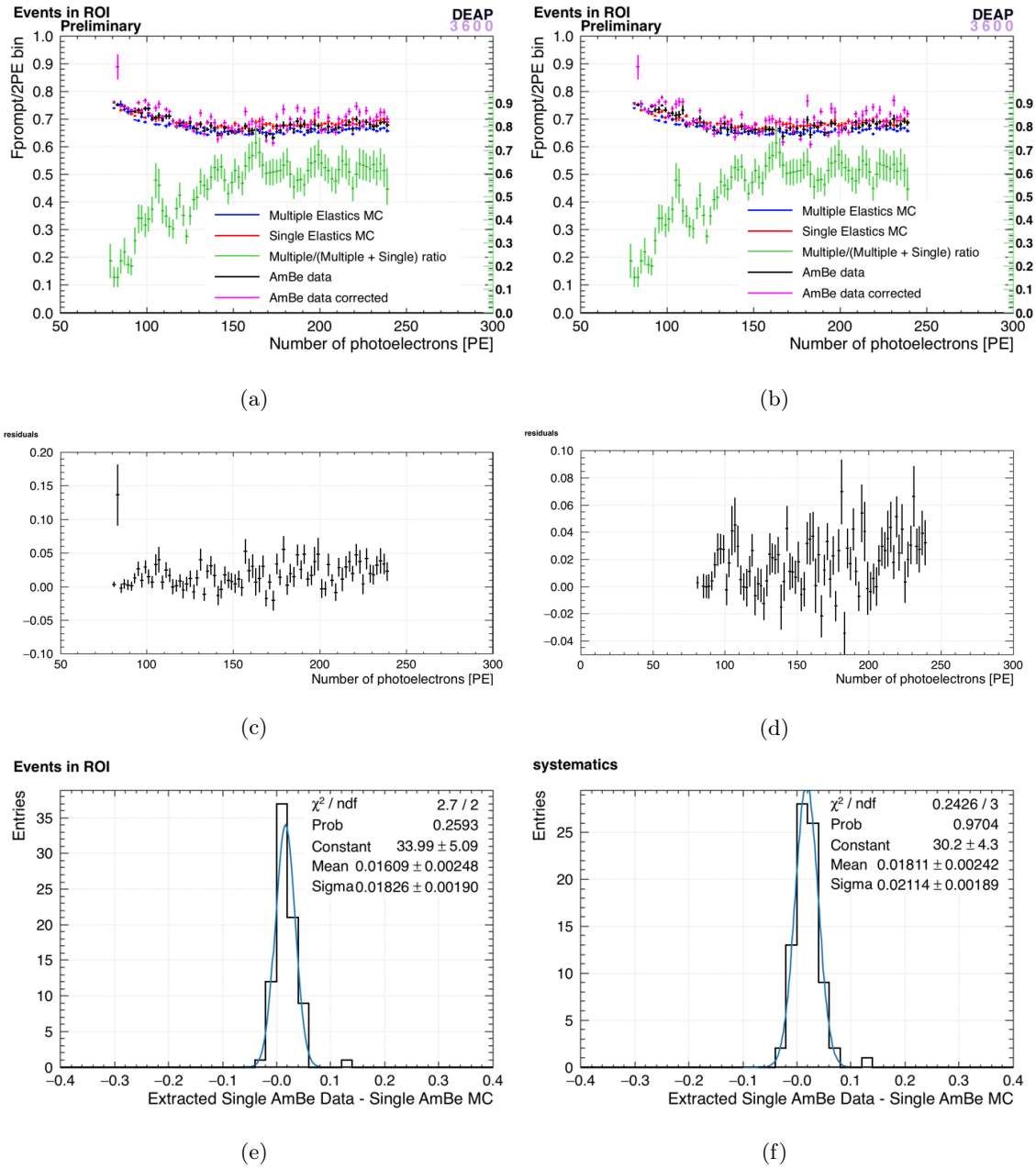


Figure 5.32: (a) Extracted "single" scatter spectrum for cut 5 and 7 (b) in the ROI and after background subtraction. The right axis shows the multiple scatter fraction (green) in both plots. In black is shown the mean F_{prompt} value for each 2 PE bin for data (black), single scatter MC (red), multiple scatter MC (blue) and the extracted "single" scatter data (magenta). (c) and (d) show the residual between the magenta and red curves for cuts 5 and 7 respectively. (e) and (f) show how the systematic can be extracted comparing magenta and red curves over the ROI.

5.8 Summary

To summarise, the two stage AmBe MC generated with the SCENE data used to define the singlet-to-triplet ratios, quenching factors and with a fit to the alpha data does not fully agree with the AmBe data across the whole energy range. The agreement is quite good in the basic ROI ($120 < \text{PE} < 240$) but there is disagreement when looking at very low PE and for PEs > 240. To use this MC to determine our acceptance and set systematics for the analysis of the first physics data taken by DEAP-3600, we need our data to be well described at least across the region of interest. However, an excess in data at high F_{prompt} was found when comparing data and MC in the nominal ROI ($80 < \text{PE} < 240$). We believe that this is due to the fact that the AmBe source produces a lot of Cherenkov events at high F_{prompt} that survive both cut 5 and 7 cuts and leak into the ROI. The fact that the spectrum of these events matches the shape of Cherenkov events selected from a Thorium source run validates this hypothesis. We have shown that a combination of the "simplified" AmBe stage two MC, 4.4 MeV gammas MC and the use of a Cherenkov spectrum from a thorium source does a good job in matching the AmBe data in the whole ROI. This work has also estimated systematic shift on the F_{prompt} distribution in MC of 0.03. Work is currently on going to generate a more complete stage one MC including 4.4 MeV gammas and in selecting more gammas to get a more realistic simulation of the AmBe source with the detector. Efforts are underway to better understand the Cherenkov production in DEAP-3600 from data and in developing a more stringent set of cuts to remove these events as this work has shown that this is the main reason why the AmBe MC cannot fully reproduce real data.

Chapter 6

Single scatters extraction with Machine Learning

In this chapter a brief introduction on Machine Learning (ML), mainly focusing on some classification techniques as they will be relevant to the analysis that will be presented - including the results of applying these techniques to neutron calibration data with the goal of extracting a clean sample of single scatters - will be given. The datasets used for the studies presented in this chapter differ from those in Chapter 5, the AmBe data includes runs collected up to July 2018 with updated AmBe Monte Carlo (both in stage one and two) with respect to what was presented in Chapter 5. In particular, the new AmBe Monte Carlo, includes the 4.4 MeV gamma in stage one and has been reprocessed with the same processors and RAT version as for data.

6.1 Introduction

ML has become over the last decade a powerful and popular tool for multidisciplinary scientific research. The goal of ML is to build predictive models that "learn" from some data sample and generalise the underlying distribution that generated the sampled data in order to make predictions on new un-seen data. Machine learning is used in a variety of fields and its popularity among the physics community is increasing. A toolkit for machine learning, TMVA, exists within CERN's ROOT package and many open source libraries are becoming available and easy to use. An example is the Scikit-learn library, which is an open source Python based machine learning library that with its API makes it very straightforward to use a huge variety of machine learning algorithms to tackle any kind of

problem. Both TMVA and Scikit-learn have been used to develop a machine learning cut to isolate single scatters and the results will be presented in the following sections.

6.2 Machine Learning with TMVA and Scikit-learn

The Toolkit for Multivariate Data Analysis (TMVA) is a Machine Learning tool in ROOT (the CERN toolkit largely used in the high energy physics community for data analysis), that provides several multivariate analysis techniques for signal/background classification. Some of algorithms included in the TMVA packages are [72]:

- Multi-dimensional likelihood estimation (K-Nearest Neighbour (KNN) and probability density estimator range-search (PDERS));
- Linear and nonlinear discriminant analysis (Fisher, H-Matrix, Functional Discriminant Analysis);
- Artificial neural networks;
- Support Vector Machine (SVM);
- Boosted/bagged decision trees with pruning (BDT);

Scikit-learn [73] is a very popular open source Python machine learning library that features many classification, regression and clustering algorithms for supervised and unsupervised learning. Scikit-learn offers a unified API shared by all algorithms in the library that follows the *fit, predict/transform* paradigm and makes it very easy to use. It only accepts numpy (another popular Python library for science) arrays or sparse matrix as input/output. A library called *root_numpy* makes it possible for root trees to be easily transformed into numpy arrays and transform them back into ROOT tree objects at the end of the learning/predicting process.
Both TMVA and scikit-learn have been used in this work. A description of the strategy and of some of the algorithms used will be given in the next sections and results will be discussed.

6.3 Binary classification

Binary classification refers to the task of classifying elements that belong to one of two classes. An example can be a program that aims to identify whether an email is spam or

not spam, or a medical test trying to determine if a patient has a disease or not, etc. In machine learning, binary classification is a type of supervised learning. A machine learning problem is called supervised when the input data are labelled, i.e. when both the input variable $\vec{x}$ and the output variable $\vec{y}$ are known in the training phase, and the model has to "learn" the mapping function $\vec{y} = f(\vec{x})$ between them from labelled training data, as opposed to unsupervised learning when only the input variable $\vec{x}$ is known and the goal is to model the underlying structure in the data (an algorithm that aims to categorise documents for topics without knowing them *a priori* is an example of unsupersived learning). The task of identifying single scatters among neutron scatters from the AmBe source that can also (and mainly) scatter multiple times is a binary classification problem. A variety of algorithms can be used for this type of classification, from K-Nearest Neighbour to boosted decision trees. An in depth discussion of these algorithms is beyond the scope of this work, however, a brief discussion of the models used in this work will be given in the following sections.

Evaluation metrics

Evaluating the performance of a model requires the use of some metrics. The confusion matrix is often used for this purpose in classification tasks. It is a 2×2 table that contains the four outcomes of a binary classification as in Table 6.1

		Predicted	
		Negative	Positive
Actual	Negative	TN	FP
	Positive	FN	TP

Table 6.1: Confusion matrix for binary classification.

If we imagine having two classes, positive and negative, true positive (TP) are those correctly identified, true negative (TN) those correctly rejected, false positive (FP) the incorrectly identified and false negative (FN) the incorrectly rejected. Some of the metrics that can be extracted from a two class matrix are as below:

- **Accuracy**: (TP + TN)/total, defines how often a classifier is correct
- **Misclassification rate**: 1 - accuracy, defines how often a classifier is wrong,
- **Recall or True Positive Rate (TPR)**: TP/(TP + FN), measures the proportion of actual positives that are correctly classified,

- **Precision**: TP/(TP + FP), when the predicted classes is positive, measures how often the prediction is correct,
- **False Positive Rate (FPR)**: FP/(FP + TN), when the actual class is negative, measures how often the classifier predicts a positive result instead.

Another way to examine the performance of a classifier is by plotting TPR versus FPR in what is known as a Receiver Operating Characteristic (ROC) curve (see figure 6.1). The point (0,1) in this graph represents a perfect classifier: all positive and negative cases are correctly classified. The red line in the graph shows the performance of a random guess. Any model that should be taken under consideration (that has "learnt" from training data) should do better than a random guess and its ROC curve should be above this red line. The closer to the top left corner the kink of this curve is the better the classifier is performing. Another term that can be extracted from the ROC curve and used as an evaluation metric is the area under the ROC curve (AUC). From what has just been stated, it is easy to understand that the closer to 1 the AUC is the better the classifier is performing. It should be noticed that ROC and AUC can be misleading in the presence of an imbalance in the two classes. If for example we are in presence of a very imbalanced dataset where 99% of the data belong to the first class and 1% to the other, and a classifier is predicting all observations to fall into the first class, we would get a very high TPR (thus a high AUC), but none of the observations belonging to the other class would actually be predicted correctly. Precision and Recall are preferable in this case.

Overfitting and Bias-Variance tradeoff

The ultimate goal of the machine learning process is to train a model which has the ability to make predictions that generalise well to future and unseen data. What can often happen in the model building process is to build a model that learns the structure of the training data so well that it can correctly predicts every example of this data set but it has a very low prediction ability on new data. This is known as overfitting. In general, the more complex the model, the more likely it overfits the data. This is because a complex model will capture every single detail (noise) in the training data rather than the "general" relationships between variables in the data. The model has in this case high variance, meaning that it is too sensitive to small fluctuations in the training set. The opposite scenario is when the model does not fit well the training data and hence is missing any

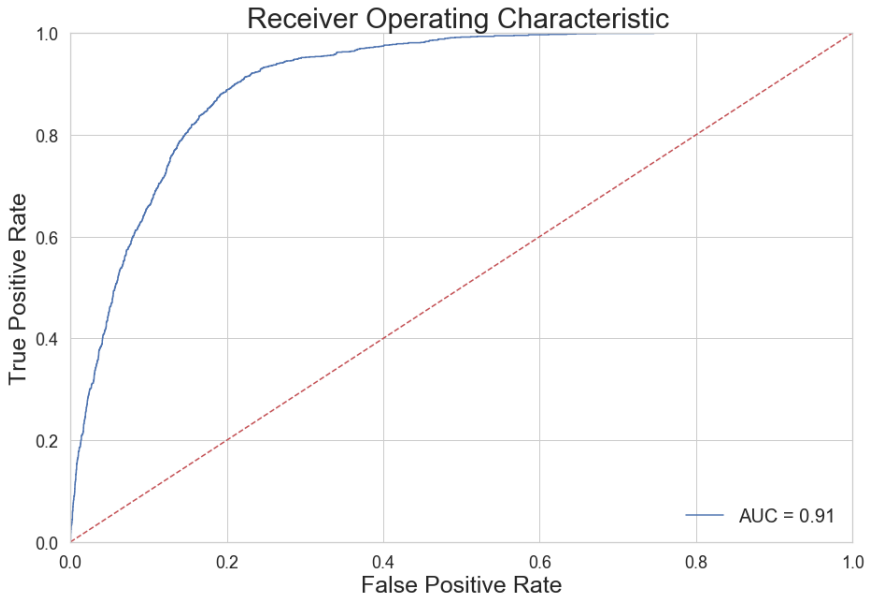


Figure 6.1: Receiver Operating Characteristic curve. Top left corner (0,1) would give a perfect classifier (AUC = 1). Dashed red line represents the performance of a random guess.

trend in them. Again, this does not allow any generalisation to new data. A model that is too simple will underfit the data. An underfitting model has high bias, which comes from erroneous assumptions in the learning algorithm. The bias-variance behaviour as a function of model complexity is sketched in Fig. 6.2 which shows how in order to avoid both underfitting and overfitting, one needs to choose a model whose complexity gives the best trade-off between variance and bias to get a generalisation error as small as possible. This will be discussed in more details in the following sections.

Train/test split and Cross Validation

A common way to check for the occurrence of an over/under-fit and to test the model performance is to split the data set into two separate chunks: a training set used to train the model, and a test set used to test its performance. A train/test split is usually performed using a 70-30 ratio but this purely depends on the dataset size and is a choice of the model builder. The scheme of Fig. 6.3 shows a resampling technique known as a k-fold cross validation that is also used for estimating model performance and operates in a similar way as the train/test split. In k-fold cross validation, the dataset is randomly split intos k sets of roughly the same size. At the first iteration k=1 the model is trained using all the samples except the first subset that is used to estimate its performance, the second

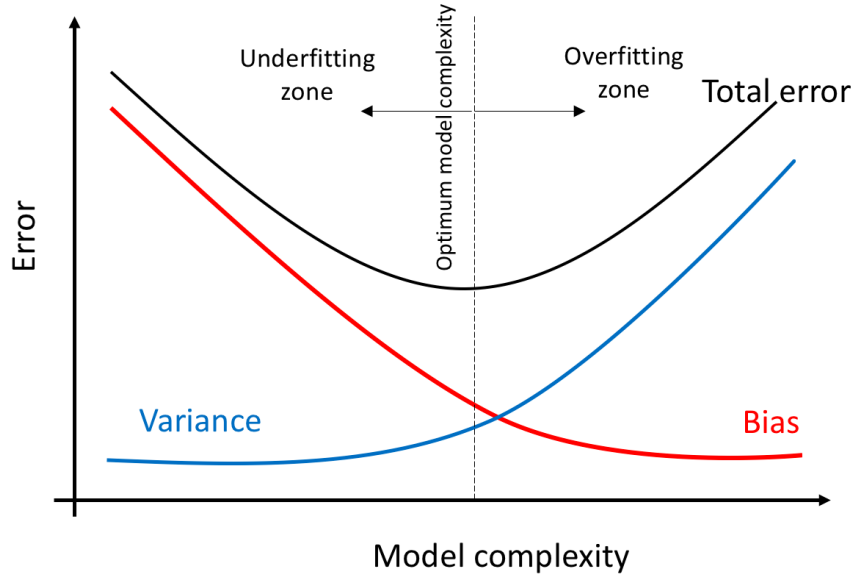


Figure 6.2: Bias and Variance contribution to the total error. A tradeoff between bias and variance is important to avoid overfitting or underfitting.

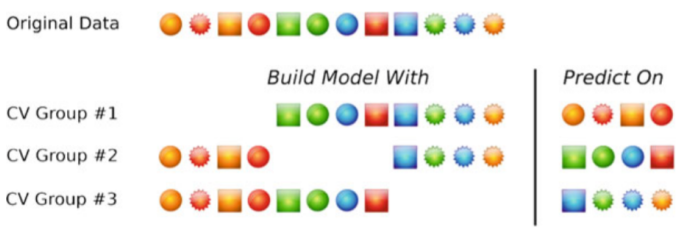


Figure 6.3: Schematic representation of a three-fold cross-validation. From the 12 examples in the training set 4 of them are left out each of the 3 groups the training set is divided into and used to evaluate the model performance recursively. The cross-validation estimate of model performance is the average of the 3 performance estimates [74].

iteration uses all the samples (now including the first subset) except for the second subset where the model will be evaluated again and so on for all the k folds. In this way, the true error is estimated as an average error at the end of the k^{th}-iteration giving a more accurate estimation and all the examples in the dataset are eventually used for both training and testing. Common choices for k are 3, 5 or 10 but this highly depends on the dataset size.

K-Nearest Neighbour (KNN)

A KNN algorithm can be used both for classification and regression tasks. In KNN for binary classification, the output is one of the two classes. The idea behind this algorithm is very simple: an element is classified as belonging to class A or B by a majority vote

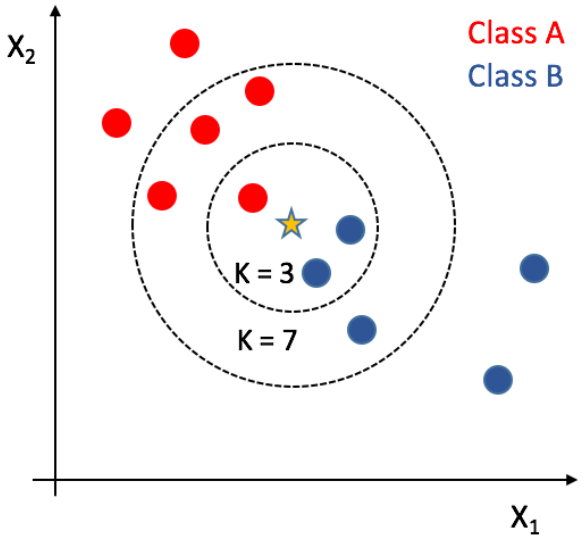


Figure 6.4: Schematic representation of KNN classification for a binary problem.

of its neighbours. K is the number of neighbours to consider and is an integer value that can be chosen and optimised. For example, the test point (yellow star) in Fig. 6.4 will be classified as belonging to class B with K=3 and to class A when K=7. A good choice of K is important for better performance and can be achieved by looking at the classification error in a given validation set as a function of K. KNN is very easy to understand and interpret but computationally expensive as it needs to store all the training data each time and the prediction stage may be slow. Because of its simplicity, it is often used as a baseline model for classification and regression.

Logistic Regression

Logistic regression is a statistical method often used in machine learning for binary classification. During the training process, a weight vector $\vec{w}$ and an offset b are computed such that for any input $\vec{x}$ the function

$$f(\vec{x}) = \frac{1}{1 + e^{-(\langle \vec{x} | \vec{w} - b \rangle)}} \tag{6.1}$$

returns values closer to zero for one class and closer to one for the other as shown in Fig.6.1 together with the decision bound that determines to which class the element belongs to, set to 0.5. This function is referred to as sigmoid function.

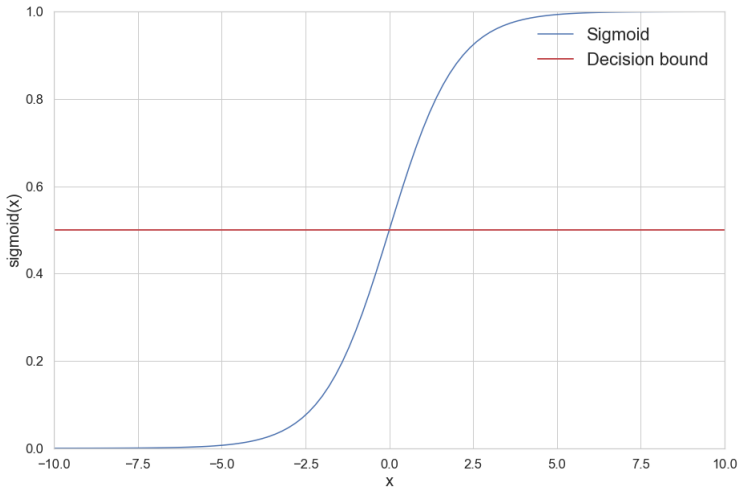


Figure 6.5: Sigmoid activation function which takes any real value and outputs a value between zero and one. The decision bound (red horizontal line) determines which classes an element belongs to in a binary classification problem.

Fisher discriminant

Linear Discriminant Analysis (LDA) is a commonly used technique for data classification and dimensionality reduction. Fisher's linear discriminant is a classification method that reduce high-dimensional data by projecting them onto a line in a one-dimensional space. The projection is done by simultaneously maximising the distance between the means of the two classes one wants to separate, and minimising the variance within each class. This makes it possible to find a line such that samples from each class are well separated. A detailed derivation of this method can be found in [75]. Restricting our attention to the two-classes case (the method can be generalised to C classes), let us imagine that of the D-dimensional samples $X = \{x^{(1)}, x^{(2)}, ...x^{(m)}\}$, N_1 belong to the class C_1 and N_2 to the class C_2. The goal is to obtain a scalar y by projecting the samples X onto a line:

$$y = \theta^T X \tag{6.2}$$

The objective function $J(\theta)$ for these two classes is then defined as:

$$J(\theta) = \left(\frac{\hat{\mu_2} - \hat{\mu_1}}{\hat{s_1^2} + \hat{s_2^2}} \right)^2 \tag{6.3}$$

where $\hat{\mu_k}$ (with $k = 1, 2$) represent the mean vectors of the two classes and $\hat{s_k}$ ($k = 1, 2$) the within-class variance for each class C_k. Solving the Fisher problem means maximising the objective function $J(\theta)$, such that the points belonging to the same class get projected

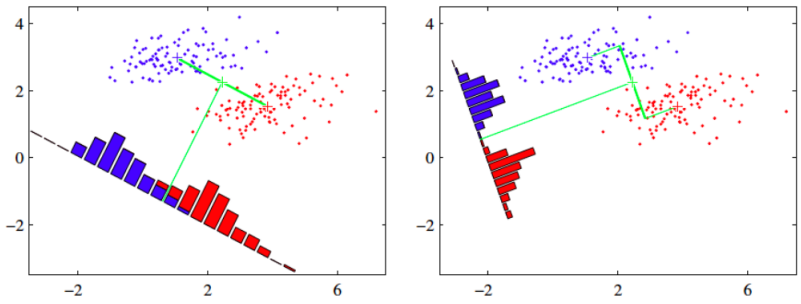


Figure 6.6: Left: two classes (red and blue) along with the histograms resulting from projection onto the line joining the class means overlapping in the projected space. Right: projection based on the Fisher linear discriminant, with improved class separation [75].

very close to each other while their projected means are as far apart as possible.

The optimum θ^* that achieves this goal is known as Fisher's linear discriminant.

Decision trees

A decision tree is a very common and relatively easy predictive model used for classification and regression problems. The algorithm consists of simple decision rules inferred from the data features that splits the data sequentially until a stop criterion is satisfied and the value of a target variable is predicted as belonging to one of two classes[1] (in the case of a binary classification problem). Fig. 6.7 shows a schematic view of a decision tree. Given a dataset of the form: $(\mathbf{x}, Y) = (x_1, x_2, ..., x_k, Y)$, where x_k are the features and Y the target variable to predict, the scheme in the figure shows how repeated yes/no decisions are taken on one feature variable at a time until the stop criterion gets satisfied. This sequence of binary splits leads to a split of the phase space into many regions that will be eventually classified as belonging to the positive or negative class by using a majority vote rule. In fact, the number of training events that end up in the final leaf node will determine whether that node should be classified as positive or negative. Decision trees can handle huge datasets and in the case of small trees they are easy to interpret, but they are very sensitive to the specific data on which they are trained and so have high variance.

[1] In general, classification tree can be used with multiple classes and regression tree to predict outcomes that take continuous values.

Ensemble learning

Ensemble learning methods combine several machine learning algorithms into one predictive model to obtain better predictive power and stability than any single constituent algorithm alone. Voting and averaging ensemble methods are used for classification and regression respectively. In a classification task, the final output prediction of an ensemble learner algorithm will be the one that receives more than 50% of the votes from each constituent algorithm. Random Forest, Extra Trees and Boosted decision trees are common ensemble learning methods and will be discussed in some detail in what follows.

Random Forest

Random Forest utilises a multitude of decision trees in the training phase and the bootstrap aggregating (bagging) technique to the tree learners. Bagging is a statistical procedure that can successfully reduce the variance for algorithms such as Decision Trees that have high variance. It consists of generating m new training sets D_i out of the original dataset D of size n, by randomly sampling from it uniformly with replacement. m "bags" of size $n^{'}$ are generated in this fashion. Given a dataset $(\mathbf{x}, Y) = (x_1, x_2, ..., x_k, Y)$ as the one described in the Decision Trees section, the Random Forest method operates as follow [74]:

1. For $b = 1, 2, ...B$:
   (a) Sample n training examples from $(\mathbf{x}, Y)$ with replacement (bootstrapped data)
   (b) Train a classification or regression tree T_b to the bootstrapped data, by recursively repeating the next steps for each terminal node of the tree, until the minimum node size n_{min} is reached
      i. Select randomly r variables from the p input variables
      ii. Pick the best input variable/split-point among the r according to some criteria (Gini/entropy)
      iii. Split the node into 2 daughter nodes
2. Output the B ensemble of trees $\{T_b\}_1^B$

For unseen new samples x, predictions are made by taking the majority vote in the case of classification or by averaging the predictions from all the individual regression trees on the x samples for regression.

Extra Trees

Extra Trees or extremely randomised trees, are the same as Random Forests but with the addition of another step of randomisation. Namely, in the Random Forest the splits of the trees are deterministic whereas they are random in the case of Extra Trees. Also, Extra Trees do not use the bagging process such that samples are drawn from the entire training set.

Boosted Decision Trees (BDT)

Decision tree algorithm are very powerful but unstable. In fact, small changes in the training data can result in large changes in the tree. To overcome this issue a technique formulated by Schapire in 1990 called boosting was introduced [76]. Boosting consists of increasing the weight of misclassified training events. In this way a re-weighed training event sample is generated and a new decision tree can be developed. BDT in TMVA (and often in general) utilises the so-called *AdaBoost* (adaptive boost) as its boosting algorithm. Given a training sample $\vec{x} = x_1, ..., x_N$ consisting of N events x_i with known classification result y_i and weight w_i for every $x_i (i = 1, ..., N)$, and considering a classifier $G(x)$ that will be boosted M times, the BDT algorithm can be summarise as follow [77]:

1. For $m = 1$ to M

   (a) Train a classifier $G_m(x)$ on the training sample

   (b) Compute the misclassification rate

   $$err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_m(x_i)}{\sum_{i=1}^{N} w_i} \tag{6.4}$$

   where $I(true) = 1$ and $I(false) = 0$. If $err_m \geq 0.5$ set $M \leftarrow m$, i.e. stop the loop after this iteration.

   (c) Compute classifier weight $\alpha_m = log(\frac{1-err_m}{err_m})$

   (d) Re-weight training sample

   $$w_i \leftarrow w_i exp[\alpha_m I(y_i \neq G_m(x_i))] \tag{6.5}$$

   for $i = 1, ..., N$.

2. Output $G(x) = \sum_{m=1}^{M} \alpha_m G_m(x)$

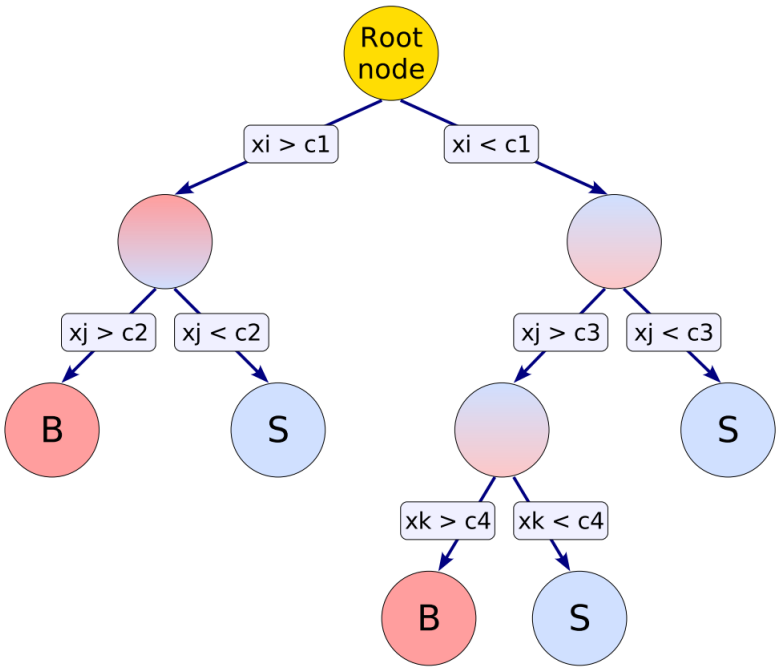
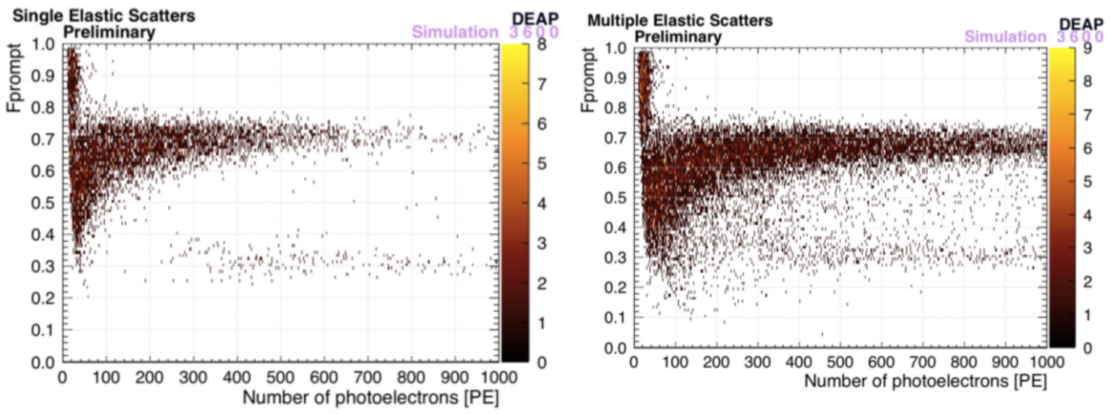


Figure 6.7: Schematic of a decision tree showing a sequence of binary splits using the feature variables x_i. At the bottom of the tree are the leaf nodes that predict the target variables nature. They are labelled "S" for signal and "B" for background depending on the majority of events that end up in the respective nodes [72].

A schematic of the split process in a decision tree is shown in Fig. 6.7.

6.4 TMVA studies

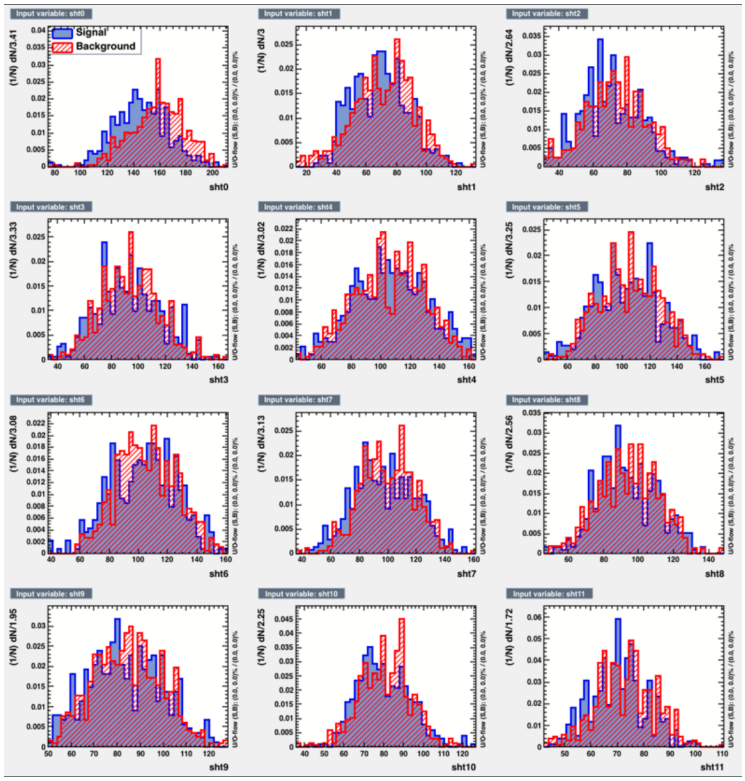
TMVA was the first library used to idenitfy single scatters among the nuclear recoil band populated by the AmBe neutron source. For a binary classification problem, TMVA needs two types of input files: *signal* and *background*. During the training process, these two types of files will be used as labelled data to "teach" the algorithms what kind of events it is dealing with. It was decided to use the Monte Carlo truth information to train the models. Fig. 6.8 shows the populations of Monte Carlo singles (6.8(a)) and multiples (6.8(b)) used to train the algorithms, with singles being the signal and multiples the background for TMVA.



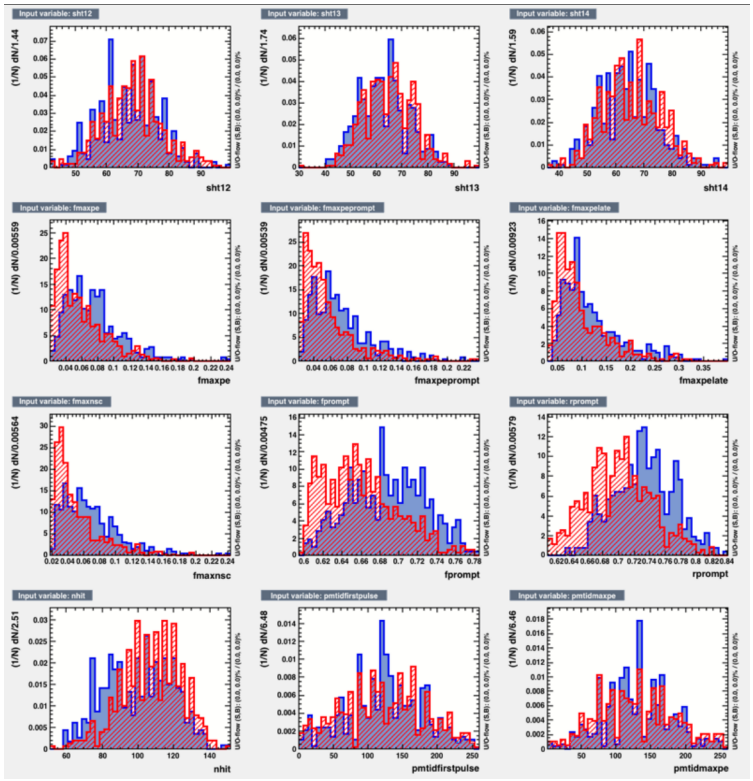
(a) (b)

Figure 6.8: F_{prompt} versus number of photoelectrons for the AmBe Monte Carlo single scatters (a) and multiple scatters (b).

As mentioned above, TMVA offers the possibility to choose between many classification algorithms and produces a set of sanity checks and informative plots together with creating root output files for further analysis. It also gives the ability to choose how to split the sample between train and test sets and apply cuts. The results presented in this chapter focus on the first analysis paper ROI and make use of the whole 2017 AmBe data set (Table 4.1). The train/test split is set to be 70:30 and the Monte Carlo is the same as in Sec. 5.4.



(a)



(b)

Figure 6.9: Input variables utilised to train the algorithms signal (blue) and background (red) distributions.

The set of selected features is as follows:

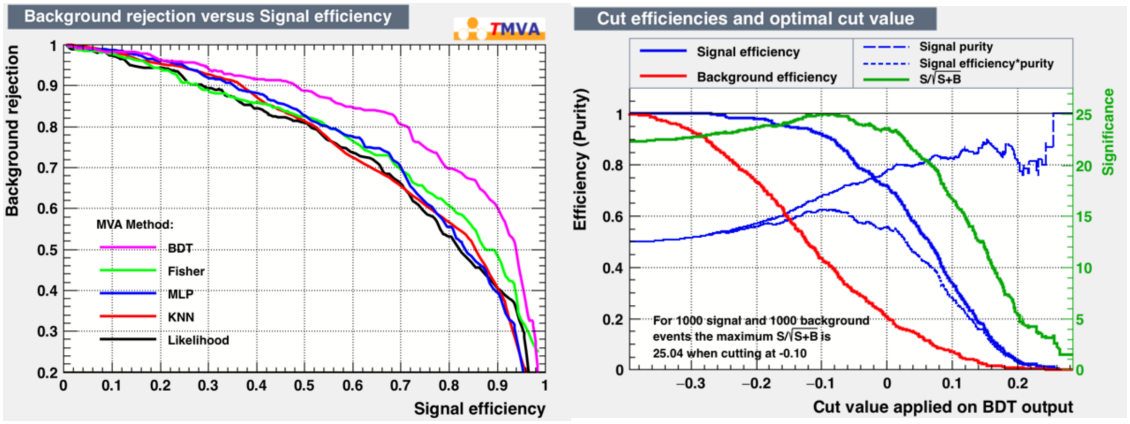
- *15 spherical harmonic coefficients*
- *fmaxpe*
- *fmaxpeprompt*
- *fmaxpelate*
- *fmaxnsc*
- *fprompt*
- *rprompt*
- *nhit*
- *pmtidfirstpulse*
- *pmtidmaxpe*
- *pmtidmaxnsc*

Their distributions in both signal (blue) and background (red) can be seen in Fig. 6.9(a) and Fig. 6.9(b). A set of different algorithms was also used and their performances compared in order to select the one that performs best on the dataset. They are:

- *Fisher discriminant*
- *Likelihood*
- *Boost Decision Tree (BDT)*
- *K nearest neighbor (KNN*
- *Multilayer perceptron (MLP)*

MLP, is a type of feedforward artificial neural network with at least 3 layers of nodes. In order to select the best performing algorithm, one can look at the Receiver Operating Characteristic (ROC) and Efficiency curves. These are shown in Fig. 6.10 and from Fig. 6.10(a). It can be seen that the BDT model outperforms all of the other algorithms, assuring the highest signal efficiency and background rejection. Fig. 6.10(b) shows the cut efficiencies and the optimal cut value for the BDT model. In particular, a BDT cut at -0.1

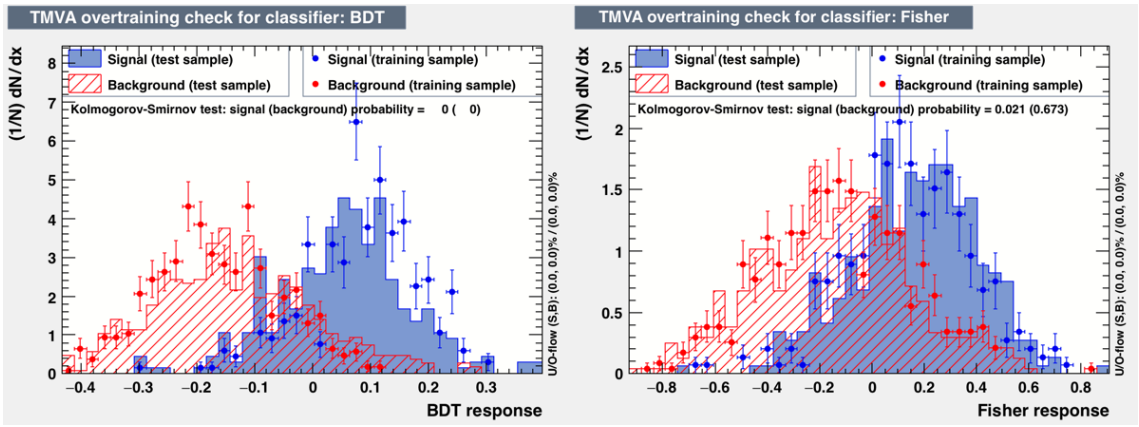
ensures the maximum signal to background ratio with $\frac{S}{\sqrt{S+B}}$ being 25.04 for this value of the cut applied on the BDT output.



(a) (b)

Figure 6.10: Background rejection versus Signal efficiency (ROC) curves for the five classifiers tested (a); cut efficiencies for the best performing classifier (BDT) and optimal cut value (b).

Moving this threshold a bit further to the right would ensure a higher signal purity which would be ideal for the case under investigation. However, this would required more statistics to be left with a significant population after applying the optimised cut. The effect of moving the BDT cut threshold is shown below.



(a) (b)

Figure 6.11: Overtraining check for two classifiers: (a) BDT; (b) Fisher discriminant. The dotted red and blue distributions are singles and multiples in the train set respectively. The solid line distributions of the same colour represent the test set samples. An overlapping of the solid line and dotted distributions indicates that the algorithm is not overtraining.

As mentioned in the introductory paragraphs of this chapter, machine learning algorithms

can often overfit losing their ability to generalise to new and unseen data. Making sure that the algorithms don't overfit is even more crucial in this case given that the samples used in the training process (Monte Carlo data) and those used in the "test" or "application" phase (data) are sampled from two (*a priori*) different processes and the fact that the ML cut from the training model will be used to cut background events on data (unseen from the model), consequently relies on the good match between data and Monte Carlo observed in the ROI. TMVA comes with a tool to check for overfitting (or overtraining) that is summarised in Fig. 6.11 for the two best performing algorithms. The two plots show the response variable for BDT (Fig. 6.11(a)) and Fisher (Fig. 6.11(b)) with the training samples for signal and background (dots) overlapping the test sample distributions (solid lines) ensuring the absence of overtraining. Given these results, the BDT algorithm was chosen as the algorithm to train on Monte Carlo and also be used to cut background events on data.

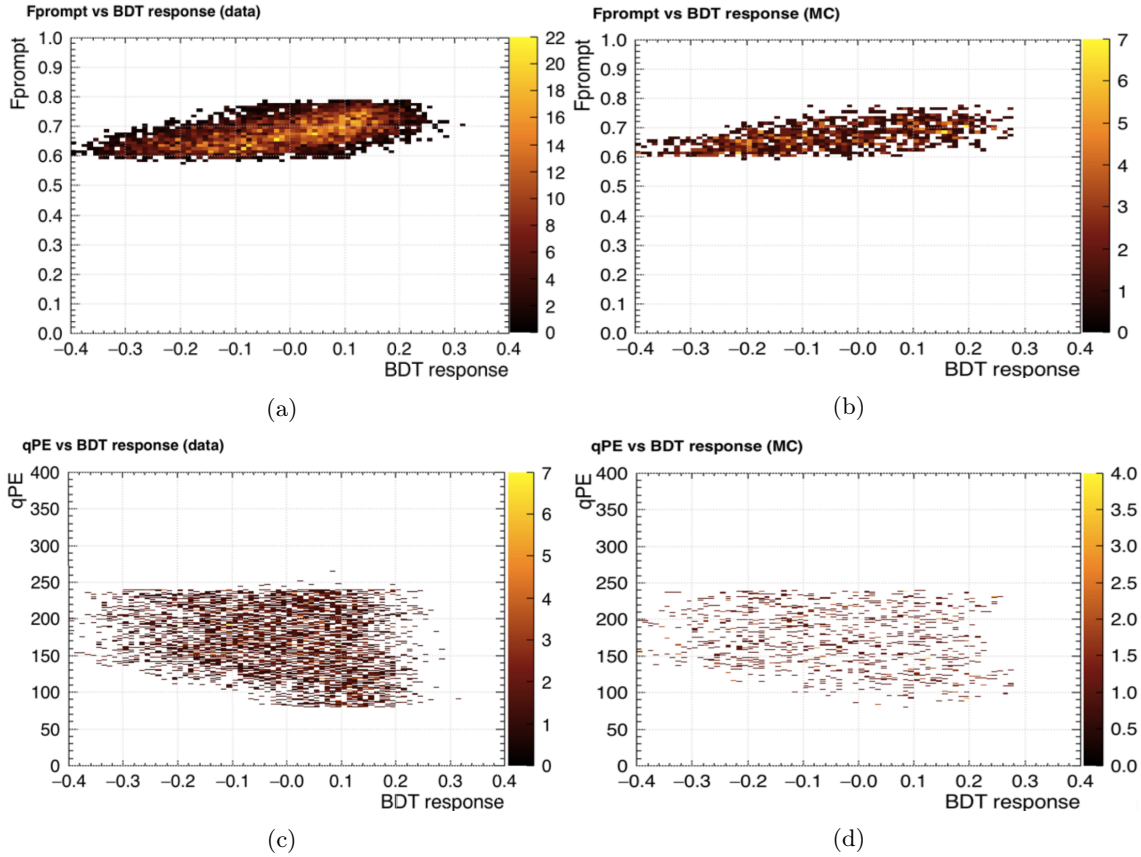


Figure 6.12: F_{prompt} and PE versus BDT response comparison between training AmBe Monte Carlo data and AmBe data. (a) F_{prompt} versus BDT response AmBe data; (b) F_{prompt} versus BDT response AmBe Monte Carlo; (c) PE versus BDT response AmBe data; (d) PE versus BDT response AmBe Monte Carlo.

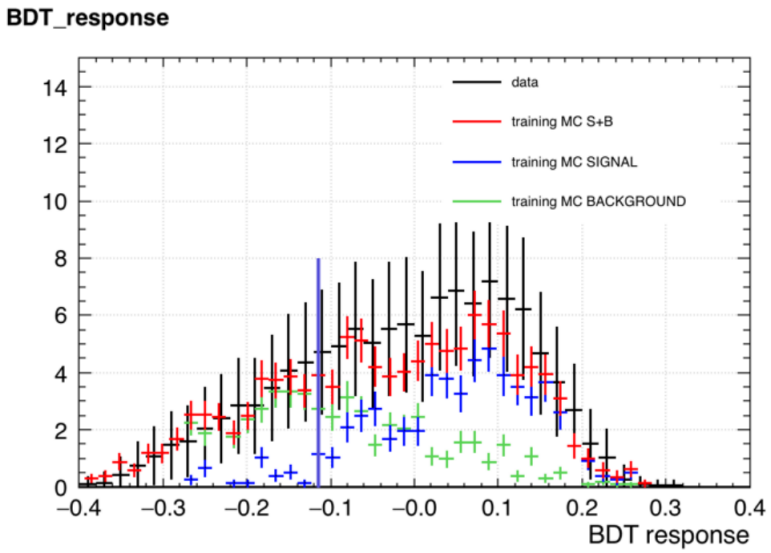
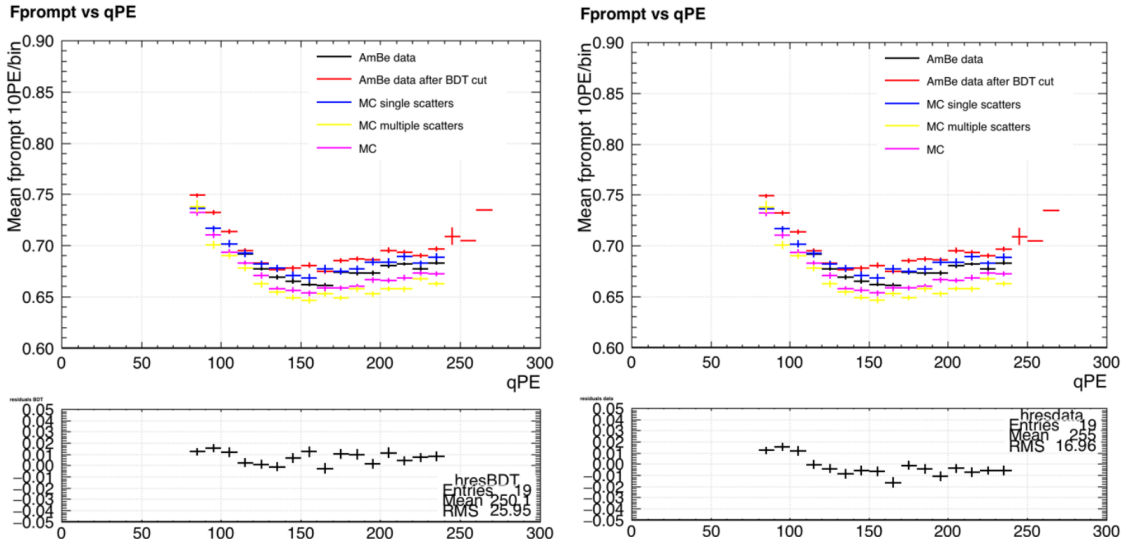


Figure 6.13: BDT response distributions for the AmBe neutron data in black, the mixture of Monte Carlo single and multiple scatters used for the training phase in red, the training Monte Carlo signal (singles) in blue and background (multiples) in light green.

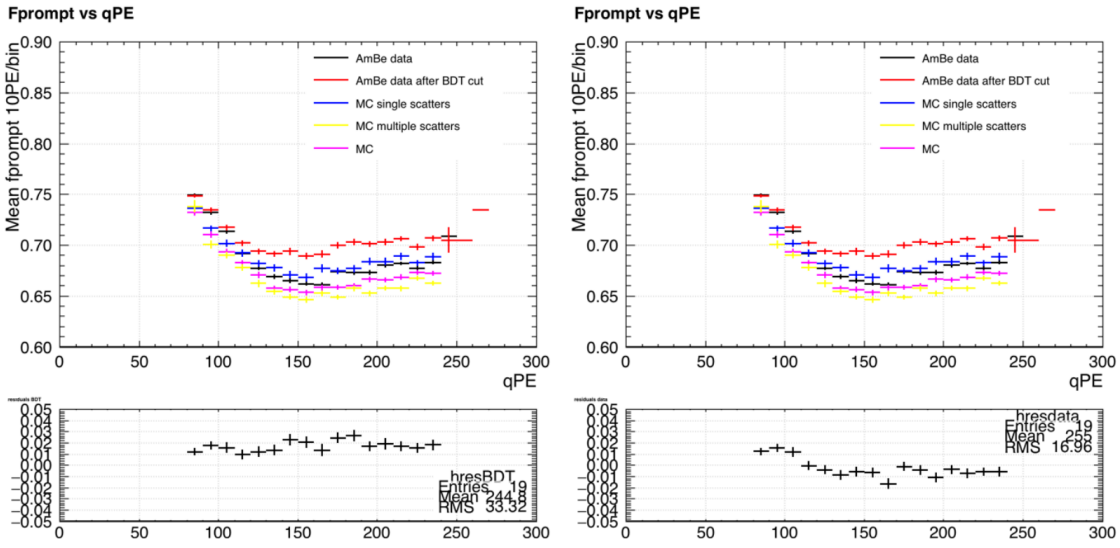
A further check that the two a priori independent samples (data and Monte Carlo) are describing the same phenomenon was performed by looking at the F_{prompt} and qPE distributions as a function of the chosen ML response (BDT). These are shown in Fig. 6.12 where no substantial differences can be seen. In this context, an even more important consistency check is to look at the BDT response variable for both the sample used to train the model and the sample of unseen data. A match between these two distributions would ensure that the algorithm is modelling the same phenomenon and would "authorise" the use of this ML cut developed on simulated data, on AmBe calibration neutron data.
These distributions are summarised in the plot of Fig. 6.13 where the distribution for the BDT response of data (black), training signal and background Monte Carlo (red), training Monte Carlo signal (blue) and training Monte Carlo background (green) are shown. The figure shows how the Monte Carlo distribution matches the data as one would expect, given that the data is a mixed distribution of signal and background events prior to any cut being applied. In what follows, the results of applying two BDT cuts will be shown. In particular, Fig. 6.14 shows the result for the optimal cut (*optimal cut*) that gives a 61% signal efficiency and 39% background efficiency, and Fig. 6.15 for a cut (*efficiency cut*) that increases the signal purity to a 73% signal efficiency and 27% background efficiency.



(a) (b)

Figure 6.14: Mean F_{prompt} versus qPE for AmBe data (black), AmBe data after the BDT *optimal cut* has been applied (red), Monte Carlo single scatters (blue), Monte Carlo multiple scatters (yellow) and total Monte Carlo (magenta). Residuals as function of qPE between mean F_{prompt} for AmBe data after BDT *optimal cut* and AmBe Monte Carlo singles (a) and between mean F_{prompt} for AmBe data and AmBe Monte Carlo singles.

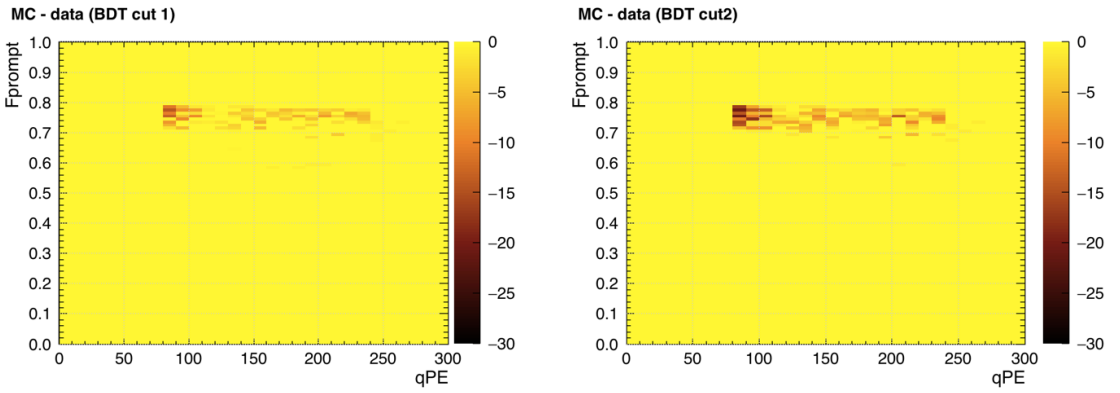
In both figures the top two plots are the same and represent the mean F_{prompt} in 10 PE/bin versus qPE for the real AmBe data (black), the AmBe data after applying for the BDT cut from the ML model (red), the Monte Carlo single scatters distribution (blue), the Monte Carlo multiple scatters (yellow) and the total AmBe Monte Carlo distribution (magenta). In the bottom plots the residuals between the AmBe data after cutting for the BDT cut and the Monte Carlo distribution (red minus blue curves) are shown on the left hand side (a) and the residuals between the AmBe data and the Monte Carlo distribution (black and blue curves) on the right hand side (b). The profile distributions of the residuals of (a) and (b) show an RMS of 0.005 for (a) and 0.008 for (b), improving the ability to isolate single scatters among the mixed sample of singles and multiples in the AmBe neutron data in the ROI.



(a) (b)

Figure 6.15: Mean F_{prompt} versus qPE for AmBe data (black), AmBe data after the BDT *efficiency cut* has been applied (red), Monte Carlo single scatters (blue), Monte Carlo multiple scatters (yellow) and total Monte Carlo (magenta). Residuals as function of qPE between mean F_{prompt} for AmBe data after BDT *efficiency cut* and AmBe Monte Carlo singles (a) and between mean F_{prompt} for AmBe data and AmBe Monte Carlo singles.

Increasing the signal purity by applying the BDT *efficiency cut* does not increase the signal extraction power from data as can be seen from Fig. 6.15(a). From the top plot it is possible to see how, especially from $qPE > 120$, the curve is consistently higher than the blue MC single scatter points and the overall behaviour in the entire ROI energy region is summarised by an RMS of the profile residual distribution of 0.007 versus 0.005 for the *optimal cut* and 0.008 for the no cut scenario.



(a) (b)

Figure 6.16: F_{prompt} versus qPE AmBe Monte Carlo minus AmBe data after BDT *optimal cut* (a) and after BDT *efficiency cut*.

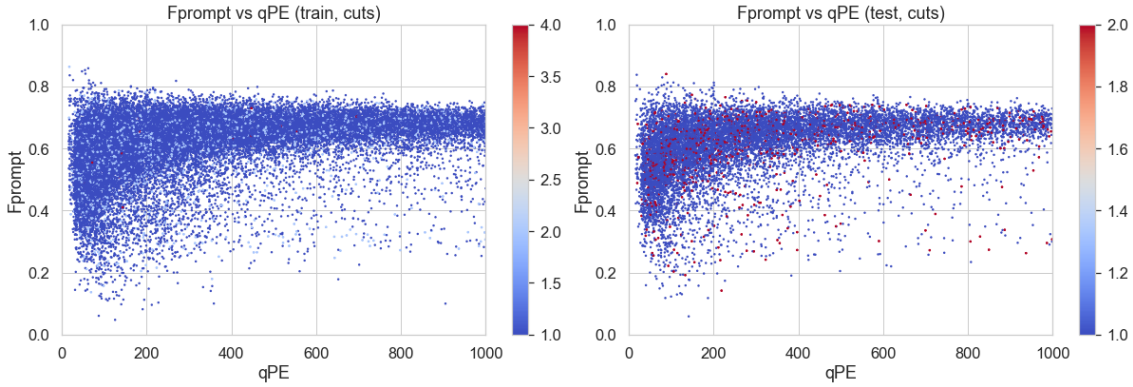
The reason for this is to be found in a deficit of events in Monte Carlo with respect to data as can be seen from Fig. 6.16. Fig. 6.16(a) plot shows the subtraction of the AmBe data after BDT *optimal cut* from the AmBe Monte Carlo sample. The same is shown in Fig. 6.16(b) but for the BDT *efficiency cut*with the colour scale representing the difference in events between the two samples. A clear deficit of events in Monte Carlo can be seen when increasing the signal purity with BDT *efficiency cut*, which leads to a higher mean F_{prompt} as observed in Fig. 6.15(a).

6.5 Scikit-learn and new AmBe Monte Carlo

A new set of AmBe Monte Carlo that includes the coincident 4.4 MeV gammas in stage 1, an increased radial acceptance for neutrons to be generated in stage 2 and an updated version of RAT and its analysis processors was generated. In total 7.5 billion stage 1 neutrons were generated. The scikit-learn studies that will be presented in the following sections, make use of 120.65 livetime hours of thenew Monte Carlo (equivalent to 2 billion stage 1 neutrons) as well as AmBe neutron data collected in March and April 2018 for a total livetime of 132.11 hours. Both MC and dataand reprocessed with the same processors and with the same updated version of RAT.

6.6 Data preparation

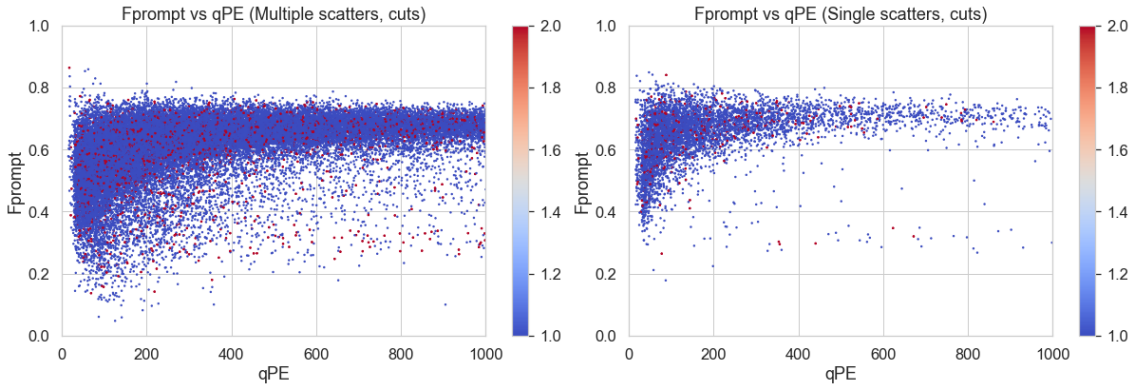
As mentioned in Section 6.2, a dedicated library allows for ROOT ntuples to be read and transformed into numpy arrays which is the format in which scikit-learn accepts data. Given a dataset $(\mathbf{x}, Y) = (x_1, x_2, ..., x_k, Y)$, we will refer from now and on to $\mathbf{x} = x_1, x_2, ..., x_k$ as the k *features* of the dataset and to Y as the *target* variable. Stage2 AmBe MC data are used to train a model which is able to discriminate between single and multiple scatters. The size of the dataset used to build a model is of 24112 entries and made of 157 features. The number of predictive features is then consistently larger than the number of predictive variables used within TMVA. This dataset is split at the very beginning into a training and testing set with a 70-30 ratio. The F_{prompt} versus qPE distribution for training set (a) and the testing set (b) are shown in Fig. 6.17.



(a) (b)

Figure 6.17: F_{prompt} versus qPE for train set (a) and test set (b).

The output Y, whether the event is a single or a multiple scatter (equivalent of TMVA's signal and background), is known for both sets. Treating the simulated test set as it were data allows a baseline for the performance of the algorithm with future data to be compared against. In particular, having shown that the model does not overfit and that there is a good match between real and simulated data, will ensure good performance of the model trained on simulated data with data. Fig. 6.18 show the F_{prompt} versus qPE distribution for multiple scatters (a) and single scatters (b) after standard cuts to remove bad events have been applied.

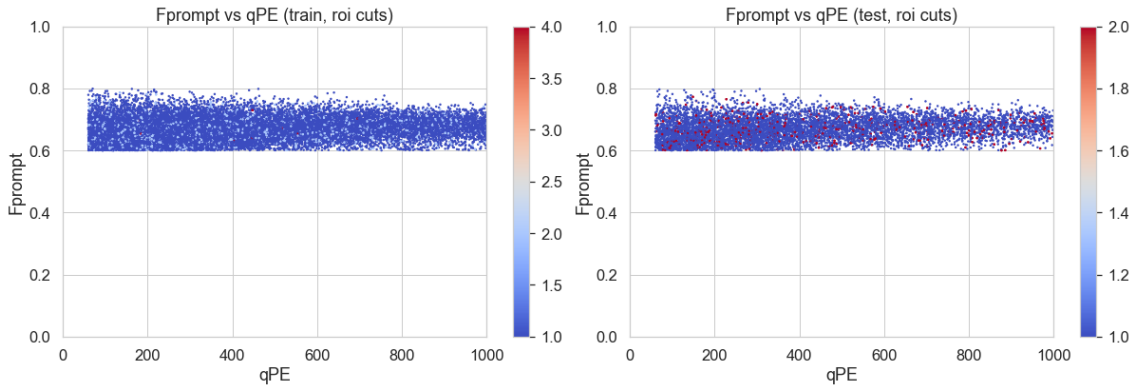


(a) (b)

Figure 6.18: F_{prompt} versus qPE for multiple scatters (a) and single scatters (b).

However, given the nature of the simulated data that does not include ^{39}Ar and underestimates Cherenkov light as discussed in Section 5.5, a decision was made to restrict the model building to a subset of the F_{prompt}-qPE space where real and simulated data are in better agreement< the region of interest for the dark matter search. In particular the

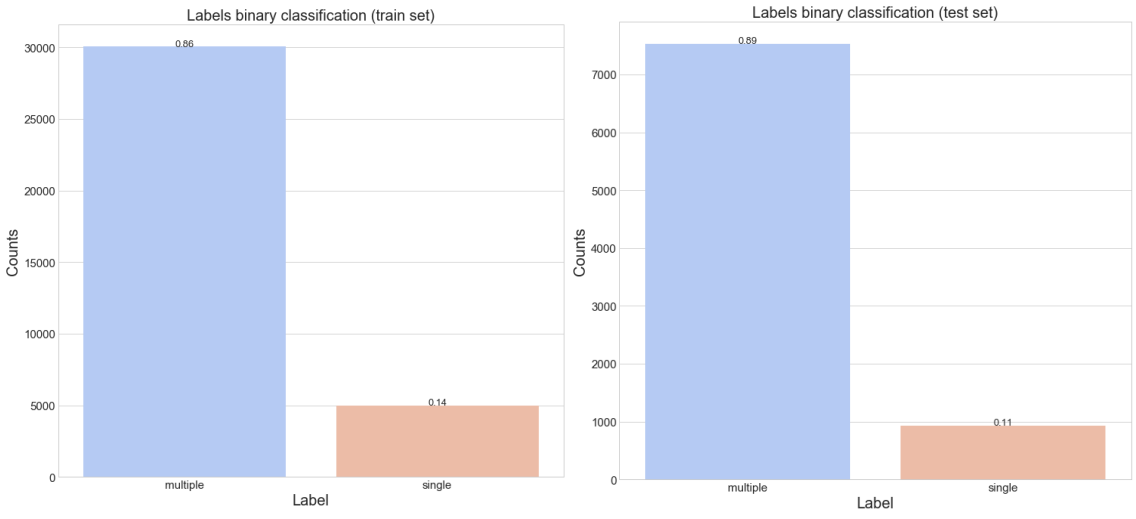
the box region $0.6 < \mathrm{F_{prompt}} < 0.8$ and $60 < \mathrm{qPE} < 1000$ was chosen for this purpose as shown in Fig. 6.19 for the train (a) and test (b) set.



(a) (b)

Figure 6.19: $\mathrm{F_{prompt}}$ versus qPE for train (a) and test (b) after cuts in the $\mathrm{F_{prompt}}$ region between 0.6 and 0.8 and qPE between 60 and 1000.

Fig. 6.20 shows the target Y distribution among the training and testing set. It is clear that multiple scatters represent the majority class with roughly 80% of the events being multiple scatters in both sets.



(a) (b)

Figure 6.20: Class distributions among the train (a) and test (b) set. The majority class is multiple scatters in both train and test sets.

Imbalances between the two classes in ML can be a problem both in the model building process and when evaluating it. A ML learning algorithm trained on an imbalanced dataset tends to learn very well the characteristics of the majority class of which it has plenty of

examples to learn from, and not of the minority class, resulting in a bad classifier overall when it comes to predicting the output class, as the majority class will be the one most often predicted. Furthermore, as previously stated, evaluation metrics like accuracy often used in binary classification problems can be very unreliable because even by predicting the majority class in all instances one can get a pretty high accuracy without having been able in fact to predict a single time when an instance of the minority class occurred. For this reason it is better in these cases to use other metrics for evaluating the model performance. A comparison of some of the features is showed in Fig. 6.21.

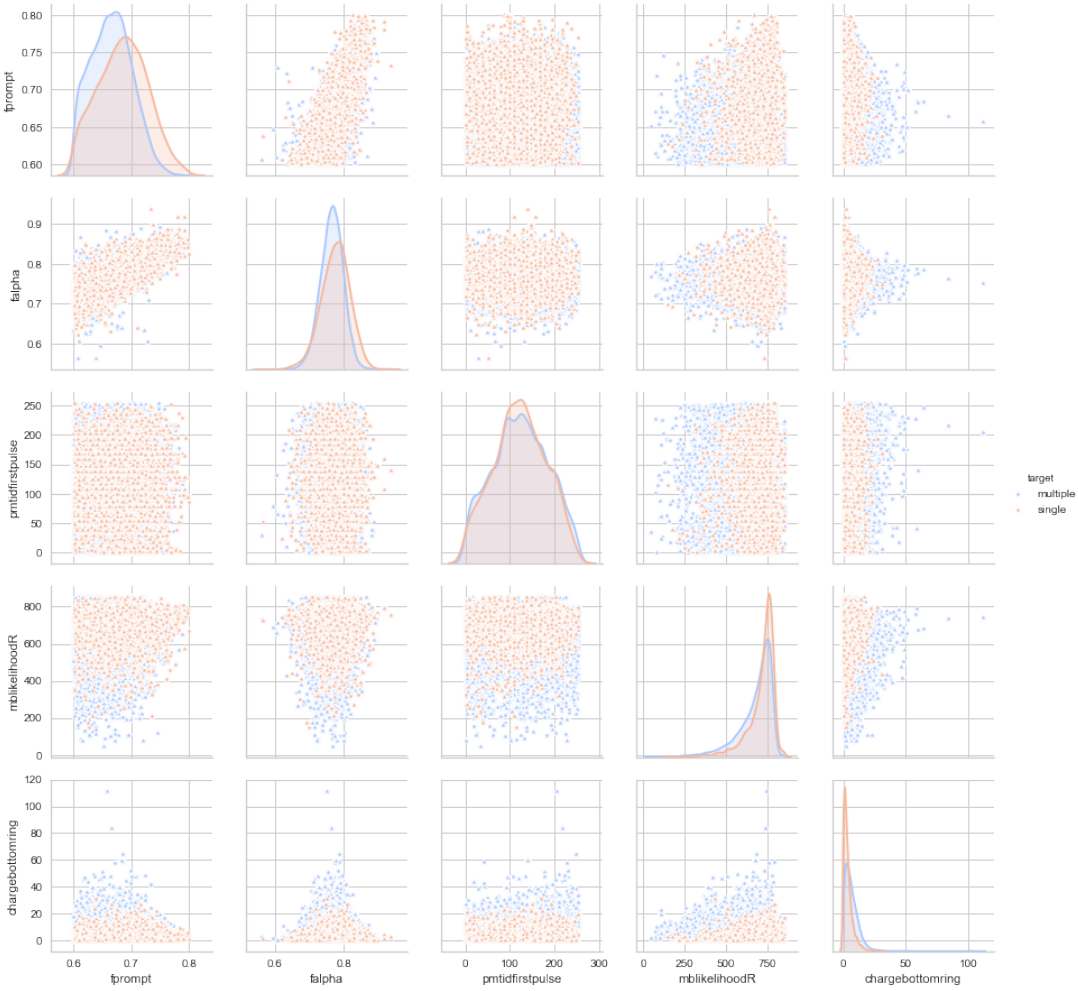


Figure 6.21: Distributions of some features for the two classes, multiple scatters in light blue, single scatters in light orange.

Feature selection

Even though the availability of a large number of predictive variables can seem ideal, high dimensionality in ML can also be problematic because points in high-dimensional space

are highly sparse and this can cause many ML algorithms to underperform. This problem is known as the *the curse of dimensionality* and can be avoided either by increasing the size of the dataset (not always possible) or through feature selection. Feature selection refers to the process of reducing the number of input variables that are fed to the ML algorithm. This will also reduce the chance to overfit. There are many techniques in use to reduce the space dimensionality, the most common being principal component analysis (PCA) or linear discriminant analysis (LDA). A solid domain knowledge is also important when building a ML model. Particular features can in fact be discarded just because the model builder knows that they are not relevant in describing the phenomenon under investigation. Feature selection in this study followed a mixture of domain knowledge and statistical observations. Some of the initial 157 features were discarded because they were not carrying any information that would be useful to distinguish between the two classes (for example features like event ID, trigger status related features, constant features etc.), and some by looking at the correlation between variables as shown in Fig. 6.22. The colour scale in this plot represents the correlation between features from 0 to 1, with blue indicating small (or zero) correlation and red high (or maximum) correlation between features. Highly correlated features basically carry a duplication of information which will not improve the model. One can then set a threshold above which only one of the features in the highly correlated pair is kept and the other is dropped.

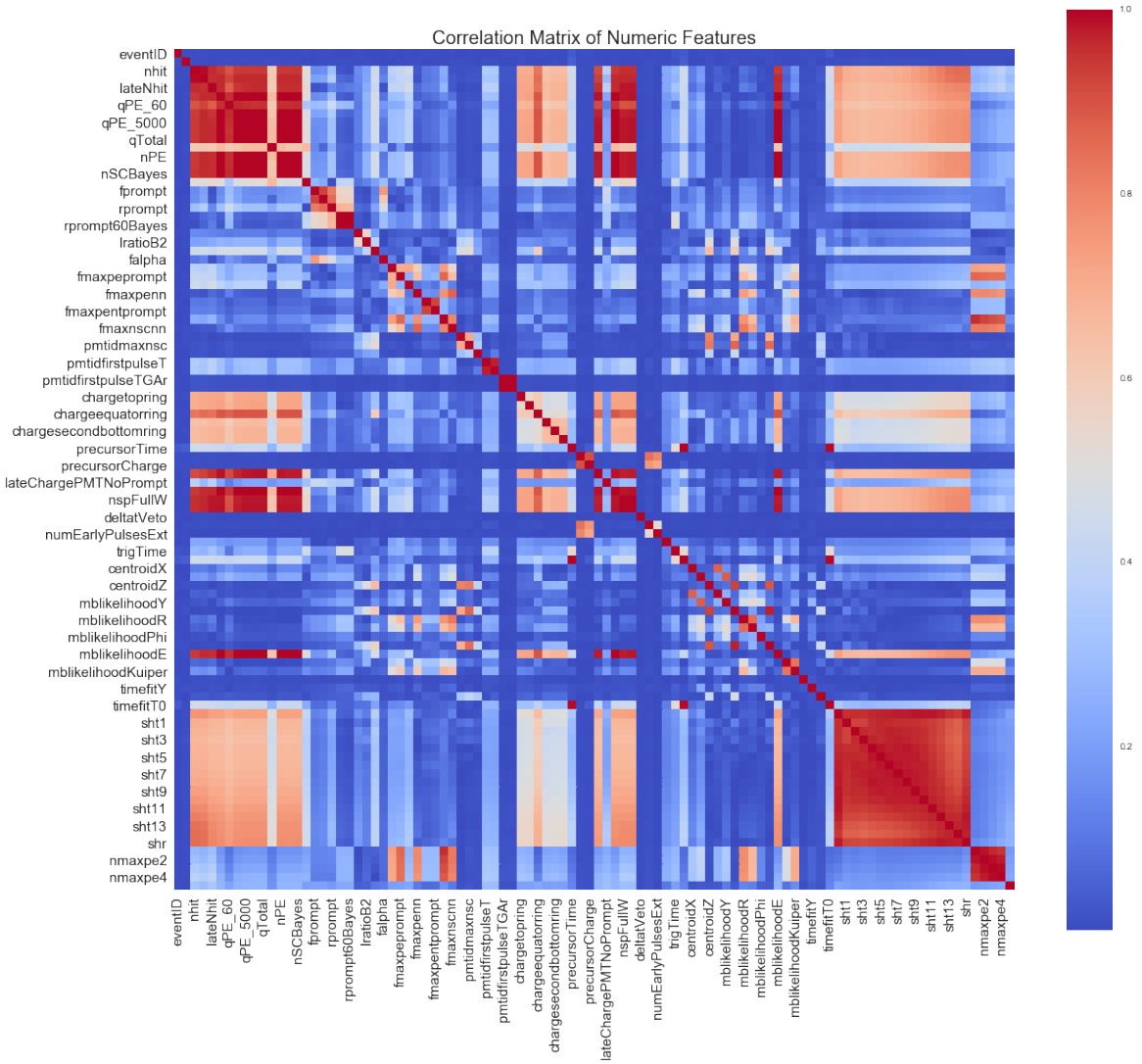


Figure 6.22: Correlation matrix between all the numeric features. Dark blue indicates little or no correlation, dark red high or maximum correlation.

This threshold was set to 0.95 in this study and the result of dropping the highly correlated features is shown in the correlation plot in Fig. 6.23.

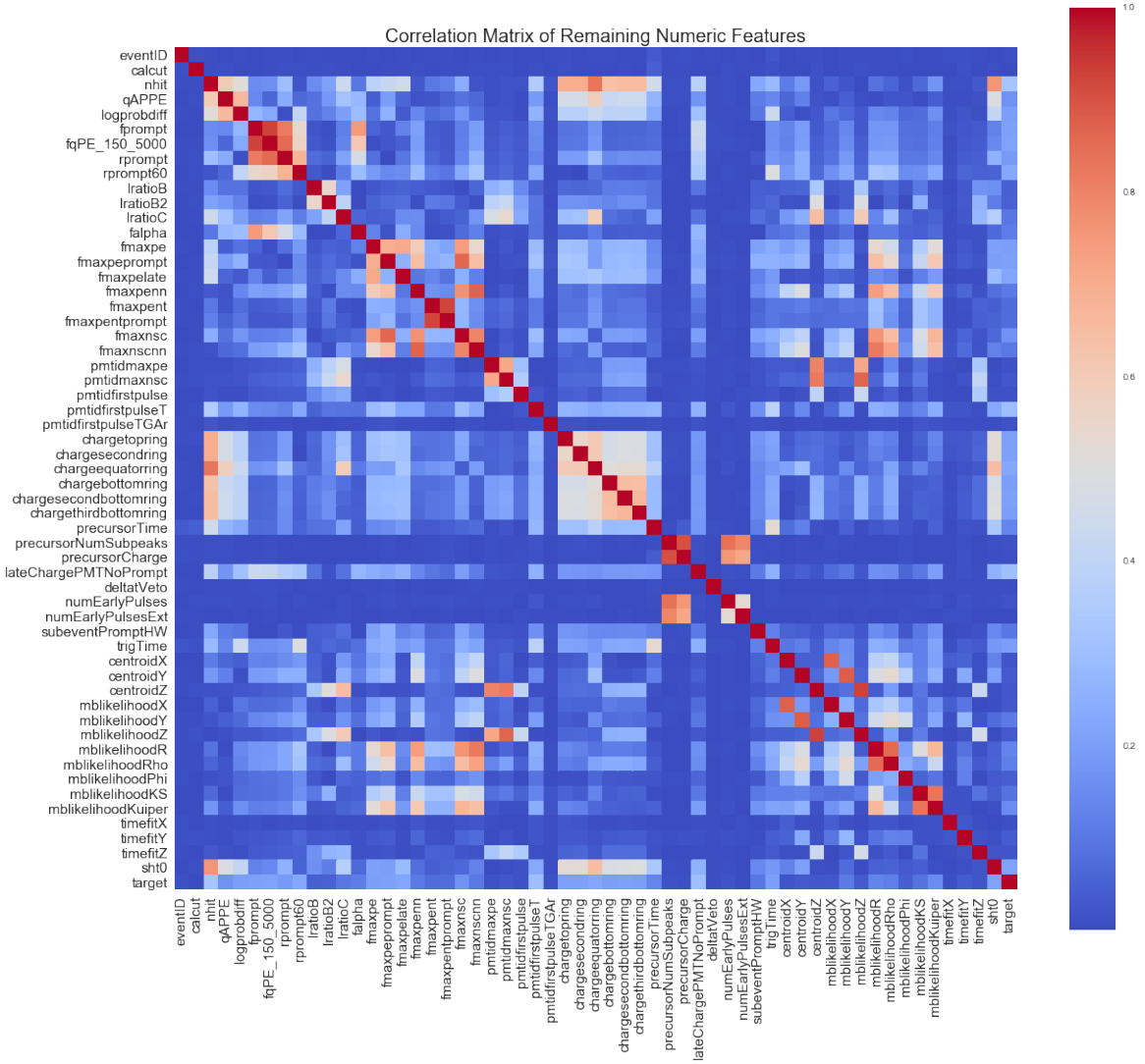


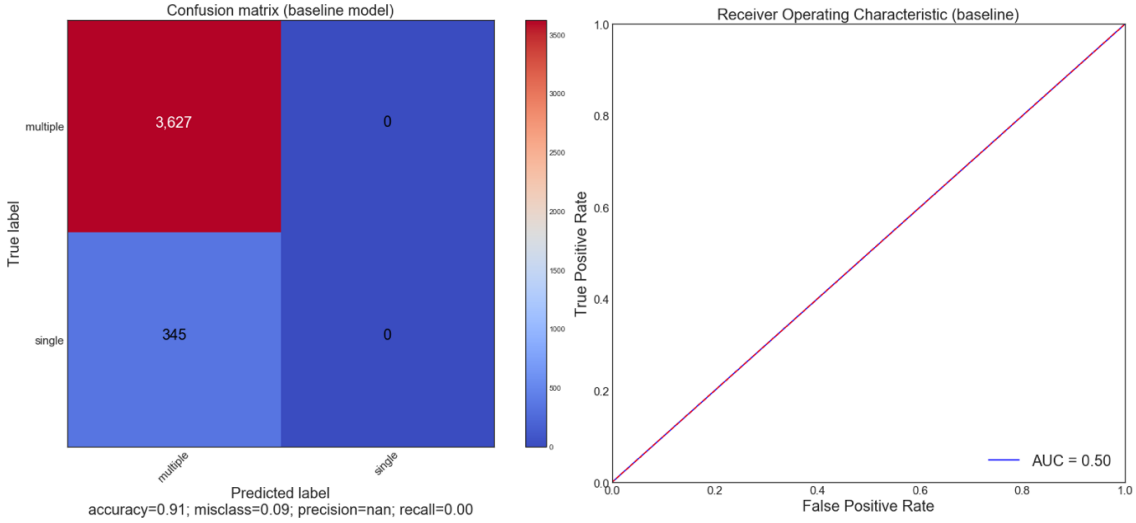
Figure 6.23: Correlation matrix between all the numeric features after having removed highly correlated features (correlation greater than 95%).

The dimension of the dataset after feature selection is 51. It should be noted that both training and test sets are treated during the data preparation phase such that they have the same number of features and are transformed similarly such that if any transformation is performed in the training set it will be reflected in the testing set.

Baseline model and evaluation metrics choice

In the metrics section, the confusion matrix and the most common metrics used to evaluate the goodness of a classification algorithm were presented. It is good practise, when building a model, to have a baseline model to compare to. Deviations from it will reveal whether or not the final model is an improvement to the baseline model. Fig. 6.26 shows the confusion

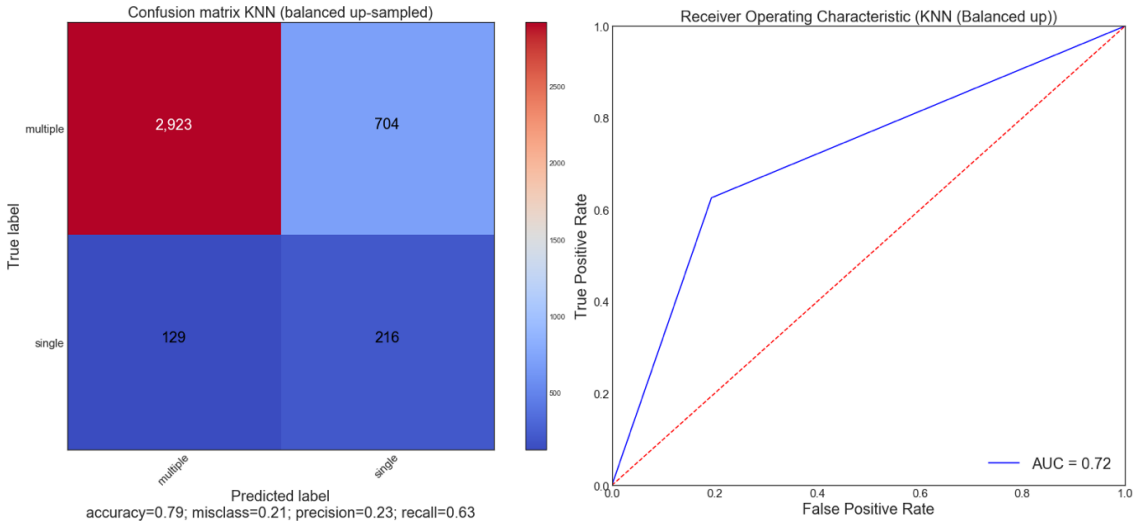
matrix (a) and the ROC curve for the most simple model one can come up with, that is a model that always predicts the majority class which is multiple scatters in this case.



(a) (b)

Figure 6.24: Confusion matrix (a) and Receiver Operating Characteristic curve for the baseline classification model.

Given the nature of the dataset, which is strongly imbalanced, it is not surprising that the accuracy score is very high even with a model that has not even seen the training data. The ROC curve gives a better representation of the goodness of this model. It lies on top the the curve that represents the good how model would do just guessing, which is what this model is doing as also confirmed by the AUC score of 0.5. No precision can be calculated as the value of both TP and FP are zero and the recall is zero as TP is zero.



(a) (b)

Figure 6.25: Confusion matrix (a) and Receiver Operating Characteristic curve for the KNN classifier.

When tackling a binary classification problem, KNN is often chosen as a baseline model for its simplicity. The results of training a simple KNN algorithm to a balanced training set are shown in Fig. 6.25. The training set was resampled such that the number of examples of the minority class (singles) would match the number of majority class examples (multiples). By looking at the confusion matrix, it can be seen that the accuracy score drops to 79% for KNN, precision and recall are now different from zero and that even though the model is not doing a great job it is performing better than a random guess would perform as shown by the ROC curve plot in Fig. 6.25(b). A more complex algorithm should do better than KNN. Improved models that perform better than KNN are described in the next section. It is good practice to choose metrics for evaluating the model before starting the building process and sticking with them. The reason why accuracy would not be a good metric in this case has already been discussed. Moreover, the most suitable choice here is suggested also by the ultimate goal of the study: to obtain as clean a sample of single scatters as possible. For this reason, the metric one would want to consider is precision. Precision, defined as the ratio of true positive over the sum of true positives and false positives ($\frac{TP}{TP+FP}$), gives an indication of how pure the prediction is, with 100% precision indicating that all the events predicted as single scatters are actually single scatters and no multiples were predicted, and 0% indicating that either all the events predicted as singles are actually multiples or that nothing was predicted to belong to the positive class (singles

in this instance).

Model selection

KNN did a very poor job with respect to the precision score. Among the many classification algorithms that exist, five were chosen and their performances tested on the training set using a five-fold cross validation. It should be noted that for all five algorithms scikit-learn provides a method that handles imbalanced data (class_weight), so the cross validation was performed on the original imbalanced training set with the class_weight method set to *balanced*. Table 6.2 reports the mean precision scores and the standard deviations for the five algorithms order in descending order with respect to precision.

Algorithm	**Mean precision score**	**Standard deviation**
Random Forest	0.601182	0.019643
Extra Trees	0.595211	0.039320
CART	0.431714	0.024209
Bagging Classifier	0.430122	0.020417
Logistic Regression	0.352950	0.017490

Table 6.2: 5-fold cross validated mean precision score for the five algorithm tested in order of performance (ranked best to worst).

The same information is shown with the box plot in Fig. 6.26. All these algorithms represent a significant improvement compared to the KNN algorithm, with Extra Trees and Random Forest outperforming simple methods such as Logistic Regression and CART.

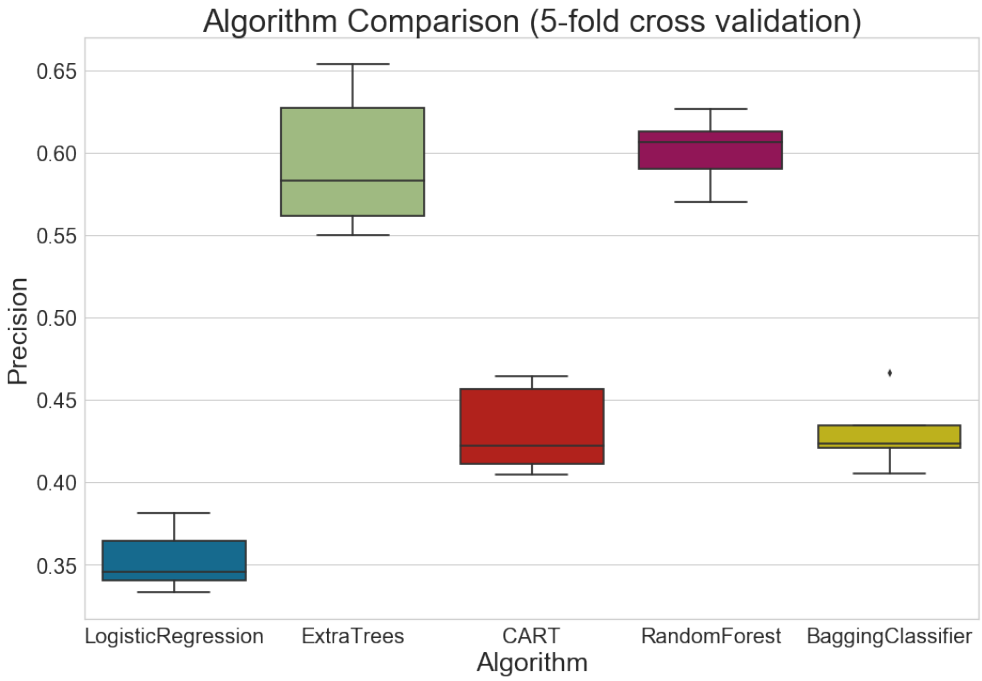


Figure 6.26: Machine learning algorithms performance comparison. The 5-fold cross validated precision score for the five algorithms tried are represented through box plots. Random Forest and Extra Trees overperform a classic decision tree (CART) algorithm, a logistic regression and a bagging classifier.

None of the hyperparameters of these algorithms were tuned at this stage, with the scikit-learn default values being used for what is shown in this section. Tuning of these parameters can improve performance even further or reveal signs of overfitting. This will be discussed in the following sections. Given the precision score results a choice was made to use and optimise Random Forest as the ML algorithm for this study.

RFE

RFE stands for recursive feature elimination and is a technique to select features by recursively considering a smaller sets of features, that is the algorithm is trained first by using all the features and the importance of each feature is evaluated. The least important feature is then pruned and the estimator re-trained on this reduced set of features. This procedure is repeated recursively. Fig. 6.27 show the five-fold cross validated precision score for Random Forest as a function of the number of features selected. The optimal number of features to use in this case turned out to be 25 but given the randomness that is intrinsic to Random Forest this value can change if the cross validation is repeated. What it does not change is the trend of this curve that is telling us that a plateau is reached once

the model is trained with more than 10 features.

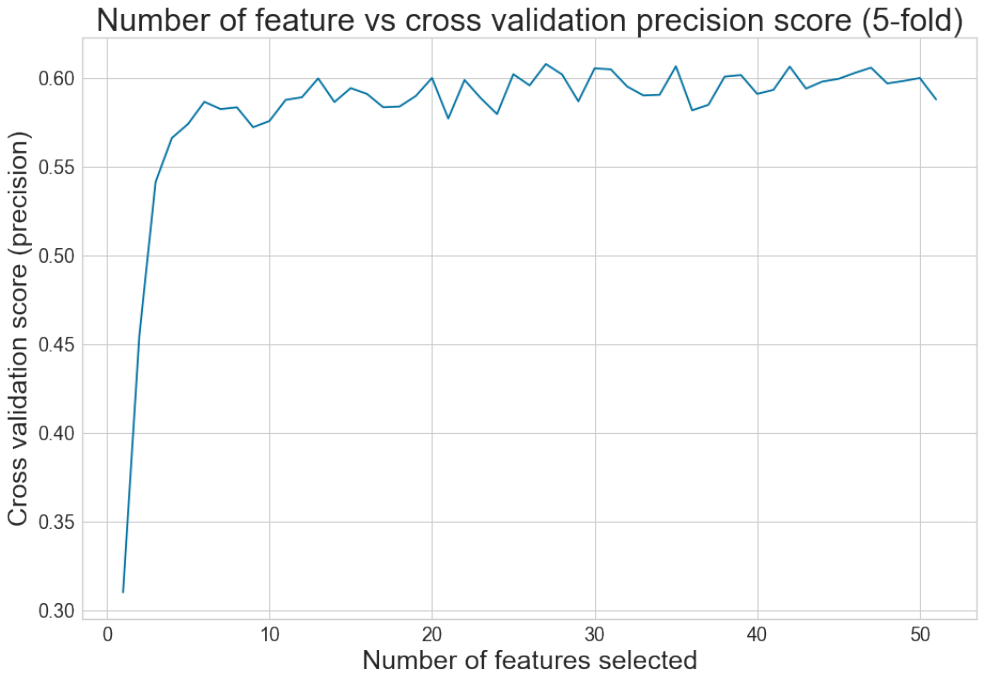


Figure 6.27: 5-fold cross validated precision score as a function of the number of features selected to train the model. Once a plateau is reached, adding more features will not improve the model performances.

The structure of Fig. 6.27 can vary in an algorithm such as Random Forest when regenerated, however the overall shape will largely be the same. As such, the number of features should not be reduced below ten (the edge of the plateau in Fig. 6.27) to keep a good precision score. Fig. 6.28 shows how each of the 51 features ranked in the RFE of Fig. 6.27 - features ranked 1 are desired by RTE and those ranked above 1 can be discarded.

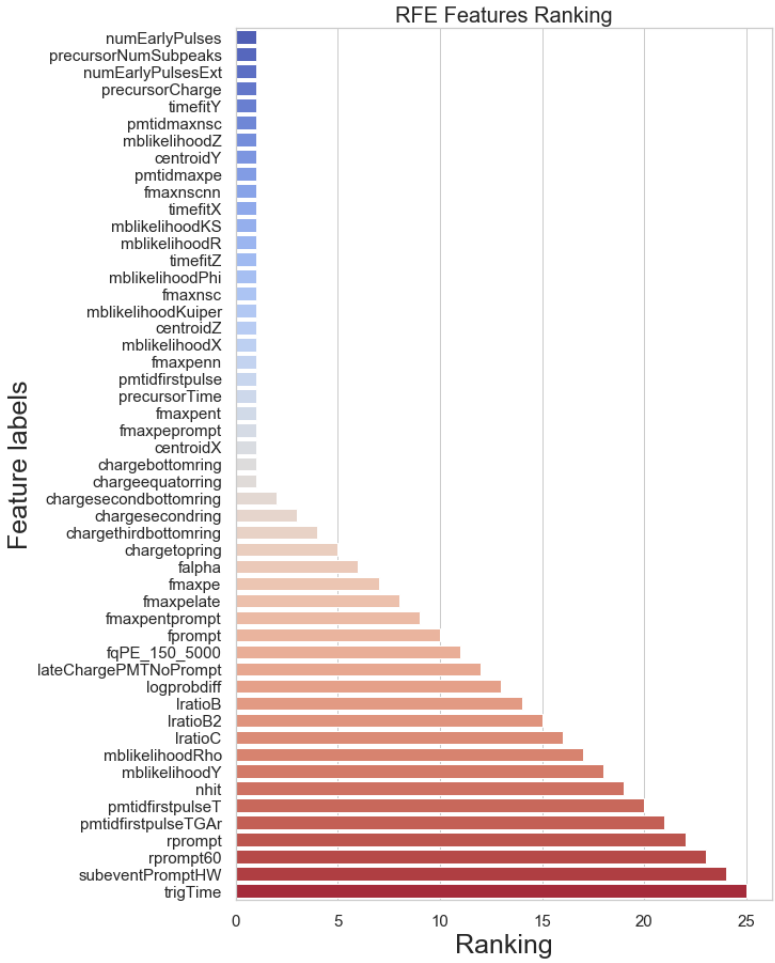


Figure 6.28: Feature labels order by their importance for the model as ranked by the recursive feature elimination (RFE) technique.

A decision was made to leave all the features as they are at this point and deal with further feature reduction when building the model that will select signal in data.

Tuning hyperparameters

The cross validated average precision score of roughly 60% obtained by the untuned Random Forest drove the decision to choose this algorithm among the others for this study. However, although it seems a huge improvement with respect to KNN or Logistic Regression, this may come with a cost which is overfitting. Overfitting is always a risk when using decision trees that are grown very deep. Random Forest, like many other algorithms, has many hyperparameters that can be optimised. Hyperparameters are parameters that are set before the learning process starts as opposed to other parameters that are learned during training by the ML algorithm. Optimisation of these hyperparameters can make

an important difference in the model performance or in avoiding overfitting. Scikit-learn comes with default values for algorithm hyperparameters that can be changed once the algorithm object is instantiated. The Random Forest classifier hyperparameters are in Table 6.3.

Table 6.3: Random Forest classifier hyperparameters.

Hyperparameter	**Value**
bootstrap	True
class_weight	None
criterion	gini
max_depth	None
max_features	auto
max_leaf_nodes	None
min_impurity_decrease	0
min_impurity_split	None
min_samples_leaf	1
min_samples_split	2
min_weight_fraction_leaf	0
n_estimators	10
n_jobs	1
oob_score	False
random_state	None
verbose	0
warm_start	False

An in depth description of all the parameters in Table 6.3 goes beyond the scope of this section but a description of a few key hyperparameters and the consequences of changing their values follows. Some such as `bootstrap` or `criterion` are self explanatory, determining whether bootstrap samples will be used when building trees or which criterion to use to measure the quality of a split respectively. The majority of the remaining hyperparameters describe characteristics specific to the tree structures and play a big role in the ability of the model to perform well and not overfit. For example the `n_estimators` hyperparameters, which determines how many decision trees will be used, generally decreases the chances of overfitting with increasing value. Setting the maximum depth of the tree, `max_depth`, to a finite and small value different from the default value of `None`, will also reduce the chance to overfit as the trees will not be developed very deeply. The fine tuning of the model hyperparameters can be a tricky and computationally very demanding task given the infinite combination of possible values they can take on. The traditional way of doing this tuning in ML is through what is called êxtitgrid search. A grid search

searches for the best set of hyperparameters according to some specified metrics, through a subset of the hyperparameter space of a learning algorithm that has been specified manually. GridSearchCV is a method implemented by scikit-leran that performs a grid search with the performance metric to guide the search being evaluated by cross-validation. The fine tuning of the Random Forest model used for this study was performed through a grid search cross validated on the training set. GridSearchCV can require a lot of CPU time depending on the dataset/grid size and the values of the hyperparameters to search into. The best set that GridSearchCV returns is then not necessarely the best set possible in the hyperparameter space but the best among the values specified in the grid. For the Random Forest described here, the specified grid was as summarised in Table 6.4.

Table 6.4: GridSearchCV hyperparameters.

Hyperparameter	**Value**
n_estimators	[500, 1000, 3000]
max_depth	[3,10, None]
max_features	[1, 5]
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 3, 10]
bootstrap	[True, False]
criterion	[gini, entropy]

where `max_features` is the number of features that the tree should consider when looking for the best split, `min_samples_split` is the minimum number of samples required to split an internal node and `min_samples_leaf` the minimum number of samples required at a leaf node. A three-fold cross validation with precision score being the metric to maximise gave the best hyperparameters on the training set reported on Table 6.5.

Table 6.5: Best hyperparameters from GridSearchCV.

Hyperparameter	**Value**
bootstrap	False
criterion	entropy
max_depth	None
max_features	10
min_samples_leaf	1
min_samples_split	5
n_estimators	1000

With this set of hyperparameters the average precision score was 62%, a 2% increase with respect to the *default* model, which is a good improvement. However, no checks have been

done yet to see whether these models are overfitting. A way to check whether a classifier is overfitting (or overtraining) borrowed from ROOT's TMVA routine is to superimpose the classifier output for the two classes for the training and the held-out test sets. If the algorithm is not overfitting the two distributions should sit on top of each other or be very similar. A closer look at the Random Forest model with the default hyperparameters shows that it does overfit the training set as shown in Fig. 6.30 where the two distributions are completely in disagreement. It should be noted that in the model that has been defined as *default* so far, the class weight hyperparameter is set to balanced to take into account the imbalanced dataset (this is true for all the classifiers tested). The classification reported for the *default* Random Forest model can be seen in Fig. 6.29. It shows a summary of the model performance for the two classes with respect to precision, recall and f1-score, which is the harmonic average of the two. The metric "support" simply specifies of many examples for each class were present in the dataset.

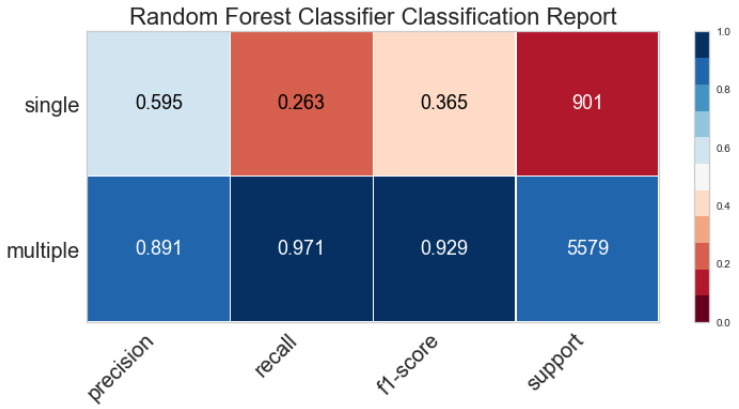


Figure 6.29: Classification report for the Random Forest classifier. The precision, recall, f1-score and support are reported for single and multiple scatters as classified by the Random Forest algorithm.

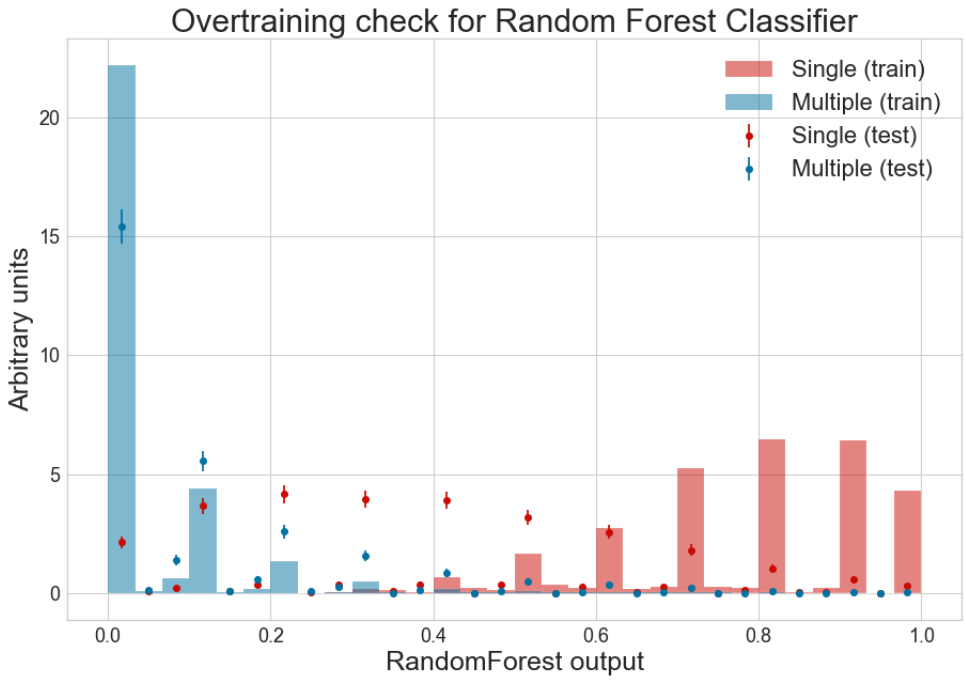


Figure 6.30: Single and multiple scatter distributions for train and test sets. Solid red and blue line distributions are singles and multiples in the train set respectively. Dotted distributions of the same colour represent the test set samples. The differences between the solid line and dotted distributions are indicating that the Random Forest algorithm is overtraining being its output for the test set very different from the train one.

The *default* model hasn't the ability to generalise the unseen data and its 60% precision, as much as a promising improvement with respect to the baseline model, is not really an accurate description of how well this model would perform on new data. The same problem was noticed with the model tuned with the three-fold GridSearchCV even though the plots are not reported here. Both models were overfitting. There are many ways to avoid the Random Forest from overfitting. As mentioned earlier, not allowing the trees to grow very deeply is one of them. The hyperparameter that can control the tree's growth is `max_depth`. A few values for the `max_depth` different than `None` (i.e. the tree is left to grow) were tested and the overfitting check was performed for these values of `max_depth` whilst leaving the other hyperparameters (those obtained through GridSearchCV) unchanged. The value of `max_depth` that gave the highest precision score without overfitting was `max_depth = 7`. The precision score on the held-out test set is lower than what both the previous two models would reach (42%, see classification report in Fig. 6.31), but the overtraining check in Fig.6.33 shows no evidence of overfitting with these set of hyperparameters.

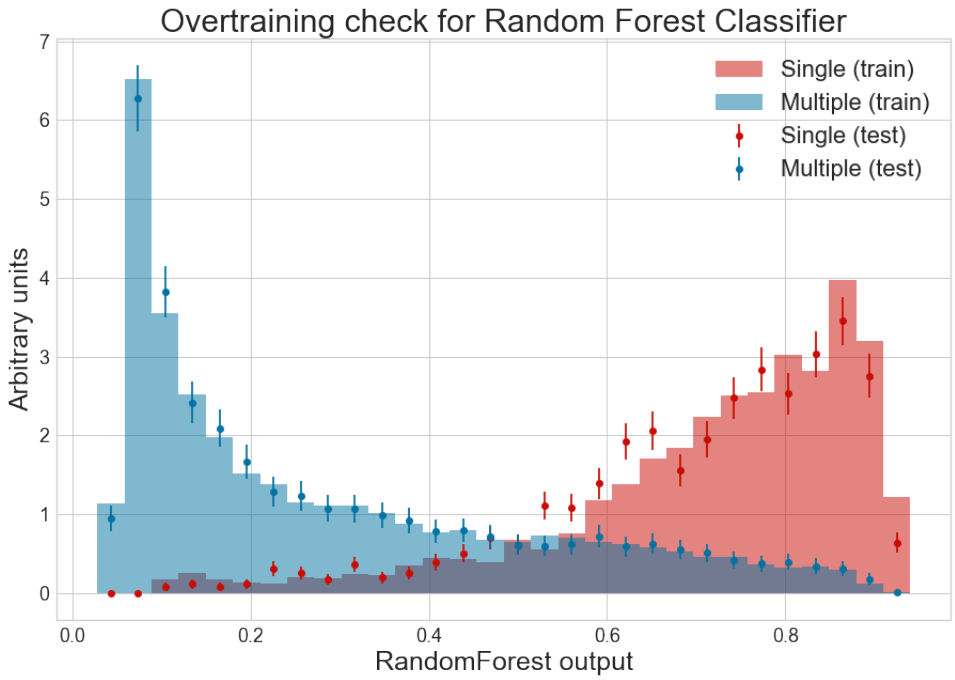


Figure 6.32: Single and multiple scatter distributions for train and test sets. Solid red and blue line distributions are singles and multiples in the train set respectively. Dotted distributions of the same colour represent the test set samples. An overlapping of the solid line and dotted distributions is an indication that the algorithm is not overtraining.

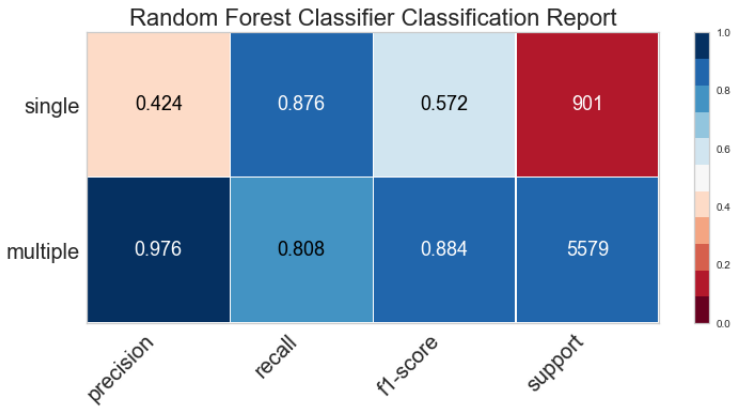


Figure 6.31: Classification report for the Random Forest classifier. The precision, recall, f1-score and support are reported for single and multiple scatters as classified by the Random Forest algorithm.

This model performances will be further investigated in the next section still using the held-out test set of simulated data as if they were the unseen data that the model is getting built for. In particular, the possibility to move the algorithm decision threshold (how the algorithm chooses the class to assign elements to) and how this can improve performances will be investigated.

Model Evaluation

The Random Forest model has been optimised to reach the highest precision score possible and tuned such that it does not overfit. The precision score of 42% presented in Fig. 6.31 refers to what the algorithm would do if the decision threshold is set to 0.5. Every ML algorithm that deals with binary classification requires a way to map some continuous prediction output to one of the two classes. In the case of Random Forest, scikit-learn comes with a method called `predict_proba` that returns the probability that an element belongs to one class or another. By default this threshold is usually set to 0.5 which means that an element will be classified as belonging to the positive class if `predict_proba` is greater than 0.5 and to the negative class otherwise. However, one can vary this threshold according to whether reaching a higher precision or a higher recall is more important for the problem being considered.

This concept is illustrated in Fig. 6.33 and Fig. 6.34 where Precision and Recall are shown as functions of the decision threshold.

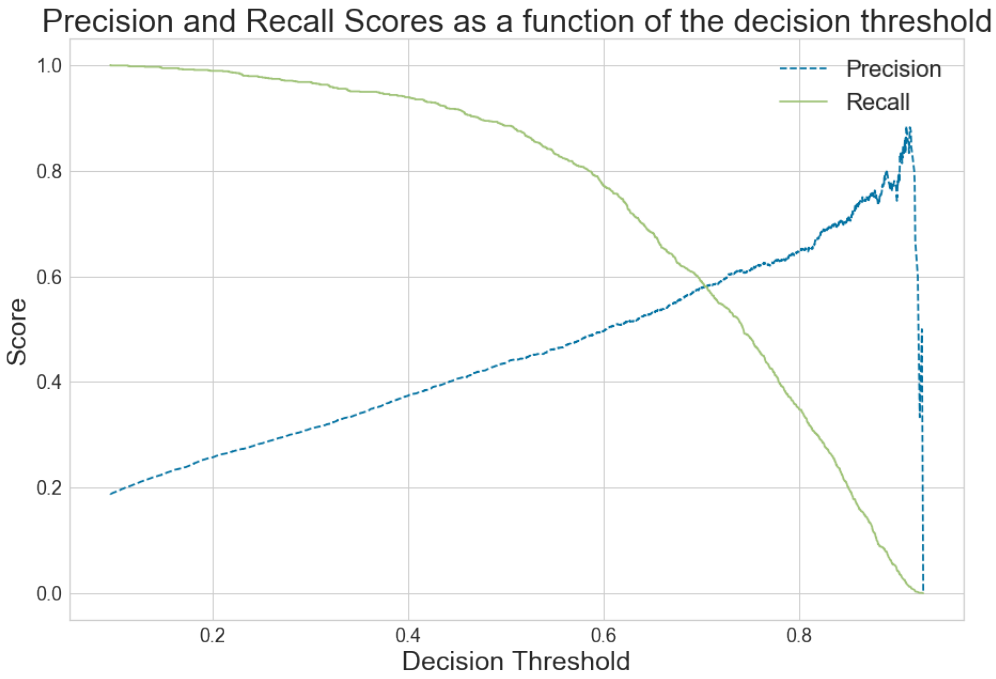


Figure 6.33: Precision and Recall scores as a function of the Random Forest decision threshold.

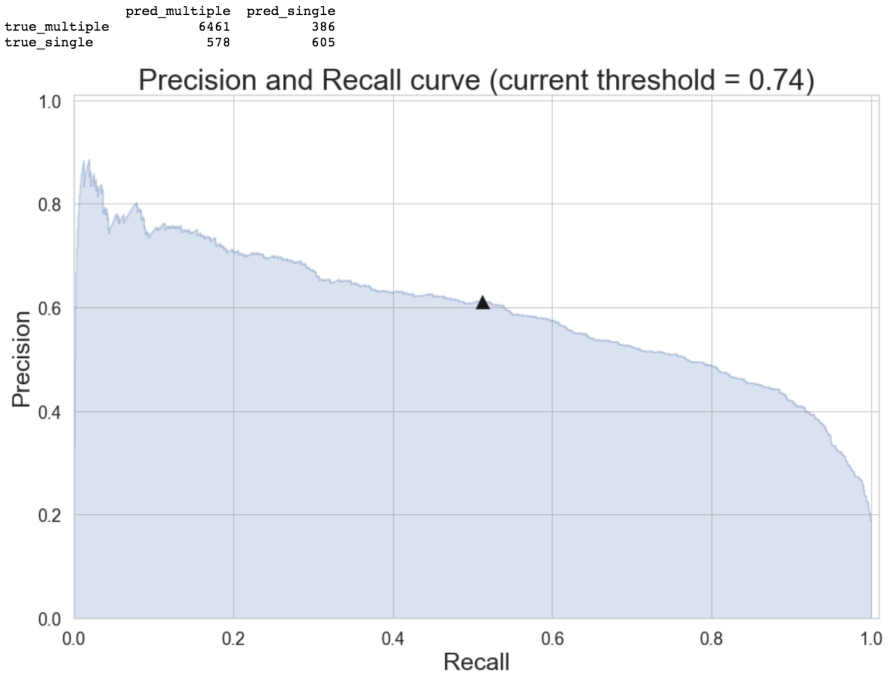


Figure 6.34: Precision versus Recall curve for the Random Forest classifier and confusion matrix for the chosen decision threshold at 0.74.

From Fig. 6.33 in particular, it can be seen that the default threshold of 0.5 does not do a good job in retrieving a clean signal as the precision is 0.5. Moving this threshold up does increase the precision but consequently reduce the recall score. The choice of where to set the decision boundary obviously depends on the particular problem under investigation and the desired outcome. Fig. 6.34 shows the precision versus recall curve together with the confusion matrix (top left) obtained by setting the decision threshold to 0.74 (represented by the triangle shape on the plot) which is the optimal chosen value for the ML decision cut from this study. The reason for this will be explained with the help of Fig. 6.35 where the effect of increasing the precision score by setting higher thresholds is summarised in similar precision-recall curve plots. The price to pay for reaching an almost 80% precision, i.e. a pretty clean sample of true single scatters (6.35(d)), is to lose a large amount of them as they will be incorrectly classified as multiple scatters. The confusion matrix on the top left side of Fig. 6.35(d) illustrates this concept pretty clearly: about 95% of the total true single scatters would be lost if the decision threshold would be set at 0.9 against 40% by setting it to 0.7 (6.35(b)). The first scenario would be ideal if a very large amount of data was easily available. As already discussed through this thesis, populating the nuclear recoil band with sufficient statistics is not a trivial task for either data or Monte Carlo

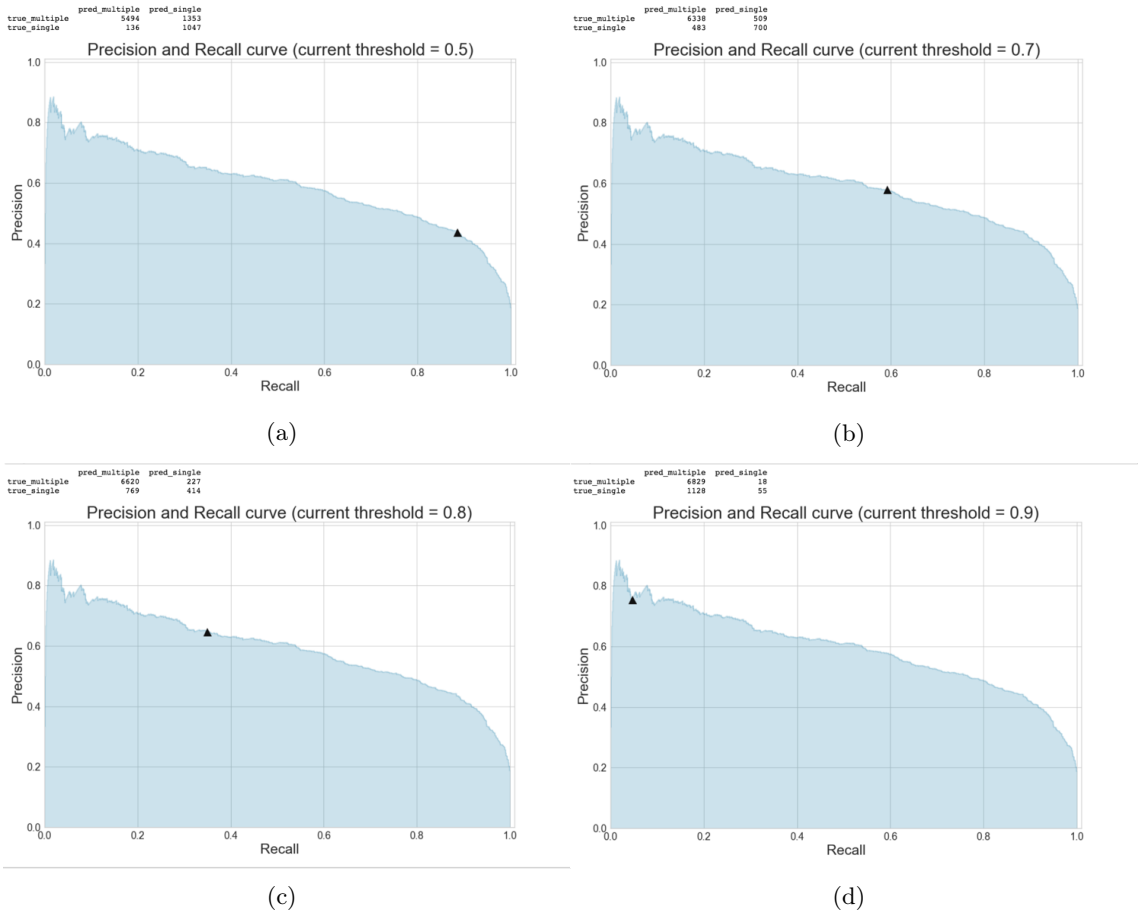


Figure 6.35: Precision versus Recall curves for the Random Forest classifier and confusion matrices as a function of the decision threshold. Threshold 0.5 (a), threshold 0.7 (b), threshold 0.8 (c) and threshold 0.9 (d).

simulated data. A decision threshold at 0.74 seems a good compromise between precision and recall in this particular case. The ROC curve for the Random Forest algorithm with this threshold is shown in Fig. 6.36 and presents an area under the curve (AUC) score of 0.91.

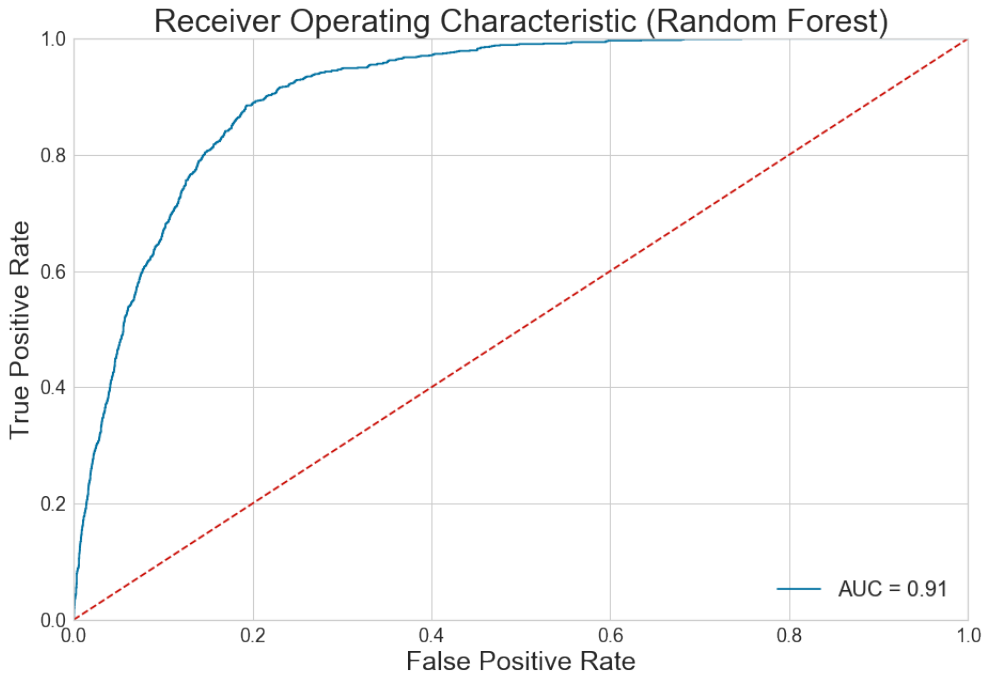


Figure 6.36: Receiver Operating Characteristic curve for the Random Forest classifier and it AUC score.

Let us focus on Fig. 6.37 and Fig. 6.38 to have a better understanding of the discrimination power of the Random Forest cut applied to the AmBe neutron Monte Carlo. The first plot shows the mean F_{prompt} (in 10 PE bins) versus qPE for Monte Carlo single scatters (red), Monte Carlo multiple scatters (blue), and "data" before (green) and after (black) applying the ML cut, where the word data is in quotation marks in both cases because what is plotted is actually simulated data that have been treated as if they were real for the purpose of exploring the goodness of this approach as explained in the previous sections. The first thing to notice when looking at Fig.6.37 is how close the green (data before ML cut) and blue (multiple MC) points are due to the presence of several multiple scatters in the data sample that pulls down the mean F_{prompt} distribution as already seen in Chapter 5 and the TMVA section of this Chapter. On the contrary, single scatters have a higher mean F_{prompt} across the whole qPE range, and the effect of applying the Random Forest cut can be seen in the black points whose mean F_{prompt} gets higher as multiple scatters are removed by the cut and either matches or gets very close to the red curve of single scatters.

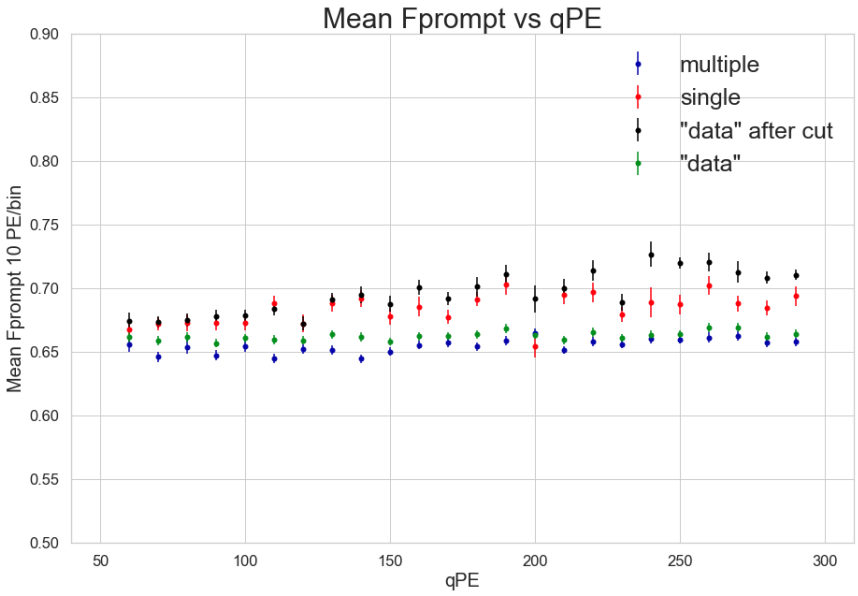
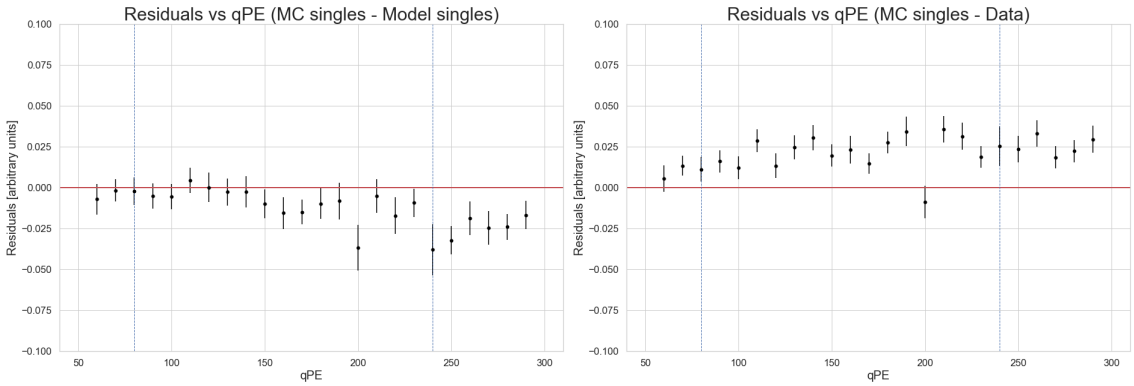


Figure 6.37: Mean F_{prompt} in 10 PE/bin versus qPE comparison across the ROI energy region between Monte Carlo multiple scatters (blue), Monte Carlo single scatters (red), AmBe Monte Carlo (treated as data in green) and AmBe Monte Carlo after Random Forest cut (black).

The goodness of the agreement can be seen in Fig. 6.38 where the residuals between the mean F_{prompt} single scatters Monte Carlo (red points) and the mean F_{prompt} of "data" selected by the Random Forest cut (black points) is shown in 6.38(a) versus the residuals between the mean F_{prompt} distribution of single scatters and the mean F_{prompt} of "data" without any cut applied in 6.38(b).



(a) (b)

Figure 6.38: Residuals versus qPE for single scatters selected by the Random Forest model (a) and for the mixture of single and multiple scatters from Monte Carlo "data".

The blue dashed vertical lines in both plots delimit the ROI energy region whilst the horizontal red line is where the points should lie if all of the single scatters present in the

sample would have been correctly identified. The Random Forest cut does a good job in the whole ROI energy region except for two bins in the high energy side at 200 and 240 qPE. The residual plot in Fig. 6.38(b) presents, as expected, an excess of points in the positive side of the residuals range. This is due to the fact that the data distribution used in this case is the original mixture of single and multiple scatters. In other words, it can be seen as the worst case scenario where no selection has been done and it can be used as a baseline to compare the effectiveness of the ML approach against.

New Monte Carlo and Data comparison

Being the final goal of this study the extraction of a clean sample of neutron single scatters from real neutron AmBe data by using the simulated data to train ML algorithms, it is important to make sure that the new set of simulations considered in this chapter and the data are describing the same phenomenon. In this section, a comparison of the sets of Monte Carlo and data utilised will be given. In Fig. 6.40 a comparison at three different energy thresholds up to 240 PE between data and Monte Carlo of the F_{prompt} distribution is shown. The very poor agreement at F_{prompt} smaller than 0.6 is expected as the new Monte Carlo does not include ^{39}Ar pile-up. The level of agreement in the F_{prompt} region between 0.6 and 0.8 (ROI) is instead pretty good. The leakage of high F_{prompt} events, especially when the energy threshold is lowered down to 80 PE (6.40(d)), confirm that the new Monte Carlo is still underestimating the Cherenkov light production in the real detector. It should be noted that the set of cuts applied differs from what was presented in the previous chapter 5. In particular, a cut on the fraction of scintillation photons seen by the PMTs and the charge seen by the two top rings of PMTs (those not under liquid argon) has been applied together with a fiducial volume cut (radial cut) to get a fiducial volume of 1 tonne.

Features selection based on match with data

In Section 6.6, a few methods for selecting features to use for building the ML model have been presented and implemented to obtain the results shown so far. Moving towards the final goal of extracting single scatters from data rather than from Monte Carlo "data" as has been done so far, it is important to make sure that the features used to build the model from the simulated single and multiple distributions have the same "shape" in data as well. For this reason, together with the feature selection arguments already presented,

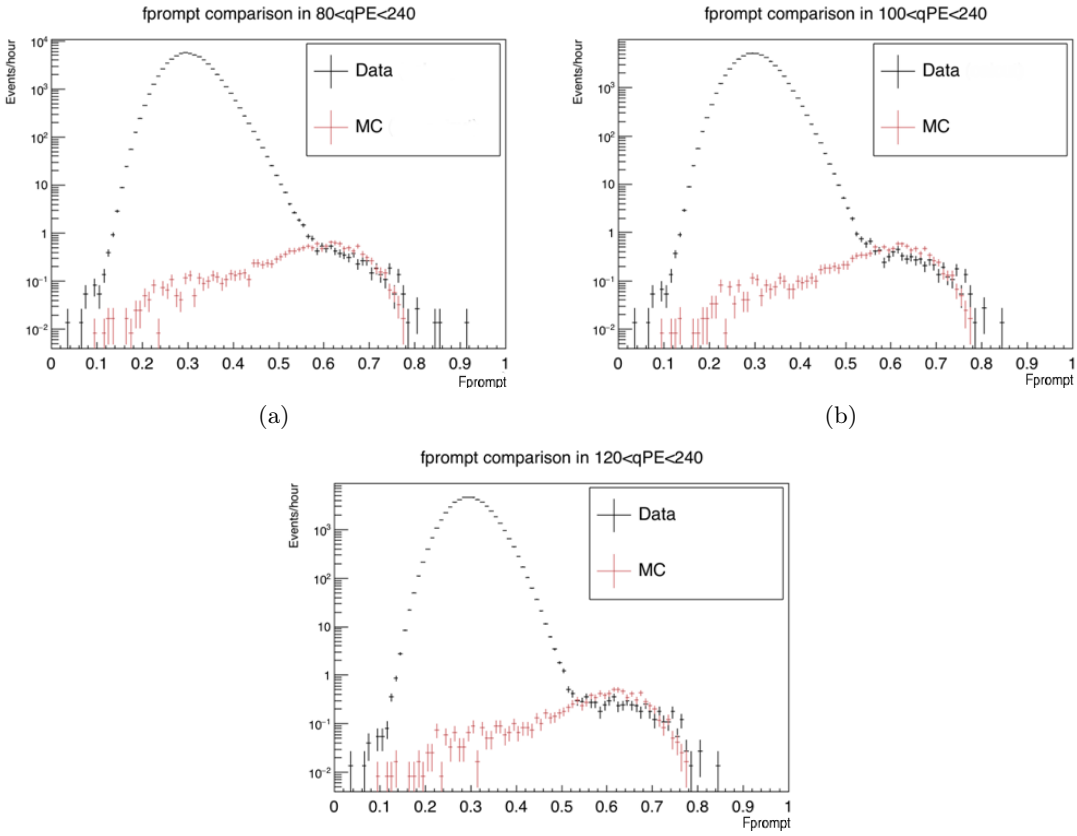


Figure 6.39: F_{prompt} distribution comparison between AmBe data and Monte Carlo in 3 energy region near the ROI: (a) $80 < qPE < 240$, (b) $100 < qPE < 240$, (c) $120 < qPE < 240$.

another criterion for selecting the model's input variables is introduced based on how well the variables in data and simulation match. There can be geometrical or physical reasons for which some variables can have completely different shapes and behaviour and these cases need to be highlighted and the features not included in the process of building the model.

A comparison between the new AmBe Monte Carlo and the AmBe data collected in March and April 2018 is shown for some variables in Fig. 6.40. What can immediately be gathered from this set of distributions is that one has to be careful to use position related features as there can be differences between the source position with respect to the detector between simulation and reality. In particular, the set of new Monte Carlo was simulated with the AmBe source always placed in the vertical calibration tube CAL A, whereas the data the dataset includes data from all three calibration tubes. This is reflected in the large misalignment between the peaks of the reconstructed angular ϕ distributions in

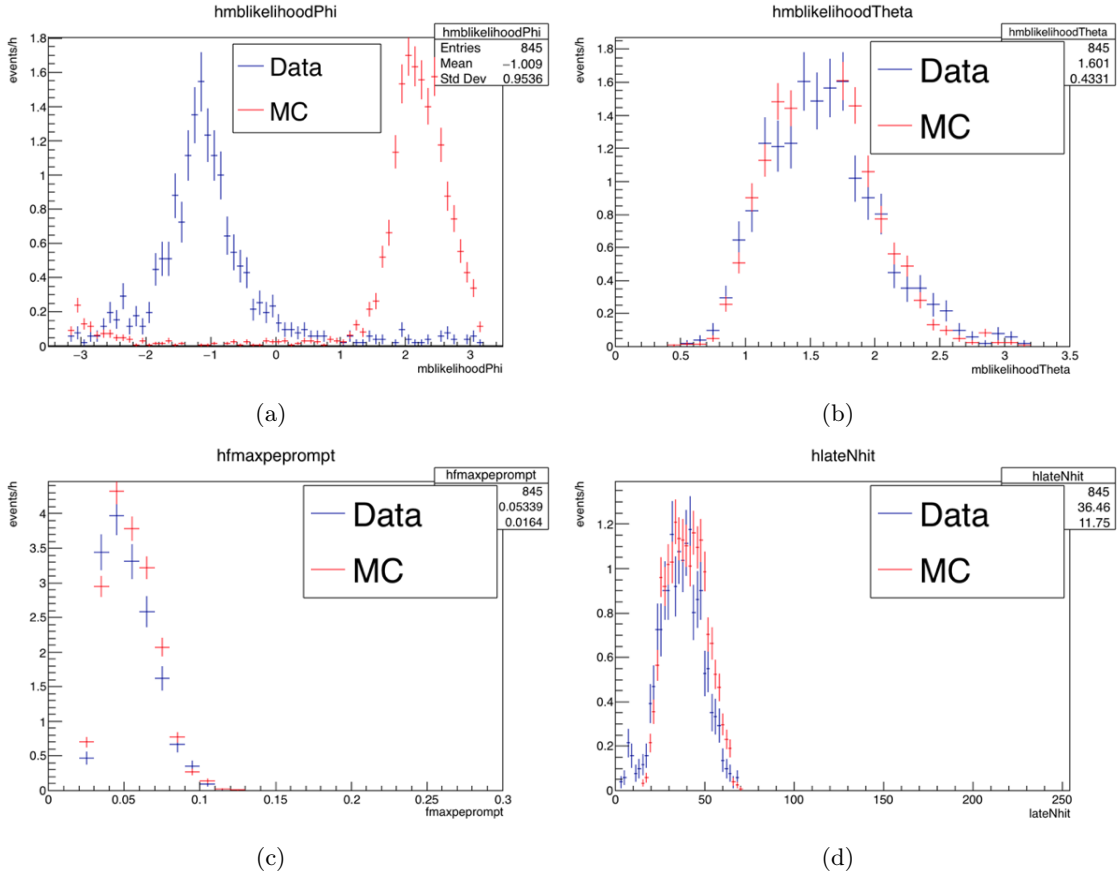


Figure 6.40: AmBe Monte Carlo and AmBe data comparison for specific variables. (a) Phi reconstructed position; (b) Theta reconstructed position; (c) prompt max charge fraction per PMT; (d) late number of PMTs hit.

6.40(d). This variable and other variables that show a similar behaviour, not necessarily excluded a priori from the model building process so far, will be excluded at this point as the finalised model will have to deal with unseen data rather than unseen simulated data as such it is important to make use of only Monte Carlo input variables that reflect the behaviour of data. That is the case for another position reconstruction related features such as the theta distribution in 6.40(b) that is not affected by the different calibration tubes utilised in collecting or generating data as the source is always place at the same height (detector's equator) in both cases. At the end of this selection process 34 out of the initial 157 features are selected for training the ML model. They are: *nhit, fmaxpe, fmaxnsc, lateNhit, logprobdiff, fprompt, frontHalfFraction, lateChargePMTNoPrompt, falpha, fmaxpeprompt, fmaxpelate, fmaxpenn, fmaxnscnn, chargetopring, chargesecondring, chargebottomring, precursorCharge, lateChargePMTHadPrompt, subeventPromptHW, mblikelihoodZ, mblikelihoodR, mblikelihoodRho, mblikelihoodTheta, mblikelihoodE, mblikeli-*

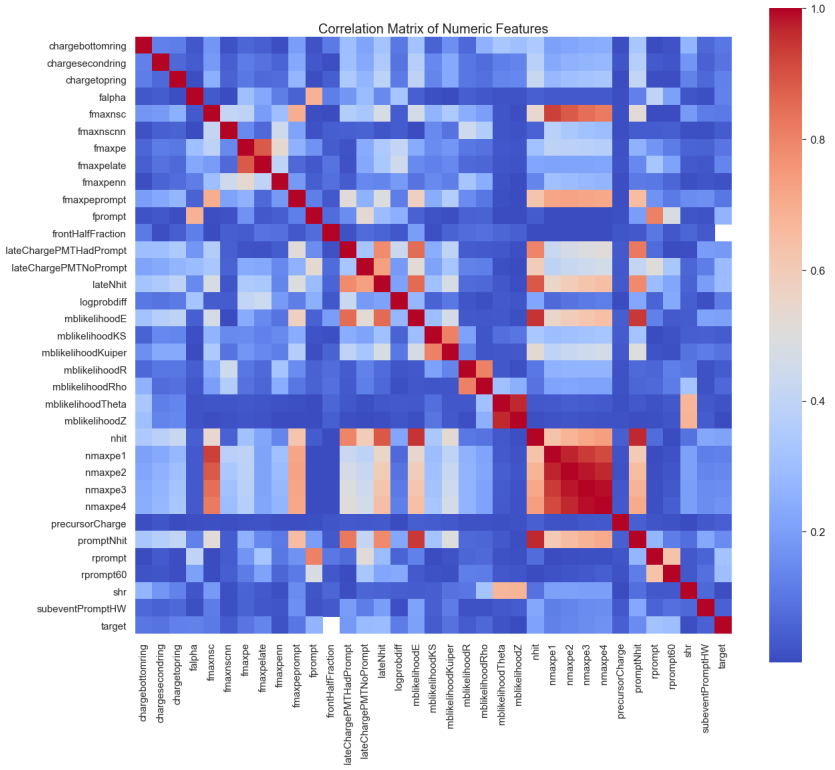


Figure 6.41: Correlation Matrix of features that shows agreement between AmBe data and AmBe Monte Carlo. Features with correlation higher than 95% are not used to train the model.

hoodKS, mblikelihoodKuiper, promptNhit, rprompt, rprompt60, shr, nmaxpe1, nmaxpe2, nmaxpe3, nmaxpe4.

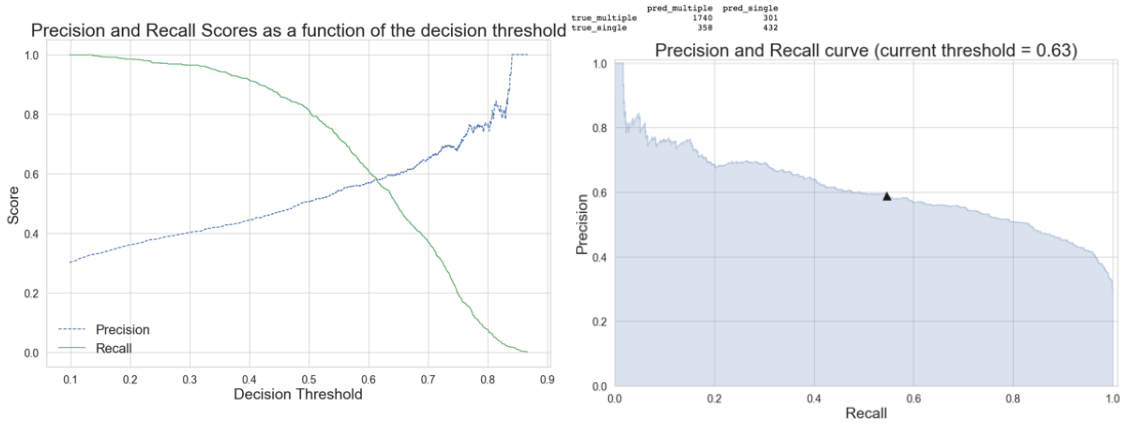
Five of them, namely *mblikelihoodZ, nmaxpe2, nmaxpe3, nmaxpe4* and *promptNhit* are discarded due to their correlation coefficient being greater than 95% as shown in Fig. 6.41, hence the final model is built by using 29 input variables.

A 3-fold cross validation was performed again with the dataset with smaller dimension. The best hyperparameters on the training set were as in Table 6.6

Table 6.6: Best hyperparameters set.

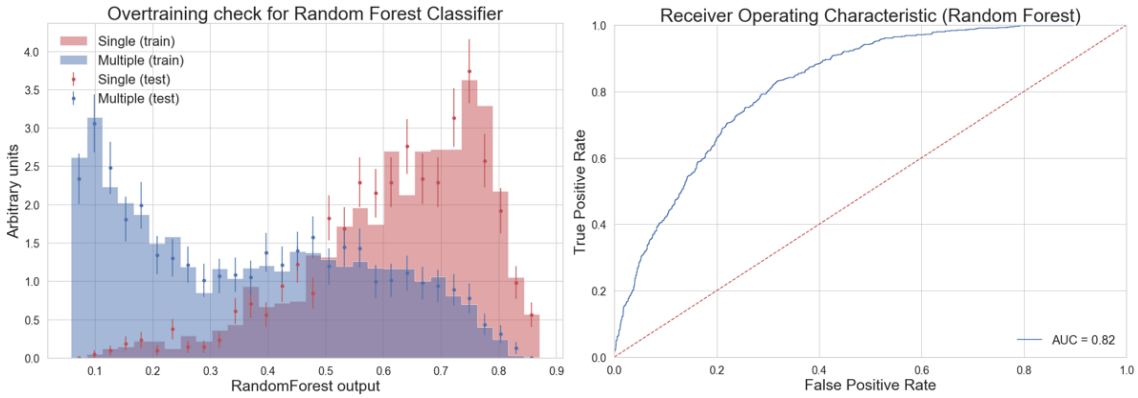
Hyperparameter	**Value**
bootstrap	False
criterion	entropy
max_depth	4
max_features	5
min_samples_leaf	1
min_samples_split	3
n_estimators	2500

The Random Forest model, re-trained by using the 29 variables only and the new hyper-parameters gives the Precision-Recall as a function of the decision threshold shown in Fig. 6.42.



(a) (b)

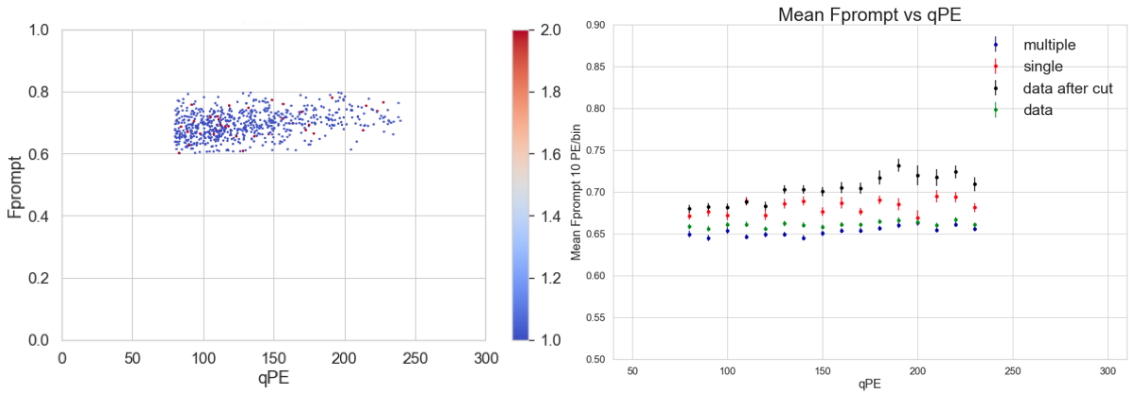
Figure 6.42: (a) Precision and Recall as a function of the Random Forest decision threshold for model built with features that matches between data and Monte Carlo (b) Precision and Recall for the chosen threshold.



(a) (b)

Figure 6.43: (a) Overtaining check for the Random Forest classifier with the final set of input variables. (b) Receiver Operating Characteristic Curve for the Random Forest and its AUC score.

Fig. 6.44(a) shows the F_{prompt} versus qPE distribution of the events that survive the Random Forest Cut. Fig. 6.44(b) the mean F_{prompt}versus qPE for the four distributions taken into consideration when first building the model.

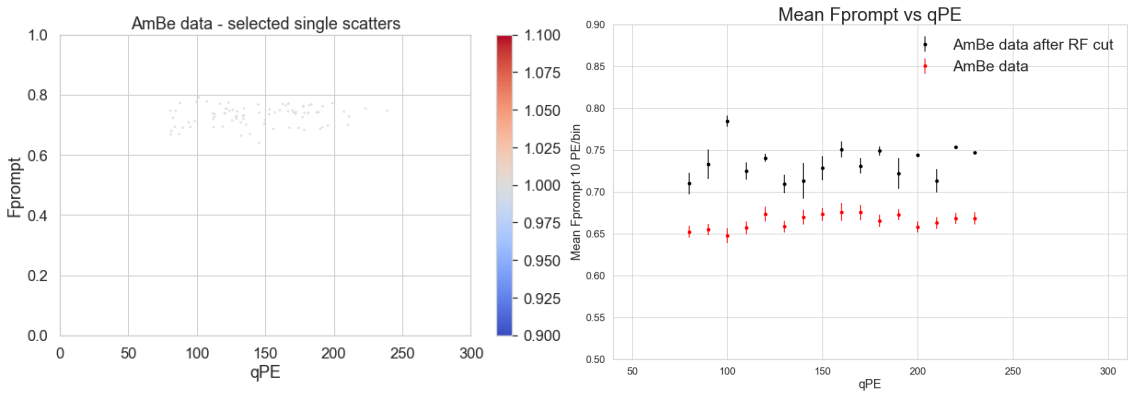


(a) (b)

Figure 6.44: (a) F_{prompt}versus qPE for the AmBe "data" events surviving the Random Forest cut; (b) Mean F_{prompt} versus qPE for single scatters Monte Carlo (red), multiple scatters Monte Carlo (black), AmBe "data" (green) and AmBe "data" after Random Forest cut.

6.7 Single scatters extraction from real AmBe data

Having trained a model to Monte Carlo data by using only input variables that reproduce as closely as possible the source interaction with the detector in the real world, the last step in the process of extracting a clean sample of single scatters from real AmBe neutron data is to apply the machine learning cut to these data and select only events that belong to the 'signal' class. By doing this to the AmBe neutron dataset collected in March and April 2018, the sample in Fig. 6.45(a) is obtained, where the F_{prompt} versus qPE for those events in the ROI is plotted. Fig. 6.45(b) shows the mean F_{prompt} versus qPE in 10 PE per bin for the same AmBe dataset after single scatters extraction by the Random Forest cut in black and before in red. The black distribution has a higher mean F_{prompt} than red as expected for neutrons that interact through a single scatter. Both plots highlight a clear deficit in statistics that is due to the fact that at the time of writing only the March and April 2018 dataset was available with calibration processing to be used for this analysis. Moreover, on the contrary of what has been done so far by treating AmBe Monte Carlo as data, dealing with real unlabelled data, it is impossible to give an *a posteriori* score of the model performance. However, given the results of the previous section (see Fig. 6.42) it is safe to state that this selection method ensures that 60% of selected events in the sample of Fig. 6.45(a) are true single scatters.



(a) (b)

Figure 6.45: (a) F_{prompt} versus qPE for the AmBe (data from March and April 2018) events surviving the Random Forest cut; (b) Mean F_{prompt} versus qPE for single scatters extracted from the AmBe data after Random Forest cut (black), AmBe data before Random Forest cut.

Chapter 7

Conclusions

7.1 Summary and Conclusions

The DEAP-3600 detector is a world leading [30] liquid argon direct dark matter search experiment that is currently collecting physics data. This work has presented the effort done to calibrate the detector both with optical systems and external radioactive sources. The laser source (laserball), deployed within the inner acrylic vessel prior to liquid argon filling, was used to remove the PMTs channel-to-channel time offsets and extract a common T_0. The time offsets extracted are calibration constants applied to calibrate raw data during the processing phase. The combination of laserball runs with the source at the centre of the detector and the use of a PPG source was used to get an alignment of all channel pulses at the sub-nanosecond level.

In order to achieve a low energy WIMP search threshold, pulse shape discrimination has been characterised using a 74 MBq AmBe neutron source. The complexity of neutron interactions has required developing a method of simulation to achieve the large statistics required to populate the detector with neutron scatters. The DEAP-3600 detector is designed and built in order to shield against neutrons as they can mimic a WIMP-like interaction, that would constitute a major source of background. This property results in the necessity to simulate a very large number of neutrons to guarantee that a statistically significant number of them make it to the liquid argon volume. The achievement of this goal with a "classic" Geant4 simulation would be impossible in a reasonable timeframe given the very high number of steps and different interactions that each particle makes travelling from the source through the several volumes of the DEAP-3600 Geant4 geometry. Moreover, a very large amount of disk space would be required. To overcome this

limitation, a two stages Monte Carlo generation method for the neutron AmBe source, that allows the generation of billions of events in a limited amount of time with reduced disk space requirements, has been developed. In the first stage billions of neutrons have been simulated from one of the three external calibration tubes with a simplified detector geometry where some volumes are removed such that the number of interaction steps each particles undergoes are reduced. The electronic/DAQ response has also been switched off at this stage. The reduction of events recorded to disk has been implemented by developing a dedicated processor ("online trigger") that allows the retention, while running the simulation, of only events where a neutron makes it into the AV volume whilst discarding all others. The most energetic inelastic gammas from neutron interactions in the detector materials have also been recorded at this stage for being re-simulated in the second stage. This method has allowed the generation of simulated neutrons from an AmBe spectrum for an equivalent livetime of approximately 180 hrs. A full simulation that includes the complete DEAP-3600 geometry has been performed by re-simulating neutrons and inelastic gammas from stage 1 with their momentum as recorded in the last GEANT4 step before entering the detector and their position extrapolated to the edge of the detector volume using their first step in the detector volume and their last step outside it. The neutron AmBe Monte Carlo thus generated has been compared to the data collected during the 74 MBq AmBe neutron source calibration runs in DEAP-3600. It has emerged that even though the Monte Carlo does not fully match AmBe data across the entire energy range explored, the agreement is quite good in the energy region defined as the basic ROI ($120<PE<240$) for the WIMP search. A detailed study has shown that the disagreement in the low energy range, particularly at $PE<80$, can be explained by the fact that the Monte Carlo underestimates the Cherenkov events produced by the source in data. The agreement in the region from 80 PE up to 240 PE has been shown to improve considerably when 4.4 MeV gamma simulation that are associated with the AmBe source and a selection of Cherenkov events from a Thorium source run are added. This evidences, reported in Chapter 5, has led to a refinement of the Monte Carlo generation process where the 4.4 MeV gamma has been simulated in stage 1 though the use of a "coincidence" generator and the radius of acceptance and storage of neutrons that interact with the AV has been increased. The two sets of Monte Carlo thus generated have been used to develop a Machine Learning model that aims to isolate single scatter (WIMP-like) events that, due to the nature of neutron interactions that can scatter multiple times as they traverse the detector, are very chal-

lenging to identify. In particular, after having shown that the simulation agrees with the AmBe data in the region of interest, different binary classification algorithms have been trained within this energy region on simulated data where it is possible to easily identify single (signal) and multiple (background) scatters, in order to produce a ML based cut that reduces the presence of multiple scatters in AmBe data when applied to them. Two independent libraries (TMVA and ScikitLearn) have been used for this purpose both showing that a 60 % signal efficiency in extracting single scatter events among those populating the nuclear recoil band near the WIMP search energy region of interest can be achieved by using an ensemble algorithm such as BDT and Random Forest. Having the ability of isolate a clean sample of neutron single scatters from the AmBe data nuclear recoil band is important for defining the WIMP acceptance for the dark matter search. This work has shown that a combination of Monte Carlo simulation and machine learning techniques can help in this direction. In particular, increasing the AmBe data statistic would allow to use a more stringent ML cut increasing the signal efficiency, overcoming the issue of having a very limited number of single scatter events in the region of interest. Work to improve the Monte Carlo DEAP-3600 optical model, the fluorescence model in acrylic, would lead to better modelling neutron interactions in the detector thus improving the quality of the AmBe simulation which are at the core of the ML process. In a future prospective and with larger statistics, the nuclear recoil mean F_{prompt} obtained by using the methodology developed in this work, can be also implemented in the profile likelihood ratio (PLR) method to improve the dark matter search.

Bibliography

[1] F. Zwicky, Helvetica Physica Acta **6**, 110 (1933).

[2] F. Zwicky, Astrophys. J. **86**, 217 (1937).

[3] V. C. Rubin and W. K. Ford, Jr., Astrophys. J. **159**, 379 (1970).

[4] P. C. 2015A, Astronomy and Astrophysics .

[5] K. G. Begeman, A. H. Broeils, and R. H. Sanders, Monthly Notices of the Royal Astronomical Society **249**, 523 (1991), /oup/backfile/content_public/journal/mnras/249/3/10.1093_mnras_249.3.523/2/mnras249-0523.pdf.

[6] D. Perkins, *Particle Astrophysics*, 2 ed. (Oxford Master Series In Particle Physics, Astrophysics, and Cosmology).

[7] D. Clowe *et al.*, Astrophys. J. **648**, L109 (2006), astro-ph/0608407.

[8] L. Roszkowski, E. M. Sessolo, and S. Trojanowski, Rept. Prog. Phys. **81**, 066201 (2018), 1707.06277.

[9] V. Rubakov, p. 239 (2018), 56 pages, presented at the 2nd Asia-Europe-Pacific School of High-Energy Physics, Puri, India, 04 - 17 Nov 2014. arXiv admin note: substantial text overlap with arXiv:1504.03587.

[10] R. H. Cyburt, B. D. Fields, and K. A. Olive, JCAP **0811**, 012 (2008), arXiv/0808.2818.

[11] EROS-2, P. Tisserand *et al.*, Astron. Astrophys. **469**, 387 (2007), astro-ph/0607207.

[12] G. Jungman, M. Kamionkowski, and K. Griest, Physics Reports **267**, 195 (1996).

[13] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rept. **267**, 195 (1996), hep-ph/9506380.

[14] PAMELA, O. Adriani *et al.*, Nature **458**, 607 (2009), arXiv/0810.4995.

[15] AMS Collaboration, M. Aguilar *et al.*, Phys. Rev. Lett. **117**, 091103 (2016).

[16] IceCube, M. G. Aartsen *et al.*, Phys. Rev. **D88**, 122001 (2013), arXiv/1307.3473.

[17] T. Marrodán Undagoitia and L. Rauch, J. Phys. **G43**, 013001 (2016), arXiv/1509.08767.

[18] PICO, C. Amole *et al.*, Phys. Rev. Lett. **114**, 231302 (2015), arXiv/1503.00008.

[19] P. L. Brink *et al.*, Journal of Physics: Conference Series **150**, 012006 (2009).

[20] E. A. and, Journal of Physics: Conference Series **375**, 012004 (2012).

[21] SuperCDMS, R. Agnese *et al.*, Phys. Rev. **D91**, 052021 (2015), arXiv/1410.1003.

[22] CRESST-II, G. Angloher *et al.*, Eur. Phys. J. **C74**, 3184 (2014), arXiv/1407.3146.

[23] CoGeNT, C. E. Aalseth *et al.*, (2014), arXiv/1401.3295.

[24] DRIFT, J. B. R. Battat *et al.*, Phys. Dark Univ. **9-10**, 1 (2015), arXiv/1410.7821.

[25] XENON, E. Aprile *et al.*, Phys. Rev. Lett. **119**, 181301 (2017), arXiv/1705.06655.

[26] J. Li, Physics World **27**, 9 (2014).

[27] LUX, C. F. P. da Silva, Dark Matter Searches with LUX, in *Proceedings, 52nd Rencontres de Moriond on Very High Energy Phenomena in the Universe: La Thuile, Italy, March 18-25, 2017*, pp. 199–209, 2017, 1710.03572.

[28] LUX, D. S. Akerib *et al.*, Phys. Rev. Lett. **118**, 021303 (2017), arXiv/1608.07648.

[29] PandaX, M. Xiao *et al.*, Sci. China Phys. Mech. Astron. **57**, 2024 (2014), arXiv/1408.5114.

[30] DEAP-3600 Collaboration, P.-A. Amaudruz *et al.*, Phys. Rev. Lett. **121**, 071801 (2018).

[31] E. Shields, J. Xu, and F. Calaprice, Phys. Procedia **61**, 169 (2015).

[32] J. Amare *et al.*, Eur. Phys. J. **C76**, 429 (2016), arXiv/1604.05587.

[33] DM-Ice, E. Barbosa de Souza *et al.*, Phys. Rev. **D95**, 032006 (2017), arXiv/1602.05939.

[34] PICO-LON, K. Fushimi *et al.*, J. Phys. Conf. Ser. **718**, 042022 (2016), arXiv/1512.04645.

[35] Particle Data Group, M. Tanabashi *et al.*, Phys. Rev. **D98**, 030001 (2018).

[36] A. I. Bolozdynya, Two-phase emission detectors and their applications, 1999.

[37] M. Schumann, JINST **9**, C08004 (2014), arXiv/1405.7600.

[38] M. M. E. Lemmon and D. Friend, Thermophysical properties of fluid systems, NIST Chemical Webbook, 2014, NIST Standard Reference Database Number 69.

[39] S. Kubota, M. Hishida, and J. Raun, Journal of Physics C: Solid State Physics **11**, 2645 (1978).

[40] A. Hitachi *et al.*, Phys. Rev. **B27**, 5279 (1983).

[41] V. Chepel and H. Araujo, JINST **8**, R04001 (2013), arXiv/1207.2292.

[42] J. Nikkel, Oddities of light production in the noble elements - seminar at university of yale, URL: `https://wlab.yale.edu/sites/default/files/seminar%20pdfs/nikkel_slides.pdf`, 2015.

[43] S. Kubota, M. Hishida, M. Suzuki, and J.-z. Ruan(Gen), Phys. Rev. B **20**, 3486 (1979).

[44] SCENE, T. Alexander *et al.*, Phys. Rev. **D88**, 092006 (2013), arXiv/1306.5675.

[45] T. Doke *et al.*, Jap. J. Appl. Phys. **41**, 1538 (2002).

[46] D. Gastler *et al.*, Phys. Rev. **C85**, 065811 (2012), arXiv/1004.0373.

[47] SCENE, H. Cao *et al.*, Phys. Rev. **D91**, 092007 (2015), arXiv/1406.4825.

[48] A. Hitachi, T. Doke, and A. Mozumder, Phys. Rev. **B46**, 11463 (1992).

[49] D. M. Mei, Z. B. Yin, L. C. Stonehill, and A. Hime, Astropart. Phys. **30**, 12 (2008), arXiv/0712.2470.

[50] W. H. Lippincott, *Direct detection of dark matter with liquid argon and neon*, PhD thesis, Yale University, USA, 2010.

[51] J. Lindhard, M. Scharff, and H. Schioett, Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd. .

[52] J. B. Birks and F. A. Black, Proceedings of the Physical Society. Section A **64**, 511 (1951).

[53] W. Brandt and M. Kitagawa, Phys. Rev. B **25**, 5631 (1982).

[54] B. Smith, Deap underground in canada - susy 2015, URL: `http://deap3600.ca/Contents/presentations/DEAP-3600SUSY2015Smith.pdf`, 2015.

[55] DEAP Collaboration, P. A. Amaudruz *et al.*, Astropart. Phys. **85**, 1 (2016), arXiv/0904.2930.

[56] DEAP-3600, P. A. Amaudruz *et al.*, Submitted to: Astropart. Phys. (2017), 1712.01982.

[57] V. M. Gehman *et al.*, Nucl. Instrum. Meth. **A654**, 116 (2011), arXiv/1104.3259.

[58] B. Broerman *et al.*, JINST **12**, P04017 (2017), arXiv/1704.01882.

[59] SNOLAB, *SNOLAB User's Handbook*, 2 ed. (, 2006).

[60] J. Walding, DEAP-3600 Neutron Source Safety Review Document for the 74 MBq AmBe Source, DEAP-STR2013x009, private communication.

[61] A. Butcher, *Searching For Dark Matter with DEAP-3600*, PhD thesis, Royal Holloway University of London, UK, 2015.

[62] B. Singh and J. A. Cameron, Nuclear Data Sheets **107**, 225 (2006).

[63] WARP, P. Benetti *et al.*, Nucl. Instrum. Meth. **A574**, 83 (2007), astro-ph/0603131.

[64] E. Vazquez-Jauregui, Neutron backgrounds in DEAP-3600: simulations, DEAP-STR2017x008, private communication.

[65] Hamamatsu, *PLP-10Laser diode head Series* (HAMAMATSU PHOTONICS K.K., Systems Division, 2014).

[66] A. Hallin, PPG/DNF Triggers and the Time Response in Run 9285, DEAP-3600 Analysis Note, private communication.

[67] R. Acciarri *et al.*, Phys. Procedia **37**, 1113 (2012).

[68] M. Bhike, B. Fallin, and W. Tornow, Physics Letters B **736**, 361 (2014).

[69] Eckert and ziegler industrial sources, `https://www.ezag.com/fileadmin/ezag/user-uploads/pdf/isotope/5_industrial_sources.pdf`, Accessed: 2018-08-22.

[70] T. Caldwell, Simulation of Noble Liquid Detectors Using RAT, Presentation, 2014, AARM Meeting, Fermilab March 19-22.

[71] M. Kimura, M. Tanaka, T. Washimi, and K. Yorita, (2019), 1902.01501.

[72] A. Hocker *et al.*, PoS **ACAT**, 040 (2007), arXiv/0703039.

[73] F. Pedregosa *et al.*, Journal of Machine Learning Research **12**, 2825 (2011).

[74] M. Kuhn and K. Johnson, *Applied Predictive Modeling* (Springer, New York, Heidelberg, Dordrecht, London, 2013).

[75] C. Li and B. Wang, *Fisher Linear Discriminant Analysis*.

[76] B. P. Roe, H.-J. Yang, and J. Zhu, Boosted decision trees, a powerful event classifier, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT 05): Proceedings, Oxford, UK, September 12-15, 2005*, pp. 139–142, 2005.

[77] A. Voigt, *Boosting PDE-Foam*, Summer project at CERN Jul Aug 2010.