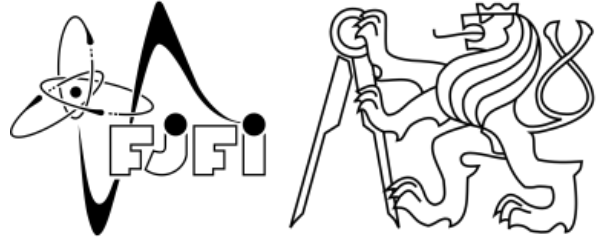Czech Technical University in Prague

Faculty of Nuclear Sciences and
Physical Engineering

Department of Physical Electronics

Master's Thesis

Advanced Methodology for Radiation Field Decomposition with Hybrid Pixel Detectors
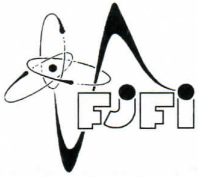
Declan Garvey

Supervisor:
Dr. Benedikt Bergmann

Co-supervisor:
Doc. Dr. Milan Šiňor

May 2, 2023

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ

Katedra fyzikální elektroniky

MASTER THESIS ASSIGNMENT

Student:	**Bc. Declan G a r v e y**
Study program:	**Physical Electronics**
Specialization:	**Computational Physics**
Academic year:	**2022/2023**
Title: (in Czech)	**Pokročilé metody dekompozice radiačního pole hybridními pixelovými detektory**
Title: (in English)	**Advanced methodology for radiation field decomposition with hybrid pixel detectors**
Language of thesis:	**English**

Thesis objective:

The aim of the work is to improve currently used methodology for analysis of radiation fields with hybrid pixel detectors of the Timepix family using statistical and machine learning approaches. Developed algorithms should then be applied to data taken in particle physics experiments (e.g. ATLAS or MoEDAL) or in space.

Tasks to complete:

1. Familiarise yourself with the detector technology, currently used analysis methodology and the Allpix^2 detector response simulation;
2. Create a database of simulated ground truth data and validate the simulated response with available measurements;
3. Understand the signatures of different particles and evaluate features sensitive to particle type and energy;
4. Propose statistical and/or machine learning models for particle separation and energy estimation, evaluate their performance with the help of the previously created database;
5. The thesis should be written in English.

Recommended literature:

1. G. Knoll, Radiation Detection and Measurment, 4th edition, 2010, John Wiley & Sons, ISBN: 978-0-470-13148-0.
2. T. Holy et al., Pattern recognition of tracks induced by individual quanta of ionizing radiation in Medipix2 silicon detector, Nucl. Instrum. Meth. A, Vol. 591, pp. 287-290, 2008.
3. T. Poikela et al., Timepix3: A 65 k channel hybrid pixel readout chip with simultaneous toa/tot and sparse readout, JINST 9 C05013, 2014.
4. B. Bergmann et al., 3D track reconstruction capability of a silicon hybrid active pixel detector, Eur. Phys. J. C 77, 421, 2017.
5. B. Bergmann et al., 3D reconstruction of particle tracks in a 2 mm thick CdTe hybrid pixel detector, Eur. Phys. J. C 79, 165, 2019.
6. P. Mánek et al., Randomized Computer Vision Approaches for Pattern Recognition in Timepix and Timepix3 Detectors, in proceeding of CDT/WIT 2019, arxiv: 1911.02367, 2019.
7. I. Goodfellow et al., Deep Learning, MIT Press, 2016.
8. S. Spannagel et al., Allpix2: A modular simulation framework for silicon detectors, Nucl. Inst. Meth. A, Vol. 901, p. 164-172, 2018.

Name and institution of the supervisor:

Dr. rer. nat. Benedikt Bergmann
Institute of Experimental and Applied Physics, CTU in Prague.

Name and institution of the co-supervisor:

doc. Dr. Ing. Milan Šiňor
Department of Physical Electronics, Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague.

Assignment date:	13th October 2022
Submission date:	3rd May 2023

The validity of the assignment is two years after the assignment date.

Garant of the study program

Head of department   *Dean*

In Prague, 13th of October 2022

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

   I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilisation of this thesis as a school work under the provisions of Article 60 (1) of the Act.

Signed: ....

Dated: ....

Acknowledgements

First, I would like to thank my supervisors, Benedikt Bergmann and Milan Šiňor, for their guidance and support throughout the work on this thesis. Benedikt for all his help and always being always being available for meetings throughout the year, and Milan in aiding with university administration and submission of the thesis. Second, I would like to thank the Institute of Experimental and Applied Physics in Prague for their abundance of support and resources throughout my thesis. More specifically I would like to thank my colleagues from the office, especially Petr Smolyanskiy for his help with the use of the Allpix^2 simulation framework and Stanislav Pospíšil for his physics and Czech knowledge throughout the year. Third, and most of all, I would like to thank Corey Fletcher for her help and support throughout the year, without which this work would not have been possible.

Abstrakt

Hybridní pixelové detektory vyvolaly zájem fyziky vysokých energií, medicíny a věd o vesmíru pro jejich schopnosti registrovat bezpozaďově jednotlivé ionizující částice a zaznamenávat jejich stopy či dráhy v detektorovém polovodičovém sensoru. Předložená práce se zaměřuje jmenovitě na hybridní detektory typu Timepix3, jež umožňují identifikovat částice ve složitých radiačních polích v širokém rozsahu jejich energií.

V předložené práci byly simulovány neidealizované datové sady popisující dráhy ionizujících částic interagujících v pixelovém senzoru, které byly následně experimentálně ověřeny. Na základě těchto datových sad byly vyvinuty nové algoritmy pro identifikaci částic, analýzu jejich drah v senzoru a pro určování jejich toků. Pro sledování drah částic byl vyvinut Random Forest Regressor, se kterým byla dosahována průměrná absolutní chyba $8,65°$, což představuje zlepšení přesnosti o 38% oproti jinému, dosud užívanému nejvýkonnějšímu algoritmu. Pro klasifikaci částic byly vyvinuty dva nové algoritmy: jeden využívající Bayesiánskou dekonvoluci a druhý, založený na využití strojového učení v kombinaci s klasifikátorem XGBoosted a Random Forest Regressor. Zatímco algoritmus vycházející z Bayesiánské dekonvoluce dovolil dosáhnout přesnosti klasifikace elektron/proton $93,06\%$ a předpovědi celkové energie protonu s neurčitostí $54,65\ MeV$, v případě využití strojového učení byla přesnost/správnost klasifikace elektron/proton rovna $96,29\%$, a neurčitost předpovědi celkové energie protonu $42,15\ MeV$.

Vyvinuté algoritmy byly poté experimentálně testovány na svazku protonů v Dánském centru pro částicovou terapii. Experimentální testy přivedly ke konzistentním výsledkům. Tyto algoritmy byly pak využity k analýze dat naměřených v neznámých radiačních polích, a to jak v mimozemském prostředí s detektorem SATRAM umístěným na družici Evropské kosmické agentury Proba-V obíhající Zemi na nízké oběžné dráze, tak z experimentu MoEDAL na Large Hadron Collider v CERN. V práci předložená analýza dat ze SATRAMu, na kterém byl použit detektor Timepix v otevřeném prostoru, doplněná o jejich statistické zpracování, umožnila poprvé úspěšně určit protonové spektrum v podmínkách tzv. radiační anomálie nad Jižním Atlantikem. Tyto výsledky navíc dokládají možnosti využití detektorů typu Timepix k vývoji modelových představ o radiačních polích kolem Země. Pokud jde o experimentální data získaná systémem detektorů Timepix3 v experimentu MoEDAL na LHC,

algoritmy sledování drah částic v sensoru dovolily určovat úhly $\theta = 56 \pm 1°$ a $\varphi = 139 \pm 1°$, které odpovídají 3D-směrům letu částic při vstupu do sensoru. To umožnilo rozpoznat interakční bod, ve kterém se srážejí vstřícné svazky urychlených iontů a jeho rozměry s lepším rozlišením než dosud a registrovat jeho rozšíření při srážkách iontů olova ve srovnání s velikostí interakčního bodu při srážkách protonových svazků. Sledování úhlů dopadu jednotlivých částic dovolilo také odlišit primární částice zrozené v interakčním bodě od částic, které jsou součástí radiačního pozadí v experimentu. Oddělené využití klasifikačních algoritmů na částice pocházející z interakčního bodu a částice z pozadí umožnilo také určit složení i spektrální charakteristiky částic tvořících celkové radiační pole v tomto experimentu.

Klíčová slova   Medipix, Timepix, Timepix3, hybridní pixelové detektory, smíšená radiační pole, sledování částic, identifikace částic, tok, fluence, vesmírné počasí, CERN, MoEDAL

Abstract

Hybrid pixel detectors have gathered interest in the fields of high energy physics, medicine, and space science for their noiseless single particle processing capabilities. The presented work concentrates on hybrid detectors of the Timepix3 type, and their capability to identify and track particles of the entire radiation field.

In the presented work, non-idealised datasets describing traces of ionising particles interacting in a pixel sensor were simulated and subsequently validated through experiment. Based on these datasets, novel algorithms for particle fluence, identification, and tracking were developed. For particle tracking, a Random Forest Regressor was developed, producing a mean absolute error of $8.65°$, and thus a 38% increase in accuracy over the next-best performing state-of-the-art algorithm. For particle classification, two novel algorithms were developed: one utilising Bayesian deconvolution and the other utilising machine learning with a combination of an XGBoosted classifier and Random Forest Regressor. The deconvolution and machine learning algorithms produced an electron/proton classification accuracy of 93.06% and 96.29%, respectively, and a total proton energy prediction accuracy of 54.65 MeV and 42.15 MeV, respectively.

The developed algorithms were then tested in an experimental environment from proton beam exposure at the Danish Centre for Particle Therapy, from which consistent results were obtained. The algorithms were later applied to unknown complex radiation fields in the extra-terrestrial environment from the SATRAM detector, located on the European Space Agency satellite Proba-V in low Earth orbit, and at the MoEDAL experiment at the Large Hadron Collider located in CERN. The South Atlantic Anomaly was successfully isolated by analysing the data obtained from SATRAM, on which a Timepix detector was used. For the first time using Timepix detectors, following statistical manipulation, a physically reasonable proton spectrum was extracted. These results present the possibility of using Timepix-type detectors to develop models of the space environment in low Earth orbit. Regarding the experimental data obtained by the Timepix3 detector system in the MoEDAL experiment, the particle tracking algorithms allowed for the determination of the 3D-orientation of particles born at the point of collisions of opposing beams at the Large Hadron Collider to be $\theta = 56 \pm 1°$ and $\varphi = 139 \pm 1°$ angles. At the same time, the interaction point was determined

with a better resolution than previously observed. This then allowed for the demonstration of the increase in the size of the interaction point during lead-lead collisions compared to the size of the interaction point during proton-proton collisions. Subsequently, the tracking of the angles of incidence of individual particles made it possible to distinguish primary particles born at the interaction point from particles that are part of the radiation background in the experiment. The classification algorithms were then applied to the interaction point and background separately, producing physically reasonable particle spectrums.

Acronyms and Abbreviations

θ Angle of incidence relative to the normal

φ Azimuthal angle of incidence

CERN   European Organization for Nuclear Research

CNN   Convolution Neural Network

DCPT   Danish Centre for Particle Therapy

fToA   fast Time-of-Arrival

IP   Interaction Point

L_{eff} effective path length

LHC   Large Hadron Collider

LLM   Left Lower-most, Right Upper-most

ML   Machine Learning

MoEDAL   The Monopole and Exotic Detector at the LHC

PbPb   Lead-Lead

pp   Proton-Proton

SAA   South Atlantic Anomaly

SATRAM   Space Application of Timepix based Radiation Monitor

sToA   slow Time-of-Arrival

THL   Threshold Level

ToA   Time-of-Arrival

ToT   Time-over-Threshold

V_{eff} Effective Volume

Contents

x

1 | Introduction

Figure 1.1: Picture of a Timepix3 detector with a silicon sensor layer consisting of 256×256 pixels with an area of 55×55 μm. Ionising radiation interacting in the sensor creates traces whose shapes and energy losses are characteristic of different types of particle interactions [1].

1.1 Hybrid Pixel Detectors

Hybrid pixel detectors are the foundation of the work presented in this paper. First developed for the particle tracking of the inner vertex of the Large Hadron Collider (LHC), their applica-

tion in many areas of physics quickly became apparent. In 1997, the Medipix Collaboration was formed with the goal of bringing pixel detectors to the field of biomedicine [2]. Over the years, the Medipix Collaboration has overseen the production of many different detectors, from pure photon counting detectors (Medipix, Medipix2, and Medipix3) to detectors where collected charge and/or time of interaction in each pixel is recorded (Timepix and Timepix3 - for illustration see Figure 1.1). This paper will concentrate on Timepix and Timepix3, and their application to experiment. These detectors operate on the same principles as a gas chamber, but with higher sensitivity and lower dead time. Their high time resolution and accurate spatial information allow for the separation and visualisation of individual particles of radiation (Figure 1.2), thus opening the door to the possibility of classifying and tracking individual particles of radiation within a field.
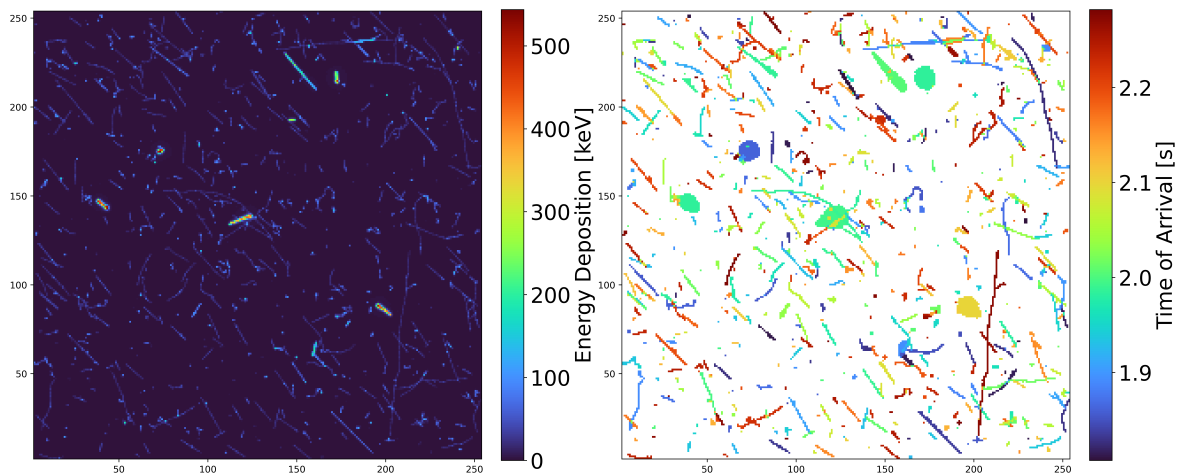

Figure 1.2: Timepix3 detector track visualisation from experimental acquisition at the Large Hadron Collider (MoEDAL experiment), showing energy deposition (left) and time of arrival (right) information of individual particles of radiation.

1.2   Thesis Outline

The goal of this thesis is to develop simulated datasets which compare to experimental results. This should then allow for novel algorithms in particle tracking, classification, and fluence measurements to be developed, with the intention of out-performing current state-of-the-art methods in areas such as accuracy, stability, and/or computational intensity. These algorithms will then be combined to decompose and analyse complex radiation environments.

The rest of the text is divided into 6 main chapters:

- Chapter 2 introduces the Timepix detector family and their working principles.

- Chapter 3 reviews the three main aspects of radiation field decomposition: particle identification, tracking, and fluence measurements. The recent developments in the field and motivations for this paper are outlined.

- Chapter 4 introduces all tools and techniques used throughout this paper.

- Chapter 5 discusses the development and validation of ground truth data by simulation.

- Chapter 6 describes the development of novel algorithms for the decomposition of radiation fields.

- Chapter 7 evaluates the performance of the newly-developed algorithms when applied to real-world experiments (known proton fields at the Danish Centre for Particle Therapy [DCPT] and unknown mixed fields in space and at the LHC).

2 | Timepix Detectors

2.1 Working Principles

Hybrid pixel detectors are a form of ionising radiation detector, consisting of a highly resistive sensor layer (typically silicon but can be other semi-conductor materials) bump bonded to a readout chip. The bump bonding points are referred to as pixels. During operation, a reverse bias voltage is applied at the bump bonding points and the opposite face, creating a fully depleted sensor. The main principle of detection is as follows: as ionising radiation passes through the fully depleted sensor, free charge carriers are created which then begin to drift due to the applied voltage. Electrons drift towards the common backside contact; holes drift towards the pixel electrodes. During this drift motion, current signals are induced in the bump bonding points closest to the charge cloud [3]. The current signal is then converted to a voltage output. When this voltage output crosses a pre-calibrated threshold level (THL), the signal is registered in the readout. The method of signal recording in the readout depends on the individual pixel detector and readout mode used. A basic illustration of this process can be seen in Figure 2.1.
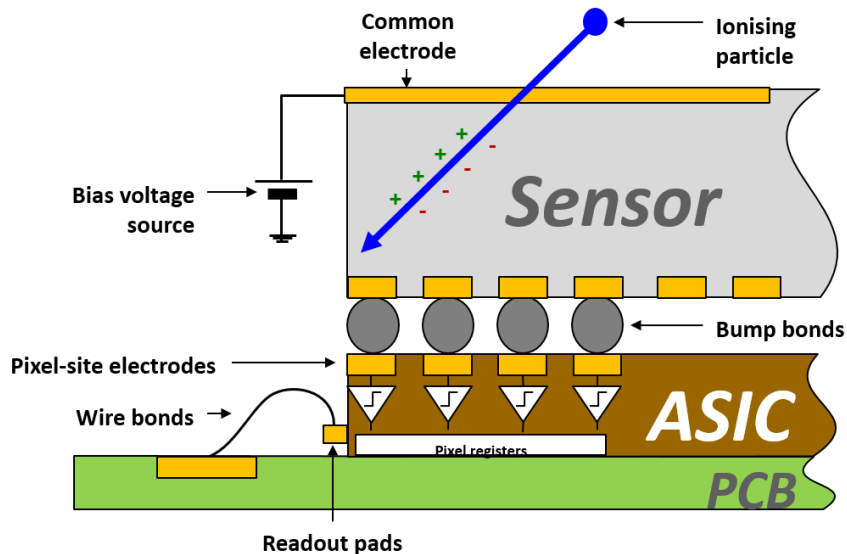
Figure 2.1: Cross section of a hybrid pixel detector. Ionising radiation creates free charge carriers which drift towards the electrodes due to the applied bias voltage. During their drift, they induce currents at the pixel sites, which are amplified and analysed in the pixel electronics. See text for details.

2.2 Timepix

Timepix is a type of hybrid pixel detector developed by the Medipix Collaboration at CERN [4]. The radiation-sensitive layer of these detectors is typically made of silicon with dimensions $1.408 \times 1.408 \times 0.05$ cm. The active material is then bump bonded to a readout chip at 256×256 points with a pitch of 55 μm. During experimental acquisition, Timepix operates in a scheme referred to as frame-based readout mode. This mode operates on a similar concept to that of a photographic camera in which the detector is fully active (the shutter is open) for a set amount of time (referred to as frame acquisition time), after which the detector is made inactive (the shutter is closed). The status of each pixel is then read out, resulting in a dead time of ≈ 11 ms. Based on the spatial information of each pixel, a 2D "image" displaying the status of each pixel can constructed. This image is typically referred to as a frame.

2.2.1 Readout Modes

In Timepix, each pixel is equipped with a 14-bit pseudo-random counter that registers values sampled by an externally fed readout clock, which is typically below 50 MHz. The registered values depend on the readout mode chosen. With this setup, there are three readout modes available, as illustrated in Figure 2.2 and defined below.

1. **Counting mode**: each time the voltage output crosses the THL, the pixel counter is incremented

2. **Time-of-Arrival (ToA) mode**: when the voltage output crosses the THL, the pixel counter begins incrementing with each clock cycle until the end of frame acquisition

3. **Time-over-Threshold (ToT) mode**: the time interval the voltage output remains above the THL is recorded (Time-over-Threshold [ToT]). If this occurs more than once in a single frame acquisition, the ToT values accumulate.
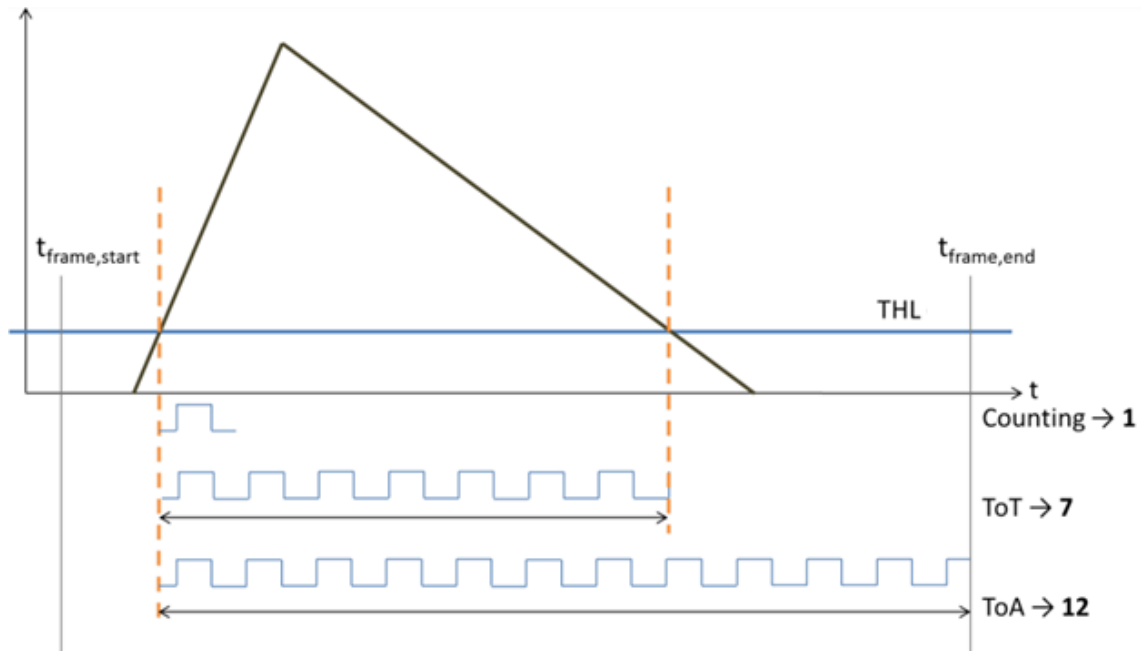
.

Figure 2.2: Illustration of behaviour of the counter register in a Timepix pixel in the various operational modes [5]. THL, threshold level; ToT, Time-over-Threshold; ToA, Time-of-Arrival.

6

2.3   Timepix3

Timepix3 detectors, the successors to Timepix, are another version of hybrid pixel detector, also developed by the Medipix Collaboration at CERN [6]. The active layer of these detectors is also typically made of silicon with dimensions $1.408 \times 1.408 \times 0.05$ cm. Both Timepix and Timepix3 detectors operate on the same physical principles; therefore, data analysis can be treated almost identically.

2.3.1   Readout Modes

In Timepix3, each pixel is equipped with an 18-bit (ToA) and 10-bit (ToT) pseudo-random counter. There is a base clock of 40 MHz (slow ToA [sToA]), which is running continuously throughout the measurement and a 640 MHz clock from a ring oscillator (fast ToA [fToA]) with time resolutions $\approx$25 ns and 1.5625 ns, respectively. The precise time measurement proceeds as follows. In experiment, when the voltage output within a pixel crosses the THL on its upwards slope, the fToA starts to count until the next rising edge of the continuously running sToA clock, as illustrated in Figure 2.3. Considering sTOA is measured from the start of the acquisition time, the final ToA value is given by $t[ns] = sToA \times 25 - fToA \times 1.5625$. The ToT is measured solely with sTOA using the moments of crossing on the upward and downward slopes of the induced voltage signal.

    With these new features, Timepix3 has all the same available readout modes as Timepix, but with higher accuracy and an additional readout scheme, known as data-driven mode. In this mode, it is possible to simultaneously measure the ToA and ToT information of each pixel (quasi-)continuously. Throughout the measurement, only pixels that have seen an event are inactive for the time that event is being read out; this takes approximately 475 ns.
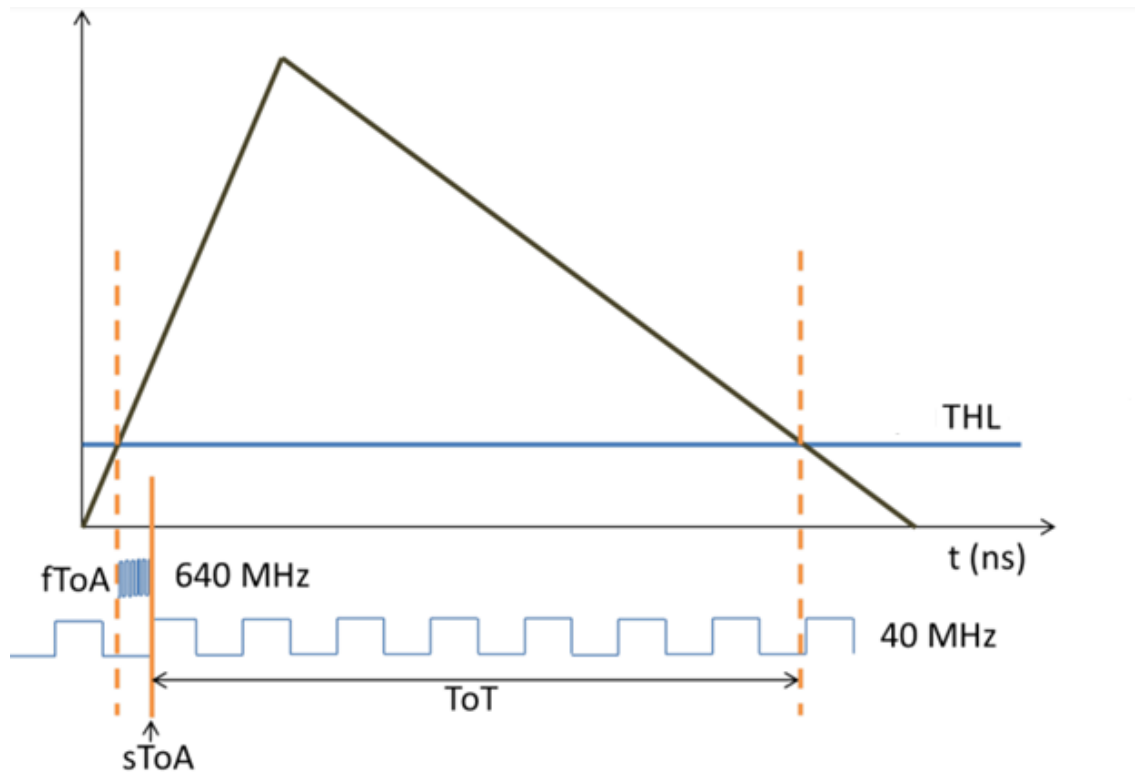
Figure 2.3: Illustration of the pulse processing done in each pixel of a Timepix3 detector. A 40 MHz clock running continuously from the measurement start determines sToA. To increase time precision, a 640 MHz clock from a local oscillator determines the time between the actual threshold level crossing and the next rising edge of the sToA clock [5]. THL, threshold level; t, time; sToA, slow Time-of-Arrival; fToA, fast Time-of-Arrival; ToT, Time-over-Threshold.

2.4 Time-over-Threshold to Energy Deposition

As mentioned previously,ToT is measured during typical experimental acquisition using Timepix detectors. This quantity initially seems somewhat abstract; however, ToT has a particular significance as it can be directly related to the energy deposition of the particle of radiation creating the signal [7]. This relation can be summarised using the following equation (also illustrated in Figure 2.4):

$$f(x) = ax + b + \frac{c}{x - t} [keV] \tag{2.1}$$

8

where $a, b, c, t \in \mathbb{R}$ are obtained for each pixel through calibration using known X-ray fluorescence and γ-photons [8]. The calibration process for Timepix and Timepix3 is identical as both operate on the same pulse-processing concept.
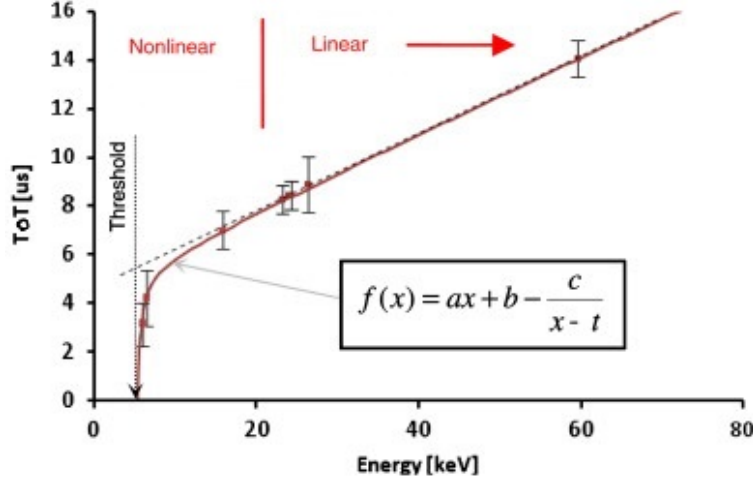

Figure 2.4: The dependence of the ToT value on energy deposition in a Timepix(3) pixel [7]. The parameters a, b, c and t of the surrogate function are determined by per-pixel calibration. ToT, Time-over-Threshold.

2.5 Particle Clusters

During typical data acquisition, a large number of pixel hits is generated which can be grouped together into so-called "particle clusters". In the presented work, it is assumed that each particle cluster corresponds to an individual particle of radiation [9]. In the case of Timepix, this is performed solely using the spatial information; however, both spatial and ToA information are utilised for Timepix3. For example, Figure 2.5 shows a clear separation of cluster 1 from clusters 2 and 3 based on the spatial information. Clusters 2 and 3 are separated by ToA information.
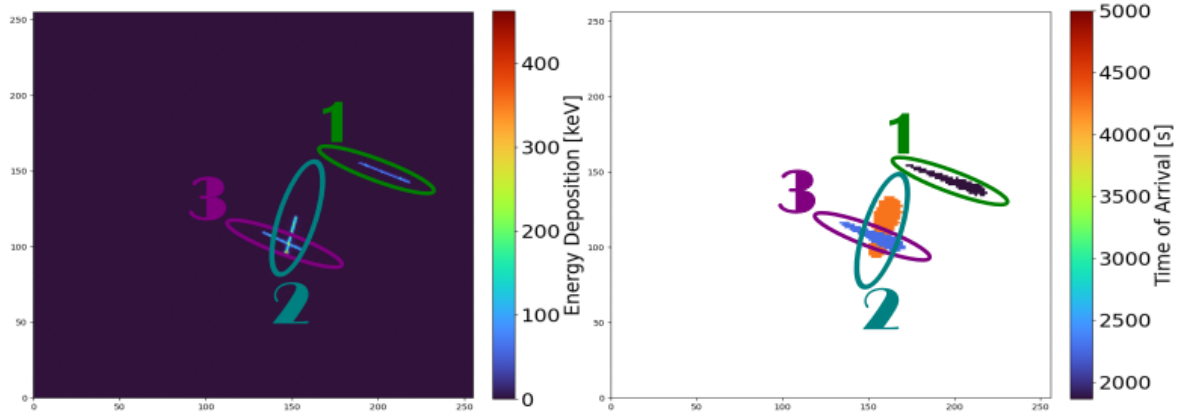
Figure 2.5: Example of the separation of three particle clusters using spatial and ToA information of the pixel hits. Cluster 1 is separated by spatial information and clusters 2 and 3 are separated by ToA information. ToA, Time-of-Arrival.

The reduced separation capabilities in Timepix lead to major problems in analysis if multiple overlapping particles are merged in a single cluster, ultimately resulting in an underestimation of the particle rate and misclassification of the particles within the cluster.

3 | Motivations

The aim of this chapter is to outline the current state-of-the-art methods in radiation field decomposition using Timepix and Timepix3 detectors.

3.1 Particle Classification

Medipix2 silicon detectors were the first hybrid pixel detectors to decompose radiation fields. For example in [10], particles of radiation were classified based solely on their morphological features, leading to the development of the classification scheme illustrated in Figure 3.1.
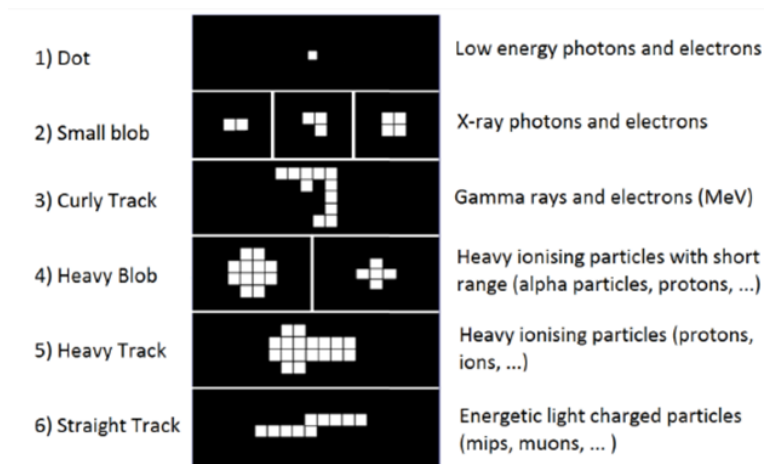

Figure 3.1: Classification scheme first proposed by Holy et al [10].

This classification scheme has continued in popularity even after the introduction of Timepix detectors with the additional ability of measuring the ToT signals. With this new information, attempts to further increase accuracy were made [11]. This classification scheme, although effective, does have its limitations, as it appears somewhat abstract and vague to those not

in the Medipix community. In recent years, there has been a push to move away from this form of classification into more physical groups, namely classes based on particle types and energies. The most recent attempt involved the use of convolution neural networks (CNNs) to classify particles into electron and several proton energy interval classes [12]. With the introduction of Timepix3 and its increased time resolution and simultaneous measurement of energy and time, more information is available. This allows for higher accuracy classification and reduction of systematic uncertainties, for example by reducing the chance of track overlap. To date, there has been little work on this topic related to Timepix3.

3.2   Particle Tracking

Particle tracking in Timepix3 refers to the process of using pixel ToT, ToA, and spatial information for the approximate 3D reconstruction of a particle's trajectory as it passes through the active silicon layer. The capabilities of Timepix3 for 3D reconstruction of particles is well-documented [13, 14]. In these papers, Timepix3 is operated in a similar fashion to a time projection chamber, where it utilises the ToA and spatial information to reconstruct the particle trajectories within the detector with the aid of approximate drift time equations. This new information was quickly applied to determine a particle's initial directionality, and locate the point of interaction in the LHC [15]. The most recent paper released on this topic [16] performed an exhaustive search of the best analytical tracking algorithm, with emphasis put on high energy particles. It will be shown in later sections (6.2) that these algorithms, although effective, have limitations when applied to low energy particles. An accurate particle tracking algorithm over a wide range of particles has countless applications, such as localisation, size measurement, and isolation of particle sources in complex radiation fields. It is also worth noting that an accurate and stable particle tracking algorithm is necessary for the measurement of particle fluence (outlined in the next section).

3.3   Particle Fluence

In simple terms, particle fluence is defined as the number of particles crossing a perpendicular surface per unit area. Unfortunately, this definition is only sufficient for the case of particle beams as there is a clear perpendicular area that can be measured. An accurate measurement of particle fluence that is independent of the directionality of the field will be important to this paper as the classification and tracking of the radiation field has no real significance unless quantified, and biases associated with detector orientation are removed. One approach to this problem was introduced in [15] where the "effective area" method was used. In this method,

measured particles were weighted with the equation

$$\alpha_{weighting} = \frac{1}{l \cdot l \cdot \cos(\theta)} \tag{3.1}$$

where l is the length of the side of the detector (1.408 cm) and θ is the angle of incidence. As will be seen in Section 6.2, this equation, although effective, is not sufficient for the applications of this paper as particles that enter/exit along the edges of the detector (edge clusters) must be removed to ensure accurate tracking and classification algorithms. The development of an accurate particle fluence measurement that is independent of detector orientation would allow for unbiased classification and radiation tracking, and increased accuracy of radiation field monitoring.

4 | Materials and Methods

In this chapter, the frameworks and algorithms used throughout this paper will be introduced.

4.1 Allpix^2

Allpix^2 is a general purpose simulation framework for simulating the performance of hybrid silicon detectors [17]. The software is based on Geant4 libraries [18] and written in C++. It is the standard simulation software for the simulation of hybrid silicon detectors. This framework, unless otherwise stated, will be used for all simulations throughout this paper.

4.1.1 Simulation Configuration

Rather than explicitly writing the simulation set-up in C++, when simulating in Allpix^2 the user need only create a configuration file, which provides a large amount of abstraction to the user. Despite this abstraction, the user still needs a level of understanding of the parameters they are setting.

DespositionGeant4

In this section of the configuration file, the physics list and the radiation sources are specified. The physics list used affects the calculation of the cross-sections and associated interactions, and will vary according to the type of particles being simulated. A guide to all physics lists available can be found at [19]. The source shape is specified by "source_type". In the example below, a sphere is defined. By default, this produces particles randomly along the sphere's surface with momentum direction determined by cosine distribution. This ultimately produces an omnidirectional isotropic field within the spherical volume [20].

```
[DepositionGeant4]
physics_list = FTFP_BERT_LIV
particle_type = "proton"
```

```
source_energy = 100.00MeV
number_of_particles = 1
source_type = "sphere"
source_position = 0 0 0
sphere_radius = 15cm
```

If a particular energy distribution aside from mono-energetic or Gaussian distribution is required, an external macro-file written in standard Geant4 language must be included using the commands "source_type = "macro"" and "file_name = "SourceMacro.mac"".

ElectricFieldReader

This section of the configuration file specifies the exact model used for the electric field. In most cases a linear field is sufficient, but more specific models are available [17]. The bias voltage is adjusted according to the desired detector setup. Provided the silicon is fully depleted, the depletion voltage has no significant effect on results when simulating Timepix or Timepix3.

```
[ElectricFieldReader]
model = "linear"
bias_voltage = 230V
depletion_voltage = 80V
```

Charge Propagation

The charge propagation inside the detector is specified using either "GenericPropagation" or "TransientPropagation". When simulating low ionising particles such as electrons or high energy protons it is typically sufficient to use "GenericPropagation". However, due to the higher amount of free charge created in the detector while simulating high ionising particles such as low energy protons, there is a large amount of diffusion and charge sharing between pixels. These phenomena cannot be modelled in "GenericPropagation", calling for the need of "TransientPropagation". In this scheme, propagation is still calculated but at each time-step the associated diffusion is also accounted for. Both schemes use Runge-Kutta integration to calculate the charge propagation. Whenever possible the generic propagation is favoured over transient due to its dramatically lower computational intensity.

```
[GenericPropagation]
name = "detector"
temperature = 315K
charge_per_step = 25
```

15

```
integration_time = 50ns
propagate_holes = true
propagate_electrons = false

[TransientPropagation]
name = "detector"
temperature = 315K
charge_per_step = 200
timestep = 1ns
integration_time = 40ns
induction_matrix = 3 3
```

CSADigitizer

This section of the configuration file is where the data readout method is specified, i.e., how the current is induced at the pixel electrodes and converted into a pixel hit. Most of these physical parameters can remain the same across a large range of simulations; fine-tuning is only necessary if simulation errors arise. However, the following parameters must always be checked and adjusted according to simulation: "integration_time", "threshold", and "clock_bin". To avoid redundant computations, "integration_time" is typically reduced for lower energy deposition particles because the voltage output produced in bump bonding points is much shorter than that produced for higher energy deposition particles. The threshold value must be calibrated for each simulation setup. Appendix A details threshold calibration along with ToT energy calibration. The clock bin is typically set to 1.525 ns (the resolution of Timepix3), but must be adjusted according to the detector being simulated.

```
[CSADigitizer]
model = "csa"
krummenacher_current = 1e-9C/s
detector_capacitance = 50e-15C/V
amp_output_capacitance = 20e-15C/V
transconductance = 50e-6C/s/V
feedback_capacitance = 10e-15C/V
temperature = 315K
integration_time = 15000ns
threshold = 15e-3V
sigma_noise = 0.1e-3V
clock_bin_toa = 1.5625ns
clock_bin_tot = 1.5625ns
```

4.2   Bayesian Deconvolution

Spectrum deconvolution refers to the decomposition of a complex signal into its contributing spectrum components. There are many different iterative and statistical schemes that can be chosen for this process. For this paper, the Bayesian deconvolution scheme was been chosen [21], which utilises the Bayesian probability formula:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{4.1}$$

where p represents a generic probability function, $|$ the given operator, and A and B some arbitrary variables or system states. Despite the simplicity of the Bayesian formula it is quite powerfully and used in many areas of physics and statistics [22]. The formula conveys the probability of A having a particular value or state given that B has a particular value or state, essentially relating two otherwise unrelated states. The states A and B can be assigned some arbitrary distribution of two variables that will be referred to as the cause vector (x_C) and the effect vector (x_E). It can then be assumed there exists an arbitrary probability distribution $p(x_E|x_C)$ given by the formula

$$p(x_c|x_E) = \frac{p(x_E|x_C)p(x_C)}{p(x_E)}. \tag{4.2}$$

The approximate values of $p(x_E|x_C)$ can be achieved through simulation. The remaining probability values for a specific experiment can be obtained through the Bayesian iterative deconvolution algorithm that is implemented using the library [23] written in C++.

4.3   Machine Learning

Machine-learning (ML) algorithms are arbitrary algorithms which are optimised to reduce a prediction error by "learning" from a given dataset without being explicitly programmed. The numbers of such algorithms are countless; thus, to reduce the scope of this paper, only a subset of these algorithms will be considered. The ones of utmost importance to this project are as follows:

Neural network   This is a type of ML algorithm that operates on the same principle as neurons in the brain. Neurons of the neural network are arranged most commonly in layers known as dense layers in which each neuron of one layer is fully connected to every neuron of the previous layer. The output of a neuron is then calculated using a weighted sum of the

previous layer with additional biases. The weights and biases associated with each neuron are determined by learning from training data. On each neuron there is typically an addition of non-linear activation functions (sigmoid, ReLU, etc.) that switch the corresponding neurons on and off. This adds an additional non-linear capability of the solution [24].

Convolution neural networks (CNN) This a specific form of neural network that is specialised for computer vision tasks with an addition of convolution layers. Convolution layers apply convolution operations using kernel matrices on the input structures (typically 2D/3D matrices) to compute the output. The elements of the kernel matrices are determined through learning from training data. Convolution operations have an advantage over total weighted sum operations in dense layers as they emphasis the relationship between neighbouring neurons. CNNs are typically highly labour intensive, particularly when applied to Timepix3 data, as they require the construction of fixed size image inputs [24].

Random Forest To construct this algorithm, multiple decision trees are constructed independently of one another using unique bootstrapped datasets (reduced dataset by random sampling) for each tree. A random sample of features at each decision node is used. Once the trees are constructed, the algorithm makes a decision based on the average of all decision trees. The use of multiple tree predictions, as opposed to a greedy algorithm, allows for an optimal solution, thus making it more robust [25].

Gradient boosted algorithms These algorithms once again use an ensemble of decision trees, but this time the construction of each tree is affected by its predecessors. Each new tree is built to improve the errors of the previous trees; this concept is known as boosting. Each construction is affected by its predecessors by calculating weights determined by the value of their lost contribution for each data point of the training data at the beginning of each construction. These weights are then used either to make a weighted bootstrapped dataset or to use weighted loss functions. Gradient boosted algorithms typically outperform Random Forrest algorithms; however, they are more sensitive to noise in the dataset [26].

XGBoosted algorithm This algorithm follows the same principles as gradient boosted algorithms but with additional L1 and L2 regularisation. This reduces the chance of over-fitting of the solution to the training set by reducing the complexity of the algorithm if the corresponding gain is not sufficient [26].

Implementation of these algorithms in python with documentation can be found at the following library websites: Keras [27] and Scikit-learn [28].

As outlined in Section 3.1, CNNs are the current state-of-the-art method in classification. However, CNNs require the construction of an individual image for every incident particle before passing said images through the CNN model. In most cases, there are many redundant variable calculations due to the varying size of particle clusters (anywhere from 1 pixel to 256 pixels in any dimension) and the requirement for a fixed-sized input into the CNN. These image constructions and the CNN calculation itself contribute to CNNs being very computationally intensive algorithms in terms of both memory and time. This computational intensity thus limits their application to fields of low flux and/or in cases with high computational resources. The algorithms being developed in the presented work are intended fast data processing of high flux field, and therefore, the development of a new ML algorithm with manual feature calculation will be beneficial.

4.4 Cluster Features

As shown in Section 2.5, it is possible to separate individual particles of radiation into particle clusters that can then be visualised as images. Thus, the next step was to determine if particle clusters have a characteristic structure dependent upon the particle type and primary energy. Figures 4.1 and 4.2 depict a basic illustration of the change in characteristic structure with increasing primary energy for electrons and protons, respectively.
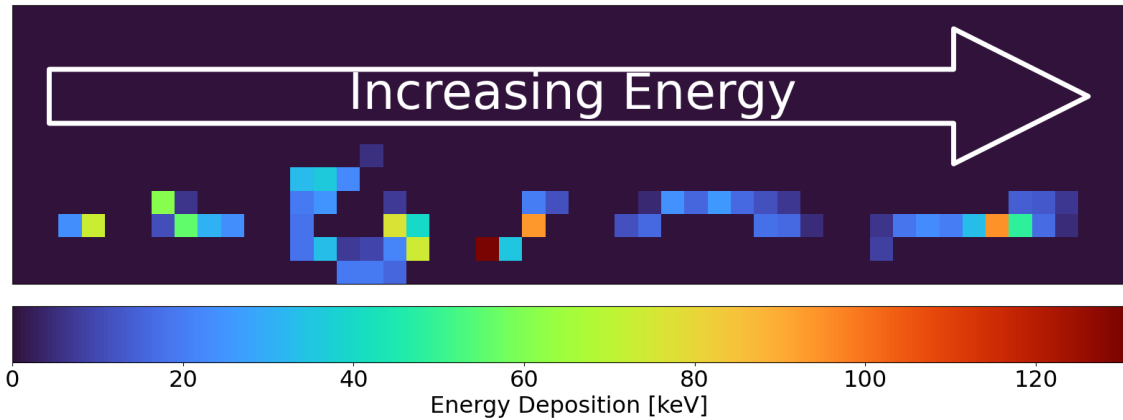

Figure 4.1: Basic illustration of typical electron tracks in Timepix3 and how they change with increasing primary energy from $0 \ MeV$ to $10.0 \ MeV$ at a $45°$ angle of incidence.
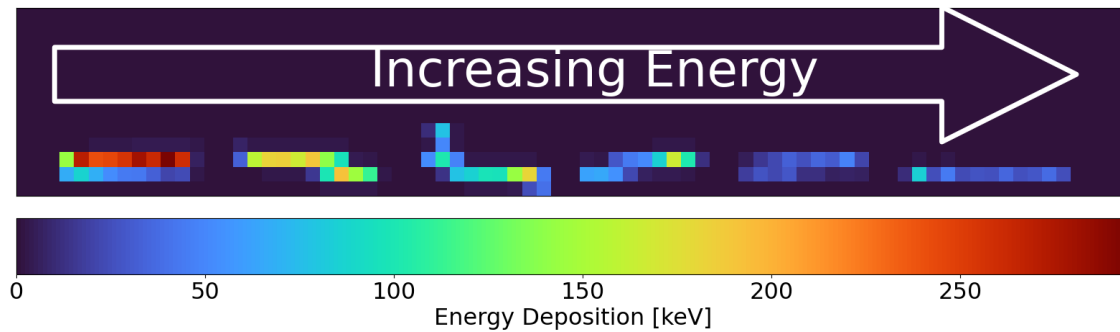
Figure 4.2: Basic illustration of typical proton tracks in Timepix3 and how they change with increasing primary energy from 10 MeV to 500 MeV at a $45°$ angle of incidence.

As shown in Figure 4.1, electrons tend to be thin curly tracks increasing in linearity with increasing primary energy. On the contrary, Figure 4.2 shows protons are always linear, with the occasional delta ray. A delta ray is a secondary track resulting from the primary due to electron scattering and a surrounding electron cloud caused by the diffusion of charge carriers in the detector. The size of the electron cloud decreases with increasing primary energy. In this regard, a manual search and calculation of a large number of features which best quantify these behaviours was performed. The features found to best describe these characteristics were as follows:

- **Height**: The energy deposition value of the maximum energy pixel within any one cluster.

- **Size**: The number of pixels present in a particle cluster.

- **Energy deposition**: The sum of the energy deposition from all pixels within a cluster.

- **Standard deviation (std) of energy**: The standard deviation of energy deposition recorded in the pixels.

- **$\Delta(\mathbf{ToA})$**: The difference in ToA value between the maximum ToA pixel and the minimum ToA pixel.

- **Chi**: The average perpendicular distance of the pixels from the energy-weighted line fit of the cluster. This average can be weighted or unweighted.

- **Std of chi**: The std of the perpendicular distance of the pixels from the energy-weighted line fit of the cluster. This can be weighted or unweighted.

20

- **Linearity**: The number of pixels within the cluster that the energy-weighted line fit crosses over.

- **Length**: The largest distance between pixels in the cluster.

- **Stopping power** ($\frac{dE}{dx}$): The energy deposition per unit distance travelled by the particle in the sensor layer. Calculated by the equation

$$\frac{dE}{dx} = \frac{E_{dep}\cos\theta}{d\rho} \tag{4.3}$$

  where E_{dep} is the total energy deposition, θ is the approximate angle of incidence of the particle, d is the thickness of the detector (typically 0.05 cm), and ρ is the density of the sensor layer (2.33 $g\ cm^{-3}$ for silicon).

- **Std along line**: The std of the positions of the pixels when projected to the weighted line fit. This std can be weighted or unweighted.

- **Box dimensions**: The dimension of the sides of a quadrilateral drawn around the cluster. This can be the maximum or minimum dimension.

- **ToA gradient**: The slope of the 2D line fit of the reconstructed coordinate based on ToA and drift time equations [13] against their distance from the energy-weighted average pixel position. The 3D reconstruction is normalised to the detector thickness to ensure consistency.

- **Average number of nearest neighbours 4-fold**: The average number of nearest neighbouring pixels which are directly vertical or horizontal to each pixel in the cluster. Two pixels are considered nearest neighbours if they are touching.

- **Average number of nearest neighbours 8-fold**: The average number of nearest neighbouring pixels which are directly vertical, horizontal, and/or diagonal to each pixel in the cluster.

- **Edge to inner ratio**: The ratio of the number of pixels along the edge of the cluster to the number of pixels fully surrounded by neighbouring pixels.

It is worth noting there can be a large discrepancy between experiment and simulation for some features, for example, ΔToA. To avoid a large discrepancy in later applications, only the features that are shown to agree with experiment will be used in the final algorithms.

4.5 Machine Learning Feature Selection

Aside from the CNN, all other ML algorithms outlined require manual feature selection and calculation. It is possible calculate all the features discussed in the previous section for every particle cluster and use that as an input. However, aside from being very computationally intensive, this method may negatively effect the accuracy of the algorithm resulting from problems due to high dimensionality [22]. Therefore, at several points throughout this paper, feature selection will be performed. To determine which features are optimal for a particular ML algorithm, the following search algorithm was implemented.

1. All possible features for the training and validation set were calculated.

2. Multiple instances of the same ML algorithm were then trained and tested, with one feature removed from the input data at each instance.

3. The testing accuracies of all models from step 2 were then compared, and the model with the highest accuracy was selected.

4. The feature missing from this model was permanently removed from the dataset.

5. Steps 2-3 were repeated until no more features were present.

6. The accuracy of every selected model in step 3 was graphed.

7. The position before a rapid drop in accuracy was selected and the features at this point were assumed to be optimal.

This is clearly a very labour-intensive algorithm but it is worth noting that it only needs to be performed once to determine the optimal algorithm.

5 | Dataset Creation

5.1 Simulation of Omnidirectional Fields

A basic diagram of the simulation setup used for data generation can be seen in Figure 5.1. A Timepix3 detector is placed in the centre of a 10 cm radius spherical source. This spherical source produces particles randomly along its surface with the initial momentum direction chosen according to a cosine probability which produces an omnidirectional isotropic field within the spherical volume [20].
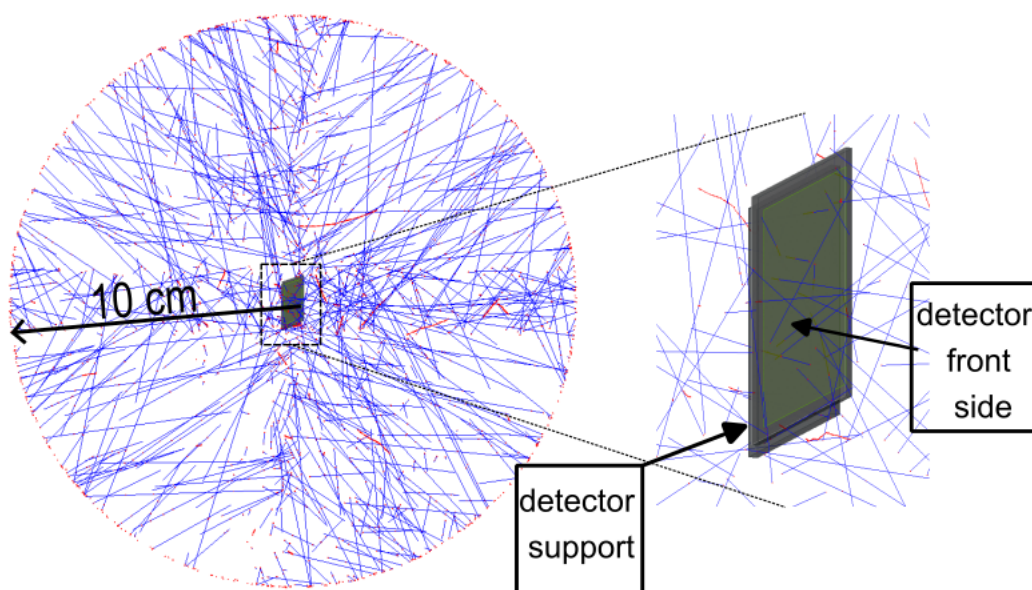

Figure 5.1: Basic illustration of the simulation setup used for dataset simulation and algorithm testing. The dimensions of the detector are omitted as the standard Timepix3 dimensions were used.

5.1.1 Variable Distributions

It might be thought that, because the simulation is performed in an omnidirectional field, all variables affecting the detector would be uniformly distributed, however, this would only be the case if Timepix3 was spherical. In reality, there are many variables which are biased due to the planar shape of Timepix3; only those relevant to this paper will be outlined and analysed. Specifically, this paper will cover the distributions of the following variables: angle of incidence to the normal (θ), azimuthal angle (φ), and path length within the active volume (L_{eff}). This analysis was performed using the simulation framework outlined in Section 5.1 using arbitrary high energy particles, and by recording the relevant variables of each particle incident on the detector directly from the simulation information. These results were then graphed into histograms as seen in Figure 5.2, which shows the variable φ was uniformly distributed whilst θ and L_{eff} show quite a large bias.
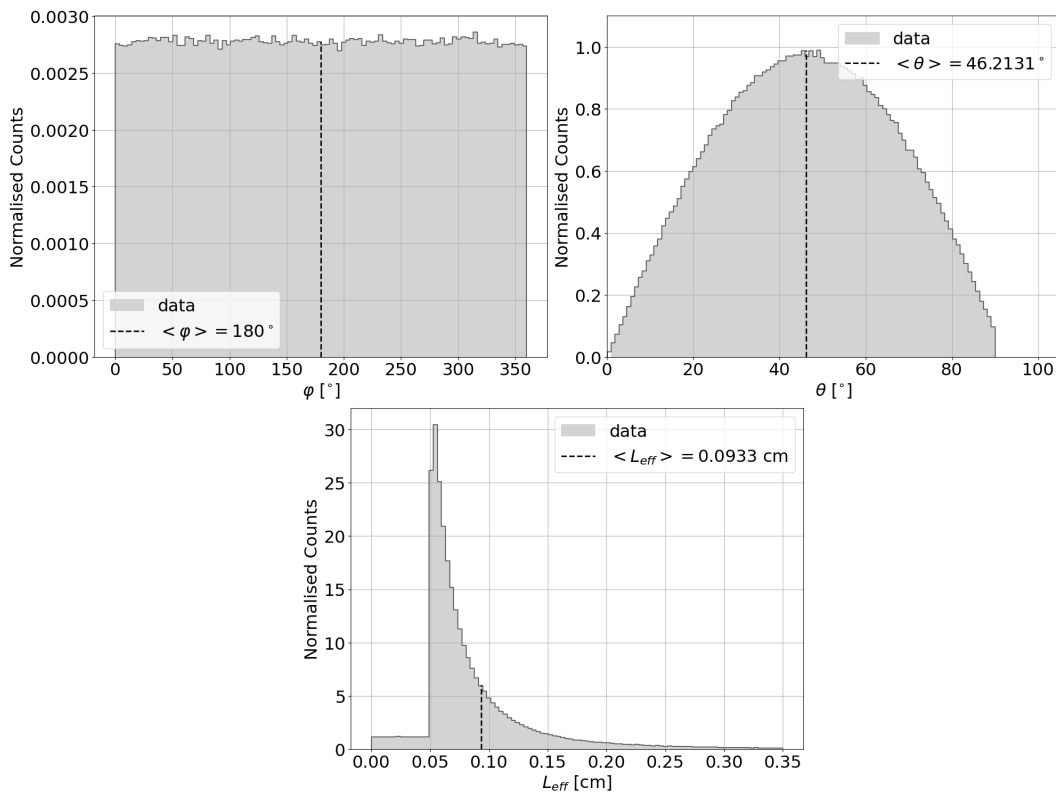
Figure 5.2: Graphs of the distributions of the angle of incidence to the normal (θ), azimuthal angle (φ), and effective path length within the active volume (L_{eff}), as measured from particles incident on a Timepix3 detector in an omnidirectional field simulation.

5.2 Construction

All training, validation, and testing datasets present in this paper were simulated in Allpix[2] using a similar omnidirectional field simulation setup as mentioned in Section 5.1. For each dataset, only the particle type and primary energy was changed. The training and validation datasets were constructed using two separate simulations of electrons and protons with flat primary energy spectrums from 0 to 10 MeV and 0 to 500 MeV, respectively. The datasets were constructed at equal sizes of approximately 250,000 particles to reduce training biases in later sections. Approximately 30% of both datasets was reserved for testing algorithms. Similarly, multiple mono-energetic fields were simulated for testing. Table 5.1 outlines an overview of the simulated datasets used for the comparison of algorithms and determination of the algorithm response to different particle types and energies.

Particle type	Primary energy (MeV)	Fluence (cm^{-2})
Protons	10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 130, 150, 200, 300, 400, 450, 500, 730, 850, 10000	31,831
Electrons	0.10, 0.25, 0.50, 0.75, 1.00, 1.50, 2.00, 3.00, 4.00, 5.00, 10.00, 100.00	31,831
Pions	10, 15, 25, 30, 40, 50, 75, 100, 150, 200	31,831
Muons	10, 15, 25, 30, 40, 50, 75, 100, 150, 200	31,831
Gammas	0.05, 0.10, 0.15, 0.20, 0.25, 0.50, 1.00, 1.5, 2.00, 2.50, 5.0, 10.0, 25.0, 50.0, 100.0, 200.0, 350.0	318,000,000

Table 5.1: Table of the mono-energetic fields that were simulated for model development. All energies were simulated separately.

5.3 Flat Incident Angle Distribution

As shown in Section 5.1.1, when simulating particles in an omnidirectional field environment a flat θ distribution is not produced. This will later cause issues when determining the optimal particle tracking algorithm in Section 6.2 and training bias when developing an ML particle tracking algorithm [24]. To counteract this problem, a reduced dataset with an approximately flat θ must be constructed. To achieve this, a weighted random sampling given by the formula

$$\alpha(\theta) = \frac{1}{V_{eff}(\theta)} \cdot \frac{1}{\sin(\theta)} \tag{5.1}$$

was used. The value of $V_{eff}(\theta)$ and its underlining theory will be outlined in Section 6.1. The sampled datasets were then made sufficiently small such that an approximately flat spectrum was achieved. In Figure 5.3 it can be seen that a sample size of 30% is sufficient to achieve an approximately flat θ distribution. Such a distribution gives approximately 150,000 particles for development of particle tracking algorithms.
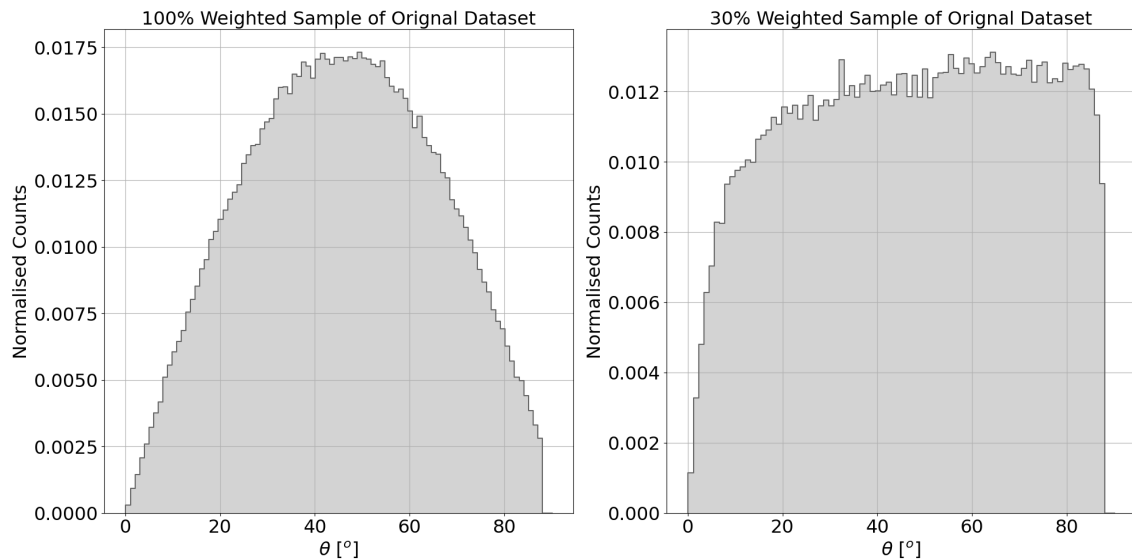

Figure 5.3: Demonstration of the affect of fluence weighted sampling on θ distribution of a simulated omnidirectional field dataset.

5.4 Simulation Consistency Tests

All training, validation, and testing steps of the methods developed throughout this paper use simulated datasets; therefore, it is important to ensure that the simulations are consistent with experimental data. For this part of the analysis only stopping power, energy deposition, cluster height, and $\Delta(ToA)$ checks were performed. An approximation of the experimental setup during a 125.51 MeV proton beam exposure at the DCPT in Aarhus, Denmark was chosen for simulation. Results can be seen in Figure 5.4, which illustrates there was reasonable agreement between experiment and simulation.
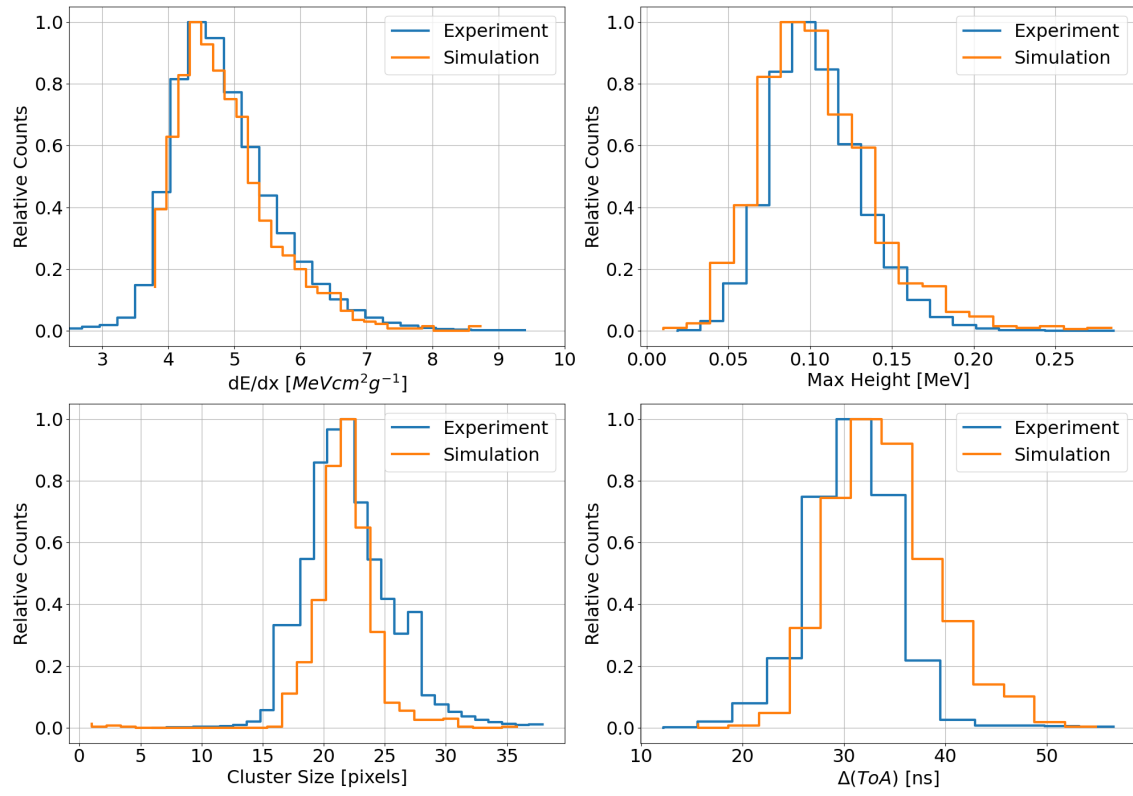
Figure 5.4: Demonstration of simulation consistency by comparing major cluster features with experimental results from the 125.51 MeV proton beam exposure at the Danish Centre for Particle Therapy in Aarhus, Denmark.

6 | Algorithm Development

In this chapter, novel algorithms will be developed and their performance will be compared with the state-of-the-art algorithms outlined in the previous section.

6.1   Particle Fluence

The naive approach to calculating particle fluence using Timepix3 would be to sum the approximate track lengths (outlined in Section 6.2) and divide this by the active silicon volume of Timepix3 (1.408 x 1.408 x 0.05 cm). Although the full volume of the detector is exposed to the radiation field, a portion of this volume is unable to detect particles between some given interval of incidence angles. This is due to the removal of edge clusters; however, removal is required for accurate calculation of a particle's impact angle and energy deposition. This point is made more clear in Figure 6.1.
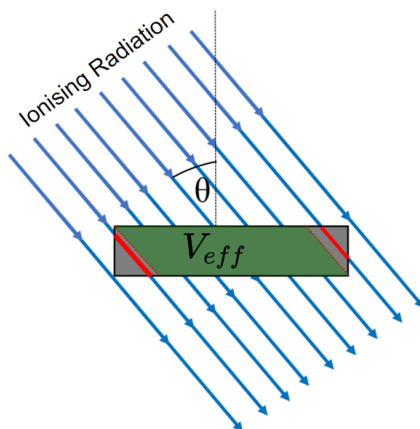
Figure 6.1: Schematic illustration of the effective volume (green) able to detect particles at an angle of incidence θ. Red-colour traces indicate particles which are lost when edge clusters are removed.

At an arbitrary angle of incidence θ, the effective volume of the detector (the volume capable of detecting radiation) was calculated as given by a parallelepiped of volume:

$$V_{eff} = l \cdot d \cdot (l - d \cdot tan(\theta)) \tag{6.1}$$

where l denotes the length of the face of the detector (1.408 cm), and d denotes the thickness of the detector (0.05 cm). It can then be presumed that the particle follows a straight trajectory through the active layer of the detector - a fair approximation in most cases - giving the final approximate fluence contribution of a particle of radiation incident on the detector as follows:

$$\alpha_{fluence} = \frac{1}{l \cdot d \cdot (l - d \cdot \tan(\theta))} \cdot \frac{d}{\cos(\theta)}. \tag{6.2}$$

This fluence approximation was then tested in an omnidirectional field simulation of known fluence of arbitrary high energy particles (see Section 5.1). The fluence contribution of each incident particle on the detector that did not enter along the edge of the detector was calculated from the simulation information and summed to give the total fluence prediction. This simulation was done multiple times with varying particle fluence; the result is depicted in Figure 6.2.
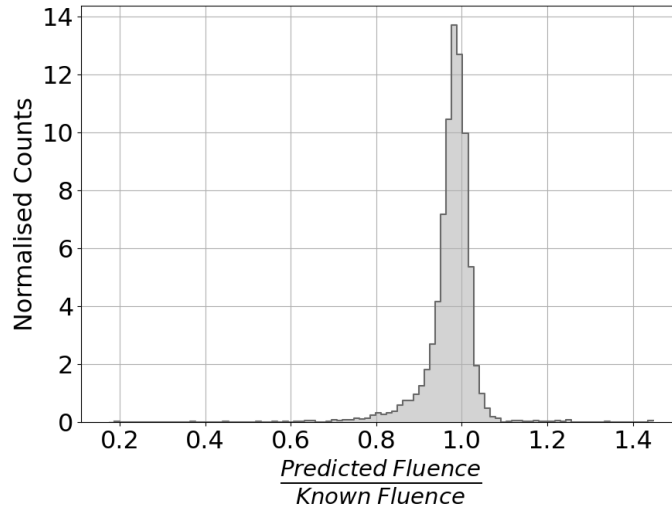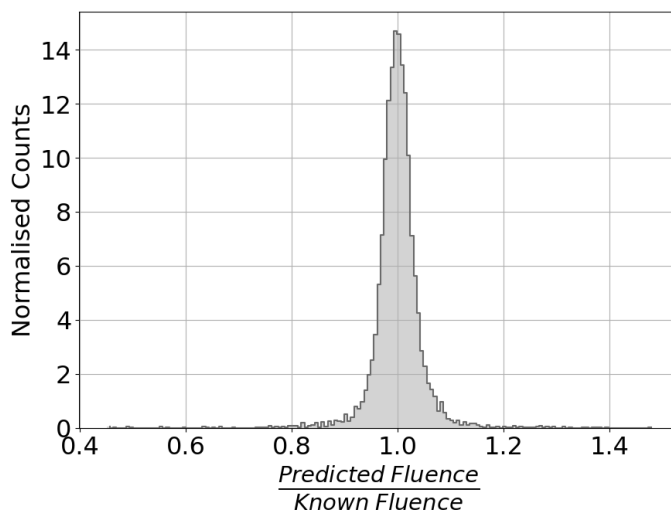

Figure 6.2: Histogram of the ratio of measured particle fluence (determined using the effective volume approximation for a planar geometry detector) to the known simulated field fluence obtained through simulation information. The fluence was calculated by dividing the approximate track length by effective volume approximation (Equation 6.2).

Figure 6.2 illustrates that this fluence approximation method underestimates the field fluence by a mean value of 97%. This discrepancy is due to the formula neglecting the effect

of the dependence of φ on the effective volume. The full formula for the effective volume is actually given by

$$V_{eff} = d \cdot (l - |d \cdot tan(\theta) \cdot sin(\varphi)|) \cdot (l - |d \cdot tan(\theta) \cdot cos(\varphi)|) \qquad (6.3)$$

where l is the length of the edge of the Timepix3 detector (1.408 cm), d is the thickness of the detector (typically 0.05 cm), θ is the angle of incidence, and φ is the azimuthal angle (this approximation is illustrated in Figure 6.3).

Figure 6.3: Histogram of the ratio of measured particle fluence (determined using the effective volume approximation for a planar geometry detector) to the known simulated field fluence obtained through simulation information. The fluence was calculated by dividing the approximate track length ($d/cos(\theta_i)$) by the effective volume approximation with inclusion of φ (Equation 6.3).

This formula is not considered in this paper due to the relatively large error which is associated with the approximation of φ [16]. For all results present in this project, equation 6.1 is used. Thus, the particle fluence contribution of each individual particle is given by its approximate track length divided by its corresponding effective volume.

6.1.1 Weighted Particle Tracking

The main advantage of having a unique particle fluence contribution at every angle of incidence is that it removes any angular bias from the dataset. This is especially prevalent when the goal is to track the majority direction of particles within a field. To demonstrate

the importance of particle fluence weighting on particle tracking, the simulation in Section 6.1 with arbitrary high energy particles was performed. Each time a particle was incident on the detector, the θ angle was recorded in two separate histograms: one with effective volume weighting, and one without. Figure 6.4 depicts the results of this simulation, illustrating that all bias associated with the angle of incidence has been effectively removed and a flat spectrum is achieved when the particle hits are weighted with particle fluence. It is for this reason that all histograms present throughout this paper will be weighted with their approximate particle fluence value.
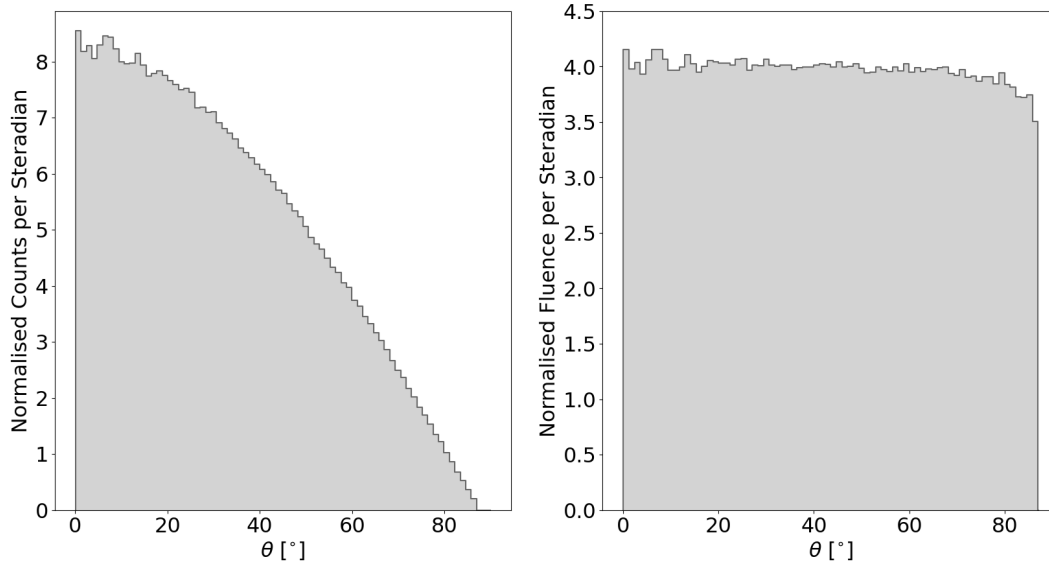
Figure 6.4: Demonstration of the effect of particle fluence weighting on the recorded θ distribution in an omnidirectional field simulation.

6.2 Analytical Particle Tracking

There are many algorithms that can be used for particle tracking in Timepix3 [16], including, but not limited to, the following:

- **Left Lower-most, Right Upper-most (LLM)**: This method assumes the particle follows a straight trajectory. The left lower-most and right upper-most represent the entry and exit points. The θ value is then calculated using trigonometric functions.

- **LLM-improved**: This method is an improved version of the LLM algorithm, as it prioritises horizontal or vertical position when searching for end points depending on

31

the cluster's majority orientation. The majority orientation of the cluster is determined through a weighted line fit.

- **Time Min-Max**: This method once again assumes a straight trajectory but also assumes that the min and max ToA pixels represent the entry and exit points of the particle.

- **Line-fit**: This method reconstructs the particle's 3D trajectory using ToA and approximate drift time equations [13]. Subsequently, the particle's path is determined by finding the major axis using eigenvalue operations.

- **Time gradient descent**: Once again, the particle's 3D trajectory is reconstructed using ToA and approximate drift time equation [13]. However, the track is normalised to the detector thickness which increases robustness of the algorithm. The particle's path is then approximated using a basic 2D line fit of the distance from mean position and reconstructed drift time position.

In the paper by Manek et al. [16], a somewhat idealised dataset of majority 1 GeV pions was used in testing. Thus, provided all particle clusters from the studied dataset that do not obey certain criteria are removed, the results in the paper can be presumed valid. The algorithms developed in the present work are intended for study of entire radiation fields without applying radiation cuts (aside from edge cluster removal); therefore, the accuracy of the results in [16] can no longer be assumed. The algorithms developed in [16] were then applied to the proton-electron dataset developed in this paper (results seen in Figure 6.5). This figure shows that the best-performing algorithm for this dataset was the LLM-improved algorithm, with a mean absolute error of $14.02°$. It is worth noting that although this algorithm achieves the best results of all the tested algorithms, it still performs very poorly on low energy particles, indicating the need for further development.
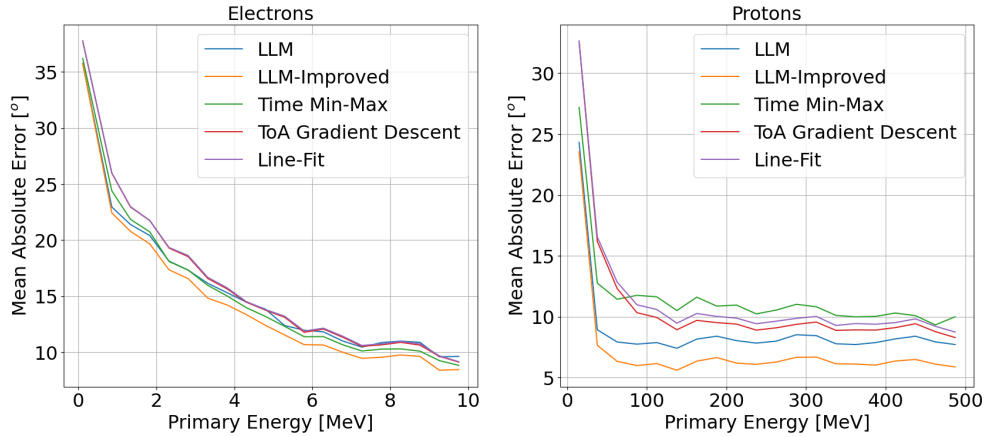
Figure 6.5: Mean absolute error of analytic algorithms for the approximation of θ as seen in [16] as a function of particle primary energy for electrons and protons. LLM, Left Lower-most, Right Upper-most; ToA, Time-of-Arrival.

To ensure there are no significant systematic errors associated with different angles of incidence, a 2D histogram of predicted angles of incidence versus the true angle of incidence of the testing dataset was constructed (Figure 6.6). A clear binning issue at low angles of incidence is visible in Figure 6.6, due to the pixel pitch of 55 μm creating a limitation in accuracy.
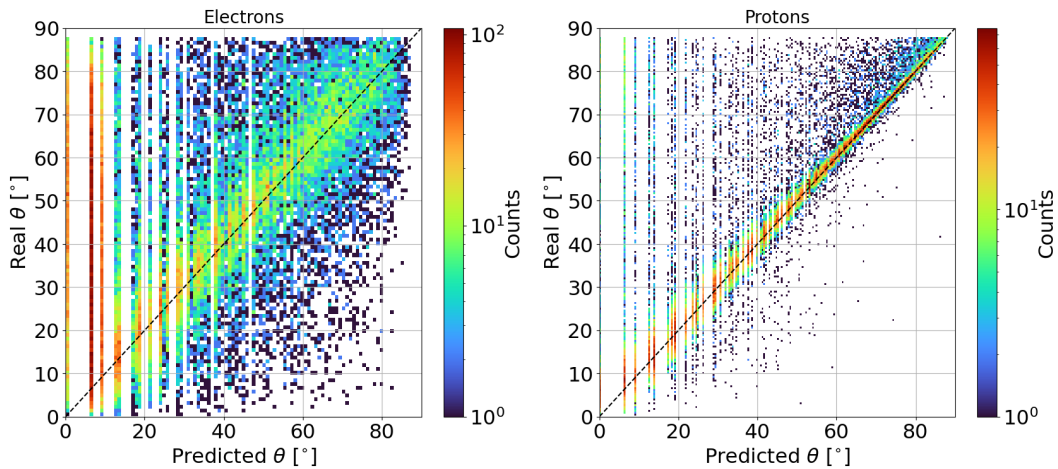
Figure 6.6: Two-dimensional histogram of predicted angle of incidence vs. true angle of incidence as predicted by the Left Lower-most, Right Upper-most-Improved algorithm [16] for electrons and protons.

33

6.3 Deconvolution based Particle Classification

As outlined in Section 4.2, in order to use the Bayesian deconvolution algorithm, a true spectrum (x_C) and an observed spectrum (x_E) must be defined. For the true spectrum, N_e electron primary energy bins and N_p proton primary energy bins from 0 to 10 MeV and 0 to 500 MeV, respectively, are first defined. These two vectors are then concatenated together to form the x_C vector. The observed spectrum x_E is then defined as the stopping power histogram weighted with particle fluence. The LLM-improved algorithm was chosen for the calculation of the θ value required for particle fluence as the errors associated with this algorithm should be modelled effectively by the deconvolution step using corresponding detection efficiencies. Stopping power was chosen for this algorithm as it is a well-researched feature for many different materials and particles [29]. In using such a well-documented feature it becomes easier to detect errors during development and application. From the newly defined x_C and x_E, each row of the response matrix $p(x_C|x_E)$ can be obtained from simulation. As an example, to obtain the i-th row of the matrix, the stopping power histogram for all data in the dataset with primary energy and particle type corresponding to the values defined by bin i are calculated and the is histogram normalised to sum up to 1. By converting the values of the histogram into a vector, the i-th row is then obtained. The results of such a calculation when applied to the particle type and energy bins defined in the dataset developed in Section 5.2 is illustrated in Figure 6.7. From this figure, it is worth noting that all electrons $\geq 1\ MeV$ are degenerate and protons are asymptotically electron-like with increasing energy.
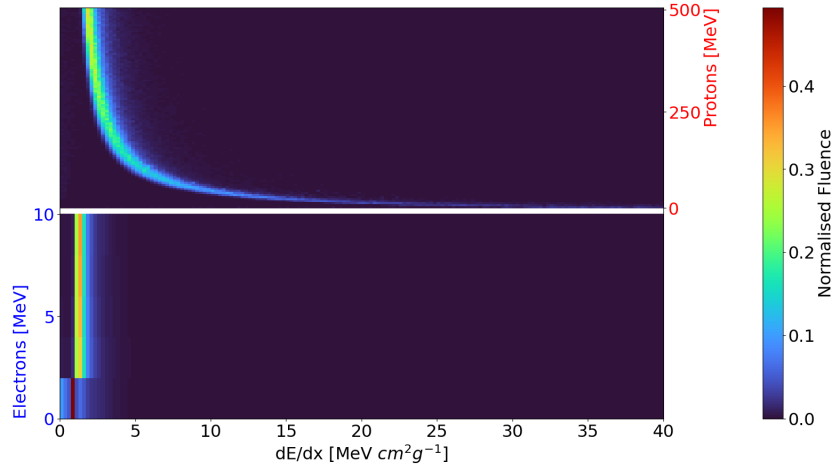
Figure 6.7: Graphical visualisation of the response matrix ($p(x_C|x_E)$) used for the deconvolution classification algorithm obtained through simulation.

Using the probability distribution $p(x_C|x_E)$ and the measured x_E curve for some arbitrary

experiment, the Bayesian deconvolution algorithm can then be applied to approximate x_C for any Timepix3 experiment. x_C provides an approximation of the classified particle fluence [21]. The effectiveness of this classification method was tested using the simulated mono-energetic omnidirectional proton and electron fields from Section 5.2. First, the deconvolution algorithm was applied to the datasets. The accuracy of this method in terms of reconstructing the particle type and energy distribution was observed, as illustrated in Figures 6.8 and 6.9.
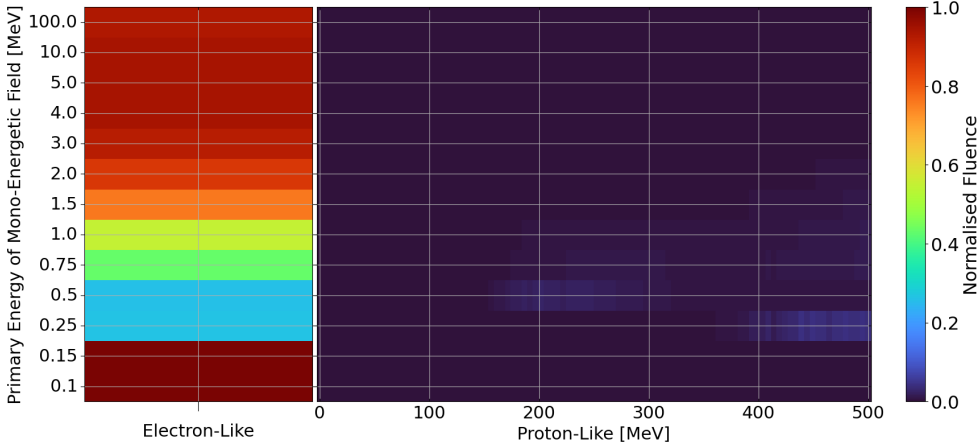
Figure 6.8: Predicted particle spectrum obtained using the deconvolution algorithm for different mono-energetic omnidirectional electron fields.
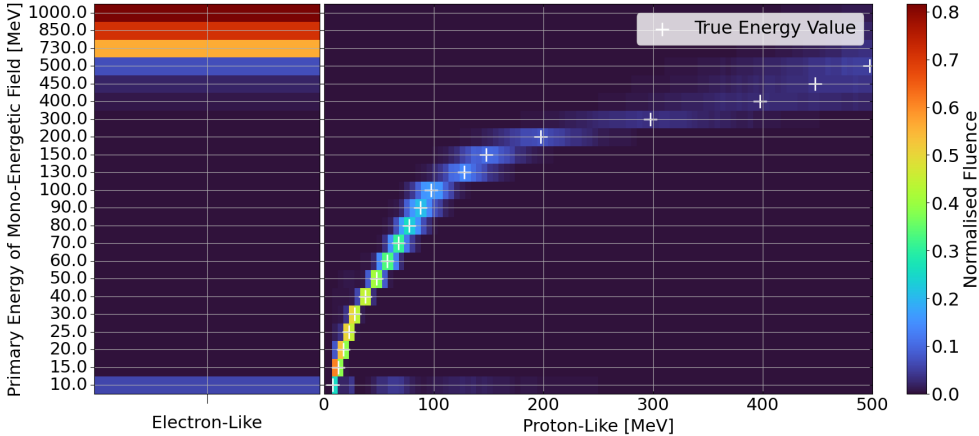
Figure 6.9: Predicted particle spectrum obtained using the deconvolution algorithm for different mono-energetic omnidirectional proton fields.

35

Second, to approximate the accuracy of the algorithm in terms of classifying the electrons and protons, the sum of the measured fluence response associated with the true particle class divided by the sum of the total recorded fluence is calculated, giving the percentage classification accuracy. The results of this can be seen in Figure 6.10. The figure shows electrons with primary energy between 0.5-1.0 MeV are significantly miss-classified. The average of classification across the spectrums was found to be 99.11% for protons and 88.00% for electrons.
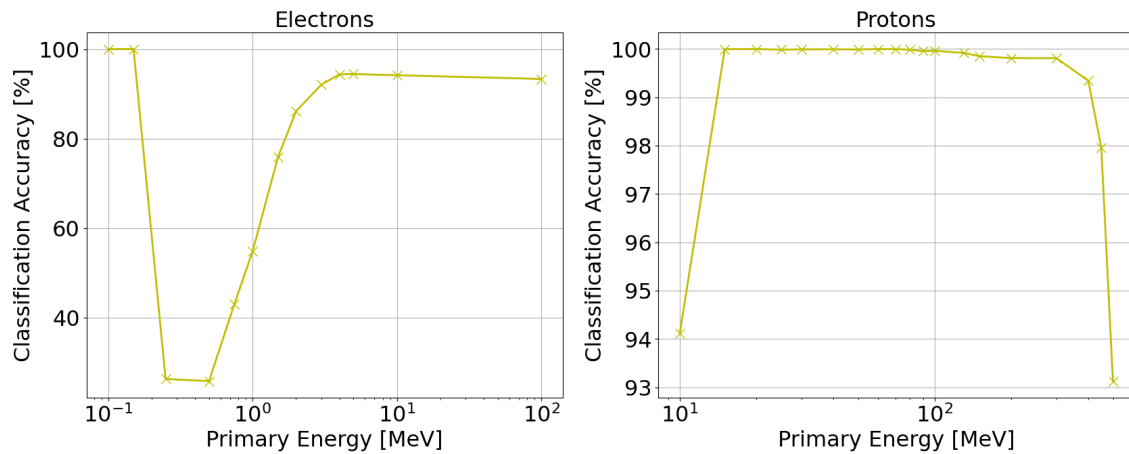
Figure 6.10: Proton/electron classification accuracy of the deconvolution algorithm for omnidirectional protons and electrons as a function of primary energy.

Third, to approximate the algorithm's proton energy prediction (spectroscope) accuracy, where applicable, a Gaussian curve was fitted to the proton response. The deviation of the mean of the Gaussian fit from the true energy mean can then be taken as the systematic error and the standard deviation of the fit as the statistical error (shown in Figure 6.11). The mean systematic and statistical errors calculated for a flat spectrum from 0 to 500 MeV protons are given by 8.85 MeV and 45.80 MeV respectively.
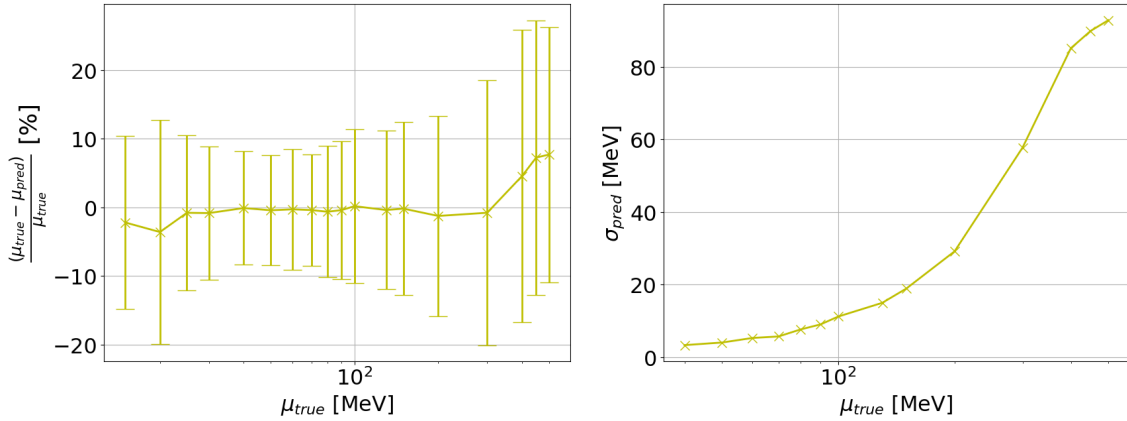
Figure 6.11: Graph of the deviation of the predicted mean from the true mean (left) and standard deviation of the predicted distribution (right) when applied to multiple mono-energetic omnidirectional proton fields as a function of the mono-energetic field energy using the developed deconvolution algorithm, illustrating the systematic error and statistical errors respectively. Predicted mean and standard deviation were obtained via Gaussian fitting.

6.4   Machine Learning-Based Particle Tracking

In Section 6.2, it was shown that, although the optimal analytical particle tracking algorithm was found to be the LLM-improved algorithm, this algorithm is not without its faults. When applied to the datasets used in this paper, issues such as the low accuracy for electrons with primary energy $\leq 1 \ MeV$ and binning issues when predicting low angles of incidence were observed with the LLM-improved algorithm. This called for the need for an alternate algorithm. As an exhaustive investigation into analytical approaches has already been performed in Section 6.2, an investigation into more advanced ML approaches was needed. Initially, it was impossible to predict which ML algorithm or particle features would be optimal. Therefore, all models with a large number of calculable features as input, excluding the CNNs discussed in Section 4.3, were tested. Figure 6.12 depicts the results when applied to the validation, showing a Random Forest regressor to be the optimal model.
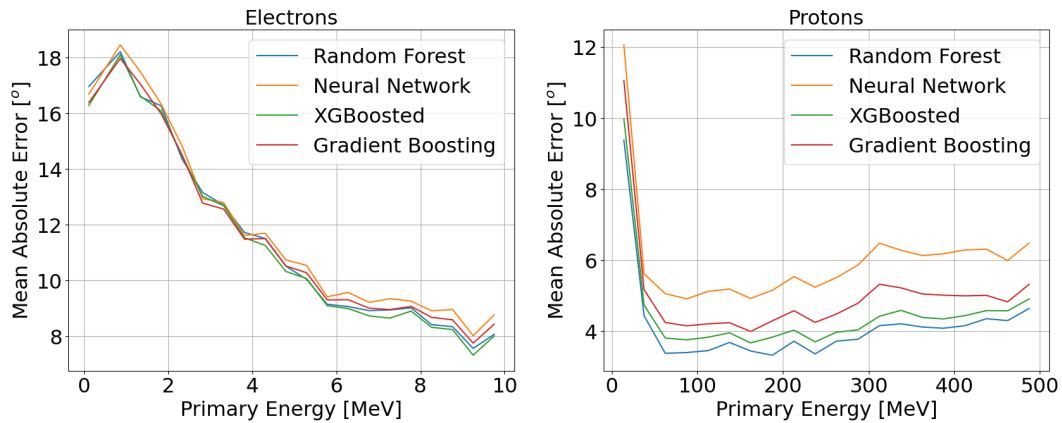
Figure 6.12: Mean absolute error of the predictions of θ using various machine learning algorithms as a function of particle primary energy for electrons and protons when applied to the validation dataset.

An optimal feature search, as discussed in Section 4.5, was then performed (see Figure 6.13). The features found to be most distinguishing were energy deposition, maximum box dimension, minimum box dimension, and weighted std along line (see Section 4.4).
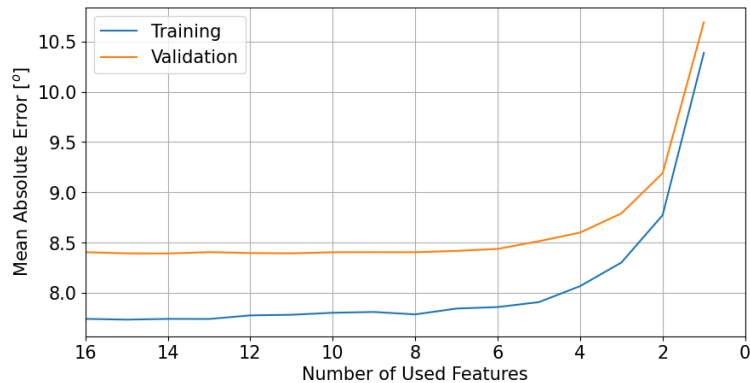

Figure 6.13: Graph of mean absolute error of the predictions of θ using a Random Forest regressor as the least distinguishing features are iteratively removed.

The final algorithm was constructed based on the extensive optimal model search. The finalised algorithm was then applied to the testing dataset (Figure 6.14), showing a significant increase in accuracy at all energies. When applied to the testing dataset, the mean absolute error of the final algorithm was found to be $8.65°$.
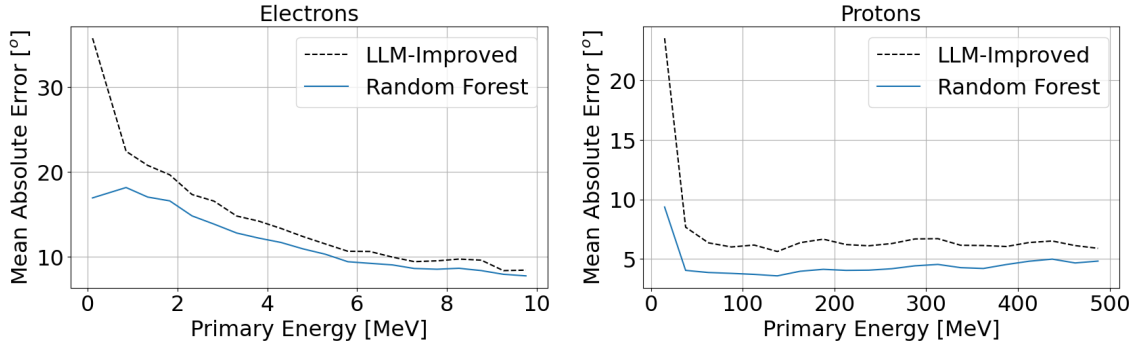
Figure 6.14: Graph of mean absolute error of the predictions of θ using the Random Forest regressor compared with the LLM-improved algorithm as a function of primary energy for electrons and protons. LLM, Left Lower-most, Right Upper-most.

Two-dimensional histograms of the true θ value versus the predicted θ value for electrons and protons can be seen in Figure 6.15. In the figure, it can be seen that the binning issue at low angles of incidences has been solved.
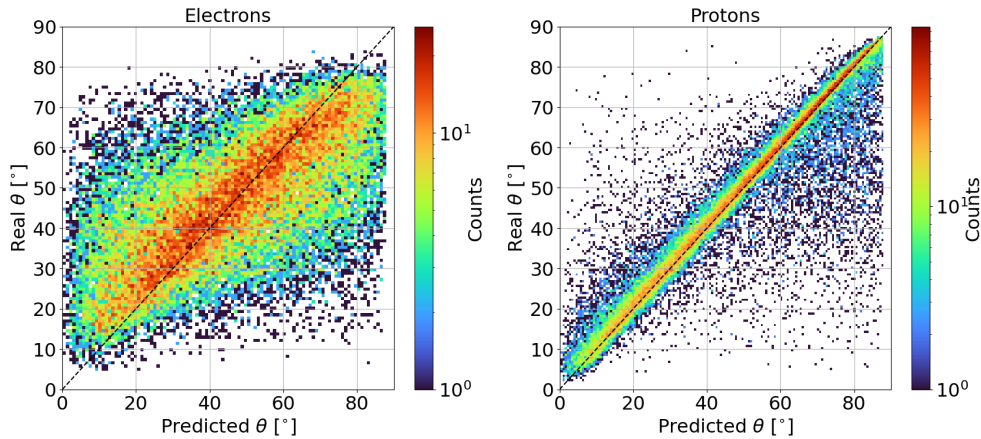

Figure 6.15: Two-dimensional histogram of predicted angle of incidence versus real angle of incidence as predicted by the Random Forest algorithm for electrons and protons.

6.5   Machine Learning-Based Particle Classification

As outlined in Section 4.3, the current state-of-the-art in ML classification (CNN) is unsuitable for the scope of this paper and hence the production of a new ML algorithm is favourable.

To avoid large training bias or a dramatic reduction in the amount of training data due to the unequal particle classes, a slightly different approach to that used in [12] is implemented. The problem will be split into two separate parts: (1) the initial classification of protons and electron and (2) the approximation of proton primary energy. Similarly to Section 6.4, an optimal model and feature search for both algorithms was then performed. For the classification optimal algorithm search (seen in Figures 6.16 and 6.17) the optimal setup was found to be an XGBoosted classifier with input features: energy deposition, average number of nearest neighbours 8-fold, height, weighted maximum std chi, and std along line (see Section 4.4).
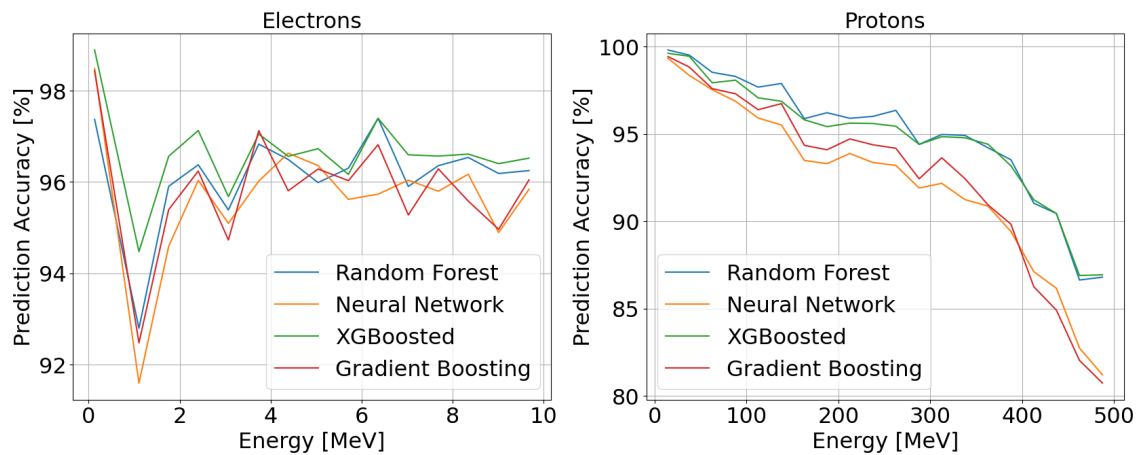
Figure 6.16: Electron/proton classification accuracy using various machine learning algorithms as a function of particle primary energy for the electron and proton dataset.
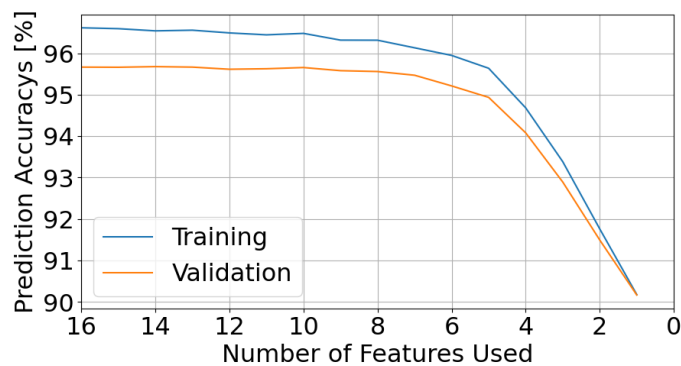
Figure 6.17: Graph of electron/proton classification accuracy using a XGBoosted classifier with the least distinguishing features iteratively removed.

The aforementioned optimal algorithm search was then applied for the proton spectroscope

model, illustrated in Figures 6.18 and 6.19. The optimal model was found to be a Random Forest regressor with input features: energy deposition, std of energy, weighted minimum std chi, weighted std along line, and average number of nearest neighbours 8-fold.
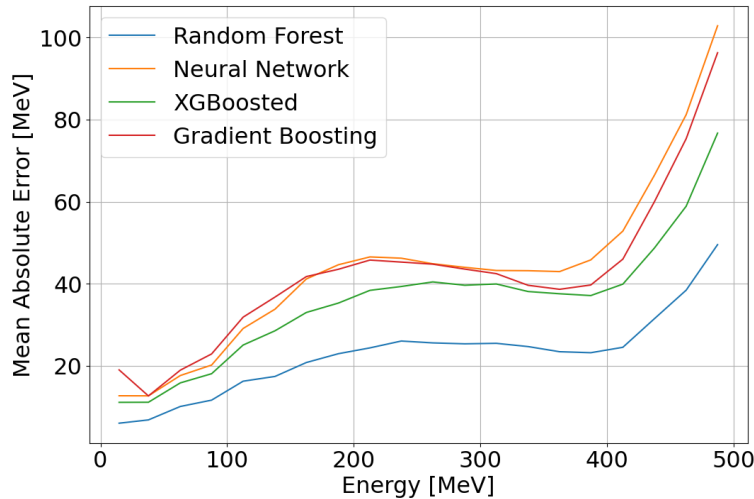

Figure 6.18: Mean absolute error of predicted proton primary energy using various machine learning algorithms as a function of true proton primary energy.
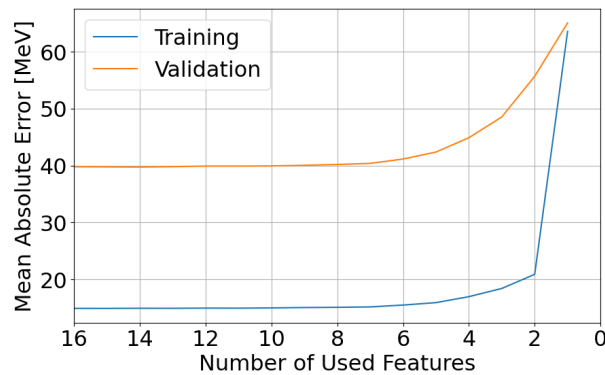

Figure 6.19: Graph of mean absolute error of proton energy prediction using a Random Forest regressor with the least distinguishing features iteratively removed.

After both particle identification algorithms had been developed, the algorithms were combined such that any particle intended for classification was initially classified into proton-like or electron-like; if classified as proton-like its proton-like energy was approximated. This

combined algorithm was then tested. For consistent comparison the same mono-energetic proton and electron datasets as used in Section 6.3 were used. First, the combined algorithm was applied to the mono-energetic fields individually to determine the accuracy of the algorithms in terms of radiation field reconstruction (see in Figures 6.20 and 6.21).
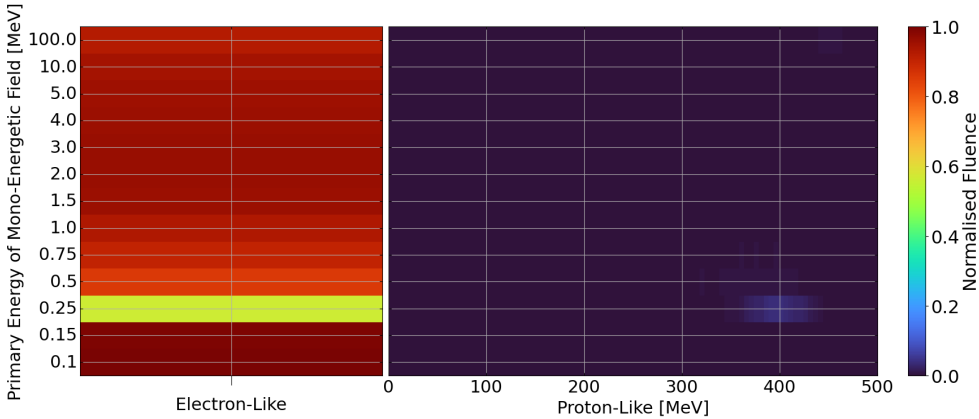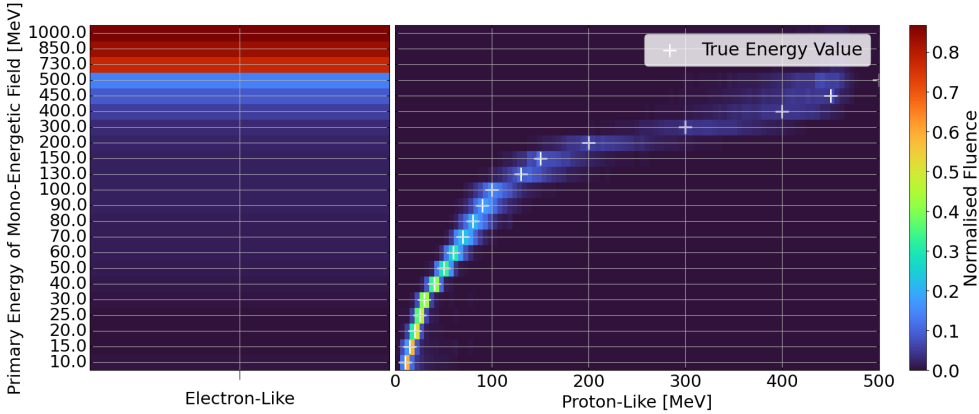

Figure 6.20: Predicted particle spectrum using the deconvolution algorithm for different mono-energetic omnidirectional electron fields to demonstrate statistical errors.

Figure 6.21: Predicted particle spectrum using the deconvolution algorithm for different mono-energetic omnidirectional proton fields to demonstrate statistical errors.

Second, the classification accuracy was calculated and compared with the deconvolution algorithm, as displayed in Figure 6.22. It should be noted from this figure that neither algorithm proved superior over the other in all cases of classification. Electrons were more

effectively classified using the XGBoosted classifier whilst protons were more effectively classified using the deconvolution algorithm. The average classification accuracy of the XG-Boosted classifier across the flat spectrums was found to be 96.7% for electrons and 96.6% for protons.
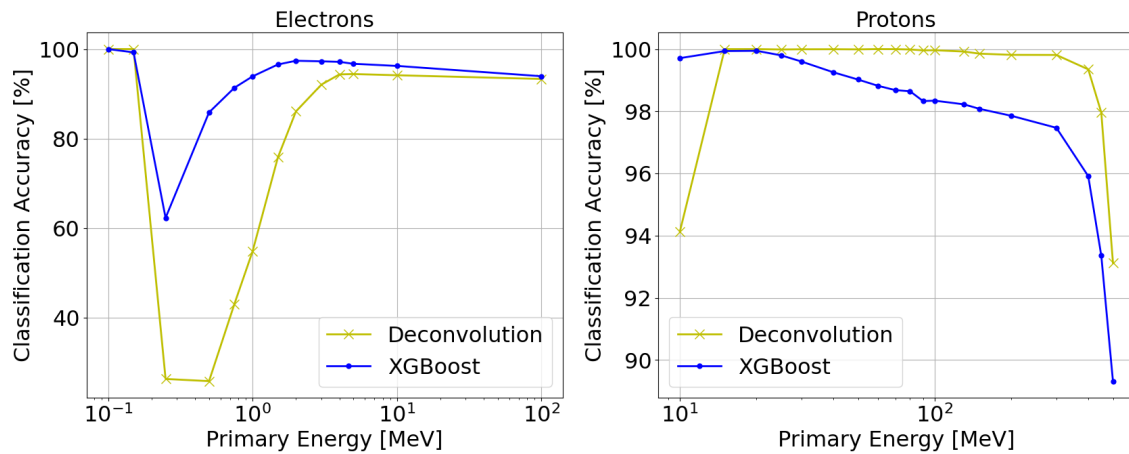

Figure 6.22: The electron/proton classification accuracy of the developed XGBoosted classifier as function of primary energy compared with the results of the deconvolution algorithm.

Third, the effectiveness of the developed ML algorithm as a proton spectroscope was determined by fitting a Gaussian curve to the predicted energy spectrum in each case. Once again, the systematic error was determined to be the deviation of the fitted mean from the true mean; statistical error was determined by the predicted standard deviation of the fit. The results of this can be seen in Figure 6.23. The average systematic and statistical errors across the flat spectrum were found to be 11.54 MeV and 30.24 MeV, respectively. It is worth noting from these results that the Random Forest algorithm has a lower statistical but higher systematic error than that of the deconvolution algorithm. A common problem of ML algorithms is having high variance but low bias when applied to testing sets [24]. The results in Figure 6.23 were achieved post regularisation by reducing the complexity of the trees.
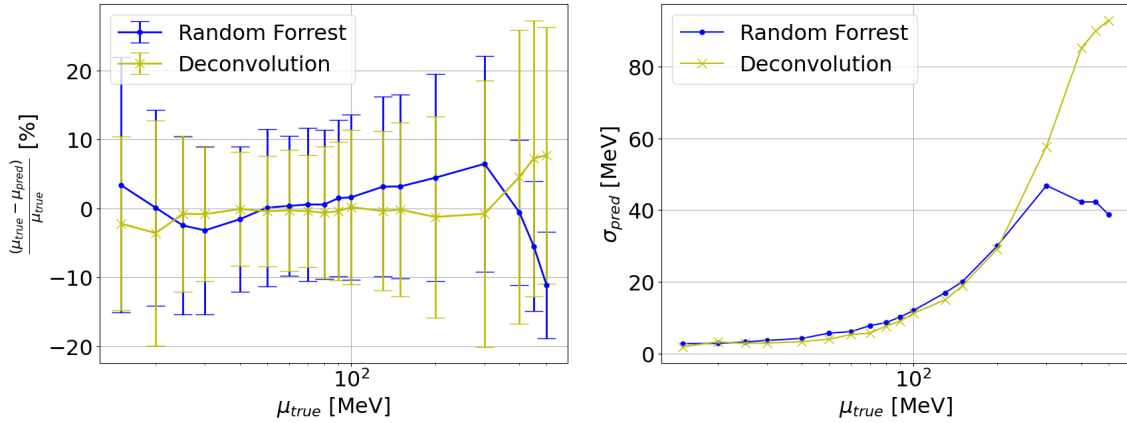
Figure 6.23: Graph of the deviation of the prediction mean from the true mean (left) and standard deviation of the predicted distribution (right) when applied to multiple mono-energetic proton fields as a function of mono-energetic field energy using the developed Random Forest and deconvolution algorithms, illustrating the systematic and statistical errors, respectively. Predicted mean and standard deviation were obtained through Gaussian fitting.

As final note, it can be observed that the developed ML and deconvolution field classification algorithms used different approaches regarding the classification of radiation fields, thus presenting differing behaviour during testing. This will allow for the detection of systematic errors when applied to unknown fields via measurement of the deviation of the two algorithms.

6.6   Classification Outside of Electrons and Protons

In terms of classification, it may seem somewhat naive to classify all particles as either proton or electron. However, there is an upper limit of classification due to the finite information provided by the Timepix3 detectors. This limitation in classification can lead to instability of the algorithms due to degeneracy of solutions in complex radiation fields. Regardless, the reduction of all particles into protons and electrons remains a valid approach provided that, when applied to all other particle types and energies under investigation, the results vary smoothly and predictably. There is a countless number of particles and energies that could be tested; therefore, to reduce the search space, only the most probable radiation types to be found in the radiation environments intended for analysis will be tested. To determine the algorithm's response to a particular particle type, multiple different mono-energetic fields of the chosen type were initially simulated. These mono-energetic fields were then passed through the algorithms and a response matrix was constructed. Subsequently, where possible,

a Gaussian curve was fitted to predicted proton fluence. An example of using this method for pions can be seen in Figure 6.24, showing that all pions above approximately 60 MeV appear as electron-like.
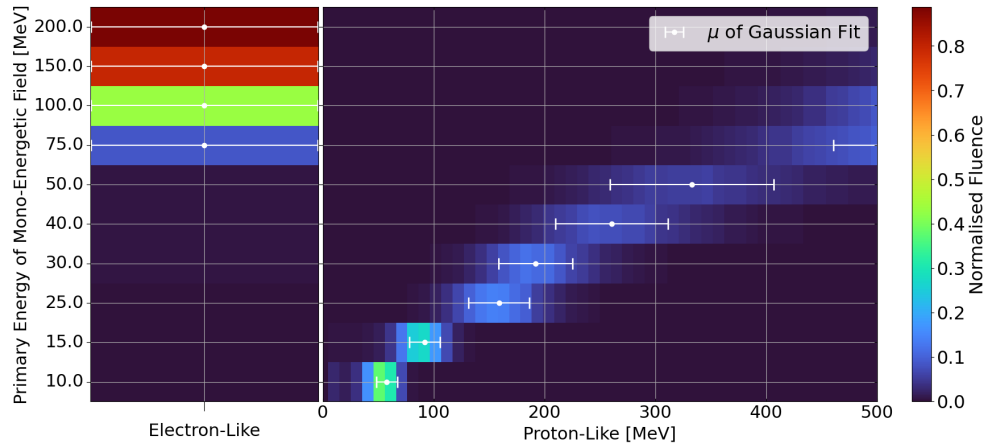

Figure 6.24: Plot of the response of the deconvolution algorithm to different mono-energetic pion fields. Where possible, Gaussian fits were applied to the proton-like response. The mean and standard deviation of the Gaussian fits are indicated with white points and error bars, respectively.

This method of response analysis was then performed for pion, muon, and gamma fields. The percentage of the field measured as electron-like was recorded, and where applicable a Gaussian curve was fitted to the proton spectrum, with the mean and standard deviation of the Gaussian fit recorded. The results of this analysis were then graphed in Figures 6.25 and 6.26 for the deconvolution and ML algorithms, respectively.
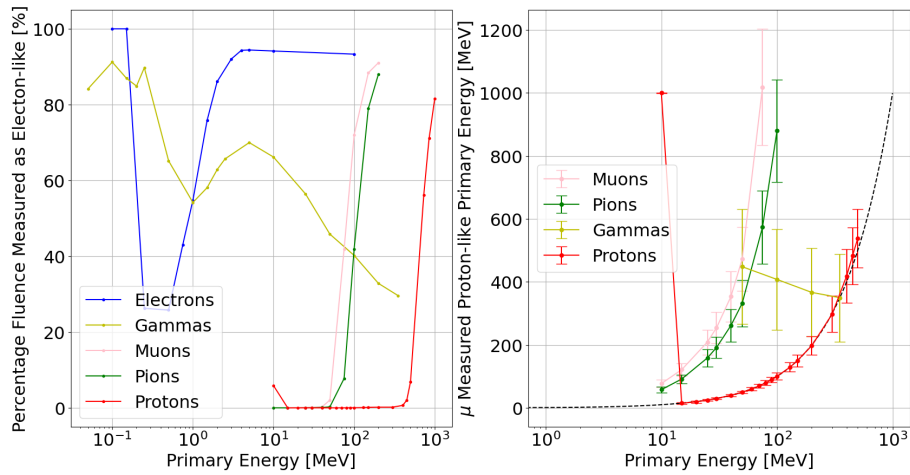
Figure 6.25: Response of the deconvolution algorithm to multiple mono-energetic particle types showing the percentage of the field detected as electron-like (left) and where applicable (proton-like fluence $> 50\%$) mean detected proton-like energy of the field (right).
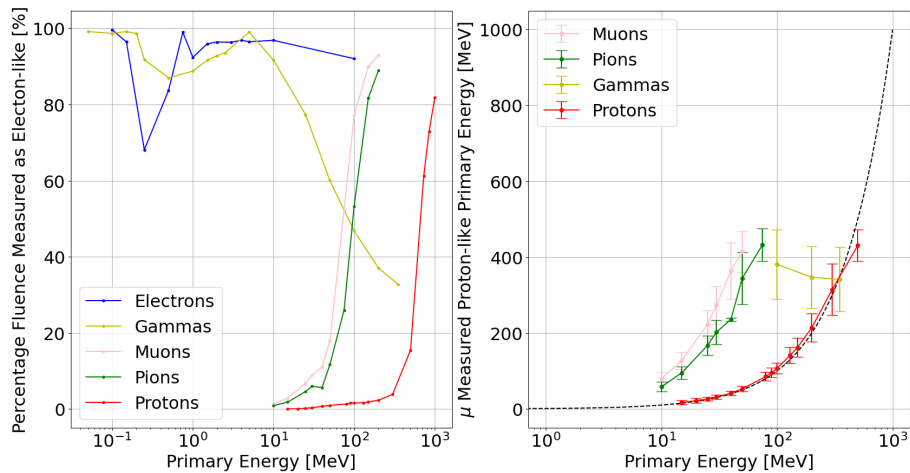

Figure 6.26: Response of the machine learning algorithm to multiple mono-energetic types showing the percentage of the field detected as electron-like (left) and where applicable (proton-like fluence $> 50\%$) mean detected proton-like energy of the field (right).

These results clearly show that muon and pion primary energies are directly correlated to their predicted proton-like primary energy suggesting a valid classification. However, this was not the case for gamma rays. Gamma rays do not vary smoothly with proton-like energy, indicating that introducing a gamma ray class may be beneficial. Yet, the introduction of said

class would present new obstacles. For example, there is a very large error associated with approximating a gamma ray's angle of incidence. Furthermore, the detection efficiency and cluster features of gamma rays are strongly correlated to their angle of incidence. One possible solution would be to consider gammas rays at different angles of incidence as separate particle classes, but this introduces yet more obstacles, such large degeneracy of solutions. It is for this reason that the rest of this paper concentrates on the proton-electron classification only, but these possible sources of error should be noted.

7 | Applications

In this chapter, the algorithms developed throughout this paper will be applied to real-world datasets. The algorithms will be initially tested with mono-energetic proton beams of known directionality. Following successful application of the algorithms, they will be applied to complex unknown radiation environments, such as those found in low Earth orbit and at the LHC at CERN, Switzerland.

7.1 The Danish Centre for Particle Therapy

Before moving to unknown fields, the developed algorithms must first be tested in a known radiation field. Thus, the experimental data from the DCPT in Aarhus, Denmark was utilised. A Timepix3 detector was placed on a rotating platform, at approximately a $1\ m$ distance from a proton beam of known primary energy. The experimental setup used can be seen in Figure 7.1.
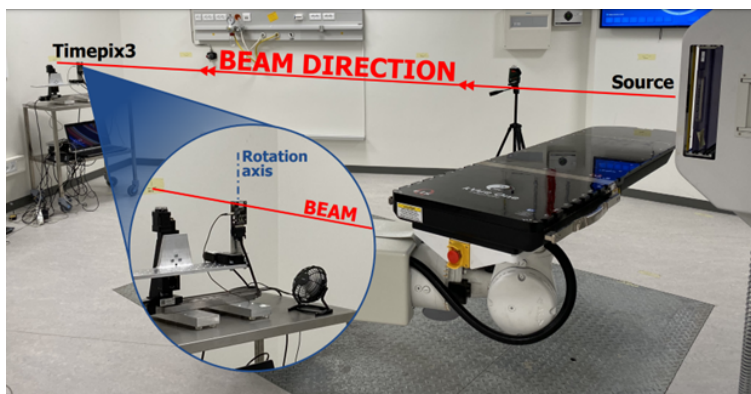

Figure 7.1: Picture of the experimental setup used during the proton beam exposure experiment at the Danish Centre for Particle Therapy. The location of the detector and the beam direction are indicated.

In this experiment, three primary energies were tested at four different incident angles. These are given in Table 7.1.

Primary Energy [MeV]	Incident Angles [$°$]
125.51 ± 1.66	$0, 41, 71, 85$
170.94 ± 1.50	$0, 41, 71, 85$
218.51 ± 0.95	$0, 41, 71, 85$

Table 7.1: Values used for the Danish Centre for Particle Therapy proton beam experiment.

Before application of the developed algorithms, data cleaning must be performed to ensure there are no noisy pixels and the operation parameters (bias voltage, etc.) are correctly set. To achieve this, it is enough to visualise several acquisition frames, and then inspect these for any pixels that are continuously activated and/or to ensure clusters are of the correct size and shape (see Figure 7.2).
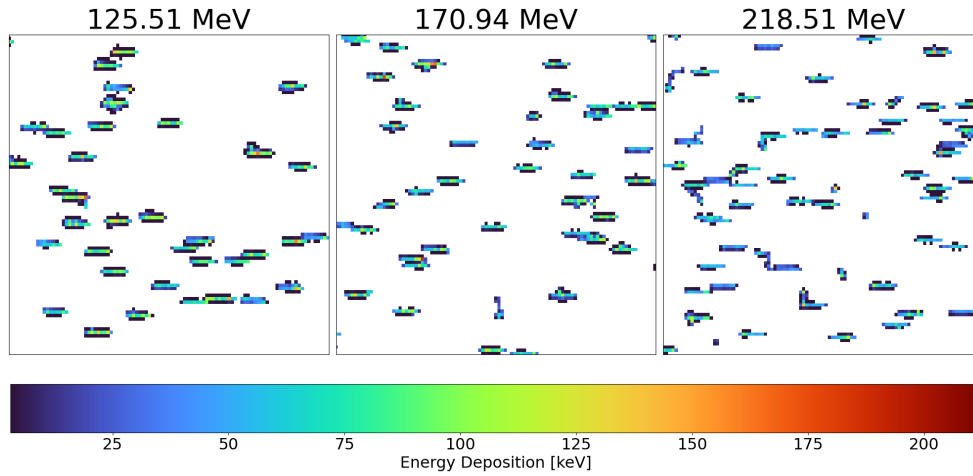

Figure 7.2: Timepix3 detector track visualisation from proton beam exposure for three primary energies(125.51 MeV, 170.94 MeV, and 218.51 MeV) obtained at the Danish Centre for Particle Theory.

7.1.1 Particle Tracking

The ML particle tracking algorithm developed in this paper was then tested on the DCPT dataset (results in Figure 7.3). The ML algorithm was found to have similar predictions to that of the original LLM-improved algorithm.
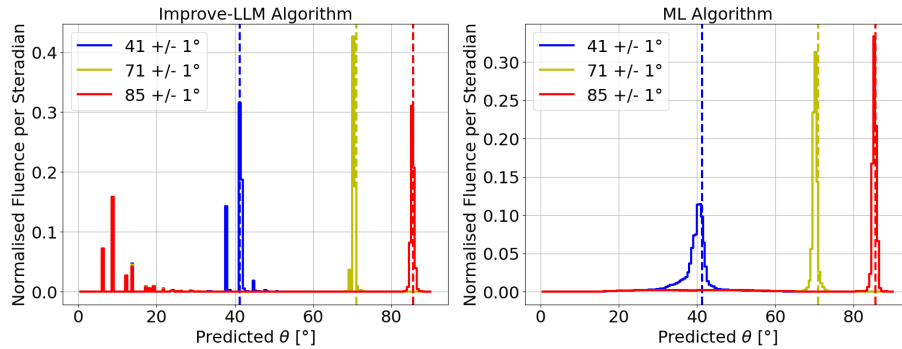
Figure 7.3: Predicted angles as determined by the analytical LLM-improved algorithm and the ML algorithm for a 170.94 MeV proton beam exposure with varying angles of incidence at the Danish Centre for Particle Therapy. The true angle of incidence is indicated by dashed lines of the same colour. LLM, Left Lower-most, Right Upper-most; ML, machine learning.

Figure 7.3 also illustrates the LLM-improved algorithm produces a non-negligible portion of particle fluence estimated to have $\theta \leq 20°$. The hypothesis for this was that low energy electrons were emitted from either the proton source or from the general background; these outliers are not present in the ML algorithm as it was trained on these particle types. To further investigate this further, a visualisation of the tracks that have a predicted angle of incidence $\theta \leq 20°$ was created (Figure 7.4). From this visualisation, it can be observed that these tracks are very similar in shape and energy deposition to that of low energy electrons (see Figure 4.1).
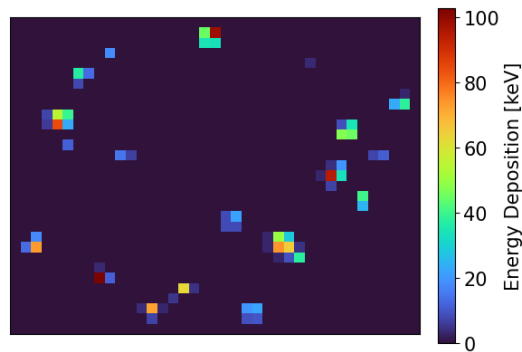

Figure 7.4: Visualisation of some tracks with an outlier predicted $\theta \leq 20°$ for a 170.94 MeV proton beam exposure at the Danish Centre for Particle Therapy.

50

These outliers may account for the slightly wider peak from the ML algorithm, as it is predicting these values to be closer to the correct angle of incidence with larger spread. It should be noted that the detector angle alignment was performed manually, and considering the particle beam is not visible, angle alignment errors due to precision of the human eye of up to $3°$ can be expected. Similarly, a spread of around $1 - 2°$ is expected based on the distance from the source and the finite size of the detector.

7.1.2 Particle Classification

The deconvolution and ML algorithms for classification of the radiation fields were then applied to the DCPT data to determine if they produce consistent results, as shown in Figures 7.5 and 7.6.
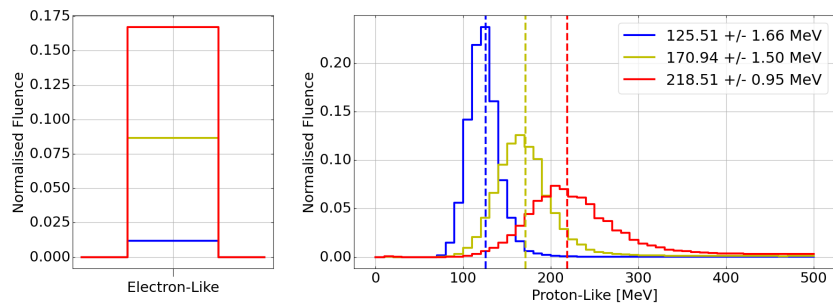

Figure 7.5: Predicted primary energy distribution using the deconvolution algorithm applied to the proton beam data obtained at the Danish Centre for Particle Therapy. The expected true primary energy is indicated by dashed lines of the same colour.
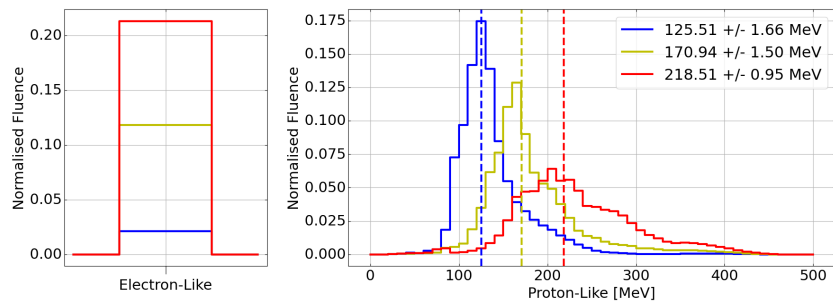

Figure 7.6: Predicted primary energy distribution using the machine learning algorithm applied to proton beam data obtained at the Danish Centre for Particle Therapy. The expected true primary energy is indicated by dashed lines of the same colour.

Figures 7.5 and 7.6 clearly show that the predictions are consistent, giving similar energy distributions around the expected mean primary energy values. Both algorithms detected an electron background within the beam which is consistent with the observed tracks in Figure 7.4. Once the shape of the predicted primary energy distribution was shown to be consistent for both algorithms and approximately Gaussian, the systematic and statistical errors can be directly analysed. A Gaussian curve was fitted to the predicted primary energy distributions in all cases. The deviation of the mean of the fit from the expected true mean can be considered as the systematic error and standard deviation of the fit as the systematic error. The results of these calculations are seen in Figure 7.7. Both algorithms are in excellent agreement with each other, showing very low systematic error.
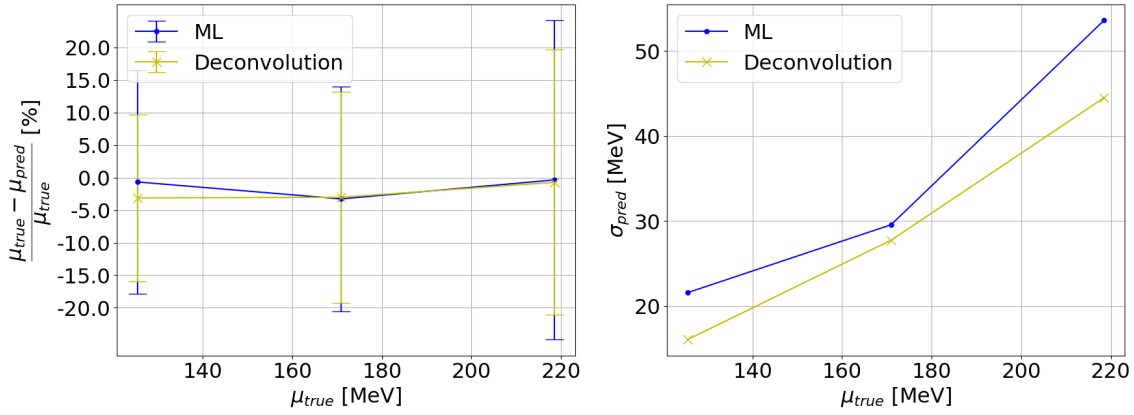

Figure 7.7: Graph of the deviation of the predicted mean from the true mean (left) and standard deviation of the predicted distribution (right) as a function of the true mean by application of the developed ML and deconvolution algorithms to proton data obtained at the Danish Centre for Proton Therapy. Predicted mean and standard deviation were obtained through Gaussian fitting. ML, machine learning.

7.2 Space Application of Timepix based Radiation Monitor (SATRAM)

The Space Application of Timepix based Radiation Monitor (SATRAM) is a radiation monitoring detector consisting of a Timepix detector with a 300 μm-thick silicon sensor placed on board the European Space Agency satellite Proba-V, which was launched into 820 km altitude (low Earth orbit) in 2013. The main goal of SATRAM was to demonstrate the capabilities of Timepix detectors in space [30]. However, since its launch, the capabilities of the

Timepix in areas such as ionising dose monitoring [31] and simulation bench-marking [12] have become quickly apparent. Due to the low weight, low power consumption, and radiation hardness of Timepix, it is an ideal candidate for the aforementioned applications. Arbitrarily, data from the first half of 2016 was chosen for this analysis.

7.2.1 SATRAM Environment

The majority of particles present in the radiation environment in low Earth orbit are protons and electrons that are trapped by the Earth's magnetic field. These particles are expected to have a primary energy much lower than those studied throughout this paper, with electrons having energies between 0 and 7 MeV and protons between 0 and 100 MeV [32]. Protons above 100 MeV and electrons above 7 MeV are present in the data, but their fluxes are orders of magnitude lower; thus, they can be considered negligible.

7.2.2 Low Orbit Field Structure

Initially, as a demonstration of the structure present in low Earth orbit, the total energy deposition recorded by the Timepix detector as a function of satellite position was produced, as shown in Figure 7.8. Three clear structures are visible: the northern polar horn, the southern polar horn, and the South Atlantic Anomaly (SAA). This observation is in agreement with the current expectations and previous measurements [32].
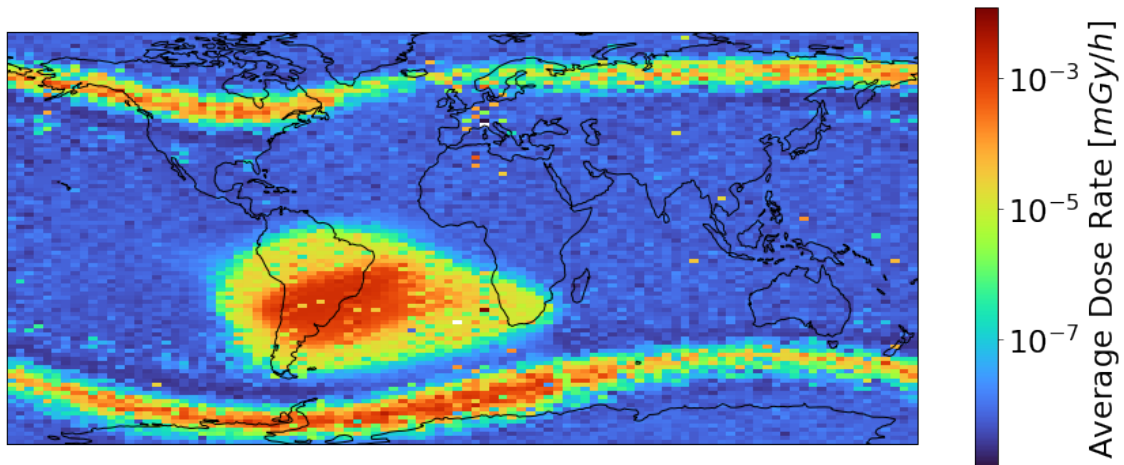

Figure 7.8: The average dose rate as measured by the Space Application of Timepix based Radiation Monitor at 820 km altitude in a low Earth orbit.

The northern and southern polar horns correspond to the points at which the satellite passes through the Earth's outer radiation belt, where electrons are trapped. The SAA is present due to the satellite crossing the Earth's inner radiation belt, which contains a mixed field of electrons and protons. This crossing is possible due to the incline of the Earth's magnetic dipole combined with the deviation of the Earth's magnetic centre with respect to the Earth's centre of mass. Figure 7.9 illustrates this magnetic field structure.
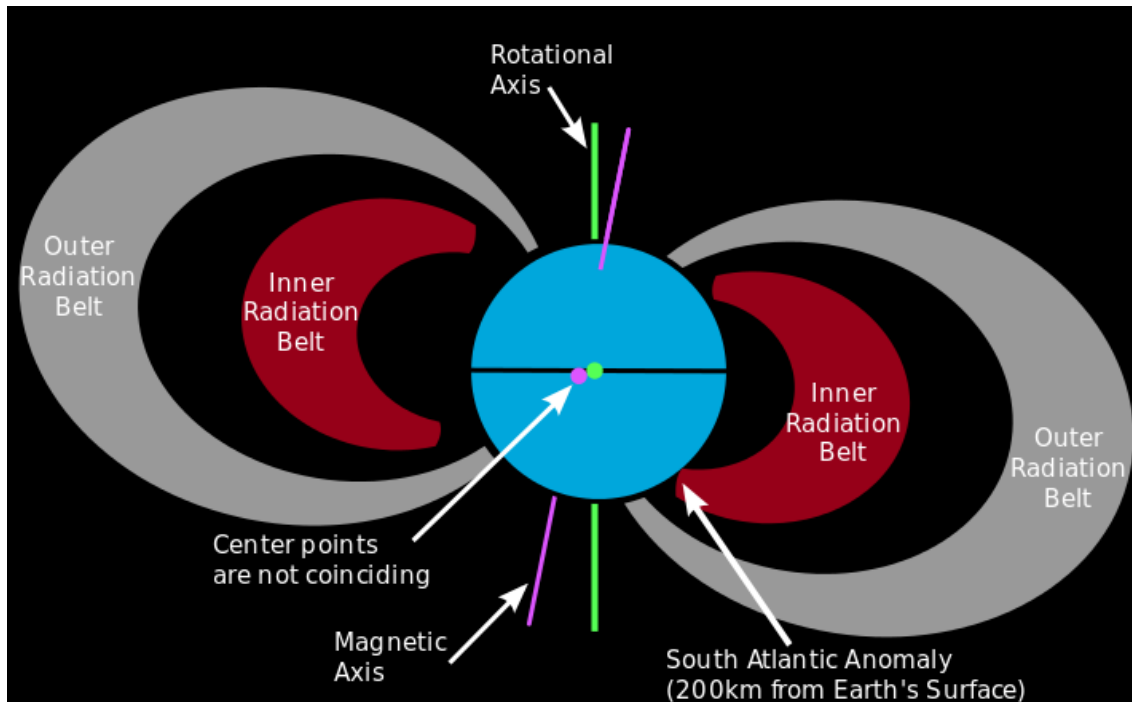

Figure 7.9: Basic illustration of the Earth's magnetic field structure [33].

It should be noted that the SAA is the only position in SATRAM's orbit that a non-negligible flux of protons should be present [32].

7.2.3 Preliminary Results

Once the position of the SAA in Figure 7.8 was identified, it was then possible to separate the radiation into two categories: radiation inside and radiation outside of the SAA based on the satellite position. As the SATRAM experiment uses Timepix in frame-based mode, there is a problem of overlapping tracks. As discussed in Section 2.5, Timepix allows only the

measurement of either ToA or ToT; therefore, ToA can no longer be used to distinguish particle clusters. To avoid detrimental amounts of overlapping clusters, only frames with ≤20% occupancy were used. Henceforth, occupancy will be defined as the number of activated pixels in a frame divided by the number of available pixels. This data was then divided into two separate sets: data accumulated inside and data accumulated outside of the SAA. The deconvolution and ML algorithms were then applied to both datasets (see Figures 7.10 and 7.11).
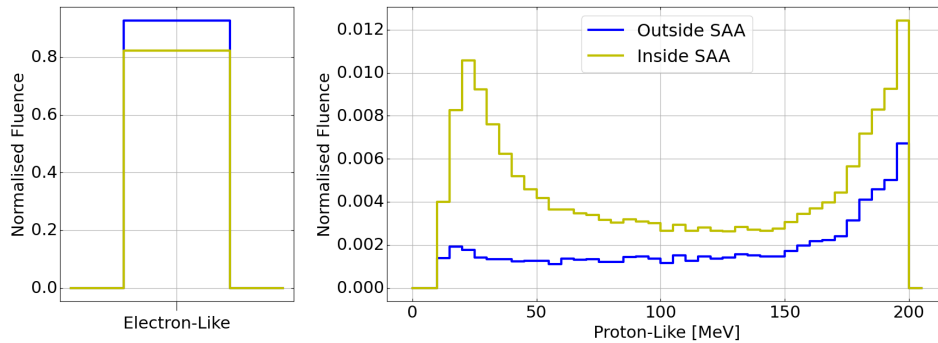

Figure 7.10: Classified particle fluence into the classes electron-like and proton-like (see Figure 6.25 for other particles) for inside and outside the SAA using the developed deconvolution algorithm. SAA, South Atlantic Anomaly.
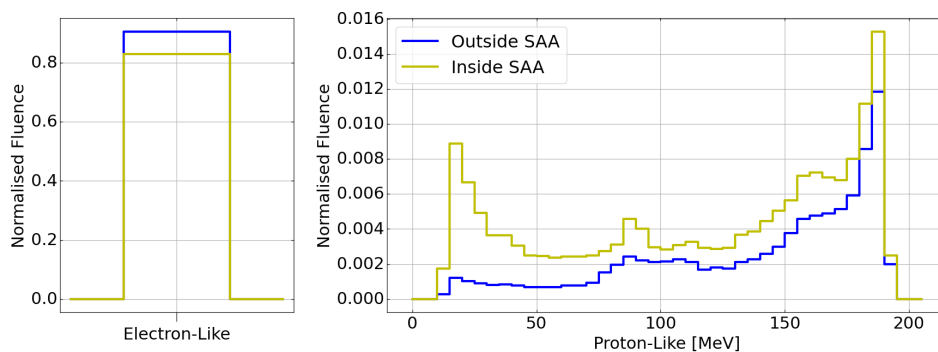

Figure 7.11: Classified particle fluence into the classes electron-like and proton-like (see Figure 6.26 for other particles) for inside and outside the SAA using the developed machine learning algorithm. SAA, South Atlantic Anomaly.

The algorithms were shown to be in good agreement; however, both predict a large amount of proton fluence outside of the SAA and a large amount of fluence of protons above

100 MeV, which is known to be incorrect [32]. The origin of this error is discernible when a simulated Timepix frame of 10% occupancy is plotted (Figure 7.12). Even in a low occupancy frame, overlapping tracks are still prevalent, causing multiple particles of radiation to be recorded in one particle cluster, thus ultimately leading to misclassification. This is a problem in Timepix and was later solved in its successor Timepix3 (See Section 2.5).
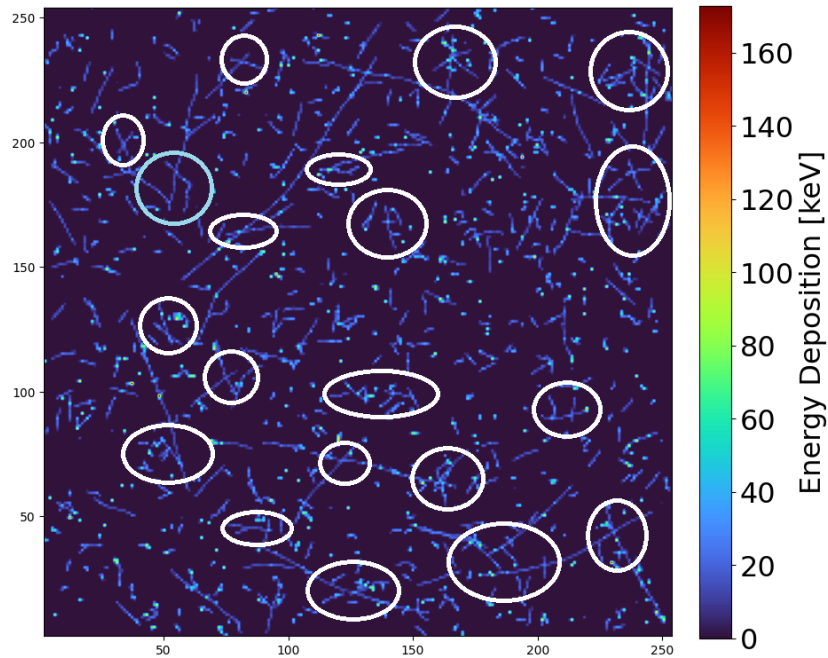

Figure 7.12: Example of a Timepix frame with a 10% occupancy, demonstrating a high percentage of overlapping tracks. Some overlap points are circled in white. The overlap is expected to lead to a confusion of electrons with protons.

Another possible contribution to this misclassification error is that, despite the expected classification accuracy of $\approx 98\%$ between electrons and protons within the energy ranges of 0 to 10 MeV and 0 to 200 MeV, there is a large flux of electrons relative to protons - a ratio of approximately 10^4. Thus, an error of only 2% will have a non-negligible effect on the predicted proton energy spectrum.

7.2.4   Proton Spectrum in the South Atlantic Anomaly

As discussed in the previous section, there are clear errors in the predicted proton spectrum. To combat this, it is presumed that there are a negligible number of protons above 100 MeV in either field. The fluence in this range can then be assumed to be solely due to "noise" and

the amount of fluence predicted in this range is therefore approximately proportional to the total noise in the field. Initially, both the curves outside and inside the SAA are multiplied by a normalisation factor such that the fluence of protons above 100 MeV sums up to 1. The overlay of the resulting spectra is displayed in Figures 7.13 and 7.14 for the deconvolution and ML algorithms, respectively. For both algorithms, the predicted proton spectrums >100 MeV have very similar structures outside and inside the SAA, which is indicative of the same phenomena causing this error.
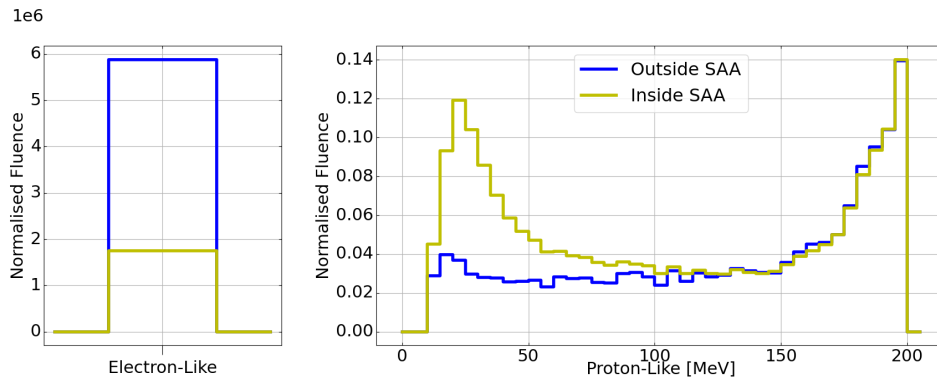

Figure 7.13: Particle fluence classified into electron-like and proton-like categories (see Figure 6.25 for other particles) for inside and outside of the SAA using the developed deconvolution algorithm after normalisation of protons with fluence >100 MeV. SAA, South Atlantic Anomaly.
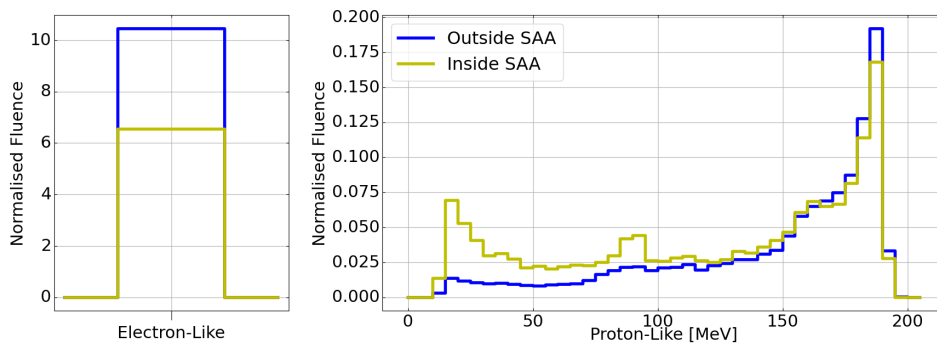

Figure 7.14: Particle fluence classified into electron-like and proton-like categories (see Figure 6.26 for other particles) for inside and outside of the SAA using the developed machine learning algorithm after normalisation of protons with fluence >100 MeV. SAA, South Atlantic Anomaly.

57

The curves inside and outside the SAA are then subtracted from each other and re-normalised. The resulting spectrum can then be presumed to be the true proton energy distribution. The results from this process for the deconvolution and ML algorithms can be seen in Figure 7.15. The error for these curves was calculated using the subtracted fluence at each point with an additional Poisson error for weighted histograms calculated by the square root of the sum of squares of all the weights [34]. Both algorithms are in reasonable agreement.
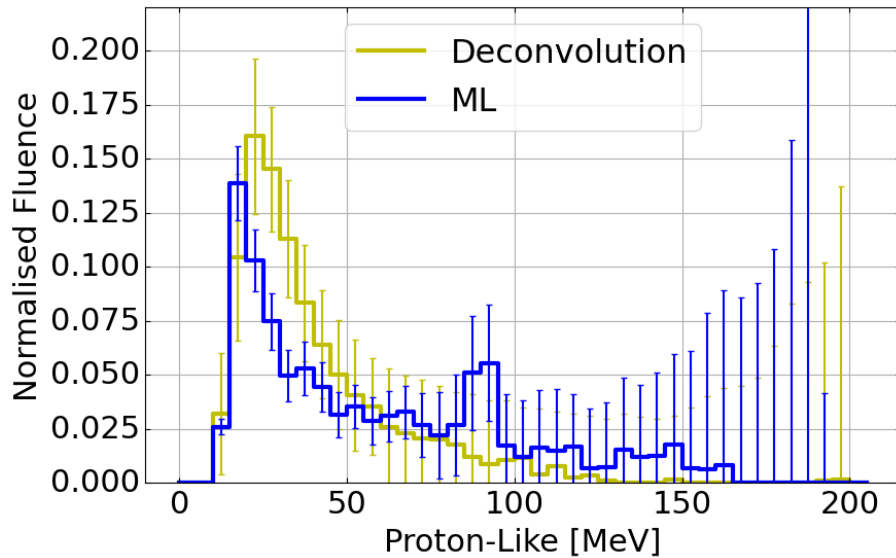

Figure 7.15: Comparison of the predicted proton energy spectra for the South Atlantic Anomaly obtained with the deconvolution and ML algorithms. The statistical uncertainty was determined via Poissonian statistics. ML, machine learning.

To obtain the final result, the average of both algorithms was taken (Figure 7.16). Protons below 10 MeV cannot be detected due to the casing around the detector. The flux quickly decreases with increasing energy. It is worth noting that this result is of particular significance as it is the first time a physically reasonable proton energy spectrum inside the SAA has been measured using Timepix detectors, opening the door to potential work in this field.
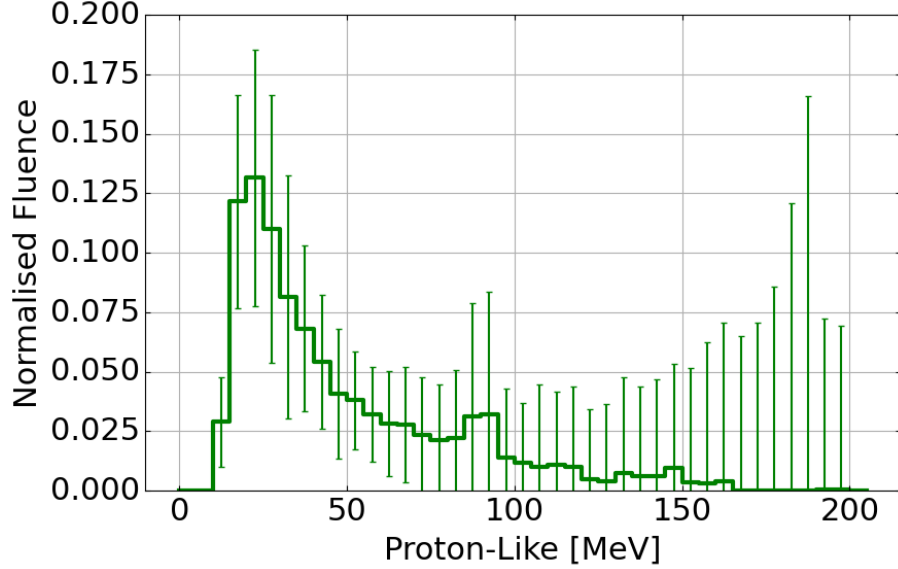
Figure 7.16: Predicted proton fluence distribution for the South Atlantic Anomaly using the average of the deconvolution and machine learning algorithms. Systematic errors were calculated using the difference between the two algorithms while statistical uncertainty was determined using Poissonian statistics.

7.3 Monopole and Exotic Detector at the LHC (MoEDAL)

The Monopole and Exotic Detector at the LHC (MoEDAL) is the seventh detector system located at the LHC in CERN, Switzerland, coordinated by the European Organisation for Nuclear Research. The primary objective of this experiment is the detection of magnetic monopoles and other exotic highly ionising stable particles [35].

7.3.1 Experimental Setup

During Run-2, Timepix3 was the only active detector system present in this experiment -all others were passive - giving it an important role for real-time monitoring of the radiation field. For this specific analysis two collision periods during 2018 were chosen: a lead-lead (PbPb) collision period on the 25th of November, and a proton-proton (pp) collision period on the 24th of September. During these periods, two Timepix3 detectors were actively measuring the radiation field. A visual representation of the experimental setup can be seen in Figure 7.17. The Timepix3 detectors are within an aluminium casing at the positions labelled

"MIDDLE1" and "MIDDLE2" in the figure. Within the scope of this work, data was analysed from the detector located at "MIDDLE2".
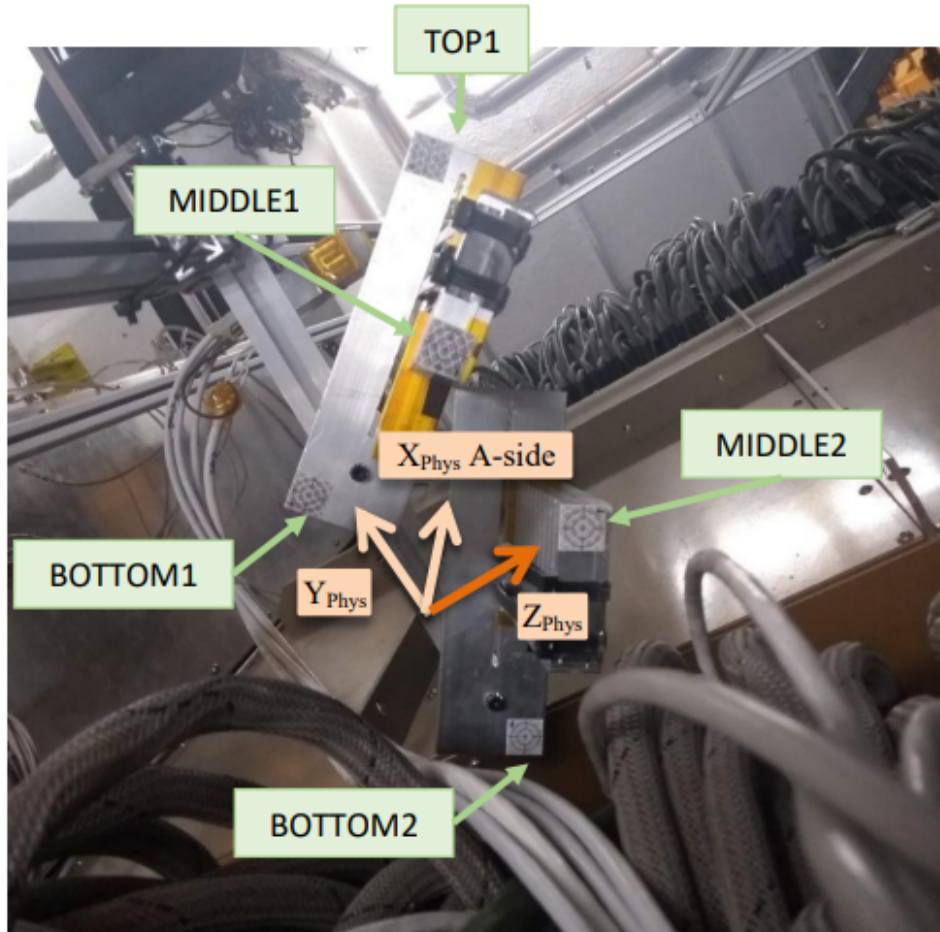

Figure 7.17: Visual representation of the experimental set up of the Timepix3 detectors present at the MoEDAL experiment. The detectors are located at a distance of $\sim 1\,m$ from the interaction point (IP8).

7.3.2 Particle Tracking

To ensure the algorithms developed agree with current results [16, 36], polar histograms of the particle fluence were created for both periods. To create these histograms, the ML algorithm developed in this paper was used to calculate θ; the line phi algorithm as mentioned in [16]

was used to calculate the φ. The results can be seen in Figures 7.18 and 7.19 for PbPb and pp collision periods, respectively. A clear interaction point (the point of collisions of the opposing beams [IP]) is visible, the position of which is consistent with current results, $\theta = 56 \pm 1°, \varphi = 139 \pm 1°$. In Figure 7.18, a clear background can also be seen around $\varphi = 300°, \theta = 60°$. As this background is not present in the pp-collision polar histogram shown in Figure 7.19, the current hypothesis is that these structures are due to a larger halo around the Pb beams, which can interact or scatter off collimators.
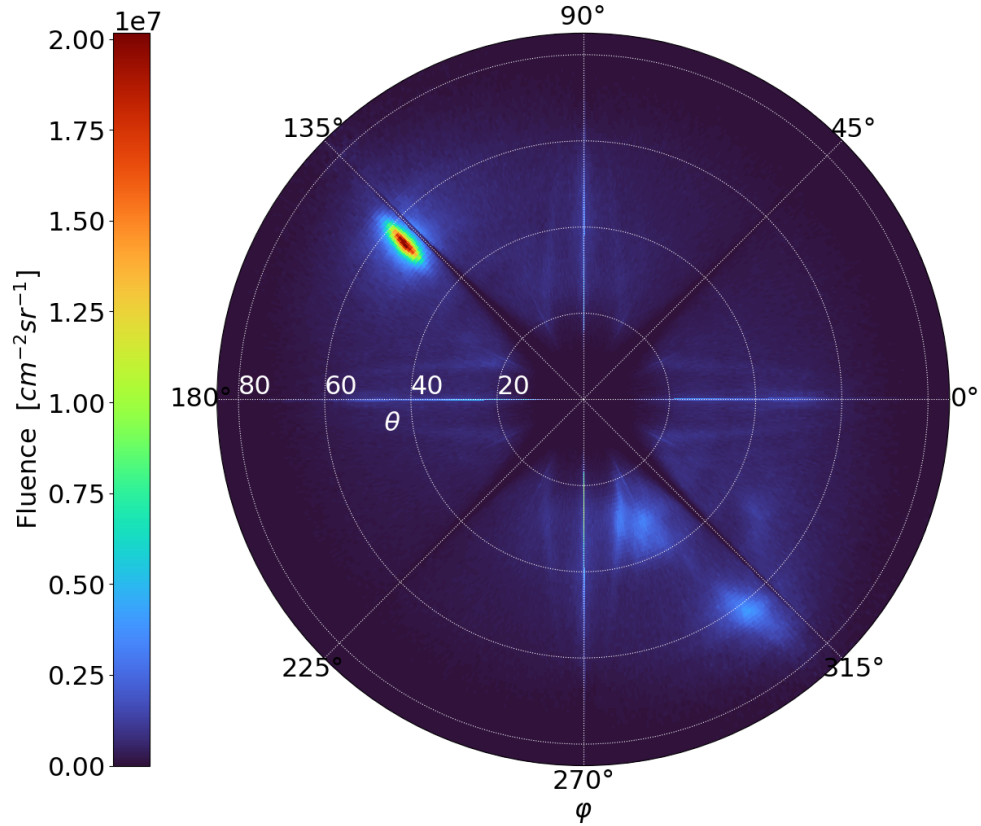

Figure 7.18: Polar histogram of the measured impact angles θ, using the developed machine learning algorithm and φ using the line fit algorithm described in [16]. Particle clusters of sizes greater than 5 pixels were selected. The detector was located at a distance of $1\ m$ from IP8 within the MoEDAL experiment. Data was acquired during the lead-lead collision period on the 25th of November 2018.
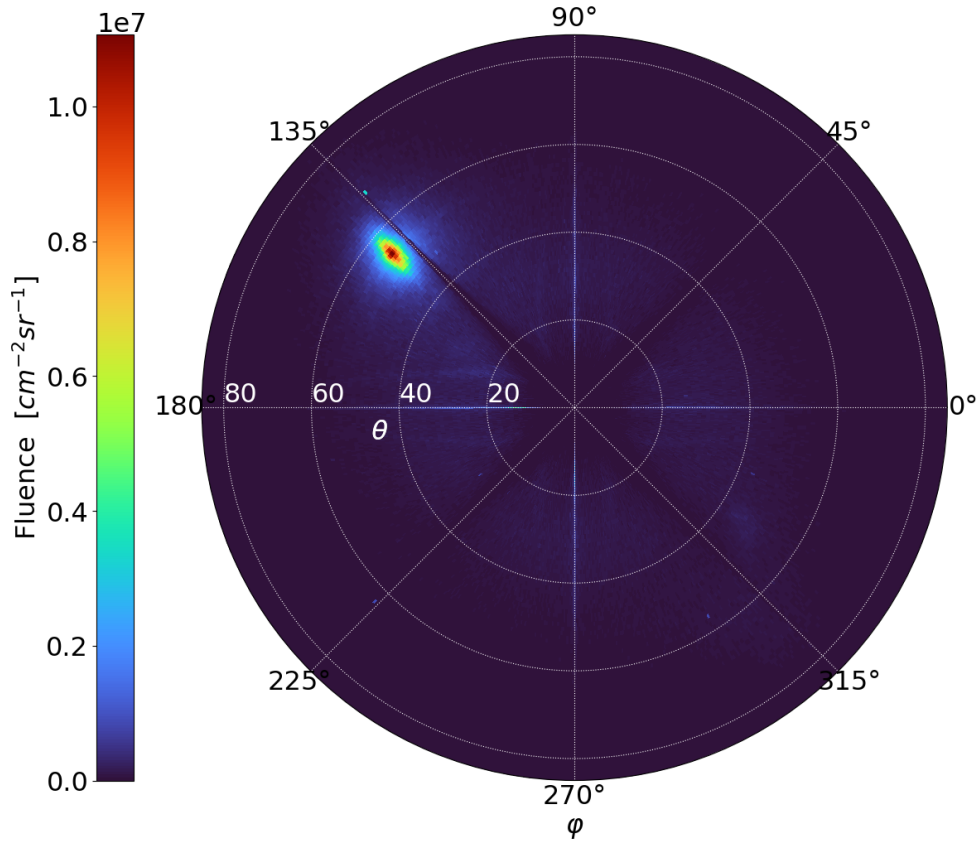
Figure 7.19: Polar histogram of the impact angles θ determined with the developed machine learning algorithm and φ using the line fit algorithm from [16]. Particle clusters of size greater than 5 pixels were selected. Data was acquired during proton-proton collision periods at a distance of $1\ m$ from IP8 within the MoEDAL experiment on the 24th of September 2018.

Due to the increased accuracy of the θ predictions, the structure of the IP can now be resolved. A deeper investigation into the structure of the IPs was performed by zooming in on the two IPs to determine how they differ from each other (Figure 7.20). The PbPb point of interaction is slightly wider than that of its pp counterpart. This is expected, because, due to the higher electronic charge of Pb versus protons there is higher Coulomb repulsion of the particles within the beam, resulting in a wider beam and a larger IP. Nonetheless, it is surprising that this difference in IP size is measurable by the Timepix3 detector at such a large distance ($\approx 1\ m$).
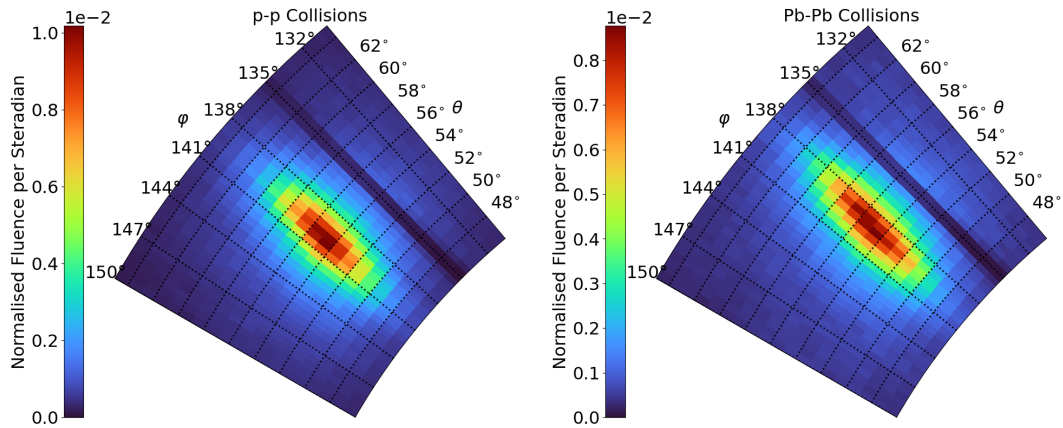
Figure 7.20: Magnification of the region of interest corresponding to the interaction points in the polar histograms from Figures 7.18 and 7.19.

7.3.3 Particle Fluence Monitoring

Once the IP position for both PbPb and pp collision periods was successfully identified based on the measured incidence angles of the particles, it was possible to split the measured field into contributions from the IP and background. Radiation from the IP was defined as all particles with measured angles of incidence, $\varphi \in [135°, 144°]$, and $\theta \in [50°, 64°]$; the background was defined as all radiation with θ or φ values outside of these intervals. This allows for the measurement of the flux from the interaction and background separately. Flux measurements were performed for both collision periods (see Figures 7.21 and 7.22). During PbPb collisions, data was available throughout the entire run duration, starting with the beam tuning. This is resembled in the graph by the ramping up of intensity: the initially low flux is followed by a step-wise and steep increase as the particle beams are being brought to collide. Interestingly, the background signal was already appearing before the actual start of the collision period. This indicates that the background had already been created when the beams were injected. This hints at its cause being related to a beam-halo effect. Moreover, a delay of the background flux with respect to the IP-flux increase is observed, the origin of which will be subject of further study. For the pp collisions, data was only taken after stable collisions were achieved and only for a short period of time. Thus, the measured fluxes remain stable throughout the full measurement period.
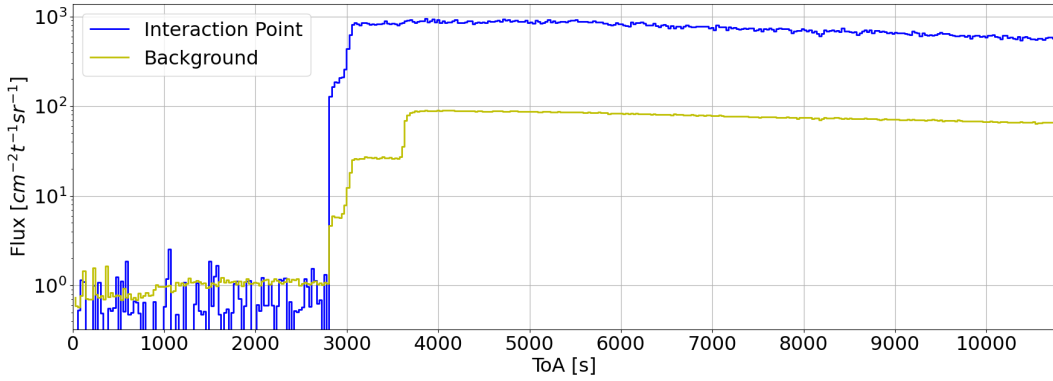
Figure 7.21: Flux recorded by a Timepix3 detector at MoEDAL during a lead-lead collision period using the developed machine learning algorithm for calculation of incident angle as input into the effective volume algorithm for flux measurement. Particles from the interaction point are separated by their angle of incidences within the intervals $\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°]$.
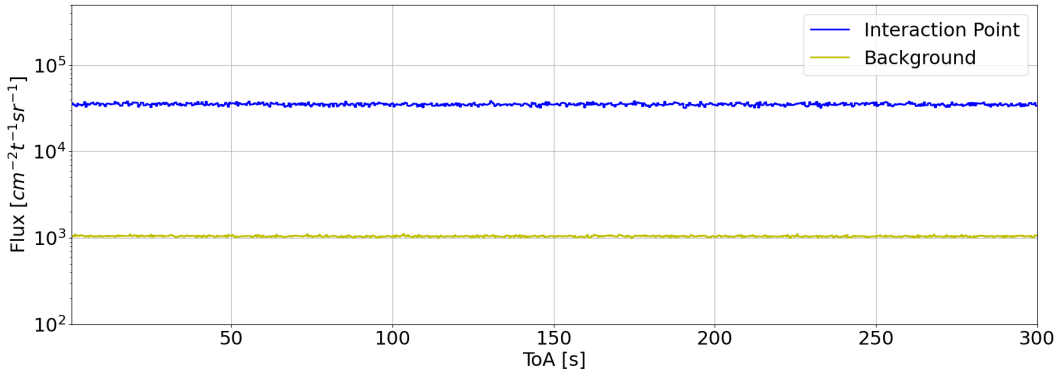

Figure 7.22: Flux recorded by a Timepix3 detector at MoEDAL during a proton-proton collision period using the developed machine learning algorithm for calculation of incident angle as input into the effective volume algorithm for flux measurement. Particles from the interaction point are separated by their angle of incidences within the intervals $\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°]$.

7.3.4 Radiation Field Classification

For brevity, only the analysis of the PbPb collision period is outlined in full in this section, as the analyses for the two periods are identical. Before classification is performed, it was first

64

necessary to plot the stopping power distribution for both the IP and background. Results of this can be seen in Figure 7.23.
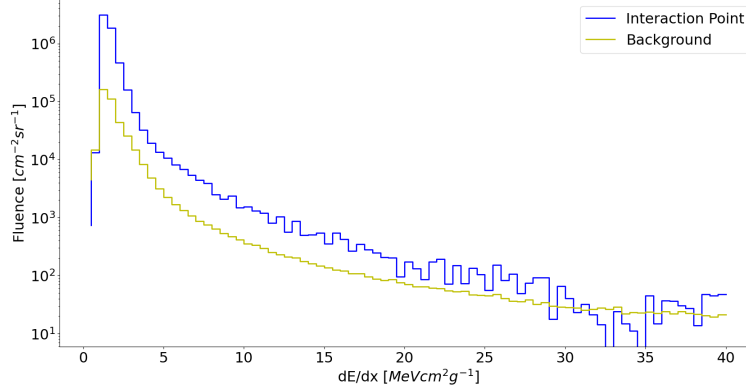

Figure 7.23: Stopping power distribution by Left Lower-most, Right Upper-most-improved algorithm [16] for a lead-lead collision period. The interaction point is defined by angle of incidence $\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°])$.

Figure 7.23 shows a peak position at $1.25 \pm 0.25\ MeV cm^2 g^{-1}$ for both regions which is in good agreement with previous results [16, 36]. Furthermore, there is a lower amount of particle fluence at higher stopping powers from the IP when compared with the background. This is to be expected, as the particles coming from the IP would be of much higher energy, and hence be less ionising in the detector. Showing that this distribution agrees with previous results further increases the confidence of the deconvolution algorithm which will use this curve to classify the field. The classification algorithms were then applied to the background and IP radiation, separately for the PbPb collision period. The obtained results are shown in Figures 7.24 and 7.25, respectively. In both figures, a peak in the fluence of proton-like particles at 500 MeV for the deconvolution and 440 MeV for the ML algorithms can be seen. This is due to the error in the misclassification of electron-like particles as proton-like. These particles are then pushed to the high end of the energy spectrum, as proton particles become increasingly electron-like with increasing energy. The peak is slightly lower in the ML algorithm, as Random Forest algorithms have difficulties in the prediction of edge values. As expected, the flux of electron-like particles is much higher than that of the proton-like particles. Most particles in the field have very high energies, causing them to be typically electron-like (see Section 6.6). Figure 7.24 illustrates that the ML predictions show a second peak at approximately 400 MeV, which is not present in the deconvolution predictions. This second peak can be explained by referring to Figure 6.20, which demonstrates that electrons with primary energy between 0.15-0.5 MeV typically produce a peak at approximately 400 MeV. This is indicative of the need for further development of ML algorithms to offset

this problem. This peak is not expected to appear in the deconvolution predictions as it did not appear during testing (see Section 6.8).
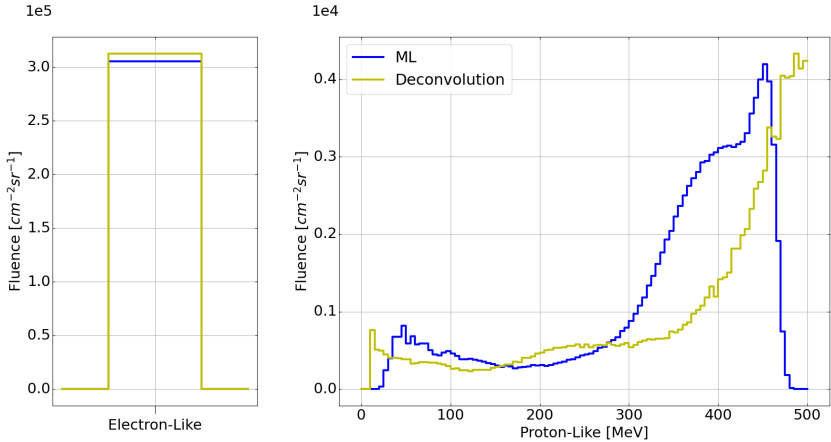

Figure 7.24: Results of the particle classification into the classes electron-like and proton-like for the background radiation ($\varphi \notin [135°, 144°]$ and $\theta \notin [50°, 64°]$) at MoEDAL during the lead-lead collision period using the developed deconvolution and ML algorithms. ML, machine learning.

Figure 7.25: Results of the particle classification into the classes electron-like and proton-like for radiation from the interaction point ($\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°]$) at MoEDAL during the lead-lead collision periods using the developed deconvolution and machine learning algorithms. ML, machine learning.

66

For the final prediction of the radiation field decomposition and systematic error assessment, the average of the predictions of the deconvolution and ML algorithms was taken. The systematic error was determined to be the absolute difference between the two predictions. The statistical uncertainty was calculated using Poissonian errors for the weighted histograms, calculated by the square root of the sum of squares of all the weights [34]. The obtained results are shown in Figures 7.26 and 7.27

Figure 7.26: Results of particle classification into the classes electron-like and proton-like for the background radiation ($\varphi \notin [135°, 144°]$ and $\theta \notin [50°, 64°]$) at MoEDAL during the lead-lead collision period. Results are represented by the average of the deconvolution and machine learning algorithm predictions. The systematic errors were calculated using the difference between the two algorithms, while the statistical uncertainty was determined by Poissonian statistics.
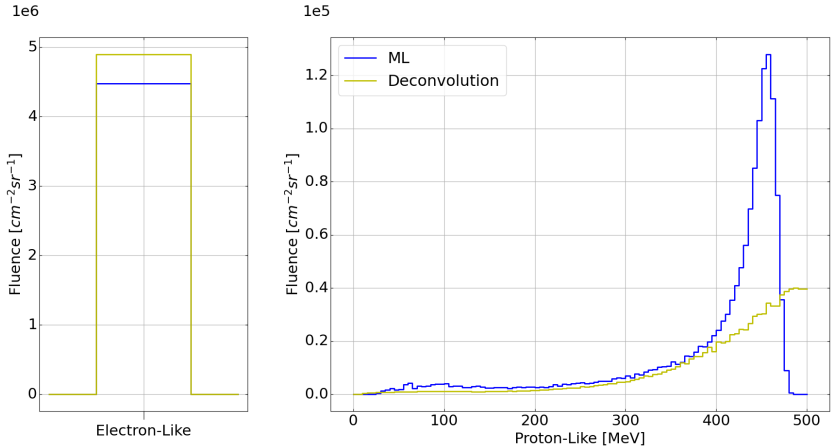
Figure 7.27: Results of particle classification into the classes electron-like and proton-like for radiation from the interaction point ($\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°]$) at MoEDAL during the lead-lead collision period. Results are represented by the average of the deconvolution and machine learning algorithm predictions. The systematic errors were calculated using the difference between the two algorithms, while the statistical uncertainty was determined by Poissonian statistics.

The classification algorithms were then applied to the pp collision period, as seen in Figures 7.28 and 7.29. The prepossessing associated with these final results can be seen in Appendix B.
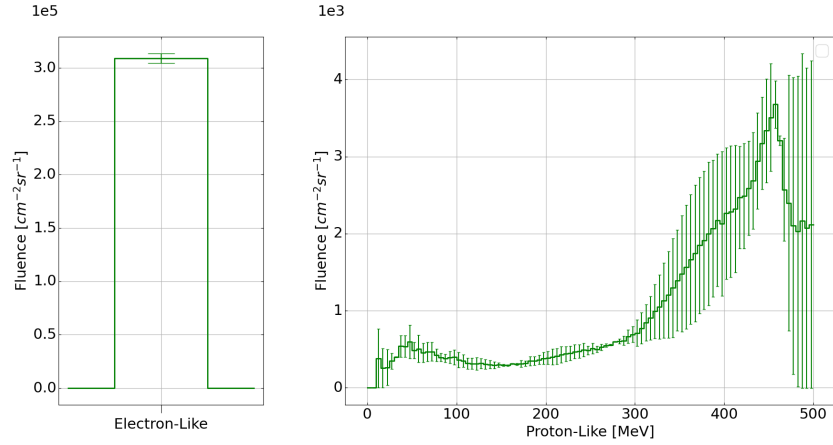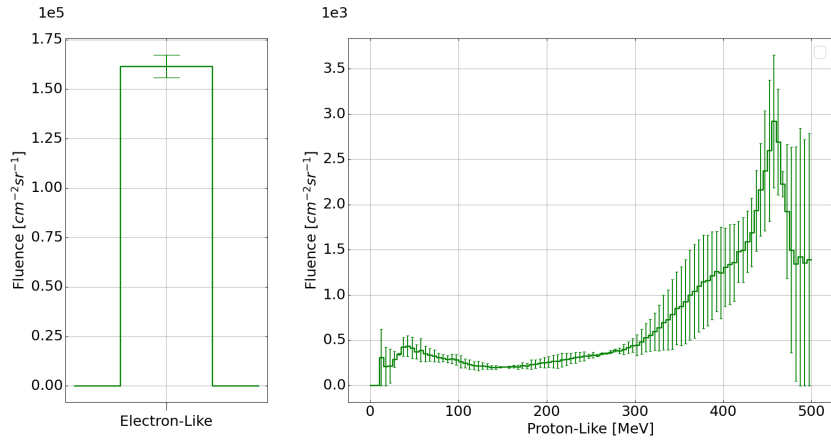
Figure 7.28: Results of particle classification into the classes electron-like and proton-like for the background radiation ($\varphi \notin [135°, 144°]$ and $\theta \notin [50°, 64°]$ at MoEDAL during the proton-proton collision period. Results are represented by the average of the deconvolution and machine learning algorithm predictions. The systematic errors were calculated using the difference between the two algorithms, while the statistical uncertainty was determined by Poissonian statistics.
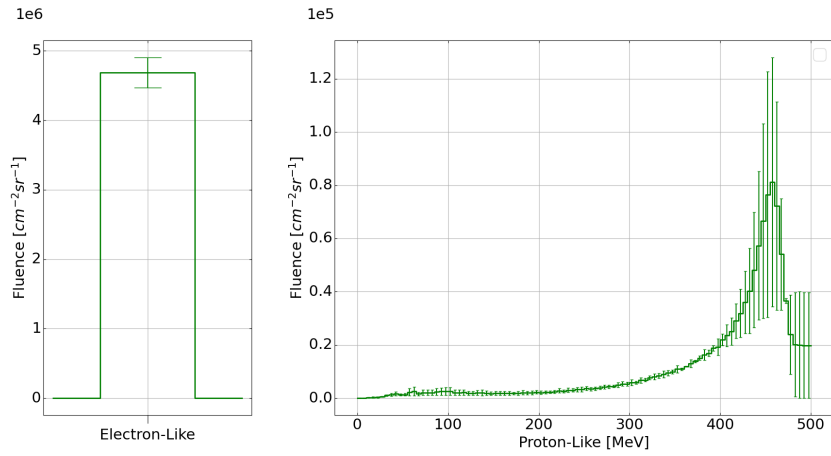

Figure 7.29: Results of particle classification into the classes electron-like and proton-like for radiation from the interaction point ($\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°]$) at MoEDAL during the proton-proton collision period. Results are represented by the average of the deconvolution and machine algorithm predictions. The systematic errors were calculated using the difference between the two algorithms, while the statistical uncertainty was determined by Poissonian statistics.

69

8 | Conclusions

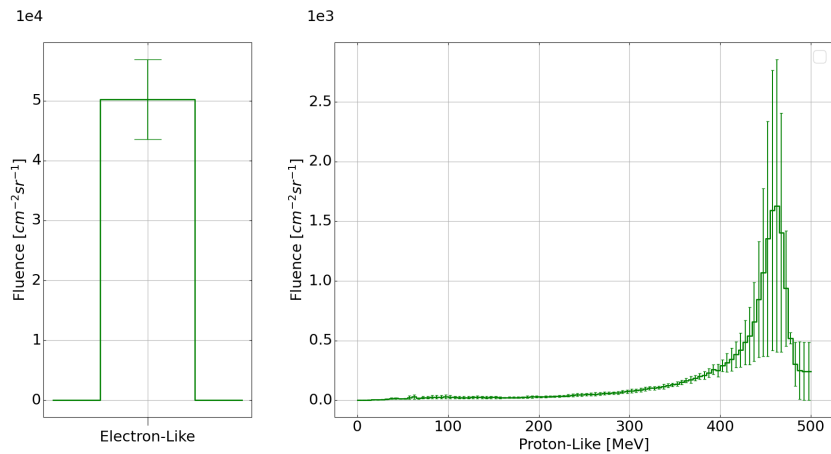The goal of this thesis was to demonstrate the capabilities of Timepix and Timepix3 detectors for radiation field decomposition, particularly in areas such as particle tracking, classification, and fluence measurement. These goals were successfully accomplished.

Ground truth data was created using Allpix^2 simulations; the simulation response was later validated through comparison with available measurements. The properties of planar detectors were then analysed.

State-of-the-art algorithms for particle tracking were reproduced and applied to the created dataset. These algorithms were proven to be insufficient when applied to the diverse datasets. The most optimal algorithm was found to be the LLM-improved algorithm with a mean absolute error of $14.02°$, indicating more advanced algorithms were required. A Random Forest Regressor was shown to be the best model, solving previous issues with a mean absolute error of $8.65°$.

Novel methods for particle classification were then investigated, leading to the creation of two novel methods - the deconvolution and ML methods. The newly developed methods were analysed in two aspects: proton/electron classification accuracy and mean proton energy prediction error calculated by the sum of statistical and systematic errors. The deconvolution had a mean classification accuracy of 93.06%, and a proton spectroscope systematic error of 8.85 MeV and a statistical error of 45.80 MeV. Following an extensive optimal model search, the ML algorithm achieved a mean classification accuracy of 96.29% and a proton spectroscope systematic error of 11.54 MeV and a statistical error of 30.24 MeV.

The newly developed algorithms were then tested in an experimental environment at DCPT where three different proton energies were tested at three angles of incidence. The particle tracking algorithm agreed with the predictions of the LLM-improved algorithm, with no outliers. The particle classification algorithms were also tested - both the ML and deconvolution algorithms successfully separated the background low energy electrons from the primary beam. Both algorithms displayed the expected Gaussian distributions with means around the predicted energy values.

Next, the algorithms were applied to two separate complex unknown radiation fields: SATRAM and MoEDAL. First, analysis of the SATRAM data was performed. The analysis

70

of this dataset proved problematic and showed clear limitations of Timepix when compared with Timepix3. Despite this difficulty, following statistical manipulation, a physically reasonable proton spectrum was extracted. This is the first time a proton spectrum as measured by SATRAM has been achieved, so comparison with previously produced results was not possible.

Second, analysis of the Timepix3 data at MoEDAL was performed. The following time periods during 2018 were chosen for this analysis: the 25th of November at which PbPb collisions were being performed and the 24th of September during a pp collision period. The IP structure for both periods were analysed demonstrating an increase in accuracy from previous measurements. The IP was successfully isolated and from the increase in accuracy its size could now be resolved, showing a larger IP during PbPb collision periods when compared with the IP during pp collision periods. Clear back-scattering in the PbPb period was obtained, which was consistent with current results but whose origin remains unclear. Once the IP and background had been isolated, particle flux monitoring was performed. The results exhibited a very stable flux during pp collision periods; however, for PbPb collision periods, a sharp increase in flux was obtained which is consistent with PbPb bunch collisions. Interestingly, an initial plateau in the background flux was observed, which was ultimately discovered to be due to the delay in the back-scattering relative to the IP which could be indicative of its origin. Finally, using the previously developed deconvolution and ML algorithms, the field was fully classified for both background and IP. An excellent agreement between both algorithms and a physically reasonable spectrum was obtained, indicating no major systematic errors.

8.1   Future Work

The presented work has an expansive scope, and as such there are numerable potential areas of future work. In relation to the development of classification, classes outside of electron-like and proton-like could be introduced, for example, gamma rays. However, algorithms would need to be adjusted to offset new errors associated with this class. If more experimental data was obtained, there would be the possibility of training the deconvolution algorithm entirely on experimental data; this would dramatically increase the confidence of the results produced. It is unrealistic to suggest a fully ML algorithm for this purpose as a much more diverse dataset would be needed with a wider range of angles and energies, otherwise the developed algorithms would be inherently biased. Similarly, using the distinguishing cluster features discovered in this paper, algorithms could be developed to pursue exotic particles which is the goal in experiments such as MoEDAL [35]. In relation to particle tracking, a similar ML algorithm could be developed for approximating φ values which would further benefit

the tracking of particles within the field. Both the deconvolution and ML algorithms could be extended to account for two Timepix3 detectors in coincidence, which could significantly increase the confidence of tracking. Using this in combination with particle converters could aid in classification and extension of more particle classes [37].

Regarding SATRAM, the comparison of the produced proton spectrum with that of other detectors on the satellite could bestow higher confidence in the results. Following this comparison, considering an approximate proton spectrum has been produced, the bench marking of simulations of low Earth radiation models, such as the AP-9 and the AE-9 models [38], is now possible.

The MoEDAL data analysis has the most possibilities for further development. In relation to the IP, by comparing PbPb and pp IP sizes, an increase in size of the IP was shown. This indicates there is a potential to resolve the size of the IP which previously has only been known approximately [39]. Possible simulation of omnidirectional point sources along the length beam pipe could allow for unfolding of the polar histogram and thus give an accurate determination of the size of the IP. Similarly, following an in-depth investigation into the experimental setup and using the developed deconvolution and ML algorithms, the exact source of back-scattering may be identified.

9 | Contributions to Publications

Some of the methodological work performed within the scope of this thesis has already been disseminated in publications.

9.1 Timepix3 as Solid-State Time Projection Chamber in Particle and Nuclear Physics

The following work was published as Benedikt Bergmann et al. "Timepix3 as solid-state time-projection chamber in particle and nuclear physics". In: *PoS* ICHEP2020 (2021), p. 720. DOI: 10.22323/1.390.0720. URL: https://cds.cern.ch/record/2784918

The capabilities of Timepix3 as a time projection chamber are presented. The 3D trajectory reconstruction algorithms were validated through simulation. For the first time, data taken with Timepix3 in the MoEDAL experiment is presented. The position of the IP relative to the orientation of the detector was approximated using developed algorithms and simulated omnidirectional field response.

My contributions to this work were:

- Processing of Timepix3 data

- Contributions to development of analysis methodology

- Preparation of graphs and figures

9.2 Improved Algorithms for Determination of Particle Directions with Timepix3

The capabilities of Timepix3 to accurately measure the directionality of individual particles of radiation was demonstrated by the development of a large number of analytical, statistical, and machine learning methods. An exhaustive search of the optimal algorithm in areas such as speed, accuracy, and stability was performed. These algorithms were developed and tested using simulated datasets with later application to experiments of known and unknown results.

My contributions to this work were:

- Preparation of the simulated datasets

- Significant contributions to the methodology development of the algorithms used for particle tracking

- Creation of graphs and figures

- Contribution to the writing of the publication

Bibliography

[1] B Bergmann. "Timepix detectors in space: From radiation monitoring in low earth orbit to astroparticle physics". In: *presentation at the 23rd IEEE real-time conference, online* (2022).

[2] Rafael Ballabriga, Michael Campbell, and Xavier Llopart. "Asic developments for radiation imaging applications: The medipix and timepix family". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018). Radiation Imaging Techniques and Applications, pp. 10–23. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2017.07.029`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900217307714`.

[3] W. Shockley. "Currents to Conductors Induced by a Moving Point Charge". In: *Journal of Applied Physics* 9.10 (Oct. 1938), pp. 635–636. DOI: `10.1063/1.1710367`.

[4] X. Llopart et al. "Timepix, a 65k programmable pixel readout chip for arrival time, energy and/or photon counting measurements". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 581.1 (2007). VCI 2007, pp. 485–494. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2007.08.079`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900207017020`.

[5] Benedikt Bergmann. "Detection and separation of x-rays, neutrons and charged particles using hybrid pixel detectors". PhD dissertation. Friedrich-Alexander University Erlangen, 2019.

[6] E Frojdh et al. "Timepix3: first measurements and characterization of a hybrid-pixel detector working in event driven mode". In: *Journal of Instrumentation* 10.01 (Jan. 2015), p. C01039. DOI: `10.1088/1748-0221/10/01/C01039`. URL: `https://dx.doi.org/10.1088/1748-0221/10/01/C01039`.

[7]   Jan Jakubek. "Precise energy calibration of pixel detector working in time-over-threshold mode". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 633 (2011). 11th International Workshop on Radiation Imaging Detectors (IWORID), S262–S266. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2010.06.183`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900210013732`.

[8]   Florian Michael Pitters et al. "Time and Energy Calibration of Timepix3 Assemblies with Thin Silicon Sensors". In: 2018.

[9]   Lukáš Meduna et al. *Real-time Timepix3 data clustering, visualization and classification with a new Clusterer framework*. 2019. DOI: `10.48550/ARXIV.1910.13356`. URL: `https://arxiv.org/abs/1910.13356`.

[10]   T. Holy et al. "Pattern recognition of tracks induced by individual quanta of ionizing radiation in Medipix2 silicon detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 591.1 (2008). Radiation Imaging Detectors 2007, pp. 287–290. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2008.03.074`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900208004592`.

[11]   J. Bartovsky et al. "Morphological classification of particles recorded by the Timepix detector". In: *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*. 2011, pp. 343–348.

[12]   Marine Ruffenach et al. "A New Technique Based on Convolutional Neural Networks to Measure the Energy of Protons and Electrons With a Single Timepix Detector". In: *IEEE Transactions on Nuclear Science* PP (Apr. 2021), pp. 1–1. DOI: `10.1109/TNS.2021.3071583`.

[13]   Benedikt Bergmann et al. "3D track reconstruction capability of a silicon hybrid active pixel detector". In: *The European Physical Journal C* 77.6 (June 2017), p. 421. ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-017-4993-4`. URL: `https://doi.org/10.1140/epjc/s10052-017-4993-4`.

[14]   Benedikt Bergmann et al. "3D reconstruction of particle tracks in a 2 mm thick CdTe hybrid pixel detector". In: *The European Physical Journal C* 79.2 (Feb. 2019), p. 165. ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-019-6673-z`. URL: `https://doi.org/10.1140/epjc/s10052-019-6673-z`.

[15] "Particle tracking and radiation field characterization with Timepix3 in ATLAS". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 978 (2020), p. 164401. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2020.164401`.

[16] P. Mánek et al. "Improved algorithms for determination of particle directions with Timepix3". In: *Journal of Instrumentation* 17.01 (Jan. 2022), p. C01062. DOI: `10.1088/1748-0221/17/01/C01062`. URL: `https://dx.doi.org/10.1088/1748-0221/17/01/C01062`.

[17] Koen Wolters, Simon Spannagel, and Daniel Hynds. "User Manual for the Allpix^2 Simulation Framework". In: (Dec. 2017). URL: `https://cds.cern.ch/record/2295206`.

[18] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/S0168-9002(03)01368-8`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900203013688`.

[19] Geant4 Collaboration. *Guide For Physics Lists*. URL: `https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsListGuide/fo/PhysicsListGuide.pdf` (visited on 04/14/2023).

[20] Sungmin Pak et al. "A Numerical Method to Analyze Geometric Factors of a Space Particle Detector Relative to Omnidirectional Proton and Electron Fluxes". In: *Journal of the Korean Astronomical Society* 51 (Aug. 2018), pp. 111–117. DOI: `10.5303/JKAS.2018.51.4.111`.

[21] G. D'Agostini. *Improved iterative Bayesian unfolding*. 2010. DOI: `10.48550/ARXIV.1010.0632`. URL: `https://arxiv.org/abs/1010.0632`.

[22] Richard Duda, Peter Hart, and David G.Stork. "Pattern Classification". In: vol. xx. Jan. 2001. ISBN: 0-471-05669-3.

[23] Lydia Brenner et al. *RooUnfold: ROOT Unfolding Framework*. `https://gitlab.cern.ch/RooUnfold/RooUnfold`. 2020.

[24] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[25] Leo Breiman. "Random forests". In: *Machine learning* 45 (2001), pp. 5–32.

[26] Tianqi Chen and Carlos Guestrin. "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016. DOI: 10.1145/2939672.2939785. URL: https://doi.org/10.1145%5C%2F2939672.2939785.

[27] Google Services. *Scikit-learn, Machine Learning in Python*. URL: https://keras.io/api/ (visited on 04/06/2023).

[28] David Cournapeau. *Scikit-learn, Machine Learning in Python*. 2007. URL: https://scikit-learn.org/stable/ (visited on 04/06/2023).

[29] D.E. Groom and S.R. Klein. "Passage of particles through matter". In: *The European Physical Journal C - Particles and Fields* 15.01 (Oct. 2000), pp. 163–173. DOI: 10.1007/BF02683419.

[30] Carlos Granja et al. "The SATRAM Timepix spacecraft payload in open space on board the Proba-V satellite for wide range radiation monitoring in LEO orbit". In: *Planetary and Space Science* 125 (2016), pp. 114–129. ISSN: 0032-0633. DOI: https://doi.org/10.1016/j.pss.2016.03.009. URL: https://www.sciencedirect.com/science/article/pii/S0032063316300216.

[31] Nicholas Stoffle et al. "Timepix-based radiation environment monitor measurements aboard the International Space Station". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 782 (2015), pp. 143–148. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.02.016. URL: https://www.sciencedirect.com/science/article/pii/S0168900215001977.

[32] Sebastien Bourdarie and Michael Xapsos. "The Space Radiation Environment". In: 2008.

[33] Marko Markovic. *South Atlantic Anomaly*. 2007. URL: https://commons.wikimedia.org/wiki/File:South%5C_Atlantic%5C_Anomaly.svg.

[34] Glen Cowan. *Error analysis with weighted events*. June 2014. URL: https://www.pp.rhul.ac.uk/~cowan/stat/notes/errors_with_weights.pdf.

[35] B. Acharya et al. "The physics programme of the MoEDAL experiment at the LHC". In: *International Journal of Modern Physics A* 29.23 (Sept. 2014), p. 1430050. DOI: 10.1142/s0217751x14300506. URL: https://www.worldscientific.com/doi/abs/10.1142/S0217751X14300506.

[36] Benedikt Bergmann et al. "Timepix3 as solid-state time-projection chamber in particle and nuclear physics". In: *PoS* ICHEP2020 (2021), p. 720. DOI: 10.22323/1.390.0720. URL: https://cds.cern.ch/record/2784918.

[37] B. Bergmann et al. "ATLAS-TPX: a two-layer pixel detector setup for neutron detection and radiation field characterization". In: *Journal of Instrumentation* 11.10 (Oct. 2016), P10002. DOI: `10.1088/1748-0221/11/10/P10002`. URL: `https://dx.doi.org/10.1088/1748-0221/11/10/P10002`.

[38] Gregory Ginet et al. "AE9, AP9 and SPM: New Models for Specifying the Trapped Energetic Particle and Space Plasma Environment". In: *Space Science Reviews* 179 (Nov. 2013). DOI: `10.1007/s11214-013-9964y`.

[39] CERN Courier. *LHC p collisions, Taking a closer look at LHC*. 2010. URL: `https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_p_collisions` (visited on 03/27/2023).

A | Simulation Calibration

Simulation calibration must be performed each time an aspect of the detector setup is changed. To perform simulation calibration, the entire setup intended for calibration is simulated but the "[DespositionGeant4]" section of the configuration file is replaced with the following configuration:

```
[DepositionPointCharge]
source_type = "point"
model = "spot"
position = 110um 110um 200um # position should correspond
                             # to the detector layer
spot_size = 1um
number_of_charges = 2000 # /2500/2800/...
```

This produces a number of free electrons at the point specified. Then, considering the energy of pair production in silicon is known to be $3.6\ eV$, the amount of charge simulated can be directly related to the energy deposition of a typical simulation with this setup. From these results, a graph of the ToT versus energy deposition is created to calibrate the ToT to energy deposition variables. A linear graph of voltage output versus energy deposition is then made to calibrate the threshold of the experiment. The THL is set to correspond to $4\ keV$ (the typical threshold of detectability in a Timepix3 detector). Calibration curves for the simulation setup used throughout this paper can be seen in Figure A.1. The figure shows clear smooth curves, giving calibration variables as follows,

$$E_{dep} = \exp((\text{ToT}) \times 1.22736 + 0.000613675) \tag{A.1}$$
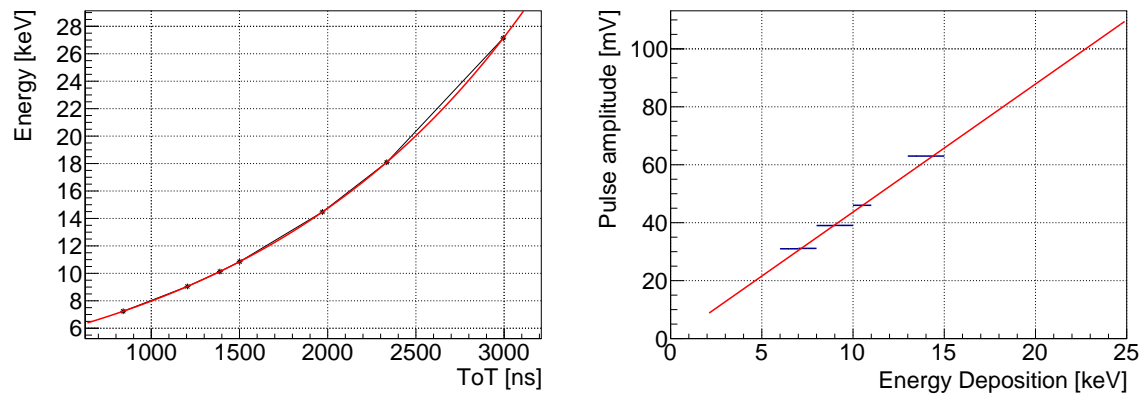
$$\text{THL}_{CSA} = 15 \text{ mV} \tag{A.2}$$

Figure A.1: ToT (left) and threshold (right) calibration curves for simulation of 500 μm silicon detector set to 230 V bias voltage and 80 V depletion voltage. The associated calibration quantities are shown in Equations A.1 and A.2. ToT, Time-over-Threshold.

B | MoEDAL pp Collisional Period Preliminary Analysis

Figures B.1, B.2, and B.3 are the results associated with the prepossessing of the pp collisional period. These results were identical to the analysis of the PbPb collisional period in Section 7.3.4, giving the final results as seen in Figures 7.28 and 7.29.
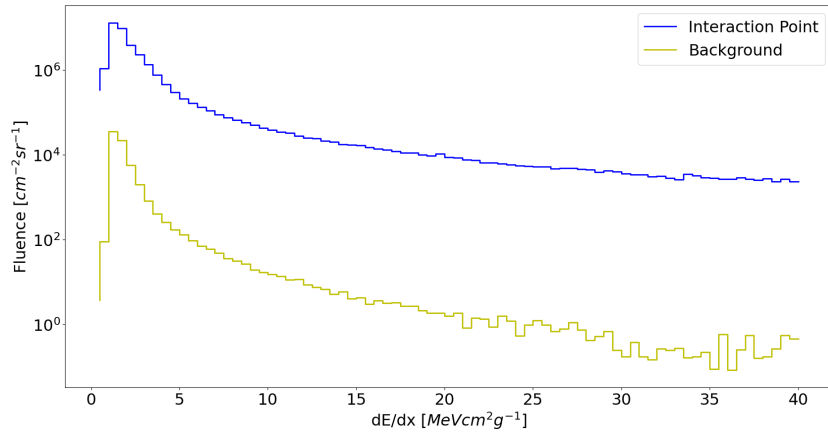

Figure B.1: Stopping power distribution by Left Lower-most, Right Upper-most-improved algorithm [16] for a proton-proton collision period. The interaction point is defined by angle of incidence $\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°]$).
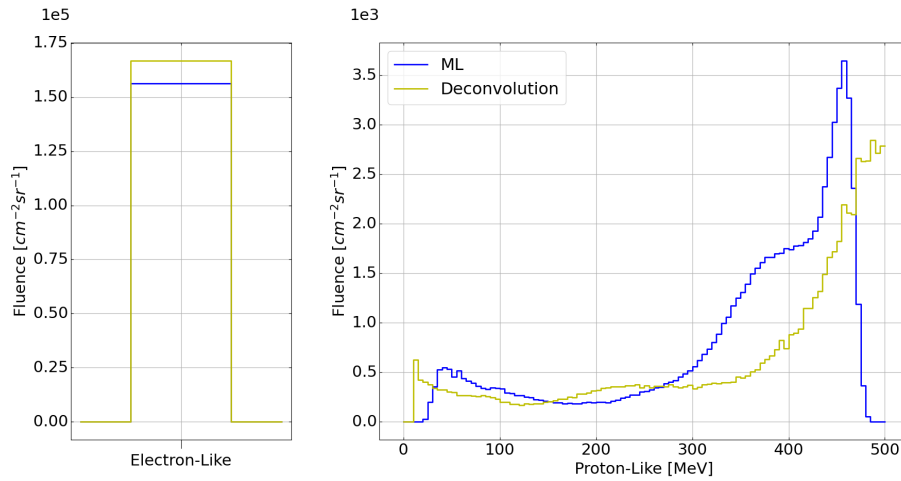
Figure B.2: Results of the particle classification into the classes electron-like and proton-like for the background radiation ($\varphi \notin [135°, 144°]$ and $\theta \notin [50°, 64°]$) at MoEDAL during the proton-proton collision period using the developed deconvolution and machine learning algorithms.
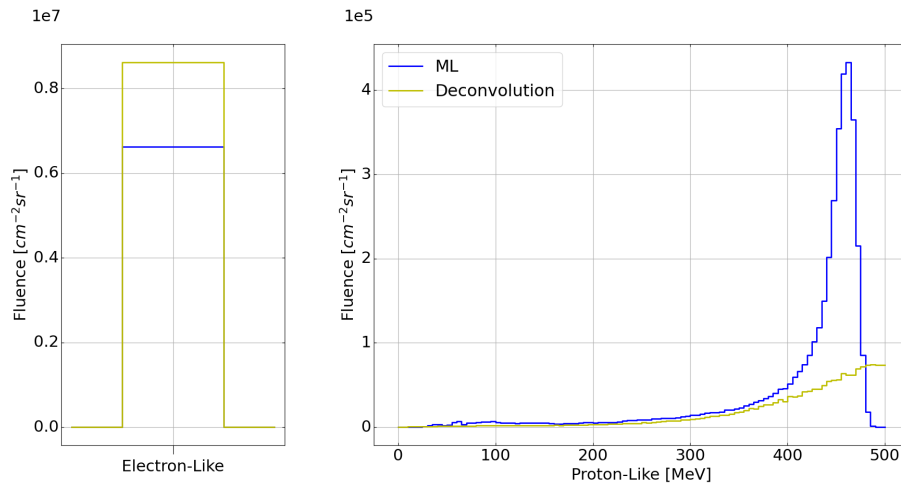

Figure B.3: Classified particle fluence into the classes electron-like and proton-like for interaction point radiation ($\varphi \in [135°, 144°]$ and $\theta \in [50°, 64°]$) at MoEDAL during the proton-proton collisional period using developed deconvolution and machine learning algorithms.