Machine Learning Methods for Antihydrogen Detection.

Lukas M. Golino

Swansea University
Department of Physics

Submitted to Swansea University in fulfilment of the requirements for the Degree
of
Doctor of Philosophy

September 2024

Abstract

Antihydrogen, composing an antiproton and positron, is the only bound state of two antiparticles yet to be synthesised, making for an enticing system to study the purported symmetry of matter and antimatter. As antihydrogen does not occur naturally in the observable universe, any study of this atom requires it to be synthesised in a lab, which the ALPHA experiment is routinely able to do. However, the absolute numbers are small and efficient detection is crucial for the experiment.

To detect these atoms, ALPHA deploys two main annihilation detectors: a silicon vertex detector and a new time projection chamber installed in 2018. A key challenge for both detectors is distinguishing between antimatter annihilations and background events (e.g. cosmic radiation), a task for which machine learning is well suited. Presently, for the silicon vertex detector, this is done with high-level variables, while the time projection chamber has no way of filtering these events.

In the present work, we have developed the first models capable of filtering events in the time projection chamber, which proved vital in the first measurement of the effect of gravity on the motion of antimatter. A first-of-its-kind deep learning model trained on low-level data from the silicon vertex detector has been developed, and it can successfully classify events to a high degree of accuracy. Further, the newest models trained for the silicon vertex detector are presented. The use of these models on real data is included, and all results generated by ALPHA from the 2022-2024 experimental runs will use the models described in this thesis.

Finally, the transverse beam profile in the accelerators throughout CERN (such as the one used to provide antiprotons to ALPHA) is an important metric for successful operation. The significant increase in beam intensities poses a challenge that make the currently deployed correcting magnetic fields undesirable. The possibility of using machine learning to reconstruct beam profiles in the Proton Synchrotron is presented, and a first attempt at applying these models to real data is included which, despite a troubled dataset, shows promising results.

Declarations

- This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

- This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references, and a bibliography is appended.

- I hereby give consent for my thesis, if accepted, to be available for electronic sharing.

- The University's ethical procedures have been followed and, where appropriate, ethical approval has been granted.

Signed:
Lukas Golino

Lukas M. Golino
September 2024

Acknowledgements and Dedication

First, I would like to thank my supervisor, Prof. Niels Madsen, for the opportunity he gave me with this PhD, and for always being there to push me to do more. Your work ethic, spirit, and enthusiasm is something that I hope to one day be able to match. The experience of working at CERN on the ALPHA experiment has truly been wonderful, and I am immensely grateful to have had the opportunity to work alongside such hard-working and inspiring people.

I want to thank Dr. Joseph T. K. McKenna for teaching me everything about the ALPHA experiment and helping guide me through the jungle of `alphasoft`. The patience and encouragement he provided me during my PhD has been instrumental.

For the people at ALPHA who always seem to have the answer: Dr. Ina Carli, Dr. Kurt Thompson, Dr. Chris Ø. Rasmussen, Dr. Adam Powell, Dr. Álvaro Nunes de Oliveira, and Dr. Janko Nauta thank you for always providing it (along with an accompanying explanation).

To all the other summer students, PhD students, post-docs, and professors (of which there are too many to name): thank you for making ALPHA the place it is - I couldn't have asked for a better group.

To Dr. James Storey, thank you for the opportunity to work in the BI group and for offering me the chance to return to CERN after this PhD.

This work and many months of my time at CERN were funded by the AIMLAC CDT for which I am grateful, allowing me to engage directly with the experiment.

I want to thank my school tutor, Mr Jackson, for always showing me the curious side of mathematics, sparking an interest in maths and physics that lasts to this day.

To my grandparents, thank you for the constant support and love over not just the last four years, but my whole life.

I must now, unfortunately, thank my brothers for their support, as well as that of my family and friends who pushed me into doing this.

To Neele, thank you for sharing the last four years of your life with me and for your constant support, advice, and daily motivation throughout, which have all been invaluable.

Finally, I dedicate this thesis to my parents, Alison and Michelo Golino. Their sacrifices have afforded me the life I am lucky enough to lead today, and without their support, encouragement, and motivation none of this would have happened. Thank you.

Pencoed, Cymru - Genève, Suisse
September 2024

Contents

List of Figures

List of Tables

List of Acronyms, Abbreviations, and Symbols

List of Acronyms

Table 0.1: List of acronyms and their first appearance.

SM	The **S**tandard **M**odel	Chapter 1	Page 2
ALPHA	**A**ntihydrogen **L**aser **PH**ysics **A**pparatus	Chapter 1	Page 2
CERN	**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire	Section 1.1	Page 4
AA	**A**ntiproton **A**ccumulator	Section 1.1	Page 4
AC	**A**ntiproton **C**ollector	Section 1.1	Page 4
LEAR	**L**ow **E**nergy **A**ntiproton **R**ing	Section 1.1	Page 4
GR	**G**eneral **R**elativity	Section 1.2	Page 6
WEP	The **W**eak **E**quivalence **P**rinciple	Section 1.2	Page 6
CPT	**C**harge-**P**arity-**T**ime Symmetry/Theorem	Section 1.2	Page 6
PET	**P**ositron **E**mission **T**omography	Section 1.2	Page 6
UHV	**U**ltra-**H**igh **V**acuum	Section 1.4	Page 10
ATHENA	**A**n**T**i**H**ydrog**EN** **A**pparatus	Section 2.2	Page 12
ATRAP	**A**ntihydrogen **TRAP**	Section 2.2	Page 12
SRR	**S**pontaneous **R**adiative **R**ecombination	Section 2.2	Page 13
LIR	**L**aser **I**nduced **R**ecombination	Section 2.2	Page 14
GBAR	**G**ravitational **B**ehaviour of **A**ntihydrogen at **R**est	Section 2.2	Page 15
TBR	**T**hree-**B**ody **R**ecombination	Section 2.2	Page 15
DCA	**D**istance of **C**losest **A**pproach	Section 2.2	Page 16
PM	**P**enning-**M**almberg [trap]	Section 2.3.2	Page 22
IP	**I**offe-**P**ritchard [trap]	Section 2.3.3	Page 27
FSI	**F**inal **S**tate **I**nteraction	Section 2.4	Page 31
AD	**A**ntiproton **D**ecelerator	Chapter 3	Page 33
MCP	**M**icro-**C**hannel **P**late	Section 3.1.2	Page 35
CCD	**C**harge **C**oupled **D**evice	Section 3.1.2	Page 36
PMT	**P**hoto**m**ultiplier **T**ube	Section 3.2	Page 38
SiPM	**Si**licon **P**hoto**m**ultiplier	Section 3.2	Page 38

Continued on next page

Table 0.1: List of acronyms and their first appearance. (Continued)

SPAD	**S**ingle-**P**hoton **A**valanche **D**iode	Section 3.2	Page 38
SNR	**S**ignal-to-**N**oise **R**atio	Section 3.2	Page 38
KE	**K**inetic **E**nergy	Section 3.4	Page 39
PS	**P**roton **S**ynchrotron	Section 3.5.1	Page 42
BGT	**B**uffer-**G**as **T**rap	Section 3.5.2	Page 44
ELENA	**E**xtra **L**ow **EN**ergy **A**ntiprotons	Section 3.6	Page 46
CT	**C**atching **T**rap	Section 3.7	Page 49
SDR	**S**trong **D**rive **R**egime	Section 3.8.2	Page 55
EVC	**EV**aporative **C**ooling	Section 3.8.4	Page 58
MB	**M**axwell-**B**oltzmann	Section 3.8.4	Page 58
SDREVC	**S**trong **D**rive **R**egime **EV**aporative **C**ooling	Section 3.8.5	Page 59
ECR	**E**lectron **C**yclotron **R**esonance	Section 3.8.5	Page 60
RCT	**R**e**C**atching **T**rap	Section 3.10.1	Page 62
ASACUSA	**A**tomic **S**pectroscopy **A**nd **C**ollisions **U**sing **S**low **A**ntiprotons	Section 3.11	Page 66
SVD	**S**ilicon **V**ertex **D**etector	Chapter 4	Page 72
TPC	**T**ime **P**rojection **C**hamber	Chapter 4	Page 72
SDD	**S**emiconducter **D**iode **D**etectors	Section 4.1.1	Page 72
MIP	**M**inimum **I**onizing **P**articles	Section 4.1.7	Page 82
ASIC	**A**pplication **S**pecific **I**ntegrated **C**ircuit	Section 4.2	Page 84
BV	**B**arrel **V**eto	Section 4.4	Page 92
TOF	**T**ime-**O**f-**F**light	Section 4.4.2	Page 95
VMC	**V**irtual **M**onte **C**arlo	Section 5.1.1	Page 102
VGM	**V**irtual **G**eometry **M**odel	Section 5.1.1	Page 102
DSAT	**D**own-**S**tream **A**tom **T**rap [stick]	Section 5.1.3	Page 106
ML	**M**achine **L**earning	Chapter 6	Page 119
k-NN	**k**-**N**earest **N**eighbour	Section 6.2.1	Page 126
BDT	**B**oosted **D**ecision **T**rees	Section 6.2.5	Page 130
SVM	**S**upport **V**ector **M**achines	Section 6.2.6	Page 131
LR	**L**inear **R**egression	Section 6.3.1	Page 132
MSE	**M**ean-**S**quared **E**rror	Section 6.3.1	Page 132
RR	**R**idge **R**egression	Section 6.3.2	Page 133
KRR	**K**ernel **R**idge **R**egression	Section 6.3.3	Page 133
RBF	**R**adial **B**asis **F**unction	Section 6.3.3	Page 134
SVR	**S**upport **V**ector **R**egression	Section 6.3.4	Page 135

Continued on next page

Table 0.1: List of acronyms and their first appearance. (Continued)

NN	**N**eural **N**etwork	Section 6.4	Page 136
ANN	**A**rtificial **N**eural **N**etwork	Section 6.4	Page 136
MLP	**M**ulti-**L**ayer **P**erceptron	Section 6.4	Page 137
ReLU	**Re**ctified **L**inear **U**nit	Section 6.4.1	Page 138
ROC	**R**eceiver **O**perating **C**haracteristic	Section 6.5.1	Page 143
AUC	**A**rea **U**nder the **C**urve	Section 6.5.1	Page 143
BI	**B**eam **I**nstrumentation	Chapter 7	Page 145
IPM	**I**onisation **P**rofile **M**onitor	Chapter 7	Page 145
FCC	**F**uture **C**ircular **C**ollider	Section 7.4.2	Page 150
BGI	**B**eam **G**as **I**onisation	Section 7.5	Page 151
VIPM	**V**irtual **I**onisation **P**rofile **M**onitor	Section 7.6.1	Page 152
PDF	**P**robability **D**ensity **F**unction	Section 7.6.2	Page 154
MVA	**M**ulti-**V**ariate **A**nalysis	Section 8.1	Page 166
PCC	**P**earson **C**orrelation **C**oefficient	Section 8.4.1	Page 174
FOM	**F**igure-**O**f-**M**erit	Section 8.6	Page 185
CNN	**C**onvolutional **N**eural **N**et	Section 10.2	Page 208
SEA-1	**S**ignal **E**fficiency **A**t **1**% [Background]	Section 10.3	Page 211

List of Abbreviations

Table 0.2: List of abbreviations and their first appearance.

Ps	Positronium	Section 2.2	Page 15
AD hall	Antimatter facility housing the antimatter experiments at CERN, Geneva	Section 3.6	Page 46
rf	radio-frequency	Section 3.6.1	Page 46
HVA/B	High voltage A/B - the high voltage electrodes surrounding the PM trap in the CT	Section 3.7	Page 50
Ekick	Electron kick-out	Section 3.8.3	Page 58
Smerge	Slow merge [of positrons and antiprotons for antihydrogen production]	Section 3.12.1	Page 68
AdaBoost	Adaptive boosting	Section 6.2.3	Page 129
Bagging	**B**ootstrap **agg**regat**ing**	Section 6.2.3	Page 129
PS-BGI-82	The horizontal BGI installed in the PS	Chapter 7	Page 145
ppp	Particles per profile, in respect to BGIs	Section 7.5	Page 151

List of Symbols

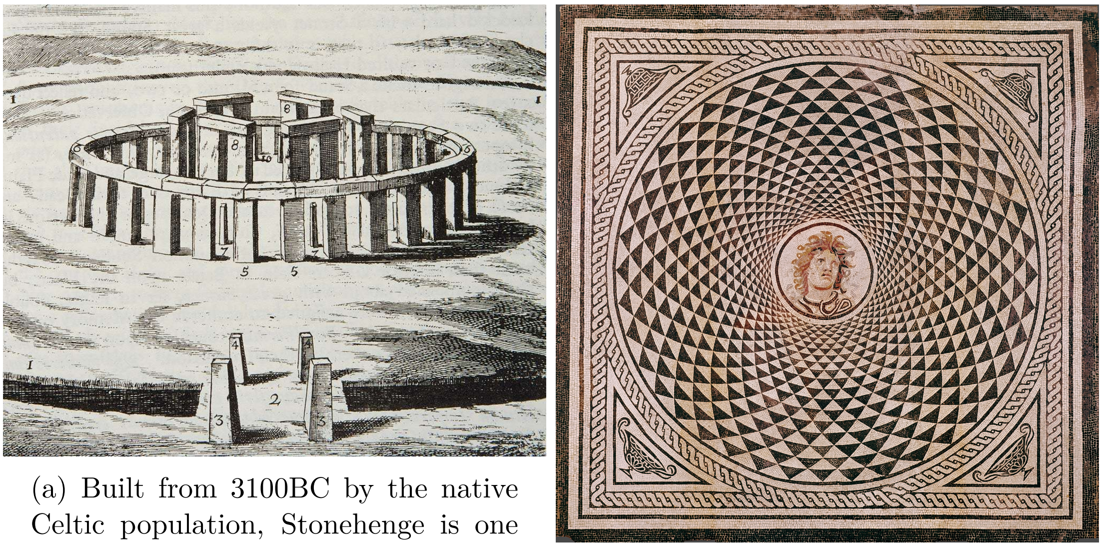
Table 0.3: List of symbols and their first appearance if applicable.

(x, y, z)	Cartesian Coordinates	N/A	N/A
(r, θ, ϕ)	Polar Coordinates	N/A	N/A
h	Planck constant $(4.135667696 \times 10^{-15}\text{eV/Hz})$	Equation 1.1	Page 3
$\hbar$	Reduced Planck constant $(h/2\pi)$	Equation 1.1	Page 3
$\bar{H}$	Antihydrogen	Section 1.2.2	Page 7
H	Hydrogen	Section 1.2.2	Page 7
e^+	Positron	Section 2.1	Page 11
e^-	Electron	Section 2.1	Page 11
$\bar{p}$	Antiproton	Section 2.1	Page 11
p	Proton	Section 2.1	Page 11
k_B	Boltzmann's constant $(1.380649 \times 10^{-23} J/K)$	Section 4	Page 16
$\vec{E}$	Electric field	Section 2.3.1	Page 18
$\vec{B}$	Magnetic field	Section 2.3.1	Page 18
ϵ_0	Vacuum permittivity $(8.8541878128(13) \times 10^{-12} F/m)$	Section 2.3.1	Page 21
λ_D	Debye Length	Section 2.3.2	Page 23
μ_B	Bohr Magneton $(e\hbar/2m_e)$	Section 2.3.3	Page 26
ν	Neutrino	Section 2.4	Page 29
$\bar{\nu}$	Antineutrino	Section 2.4	Page 29
π^+	Positively charged pion	Section 2.4	Page 31
π^-	Negatively charged pion	Section 2.4	Page 31
π^0	Neutral pion	Section 2.4	Page 31
N_A	Avogadro's constant $(6.02214076 \times 10^{23}\text{mol}^{-1})$	Section 4.1.7	Page 82
M_u	Molar mass constant $(\text{M}(\text{C}^{12})/12$ - one twelfth the molar mass of carbon-12)	Section 4.1.7	Page 82

Chapter 1

Introduction

For thousands of years, humans have been fascinated by symmetry, the proof of which is reflected in our architecture (Figure 1.1a), art (Figure 1.1b), pottery, and even music. The reasons for our captivation are likely due to its prevalence in nature as a signal for health and harmony, and it may be associated with the way the brain processes visual information, mediated by evolution [1]. Perhaps initially an aesthetic pursuit for humans, symmetry has since taken up a central role in how we actually understand the universe around us and even how we explain its inner workings.



(a) Built from 3100BC by the native Celtic population, Stonehenge is one of the earliest examples of symmetry in human architecture. This reconstructed sketch was drawn in 1725 by Inigo Jones.

(b) Roman floor mosaic: *Head of Medusa*, ca. 115–150AD, stone tesserae. The Romans were famous for using symmetry in all aspects of life.

Figure 1.1: Examples of symmetry in early human architecture and art.

In physics, the term symmetry has become synonymous with *invariance*; an absence of change under specific kinds of transformations [2]. So when, in 1918, Noether proved the link between invariance of various actions under a symmetric transformation and the laws of conservation already recognised at the time (namely classical mechanics and electrodynamics) [3], the concept was cemented in the minds of physicists as a fundamental property of nature. Since then, this concept of symmetry in nature has become a powerful tool of theoretical physics, and it has since been shown that many laws of nature originate in symmetries of some kind.

Three significant discrete symmetries currently being studied in fundamental physics are:

- The charge conjugation $\mathcal{C} : q \mapsto -q$, where all fundamental charges change sign;
- the parity transformation $\mathcal{P} : x \mapsto -x$ which describes an inversion of spatial coordinates;
- and time reversal $\mathcal{T} : t \mapsto -t$.

The combination of which is known as CPT symmetry and describes a particle (p) turning into its antiparticle ($\bar{p}$). Individually, these are sometimes known as *near-symmetries* as each is broken in the universe. However, the CPT theorem says that CPT symmetry must hold for all physical phenomena, which implies, among many other things, that particles and antiparticles should demonstrate the same (or sign-opposite) properties when probed. The standard model (SM) is the closest to a *theory of everything* that we have today, accurately describing three of the four fundamental forces of nature[1], and classifying all known elementary particles; and the CPT theorem[2] is a cornerstone of this model. It is through this theorem that the study of antimatter is intimately linked with the study of symmetry as a concept, and therefore, our understanding of antimatter is inherently linked with the standard model and how we understand the laws of nature as a whole.

The Antihydrogen Laser Physics Apparatus (ALPHA) experiment is split into two main components: ALPHA-2, the goal of which is to test CPT symmetry by probing antimatter directly, and ALPHA-g, to test the theory of gravity on antimatter, one of the few missing pieces of the standard model. The apparatus used for these studies and the methods by which the particles are detected and analysed will be the main focus of the following thesis, with a specific focus on how machine learning is used to classify events as accurately as possible.

---

[1] The electromagnetic, weak, and strong forces – gravity is not described in the standard model.

[2] The CPT theorem says that any Lorentz invariant local quantum field theory with a Hermitian Hamiltonian must have CPT symmetry.

1.1 History of Antimatter

In 1928, Dirac derived what is now known as the “Dirac equation” [4] [3]:

$$(i\hbar\gamma^{\mu}\partial_{\mu} - mc)\psi = 0 \tag{1.1}$$

as a relativistic version of the Schrödinger wave equation:

$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle. \tag{1.2}$$

This result, however, allows for negative energy solutions to the already well-established energy-momentum relation:

$$E^2 = p^2c^2 + m^2c^4. \tag{1.3}$$

Dirac commented on this supposed discrepancy in his work, stating:

> “The second difficulty in Gordon’s interpretation arises from the fact that if one takes the conjugate imaginary of equation (1)[4], one gets [(1) in another form] which is the same as one would get if one put $-e$ for e. The wave equation (1) thus refers equally well to an electron with charge e as to one with charge $-e$.” [4]

While this quote already seems to suggest the idea of a negative electron, it wasn’t until 1931 [5] when Dirac formally proposed the idea of an antiparticle, suggesting that the *negative-energy* solutions might instead correspond to some new form of particle, equal to the electron in mass but opposite in charge; an “anti-electron”.

In 1933, Anderson was able to observe this particle for the first time by examining images of cosmic rays passing through cloud chambers under the influence of a magnetic field [6]. In these images, some of the tracks produced match the charge-to-mass ratio of the electron but curve in the opposite direction than those produced by electrons, indicating that the observed particle has the same mass as the electron but an opposite charge; an anti-electron- or “positron” as Anderson

---

[3] Though Dirac never wrote the equation in this form, his result was analogous.

[4] Throughout this quote (1) refers to Dirac’s form of our Equation 1.1

named it.

Both of these discoveries earned Nobel Prizes, one for Dirac in 1933 “for the discovery of new productive forms of atomic theory,”[5] and one for Anderson in 1936 “for his discovery of the positron”.

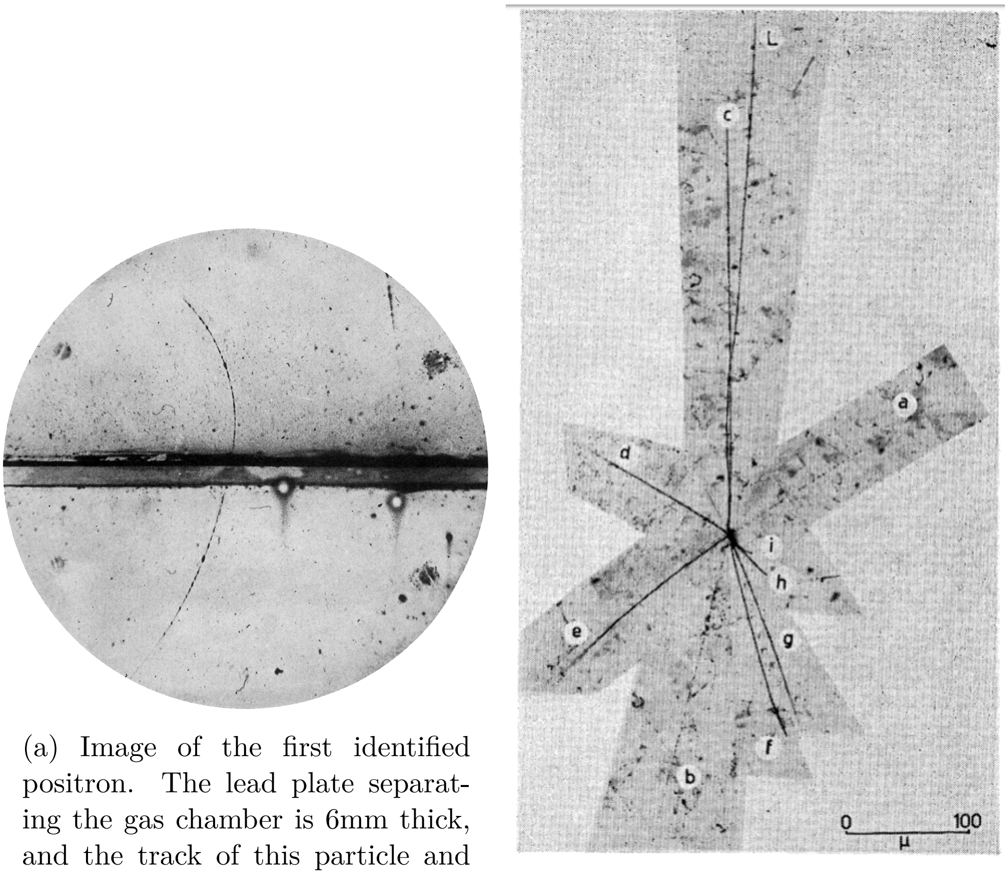
While the first reference to the “antiproton” also occurs in Dirac’s 1931 article [5], it took until 1955 for their discovery at the Bevatron in Berkeley, California. Between the years of 1954 and 1993, the Bevatron was operated as a proton synchrotron, allowing high-energy collisions between proton beams and a copper target, reaching collisions of up to 6.2GeV[6], the calculated optimum energy for producing antiprotons. A paper titled “Observation of antiprotons” by Owen Chamberlain, Emilio Segrè, Clyde Wiegand, and Thomas Ypsilantis [7] was published in November 1955 announcing the discovery of a new particle which, much like the positron, was found to be identical to the proton except with negative charge. To detect the antiprotons, three scintillating panels and two Cherenkov counters were deployed along a secondary beamline beyond the copper target, designed to observe the velocity of particles that passed through them, the resulting signal of which (curve and momentum) could have only come from an antiproton. Another Nobel Prize was awarded for this discovery to Segrè and Chamberlain in 1959 “for their discovery of the antiproton”. By 1996, CERN in Geneva, Switzerland, had been creating and accumulating antiprotons using a similar method for years, albeit with relatively high energies. After the machines required for this process: the Antiproton Accumulator (AA), the Antiproton Collector (AC), and the Low Energy Antiproton Ring (LEAR) were shut down, the community of low energy antimatter scientists requested a new ring that would enable them to study antiprotons at lower energies. The proposal for the “Antiproton Decelerator (AD)” promised dense beams of over 10^7 antiprotons per minute with a momentum of $\sim$ 100MeV/c in 200ns bunches [9]. The proposal was approved on 6 February 1997 [10], and operation began, on the site of the old AA and AC, in the year 2000.

From here antimatter research began to flourish, with ATHENA creating thousands of “cold” antihydrogen atoms in 2002 [11], its successor, ALPHA, initially trapping them in 2010 [12], confining them for over 1000 seconds in 2011 [13]

---

[5] Dirac shared this Nobel Prize with Schrödinger.

[6] Or 6.2 billion electron volts (BeV), hence the name Bevatron.



(a) Image of the first identified positron. The lead plate separating the gas chamber is 6mm thick, and the track of this particle and its secondary indicates that it is a positron. This image and other similar ones were published in [6].

(b) Antiproton star-shaped annihilation captured in emulsion. Figure from [8].

Figure 1.2: Images from the discoveries of the positron and antiproton.

and then performing precision spectroscopy of the 1S-2S transition [14], the 1S-2P (Lyman-α) transition [15] in 2018, and demonstrating successful laser cooling on the trapped antihydrogen [16] in 2022. More recently, in 2023, the ALPHA collaboration, including the author was able to experimentally measure the value of $\bar{g}$ (the local acceleration of antimatter towards the Earth) to within 25% [17], demonstrate consistent and controllable laser cooling, increase antihydrogen production to stacks of 15,000 atoms at a time, and measure the hyperfine structure, 1S-2S transition frequency, 2S-4P frequency, and the 2S-2P Lamb shift transitions.

1.2 Motivation

Despite passing all tests of its validity so far, there are some major gaps within the standard model. For example, it cannot explain the baryon asymmetry problem (the observed imbalance of matter and antimatter), it does not incorporate the theory of gravitation as per general relativity, nor does it account for the universe's accelerating expansion.

In order to understand the final force, gravity, we must turn to Einstein's theory of general relativity (GR) [18] and, similar to the CPT theorem, which claims that matter and antimatter share the same properties, the weak equivalence principle (WEP), predicts that matter and antimatter should gravitate in the same way.

A violation of either of these laws could potentially explain, or at least demonstrate, some of the issues with our current theory of everything.

Most antimatter studies consist of these types of particle/antiparticle comparisons; for example comparisons of the charge, mass, and g-factor of the electron/positron pair [19–21]; or of the charge, mass, and magnetic moment of the proton/antiproton pair [22, 23]. Since none of these measurements have uncovered a statistically significant difference between these properties, they must act as a form of validation of the CPT theorem and, therefore, the SM. However, the baryon asymmetry problem still cannot be accounted for regardless of these findings, potentially pointing to some undiscovered fundamental difference between matter and antimatter that made the universe favour one over the other.

Beyond these mysteries, antimatter has practical applications for which it is important to understand its properties. The ability to monitor and manipulate it accurately, which is central to this field of antimatter research, is vital for use in any practical sense. A typical example of these applications is positron emission tomography (PET), which is a form of medical imaging relying on positron-electron annihilations to build an image. Additionally, antiprotons have been calculated to potentially be effective in treating certain forms of cancer, much like currently employed ion therapy methods [24].

With all these considerations in mind, it is clear that the study of antimatter holds significant importance and relevance in modern physics and beyond, and any inconsistency of the SM found would represent one of the biggest breakthroughs in the history of physics.

1.2.1 Antihydrogen as a Test Subject

To test these fundamental symmetries, many groups are looking at antiprotons, positrons, or positronium (the bound state between an electron and positron). However, ALPHA uses antihydrogen for its studies, and the reasons for this are numerous:

1. It is the simplest pure bound-state antimatter atom, and atoms allow for spectroscopy.
2. It is neutral, preventing the effect of the atom's charge from dominating the other incredibly sensitive properties we would like to measure.
3. It corresponds to the very well-known hydrogen atom, providing ample resources for comparison.
4. It is (again assuming a certain similarity with hydrogen) stable.
5. Given its small magnetic moment, it is relatively easy to confine.

Given enough antihydrogen, trapped or in flight, lots of experimental opportunities arise, generally involving performing spectroscopy on the atoms or measuring the gravitational interactions of antimatter.

1.2.2 Spectroscopy

Comparing antihydrogen ($\bar{H}$) with hydrogen (H) via spectroscopy is beneficial for many reasons. For example, the $1S-2S$ transition is an ideal candidate for precision comparisons of $\bar{H}$ and H because it has a sharp natural line width of ~ 1.3Hz, and has been measured in hydrogen very precisely ($f_{1S-2S} = 2,466,061,413,187,018(10)$Hz with a relative uncertainty of 4.2×10^{-15} using Doppler-free two-photon spectroscopy [25]). This result was later confirmed with a caesium fountain [26] and allows for a very precise comparison when performing spectroscopy on $\bar{\mathrm{H}}$. Beyond this, other transitions in hydrogen have been measured to a high degree of accuracy that are also potential candidates for direct comparisons, including the $2S_{1/2} - 2P_{1/2}$ Lamb shift (measured to be $f_{2S_{1/2}-2P_{1/2}} = 1057.8446(29)$MHz with a relative uncertainty of 3.2×10^{-6} [27]), and the ground state hyperfine splitting (measured to be $1,420,405,751.7662(30)$Hz with relative uncertainty of 2.1×10^{-12} [28]). Each of these has been measured in $\bar{H}$ this year (with contribution from the author) and was analysed using the machine learning models and analysis techniques described later in this thesis, the details of which are currently being reviewed prior to publication.

1.2.3 Antimatter and Gravity

One of the attributes of our current theory of gravity, general relativity, is the universality of free-fall, which states that gravitational mass and inertial mass are the same. This principle is known as the weak equivalence principle and, if true, implies matter and antimatter should respond to gravity in the same way.

Before 2023, the gravitational interaction between antimatter and matter had not been directly measured [29], but the fact that GR and the SM can not be unified again points to some missing piece. While the theoretical consensus is that any mass must be attracted to mass regardless of charge [30], there do exist major consequences should antimatter be repelled by normal matter instead of attracted. These "antigravity" theories have been discussed at length by other authors [31–34]; however, until a direct measurement of the interaction was successfully performed these proposals could not be ruled out entirely.

This interaction was indeed measured with $\bar{H}$ in 2023 [17] with contributions from the author and was analysed using the techniques to be described throughout this thesis.

1.3 Overview

This thesis will describe the first-ever direct measurement of the effect of gravity on the motion of antimatter and the first machine learning models to be deployed utilising data from the newly commissioned time projection chamber detector (see Sec. 4.4.1) used for detection (the training and analysis of which was the responsibility of the author). Further, the models used in the silicon vertex detector will be presented with a first of its kind deep neural network, trained to work on low-level raw detector signals, being explored.

Chapter 2 will discuss the theory behind the formation, trapping, and manipulation of antimatter particles used in the ALPHA experiment. Chapter 3 describes the apparatus specific to ALPHA in more detail and explains the processes used in the production and trapping of antihydrogen. Chapter 4 turns to the detection of antihydrogen using two tracking detectors, describing the theory behind their implementation and the reconstruction algorithms used to convert the raw detector output to event information. Chapter 5 will present simulations used to calibrate signals on counting detectors to absolute numbers within the apparatus. Then, Chapter 6 describes the general theory behind some of the machine learning models tested and used. Chapter 7 will apply these models to beam reconstruction in a beam profile monitor installed at CERN, and Chapter 8 applies these models to data collected from the ALPHA detectors quantifying which performs the best. Chapter 9 will apply these models to previously unseen validation dataset, measuring the model performance, and finally, 10 presents the first attempt at deep learning for event classification at ALPHA.

As this thesis contains topics from many aspects of the scientific landscape care is taken not to take knowledge for granted and, though some discussions may seem elementary to some, this approach was taken to ensure readers from any of the fields contained in this thesis will not lack some of the field-specific knowledge required to understand the whole document.

1.4 Author Contribution

The ALPHA collaboration is a group with a diverse set of experts and scientists from different disciplines, and all scientific contributions by ALPHA are a result of teamwork. The author of this thesis spent 30 months on-site at CERN with the experiment working on a range of activities spanning various fields. This includes the commissioning and installation of the time projection chamber detector for the ALPHA-g experiment, the maintenance of the silicon vertex detector used for the ALPHA-2 experiment, the production and implementation of online data analysis tools, as well as various other software-related tasks such as offline analysis tools, and developing machine learning models. Like all members of ALPHA, shift work is a big part of being on-site and usually consists of running the experiment itself, analysing data live, and optimising procedures by scanning parameters or tuning experiments. Other day-to-day operations range from cryogenic handling to replacing broken electronics or ultra-high vacuum (UHV) valves. The author was run coordinator for one week in August 2024, acting as the lead responsible for the direction of the experiment during the week, ensuring all activities are conducted according to plan. Further to this, the author spent six months working with the CERN beam instrumentation group using machine learning to reconstruct profiles obtained from a specific instrument, the details of which are reported in Chapter 7.

The work presented in this thesis is additionally reported in the following publication with the author's participation:

- Anderson, E.K., et al. Observation of the effect of gravity on the motion of antimatter. *Nature* **621**, 716–722 (2023).

Chapter 2 Antihydrogen Theory & Formation

Experimental work with antimatter requires an understanding of many different fields of physics, the basics of which are important to be established.

This chapter will discuss the atomic theory behind the antihydrogen atom, the possible mechanisms behind its formation during synthesis, how the resulting atoms are trapped, and the process by which they annihilate. There is a focus on three-body recombination as a formation mechanism due to its importance in the ALPHA production cycle, as well as the specific traps used by the ALPHA collaboration.

2.1 Antihydrogen in Flight

The antihydrogen atom is one of the most basic antimatter systems, consisting of a single positron (e^+) and an antiproton ($\bar{p}$) in a bound state; mirroring hydrogen which contains one electron (e^-) and one proton (p).

There are multiple mechanisms by which trappable antihydrogen can form, involving the mixing of cold antiproton and positron plasmas; however, the first antihydrogen formed was detected in flight at LEAR, CERN in 1995 [35] (also slightly later at Fermilab's Antiproton Accumulator in 1998 [36]). An antiproton passing through the Coulomb field of a nucleus will create e^+/e^- pairs in a process known as "pair production," and occasionally, the antiproton will recapture the positron in a bound state, forming antihydrogen. Both experiments passed antiprotons travelling with a transverse momentum of 1.94GeV/c and $\sim$ 6GeV/c respectively through a gas target to create antihydrogen in this way. At the lowest order, two mechanisms were described for this process for a nucleus of charge Z [35]:

$$\overline{p}\mathrm{Z} \rightarrow \overline{p}\gamma\gamma\mathrm{Z} \rightarrow \overline{p}\mathrm{e}^+\mathrm{e}^-\mathrm{Z} \rightarrow \overline{\mathrm{H}}\mathrm{e}^-\mathrm{Z} \tag{2.1}$$

$$\overline{p}\mathrm{Z} \rightarrow \overline{p}\gamma^*\mathrm{Z} \rightarrow \overline{p}\mathrm{e}^+\mathrm{e}^-\mathrm{Z} \rightarrow \overline{\mathrm{H}}\mathrm{e}^-\mathrm{Z} \tag{2.2}$$

The first process (Equation 2.1) is a two-photon ($\gamma\gamma$) mechanism for $\bar{\mathrm{H}}$ production and the second process (Equation 2.2), which involves a *virtual photon* (γ^*), is much more likely to occur, with a cross-section roughly three times greater than the previous [35]. The resulting cross-section is characterised by a Z^2 and a $\ln E_{\bar{p}}/m_{\bar{p}}$ dependence, where $E_{\bar{p}}$, and $m_{\bar{p}}$ is the energy and mass of the $\bar{p}$ respectively; and $E_{\bar{p}}/m_{\bar{p}}$ is collectively known as the *Lorentz factor* of the $\bar{p}$. For the beam used at CERN in 1995 [35], this cross-section was calculated to be $\sim 2\text{pb} \times Z^2 \approx 6\text{nb}$ given the Xenon target used. The overall process which can be stated simply as: $\bar{p}\mathrm{Z} \rightarrow \bar{\mathrm{H}}e^-\mathrm{Z}$ has a cross-section $\sim 4\mathrm{Z}^2$ for a beam of momentum greater than 6GeV/c, which is not particularly high.

After the antihydrogen was formed in this experiment, it has the same momentum as the proton beam it was formed with (1.94GeV) and 10 meters later this beam impacts three consecutive silicon counters and the resulting annihilation products were detected by both the counters and a cylindrical six-fold segmented sodium iodide (NaI) detector for the two 511KeV photons that are produced as a result of electron/positron annihilations (see Sec. 2.4).

While both of these experiments observed 9 and 99 $\bar{\mathrm{H}}$ atoms respectively, in-flight antihydrogen production results in $\bar{\mathrm{H}}$ samples that are too energetic to be trapped and, as such, using in-flight antihydrogen like this is not ideal for spectroscopy and measurements on gravity.

2.2 Cold Antihydrogen

The production of in-flight antihydrogen in the mid-90s was an important stepping stone in the field of antimatter research; however, precision measurements require cold, stable antihydrogen. This requires trapping both $\bar{p}$ and e^+ and mixing them in a slow and controllable manner instead of in-flight production. In 2002, both the ATHENA and ATRAP experiments succeeded in the dynamic recapture and cooling of $\bar{p}$ plasmas in a Penning trap (see Sec. 2.3.2 for Penning traps) and mixing them with e^+ plasmas to form much colder antihydrogen [11, 37]. Producing antihydrogen in this manner potentially allows for trapping and results in higher antihydrogen production rates, sometimes peaking at multiple 100s per second. This amount of antihydrogen allowed for studies on the temperature dependence of $\bar{\mathrm{H}}$ formation, the cooling dynamics, as well as temperature modulated and laser stimulated production [38–43]. Additionally, the silicon vertex detector of the ATHENA experiment allowed for 3D reconstruction of events [44] and therefore studies on the spacial and temperature profiles of the produced $\bar{\mathrm{H}}$ [45]. Though the ATHENA experiment was completed in 2005, the detector described in [44] was the basis for the silicon vertex detector used by the ALPHA collaboration.

There are numerous mechanisms by which mixing plasmas of e^+ and $\bar{p}$ can form $\bar{\mathrm{H}}$, all of which are low-energy processes, therefore requiring low-energy plasmas and resulting in relatively cold antihydrogen. Regardless of the method however, each form of antihydrogen production requires three ingredients: e^+, $\bar{p}$ and some third party to carry away the binding energy of the process associated with the formation of $\bar{\mathrm{H}}$.

The most commonly discussed mechanisms of antihydrogen production are:

1. **Spontaneous radiative recombination (SRR)**, which occurs when a photon generated via spontaneous emission carries away the binding energy of the interaction (this process is the inverse of photoionisation). The resulting interaction can be described by:

   $$e^+ + \bar{p} \to \bar{\mathrm{H}} + \hbar\omega \tag{2.3}$$

   where $\hbar$ is Planck's constant, and ω is the angular frequency of the emitted photon. Due to the reliance on this spontaneously emitted photon, the cross-section of this interaction is small and was calculated to be

   $$\sigma_{SRR} = (2.1 \times 10^{-22}\mathrm{cm}^2)\frac{E_0^2}{nE_p(E_0 + n^2E_p)} \tag{2.4}$$

   where n is the energy level of the $\bar{\mathrm{H}}$, E_0 is the binding energy of the ground state positron around the antiproton (e.g. 13.6eV for $n = 1$), and $E_p = \frac{1}{2}mv_p^2$ is the kinetic energy of the positron-antiproton centre-of-mass frame [46].

   Alternatively, this cross-section can be given as a rate per second per antiproton [47]:

   $$\Gamma_{\mathrm{SRR}} = \left(3 \times 10^{-11}\right)\ n\ \sqrt{\frac{4.2\ \mathrm{K}}{T_{e^+}}} \tag{2.5}$$

where T and n are the temperature and density of the positron plasma.

2. **Laser induced recombination (LIR)**, which can occur when the above process (SRR) is stimulated with the addition of laser exposure. This interaction can be described by:

$$e^+ + \bar{p} + k\hbar\omega \rightarrow \bar{\mathrm{H}} + (k+1)\hbar\omega, \tag{2.6}$$

which is similar to that of SRR but with k additional photons, each with an energy of $\hbar\omega$ as a result of the laser. This process would further allow tuning the energy of the laser (and therefore the photons) to create antihydrogen in specific energy states. The ratio of the stimulated recombination probability over the spontaneous recombination probability is given by

$$g = \frac{W_{\mathrm{stim}}}{W_{\mathrm{spon}}} = \frac{Pc^2}{F\Delta\nu 8\pi h\nu^3} \tag{2.7}$$

where P is the laser power in watts, F is the cross-sectional area of the laser, ν is the frequency of the laser, and $\Delta\nu$ is the frequency spread of the interaction (laser × positrons) [48]. In practice, this frequency spread is dominated by that of the positrons[1], and therefore $\Delta\nu$ can be expressed as $(mv_{e^+}/\hbar)\Delta\nu_{e^+}$ where ν_{e^+} and $\Delta\nu_{e^+}$ is the velocity and velocity spread of the positron respectively.

This method of $\bar{\mathrm{H}}$ production was investigated by the ATHENA collaboration in 2006, targeting in the $n = 11$ state [43]. By applying Equation 2.7 to the parameters of the experiment at the time, one would expect the $\bar{\mathrm{H}}$ production rate to increase from the previously seen $24\frac{\bar{H}}{s}$ to $\sim 224\frac{\bar{H}}{s}$ (although [43] claims an expected an increase to $84\frac{\bar{H}}{s}$ due to a slight correction on Equation 2.7 to account for the re-ionisation rate given by the laser power that is comparable to the radiative decay rate - more discussions of this effect can be found in [49]). However, a 90% confidence limit was placed on an upper bound of the measured increase to be $0.3\frac{\bar{H}}{s}$ suggesting that SRR contributes negligibly to the $\bar{\mathrm{H}}$ production rates seen by the experi-

---

[1] I.e., the frequency spread of the laser is much smaller than that of the positron energy spread.

ment prior to the addition of LIR. This hints that the other mechanisms of antihydrogen production (charge transfer and three-body recombination) are the dominant production mechanisms over SRR/LIR. As a result of this experiment, LIR as a method for efficient $\bar{\text{H}}$ production was abandoned in favour of these other mechanisms.

3. **Charge transfer**, which is when, instead of mixing $\bar{p}$ and e^+ plasmas to form antihydrogen with some third body carrying away the binding energy, positronium (Ps)[2] is mixed with $\bar{p}$. Once mixed, the $\bar{p}$ is able to replace the electron in the Ps system, creating $\bar{\text{H}}$ and releasing an electron:

$$\text{Ps} + \bar{p} \rightarrow \bar{\text{H}} + e^-. \tag{2.8}$$

Producing $\bar{\text{H}}$ in this method requires first forming positronium plasmas. This can be achieved in a few ways; for example, the GBAR experiment creates Ps atoms by colliding $\sim 10^{10}$ positrons onto a target made from porous silica, converting the positrons into Ps with an efficiency of $\sim 30\%$ [50], or "laser-controlled" $\bar{\text{H}}$ production, successfully implemented by the ATRAP experiment in 2004 [42] uses collisions between excited caesium (Cs) atoms and positrons to create excited Ps. The energy level of the $\bar{\text{H}}$ produced in this method will be dominated by that of the positronium and it is therefore possible to create $\bar{\text{H}}$ in higher energy states by exciting the Ps atoms prior to mixing [51].

4. **Three-body recombination (TBR)**, which is the most relevant method for this thesis due to its high formation rates at high-density and low-temperature plasmas, and as this is the preferred method deployed by the ALPHA experiment. TBR occurs when the positron plasma is dense enough for an additional positron to carry away the binding energy of the process:

$$e^+ + e^+ + \bar{p} \rightarrow \bar{\text{H}} + e^+. \tag{2.9}$$

This process occurs when two positrons collide close to a $\bar{p}$, hence the need

---

[2] Positronium (Ps) is the bound state of an electron and a positron in orbit around each other.

for high-density plasmas. If the collision occurs in such a way that one of the positrons loses more kinetic energy than the binding energy of the $\bar{p}$, it will become trapped in a bound state with the $\bar{p}$, while the other e^+ carries away the excess energy. This process can be very efficient for dense plasmas; however, it mostly forms high-n states that are either untrappable or weakly bound [47]. Further collisions between the $\bar{\mathrm{H}}$ and the positron plasma can occur after production, knocking the positron deeper into a bound state [52], but it is also possible for further collisions to ionise the $\bar{\mathrm{H}}$ back into a $\bar{p}$.

The production rate of this procedure was first calculated by Glinsky and O'Neil [53] for an ion introduced into a positron plasma[3], in a strong magnetic field ($|\vec{\mathbf{B}}| \to \infty$). The frequency of positron-ion collisions can be given by nb^2v and the probability that there will be another positron in the vicinity to carry away the binding energy can be given by nb^3, where n is the density of the positron plasma, $v = \sqrt{k_B T/m}$ is the thermal speed of the positrons (T is temperature of the plasma, m is the positron mass, and k_B is Boltzmann's constant), and $b = e^2/k_B T$ is the classical distance of closest approach (DCA) (e is the elementary charge). Hence, the resulting rate of production will be proportional to the product of these two frequencies:

$$\begin{gathered} \Gamma_{TBR} \propto (nb^2v)(nb^3) = n^2b^5v \\ \implies \Gamma_{TBR} = Cn^2b^5v \end{gathered} \tag{2.10}$$

for some constant C. Equation 2.10 can be loosely understood as the velocity of the e^+, v, multiplied by the cross-section, b^2, multiplied once further by the probability of finding another e^+ in that volume nb^3. While it can therefore be seen that the production rate is proportional to the density and temperature of the positron plasma ($\Gamma_{TBR} \propto n^2T^{-9/2}$); the exact rate depends on the constant of proportionality, C.

Initially also calculated by Glinsky and O'Neil [53] to be $C = 0.070(10)$ and later confirmed by Robicheaux and Hanson [54] using a completely different, more precise, method giving $C = 0.072(2)$. This can be compared to the

---

[3] Actually their work considers ions into electron plasmas, but we assume the same formulas to hold for $\bar{p}$ into e^+.

rate calculated for $|\vec{\mathbf{B}}| = 0$, which was 0.76(4), showing the effect of the magnetic field on the production rate. Though by including the next order term (which essentially accounts for the $\vec{\mathbf{E}} \times \vec{\mathbf{B}}$ drift) in the guiding centre approximation, the value of C increases slightly to $C = 0.11(1)$.

This process was simulated using Monte-Carlo methods [55], and the resulting value of C was found to agree with those calculated by Robicheaux and Hanson [54]. However, this simulation also found that the formation rate of low field seeking (and therefore trappable) states of $\bar{\text{H}}$ is suppressed in stronger magnetic fields. In the paper, less than 8% of the resulting $\bar{\text{H}}$ were found to have magnetic moments large enough to be trappable, significantly increasing the loss of the process and reducing the overall efficiency.

Further, note that the $T^{-9/2}$ term in the rate suggests that colder positrons could result in higher production rates, and Sec. 3.9.1 discusses recent developments in ALPHA's experimental cycle aimed at achieving this.

N.B: There is an assumption of an infinite and homogeneous mixing of positron and antiproton plasmas in the work of Robicheaux and Hanson and other papers by, for example, Jonsell [56] which aim to better approximate experimental conditions suggest that the trappable fraction is much lower, on the order of 10^{-4}, the evidence of which is backed up by experimental results.

These processes are discussed in more detail in, for example, [52] [57].

2.3 Trapping Particles

The ALPHA collaboration employs positron/antiproton mixing as its method of antihydrogen production leading to TBR. This requires the trapping of both species prior to mixing, as well the resulting neutral antihydrogen after formation. This section will discuss the traps used by ALPHA to achieve this. When discussing the trapping of charged particles, matter and antimatter are equivalent as they are both charged, just oppositely, and the principles of their operation are identical.

2.3.1 Charged Particles

Consider initially the trapping of a single particle. Since charged particles interact with both the electric ($\vec{E}$) and magnetic fields ($\vec{B}$), a simple way to trap them would be to place them in the 3D minimum of the electrostatic potential. However, Earnshaw's theorem, which follows from Gauss' law, states that a point charge cannot be maintained in a stationary equilibrium by the electrostatic interaction of the charges alone.

Formally, this means that the electric force in 3D space ($\mathbf{F}(\mathbf{x})$) resulting from an electric potential ($U(\mathbf{x})$) will always be divergenceless, and therefore no local maxima or minima can exist in free space [58], i.e.:

$$\nabla \cdot \mathbf{F} = \nabla \cdot (-\nabla U) = -\nabla^2 U = 0. \tag{2.11}$$

As a result of this, electrostatic fields can be used to contain particles in at most two dimensions, and the third must be constrained in another way. One method of achieving this is to add a uniform magnetic field along the trap axis ($\vec{B} = B_0\hat{\mathbf{z}}$), which confines particles in the (x, y) plane.

A *Penning trap*, then, is a particle trap capable of confining the motion of charged particles by using both an electric and a magnetic field. The homogeneous magnetic field (parallel to the trap axis) confines their motion radially and stops the charged particles from escaping perpendicular to the trap axis. While an axial electric field then confines the particles along the trap axis and stops them from escaping axially. In this configuration, the particles remain trapped in all three dimensions, though not completely stationary.

The trap itself is named after F.M. Penning by Hans Dehmelt, who first proposed the idea of the Penning trap in 1955 after seeing a vacuum gauge built by Penning. Dehmelt and Wolfgang Paul finally built the first Penning trap in 1959, jointly winning the Nobel Prize in 1989 for their ion trapping techniques, which included the Penning trap [59].

Generally, a Penning trap consists of two end caps, which are biased with respect to a hyperbolic electrode surrounding the trap (see Figure 2.1). Then, given the resulting quadratic electrostatic potential

$$U(\mathbf{x}) = \frac{V_0}{R^2}(2z^2 - x^2 - y^2) \tag{2.12}$$

where V_0 is the trap potential, and R^2 is related to the size of the trap ($r_0^2 + 2z_0^2$ - see Figure 2.1), one can confine the motion of a particle within the trap.

A potential such as that found in Equation 2.12 can be generated by a configuration of electrodes found in Figure 2.1.

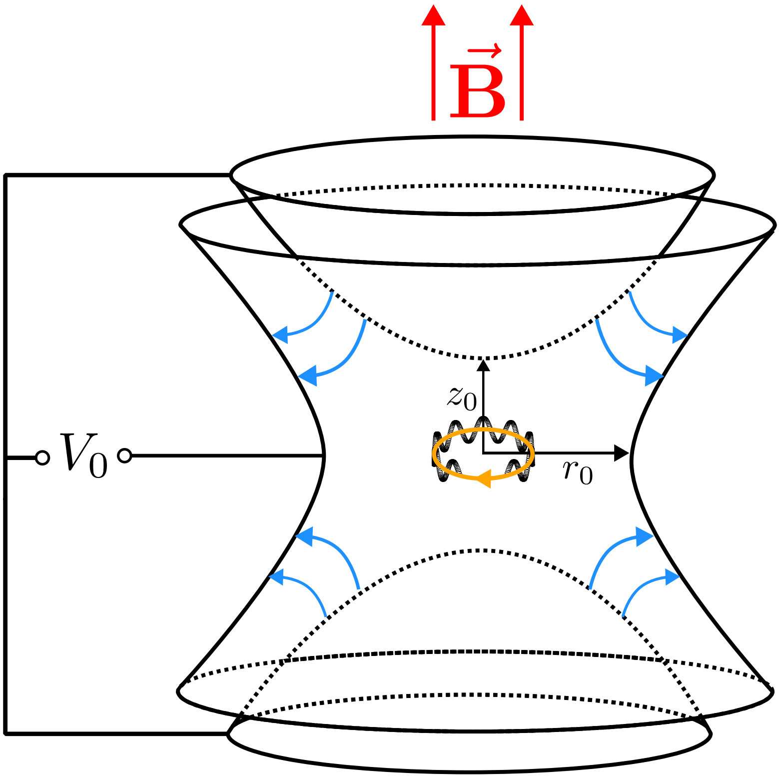


Figure 2.1: A standard Penning trap configuration formed from a hyperbolic electrode with two end caps. The central electrode is biased with a voltage of V_0. The electric field lines are displayed in blue, and the magnetic field in red. The orange circle displays the *magnetron motion* which arises in Penning traps, and the black trajectory displays the full motion of the particle after solving the equations of motion. This motion is seen in more detail in Figure 2.2.

The equations of motion of a single ion inside the trap are then described by the Lorentz force

$$\mathbf{F} = m\mathbf{a} = q(\vec{E} + \mathbf{v} \times \vec{B}) \tag{2.13}$$

where m, q, $\mathbf{v}$, and $\mathbf{a}$ are the mass, charge, velocity, and acceleration of the particle respectively. Given that $\vec{E} = -\nabla U$, and $\vec{B} = B_0\hat{\mathbf{z}}$ we can expand this equation

(Equation 2.13) into the three equations of motion for x, y, z:

$$\frac{d^2x}{dt^2} - \omega_c\frac{dy}{dt} - \frac{1}{2}\omega_z^2 x = 0 \tag{2.14}$$

$$\frac{d^2y}{dt^2} + \omega_c\frac{dx}{dt} - \frac{1}{2}\omega_z^2 y = 0 \tag{2.15}$$

$$\frac{d^2z}{dt^2} + \omega_z^2 z = 0 \tag{2.16}$$

where $\omega_z = \sqrt{4qV_0/mR^2}$, and $\omega_c = qB_0/m$.

We note that solutions to Equation 2.16 will take the form $z(t) = A\sin\omega_z t + B\cos\omega_z t$, which corresponds to a harmonic, oscillatory motion in z with a frequency quantified by ω_z, this is known as the *axial frequency.*

It can also be seen now that by setting $B = 0$ in Equation 2.13 and finding the resulting equations of motion that without the magnetic field and only the electrostatic field, the particles will still be confined in z with the same harmonic, oscillatory motion, but will have exponential solutions in x, y plane, indicating that they will not be trapped radially.

It is also worth noting here that the equation of motion of z is decoupled from the motion in the x, y directions, while the motion in x, y are dependent on each other.

Finding the solutions to the equations of motion for x, y with the magnetic field in place, however, is more difficult. Following the methods described in [60] we can substitute $u = x + iy$ into both Equation 2.14 and Equation 2.15, merging them into a single equation of motion for u:

$$u'' + i\omega_c u' - \frac{1}{2}\omega_z^2 u = 0 \tag{2.17}$$

the solution of which has general form $u = e^{-i\omega t}$. Substituting this ansatz into Equation 2.17 gives the condition:

$$\omega^2 - \omega_c\omega + \frac{1}{2}\omega_z^2 = 0 \tag{2.18}$$

whose roots are given by:

$$\omega_\pm = \frac{1}{2}\left(\omega_c \pm \sqrt{\omega_c^2 - 2\omega_z^2}\right). \tag{2.19}$$

For this to have real solutions we require that $\omega_c^2 - 2\omega_z^2 > 0$, this is sometimes referred to as the *trapping condition* [60] as this is required for solutions to Equation 2.14 and Equation 2.15 to be bounded (otherwise they are exponential).

Then for $\omega_c \gg \omega_z$ the positive root approximates to ω_c, which is known as the *cyclotron frequency*. Further, by noting the product of both roots $\omega_+\omega_- = \frac{\omega_z^2}{2}$ we can approximate the negative root as $\frac{\omega_z^2}{2\omega_c}$, this is known as the magnetron frequency, often denoted ω_m and is the motion resulting from the $\vec{E} \times \vec{B}$ drift, which occurs when a charged particle moves in the presence of both electric and magnetic fields, causing it to drift with a velocity perpendicular to both fields.

In summary, the motion of a single particle in a Penning trap can be considered as a superposition of three oscillatory motions, the frequencies of which are described by:

$$\omega_z = \sqrt{\frac{qV_0}{m}}, \qquad \text{(Axial)} \tag{2.20}$$

$$\omega_+ \approx \omega_c = \frac{qB_0}{m}, \qquad \text{(Cyclotron)} \tag{2.21}$$

$$\omega_- \approx \omega_m = \frac{\omega_z^2}{2\omega_c} \qquad \text{(Magnetron)} \tag{2.22}$$

The motion of a particle trapped in this configuration can be seen in Figure 2.2.

For two ions in a trap, the equations of motion need to be corrected to account for the Coulomb interaction between them as a result of their charge. This results in the inclusion of a $q^2/4\pi\epsilon_0 r_{ab}^2$ term, where ϵ_0 is the vacuum permittivity, and r_{ab} is the distance between the two ions. This new term couples the motion of the particles to each other, as well as the x, y motion with the previously independent z motion.

Further, for full plasmas of charged particles, which are dense and often contain millions of particles, this becomes even more involved, and the Penning trap is no longer the optimum trap. A modified version of the Penning trap (the Penning-Malmberg trap) is discussed in Sec. 2.3.2, and the equations of motion for a trapped

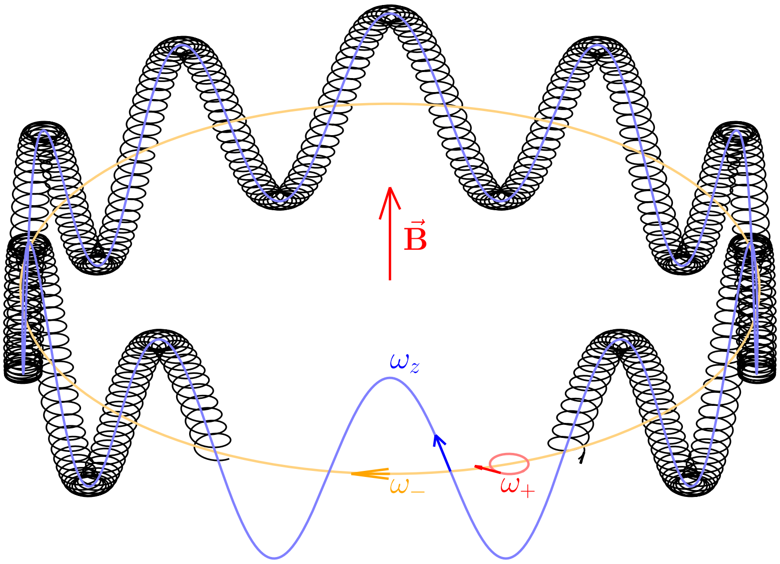


Figure 2.2: The motion of a single trapped particle in a Penning trap. Not to scale. The axial (ω_z), cyclotron (ω_+), and magnetron (ω_-) motions are displayed in blue, red, and orange respectively. The direction of the magnetic field vector is also shown in red in the centre of the plot.

plasma are derived. More detailed descriptions of the dynamics of a Penning trap can be read in, e.g. [61] or [62].

2.3.2 Charged Plasmas

The Penning-Malmberg (PM) trap is a modified version of the Penning trap that is a series of cylindrical electrodes instead of two perfectly hyperbolic end caps and the hyperbolic ring electrode.

First implemented by John Malmberg in 1975 [63], the interest in a PM trap comes from its ability to trap and store plasmas instead of singular ions. A schematic diagram of a PM trap with just three electrodes can be seen in Figure 2.3.

Although it is typically difficult to trap neutral plasmas in a PM trap, single-species, non-neutral plasmas can be trapped for considerable amounts of time with this configuration.

A plasma is a state of matter with properties slightly different from that of solid, gas, or liquid and is defined as a collection of charged particles moving freely in space under the influence of external and internal magnetic/electric fields. In a plasma, these internal fields are usually strong enough to affect the dynamics of

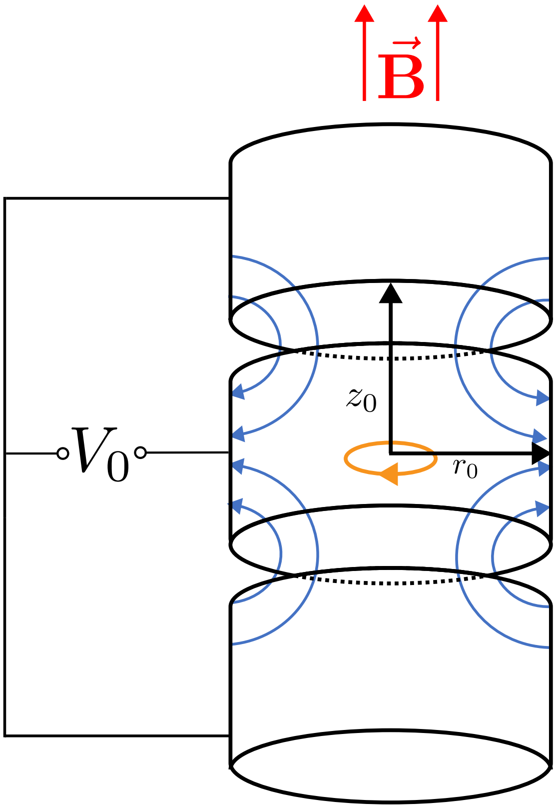


Figure 2.3: A Penning-Malmberg trap

the plasma itself, adding a layer of complexity to the kinetics.

The Debye length (λ_D) of a plasma is a measure of its net electrostatic effect and how far it persists in space [64]. The Debye length is a fundamental property when discussing plasmas and is defined as $\sqrt{\frac{\epsilon_0 k_B T}{n e^2}}$ where T, n are the temperature and density of the plasma.

If the Debye length of the plasma is much smaller than its actual length in z (L), then the cloud can be considered to be a plasma.

Further, we can assume the plasma density is described by a continuous Boltzmann distribution and is therefore semi-fluidic if and only if the density of the plasma varies slowly on the scale of the mean *inter-particle spacing*, $a \equiv n^{-1/3}$. Without this condition ($\lambda_D \gg n^{-1/3}$) the plasma cannot be treated as a continuous distribution, as there will be large variations throughout the sample. This inequality is often written as $\lambda_D n^{-1/3} \gg 1$ and the dimensionless quantity $\lambda_D n^{-1/3}$ is the "number of charges in a Debye sphere", often called the *plasma parameter* [65].

Most of the collections of particles used at ALPHA adhere to these two conditions, and are therefore treated as plasmas. They are also usually non-neutral, single-species plasmas though there are many situations where multiple species sit close, or mixed in the same well.

While the equations of motion described in Sec. 2.3.1 can hold for some of the collections of particles used by ALPHA (specifically antiprotons prior to compression), generally, we will need to adjust these equations to account for the internally generated electric and magnetic fields from the plasma cloud itself.

In order to demonstrate radial confinement of charged plasmas one looks at conservation of angular momentum of the plasma as a whole, and the individual dynamics of particles in the trap are not considered.

The methods described in [66, 67] outline the proof of principle of confining plasmas in a trap.

First, consider a plasma of N particles (of the same charge) in a cylindrical and symmetric PM trap (see Figure 2.3), with the same constant axial magnetic field as in Sec. 2.3.1. The motion of the charges is well approximated by the Hamiltonian for many particles, and it can be shown through the properties of cylindrical symmetry that the Hamiltonian is invariant under translations of θ (in cylindrical coordinates) [67] and thus the total canonical angular momentum of the system ($P_\theta \equiv \sum_{j=1}^{N} p_{\theta_j}$) is conserved (i.e. $P_\theta = L$ for some constant L). While in theory, perfect conservation is possible, in practice, this isn't the case. Neutrals in the trap (remaining due to imperfect vacuum conditions) will collide with particles in the plasma - slowly radiating energy and angular momentum away from the plasma. Further, minor imperfections in the symmetry of the trap will create slight variations on the fields, generally resulting in a small torque being applied to the plasma. With enough care, these effects can be minimised such that the plasma will reach thermal equilibrium prior to any significant loss of energy or great torque being applied.

To see why conservation of angular momentum implies radial confinement, we begin by rewriting the constant of angular momentum using velocity variables, giving:

$$P_\theta = \sum_{j=1}^{N} m_j v_{\theta_j} r_j + \frac{q_j}{c} A_\theta(r_j) r_j \tag{2.23}$$

where m_j, v_{θ_j}, q_j, and r_j are the mass, angular velocity, charge, and radius of the j-th particle respectively, and $A_\theta(r)$ is the azimuthal component of the vector potential (for the magnetic field described so far this is equal to $Br/2$). Hence, by noting that we are working with a single species plasma (i.e., $m_j = m\forall j$, and $q_j = q\forall j$) and that for a sufficiently large magnetic field the second term in Equation 2.23 will dominate, it can be approximated as:

$$P_\theta \simeq \frac{qB}{2c}\sum_{j=1}^{N} r_j^2 \tag{2.24}$$

and, since we already know $P_\theta = L$ we can say:

$$\sum_{j=1}^{N} r_j^2 \simeq \text{const.} \tag{2.25}$$

providing a constraint on the overall radius (or, more accurately, the mean square radius). Therefore, as long as angular momentum is conserved, which will hold as long as no additional external forces are applied to the plasma, the radius of the plasma will remain constant. It is worth noting here that the synonymous equation for a multi-species plasma of the same mass is:

$$\sum_{j=1}^{N} q_j r_j^2 \simeq \text{const.} \tag{2.26}$$

and hence it would be possible for e.g., an electron-positron pair to travel together radially outwards until neither is trapped, while still keeping the total angular momentum of the system constant.

While the conservation of angular momentum ensures the plasma remains trapped radially, axial confinement comes from applying a bias voltage on the two end electrodes such that the plasma cannot overcome the resulting electrostatic potential. This dip between two peaks of electrostatic potentials is often referred to as a *well*, and can be thought of classically as a ball rolling up and down a hill. If the ball (plasma) does not have enough longitudinal energy to climb the hill (well, or more accurately *potential difference*), it will remain trapped until either the blocking potential is lowered or some mechanism of inducing longitudinal energy is imposed on the plasma.

2.3.3 Neutral Atoms

Both the Penning, and the Penning-Malmberg trap rely on the non-neutrality of the particles they are trapping to confine them, and both depend upon the electrostatic field to contain the particles in two dimensions. For an electrically neutral atom such as antihydrogen, these electric potentials will not contain it. As such a new method of trapping is required to hold these atoms.

First, consider that all atoms can be thought of as small magnetic dipoles, the potential energy of which, given magnetic moment $\vec{\mu}_H$, in an externally produced magnetic field $\vec{\mathbf{B}}$, can be written as

$$E = -\vec{\mu}_H \cdot \vec{\mathbf{B}}. \tag{2.27}$$

By applying an external magnetic field on this dipole, a torque will be generated ($\vec{\tau} = -\vec{\mu}_H \times \vec{\mathbf{B}}$), causing it to either align or anti-align with the external field. This results in the magnetic moment of the antihydrogen atom taking one of two forms: $\vec{\mu}_H = \pm\vec{\mu}_B$ where μ_B is the Bohr magneton ($\mu_B = e\hbar/2m_e$). As such the resulting potential energy (Equation 2.27) can be expressed as $E = \mp\vec{\mu}_B \cdot \vec{\mathbf{B}}$. This means that depending on the state of the antihydrogen, the potential energy will either be negative or positive. For a positive dipole, this results in a negative potential energy, and hence, the energy decreases as the field increases. Antihydrogen in this configuration is known as “high-field seeking $\bar{\mathrm{H}}$”; synonymously, negative dipoles are known as “low-field seeking $\bar{\mathrm{H}}$”.

Gauss’ law again prevents any sort of field configuration with a local maximum from existing, and therefore, the “high-field seekers” are sometimes referred to as untrappable $\bar{\mathrm{H}}$. However, it is possible to create a trap with a local minimum where the centre of the trap is the minimum of the absolute magnetic potential, which increases outwards in every direction. Such a trap would be able to hold any low field seeking $\bar{\mathrm{H}}$, and this type of configuration is referred to as a *minimum-B* or *Ioffe* trap. As long as the kinetic energies of the atoms contained within it are small enough to not overcome the magnetic potential of the trap walls, the atoms will remain trapped [12].

The maximum potential energy of a minimum-B trap is given by

$$E_{\max} = \vec{\mu}_B (B_{\max} - B_{\min}), \tag{2.28}$$

or

$$E_{\max} = 0.67 \Delta B [K] \tag{2.29}$$

after dividing by k_B (to convert to temperature) where $\Delta B = B_{\max} - B_{\min}$ is the difference between the maximum and minimum of the traps magnetic field magnitude in tesla.

A minimum-B trap can be achieved simply with two mirror coils in an *anti-Helmholtz* configuration (i.e. with oppositely flowing currents); however, the fields resulting from this configuration can produce Majorana spin flips at the centre which has a region of zero field [68, 69] resulting in gradual losses of the trapped atoms as the probability of flipping to an untrappable state approaches 1. Therefore, in order to prevent this zero field inconvenience, a trap is required that contains a minimum, but does not pass directly (or very near) to zero field. In the case of an *Ioffe-Pritchard* (IP) trap, this is done by introducing a series of parallel bars that generate a transverse multipole with zero field at $r = 0$. In this setup, the two mirror coils are still required to maintain axial trapping, though no longer with oppositely flowing currents. It then becomes possible to change the curvature of the trapping potential to obtain a more homogeneous field in the axial direction, suppressing Majorana transitions [70]. Given a multipole field of order p the magnetic field strength as a function of r ($|\vec{B}(r)|$) is proportional to r^{p-1} [71] ($p = 2 \implies$ quadrupole, $p = 3 \implies$ sextupole, $p = 4 \implies$ octupole). A plot of the magnetic field magnitude as a function of r for various multipole orders can be seen in Figure 2.4.

In antihydrogen synthesis, it is necessary to utilise both a PM trap to manipulate the charged species and an IP trap to contain the resulting neutral $\bar{\mathrm{H}}$. However, as mentioned in Sec. 2.3.2, one must assume azimuthal symmetry in the confining fields in order to achieve stable trapping of plasmas. This condition will not hold with the addition of a multipole to the magnetic field. How much this perturbs the field is extremely important when deciding on which order of pole to

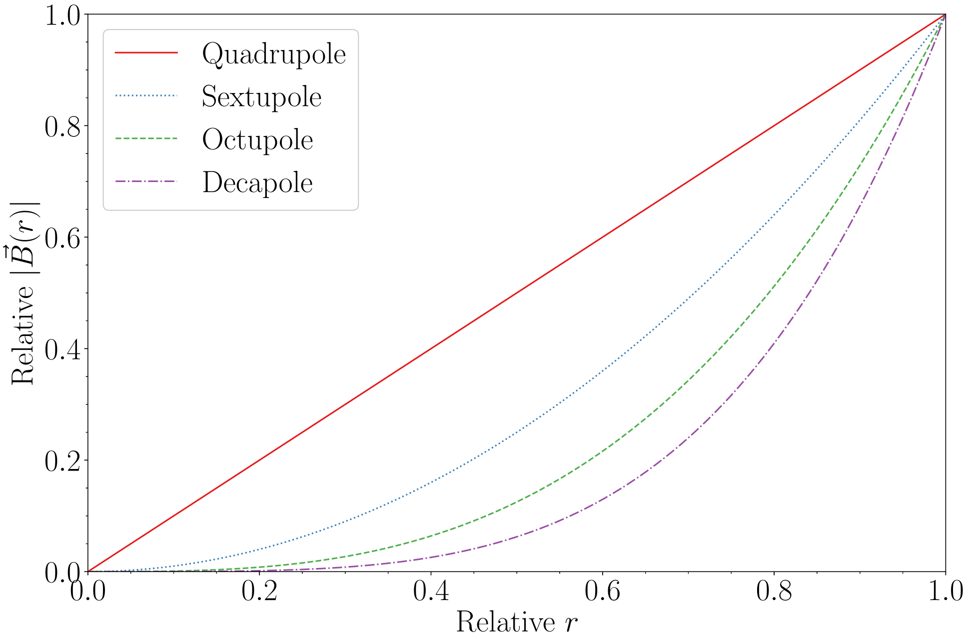


Figure 2.4: Relative radial fields for different order multipole. Plot is normalised s.t. $|\vec{B}(1)| = 1$.

implement. As can be seen from Figure 2.4, the higher the order of the pole, the smaller the perturbation is near $r = 0$. However, given that the entire structure is to be placed around a PM trap already, this means that our maximum in r will be less than the maximum displayed in the graph, and the higher order the multipole is, the smaller the depth of the resulting IP trap will be. Further to this, the distortion of otherwise homogeneous field lines caused as a result of superimposing a multipole over a solenoidal field can cause field lines to diverge/converge when moving along the axial direction. Therefore, depending on the inner radius of the trap electrodes, particles which follow the field lines could be made to intersect the walls, causing losses. This maximum radius of perturbation before loses occur (r_{max}) can be expressed as

$$r_{\text{max}} = R \exp\left(-\frac{|\vec{B}(R)|L}{2B_0 R}\right) \qquad \text{(Quadrupole)} \tag{2.30}$$

$$r_{\text{max}} = R\left(1 + \frac{|\vec{B}(R)|L}{B_0 R}\right)^{-1/2} \qquad \text{(Octupole)} \tag{2.31}$$

where R and L are the trap radius and length of confinement respectively, $|\vec{B}(R)|$

is the magnitude of the multipole field at the wall, and B_0 is the uniform axial magnetic field. Assuming a trap of $(R, L) = (2.25, 2.25)$cm, in a 1T axial field, and a 2T radial field, the values of r_{max} for a quadrupole and octupole are 0.827729, and 1.29904 respectively. This increase in r_{max}, along with the reduced perturbation at $r = 0$ motivated ALPHA to build its trap using an octupole for transverse confinement despite the decrease in trap depth that occurs as a result [71].

2.4 Annihilation

When a particle collides with its corresponding antiparticle, the resulting process is known as an *annihilation*. When this happens, the particles involved in the collision destroy each other, producing additional particles often called "secondaries". Since any particle-antiparticle pair must have opposite additive quantum numbers, any set of particles created as a result of this process must also have net-zero quantum numbers. Further, this process obeys the laws of conservation of energy, momentum, and spin, placing further restrictions on the sets of secondaries possible from the process.

In general, sufficiently low-energy collisions produce photons; however, in high-energy collisions lots of exotic and heavy particles can be generated via pair production. In the ALPHA experiment, most of the collisions that occur would be described as low energy due to the low kinetic energy of the plasmas involved, and as such, this chapter will only consider low-energy annihilations.

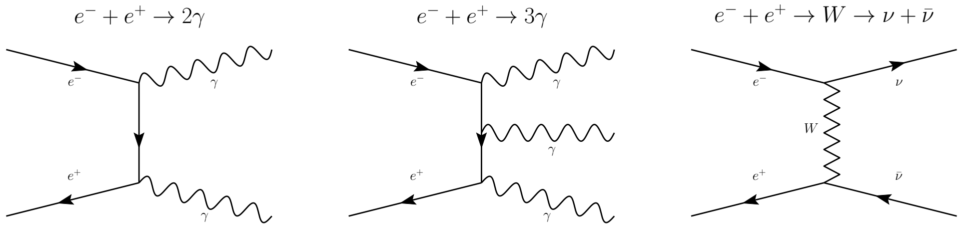
When positron and electron pairs annihilate, the most likely outcome is the production of two or more photons. Conservation of momentum forbids a single photon from being produced (as a single photon must carry some non-zero momentum), and the only other fundamental particle that the interaction has enough energy to generate are neutrinos (ν), which are 10,000 times less likely to be produced than photons. This is due to the ratio of the coupling constants for both the electromagnetic and the weak force which is $\sim 10,000$[4].

The Feynman diagrams for 2γ, 3γ, and $\nu - \bar{\nu}$ production via electron-positron

---

[4] The cross-section for photon production is proportional to the electromagnetic force coupling constant, and same for neutrino-antineutrino pair production with the weak force coupling constant.

annihilation can be seen in Figure 2.5.



(a) Feynman diagram for the 2γ e^+/e^- annihilation.

(b) Feynman diagram for the 3γ e^+/e^- annihilation

(c) Feynman diagram for the $\nu - \bar{\nu}$ production e^+/e^- annihilation

Figure 2.5: Feynman diagrams for the 2γ (Figure 2.5a,) 3γ (Figure 2.5a), and $\nu - \bar{\nu}$ production (Figure 2.5c) mechanisms for electron-positron annihilation.

In the two-photon case, the resulting photons must have each the energy of the electron/positron rest energy ($E_0 = m_e c^2 = 511$keV), and due to conservation of angular momentum, must be ejected in exactly opposite directions. These two properties provide enough of a unique characteristic to be detected and discriminated against the background with good confidence. The ATHENA experiment exploited this fact to detect antihydrogen annihilations in their apparatus [11, 72], proving they had indeed trapped antihydrogen. By using 192 pure CsI crystal scintillators (each $\sim 4\text{cm}^3$) arranged in a concentric cylinder around the trap and searching for signals around 511keV with $\cos(\theta) = -1$ they were able to unambiguously identify 131 ± 22 events resulting from antihydrogen production. However, this method of detection suffers from poor photon detection efficiency ($\sim$ 20% per crystal, and $\sim$ 0.25% total efficiency), requiring new detection methods to be developed for spectroscopy and other measurements (see Chapter 4).

While the discussion of the electron/positron annihilation is simplified by the fact that they are both elementary particles, the antiproton itself is a composite particle, comprised of two up antiquarks ($\bar{u}$) and one down antiquark ($\bar{d}$).

The antiquarks in the antiproton will therefore annihilate on contact with any other quarks such as those found in a proton, or a neutron. Further, and the more frequent process at ALPHA, is the annihilation of an antiproton on a heavy nucleus, $\mathcal{N}$ (as most of the antiprotons lost will be on the wall of the PM trap, comprising gold-plated aluminium) which can result in quite exotic behaviour during the final state interaction (FSI)[5] such as absorption of the produced secondary pions [73], or the hadronisation of the nucleus itself [74].

Regardless of the method, when an antiquark-quark annihilation occurs, usually the result is the production of a gluon which subsequently (along with the remaining quarks, antiquarks, and gluons), undergoes a complex rearrangement process known as *hadronisation*, or *fragmentation*. This process leads to the formation of several mesons, predominantly pions (π) and less often kaons (k) which distribute the total energy and momentum of the interaction throughout themselves [75].

The resulting mesons are unstable and will decay into photons, neutrinos, or electron-positron pairs unless they themselves collide with more material. The lifetime of the charged mesons is on the order of $\sim 10^{-8}s$, which is enough time to travel through the ALPHA apparatus entirely; the neutral mesons however typically have a lifetime of $\sim 10^{-16}s$ after which they decay into γ rays, which can then impact the apparatus creating more electron/positron pairs. The charged mesons can be detected as they leave the apparatus in a variety of different ways and will be the focus of Chapter 4.

While in pure $\bar{p}p$ annihilations, equal numbers of π^+ and π^- are produced (on average 1.5 of each, as well as 2 π^0), and in pure $\bar{p}n$ annihilations the number of π^- must exceed π^+ by 1 (on average 1 π^+ and 2 each of π^-, π^0) one consequence of the FSI on $\bar{p}\mathcal{N}$ annihilation is that $\bar{p}p$ annihilations can assume properties of $\bar{p}n$ annihilations and vice versa. This is caused by charge exchange in $\pi\mathcal{N}$ reactions ($\pi^-p \leftrightarrow \pi^0n$ and $\pi^+n \leftrightarrow \pi^0p$) and due to π reabsorption reactions ($\pi^-pp \leftrightarrow np$, $\pi^-pn \leftrightarrow nn$, $\pi^+nn \leftrightarrow pn$, and $\pi^+np \leftrightarrow pp$). All this makes it difficult to predict the multiplicity of a $\bar{p}\mathcal{N}$ annihilation, and as a result they are often modelled as the statistical average of both processes, of which the branching ratios are well known[6]. For example Table 2.1 shows the values for the results of an experiment performed in 1974 that measure the branching ratios of the product of $\bar{p}p$ annihilations (from [76]), and Table 2.2 shows the results from a similar experiment measuring the branching ratios of $\bar{p}n$ annihilations (from [77]).

---

[5] The FSI is the interaction of the annihilation products and the residual nucleons in the nucleus.

[6] These ratios are well known in the sense that experiments have been performed to measure them.

These branching ratios are crucial for use in simulating $\bar{p}$ annihilations, and these experimentally found branching ratios are the basis for many programs internal physics list, including the one used to generate the results found in Chapter 5. Further, accurate branching ratios allow for a more faithful comparison of true detector responses against simulation, especially where the number of resulting tracks is an important property of the detector's software.

Annihilation Products		Branching ratio (%)
$\pi^+ \pi^-$		0.375 ± 0.030
$\pi^+ \pi^- \pi^0$		6.9 ± 0.35
$\pi^+ \pi^- \mathrm{m}\pi^0$	$(m > 1)$	35.8 ± 0.8
$2\pi^+ 2\pi^-$		6.9 ± 0.6
$2\pi^+ 2\pi^- \pi^0$		19.6 ± 0.7
$2\pi^+ 2\pi^- \mathrm{m}\pi^0$	$(m > 1)$	20.8 ± 0.7
$3\pi^+ 3\pi^-$		2.1 ± 0.25
$3\pi^+ 3\pi^- \pi^0$		1.85 ± 0.15
$3\pi^+ 3\pi^- \mathrm{m}\pi^0$	$(m > 1)$	0.3 ± 0.1

Table 2.1: Branching ratios of $p\bar{p}$ annihilations, from [76].

Annihilation Products		Branching ratio (%)
$\pi^- \pi^0$		0.75 ± 0.15
$\pi^- \mathrm{m}\pi^0$	$(m > 1)$	16.9 ± 0.7
$\pi^+ 2\pi^-$		2.3 ± 0.3
$\pi^+ 2\pi^- \pi^0$		17 ± 2
$\pi^+ 2\pi^- \mathrm{m}\pi^0$	$(m > 1)$	39.7 ± 2
$2\pi^+ 3\pi^-$		4.2 ± 0.2
$2\pi^+ 3\pi^- \pi^0$		12 ± 1
$2\pi^+ 3\pi^- \mathrm{m}\pi^0$	$(m > 1)$	6.6 ± 1
$3\pi^+ 4\pi^- \mathrm{m}\pi^0$	$(m > 1)$	0.35 ± 0.03

Table 2.2: Branching ratios of $\bar{p}n$ annihilations, from [77].

Chapter 3 | ALPHA

The Antiproton Decelerator (AD) at CERN, Geneva, where the ALPHA experiment is based, is the only place where experimentation on cold, trapped antihydrogen takes place. In order to achieve this task, ALPHA makes use of many different components and experimental techniques. While the general theory of particle trapping was discussed in Chapter 2, and the methods of antihydrogen annihilation detection will be discussed in Chapter 4, this chapter describes the general procedures and components used by ALPHA during the trapping and experimentation cycle.

The ALPHA apparatus is split roughly into two major components, sometimes referred to as independent experiments: ALPHA-2 and ALPHA-g. Though the majority of *diagnostic stations* (instruments for diagnosing various properties of the beam, discussed in Sec. 3.1) are present in both sections, and, for example, the same catching trap (Sec. 3.7), and positron accumulator (Sec. 3.5.2) is used to transfer the antiprotons and positrons to both experiments there are some fundamental differences described as needed in the following chapter. A schematic diagram of the whole apparatus with appropriate labels can be found in Figure 3.1.

3.1 Sticks

Given the size of the ALPHA experiment, and the many species of plasma and laser/microwaves deployed at various points throughout, there is a need to implement many different diagnostic measurements, and to access the machine at different positions. In order to achieve this, ALPHA deploys both vertically and horizontally translating *sticks* to switch between various beamline instruments during operation.

These stations, often referred to simply as “sticks” are the housing unit of ALPHA’s many diagnostic tools, and consists of a vertically or horizontally mounted translator used to align different devices with the plasma traps axis during the cycle of the experiment.

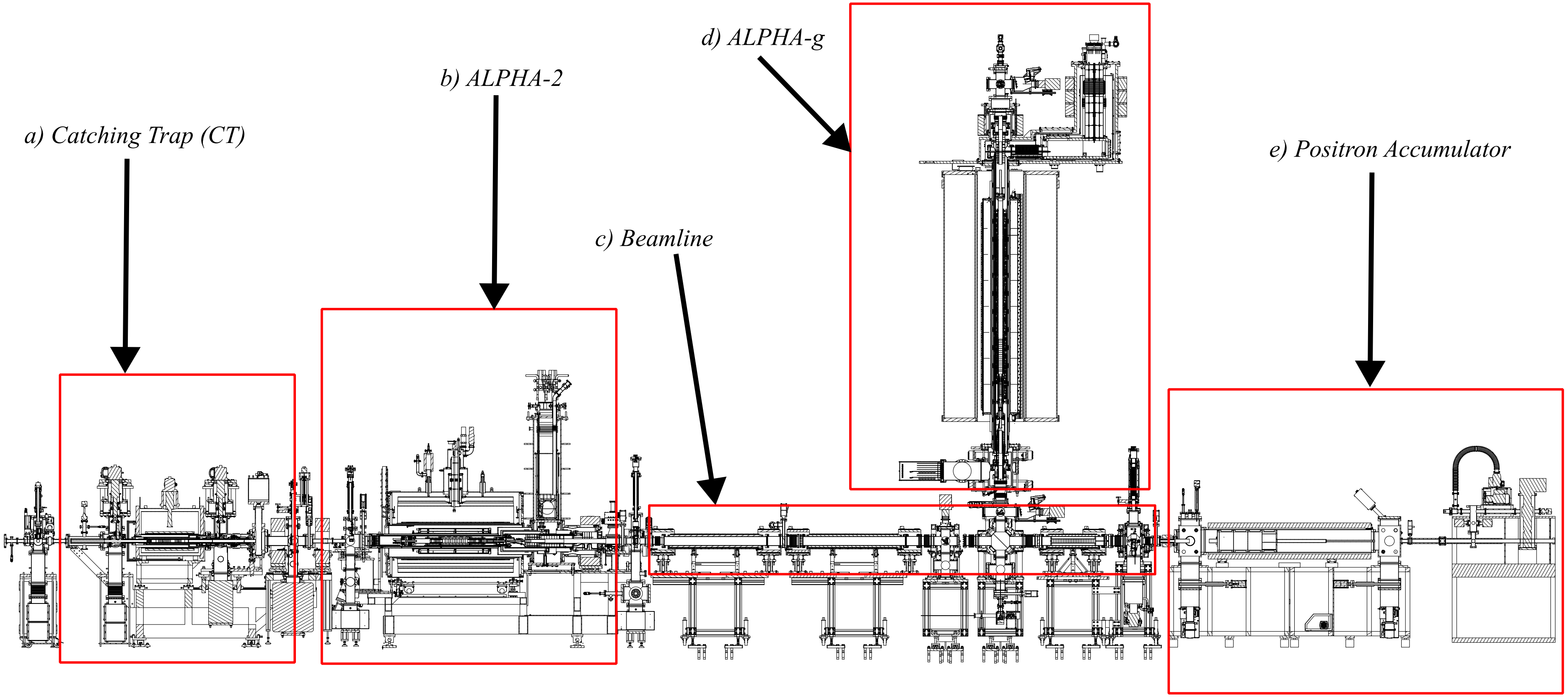


Figure 3.1: Full schematic of the ALPHA experiment, split roughly into its constituent components, from left to right they are a) the catching trap; b) ALPHA-2; c) the beamline; d) ALPHA-g; and e) the positron accumulator.

The ALPHA experiment has nine sticks placed throughout the various sections of the beamline, and though the specific instruments on each stick vary, in general, the sticks will consist of a selection from the following list: an electron gun, a microchannel place, a Faraday cup, laser/microwave mirrors/horns, a beryllium ablation source, or a “passthrough”. While some of these are fairly self-explanatory, for example, a passthrough is simply a hole to allow particles to pass through, and a laser mirror is just a mirror; other tools are described in more detail in the following sections.

3.1.1 Electron Gun

An electron gun (also called electron emitter) is an electrical component that produces a narrow, collimated electron beam. The electron guns consist of a filament, usually made of barium oxide, which releases electrons when a voltage is applied over it. The resulting electrons are then directed into Penning traps to be used for antiproton cooling or magnetometry.

3.1.2 Microchannel Plate

A microchannel plate (MCP) is a device used to detect single particles or photons. It works similarly to an electron multiplier (discussed later in Sec. 3.2) as both devices use secondary emission to amplify the intensity of a single particle. However, because the MCP has many channels arranged in a 2D plane, it can provide images which are a 2D projection of whatever impacts the MCP. This is especially useful when used with a plasma as it can allow for diagnostic of the size and shape of the plasma.

An MCP is a plate made of a highly resistive material containing many small tubes (microchannels) which lead through the plate. The plate is usually a maximum of $\sim$ 2mm thick, the channels are $\sim 5 - 20 \mu m$ in diameter, are distributed evenly throughout the plate, and are at an angle non-parallel to the plane of the plate. This ensures that when a particle enters a microchannel from the parallel plane it will hit the wall; and, by applying a strong electric field across the plate, each individual channel becomes a continuous electron multiplier, which causes the impact to release more electrons, which cascade and impact the wall again, further releasing more electrons. This continuous cascading and amplifying is what causes the amplification by several orders of magnitude of the signal initially hitting the plate.

The electrons eventually leave the reverse side of the plate and can be detected by an anode; however, in the setup at ALPHA, the electrons are directed onto a phosphor screen, which in turn becomes excited by the impact of the electrons, emitting photons. Finally, a 45° mirror directs the light onto a charge-coupled device (CCD) camera set to trigger at the same time at the MCP.

The gain of an MCP is very noisy, meaning identical particles can give widely varying signals when viewed on an MCP.

The MCPs deployed at ALPHA have an active area $\sim 13.53\text{cm}^2$ (diameter of 4.15cm) with a hexagonal array of microchannels, each with a $12\mu m$ diameter, and $15\mu m$ spacing. The voltage applied on the MCP varies with the expected signal but is usually either 1 or 1.1kV, resulting in a gain of $\sim 8 \times 10^5$.

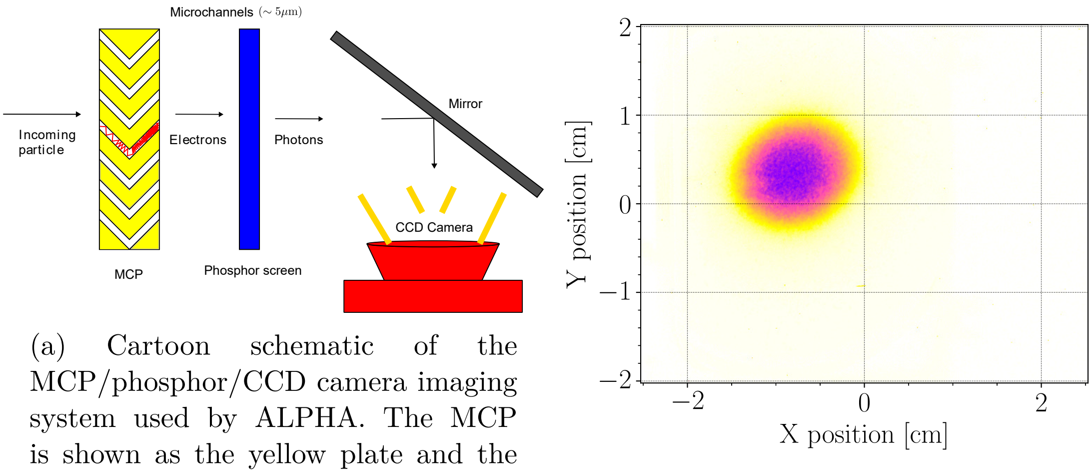
Details of the specific configuration of the MCPs at ALPHA can be found in [78], including the differences between using the MCP for species of e^+, e^-, or $\bar{p}$.

A schematic diagram of the setup can be seen in Figure 3.2a, and an example of one of the images recorded by one such MCP can be seen in Figure 3.2b.

3.1.3 Faraday Cup

A Faraday cup is a conductive cup designed to catch and measure the total charge of a batch of charged particles in a vacuum and is useful for counting the number of electrons or positrons in a given shot.

When a charged particle hits the wall of a Faraday cup it gains a small net charge as the particle impacts the wall of the cup. This cup can then be discharged to measure a small current which should be proportional to the total charge of the particle. By directing many particles into the cup it can be used to gain a rough count of how many particles were in the packet. However, since the readout noise of these is high, and the charge of elementary particles is small, in general, $> 10^6$ particles are required for a discernable measurement above noise in our setup.



(a) Cartoon schematic of the MCP/phosphor/CCD camera imaging system used by ALPHA. The MCP is shown as the yellow plate and the microchannels as chevrons through the volume. The phosphor screen is represented as the blue rectangle, the mirror is the grey rectangle, and the CCD camera is shown in red. The red cascade within the MCP chevron represents how the electrons multiply to amplify the signal that comes into the MCP.

(b) Example of an image from the CCD camera after directing a sample of antiprotons onto the MCP after the full $\bar{p}$ preparation steps have been completed.

Figure 3.2: Schematic diagram of the MCP imaging system (Figure 3.2a), and an example of an image taken using the system (Figure 3.2b).

3.1.4 Beryllium Source

One of the sticks houses a beryllium ablation source which, using 5ns pulses of a 355nm laser, produces Be^+ ions to be used for sympathetic positron cooling. The target is $> 99\%$ pure beryllium and ~ 0.5mm thick. Using this method we are able to reliably produce between 10^8 and 10^{10} ions, approximately 5% of which are directed towards a PM trap for cooling and mixing with positrons.

3.2 Scintillating Counters

A scintillating counter is an instrument used to detect and measure charged particles by using the excitation effect of incident radiation within the scintillating material.

When antiparticles collide with particles, the resulting secondary products generally have sufficient energy to ionise the molecules in a scintillating material if it should pass through some. The scintillators absorb the energy of the particle and *scintillate*, re-emitting the absorbed energy in the form of a photon, hence the name. This light is then usually amplified and converted to an electric signal by a photomultiplier tube (PMT) or a silicon photomultiplier (SiPM).

PMTs are essentially discrete electron multipliers consisting of a cascade of electrodes (often called dynodes) which cause secondary emission by the photoelectric effect, resulting in a signal large enough to be read out.

SiPMs use single-photon avalanche diodes (SPADs), which rely on $p-n$ junctions [79] to convert the photons to signal for readout. $p-n$ junctions will be discussed in more detail in Sec. 4.2 as they are also the technology behind the ALPHA silicon detector, the primary resource for detecting antihydrogen in ALPHA-2. The gain in a SiPM is generally more deterministic than those of a PMT, reducing the excess noise in the readout chain and improving the signal-to-noise ratio (SNR). Further, the bias voltages used in SiPMs is $\sim$ 100V, which is considerably less than the bias voltages required for PMTs (typically in the range of multiple kV), making them the preferential choice in most cases.

ALPHA deploys a range of both PMTs and SiPMs throughout the apparatus, mostly arranged in pairs to allow for coincidence detection, significantly reducing the noise in the readout. By using the panels in this way, we are able to count how many particles impact the scintillator. It is important to note that this is not necessarily the same as the number of antiparticles as each panel has an individual efficiency, the coincidence operations will come with an associated efficiency, and their geometric solid angle compared to the source of the antiparticles will all contribute to the total efficiency of the signal with respect to number of antiparticles. This efficiency is of great importance when using these scintillating panels for diagnostics, and the calibration of these panels when performing certain operations via simulations is the focus of Chapter 5.

3.3 CsI Detectors

Another form of scintillating counter; caesium iodide counters are crystalline blocks sensitive to the 511keV γ-rays produced after a positron annihilates. They are coupled to photodiodes that measure the yield and read out the traces. These detectors are placed throughout the experiment, specifically in places where particles are transferred between sections.

3.4 Longitudinal Temperature Diagnostic

An important parameter in antihydrogen production is the temperature of the plasmas involved, as the production rate of antihydrogen scales with both the temperature of the $\bar{p}$ and the e^+. However, a plasma doesn't have one single temperature associated with it; instead, the plasma has a distribution function ($f(\mathbf{x}, t, \mathbf{v})$) describing its energy distribution (often Maxwellian in nature) at a given time t, in position $\mathbf{x}$. In general, then, the temperature of a plasma can be considered the average thermal kinetic energy (KE) per particle in the plasma and is typically measured in eV or K. In [80], the authors describe a method for measuring the parallel energy distribution of a magnetically confined plasma by slowly lowering a confinement potential and seeing how many have enough energy to escape the confinement at any given moment. The probability density function $f(E_p)$ for finding a particle with energy E can be described as:

$$f(E_p) = 2\sqrt{\frac{E_p}{\pi}} \left[\frac{1}{k_B T_p}\right]^{\frac{3}{2}} \exp\left(-\frac{E_p}{k_B T_p}\right) \tag{3.1}$$

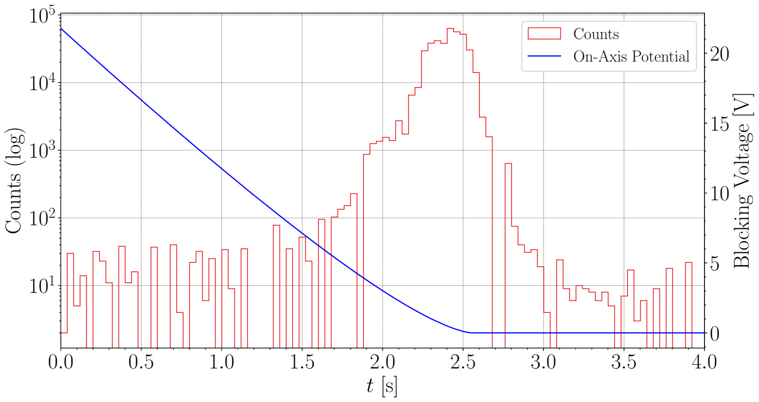
where E_p is the parallel/axial energy, T_p is the parallel temperature of the plasma, and k_B is the Boltzmann constant. However, we can assume as the particles escape on axis, their energies follow a one-dimensional Maxwell-Boltzmann distribution, i.e. the probability density function Equation 3.1 simplifies to

$$f(E_p) = \frac{1}{\sqrt{\pi k_B T_p}} \exp\left(-\frac{E_p}{k_B T_p}\right) \tag{3.2}$$

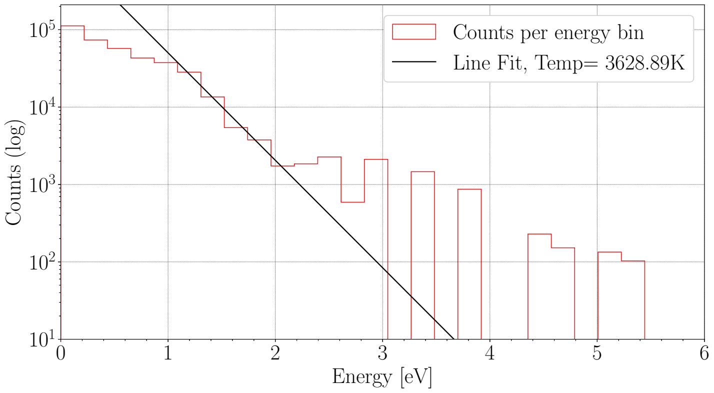
Then, the relationship (also from [80])

$$\frac{d\ln(f(E_p))}{d(E_p)} = \frac{-1.05}{k_B T_p} \tag{3.3}$$

which is accurate to about 5% while $E_p \gtrsim 2k_B T_p$, allows us to relate the parallel energy to the temperature. This derivative is easy to find experimentally in a PM trap since the confining voltages and the number of escaped particles reflect this distribution. Figure 3.3 shows one such measurement of $\bar{p}$ temperature using this method.



(a) Log plot of counts detected on a nearby scintillating panel as the on-axis blocking potential is slowly lowered, releasing contained antiprotons longitudinally. The blue line shows the voltage applied to the containing electrode over the same time. The counts occurring after the voltage has reached zero are particles that sit at a high radius compared to those in the centre. These particles are often referred to as the beam *halo*.



(b) Log plot of the counts converted to energy units. The black line is a linear fit (on a semi-log plot) to the range $[1, 2]$. Using Equation 3.3 this is converted to temperature units, and the resulting temperature of the plasma is quoted as ~ 3628K.

Figure 3.3: Example of the longitudinal temperature diagnostic procedure used by ALPHA to attribute a temperature to plasmas.

By plotting the number of escaping particles as a function of the trap depth on a log scale, one finds that the linear sections of these plots correspond to regions where Equation 3.2 accurately describes $f(E_p)$ (and hence where the energy distribution is Maxwellian), thus by fitting a line to these regions using Equation 3.3

we can find the temperature of the plasma. The non-linear sections of the curve seen for low energy particles represent the areas where the effect of space charge of the remaining particles is enough to alter their energy distribution as they leave. It is also possible to use these non-linear sections to determine the temperature if more sophisticated fitting methods (which take into account this space charge effect) are used; a detailed description of this method can be found in [81]. Further, in order to experimentally determine $f(E_p)$, it is necessary to lower the confining potential, and, in doing so, it is possible to apply some forces on the plasma, potentially changing the energy, dimensions, or density of the plasma. Thus, the temperature derived from the above method is not necessarily representative of the temperature of the plasma prior to the measurement. To correct for these "finite length effects" as they are known, one can either analytically calculate corrections for the temperature to take this into account, or one can simulate the dynamics of the dumps to reproduce the measurement and obtain a correction. Both methods have merits, and more discussion of this effect and the possible corrections to counteract it (as well as for the space charge effect) can be found in [82].

3.5 Obtaining antiparticles

Fundamental to the production of antihydrogen is the collection and trapping of the constituent particles (antiprotons and positrons - see Sec. 2.2), which can be a difficult process and, in the case of antiprotons, requires a large facility and lots of energy to produce.

3.5.1 Antiproton Production

The "usual" method of $\bar{p}$ production is to shoot a very high energy proton beam onto a conversion target and then collect the resulting $\bar{p}$ in a separated beam channel [83–85]. In this way, antiprotons can be produced via pair production (where some of the collision's energy is converted into mass, creating new (anti)particles).

The interaction can be described with the equation:

$$p(\text{beam}) + p(\text{target}) \rightarrow p + p + p + \bar{p} \tag{3.4}$$

provided the colliding protons have an energy above the production threshold for antiproton production ($\sim 7m_pc^2 \approx 6.6\text{GeV}$). Generally, a beam of high luminosity protons are directed onto a stationary target, usually a thin rod of some heavy metal, such as tungsten or iridium. Ideally, it would be long compared to the length of the collision but short compared to the absorption length so that most of the protons interact with, but also leave, the target. This is the process deployed by the AD facility at CERN, where 10^{13} protons are accelerated up to 26GeV/c by the Proton Synchrotron (PS), then directed into a 3mm diameter, 5cm long rod of iridium, creating many secondaries, some of which are antiprotons [86]. In order to capture these antiprotons, special "collector lenses" such as plasma lenses, lithium lenses, or a "magnetic horn" are needed to guide the beam. Previous studies have shown that magnetic horns are easier to repair and exchange, and, despite resulting in a 20-40% reduction in the total yield of $\bar{p}$, this was chosen as a lens for the AD facility, resulting in $\sim 5 \times 10^7 \bar{p}$ per shot [87]. These antiprotons have an energy of $\sim$ 3.57GeV/c with a momentum spread ($\Delta p/p$) of $\sim$ 1.5% and are directed into a storage ring for cooling and deceleration.

In this context, cooling refers to a decrease in the beam emittance. The beam emittance (often denoted ϵ) is a way to measure the quality of a beam, specifically its "laminarity". Beam emittance is a quantity related to the area occupied by the beam when plotted in the position-momentum phase-space. Though each particle is technically represented by a 6D vector comprising the particles position and momentum in three dimensions (x, y, z, p_x, p_y, p_z) if we assume that the motions in each dimension are independent, then we can treat them separately thus the emittance in, for example x, denoted ϵ_x, is related to the area of the 2D plot of x vs p_x.

Deceleration refers simply to a reduction in momentum.

Further discussions of how the antiprotons are cooled and decelerated prior to injection to the experiments can be found in Sec. 3.6.

3.5.2 Positron Accumulation

Accumulating positrons is a far simpler process than accumulating antiprotons, and they can be accumulated in much larger numbers and at much lower energies using a radioactive source that emits positrons, and an accumulator.

The ALPHA experiment (as well as many others that collect positrons) uses a ^{22}Na isotope as a positron source, which emits positrons via the following decay:

$$^{22}\text{Na} \rightarrow {}^{22*}\text{Ne} + e^{+} + \nu_e \tag{3.5}$$

where a proton in the sodium-22 nucleus converts to a neutron and emits a positron and an electron-neutrino in the process. The excited neon atom (^{22*}Ne) then decays to the ground state, emitting a 1.274MeV gamma ray ($^{22*}\text{Ne} \rightarrow {}^{22}\text{Ne} + \gamma$). The branching ratio of this beta decay (Equation 3.5) is $\sim 90\%$ and the source has a half-life of ~ 2.6 years [88] allowing the use of one source for multiple experimental seasons without the need for replacement.

The positrons emitted from this process have a very large energy spread, typically on the order of MeV, which renders them unusable for low-energy physics. In order to reduce their energy enough to be trapped, first they are implanted into some solid material known as a *moderator* to lose enough energy for trapping. When positrons are implanted into a material they will slow down through a few different processes (mostly involving the creation of electron-hole pairs, phonons, and core electronic excitations) and, depending on the properties of the moderator used, the time for the positrons to reach thermalisation can approach the annihilation lifetime. However, by selecting its initial implantation depth and diffusion length (a characteristic of the material), the positrons can reach the surface of the other side of the moderator before annihilation. Thus, positrons implanted into materials and released into the trap may reach the surface in a thermal or near-thermal state. The probability of this happening is low (in the 0.1–1% range), but the resulting positrons released into the vacuum are much lower energy than those that leave the source now in the $\sim$ 50eV range [57]. The ALPHA experiment uses a layer of solid neon deposited directly on the source as a moderator and using this technique can produce $\sim 5 \times 10^{6}$ positrons per second.

Now, in order to further collect, cool, and store the positrons emitted from this source one must deploy some form of trap; and ALPHA deploys a buffer-gas trap (BGT), also known as a “Surko” trap after it’s inventor Clifford Surko [89, 90].

A Surko trap consists of a modified PM trap (see Sec. 2.3.2) where instead of multiple electrodes of constant inner radius, the radii of each consecutive electrode increases. A uniform axial magnetic field of 0.15T inhibits positron motion radially, and voltages imposed on these electrodes prevent axial loss. The number of electrodes in this stack is often referred to as *stages*, for example, ALPHA deploys a 3-stage trap. The trap has a stepped electric potential, where the well gets deeper in multiple steps to its maximum depth.

However, when positrons enter the trap, without any mechanism to lose energy, they will be reflected at the furthest end of the trap and leave. In order for them to lose enough energy to remain trapped, BGTs deploy a volume of some molecular gas which (via inelastic collisions) causes cooling via excitation. The cooling gas used by ALPHA is ordinary molecular nitrogen, N_2, which has been found to be the most efficient gas for positron capture, and has a good probability of causing an excitation over an annihilation.

Nitrogen has a prominent positron excitation transition:

$$e^+ + \mathrm{N}_2 \to e^+ + \mathrm{N}_{2^*}. \tag{3.6}$$

which, upon decay of the N_{2^*} results in the positron losing $\sim$ 9eV kinetic energy, and, since the well is set specifically to have a roughly 5eV blocking potential, once the positrons lose the 9eV after entering the trapping region above the blocking potential, they are subsequently confined axially.

At 10eV (the energy level of the positrons in the high-pressure region of the trap [57]) the cross-section of this transition is greater ($\sim 2.2a_0^2$) than that of the transition which will cause the loss of positrons into positronium ($1.2a_0^2$) [91]:

$$e^+ + \mathrm{N}_2 \to \mathrm{Ps} + \mathrm{N}_{2+}, \tag{3.7}$$

where the positron binds to an electron in the nitrogen, forming positronium and leaving an ion behind. The relative frequency of these two interactions is what makes nitrogen an efficient form of positron cooling.

Since the electrodes have varying radii, there is a pressure gradient in the buffer gas along the electrodes; the lower pressure regions correspond to the regions with the lowest electrostatic potential. As a result the positrons enter the trap in the "high potential, high pressure" region of the trap and through collisions slowly drop into the "low potential, low pressure" region of the trap.

Surrounding this final low-pressure region is an electrode that is segmented azimuthally, allowing one to apply different electric fields at different angles around the plasma. By oscillating this field around the plasma, a torque can be applied, causing a compression. This is known as a "rotating wall" electrode, and is discussed in more detail in Sec. 3.8.1. The combination of the rotating wall for compression and the small amounts of remaining buffer gas result in a dense, cool positron plasma ideal for antihydrogen production.

A diagram of the ALPHA Surko trap can be seen in Figure 3.4, which shows the setup of the trap, as well as a sketch of the energy loss mechanism.

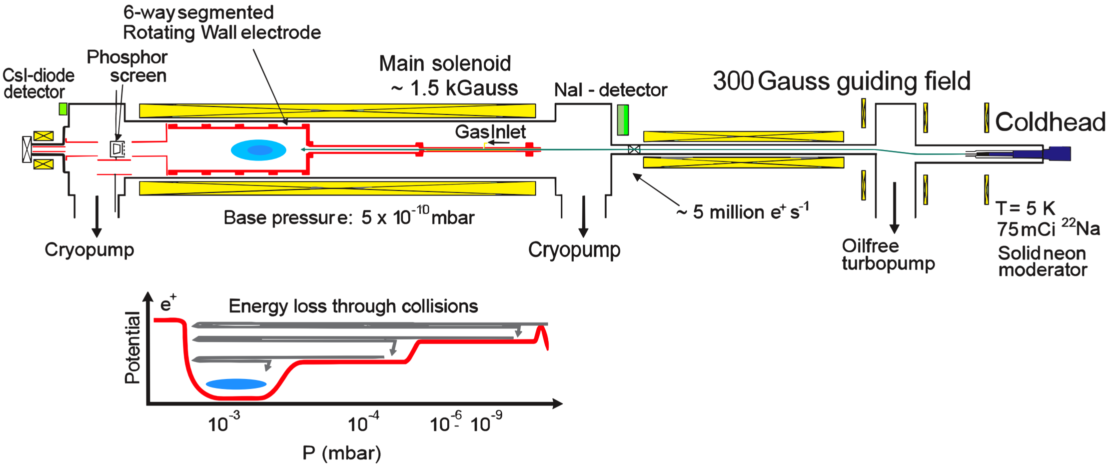


Figure 3.4: Schematic of the ALPHA Surko trap, adapted from [92], which shows the radioactive source, magnets, moderator, cryogenic equipment, and PM trap electrodes. The lower panel shows a diagram of the cooling process, where the red curve represents the electric potential, as a function of the pressure (and therefore also position). When the nitrogen line is closed, the gas is swiftly pumped out to allow for ballistic transfer of the positrons into the ALPHA-2, or ALPHA-g mixing traps.

Once the positrons are collected in the BGT, they can be transferred into one of the two mixing regions of the experiment. This process, known as the "positron transfer," requires pumping out the N_2 gas, then firing a series of pulsed magnets in the "beamline" (see Sec. 3.11) to transfer the positrons across the regions of low radial confinement into the antiproton trap. The transfer efficiency of this is $\sim 50\%$ with $\sim 10^8$ positrons arriving in both the ALPHA-2 and ALPHA-g traps. In total, the accumulation cycle takes about 100s. This accumulation and ballistic-transfer method (along with the stacking of several accumulation cycles) can give some of the highest density positron plasmas currently used in experimental physics [93].

3.6 Antiproton Deceleration and Cooling

The antiprotons resulting from the process described in Sec. 3.5.1 are too energetic to be used in antihydrogen production. For this reason the antiprotons need to be cooled before injection to the various experiments operating from the AD hall. This is done through two further storage/accelerator[1] rings; the AD, and ELENA (Extra Low ENergy Antiproton).

3.6.1 AD

The AD uses radio-frequency (rf) cavities to decelerate the beam and two separate processes to cool the antiprotons over its (roughly[2]) 100s cycle: stochastic cooling and electron cooling.

For the rf-deceleration an oscillating electric field which opposes the direction of motion of the beam is applied to the pulses as they pass through the rf-driven cavity. However, due to the conservation of phase-space density by decreasing the momentum of the particles in the beam, the position spread of the particles must expand, causing adiabatic expansion [94].

Thus, the beam must be cooled to reduce the size and energy spread of the circulating beam after any round of rf-deceleration. By applying both of these processes alternatively one can reduce both the momentum, and beam energy spread of the circulating beam.

Stochastic cooling [95, 96] involves applying a series of corrections or "kicks"

---

[1] (decelerator)

[2] The timing of this whole cycle is not completely deterministic, and can often fluctuate from 100-130s.

to the mean position of the beam using an electric field. The method requires a pickup (detector) to record the mean deviation of the beam, and an apparatus to apply the kick. Further, by placing these on opposite sides of the storage ring and running a cable directly between the two points, the signal arriving from the pickup to the kicker (with information on the optimum kick to apply) will arrive before the beam; allowing these kicks to be applied to the same particles that were measured at the pickup. Since the pickup can only record the mean deviation of the entire plasma no one single kick will be a correct correction for all particles. However, by carefully selecting the kick based on which will minimise divergent corrections to the beam, and ensuring to minimise the noise in the system, over the course of many turns the effect of this divergence is minimised, and the result is a net convergence to reduce the beam size and energy spread of the beam.

Electron cooling [97, 98] uses a beam of already cold electrons to interact with the antiproton beam, transferring energy to the electrons. To do this, cold electrons are ejected in the ring, matching the speed of the antiprotons. Then through Coulomb interaction between the two species the antiprotons transfer some of their excess energy to the electrons, reducing their energy spread, and increasing those of the electrons. These higher energy electrons are then removed from the beam later so as not to contaminate the bunch.

The two methods are complementary to each other and work well in tandem; the requirement to match the electron velocities to those of the antiprotons makes electron cooling an inefficient method of cooling when the antiprotons have a high velocity, conversely stochastic cooling is better applied to higher-energy beams, as the rate of cooling incurred from electron cooling is strongly dependent on the velocity spread of the antiprotons.

Therefore, the AD uses stochastic cooling on the two higher momentum stages, and electron cooling on the lower two stages.

Table 3.1 shows the target beam emittance (ϵ) and the momentum spread ($\Delta p/p$) of the beam before, and after each stage of the cooling process (subscript b for before, and a for after); and Figure 3.5 gives a schematic view of the cycle showing both the momentum and emittance from injection to ejection.

As some of the antiprotons are lost through the cycle, the result of this process

	Before		After			
p (GeV/c)	ϵ_b ($\pi \cdot$ mm $\cdot$ mrad)	$\Delta p_b/p_b$ (%)	ϵ_a ($\pi \cdot$ mm $\cdot$ mrad)	$\Delta p_a/p_a$ (%)	t (s)	Cooling process
3.57	200	1.5	5	0.1	17	Stochastic
2	9	0.18	5	0.03	8	Stochastic
0.3	33	0.2	2	0.1	16	Electron
0.1	6	0.3	1	0.01	9	Electron

Table 3.1: Transverse emittances and momentum spread before (b) and after (a) each cooling stage during the AD cycle.

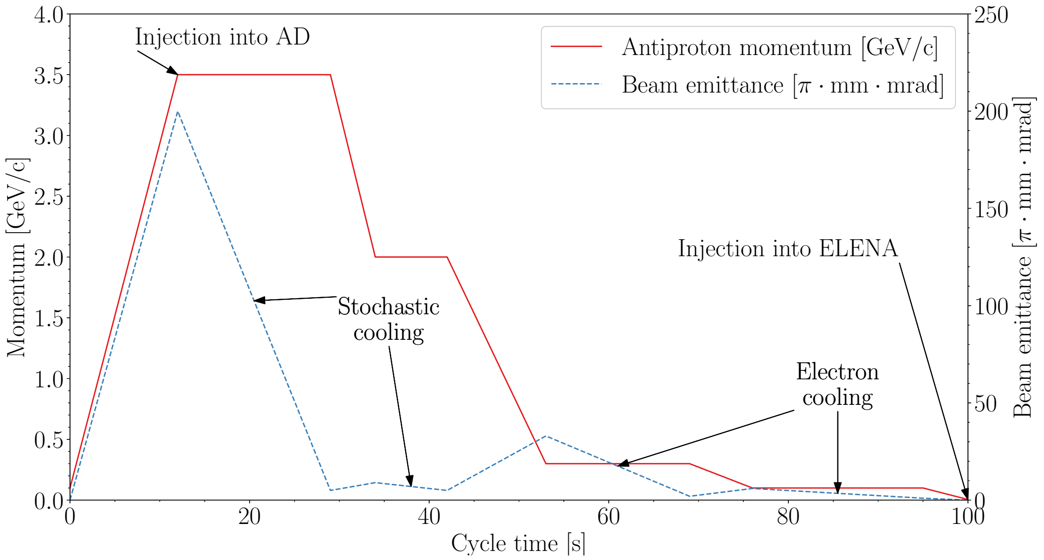


Figure 3.5: Example of AD cycle showing the momentum, and beam emittance throughout the cycle. The moments that show a decrease in momentum correspond to deceleration of the beam, and the moments that show a decrease in the emittance correspond to a cooling process (labelled on this plot). Notice on this plot that periods of deceleration also give beam "heating," i.e. they increase the emittance.

is bunches of $\sim 5 \times 10^7$ antiprotons at energies of $\sim$ 5.3MeV being delivered to ELENA for further cooling and deceleration [99, 100].

3.6.2 ELENA

While the 5.3MeV are considerably less energetic than the time of creation, the majority are still far too energetic to be trapped in a PM trap (this would require electrodes biased to 5.3 million Volts). Prior to the construction of ELENA the experiments were forced to further reduce the energy themselves before being able to trap them, and in ALPHA the loss during this process was typically 99%. To avoid these losses and allow more of the antiprotons to be trappable ELENA was constructed, reducing the energy of the 5.3MeV antiprotons even further to $\sim$ 100keV [101].

ELENA is a 30m ring located within the AD hall, and is the only ring at CERN entirely visible. It uses further rf-deceleration and electron cooling cycles to reduce the beam energy to $\sim$ 100keV over another period of $\sim$ 20s. This is then split into 4 separate bunches which can be delivered to up to four experiments simultaneously [100], allowing for 24-hour $\bar{p}$ delivery (provided fewer than four experiments are requesting beam), which, prior to ELENA, had to be shared between the experiments as only one beam was available. Losses incurred during this cycle reduces the roughly 5×10^7 antiprotons entering ELENA by about 12%, into four bunches of $\sim 1.1 \times 10^7$ antiprotons each.

ELENA has two cooling plateaus, both utilising electron cooling, and two deceleration cycles [101].

Table 3.2 shows the target beam emittance (ϵ) and the momentum spread ($\Delta p/p$) of the beam before and after the two stages of electron cooling (subscript b for before, and a for after); and Figure 3.6 gives a schematic view of the cycle showing both the momentum and emittance from injection to ejection.

	Before		After			
p (MeV/c)	ϵ_b ($\pi \cdot$ mm $\cdot$ mrad)	$\Delta p_b/p_b$ (%)	ϵ_a ($\pi \cdot$ mm $\cdot$ mrad)	$\Delta p_a/p_a$ (%)	t (s)	Cooling process
35	8	0.1	1.1	0.02	8	Electron
13.7	2.8	0.05	0.5	0.03	2	Electron

Table 3.2: Transverse emittances and momentum spread before (b) and after (a) each cooling stage of the ELENA cycle.

3.7 Catching Trap

The ALPHA catching trap (CT) is a device built to catch and cool the antiprotons from ELENA before transferring them to a separate device for antihydrogen production. The CT consists of a PM trap sat within an external 3T magnetic field

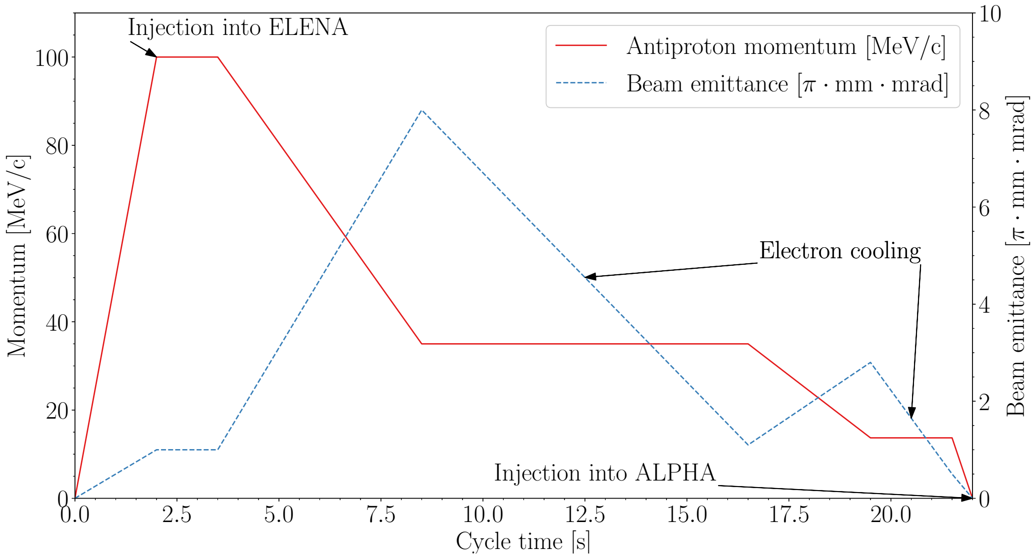


Figure 3.6: Example of ELENA cycle showing the momentum, and beam emittance throughout the cycle. The moments that show a decrease in momentum correspond to deceleration of the beam, and the moments that show a decrease in the emittance correspond to a cooling process (labelled on this plot). Notice again on this plot that periods of deceleration also give beam "heating," i.e. they increase the emittance.

provided by a solenoid known as the "Swansea magnet". The PM trap consists of 20 electrodes labelled E1-E20, two of which (E6 and E16) are segmented to allow for rotating wall compression (see Sec. 3.8.1); and two are high voltage electrodes labelled high-voltage A (HVA or E1), and high-voltage B (HVB or E13), used for dynamic trapping of antiprotons. Prior to the construction of the CT, antiprotons were caught directly in the 3T field of the mixing trap (see Sec. 3.10) however, in order to separate the high voltage electrodes from the region marked for precision spectroscopy, and to allow for better access of light axially into the trap, a dedicated catching apparatus was constructed for use in the 2014 experimental season. The catching trap also contains a special foil known as the *degrader foil* used to reduce the energy of the antiprotons even further prior to trapping (much like the positron moderator in Sec. 3.5.2).

The combination of AD and ELENA allows for $\sim 10^7$ antiprotons to be delivered to the experiments with an energy of $\sim$ 100keV with a pulse length $<$ 150ns. This is equivalent to a momentum of 13.7MeV/c and is still slightly too energetic for trapping, even with the high voltage electrodes. In order to reduce their energy further the antiprotons are directed through a sheet of $\sim 1\mu$m thick pure aluminium foil. Like the positron moderator this reduces the energy of the antiprotons further, resulting in cooler, lower energy $\bar{p}$ which are easier to catch [102]. The degrader foil can also be used as a Faraday cup, allowing diagnostic information on the number of antiprotons annihilating (see Sec. 3.1.3). Detailed discussions of the degrader foil can be found in [103]; and while the main purpose of the foil is to decrease the energy of the antiprotons passing through, some antiprotons annihilate on foil, and the secondaries of this interaction can be measured on external scintillators surrounding the experiment.

The two high-voltage ($\sim$ 4kV) electrodes are used for the initial trapping, until the antiprotons can be cooled enough via cyclotron cooling to be contained by the standard electrodes (which are limited to a voltage of 140V). The antiprotons enter the ALPHA apparatus with HVB already engaged, antiprotons with energy $<$ 4keV will not be able to overcome the electric potential barrier and will be reflected, back towards ELENA. HVA is then ramped up, triggered by the ELENA timing signal, before the antiprotons can leave the trap, confining them between HVA and HVB. A diagram of the CT PM trap can be seen in Figure 3.7 along with the trapping potentials applied when catching antiprotons.

At this point further cooling is required to reach thermal temperatures, allowing the low voltage electrodes to be sufficient for confinement.

Fortunately, charged particles will self-cool in a magnetic field via a mechanism known as *cyclotron cooling.* The Larmor formula [105] is used to calculate the total power radiated by a non-relativistic point charge as it accelerates and is given by

$$\frac{dE}{dt} = -\frac{q^2}{6\pi\epsilon_0 c^3}\left|\vec{a}\right|^2 \tag{3.8}$$

where $\vec{a}$ is the acceleration of the particle. However, for a particle in a PM trap, the cyclotron motion dominates this cooling effect, meaning we can approximate $\vec{a}$ with $\vec{\omega_c} \times \vec{v}$. Further, we can assume the energy of the system is entirely kinetic [106] and thus

$$\begin{aligned} \frac{dE}{dt} &= \frac{q^2\omega_c^2}{3\pi\epsilon_0 mc^3}E \qquad (3.9)\\ &= \frac{q^4 B_0^2}{3\pi\epsilon_0 m^3 c^3}E \quad \text{(using Equation 2.21)} \qquad (3.10) \end{aligned}$$

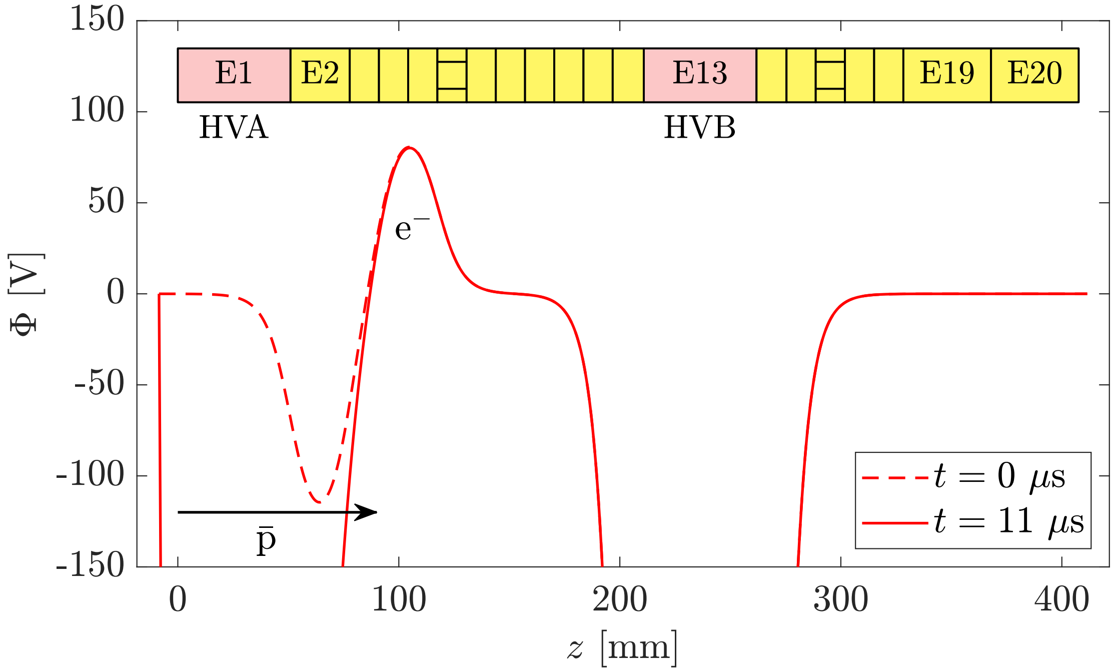


Figure 3.7: The CT electrodes and the axial potential before (dashed line) and after (solid line) catching antiprotons. Antiprotons enter from the left while HVA is off, a set time ($\sim 11\mu$s) after the bunch is injected a signal triggers HVA to ramp up, trapping the antiprotons inside. Electrons already sit inside the trap for cyclotron cooling. The peak voltage of HVA and HVB sits off the axis of the plot at -4kV. Figure from [104].

which can be rewritten as $dE/dt = -\gamma_c E$ where γ_c is the factor in Equation 3.10 and referred to at the *cyclotron cooling rate*. The solution to this equation is $E(t) = E_0 e^{-\gamma_c t}$ where E_0 is the initial kinetic energy of the particle, showing that the cooling of a particle in a magnetic trap due to the cyclotron motion is exponential. We can calculate the value of γ_c for both antiprotons and electrons (which is a constant given constant uniform magnetic field), and we see that the cooling rate for antiprotons in a 3T trap is $\gamma_{c,\bar{p}} \approx 1.953125e^{-93}/s$ which is considerable slower than that of electrons or positrons $\gamma_{c,e^-} \approx 3.333.../s$. This motivates the decision to use electrons in the cooling process whereby, instead of waiting for the antiprotons to cyclotron cool alone, we mix them with electrons, and, through the cooling of the electrons, allow the antiprotons to cool via Coulomb interactions.

It's worth noting here that while $\lim_{t\to\infty} E(t) = 0$, in reality this method has a limit where the mechanical temperature matches the cyclotron cooling rate, resulting in an equilibrium point.

Following the model and methods described in [107, 108] it can be shown that as many as 10^4 antiprotons, with initial energies in the keV range, could be cooled down to the eV range in several hundred milliseconds using an electron cloud with a density of only $10^8 e^-/\mathrm{cm}^3$

A sketch of the whole antiproton catching and electron cooling procedure can be seen in Figure 3.8.

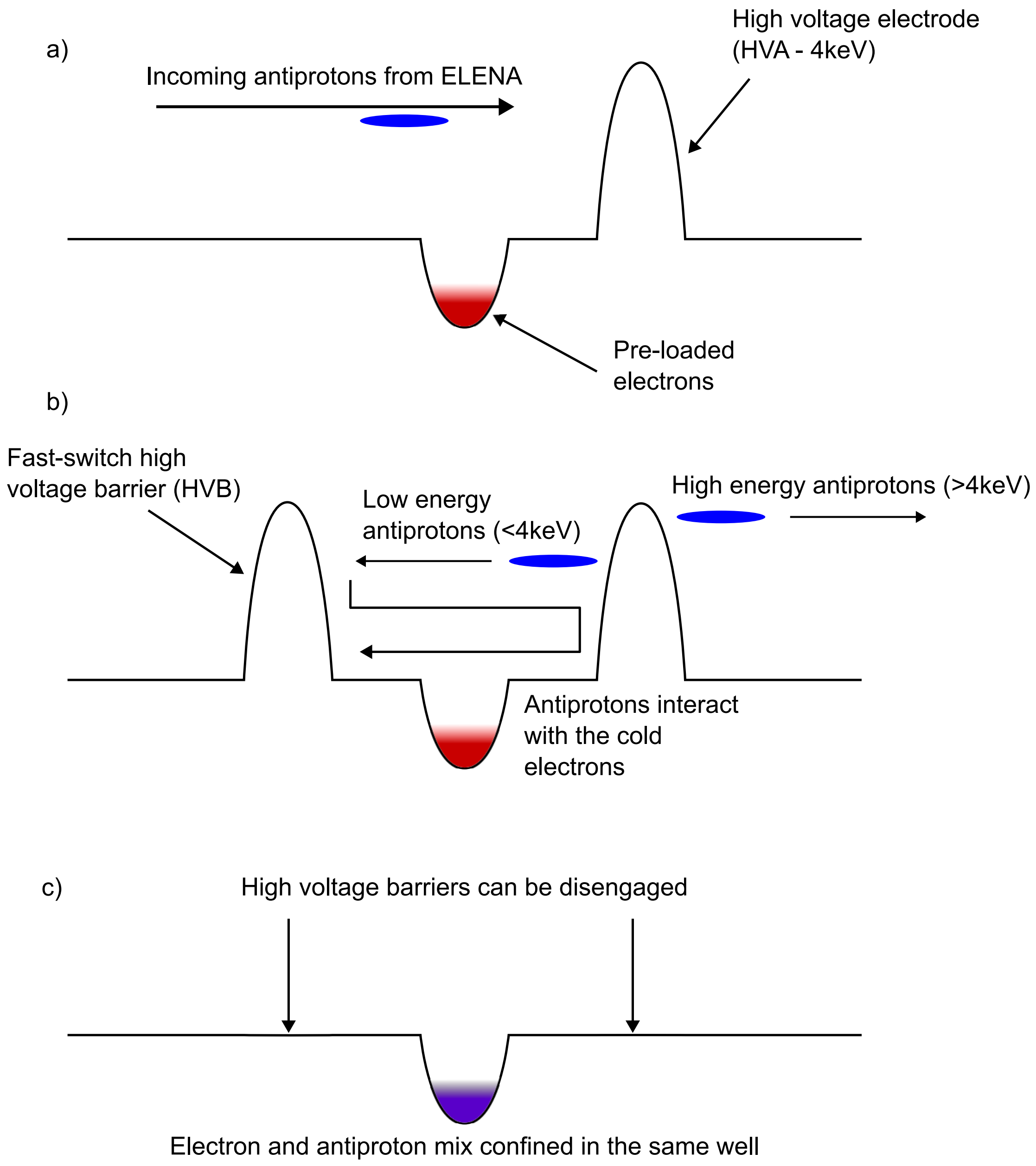


Figure 3.8: Sketch of the antiproton catching and cooling process. a) Antiprotons from ELENA enter the CT where electrons are already preloaded and HVA is already engaged; b) antiprotons with <4keV of energy are reflected back towards ELENA (those of energy >4keV are lost) whereby HVB is quickly engaged, confining the antiprotons between the two electrodes; c) Through Coulomb interactions with the elections they can cool into the low-voltage electron well and the HV barriers can be disengaged

3.8 Antiproton Preparation

Once the antiprotons have been captured and cooled to a reasonable level, they require transfer to the mixing trap where antihydrogen can be produced. Before this is done (and also after the transfer in the new trap), there are more parameters of the plasma that must be controlled. Namely, temperature, density, and size. There are multiple methods deployed by ALPHA that aim to control these properties of the trapped plasma, which are detailed in the following sections (Sec. 3.8.1 - Sec. 3.8.5).

3.8.1 Rotating Wall Compression

As mentioned in Sec. 3.5.2 plasmas often require compression or occasionally, expansion. This can be achieved through a method known as a “rotating wall”. A rotating wall is a special segmented electrode, divided azimuthally. By applying a pulsed, off-phase, periodic potential on these electrodes a torque is applied on the plasma, and, depending on the charge, will cause compression or expansion.

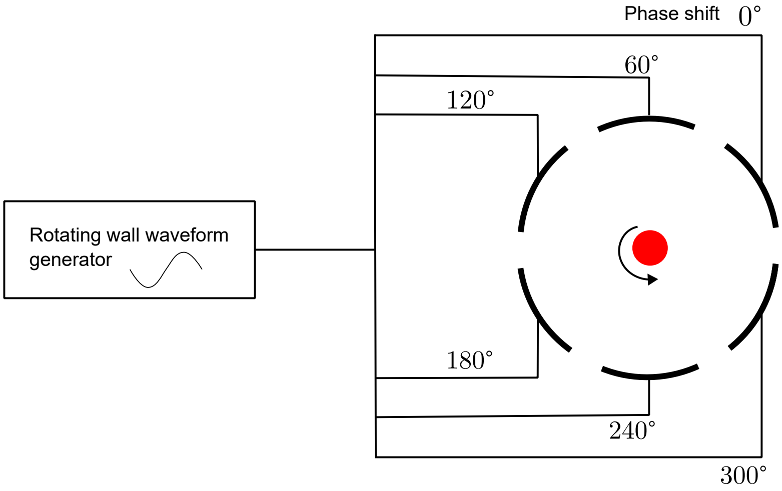


Figure 3.9: Sketch of a rotating wall electrode. A waveform generator provides a sinusoidal wave, which is split and phase-shifted, to each of the electrodes.

Referring to Equation 2.24 recall that the total canonical angular momentum applied on the plasma (P_θ) is proportional to the mean squared plasma radius ($\sum_{j=1}^{N} r_j^2$), i.e. $P_\theta \simeq \frac{qB}{2c} \langle r^2 \rangle$. Therefore if $dP_\theta/dt > 0$ and $q < 0$ (as is the case with antiprotons and electrons), then $d \langle r^2 \rangle /dt < 0$. Meaning that by applying a positive torque to a plasma of antiprotons or electrons, the change in the average of the radii will also be negative, compressing the plasma. It also follows that a negative torque will cause plasma expansion; and further that these are reversed for positively charged particles and negative torques.

The system shown in Figure 3.9 is capable of applying such torques to plasmas, where each section of the segmented electrode receive an oscillating sinusoidal potential with an offset phase, generating a rotating electric field which couples with the radial motion of the plasma.

The voltage applied to the n-th electrode in a rotating wall configuration is

$$V_n = V_0 \sin(\omega_{\mathrm{RW}} t - \theta_n) \tag{3.11}$$

where V_0 is the magnitude of the compression, ω_{RW} the angular frequency, t is time, and $\theta_n = 2n\pi/N$ where N is the total number of segments in the electrode, in the case where $N = 6$ the angles θ_n are displayed as phase shift in Figure 3.9

This rotating wall method for compression is common in particle manipulations where controlling the exact size or density of a given plasma is necessary [67, 109–113].

In ALPHA this method is generally applied to the combination of particles and antiparticles that allow cyclotron cooling; as compression invariably comes with heating [94], the combination of both in the well allows for compression via rotating wall, and then cooling through Coulomb interactions with the electrons, which radiate away the heat via cyclotron cooling.

3.8.2 The Strong Drive Regime

The strong drive regime (SDR), first discovered in 2005 [114], is a specific method of implementing the rotating wall where the rotational frequency of the plasma becomes in sync with the frequency of the field applied on the RW electrode. This results in a linear relationship between the frequency applied and the final density of the plasma, allowing for better control and more reproducibility of plasmas. The SDR has been shown to work on electron plasmas a few cm long and of frequencies from $0 - 10$MHz [115].

Reaching the SDR can be difficult, as the orientation of the plasma axis, magnetic field axis, and electrode field axis relative to each other is important; however ALPHA has been able to consistently use the SDR to obtain linear increases in plasma densities when driving from 100kHz - 1MHz.

Prior to development of SDR, in order to couple to the plasma one must couple to the plasma modes at very specific rotating wall frequencies. Which can result in very slow, gradual changes in the plasmas rotational frequency (and hence, its size), further plasma compression will only occur at discrete frequencies ω_{RW}, this is now known as a weak drive regime.

If instead one is able to reach the strong drive regime, efficient compression can occur over a much wider range of frequencies, without requiring specific tuning to the mode of each plasma.

3.8.3 Electron Kick-Out

The combination of antiprotons and electrons works well for the application of rotating wall compression and cyclotron cooling, however, the electrons must be removed prior to positron mixing for antihydrogen production.

This is achieved with an *electron kick-out*, also referred to as an “e-kick”. The large mass difference between the two species means that the axial thermal velocity of a particle is inversely proportional to its mass[3]. The thermal velocity of a particle in the axial direction at time t is given by

$$v(t) = \sqrt{\frac{k_B T}{m}} \tag{3.12}$$

where m, T are the mass, and temperature respectively. Therefore, the electrons have a thermal velocity that is ~ 42 times greater than that of the antiprotons, independent of temperature; and when the two species are no longer trapped by confining potentials they will separate almost instantly. For example, to travel the approximately 2cm distance corresponding to the width of the electrode would take the electrons $\sim 0.1625\mu$s, while the antiprotons would take $\sim 6.897\mu$s.

Given this fact we can now deploy a simple solution to remove the electrons from the well:

---

[3] Actually it’s proportional to temperature/mass but as the particles are in thermal equilibrium the temperature should be equal for both species.

1. The potential well containing the electron-antiproton mix is raised, and a single electrode separates the well from the ground potential
2. The potential applied to this single separating electrode is quickly lowered to ground.
3. Now that the plasma is unconfined the electrons (and antiprotons) are free to escape; however owing to the difference in their masses, the electrons escape much quicker, leaving most antiprotons yet to leave the trap
4. The confining potential is reapplied before the antiprotons can leave
5. Repeat until all electrons are gone.

Figure 3.10 shows a sketch of this procedure.

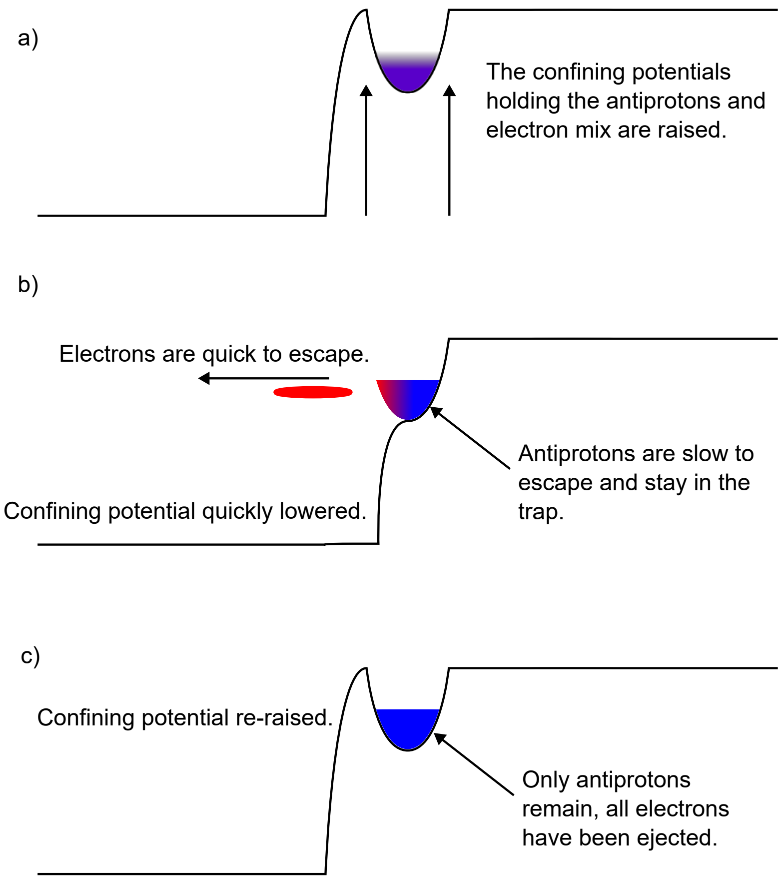


Figure 3.10: Sketch of the electron kick-out sequence. a) The confining potentials surrounding the electron/antiproton mixture are raised above ground; b) one of the confining potentials is lowered, allowing the electrons to escape; c) the confining potential is quickly re-raised after the electrons are gone, but before the antiprotons are able to leave.

The quick pulsing of this electrode, as well as the change in charge density of the mixture that comes from removing a large amount of charged particles, again causes heating. In order to reduce this effect, multiple e-kicks are applied to the plasma in increasingly intensity. The first kick removes most ($\sim 95\%$) of the electrons, which still leaves some electrons to once again cool the antiprotons through Coulomb interactions. Successive e-kicks then occur each time with a shallower well, and a lower raise from ground, causing less heating.

Often-times multiple kicks are required to completely eliminate the electrons and if all are not tuned correctly this process will heat/expand the plasma, or release too many antiprotons. By carefully tuning the parameters and alternating e-kicks with rotating walls this process can leave us with a cool, dense, pure antiproton plasma of somewhere between 200-1000K, and a radius of $\sim$ 0.8mm.

3.8.4 Evaporative Cooling

A lot of time in the preparation procedure has been dedicated to keeping the antiprotons cool, and while 200K is relatively cold, in order to trap the resulting $\bar{\mathrm{H}}$ in the roughly 0.8T field provided by the ALPHA-2 and ALPHA-G neutral traps the temperature of the resulting $\bar{\mathrm{H}}$ must be < 0.5K[4]. In order to maximise the amount of $\bar{\mathrm{H}}$ produced that is cool enough to trap, the temperature of the $\bar{p}$ must be reduced even further. Another method of cooling charged particles, which has been shown to be capable of reaching $<$ 500pK [116] is known as *evaporative cooling* (EVC).

Evaporative cooling works by lowering the potential on one side of a confining well just enough such that the hotter particles can leave, but not enough such that the cooler ones can escape. In the case where the energy distribution of the plasma is a Maxwell-Boltzmann (MB) distribution, those atoms at the higher velocity tail which have a higher KE (and therefore temperature) are the first to escape when lowering the potential. The net result is that while the total trap population decreases, so does the mean energy of the remaining population, and through collisions the energy of the population will redistribute again to a thermal equilibrium resulting in a MB distribution with a mean lower than the

---

[4] This value comes from substituting $\Delta B = 0.8T$ into Equation 2.29.

initial sample. An example of the evaporation stage of this process can be seen in Figure 3.11.

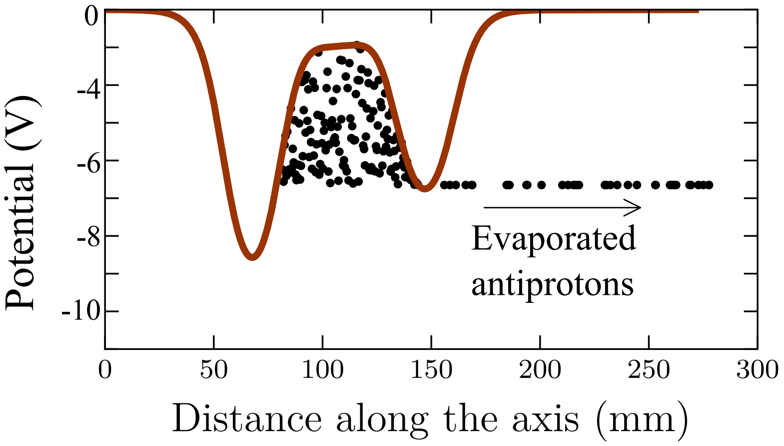


Figure 3.11: An example of an electric potential well design to confine antiprotons during evaporative cooling. The antiprotons (which are negatively charged) sit at the bottom of the potential well in black. By generating an uneven potential that is similar in magnitude to the energy spread of the antiprotons, the hottest particles are free to escape, leaving a cooler sample behind in the well. Figure adapted from [117].

EVC is a technique that relies on losses to work; something which is generally not desirable. This results in a trade-off between the antiproton number and the temperature desired. As more particles evaporate out of the well, the colder the remaining plasma will be, but also the fewer particles will remain.

ALPHA chooses to optimise this step based on antihydrogen formation rates, aiming to maximise the rate of production, not at producing the coldest possible plasma.

Using the evaporative cooling technique on antiprotons ALPHA recorded a minimum temperature of 9 ± 4K, with $6 \pm 1\%$ of the initial 45,000 antiprotons that were used remaining in the final 10 mV well [118]. ALPHA has also demonstrated evaporative cooling on positrons [119] and uses this technique to cool the positrons prior to mixing also.

3.8.5 Strong Drive Regime Evaporative Cooling

Developed in 2018 by the ALPHA collaboration [117], strong drive regime evaporative cooling (SDREVC) combines the methods of a rotating wall in the strong drive regime (Sec. 3.8.2), with evaporative cooling (Sec. 3.8.4) applied simultaneously. By performing both concurrently in this way SDR allows precise tuning of the plasma density, and EVC allows tuning of the plasmas space charge yielding incredibly consistent and reproducible plasmas.

When aspects such as the background gas levels, electrode fields, magnetic fields, and temperatures are all fluctuating, even on a shot to shot basis, having some method to maintain consistency is incredibly valuable. Two consecutive antiproton/positron shots produced under (nominally) the same conditions can often produce varying plasmas which is not good for systematic production of antihydrogen. The control and reproducibility of the plasmas at ALPHA are crucial, not only for $\bar{\mathrm{H}}$ production and physics experimentation; but also for use in systematic measurements such as electron cyclotron resonance (ECR) which is used to characterise the magnetic fields at ALPHA and requires a consistent electron plasma (often referred to as an electron *reservoir*).

SDREVC works by elongating the plasma to a “cigar” shape in the PM trap, the tail of which is sat under a RW electrode; then, by applying the RW and lowering the potential difference s.t. the space charge of the plasma is limited by the electrostatic potential of the well, the plasma will have a well-defined density and space charge.

As long as the initial number of particles is enough to both a) reach the strong drive regime; and b) fill the final well aimed for in the EVC step, this will result in reproducible plasmas with a known number of particles, density, and space charge. If the number of particles prior to the SDREVC step falls below this minimum threshold, the SDREVC will not succeed in producing an identical plasma; and in fact, in most cases where SDREVC has stopped giving reliable results, increasing the number of particles prior to the SDREVC step, generally resolves the problem.

The application of a RW to a plasma will again cause heating; and in order for SDREVC to work the plasma must be sufficiently cold. For electrons this isn’t a problem as the cyclotron cooling discussed in Sec. 3.7 is enough to counteract the heating of the RW. However, for this reason SDREVC has not been demonstrated to work on antiprotons, as the heating via the RW is too much (sometimes increasing the plasma from 100k to 3/4000K). However, the Be^{+} ions used for positron cooling (see Sec. 3.9.1) also benefit from external cooling in the form of laser cooling and recently the ALPHA collaboration was able to demonstrate the first application of SDREVC to obtain reproducible Be^+ plasmas; this being the first time that SDREVC has been demonstrated on a heavy ion.

3.9 Positron Preparation

Experimental [120] and theoretical [56] evidence has shown that for the parameters of the experiment in ALPHA, the antiprotons will thermally equilibrate with the positrons during mixing, just before antihydrogen is produced. This means the temperature of the resulting antihydrogen is essentially that of the positron plasma prior to mixing. Given that the trapping field for both the ALPHA-2 and ALPHA-g magnetic trap is ~ 0.5K deep (i.e. ΔB in Equation 2.29 is ~ 0.8T); it is essential for efficient trapping to keep the positrons as cool as possible. A newly developed method of positron cooling consists of mixing positrons with positive beryllium ions which can be cooled via laser cooling, and then in turn sympathetically cool the positrons.

3.9.1 Beryllium Assisted Positron Cooling

The Be^+ ion is the lowest mass ion that can be laser-cooled from the ground state making it an excellent candidate for sympathetic cooling [121, 122].

The Be^+ ions are generated by ablation on a piece of metal beryllium using a pulsed 355nm laser [123], regularly producing anywhere from 10^8 to 10^{10} atoms. Once the positrons are transferred from the positron accumulator and re-caught we may apply any of the methods mentioned in Sec. 3.8 - namely RW for compression, and SDREVC for better control of density and space charge. Once the positrons have been sufficiently cooled they are merged with the Be^+ ions where another round of RW compression takes place. Immediately following this step the Be^+ and positron plasmas have good radial overlap ($r \approx 0.6$mm). This mixture can then be cooled using a detuned 313nm laser, and much like using the cyclotron cooling of electrons to sympathetically cool antiprotons, here laser cooling of Be^+ ions causes sympathetic cooling of the positron plasma. Using this method positron temperatures were on average 6.8 ± 0.5K, [124] a factor of 2.6 times lower than the temperatures reported (using the same measurement method[5]) as those recorded at the cyclotron limit in the absence of any Be^+ ions: $17.3 \pm 0.5\mathrm{K}$ [124].

It was also shown that using these much cooler positrons to mix with antiprotons and form antihydrogen resulted in a factor 8 increase in trapped $\bar{\mathrm{H}}$. This allows us to accumulate over 15,000 atoms in under 7 hours making measurements that would previously take weeks, be able to be performed in under 24 hours.

While both the ALPHA-2 and ALPHA-g trap were built to allow room for laser light, (for Be^+ cooling, $\bar{\mathrm{H}}$ cooling, and spectroscopy) as of 2024 only the ALPHA-2 trap has the hardware in place for ablation and laser cooling of beryllium ions. The introduction of this method to the ALPHA-g apparatus has begun and is ongoing, which could seriously reduce the statistic error on the result presented on the gravity experiment [17].

3.10 Atom Trap(s)

Both the ALPHA-2 and ALPHA-g apparatus contain a PM trap for mixing the charged particles to produce antihydrogen, as well as a neutral trap for containing the resulting $\bar{\mathrm{H}}$. Though both the ALPHA-2 and ALPHA-g traps are very similar there are some minor differences, and the specifics of each trap (both the PM and neutral traps) is described in the following sections.

3.10.1 ALPHA-2 Trap

The ALPHA-2 atom trap consists of a PM trap containing 27 electrodes (labelled E1-E27) in a 1T external solenoid. The magnets for the neutral trap sit between the electrodes and external solenoid for the PM trap. I diagram of the A2 atom trap can be seen in Figure 3.12.

The PM trap is split roughly into three sections, the re-catching trap (RCT), the mixing trap, and the positron re-catching trap. Electrodes E1-E7 (see Figure 3.12) are part of the RCT, and are in place to "re-catch" the antiprotons

---

[5] The technique used here - described in Sec. 3.4 - may not be truly indicative of absolute temperature, however, it should at least be relatively comparable with other measurements made in the same way.

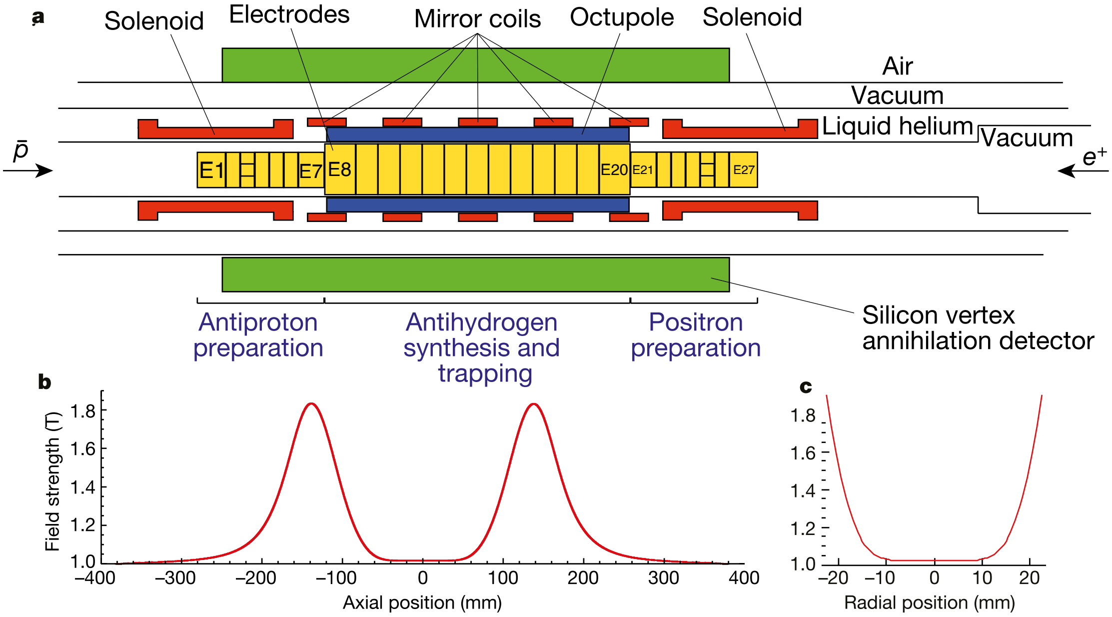


Figure 3.12: a) Sketch of the electrode and magnet layout of the ALPHA-2 atom trap. Antiprotons and positrons are re-caught at each end of the trap, before they are mixed in the central region, where the octupole and mirror coils can trap the resulting antihydrogen. The external PM solenoid is beyond the figure. Electrodes E3 and E25 are segmented to allow RW compression and E8-E20 have a larger radius than the re-catching traps; b) Axial well formed by the five mirror coils responsible for the axial confinement of neutral antihydrogen. c) Radial magnetic octupole field profile. Figure adapted from [14].

transferred to the atom trap from the catching trap, and apply the same preparation that was performed in the CT (see Sec. 3.8). E3 and E24 are segmented to allow for rotating wall compression (as described in Sec. 3.8.1). The mixing trap (E8-E20) is the section used for the mixing of antiprotons and positron for antihydrogen production (see Sec. 3.12). The neutral trap to store the resulting $\bar{\mathrm{H}}$ surrounds the electrodes. The final section (E21-E27) is the positron re-catching section, and is once again used for the re-catching of the positrons transferred from the positron accumulator and to perform preparation similar to that of the antiprotons.

The ALPHA-2 neutral trap consists of five mirror coils and an octupole [71]. The mirror coils provide a strong magnetic field gradient along the axial direction, and though there is some radial component (as well as fringe fields) their purpose is to axially trap the resulting $\bar{\mathrm{H}}$'s motion in the z-direction, and maintain a uniform field in the region where experiments will occur.

The mirror coils are placed at $\sim \pm 14$cm centred on the PM trap, giving the locations of the field maximum seen in Figure 3.12.

This corresponds to a minimum of ~ 1.06T, and a maximum of 1.82T, resulting in a trap depth of 44μeV, or ~ 0.5K (from Equation 2.28 and Equation 2.29 respectively). This value defines the overall trap-depth of the ALPHA-2 atom trap, as any atom with an energy greater than this will eventually escape the well and annihilate on the wall.

3.10.2 ALPHA-g Trap

The ALPHA-g trap is, by design, very similar to that of the ALPHA-2 trap. However, since both the antiprotons and positrons enter the apparatus from the same direction, instead of multiple re-catching traps on either end of the mixing trap, the ALPHA-g trap has an extra long RCT to allow for the preparation of both species simultaneously.

ALPHA-g was constructed specifically to measure the effect of gravity on the motion of antimatter and, due to the fact that gravity is one of the weakest fundamental forces, precise control of the magnetic fields in ALPHA-g is essential. A $\pm 1\%$ precision gravity experiment would require controlling the field at the relevant mirror coils to within $\pm 7 \times 10^{-6}$T [125], including all stray fields.

The neutral trap (and also the PM trap) are divided into three sections, the lower trapping region, the analysis region (in the centre), and the upper trapping region. The lower, and upper traps are designed to both be identical to each other (and to the ALPHA-2 trap), and are capable of performing the gravity measurement outlined in Sec. 9.1.3 alone. However, the analysis region in the centre is designed to perform a precision measurement when it is installed. Though it does not contain a PM trap like the bottom and top, and is therefore not capable of producing antihydrogen itself however, two transfer coils are in place to allow transfer of antihydrogen from either of the productions traps into the centre. Identical traps on either side of the ALPHA-g apparatus are essential for good symmetry and good detection coverage in the case that antimatter should fall in either direction.

The depth of both the upper/lower traps is also ~ 0.5K, and though technically the depth of the analysis trap is deeper it would regardless be limited by the temperature of $\bar{\mathrm{H}}$ trapped in the production traps.

A detailed diagram of the ALPHA-g lower PM and neutral trap can be seen in Figure 3.13, ALPHA-g was designed to be entirely symmetric axially, meaning that the designs for an identical, inverted trap further up the apparatus exist and was installed in August/September 2024, though it has not been tested as yet. The central analysis trap region is also not installed as of 2024 and therefore not displayed in the plot.

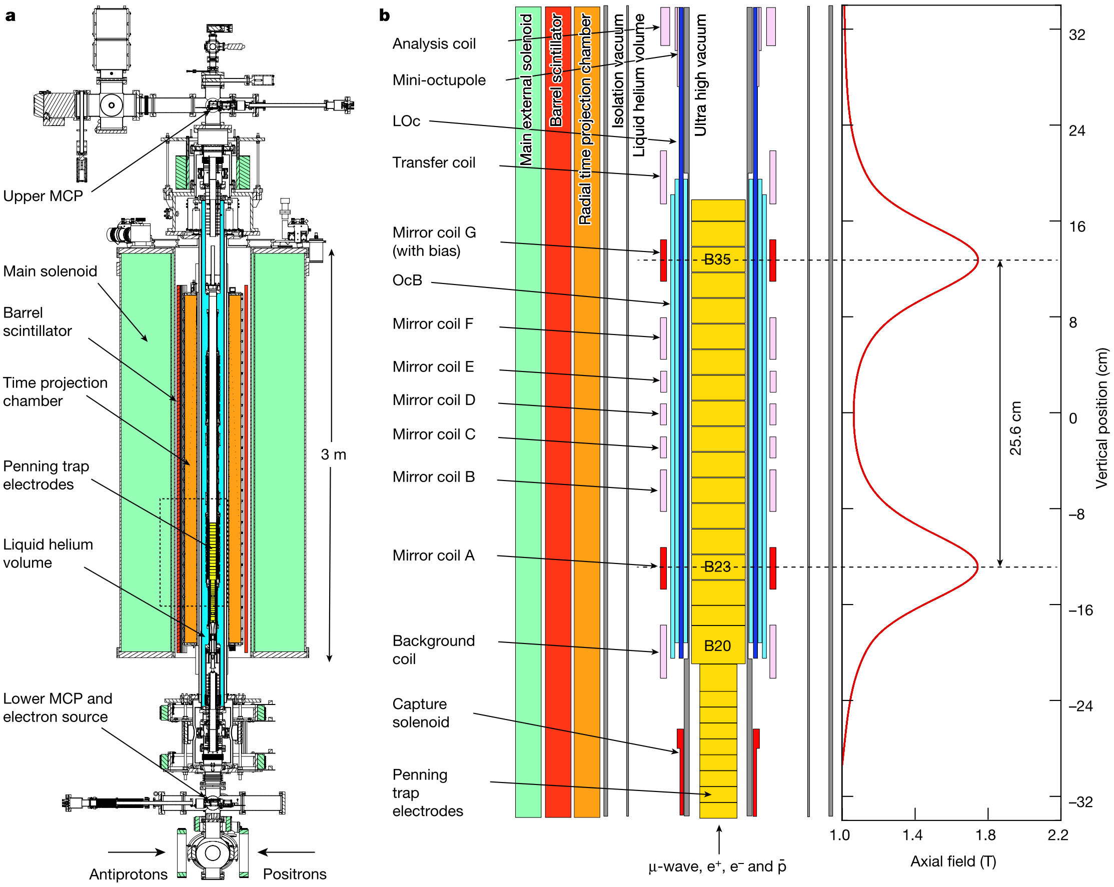


Figure 3.13: Detailed view of the ALPHA-g atom trap, a) shows a cross-section of ALPHA-g pointing out the locations of two MCPs, the detectors, and the direction of injection of antiprotons and positrons; b) shows an extended view of the dashed box in a) illustrating the PM and neutral traps (Note that the detectors and the main solenoid are not drawn to scale in this close up). The on-axis, axial field profile at full current is shown on the right. Figure from [17].

3.11 Beamline

Prior to the construction of ALPHA-g the positron accumulator was connected directly to the ALPHA-2 atom trap, meaning the ballistic transfer of the positrons into the atom trap was rather simple, and since the timing of the ejection can be controlled very accurately, the recapture was easily done using a similar re-catching system to that described in Sec. 3.7. However, after the ALPHA-g apparatus was constructed, and placed between the positron accumulator and the ALPHA-2 atom trap, it was required for some method of not only the linear (axially) transfer into ALPHA-2, but also a method of positrons being transferred upwards, around a 90° bend, into the ALPHA-g trap. Further, the beamline must be multispecies and reversible, capable of also transferring antiprotons through the ALPHA-2 atom trap upwards 90° into ALPHA-g.

A novel beamline was constructed and commissioned in 2018 to address this problem [126].

In most other beamlines used for transferring particles, the plasmas are on the order of MeV to GeV, which are less strongly effected by the magnetic field. ALPHA's catching trap and positron accumulator are only capable of transferring plasma with energies of 10-100eV[6], which can be perturbed significantly from their path in the presence of a magnetic field. For example, a positron plasma being transferred at 50eV would be perturbed radially by $\sim$ 20cm in the presence of a 10^{-4}T field [126], whereas conventional magnetic or electrostatic lattice beamlines can generate stray fields on the order of 100s of times greater than this. As such, ALPHA opted to implement a magnetically guided beamline, using a series of solenoids to focus and steer the beam in the transverse directions.

A special part of the beamline, known as the "interconnect", is used to steer positrons and antiprotons around the 90° corner into ALPHA-g. The interconnect is an arrangement of seven independent magnets that act to create a curved magnetic field, allowing for multidimensional transfers. A diagram of the interconnect, as well as the simulated field lines, can be seen in Figure 3.14.

Further, there are multiple MCPs along the beamline, allowing us to diagnose and confirm the beam is following the expected path.

Other experiments have previously implemented this type of 90° bend, and the ASACUSA experiment (also operating out of the AD hall facility) also has a beamline for the transfer of 100eV positrons around a 90° corner [127].

---

[6] Even if possible, in favour of lower temperatures ALPHA would prefer to transfer with lower energies.

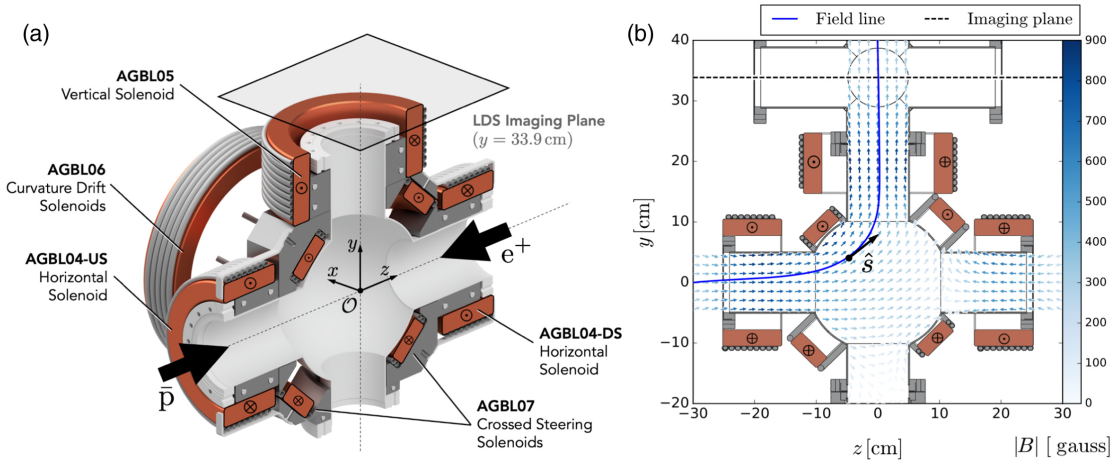


Figure 3.14: (a) Simplified schematic showing a cross-section of the interconnect. The magnet windings are highlighted in orange, while the other structures are in grey. One of the AGBL06 magnets (placed at $-x$ of the visible one) cannot be seen. Crosses (dots) indicate the current flowing into (out of) the shown cross-section of each magnet. (b) Quiver plot showing the strength (colour) and direction (arrow orientation) of the magnetic field within the mid-plane of the interconnect. The blue line shows a magnetic field line traced from the horizontal axis of the experiment, while the dashed line indicates the location of an MCP. Figure from [126]

3.12 Antihydrogen Formation - "Mixing"

Antihydrogen synthesis at ALPHA is done by mixing the positrons and the antiprotons together in a PM trap. Other experiments operating from the AD hall have also used this process to synthesise antihydrogen in the past [11, 37, 128].

Since antiprotons and positrons are oppositely charged, and therefore held in oppositely charged wells, overlapping them in a PM trap is not trivial. The two wells storing the plasmas are opposite to each other and therefore in order for the two populations to interact with one another one of the plasmas must receive enough energy to overcome the barrier holding the other plasma. Prior to the introduction of SDREVC this was done via "auto-resonant" mixing which involves using a chirp generator to pulse the electrode holding the antiprotons, driving the antiprotons into the positron cloud[7]. This resulted in about one trapped antihydrogen atom per mixing cycle.

However, simulations had shown that given consistent plasmas (something

---

[7] It was shown that driving the antiprotons into the positrons provided better results than the opposite setup [129].

now possible with SDREVC) an alternative method of mixing would be optimum for antihydrogen production, the *slow merge* or "smerge" [130].

3.12.1 Slow Merging - "Smerge"

The slow merge or "smerge" is a method of mixing the two populations which does not require the driving of extra energy into one of the plasmas. Instead, if one lowers the potential barriers of each population in a specific way as to cause EVC on both plasmas, in opposite directions, the two plasmas will evaporate into each other, causing mixing. A sketch of this process can be seen in Figure 3.15.

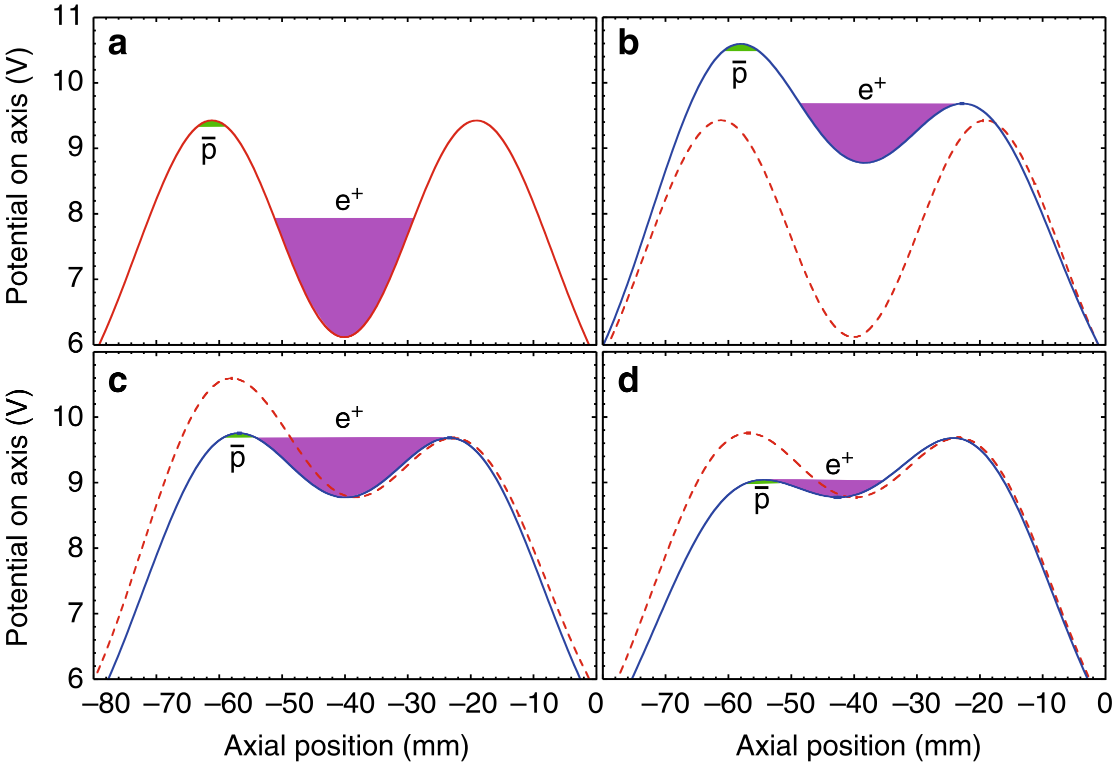


Figure 3.15: Different stages of the smerge process. The dashed line represents electrostatic potentials before each step, and the solid lines shows the potentials after the step. a) Potentials prior to mixing process. b) EVC, where positrons escape to the right. c) Potential adjustment in preparation for mixing. d) Smerge mixing, the wells slowly shrink until both plasmas have fully evaporated. This whole process can take anywhere from 0.5-5s. Figure from [120].

The result of utilising this smerging process over auto-resonant mixing was an immediate jump from 1 $\bar{\mathrm{H}}$ per cycle to $\sim$ 20. While the process of smerging has remained roughly similar since it's addition to the ALPHA experiment some variables of its exact implementation have shifted over the years. For example, while mixing was 1s in 2016-2018, with the introduction of Be^+ (see Sec. 3.9.1) it was found that 2s mixing steps gave better results. Further, the EVC step portrayed in Figure 3.15b is no longer used, and the positrons are actually mixed with the Be^+ which is simultaneously being laser cooled during this mixing process. The antihydrogen produced in this method is usually a result of TBR (see Sec. 2.2), and with the inclusion of Be^+ cooling into the experiment has peaked at 186 $\bar{H}$/cycle.

In this step, anywhere from 50-400k $\bar{H}$ can be produced with close to 100% efficiency; however, only those that are cooler than the trap depth of 0.5K will remain trapped, generally 50-120 per cycle. ALPHA is able to repeat this step over multiple cycles without incurring major losses, as demonstrated in [120], accumulating upwards of 15000+ atoms for high statistics experiments. The antihydrogen lifetime in the ALPHA trap has been demonstrated to be over 66 hours [131] allowing for the possibility of stacking over multiple days. Though this number is dependent on vacuum conditions, rarely is this a problem for stacking. Each stack takes two cycles of the AD/ELENA to complete, amounting to $\sim 50 - 120\bar{H}$ trapped every 5 minutes.

The data obtained during this mixing process provides an excellent candidate for machine learning applications. More details on the data obtained during this process, including rates, distributions, etc., can be found in Sec. 8.3.

3.13 Laser Cooling

Once the antihydrogen is trapped in the neutral well it is ready to be used for physics experiments. However, for many experiments, the line shape we are measuring is a function of the temperature of the sample. For example, in the case of $1S - 2S$ spectroscopy the transit time broadening affects the natural linewidth. Further, in any hyperfine/spin-flip experiment which is dependent on the magnetic field, the temperature of the antihydrogen atoms defines how far from the B-minimum the atoms can deviate, changing the frequency of spin-flip transition levels and broadening the spectrum.

Once trapped the energy of the $\bar{H}$ is < 0.5K, and is also Maxwellian. However, the mean and shape of this distribution is unknown, only its maximum, and it would be nice to have a method to consistently control this distribution from experiment to experiment. Doing so would place a limit on the effect of the transit-time broadening for a given experiment, or on the magnetic field contribution in microwave experiments. Further, being able to reliably reproduce a sample of $\bar{\mathrm{H}}$ in terms of temperature would allow us to systematically study these effects.

While photons are massless, they carry momentum and can therefore exert a force upon other particles they collide with [132]. Laser cooling then, is a process for cooling atoms where photons from laser light injected into a trap collide with the atoms stored within it, slowing them and reducing their transverse energy [133]. While ALPHA is only able to inject light on axis, and laser cooling only occurs in the single dimension parallel to the laser, the ALPHA neutral trap couples the motion of atoms to all three dimensions, causing an overall reduction in the energy of the particle.

Laser cooling was initially demonstrated on matter in 1978 [134, 135], on hydrogen in 1993 [136], and on antihydrogen by ALPHA in 2021 [16] after observing the 1S-2P transition in 2018 [15].

By exciting the 1S-2P$_{a+}$ transition with a Lyman-α laser and allowing the atom to decay back to the ground state, it is possible to Doppler cool the sample down from 44μeV (0.5K) to $\sim$ 12.5μeV (0.15K) [16][8].

Further, in the 2023 run, by applying laser cooling for varying amounts of time and measuring the temperature of the resulting sample through spectroscopy and time-of-flight analysis, we were able to create consistent and reproducible samples of laser-cooled $\bar{\mathrm{H}}$ with which to perform sensitive experiments on.

---

[8] In the paper the transverse energies were well-defined and reduced from $15.1 \pm 0.8\mu$eV to $1.3 \pm 0.1\mu$eV - however the longitudinal energies are more difficult to reconstruct, relying on spectral line shapes for reconstruction. Instead, an upper bound of $\sim 7.5\mu$eV is placed on the longitudinal energy.

Chapter

4 Detection & Reconstruction

Once antihydrogen has been trapped, the only method of observing its existence so far is through annihilation with matter. Upon annihilation (see Sec. 2.4) the resulting charged pions can easily be recorded given the correct detectors are in place. Both the ALPHA-2, and ALPHA-g PM traps are completely encased by their own unique detector systems which sit outside the trapping region and just inside the external solenoid to minimise travel distance (and also therefore scattering material) for the annihilation products.

Since the neutral traps have a maximum depth of 0.5K, and the hottest antiprotons and positrons get in the apparatus at any point is 2-3000K, the maximum velocity of particles annihilating at ALPHA is ~ 369km/s[1] we can treat all annihilations non-relativistically.

As discussed in Sec. 2.4, when positrons annihilate they release two, oppositely directed 511keV photons, and while the detectors that originally verified the production of antihydrogen contained the required scintillating crystals to detect these particles [11, 72], they are no longer needed; and instead, the ALPHA detectors focus on detecting the charged pions/kaons produced by annihilation with the goal to reconstruct the x, y, z position (or *vertex*) of the annihilation.

Since the detectors are external to the trap (both the PM, and neutral traps) the secondaries will have to travel through a lot of material before reaching any detector systems, this can sometimes cause scattering of the reconstructed tracks (discussed in Sec. 4.2.1).

Annihilations can occur on many different materials, usually though, they either occur on the trap wall (gold-plated electrodes) or on the residual gas in the vacuum chamber. The recent introduction of heavy beryllium ions into the experiment however has offered a new material for which annihilations can occur, though the fingerprint of this annihilation is generally similar to those on residual gas. Though the pressure in the traps is generally kept low (below 10^{-10}mbar) this does change throughout operation, and can cause some issues, especially during microwave exposure [2].

---

[1] This corresponds to a 3000K positron, antiprotons would be even slower.

[2] The presence of microwaves causes the electrodes to heat, boiling any frozen gasses off,

The ALPHA-2 trap is surrounded by a single detector known as the silicon vertex detector (SVD), while ALPHA-g has a time projection chamber (TPC) for vertex reconstruction, and additionally (owing to the large size of the ALPHA-g detector) a barrel of scintillating bars (the barrel veto), for use in cosmic background rejection.

The SVD relies on charged semiconductors to record the location of a passing charged particle through its modules to detect signals. The theory behind these semiconductor detectors is presented in Sec. 4.1, with the specifics of the SVD presented in Sec. 4.2.

The TPC is a form of *ionisation chamber*, which works similarly to semiconductor detectors but uses a gas instead of a solid material. The theory behind gas chamber detectors is presented in Sec. 4.3, and the specifics of the TPC in Sec. 4.4.1. The barrel veto which is a form of scintillating counter (already discussed in Sec. 3.2) is presented in Sec. 4.4.2.

4.1 Semiconductor Theory

4.1.1 Semiconductor Diode Detectors

Semiconductor diode detectors (SDDs) are detection devices that rely on electron-hole pairs produced by a charged particle passing through some semiconducting material for detection. The motion of these electron-hole pairs in an electric field generates the basic signal of the detectors. SDDs are beneficial for their ability to have fast timing characteristics, and to be quite compact. The most popular material for SDDs is silicon, but other possibilities exist.

4.1.2 Band Structure Model

Given that the electrons in a material all sit at a specific energy level, the lattice structure of crystalline materials allows for particular energy bands in which the electrons must be confined to. The gaps between these energy bands are known as “forbidden energies” and in general, electrons do not exist at these energy levels. A

---

temporarily reducing the vacuum in the trap, and increasing the likelihood of antihydrogen encountering a gas molecule (discussed in Chapter 9) simplified way of looking at this is to imagine each material as having two bands: the valence band and the conduction band. Electrons in the valence band are confined to specific lattice sites in the material and cannot move freely; whereas the electrons in the conduction band, are free to move throughout the material, therefore carrying charge. It follows that a material in which all electrons sit in the valence band, will not carry charge, and therefore not conduct electricity, and a material in which all electrons sit in the conduction band will conversely conduct very well. The difference between the energy levels of these two bands is what's referred to as the *energy gap* (denoted E_g), and the size of this gap is what defines the conductivity of a material. For example, a very good insulator has a large energy gap (often > 5eV), and, in absence of electric fields or thermal excitation will not have any electrons in the conduction band, and therefore nothing to carry charge. Something that conducts well, like for example metal, might instead have a zero or even negative energy gap. A semiconducting material is one where this energy gap is greater than that of metal, but not quite as large as that of a good insulator. Usually the energy gap for semiconductors is considered to be ~ 1eV. A diagram of this simplified model can be seen in Figure 4.1 showing roughly where these energy bands might sit for an insulator, a semiconductor, and a metal.

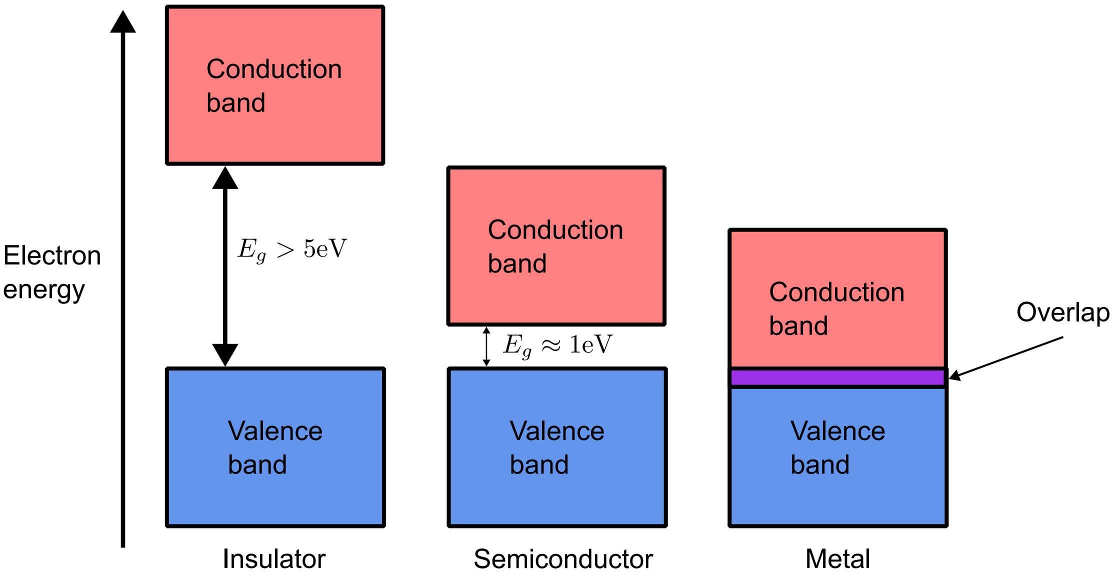


Figure 4.1: Illustration of the band structure for electron energies in metals, semiconductors, and a good insulator.

4.1.3 Pair Mobility

The arrangement of electrons in this structure is not constant. If a material has non-zero temperature, some of the thermal energy will be shared by the electrons and it is then possible for an electron in the valence band to gain enough energy to become excited into the conduction band. This process not only creates an electron that has left its bonding site and is free to travel throughout the material, but it has also left a hole behind in its original place in the valence band. This is known as an *electron-hole pair*, and by applying an electric field, both can be made to "move"[3] in opposite directions throughout the material, contributing to the materials observed conductivity. The probability of this energy jump happening at a given time is a function of the temperature of the material, T, and of the energy gap:

$$p(T) = CT^{3/2}e^{\left(-\frac{E_g}{2k_BT}\right)} \tag{4.1}$$

where C is the proportionality constant characteristic of the chosen material. As can be seen from the negative exponential term, this probability is dependent on the ratio of the energy gap, to the temperature. I.e. materials with a large E_g will be unlikely to experience this thermal excitation, and therefore will show the low conductivity properties expected of an insulator, even as you heat it. Without the application of an electric field, the electron-hole pairs will eventually recombine returning the material back to its equilibrium state. If however, an electric field is applied to the material the electron-hole pairs will undergo an overall migration which is a combination of the random thermal velocity, and of the drift-velocity coming as a result of the electric field. The electrons in the material will be drawn in the opposite direction of the electric field, causing the holes to be filled by electrons further downstream (relative to the electric field), these electrons and holes act as negative and positive point charges moving throughout the material and is what causes a material to conduct charge.

The *mobility* of both the holes and the electrons can be expressed (in a mod-

---

[3] A hole is said to have moved when an already bound electron leaves a valence spot to fill the hole, resulting in the hole "moving" to the spot where the electron once was.

erate electric field) in terms of the drift velocity (ν) to be

$$\mu_{\text{holes}} = \nu_{\text{holes}}/E_0 \tag{4.2}$$

$$\mu_{\text{electrons}} = \nu_{\text{electrons}}/E_0 \tag{4.3}$$

where E_0 is the magnitude of the electric field. In semiconductors the mobility of both the holes and the electrons are roughly similar, and as the electric field increases the drift velocity also increases logarithmically until a *saturation velocity* is reached. This usually occurs at velocities of 10^7cm/s which corresponds to the charge carriers moving through 0.1cm of silicon (the typical thickness used in SSDs) in 10ns, making SSDs one of the fastest responding detectors, and explains why most detectors are operated with fields strong enough to reach the saturation velocity [137]. The electron-hole mobility of a material is an important property when selecting a material for detectors, as fast readout (on the order of nanoseconds) is required. The collection time of the $\sim$ 0.3cm thick silicon modules used in the SVD, is $\sim$ 30ns, making it extremely fast, especially compared to other detector systems [129]. Some typical values for these properties in silicon at different temperatures are listed in Table 4.1.

	Temperature (K)			
	$\forall T$	0	77	300
Atomic number	14			
Atomic weight	28.09			
Stable isotope mass numbers	28-29-30			
Density (g/cm^3)				2.33
Atoms/cm^3	4.96×10^{22}			
Dielectric constant (relative to vacuum)	12			
E_g (eV)		1.165		1.115
Intrinsic carrier density (cm^{-3})				1.5×10^{10}
Intrinsic resistivity Ω·cm				2.3×10^5
Electron mobility cm^2/V·s			2.1×10^4	1350
Hole mobility cm^2/V·s			1.1×10^4	480
Energy per electron-hole pair (eV)			3.76	3.62

Table 4.1: Some important properties of silicon at different temperatures, from [138].

4.1.4 Pure/Dopants

In the absence of any ionising radiation, and assuming a pure semiconductor material, we must have an equal number of electrons in the conduction band as holes in the valence band, as every electron that jumps to the conduction band via thermal excitation must leave behind a hole in the valence band. A material such as this is referred to as an *intrinsic* semiconductor. However, purity at this level is very rarely achieved in reality, even in silicon, which is generally the highest-purity semiconductor that can be produced.

Formally, an intrinsic semiconductor is one in which the concentration of electrons in the conduction band, n, (for “negative”) will be exactly equal to the concentration of holes in the valence band, p, (for “positive”). These two quantities are referred to as the *intrinsic carrier densities* and represent the average carrier density over the whole material, offering a measure for the conductivity of the material. It can be seen from Table 4.1 that pure silicon contains 15 billion of these electron/hole pairs per cm^3.

When producing semiconductors, intentionally introducing impurities into the material can sometimes cause it to acquire some useful properties. This process is known as “doping”, and the resulting material is known as an *extrinsic* semiconductor.

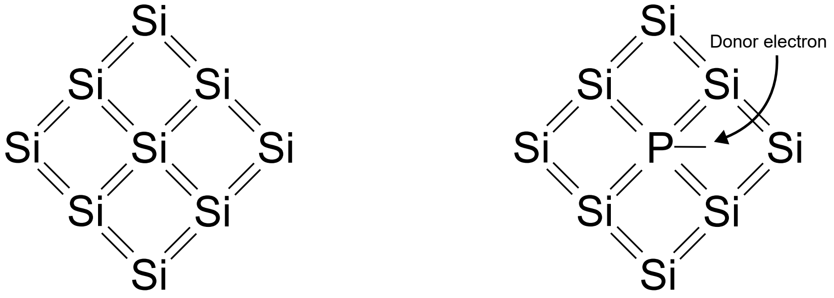
There are two types of doping used in semiconductor production, referred to as *n-type* and *p-type* doping, resulting in *n-type* and *p-type* semiconductors.

4.1.5 n-Type Semiconductors

n-type semiconductors are created by doping the pure material with extra donor electrons during manufacturing. While silicon, germanium and even diamond can be used in SDDs [139], to illustrate this process we use silicon as an example.

Since the atomic structure of silicon is tetravalent, each silicon atom forms covalent bonds with its four nearest neighbours, as illustrated in Figure 4.2a.

By introducing a *dopant* to the material in sufficiently small quantities (usually $\sim$1-5ppm) the lattice structure of the silicon will be maintained (as the dopant will take the place of a silicon atom in the lattice) and, by selecting the doping material from group V of the periodic table (i.e. it is pentavalent) the material will not only bind in place of a silicon atom, but will have an extra electron weakly bound to the impurity site, as illustrated in Figure 4.2b.



(a) Pure silicon.

(b) *n*-type doped silicon, with a donor phosphorus atom.

Figure 4.2: Silicon lattice structure for both pure (Figure 4.2a) and *n*-doped silicon (Figure 4.2b).

It follows that this electron will take much less energy to dislodge it into the conduction band, and will not leave behind a corresponding free moving hole[4]. These electrons generally have energies that lie in the usually forbidden energy gap, at a new level known as the *donor level* (see Figure 4.4). The new gap formed between these weakly bound electrons and the conduction band is sufficiently small that the probability of thermal excitation is high enough to ensure a large fraction of all impurities become ionised, and the number of conduction electrons becomes completely dominated by the contribution from the impurities.

Despite the fact that the net effect of this is a dominance of conductivity through the electrons instead of the holes (as the majority of electrons in the conduction band have not left behind a hole) the total conductivity of silicon doped in this way is much larger than that of pure silicon.

The net effect then of *n*-type doping is to create a material where the number of conduction electrons is much greater than the number of conduction holes. The conductivity of this material is now determined almost entirely by the movement of electrons, and the holes play a very small part in this determination. In this case the electrons are referred to as *majority carriers*, and the holes as *minority*

---

[4] The hole is still formed, maintaining the charge neutrality of the material, but they are no longer free to move throughout the material since the donor atoms are fixed in the lattice and cannot migrate, removing the effect of their mobility to the overall conductivity of the material.

carriers.

4.1.6 p-Type Semiconductors

Similarly to n-types, by introducing an impurity from group III of the periodic table (trivalent), when the atom replaces a silicon atom, it will instead have one fewer valence electron than the surrounding silicon, as illustrated in Figure 4.3.

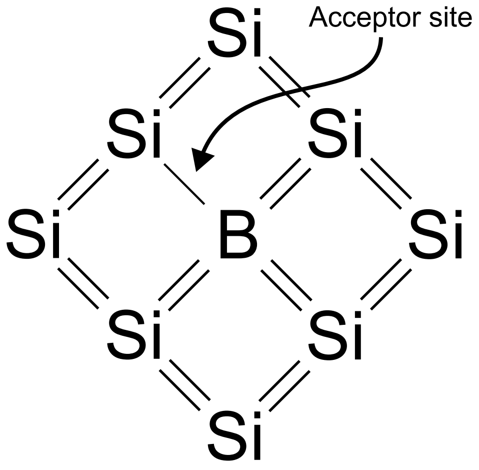


Figure 4.3: p-type doped silicon, with a donor boron atom.

If an electron is then captured for this spot the bond it is confined by is slightly weaker than that of the surrounding material. Therefore, these electrons once again reside in the previously forbidden energy gap, near the valence band known as the *accepter level* (see Figure 4.4).

It again takes less energy for the electrons in the valence band to fill these holes, and each time an electron is excited to the *acceptor* energy level, it leaves behind a corresponding hole in the valence band.

This results in a dominance of holes coming from the acceptor impurities. Further, this increased availability of holes makes it more likely for electrons to fill them, reducing the number of free conduction electrons available.

The filled acceptor sites again represent fixed negative charges that balance the positive charge of the majority holes, keeping the net charge neutrality of the material as a whole; and analogously to n-types the net result of this process is that the availability of conduction holes is greater than that of the electrons,

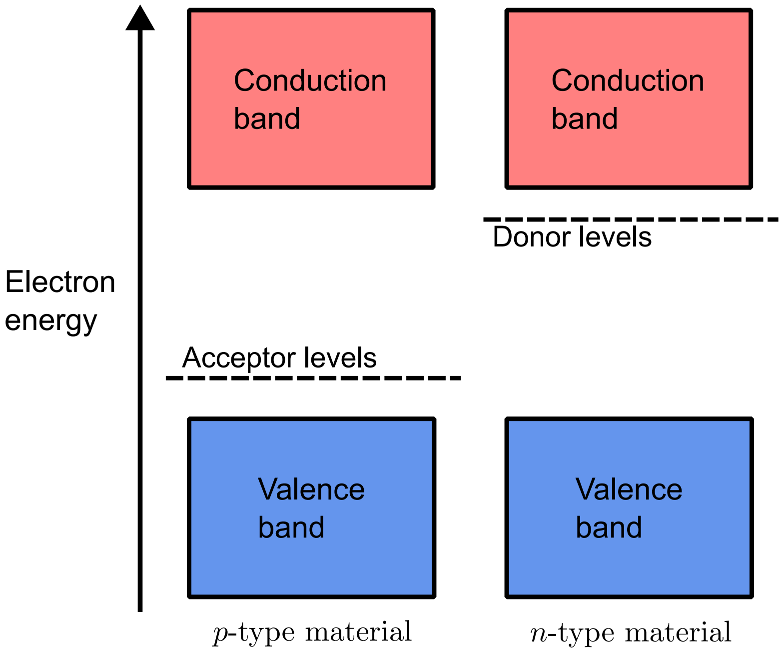


Figure 4.4: Energy bands in n-type and p-type semiconductors.

making the holes the majority carriers and the electrons the minority carriers.

4.1.7 *p*-*n* Junctions

When a material is doped such that part of it *n*-doped, and another part is *p*-doped the region where both the *p*-type and *n*-type regions meet is known as a $p-n$-junction [79]. In this region electrons from the *n*-type material diffuse across and fill holes in the *p*-type material, and vice versa, in a process known as *diffusion*. This creates a region within the material known as the *depletion layer/region* and acts as an insulator, as within this region there are no free electrons or holes to move throughout the material carrying charge.

The build-up of charge in the depletion region causes an electric field to form, the potential of which at any point within the region is described by the Poisson equation:

$$\nabla^2\varphi = -\frac{\rho}{\epsilon} \tag{4.4}$$

where ϵ is the dielectric constant of the material (~11.9 for silicon), and ρ is the charge density.

Any electrons or holes that exist in or near the depletion region will then migrate quickly back to the *n* or *p*-doped regions as required by the electric field.

Meaning, the only charges left within the depletion region are immobilised donors and filled acceptors, which do not contribute to the conductivity. As such the resistivity in these regions is high compared to that of the p and n-doped regions.

Therefore, when a charged particle travels through the depletion region, electron-hole pairs are created in great numbers, which then drift quickly according to the electric field, and create a burst of current through the material which can be detected. It is this process of electron-hole pair creation and the quick reset to equilibrium that makes sharp, fast signals in the charge of the material, allowing for the detection of particles through this region. However, without further adjustments a detector like this does not perform well. The potential, φ (usually $\sim$ 1V) which is formed across this junction is not strong enough to make the charge carriers move fast enough, resulting in charge carriers being lost due to recombination and occasional signals not being collected, causing a significant decrease in efficiency. Further the depletion region itself may not be big enough to provide enough solid angle coverage, which can reduce efficiencies even further.

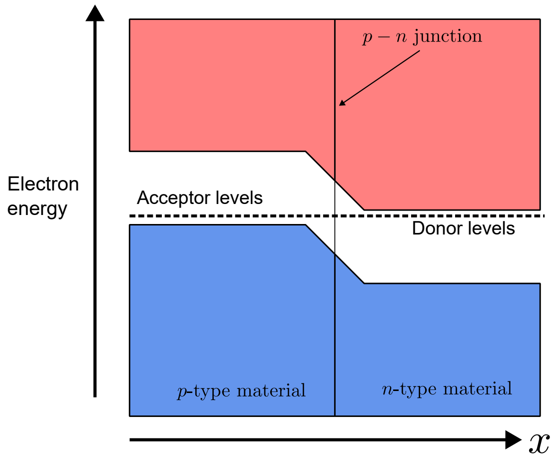
For this reason, unbiased junctions are not used. Instead, an external voltage is applied over the junction, resulting in a "reverse bias"[5] on the semiconductor. Under a reverse bias the electric field penetrates further into the silicon bulk, increasing the size of the depletion region and therefore extending the volume over which radiation-produced charge carries will be detected. Further, electron-hole pairs created as a result of a charged particle will travel faster and more forcefully towards equilibrium, allowing for greater efficiency and faster readout times.

A schematic of the band structure under no bias, forward bias, and reverse bias can be seen in Figure 4.5.

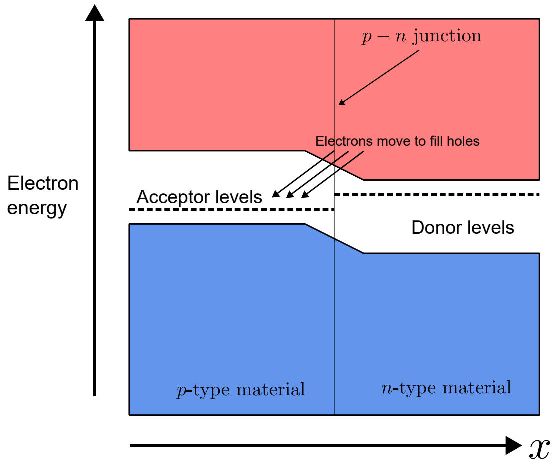
When a charged particle passes through a $p-n$ junction, it will ionise the atoms along its path causing the electron-hole pairs. The average energy loss per unit distance $\langle dE/dx \rangle$ (also referred to as the stopping power) for a moderately relativistic charged particle passing through the bulk of a material is described by

---

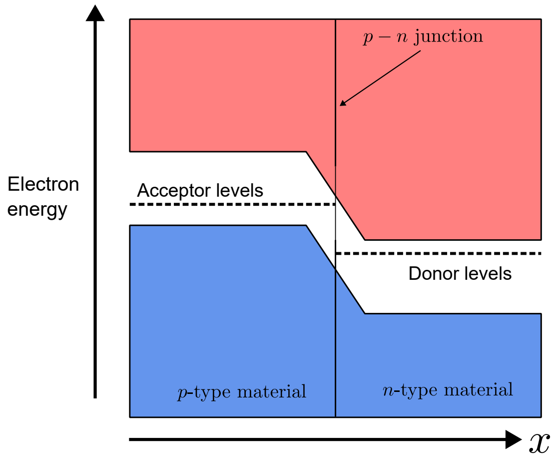
[5] Here forward means applying a positive voltage to the p-side and reverse as applying a positive voltage to the n-side.



(a) No bias - With no bias the junction is in equilibrium, and the donor level matches the acceptance level.



(b) Forward bias - With a forward bias applied the donor level moves above the acceptor level, allowing electrons to freely move “down” into the p-side from the n-side, resulting in a net current through the diode.



(c) Reverse bias - With a reverse bias applied the acceptor band is moved above the donor level, meaning some form of thermal excitation (i.e. a charged particle passing through the region) is needed to allow current to flow.

Figure 4.5: $p-n$ junction characteristic behaviour for no bias, forward bias and reverse bias.

the Bethe-Bloch formula [140]:

$$-\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi}{m_e c^2} \cdot \frac{nz^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\varepsilon_0}\right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I \cdot (1-\beta^2)}\right) - \beta^2\right] \tag{4.5}$$

where v is the velocity of the particle, z is the charge number, n is the electron density of the material ($n = \frac{N_A \cdot Z \cdot \rho}{A \cdot M_u}$ where ρ is density, Z is atomic number, A is atomic mass, N_A is the Avogadro constant, and M_u is the molar mass constant), and finally $\beta = v/c$.

A feature of this equation is that it has a global minimum, and particles near this minimum are known as *minimum ionizing particles* (MIPs).

Further, a charged particle passing through matter will undergo many small angle deflections, mostly due to the Coulomb scattering from the nuclei. This process is known as *multiple Coulomb scattering.* The distribution of the resulting angles is roughly Gaussian for “small” deflections[6], and is well described by Molière [141] demonstrating that the resulting distribution of angles of many particles is roughly Gaussian with a width given by

$$\theta_0 = \frac{13.6 \text{ MeV}}{vp} z\sqrt{x/X_0}\left[1 + 0.038\log(x/X_0)\right] \tag{4.6}$$

where p is the momentum of the incident particle and x/X_0 is the thickness of the scattering medium in radiation lengths [142, 143]. For larger angles it behaves more like Rutherford scattering which has larger tails than a Gaussian distribution [144].

This value of θ_0 comes from a fit to experimental data from Molière [141] and is accurate to at least 11% for $10^{-3} < x/X_0 < 100$.

Understanding these concepts are essential in the design of any detector as correctly simulating these properties allow one to properly estimate the expected performance/response of the detector during the design phase, and incorrectly accounting for the stopping power of the detector, or the scattering throughout could result in a detector not fit for purpose.

---

[6] “Small” in this context is less than a few θ_0, defined in Equation 4.6.

4.1.8 Double-Sided Silicon Microstrip Detectors

By using multiple strips of $p-n$ junctions arranged in a grid it is then possible to create a module/block of silicon that is sensitive to the position of charged particles passing through it. Further, by having two layers of strips per module, arranged orthogonally one can gain three-dimensional[7] information on the location of the particle passing through the module [145]. Generally, one side contains p-type strips, and the other side contains n-type strips sat orthogonally, as shown in Figure 4.6.

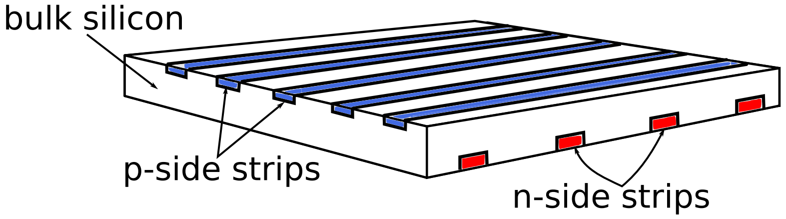


Figure 4.6: A diagram of a double-sided silicon microstrip detector. The p-side strips on one side of the module run orthogonally to the n-side strips on the opposite side, allowing for two-dimensional resolution. Figure adapted from [146].

Pixel detectors such as the Timepix [147] series use similar technology but instead of two layers of strips, they segment the junctions into squares or “pixels” to obtaining $x-y$ resolution [148].

The spacing between each strip or pixel can often range from $50-300\mu$m [149–151] and overall, double-sided silicon detectors can provide fast and stable particle detection with excellent spatial resolution.

4.2 Silicon Vertex Detector

The SVD is made up of 72 individual double-sided silicon microstrip detectors, or *modules*. The 72 modules are arranged in 3 concentric circles, split into two halves. A schematic diagram of the detector with respect to the PM trap can be seen in Figure 4.7, and a cross-section of the panels can be seen in Figure 4.8.

Each layer is staggered to allow for more solid angle coverage, and the radii of the layers are shown in Table 4.2.

---

[7] The third dimension comes from knowing where in space the modules are placed.

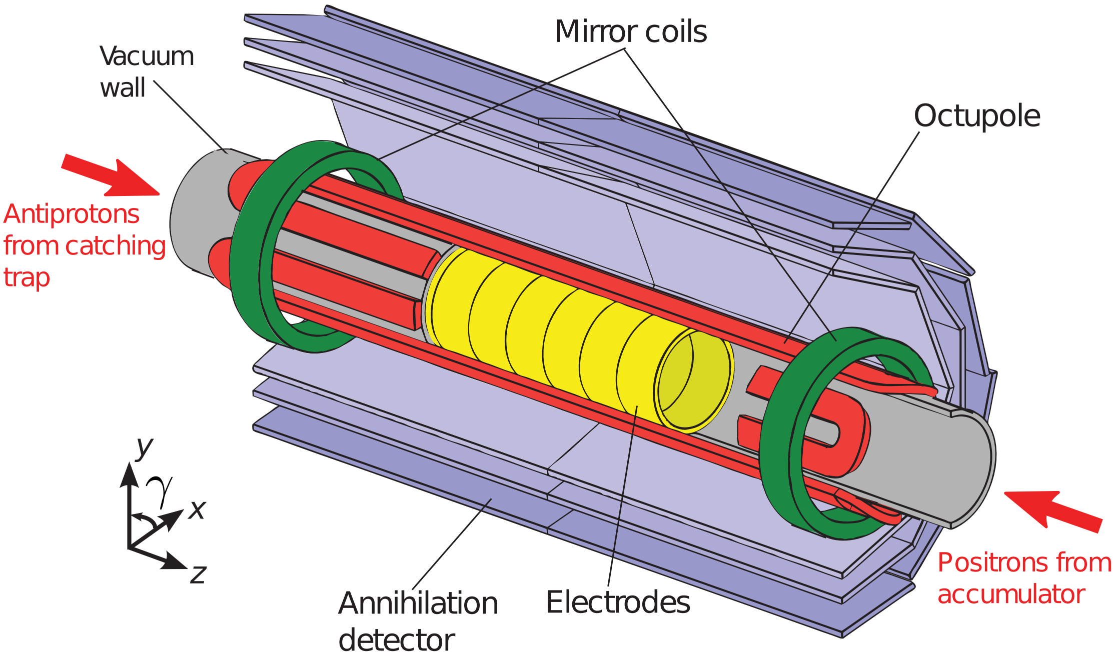


Figure 4.7: A schematic diagram of where the ALPHA-2 detector sits in respect to the ALPHA-2 PM trap. The mirror coils are shown in green, and the octupole is in red. The PM trap electrodes are shown in yellow, and the surrounding silicon detector is light blue. Though this image is of a slightly older configuration than the current, the placement of the detector with respect to the PM trap is still the same. Figure from [119]. (Not to scale.)

SVD Module Layer	Radius
Inner	89 94.5
Middle	108 113.5
Outer	127 132.5

Table 4.2: Radius of each panel in the silicon modules, per layer.

Each module contains two silicon wafers and 256 n-side and p-side strips with four Application Specific Integrated Circuit (ASIC) readout chips. The silicon wafers are 6.1cm × 11.5cm meaning each module has an active area of 6.1cm × 23.0cm = 140.3cm^2. The pitch width is 227μm for the p-side strips, and 875μm for the n-side strips, and each ASIC reads out 128 signal strips making for a total of 36,864 channels. The ASICs handle both the trigger signal, and the readout for each strip. Each channel on the ASIC contains a discriminator which generates a fixed 105nm trigger pulse if the signal exceeds the preset discriminator threshold

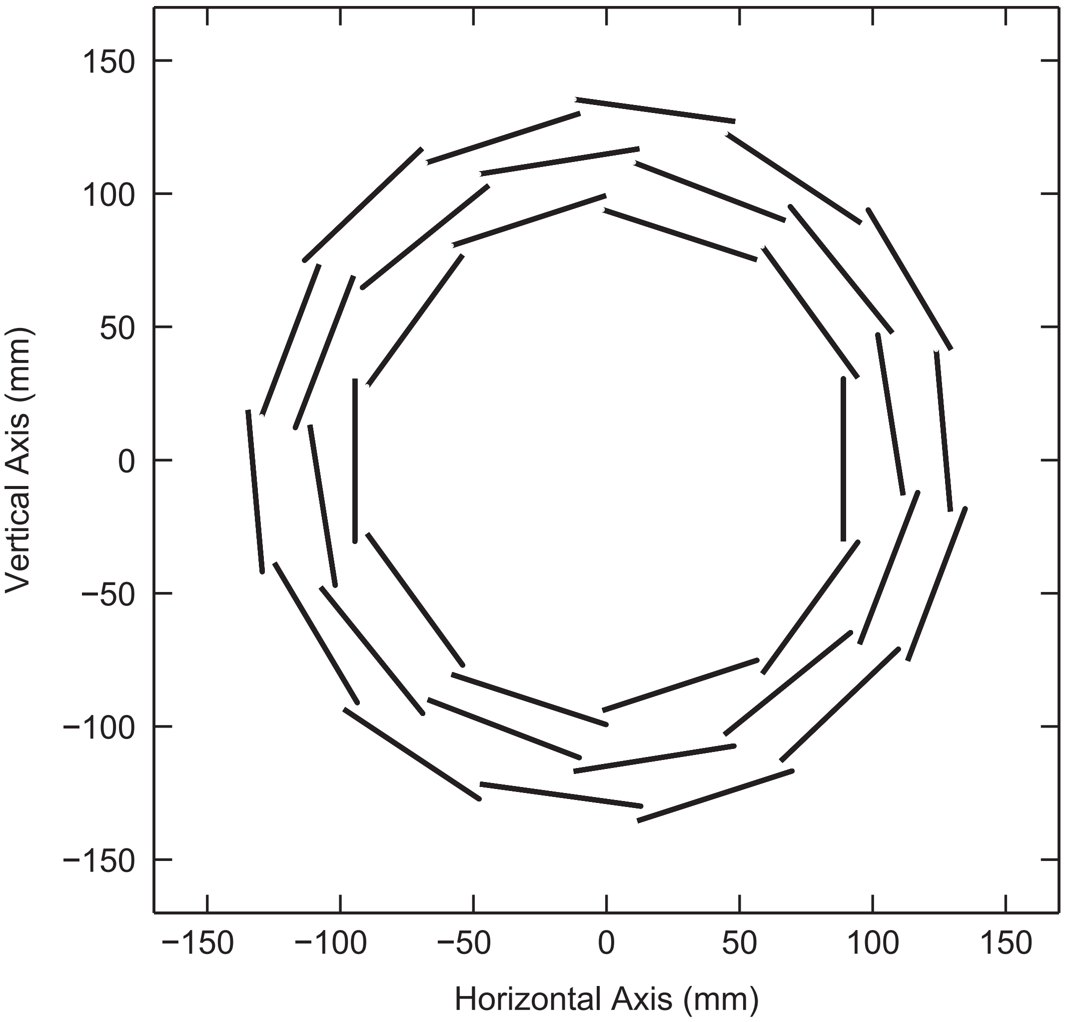


Figure 4.8: Cutaway of the SVD showing the layered module configuration. Figure from [152].

causing readout of every channel on the detector. Though each channel has its own trigger signal, they are logically ORed such that only one trigger signal is delivered out per ASIC. Further, each ASIC only has one analogue out channel requiring the signal on each channel to be read out serially. A diagram of a single silicon module can be seen in Figure 4.9.

4.2.1 Event Reconstruction

The event reconstruction algorithm attempts to determine the annihilation location (or *vertex*) of the event causing the detector trigger and read-out. This process is divided into roughly three parts: (1) the hit clustering, which attempts to find the location of hits on each silicon module; (2) the reconstruction of sec-

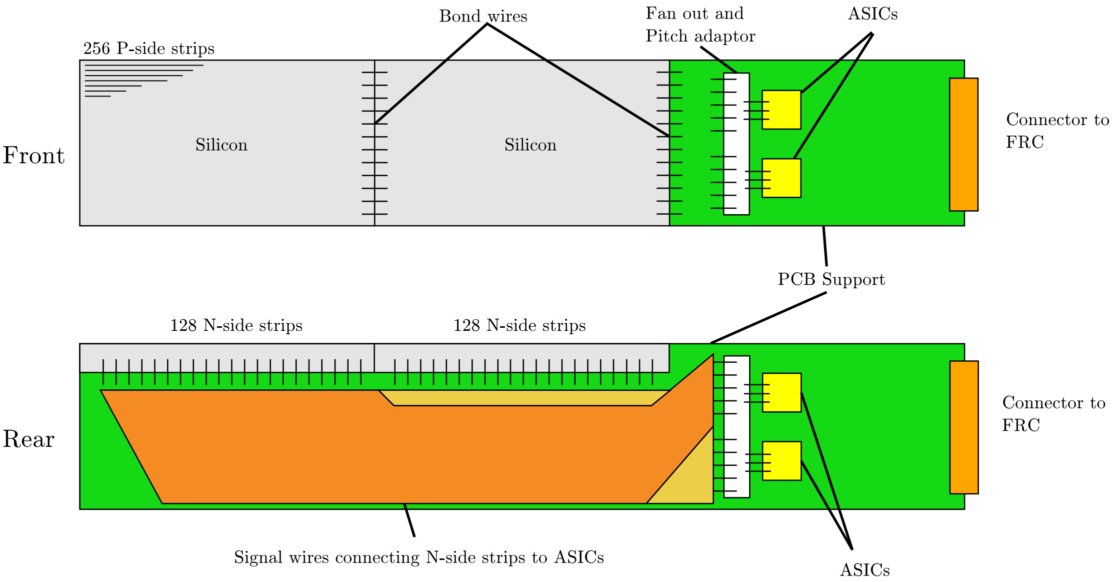


Figure 4.9: Diagram of a single silicon module (not to scale) showing each silicon wafer and the readout ASICs. Figure from [146].

ondary trajectories caused as a result of annihilation; and (3) the determination of the vertex location.

The first step to reconstructing the vertex is to reconstruct where the charged secondaries pass through the silicon modules. The location of each strip and each module is well known and therefore the read-out position can be easily reconstructed from the signals on the ASICs. However, it is possible, depending on the angle that particles travel through the module, for multiple strips to register a signal in a single event. In this case, the electron-hole pairs generated by the particle's path through the depletion region of the module will be shared between many strips. The resulting groups of hits are referred to as *clusters* and actually allow us to better determine the position of a hit as compared to a single strip by weighting the pulse height of each channel on the ASIC. The resulting position of the hit (in one dimension) is then given by

$$x = \frac{\sum_{i=0}^{127} h_i x_i}{\sum_{i=0}^{127} h_i} \tag{4.7}$$

where x_i is the x position of strip i, and h_i is the height of the recorded pulse on that strip.

By combining the location of clusters on the p-side strips, and the n-side strips we are able to completely reconstruct the location of the hit in the plane of the silicon module. Further, since the location of the module's placement in a global coordinate frame is known, we can translate these hits into absolute x, y, z positions in the lab frame. In this way each event, usually consisting of $\sim 9-15$ module hits, can be transformed into a series of points in 3D space (sometimes referred to as *space-points*).

The next step in the reconstruction algorithm is to group the hits into tracks, where each track represents the path of a charged particle through the detector.

The relatively small number of hits per event allows for a full examination of all hit combinations when attempting to reconstruct tracks, which offers a significant advantage in track finding. Each of these combinations is known as a track candidate, and consists of exactly three hits, one in each layer of the detector. Each track candidate has six degrees of freedom (as each space-point has two - x, y in the module plane); and, by using Monte Carlo simulations it has been shown the tracks resulting from $\bar{\mathrm{H}}$ annihilation are generally helical [153], therefore the actual tracks the particles take have only five degrees of freedom[8]. Thus, after determining the five helix parameters, there is only one remaining degree of freedom for tracks, corresponding to the axial projection of the helix. Track candidates can then be selected based on how well the fits correspond to helices in the axial projection. As such a χ^2 goodness-of-fit test is constructed comparing the points the predicted helix passed through the silicon module with the actual hits recorded. Note that since each candidate only has three hits, the projection of the helix into the x, y plane (here z is the direction of the magnetic field) forms a circle, which will certainly pass through each point. As such, these terms do not contribute to the goodness-of-fit test, and only its axial projection contributes. Tracks with a sufficiently low χ^2 (see Appendix A, and specifically Equation A.5) are kept for the final stage of the process, and the rest are discarded in a process known as *pruning.* The tracks that are kept for vertexing are known as the *good/used* tracks.

The efficiency of this track finding process, evaluated using Monte Carlo simulations, was found to be $88 \pm 5\%$ for all charged tracks with three hits. Further, many of the tracks that were not resolved originate from e^+ /e^- pairs produced

---

[8] Helices can be parametrised by a 5-component vector. See Appendix A.

from scattering in the silicon itself, and as such the tracks do not generally extrapolate back to the annihilation vertex, making their exclusion beneficial.

Once the tracks have been found the final stage of the reconstruction is to resolve the vertex position, which is considered to be the location where the tracks converge, and is taken to be the position where the tracks mean *distance of closest approach* (DCA) is minimised. The DCA is defined as

$$D = \frac{1}{N_{\text{tracks}}} \sum_{i=1}^{N_{\text{tracks}}} \sqrt{\min\{|\mathbf{r}_i - \mathbf{r}_v|^2\}} \tag{4.8}$$

where $\min\{|\mathbf{r}_i - \mathbf{r}_v|^2\}$ is known as the distance of closest approach between the i-th track and the reconstructed vertex position, $\mathbf{r}_v$.

The minimisation of which results in a vertex position that balances contributions from each track equally. Since all tracks are treated equally it is important to make sure that we include as many tracks as possible, and that we exclude tracks that do not converge with the others as these tracks could potentially bias the resulting vertex position away from the true annihilation position. For this reason a multistep vertex finding algorithm is implemented to reduce sensitivity to specific tracks, and ensure the vertex position is stable. The steps to the algorithm are as follows:

1. If $N_{\text{tracks}} < 2$ do not proceed. At least two tracks are needed to find a vertex. If this occurs prior to any track exclusion and vertex determination the event is considered to not have a vertex. These events may be useful for some studies but in general are not used.

2. Using all tracks find the vertex that minimises Equation 4.8 and find the resulting value of D, denoted D_0.

3. Loop through each track and repeat step 2 excluding said track, resulting in a set of DCAs: $\mathcal{D} = \{D_i \mid i \in [1, N_{\text{tracks}} + 1]\}$, the minimum of which is D_{min}.

4. Calculate $\Delta D = (D_0 - D_{\text{min}})/D_0$ which represents the *fractional* improvement in the mean DCA by excluding that specific track.

5. If $\Delta D \leq D_c$ for some D_c known as the cut-off value, then exit the algorithm and return the given $\mathbf{r}_v$ position. Otherwise, continue to the next step.

6. If $\Delta D > D_c$ then remove the track corresponding to current value of ΔD and repeat the algorithm from step one. If there are now less than two tracks return the value of $\mathbf{r}_v$ and do not proceed.

The choice of D_c used by the ALPHA-2 reconstruction (informed by Monte Carlo simulations) is 0.4 [153].

An example of a reconstructed event can be seen in Figure 4.10.

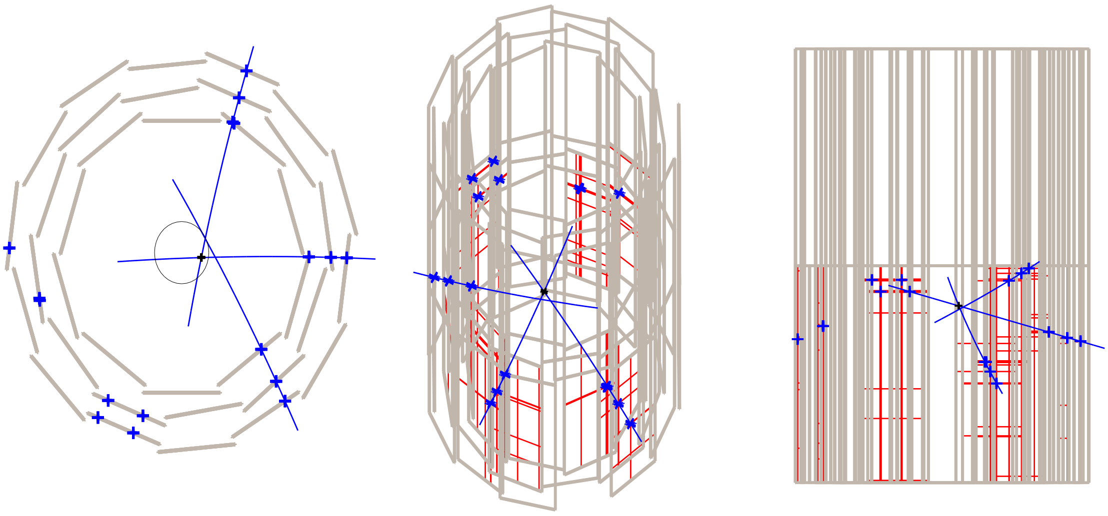


Figure 4.10: An example of a reconstructed event seen from three different angles. Three module hits (blue crosses) are combined into tracks (blue lines), and the tracks are combined into a vertex (black cross).

By simulating many events and comparing the true vertex to the reconstructed vertex we are able to find the resolution of the detector and reconstruction algorithm as a whole where we find the radial resolution to be 0.87cm, the axial resolution to be 0.56cm, and the azimuthal resolution of 21.4° which corresponds to 0.83cm given the electrode radius of 2.2275cm [153].

During this process many variables relevant to the topological fingerprint of an event (such as N_{tracks}, D_0, D_{min}) are saved and stored for machine learning purposes. Detailed descriptions of these variables and their uses can be found in Sec. 8.2.

4.3 Ionisation Chambers

Ionisation chambers are some of the oldest and most widely used variety of particle detectors [137, 154–156], and utilise the effects resulting from a charged particle passing through a gas for detection. As a charged particle passes through the gas within the chamber, the gas molecules along the track either become ionised or excited. Most ionisation chambers are based on detecting this ionisation created as a result of these charged particles.

Ion chambers are the simplest form of gas-filled detectors and are based simply on the collection of these charges through the application of an electric field; similar to the process used in SDDs to collect electron-hole pairs (Sec. 4.1.1).

4.3.1 Ionisation of Gasses

When a charged particle passes through a gas it leaves behind a trail of ionised particles in its wake. Upon ionisation, neutral molecules will convert to a positive ion, and a free negative electron. These are known as *ion pairs* and are synonymous with the electron-hole pair created when a charged particle passes through a solid. The total number of ion pairs created along the trajectory is of importance to the principle of operation and for ionisation to occur the particle must transfer at least the ionisation energy of the gas molecule used. For most gasses, this is $\sim 10 - 20$eV. However, as mentioned, excitation is also possible along the path, and though these can be detected in some configurations the gas detector deployed by ALPHA cannot do this. As such, the average energy lost by the travelling charged particle per ion pair is known as the *W-value*, and is always much higher than the actual ionisation energy usually at around $30 - 35$eV. This means that a particle with, for example, a kinetic energy of 139MeV could create anywhere from $4 - 4.6 \times 10^6$ ion pairs before being stopped completely (assuming the particle doesn't leave the gas volume before stopping).

4.3.2 Ion Pair Mobility

If an external electric field is applied over the region of ion pair creation the electrostatic forces will move the charges away from their origin. This motion is described by a combination of the random thermal velocity (the gas will be in a state of constant thermal motion which will be imparted on the ion pairs) and the net *drift velocity*.

The drift velocity for the positive ions is in the direction of the field, while the negatives go against it.

For the ions in a gas, the drift velocity (v_{ion}) can be well described as

$$v_{\text{ion}} = \mu_{\text{ion}} \frac{E_0}{p} \tag{4.9}$$

where E_0 is the magnitude of the electric field, p is the pressure of the gas, and μ_{ion} is the mobility and is proportional to the average time between collisions [157]. The mobility generally remains fairly constant for a wide range of pressures and electric fields and further, it does not really differ between positive or negative ions. Typically, this value ranges from $1 - 1.5 \times 10^{-4}\text{m}^2 \cdot \text{atm/V} \cdot \text{s}$ [158]. Given a gas at a pressure of 1atm and an electric field of 1000V, the resulting drift velocity will be 1m/s.

The electrons however have a much lower mass than the ions, which allows for a greater acceleration between encounters with neutral gas molecules, and the value of the mobility is generally $\sim$ 1000 times greater than that of the ions.

The ionisation current generated by the movement of ion pairs under the electric field can then be measured by an electrometer circuit, though the resulting current will be on the order of $10^{-12} - 10^{-15}$A. With correct amplification an analogue signal proportional to the amount of ionisation (and therefore energy deposition) can be recorded.

4.4 The ALPHA-g Detector(s)

The ALPHA-g apparatus utilises two different detector systems. The radial time projection chamber (TPC, or sometimes rTPC) which is a form of ionisation chamber, used for precise vertex reconstruction; and the barrel veto (BV) which is a barrel of scintillating bars read out on both ends by SiPMs used for both triggering the detector readout, and for background suppression. The detectors are arranged around the trap in concentric circles, with the TPC lying immediately outside the neutral trap and cryostat, and the BV surrounding that.

Due to the large size of the ALPHA-g sensitive region required for detection ($\sim$ 3m - see Chapter 3 and specifically Figure 3.13) a silicon strip detector is not a realistic solution, despite their strong overall performance. SSDs are expensive to produce, especially with the amount of panels that would be required for ALPHA-g. Further, whole tracks can be lost if a certain panel loses efficiency, and given the amount of material required between the trapping region and the detector (mainly magnet coils) there is a chance that a lot of scattering occurs, again losing tracks.

The cost and time required to manufacture a silicon strip detector is not favourable. A gas chamber detector becomes more suitable, as inexpensive gas can replace the solid-state material required in an SDD, and a gas detector allows for many more samples of the track than just the three layers of a silicon strip detector, making the loss of tracks in ALPHA-g less likely. However, the cost is a decrease in accuracy, and a slower readout time than an SSD.

4.4.1 The Time Projection Chamber

The ALPHA TPC uses a radial electric field, meaning the charged particles drift outward towards the wall of the detector.

The electric field $\vec{\mathbf{E}}(r)$ in use is given by

$$\vec{\mathbf{E}}(r) = \frac{V}{\ln \frac{b}{a}} \frac{\hat{\mathbf{r}}}{r} \tag{4.10}$$

where a and b are the inner and outer radii of the TPC, and V is the bias voltage (typically 10kV) [159]. In this coordinate frame z is the vertical axis pointing upwards, and the direction in which the constant magnetic field $\vec{\mathbf{B}}$ is directed. It is more common for TPCs to have the electric field parallel to the magnetic field, i.e: $\vec{\mathbf{E}} = f(z)\hat{\mathbf{z}}$; however, the ALPHA TPC is 3m long in this direction and this would result in long drift times, through a magnetic field which has the potential to perturb the tracks and seriously degrade the accuracy of the resulting trajectories and vertices. Further, there is a strong need to be compact in the design of the ALPHA TPC as there are space constraints that must be considered, especially at the top and bottom of the apparatus, making a radial electric field the favourable option.

The result of the electric field described in Equation 4.10 is that ion pairs created by the charged particles drift outwards radially towards the wall and need to be collected there. The outer wall of the TPC is made up of many cathodes known as *pads*, with resolution axially (z), and azimuthally (ϕ). Each pad has a 4×4mm active area and can record the deposited charge of resulting electrons. There are 576 pads segmented axially, and 32 azimuthally giving a total of 18,432 pads for readout. The pads are read out by special 72-channel ASICs in groups of 4 making for a total of 4,608 ASICs. As mentioned, without amplification, the charges deposited on these pads would be too small to read out; as such, just prior to the pads sits an array of anode wires, biased to 3.5kV, which creates an electric field that causes an electron avalanche and thereby achieving charge multiplication to be read out. Thus, the origin of each ionisation in space can be determined from which pad is registering the signal. However, as all ionisation products are collected on the outer wall where the pads reside, r information is lost. In order to reconstruct the r positions of the hits (which is required to construct the space-points in 3D space needed for reconstructing tracks) we infer the r position from the ionisation drift time t_d. That is, the r position is determined from the *time-projection* equation:

$$r = |v_{\text{ion}}| \cdot t_d \tag{4.11}$$

where v_{ion} is the drift velocity from Equation 4.9 and t_d is the *drift time*. Measured on a "per-event" basis the drift time is the time between the readout trigger[9] and the hits registered on the pad, and can be recorded for each individual hit.

Given the external magnetic field required for the PM trap to function, the electrons also experience a similar Lorentz force to those of the atoms inside the trap (Equation 2.13). Given $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ are orthogonal then the motion is curved according to the Lorenz force with an angle, α, known as the *Lorentz angle* which

---

[9] The ALPHA-g detector is triggered on two or more clusters of signal on the anode wires.

separates the drift velocity and the electric field. The magnitude of α is given by

$$\alpha = \tan^{-1}\frac{e|\vec{\mathbf{B}}|\tau}{m} \tag{4.12}$$

where e, m are the electron mass and charge respectively, and τ is the average time between collisions of an electron with gas molecules while drifting towards the wall [157]. During the reconstruction this angle must also be corrected for in the ϕ position of the space-points. If the Lorentz angle is expected to be large then one would like to choose a gas with a high mobility (μ_{ion}) for an increased drift velocity [159, 160], this can be done by adding small portions of polyatomic gasses to the mixture used in the chamber. Further, it is important to choose a gas that has low electron capture probability, that is, one in which the likelihood of the electron being re-absorbed prior to reaching the pads is small. For this reason noble gasses, and very commonly argon, are chosen to be the main gas in TPCs. The ALPHA TPC uses a mixture of 70% argon with 30% CO_2 for a good balance of drift velocity and low electron capture probability [17].

4.4.2 Barrel Veto

Surrounding the TPC is the Barrel Veto (BV) detector which is a scintillating bars detector read out on both ends by an array of SiPMs. The BV is used in combination with the TPC both for the triggering of the detector, and for the rejection of cosmic background rays. Though it is not a true veto (it will always trigger the detector when it detects charge, not attempting to veto specific events), it is used for cosmic background rejection during offline analysis.

The BV comprises a series of 64 plastic scintillating bars arranged in a cylinder around the TPC. Each bar is slightly trapezoidal ($\sim$ 0.7cm in the front and $\sim$ 0.76cm at the back) to allow for uniformity and structural integrity. Each bar is 260.4cm in length and contains one SiPM on either end of the bars for readout. The inner diameter of the BV is 44.6cm, and it is $\sim$ 2cm thick. The BV was designed to have both a long attenuation length, which is required for the 2.6m distance the light will potentially be required to traverse before readout, and a fast readout, required for triggering and vetoing.

The main purpose of the BV is to calculate the *time-of-flight* (TOF) of particle trajectories throughout the detector. To understand why this is a valuable variable: consider the situation where a cosmic ray passes directly through the TPC, against an $\bar{\text{H}}$ annihilating on the wall. Assume the cosmic ray passes directly through the detector, and enters the BV triggering a bar at some time t_1, then leaves the detector triggering another bar at time t_2, then one must have $t_1 < t_2$ as the time taken for the particle to pass the detector is (assuming speed c and a roughly 1m path[10]) is $\sim 3.3ns$ which can be distinguished from one another with fast enough read out electronics. However, in the case where an $\bar{\text{H}}$ atom has annihilated on the wall and released n secondaries, which all travel outwardly to the wall passing through both the TPC and BV. Then it can be expected that $t_{n_1} \sim t_{n_2}\ \forall\ n_1, n_2 \in n$. A diagram (not to scale) of this concept can be seen in Figure 4.11.

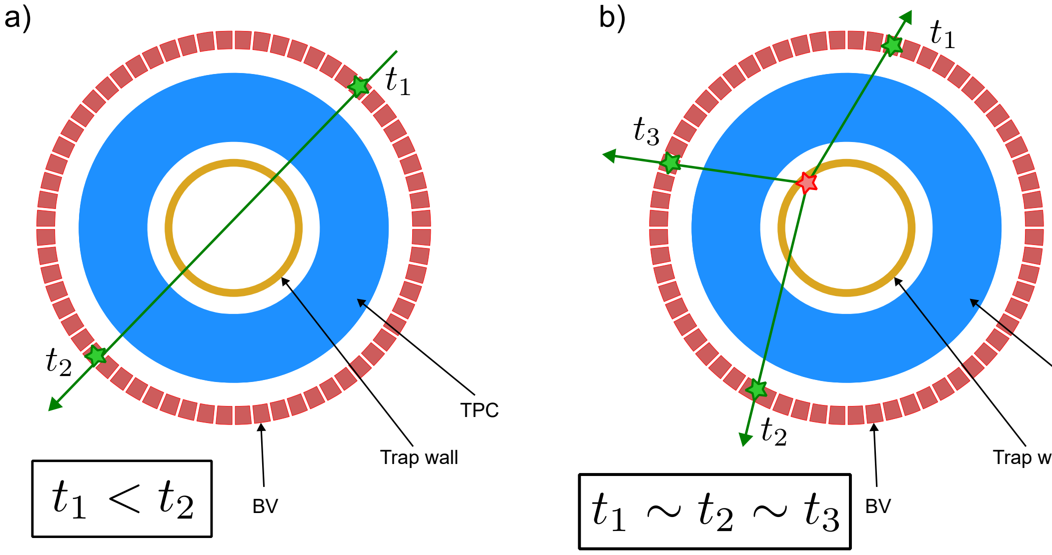


Figure 4.11: A diagram showing how TOF might vary between a) a cosmic ray; and b) an annihilation on the trap wall (denoted by the red star). Cosmic rays usually only leave two hits with distinguishable time delay between the hits, whereas antihydrogen events often produce multiple hits, with minimal delay between the hits recorded times. (Not to scale).

Sometimes a cosmic ray may scatter on material in the detector, causing a shower of two or more hits, each with similar TOFs, however the initial entry time of the ray (and therefore the first hit) should still be much sooner than the shower

---

[10] This is dependent on the angle of entry but for demonstration purposes assume 1m trajectory through the TPC and BV.

recorded later, allowing discrimination of these events by comparing the largest and smallest TOF.

Further, using the TPC itself to trigger detector readout is not ideal, drift times of electrons in the gas will vary between events meaning the trigger is not consistent between events. The propagation of light through scintillators, and the readout by SiPMs is much faster, making the BV an ideal candidate for triggering the detector readout, providing a consistent and fast trigger.

The BV signal is read out by 8 analogue-to-digital (ADC) converters, and a single time-to-digital (TDC) board and was designed to measure the TOF of particles within a precision of 200ps.

4.4.3 Event Reconstruction

The ALPHA-g event reconstruction works similarly to ALPHA-2 with some key differences to account for differences in the detector hardware and readout.

The first stage is to reconstruct the space-points from the pad hits. For this we require knowledge of which pad was hit, the time of the hit (to reverse the drift time), and the geometry of the pads in space, as well as information on the drift time for the chosen detector configuration. Consider a hit recorded on a specific pad at a given time denoted (c_z, c_ϕ, c_t) where c_z and c_ϕ represent hits in a given column/row of the pads (i.e. $c_z \in \mathbb{N}[1, 577]$, and $c_\phi \in \mathbb{N}[1, 33]$[11]), and c_t represents the time bin of the hit. Since the pad hits represent the end point of the electron trajectory, and where the charge is collected on the TPC wall, in order to convert the given pad hit to an origin of ionisation we use the following relations:

$$x = r_h \cos\phi \tag{4.13}$$

$$y = r_h \sin\phi \tag{4.14}$$

$$z = c_z \cdot 4\text{mm} \tag{4.15}$$

where $\phi = \frac{c_\phi \cdot 4\text{mm}}{r_{\text{TPC}}} - \alpha$ (r_{TPC} is the outer radius of the TPC, 203mm). r_h and the Lorentz angle (α) are functions of the drift time, and therefore of c_t. In order to obtain these values Monte Carlo simulations were run with the Garfield

---

[11] In this context $\mathbb{N}[a, b] = \{x \in \mathbb{N} \mid a \leq x \leq b\}\ \forall a, b \in \mathbb{N}$

[161] package and the results were stored in a look-up table for access during the reconstruction algorithm. Once this stage of the reconstruction is complete we are left with a set of space-points (x, y, z) for which to reconstruct the tracks (a track being defined as a collection of space-points originating from the same particle's path through the detector). In order to group the space-points into tracks, an algorithm is used that searches for nearby space-points within a sphere of given radius starting with the high-radius space-points and working inwards. Monte Carlo simulations helped settle on a radius of 11.5mm for this sphere in the ALPHA-g reconstruction.

Once the space-points have been successfully grouped into tracks, the next step is to fit a line to their trajectory. It is known that the paths of the secondaries given the magnetic field will be roughly helical, and therefore parametrised by five variables, two of which are the *axial* parameters, and three of which are the *radial* parameters. The fitting is done via the method of least-squares, and the two groups of parameters are minimised separately (see Equation A.4 and Equation A.5). To give an initial value for the fitting parameters a straight line is fitted between the first and last space-point of a track, and its parameters are fed into the minimisation algorithm.

Not all helices are a result of the initial $\bar{\mathrm{H}}$ annihilation, and a round of pruning (see Sec. 4.2.1) must take place prior to vertex fitting. This is done using the same DCA, D, as defined in Equation 4.8, but also from the axial and radial χ^2 resulting from the fitting process (Equation A.4 and Equation A.5).

The cuts deployed on these variables in the ALPHA-g reconstruction are shown in Table 4.3.

Variable	Cut
$\lvert D \rvert$	< 40mm
χ^2_r (Equation A.4)	< 15
χ^2_z (Equation A.5)	< 8

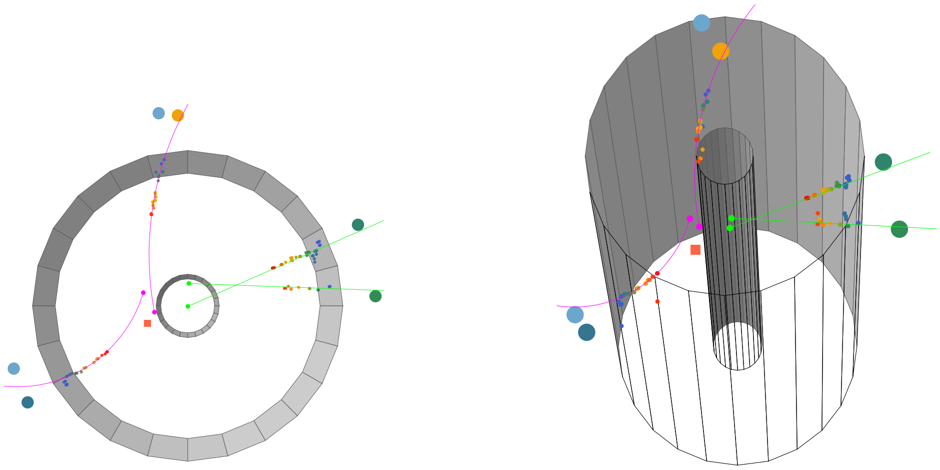
Table 4.3: Cuts used to reject or accept a helix in the ALPHA-g reconstruction algorithm.

All helices are denoted *candidates*, and only those passing the chosen cuts are designated as *used* or *good* helices.

With all this in place, the final step is to reconstruct the vertex position, which is done in a three-step process, as outlined below:

1. Seed: Loop through each pair of helices and find the pair that give the minimum DCA. The midpoint of the two points along the tracks resulting in this minimum DCA is taken as the *seed vertex*.
2. Recalculate: Extend both helices used in the previous step and minimise the distance between their trajectories, using the seed vertex as an initial value for the minimisation.
3. Improve: If there are more than two good helices include another helix and perform another round of minimisation. The resulting χ^2 of this minimisation must be less than some cut-off value (in our case $\chi^2 < 3$) to consider the tracks inclusion an improvement.

An example of a reconstructed event in the TPC can be seen in Figure 4.12.



(a) Axial view. (b) Slightly tilted view.

Figure 4.12: Two images of the ALPHA-g reconstruction. The small points are space-points, and their colour gradient reflects their time of detection (blue hits come first and are hence closer to the wall). The lines are reconstructed tracks, the two purple tracks are the initial seed helices, and the green tracks were reconstructed but not used in the "improve" step. The furthest out points are hits on the BV, again with colour reflecting hit time (not on the same scale as the space-points).

This reconstruction algorithm was found via Monte Carlo to have an efficiency of 94.1 ± 0.2%; and to have radial resolution of 6 ± 0.2mm, axial resolution of

3.9 ± 0.1mm, and an azimuthal resolution of $16.3 \pm 0.4°$ which corresponds to 6.337 ± 0.156mm given the electrode radius of 2.2275cm [159].

4.5 Conclusion

In conclusion, we have seen how detecting and reconstructing antihydrogen annihilations requires dedicated technology and techniques to understand. By utilising SDDs and ion chamber detectors, the ALPHA experiment has a wide array of detection devices, each with different pros and cons, capable of reconstructing the events annihilation vertex to millimetres of the actual position (according to simulation). ALPHA's detection, readout, and reconstruction chain is essential to all generated results, and without these dedicated methods working optimally, results can become lost to noise. The final step to the reconstruction, the event classification, is not included in this chapter and is instead the focus of Chapter 8 and Chapter 9.

Chapter

5

Simulations for Calibrating Absolute Atom Counts

Monte Carlo simulations have already proven vital for calibrating detector responses in the ALPHA experiment, particularly for the SVD and the TPC (refer to Sec. 4.2.1 and [153], or Sec. 4.4.3 and [159]). Additionally, to these tracking detectors, ALPHA deploys multiple scintillating panels throughout the experiment (see Sec. 3.2), which detect signals resulting from the manipulation and movement of plasmas, some of which lead to partial or total loss of atoms.

While uncalibrated panel signals are useful for relative measurements, they cannot provide absolute particle counts. A straightforward method to achieve this calibration involves using Monte Carlo simulations to replicate the processes that cause particle losses and comparing the simulated responses with experimental data.

Currently, only the SVD and TPC have undergone this calibration process, with trigger efficiencies (relative to true $\bar{p}$) known to be approximately 99.5% and 93.5% respectively. These calibrations enable us to convert trigger signals to absolute $\bar{p}$ counts, which is crucial for determining particle transfer efficiencies, such as between ALPHA-2 and ALPHA-g. Although direct $\bar{p}$ transfers from ALPHA-2 to ALPHA-g are rare, we can assess the transfer efficiency from the catching trap (CT) to the respective atom traps. Observations indicate that the transfer efficiency into ALPHA-g is roughly 10% of that into ALPHA-2, implying significant $\bar{p}$ losses along the beamline, with annihilation signals detectable by the caesium iodide (Sec. 3.3) and scintillating detectors (Sec. 3.2).

To address these inefficiencies, calibrating various panels throughout the ALPHA experiment using Monte Carlo simulations will enable us to obtain absolute $\bar{p}$ counts across different stages, not just within the atom traps. This calibration could help enhance transfer efficiency and potentially increase $\bar{\mathrm{H}}$ production rates.

Achieving this calibration requires both experimental data and simulations to compare observed results with expected outcomes. This chapter details four unique experimental procedures employed by ALPHA during the antihydrogen production cycle for diagnostic purposes and compares the results with Monte Carlo simulations to ensure accurate reproduction of experimental conditions.

5.1 Simulation Setup

In this section, we describe in detail the software and setup used for these simulations.

5.1.1 ROOT, Virtual Monte Carlo, and Geant4

The foundation of the simulations are built on three main software packages: ROOT [162], Virtual Monte Carlo (VMC) [163], and Geant4 [164, 165]. ROOT provides a versatile framework for data analysis, widely adopted by the physics community. VMC acts as an intermediary; providing integration of the various Monte Carlo engines, including Geant4, into the simulation. Geant4 is the physics engine powering the physics processes, and managing all physics interaction from its internal *physics list*. The geometry model is built in ROOT using the `TGeo` class and is then converted to Geant4 using the Virtual Geometry Model (VGM) package [166]. It is necessary to provide Geant4 with a geometry model, and particle generator (responsible for initialising individual events), and Geant4 handles the step-by-step solving procedure itself, tracking all particles until they either lose all kinetic energy, or exit the simulation volume.

There is an internal wrapper class that manages the initialisation of the simulations, the `a2mcVirtualMC` class. This class takes care of building the geometric model, initializing the *primary generator* and retrieving/storing the event information to an output ROOT file as particles step through the simulation. The ROOT file is broken into many `TTree` [167] objects which store various information about the simulated event, such as the primary particle's generation location and momentum, the annihilation position, and the resulting signal on the various detectors.

The coordinate system of the world geometry is Cartesian with origin at the centre of the SVD. The z-axis corresponds with the axial direction of the fields in the ALPHA-2 atom trap, the y-axis points vertically up, and the x-axis points to the "right" (or starboard side relative to $\bar{p}$) while standing at negative z and looking positively. As the beam enters from the negative z direction, this is known as the *upstream* end, conversely the other side is the *downstream* end and is relative to $\bar{p}$ direction through the apparatus; therefore the CT sits on the upstream end, and the positron accumulator sits at the downstream end of the beam line.

5.1.2 Geometry

The geometry is built within the `a2mcApparatus` class, which creates `TGeoVolume` objects for each individual component between the trap and detectors. Figure 5.1 shows a view of the overall geometry the simulation uses, and Figure 5.2 shows more detailed views of the internals of the ALPHA-2 atom trap, catching trap, and DSAT/IC region.

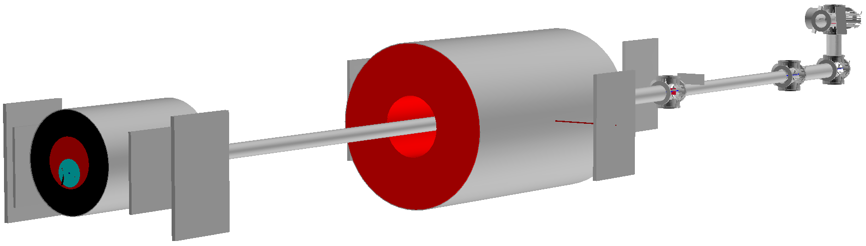


Figure 5.1: Image of the overall geometry of the simulations, from left (upstream) to right (downstream) we have the catching trap, surrounded by four scintillating panels, the atom trap with two panels, the "downstream atom trap (DSAT)" stick, then finally the interconnect surrounded by three SiPMs.

Further, VMC allows for the user to import a magnetic field map, and the map used contains the value of $\vec{\mathbf{B}}$ at each point of (x, y, z) from $(-50, -50, -444)$cm to $(50, 50, 444)$cm with the value of (B_x, B_y, B_z) given every $2 \times 2 \times 2\text{cm}^3$.

The most dense material between the trapping region and the scintillating panels are usually the magnets that supply the constant $\vec{\mathbf{B}}$-field for the PM traps, the Swansea magnet (CT), and the Carlsberg magnet (ALPHA-2) both of which consist of NbCuTi windings surrounded by steel shells. The detailed geometry of these magnets, and the geometry used for the simulations can be found in Appendix E.

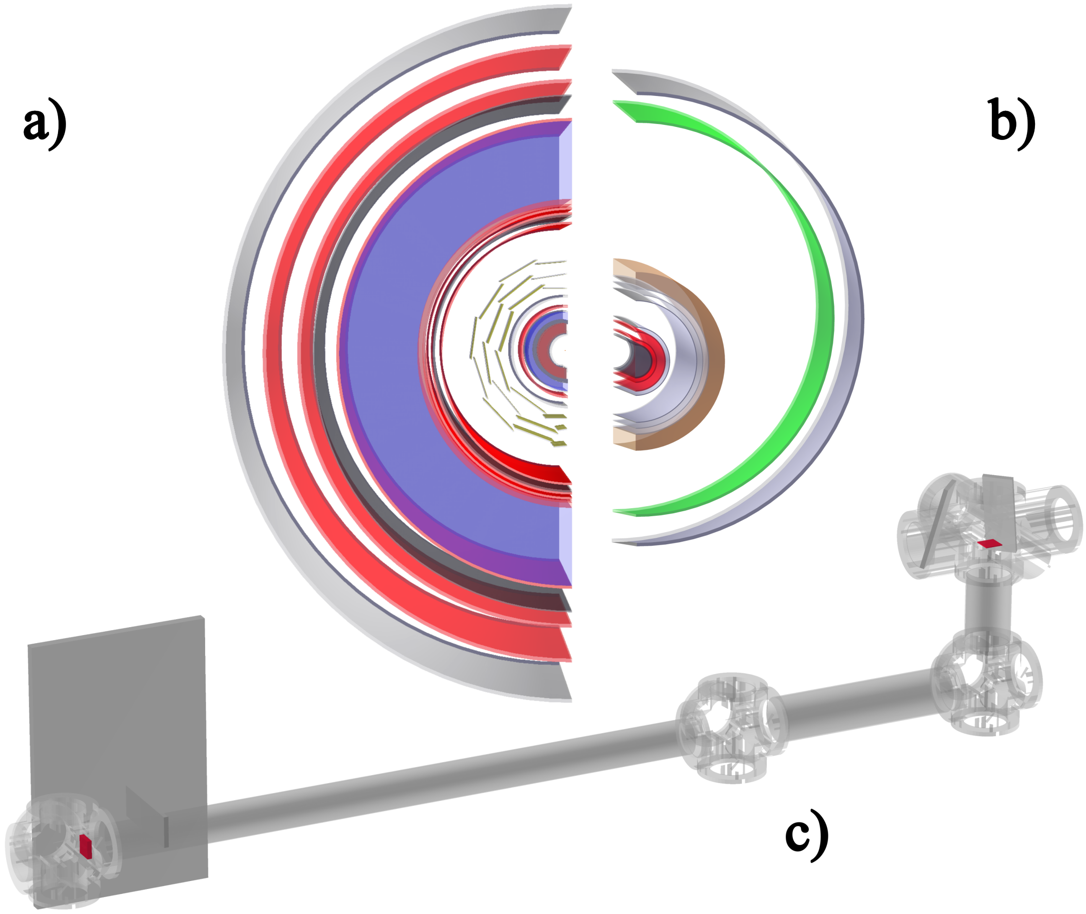


Figure 5.2: View of the simulated geometry. a) Cutaway view of the ALPHA-2 atom trap region in the simulations. b) Cutaway view of the CT region in the simulations. c) Side view of the DSAT and IC (MCPs are pictured in red). Individual images are not to scale with one another and are only placed in this configuration for display purposes.

5.1.3 Particle Generation

While the geometry of the simulations remains consistent across different setups, the location and momentum of particle generation depend on the specific experiment being simulated. Four rounds of simulations were conducted, each aimed at calibrating different scintillating counters during a distinct process. In each configuration the particles are given an initial position ($\mathbf{x}_{\text{origin}}$) and the unit vector of the momentum ($\hat{\mathbf{p}}_{\text{origin}}$), with some Gaussian spread to mimic the real spread of plasmas in the ALPHA experiment.

The four configurations currently implemented are:

1. *Slow release* of antiprotons in the atom trap. It has previously been shown

that the SVD trigger channel is $\sim 99.5\%$ efficient to true $\bar{p}$ numbers. This is one of the few places in the experiment where we have such a well-defined calibration, and surrounding the ALPHA-2 atom trap are two $60 \times 40\text{cm}^2$ scintillating panels (Sec. 3.2) read out by PMTs. By comparing the numbers recorded on each of these panels with that of the SVD trigger channel we are able to obtain a good estimate for the detection efficiency of these panels - without the need for simulation. However, simulating the same process and comparing this experimentally calculated efficiency with the simulated efficiency will provide a good benchmark with which to test the accuracy of the Monte Carlo. In this experimental configuration the $\bar{p}$ are held in the ALPHA-2 PM trap, confined to one electrode (E13) and trapped axially between the two surrounding electrodes with a 0.2V potential. By slowly raising the confining potential the $\bar{p}$ are gradually released from the trap and are free to annihilate on the wall. The gradual release of the $\bar{p}$ takes 200s to ensure that no detectors become saturated. This is known as a *slow release* of $\bar{p}$. In this configuration we set $\mathbf{x}_{\text{origin}}$ to $(0, 0, -1)$ corresponding to the location in the trap where the $\bar{p}$ are held. In principle the $\bar{p}$ are free to annihilate anywhere on the trap wall azimuthally, however when these experiments were performed a small *patch potential* [168] on the surface of the electrodes had formed, meaning the $\bar{p}$ were drawn to a location that corresponds roughly to $\hat{\mathbf{p}}_{\text{origin}} = (2/\sqrt{5}, -1/\sqrt{5}, 0)$.

2. *Cold dump* from the catching trap to the degrader foil. As mentioned in Sec. 3.4 it is possible to obtain longitudinal temperature measurements by dumping the particles near a detector. When performed with particles cooled in the catching trap this is known as a CT *cold dump* and is an important diagnostic tool in the ALPHA experimental cycle. Surrounding the CT are four scintillating panels, two of which have an active area of $60 \times 40\text{cm}^2$ and are read out by PMTs, and another two 40×40cm read out by SiPMs sitting slightly closer to the trap. Though the temperature diagnostic is independent of the absolute calibration of the detector, knowing the calibration of the panels allows for an additional diagnostic during the cold dump. By lowering the potentials asymmetrically the atoms are directed upstream and

annihilate on the degrader foil (see Sec. 3.7). In this configuration the $\bar{p}$ are generated in the centre of the catching trap ($\mathbf{x}_{\text{origin}} = (0, 0, -317)$cm), and are given negative z momentum only ($\hat{\mathbf{p}}_{\text{origin}} = (0, 0, -1)$) directing them upstream towards the foil.

3. *DSAT dump* from ALPHA-2 to the "DSAT" stick MCP. One of the sticks (see Sec. 3.1) further downstream of the atom trap, known as the "downstream atom trap stick" (DSAT) has an MCP (Sec. 3.1.2) and is surrounded by a small $20 \times 7\text{cm}^2$ scintillating panel read out by a SiPM. The true efficiency of this panel during an MCP dump could provide information on $\bar{p}$ numbers leaving the ALPHA-2 atom trap, or leaving the CT when transferred directly to ALPHA-g. For this round we set $\mathbf{x}_{\text{origin}} = (0, 0, 20)$cm which is directly upstream of the MCP; and the $\bar{p}$ are directed onto the MCP with z momentum only: $\hat{\mathbf{p}}_{\text{origin}} = (0, 0, 1)$. (See Figure 5.3)

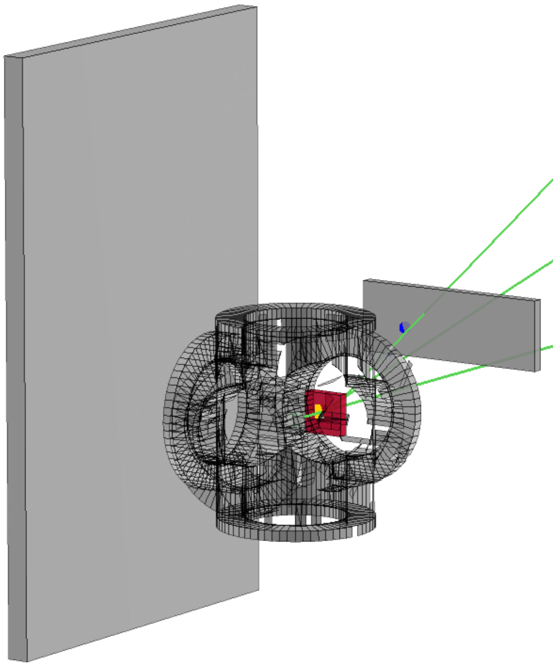


Figure 5.3: Example event at the DSAT. The red square is the MCP, and the yellow circle the annihilation location. The resulting tracks are in green, and the blue circle denotes where the scintillator registers a hit.

4. *IC dump* from ALPHA-2 to the interconnect MCP. Further along the transfer line into ALPHA-g is the *interconnect* (see Sec. 3.11) where $\bar{p}$ (and e^+) must turn a 90° corner to enter ALPHA-g. This transfer carries a high risk for

losses, and understanding how many particles survive this transfer is important and especially useful when compared with the DSAT dump, providing a potential answer to where exactly the particles are being lost, potentially improving transfers. For this configuration $\mathbf{x}_{\text{origin}} = (0, 37.7, -420)$cm which corresponds to directly below the MCP; and are given vertical momentum only, directing them upwards onto the MCP: $\hat{\mathbf{p}}_{\text{origin}} = (0, 1, 0)$.

A summary of these four simulation conditions can be seen in Table 5.1.

Experiment	Particle	# of particles	$\mathbf{x}_{\text{origin}}$ (cm)	$\hat{\mathbf{p}}_{\text{origin}}$
Slow Release	$\bar{p}$	100,000	$(0, 0, -1)$	$(2/\sqrt{5}, -1/\sqrt{5}, 0)$
Cold Dump	$\bar{p}$	100,000	$(0, 0, -317)$	$(0, 0, -1)$
DSAT Dump	$\bar{p}$	20,000	$(0, 0, 20)$	$(0, 0, -1)$
IC Dump	$\bar{p}$	50,000	$(0, 37.7, -420)$	$(0, 1, 0)$

Table 5.1: Summary of simulation conditions used to match each experimental dump.

In each case, the primary generator creates antiprotons at the chosen location with the momentum direction given by the unit vector specified and all resulting collisions are handled by Geant4's internal physics list. The overall multiplicity of the events should follow the values seen in Table 2.1 and Table 2.2 which outline the expected branching ratios and multiplicity of secondary particles resulting from antiproton annihilation on various target nuclei.

In all four experimental setups, the most likely target for annihilation is some heavy nucleus ($\mathcal{N}$), given that the foil target, ALPHA-2 electrodes, and MCP are made from gold-plated aluminium, aluminium, and glass respectively. When an antiproton annihilates with a nucleus, it results in the production of several secondary particles. The detailed physics processes governing these interactions are modelled within Geant4, which includes hadronic interactions, electromagnetic processes, and particle decays.

5.1.4 Simulation Output

Each particle interaction with either the SVD or scintillating panels placed throughout the experiment results in the creation of an `a2mcSilDIGI` or `a2mcPanelDIGI` object respectively. These objects handle the *digitization* of the hits, ensuring double hits from the same event do not count twice and correctly calculating the total energy deposited into the strips and panels respectively. Each simulated $\bar{p}$ is considered a unique event and for each event many variables are recorded and saved to a ROOT file for later analysis, these include:

- **Generation position:** The initial position where the antiproton is generated.
- **Generation momentum:** The initial momentum of the generated antiproton.
- **Annihilation position:** The position where the antiproton annihilates.
- **Annihilation products:** The particles resulting from the annihilation.
- **Multiplicity:** The number of secondary particles produced.
- **Secondary momenta:** The momenta of the secondary particles produced from the annihilation.
- **Energy deposition:** The total energy deposited in the detectors.
- **Hit information:** Detailed information about each interaction in the detectors, including time, location, and energy.

The output of the simulation is the collection of these events, stored in a structured format within the ROOT file for analysis.

5.2 Results

The results of the experiments outlined in Sec. 5.1.3 are presented here. The primary focus is on the response of the detectors during each simulation setup, with an aim to calibrate the signal on the panels to the number of simulated $\bar{p}$. By looking at the multiplicity of each event separated into particles (Figure 5.4) we are able to see the branching ratios of the charged pions match well with those presented in Table 2.1 and Table 2.2.

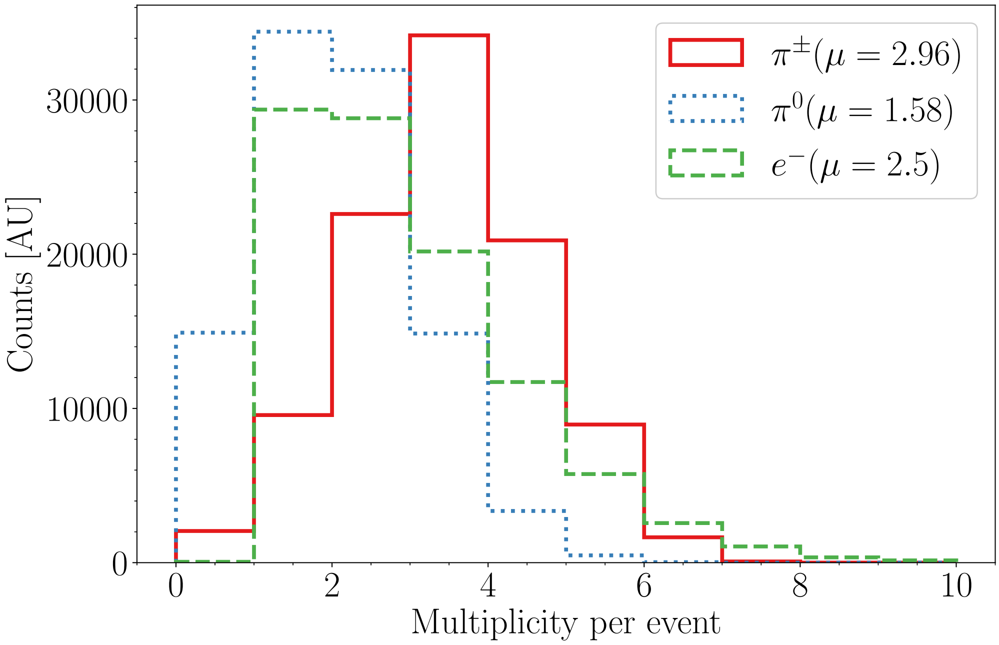


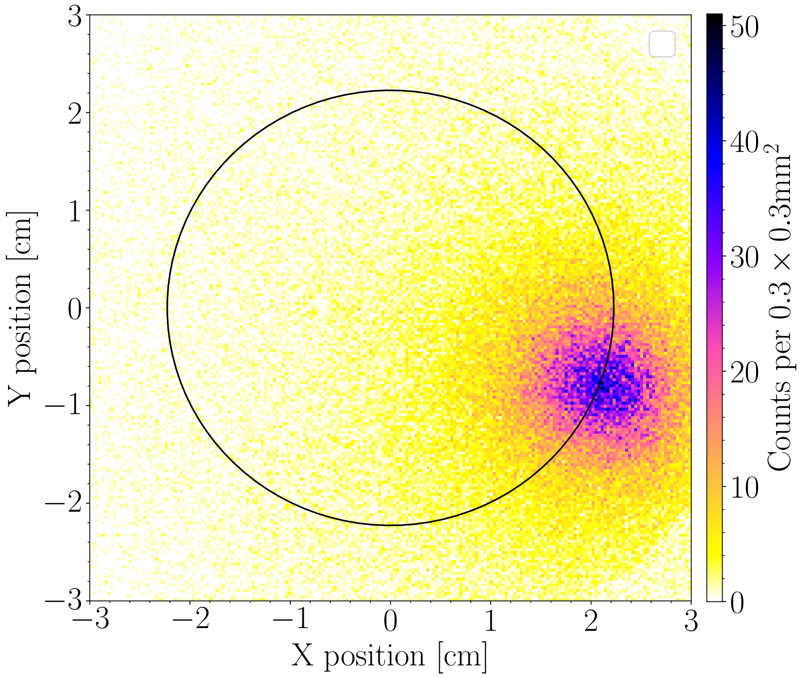
Figure 5.4: Multiplicity of each event in the simulated results, separated into π^0, $\pi^{\pm}$, and e^-. The mean number per event is denoted as μ in brackets in the legend.

5.2.1 Slow Release Experiment

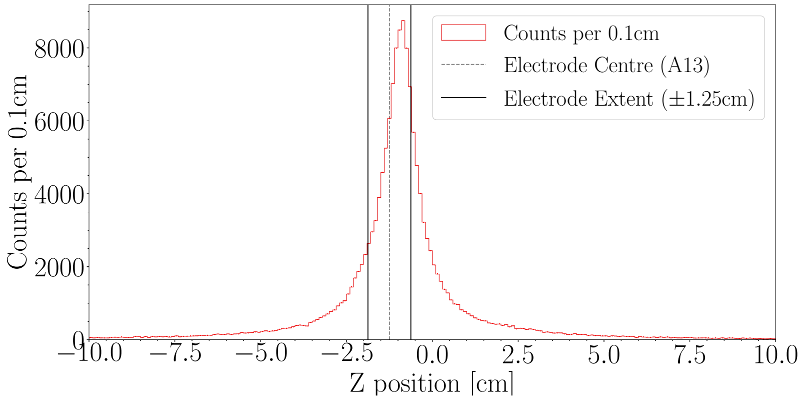
The slow release experiment takes place in the ALPHA-2 atom trap, which is surrounded by the SVD. We are therefore able to reconstruct the vertex position of the original annihilation within the trap, something not possible with the other experiments. The slow release experiment was repeated four times during the 2022 run, each time slowly raising the potential of the electrode holding the $\bar{p}$ until the potential was flat and all particles were released. The reconstructed x, y, z position of the resulting vertices over all four trials can be seen in Figure 5.5.

The peak in z is expected, and the width of the resulting histogram corresponds to the width of the electrode in which the $\bar{p}$ are held during this experiment.

The distinctive x, y "hotspot" arises due to patch potentials on the trap walls (resulting from imperfections in the metal, and amplified by prolonged operation of the apparatus, on the order of days/weeks). These patches cause localized areas of increased charge on the metal surface. Consequently, when $\bar{p}$ are free to annihilate, they tend to concentrate in the positively charged regions [168].



(a) $x - y$ vertex of all slow release experiments performed.



(b) z vertex of all slow release experiments performed.

Figure 5.5: x, y, z vertex of all slow release experiments performed. The plots contain four trials, each of 200s, resulting in 158,694 vertices. Figure 5.5a shows the x, y vertex, binned into $0.3 \times 0.3\text{mm}^2$ bins, and the black circle shows the location of the electrode at $r = 2.2275$cm. Figure 5.5b shows the z vertex, binned into 10mm bins, and the black lines shows the width and location of the electrode where the $\bar{p}$ were stored. In these plots the resolution of the detector and reconstruction algorithm are evident, as the true vertex for the majority of these events is likely directly on the wall.

Over the whole experiment we record $493,800 \pm 700$[1] SVD triggers, which, when assuming an efficiency of 99.5%, and given the 9.603(50)Hz background rate converts to $488,500 \pm 700\bar{p}$ annihilating in the trap. The two panels surrounding the SVD are logically ORed to obtain a single channel known as the "ATOM_OR" channel, which in the same period records $25,350 \pm 160$ counts above threshold. Given the background rate of this channel (12.609(57)Hz) there are $15,260 \pm 160$ counts above background, expected to come from $\bar{p}$ annihilations. This results in a detection efficiency of 3.125(33)% to true $\bar{p}$ annihilations in the trap.

To validate these findings, simulations were conducted with 100,000 simulated $\bar{p}$ particles following the same experimental conditions. The resulting x, y, x positions of the decay events are depicted in Figure 5.6

Once a simulated particle is recorded as passing through the sensitive volumes (detectors) surrounding the atom trap, the energy deposited is stored, and any further hits are digitised to ensure no double counting occurs. 99,650 particles were detected in the SVD, resulting in a trigger efficiency of 99.65(44)%, confirming the previous value obtained via other simulations.

Further, once accounting for the efficiency of both individual SiPMs in the "ATOM_OR" channel[2], and logically ORing on an event-by-event basis, a total 3189 events are recorded on the ATOM_OR during the simulation. This results in a detection efficiency of 3.189(57)% which agrees with the experimentally obtained value of 3.125(33)%.

These simulations confirm experimental outcomes, validating the setup's accuracy for applications where direct experimental verification via the SVD is not possible.

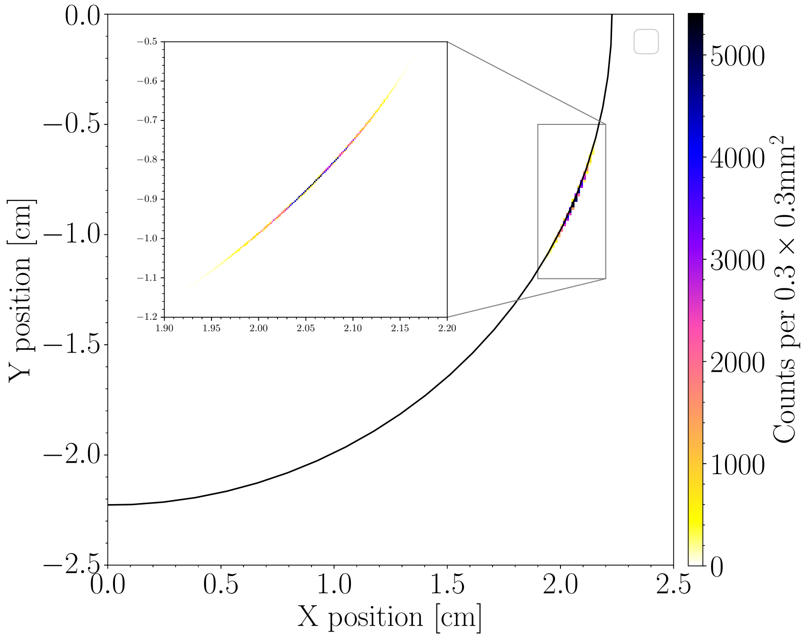
5.2.2 CT Cold Dump

The CT cold dump takes place upstream of the CT on the degrader foil. In the absence of a tracking detector, the only source of detection here is the scintillating panels themselves. This region includes two $60 \times 40\text{cm}^2$ panels read out by PMTs

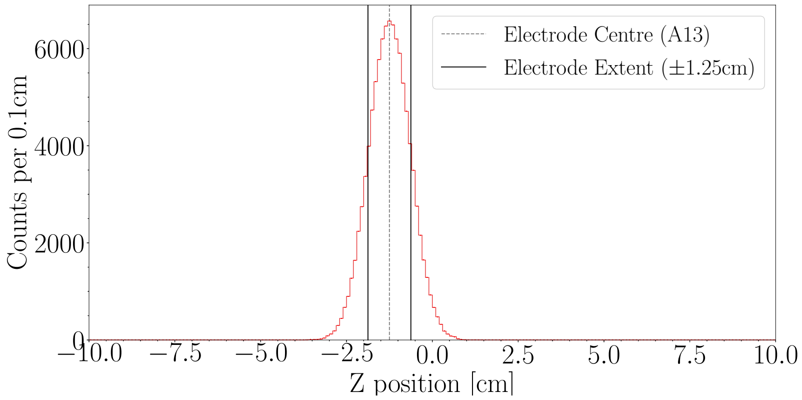
---

[1] All errors in this section come from counting statistic errors propagated with the appropriate formulas where needed.

[2] Here we are referring to the efficiency of a given panel to detect a charged particle that passes through it. The values of these efficiencies for each panel was measured using the technique outlined in Appendix D, and the results of these measurements are displayed in Table D.1



(a) $x - y$ vertex of simulated slow release.



(b) z vertex of simulated slow release.

Figure 5.6: x, y, z vertex of simulated slow release experiment. The plots contain 100,000 $\bar{p}$ simulated using the configuration described in Table 5.1. Figure 5.6a shows the x, y vertex, binned into $0.3 \times 0.3\text{mm}^2$ bins, and the black circle shows the location of the electrode at $r = 2.2275$cm. Figure 5.6b shows the z vertex, binned into 10mm bins, and the black lines shows the centre of the electrode, where the $\bar{p}$ are initiated.

which are logically ORed to create the "CT_PMTOR" channel, and two 40 × 40cm^2 panels read out by SiPMs which are both logically ORed and logically ANDed to create the "CT_SiOR" and "CT_SiAND" channels respectively.

Background rates for these channels were measured and can be seen in Table 5.2, and all final counts throughout this section include the panel's detection efficiency (see Appendix D and Table D.1).

Panel	Background Rate (Hz)
CT_PMTOR	8.772 ± 0.048
CT_SiOR	27.607 ± 0.085
CT_SiAND	0.720 ± 0.013

Table 5.2: Measured background rates for channels in the CT region.

The cold dump was repeated 11 times, lasting approximately 4s each. The resulting background subtracted counts recorded over all 11 trials can be found in Table 5.3, along with background rates themselves, the simulated counts recorded on each channel after generating 100,000 $\bar{p}$ directed onto the foil (as described in Sec. 5.1.3), and the resulting simulated efficiency of these channels.

Channel	Background Rate (Hz)	Counts		Simulated Efficiency (%)
		Background Subtracted	Simulated	
CT_PMTOR	8.772(48)	$530\,432 \pm 729$	3916 ± 63	3.92(6)
CT_SiOR	27.607(85)	$3\,783\,903 \pm 1946$	$34\,670 \pm 186$	34.67(19)
CT_SiAND	0.720(13)	$740\,128 \pm 860$	4336 ± 66	4.34(7)

Table 5.3: Background rates of each channel around the CT region, along with background subtracted counts recorded over 11 repeats of the cold dump experiment, the simulated counts recorded on each detector after simulating 100,000 $\bar{p}$toward the degrader foil, and the resulting simulated detection efficiency.

These numbers suggest that if ∼5k particles are detected on the CT_SiOR channel during a CT cold dump, the true number of $\bar{p}$ in this dump is ∼14k (5k/34.76%) particles.

While there is no detector here to calibrate these numbers as in the atom trap, one potential approach is look at the counts of each channel in this region relative to one another, forming a "matrix" of relative counts. By comparing the values of this matrix between real and simulated data it may be possible to gain some insight into whether the simulation is under or over counting, and potentially offer a method of self-calibration. This matrix for the real CT cold dump data and for this round of simulations can be seen in Table 5.4 labelled “real data”, and “simulated data” respectively. The values of these tables are of the same order of magnitude, but differ somewhat. It can be seen for example, that the simulated counts on both ORed channels is too high in relation to the real cold dump data, or alternatively, the AND channel is too low. It is possible that other effects such as noise in the DAQ system, or cosmic background (neither of which are currently included in simulations) cause erroneous readouts on the AND channel though this remains untested and further simulations would be required to confirm this.

	Relative to	Counts on		
		CT_PMTOR	CT_SiOR	CT_SiAND
Real Data	CT_PMTOR	100.00	713.36 ± 1.05	139.53 ± 0.25
	CT_SiOR	14.02 ± 0.02	100.00	19.56 ± 0.02
	CT_SiAND	71.67 ± 0.13	511.25 ± 0.65	100.00
Simulated Data	CT_PMTOR	100.00	885.34±14.93	110.73 ± 2.44
	CT_SiOR	11.30 ± 0.19	100.00	12.51 ± 0.20
	CT_SiAND	90.31 ± 1.99	799.58±12.88	100.00

Table 5.4: Relative counts on each channel in the CT region with respect to each other over all 11 trials of the cold dump experiment (real data); for the 100,000 simulated $\bar{p}$ (simulated data). Systematic errors are not included and are presumably much larger than those from statistics (which are included). The values themselves agree to within 20-30% showing our prediction should be roughly similar to reality, but further measurements would need to be done to quantify exactly the systematic uncertainty and further simulations could be run including some of the improvements mentioned in Sec. 5.3.

These results provide a value for the efficiencies of the panels surrounding the CT region, and point to a method of calibration which suggests reasonable agreement. However, more simulations could be run to more precisely estimate these values in future.

5.2.3 DSAT Dump

Above the DSAT stick there is a single scintillating panel, read out by a SiPM known as “ATOM_Stk”. After simulating 20,000 $\bar{p}$ into the MCP we observe 3979 ± 166 counts, which after correcting for the panel efficiency, results in a detection efficiency of 19.90(84)%.

Performing the usual transfer from the CT to ALPHA-g results in 32274 ± 180 counts on the SiPM during an MCP dump (the resulting MCP image can be seen in Figure 5.7), suggesting $\sim 162000 \pm 7000$ $\bar{p}$ are surviving the journey out of the CT, through ALPHA-2, and on to the MCP. Given the mean number of $\bar{p}$ that were re-caught in the atom trap during these experiments ($\sim$ 240k $\bar{p}$), this corresponds to a transfer efficiency of $\sim$ 67.5%.

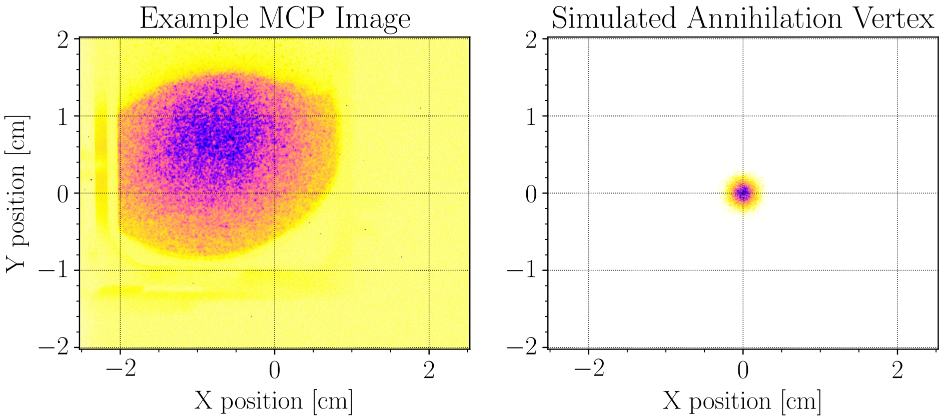


Figure 5.7: A comparison of a single MCP image containing ~ 64000 $\bar{p}$ (using the above calibration factor), with the simulated 20000 $\bar{p}$ annihilation position in the x, y plane. An aperture is visible in the MCP image which blocks the full image from being displayed, and though the size and position between the two is slightly different, differences on this level will not change the overall outcome of the simulation.

In order to verify these results, a sequence was prepared that was tuned to split the $\bar{p}$ plasma precisely in two. The variation in this splitting process was found to be $< 2\%$. One half was then directed to the MCP, and the other half slowly released in the SVD for counting. An example of two of these "half-dumps" can be seen in Figure 5.8.

Three rounds of dumping half to the MCP, and half into the SVD results in an estimated efficiency of 23.3(1)% which matches well with the simulated result.

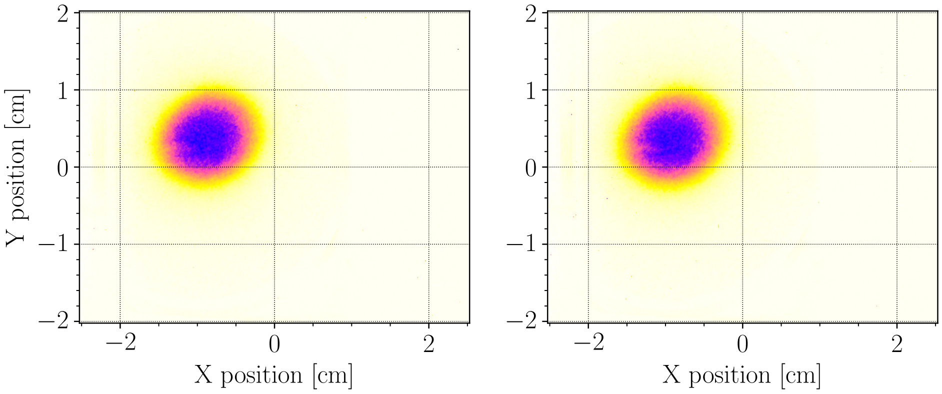


Figure 5.8: Example of two $\bar{p}$ half dumps taken in the ALPHA-2 atom trap. These two images were taken $\sim$ 600ms apart, and the estimated number of particles is $2.142e^7$ and $2.154e^7$ for the left and right images respectively.

5.2.4 Interconnect Dump

The final round of simulation has no way of cross-calibration. As such only the results of the simulations are presented without any comparison to real measurements of any kind. However, the “half-dump” experiment described above could be adapted to provide an experimental value here also.

There is a further MCP on the interconnect between the beamline and the ALPHA-g atom trap (see Sec. 3.11 for details). Once again the plot of the image recorded by the MCP during a transfer is compared to the distribution of the annihilation position of the $\bar{p}$ in Figure 5.9 which shows good general agreement.

For this setup there were three $20 \times 7\text{cm}^2$ scintillating panels read out by SiPMs known as “SiPM_H,I,J” respectively.

The simulated resulting counts and efficiency of each panel is shown in Table 5.5.

Panel	Counts	Efficiency (%)
SiPM_H	7631 ± 309	15.26 ± 0.62
SiPM_I	7793 ± 316	15.59 ± 0.63
SiPM_J	8029 ± 325	16.06 ± 0.65

Table 5.5: Simulated counts and resulting efficiencies on nearby panels after simulating 20,000 $\bar{p}$ annihilating on the IC MCP.

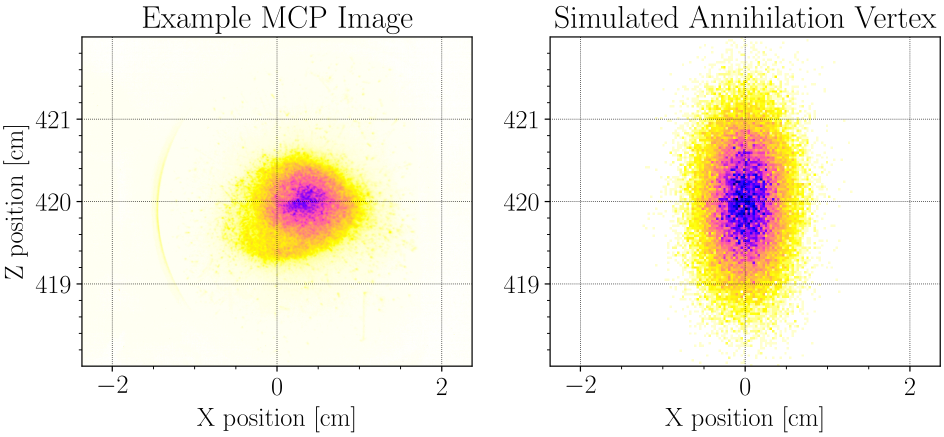


Figure 5.9: A comparison of a single MCP image on the interconnect MCP with the simulated 50000 $\bar{p}$ annihilation position in the x, z plane.

5.3 Conclusion

The description and results of four rounds of simulations, each specifically tailored to match an event within the ALPHA experimental cycle, were presented; and resulting simulated efficiencies for each panel have been estimated. These efficiencies can be used to convert detector signals into true $\bar{p}$ numbers, a useful diagnostic to be added to the many in which ALPHA already deploys on a daily basis. Through experimental results we were able to verify the results obtained in the ALPHA-2 atom trap, and on the downstream atom trap MCP dump. We were also able to provide some calibration on the CT simulations, potentially pointing to a minor disagreement between the two results. The interconnect dump remains uncalibrated and an experimental scheme designed to verify the results obtained here would be beneficial when the possibility arises. A potential scheme to verify these IC efficiencies exists by replicating the $\bar{p}$ half dump techniques used to verify the DSAT results either using the TPC as the verification of the absolute number, or taking into account the transfer efficiency from the ALPHA-2 trap to the region in question. Further, potential measurements exist to verify the CT simulations. Specifically adding the cosmic background flux to the simulations would be an important step to verifying the results, as well as measuring the noise on the channels in this system. Optimally this is done by entirely shielding the panel from particles while operating everything else in the exact same configuration. However, shielding the detector entirely such that it can be sure that all signal are a result of internal noise, and not external particles, seems unlikely to be possible and therefore a better method for measuring the readout noise would be required.

Chapter

6

Machine Learning Methods

The concepts behind antihydrogen detection within the ALPHA experiment were discussed in Chapter 4, and the SVD itself was used to verify simulated panel efficiencies in Chapter 5. However, in addition to the charged products of an antihydrogen/antiproton annihilation within the trap, the detectors are both sensitive to the passage of charged particles originating from cosmic rays. These cosmic rays are our major source of background, and account for a vertex rate of ∼7Hz in the SVD, and ∼ 70Hz in the TPC. The need to differentiate these types of event (especially for experiments with low signal) has motivated the need for some analysis capable of discriminating between these events. Fortunately, the fingerprint of antiproton annihilations, and cosmic ray events is distinct enough that it is possible to effectively filter events on a per-event basis. The preferred method to perform this analysis is generally by using traditional rectangular cuts (discussed in [153]) or *machine learning* (ML).

ML is a subset of artificial intelligence in which specific algorithms are developed that can "learn" from a set of given data, and generalize this information to unseen data. In this way, the model is capable of making decisions without explicit instructions or intervention from a human. The benefits of ML often include improved accuracy and efficiency, the ability to handle large and complex datasets, automation of repetitive tasks, and the potential to discover insights and patterns that may not be apparent through conventional analysis methods. Further, ML allows one to find if a pattern even exists when it is not necessarily clear to human eyes.

Machine learning models are generated using specific algorithms which depend on the architecture of the given model, known as the *training* process. The choice and resulting accuracy of which depends on the dataset itself, as well as the patterns that exist within it. Often, the *best-performing* algorithm is selected, where performance grading depends on application, though this choice is not always well-defined.

Two forms of ML model exist to handle two different types of problem:

1. *Regression models* are ML models that aim to predict a **continuous** variable

from the data sample provided, for example using a home's number of rooms, total ground area, and garden size to predict its value. Regression allows one to estimate the conditional expectation of the dependent variable (house price) when the independent variables take on a given set of values (e.g. when number of rooms, ground area, and garden size = (5, 50m^2, 25m^2)). Specifically, regression models are functions that take the input variable(s) and maps them to a continuous interval.

2. *Classification models* are models that aim to predict a **discrete** variable or "class" that the data sample belongs to. For example classifying emails as genuine or spam based on specific markers of the email. A classifier that only aims to place the samples in one of two classes is known as a *binary classifier*, otherwise it is a multi-class classifier. Here the model is a function that takes the input variable(s) and maps them to a specific set of discrete values.

6.1 Training

The training process is a critical step in developing an ML model, and defines how the model learns the patterns and relationships existing in a given dataset. This dataset is known as the *training data* and can come in a few forms which vary the "type" of learning performed:

1. **Supervised learning** occurs when the training sample is strictly labelled, each entry has a resulting output, or *target* variable. This would be the dependent variable for regression and the class for classification.

2. **Unsupervised learning** occurs when the specific label is not known, and the goal is for the algorithm to find patterns in the data that may not be known prior to the training. While this is more commonly used in classification (and referred to as *clustering*) technically unsupervised regression algorithms do exist.

3. **Semi-supervised learning** occurs when the training data contains a mix of labelled and unlabelled data. Generally this process will train initially on

a small amount of labelled data, then later on the larger batch of unlabelled data.

4. **Reinforcement learning** is a method which attributes positive and negative values to desired outcomes or actions, pushing the model to avoid negative samples and seek out positive outcomes. It can be used in combination with unsupervised learning as a way to direct the training process.

Generally in machine learning literature x is used to denote the input/feature variable, and y is used for the output/target. The input does not have to be scalar and can contain many variables, in this case it is referred to as the *feature vector*, and is denoted $\mathbf{x}$. The size of this feature vector must remain constant for each *sample* in the dataset. Further, the output need not be scalar, and using a vector as a target variable in this way is known as “multivariate regression.” More commonly however, when there are multiple dependent variables, one would train N different models for each variable $y \in \mathbf{y}$ instead which is known as *univariate regression.* It is important to note that both of these are different concepts to “multivariable regression” which is just any regression with multiple inputs as these concepts are often confused [169, 170].

The set of labelled examples of inputs and outputs is known as the training sample: $D = \{(\mathbf{x}_0, y_0), (\mathbf{x}_1, y_1), ..., (\mathbf{x}_k, y_k)\}$. Each member of this set is known as a *sample*, and the total number of samples, k, is an important metric when collecting data. In general, k will need to scale with the complexity of the problem; however there is no well-defined way to calculate the exact amount of data needed to achieve good results. Generally though, the more samples in the training set, the more likely the model is to converge to a solution without over-training.

The overall goal of most machine learning methods is to find the mapping f (known as *decision function*) that best approximates the true underlying relationship between the feature and target variables:

$$f(\mathbf{x}) = y \tag{6.1}$$

f can take many different forms, but let us call the set of parameters used in the function $\Theta = \{\theta_m\}$. Then for each ML method used, the structure of the decision function is fixed, such as the number of parameters in $\{\theta_m\}$, and their arrangement; however, the parameters themselves can change with the objective to find the best set of parameters s.t. $f(\mathbf{x}|\Theta)$ describes the output best.

For example, when fitting a straight line to a data set (which is a basic form of regression) one would have a decision function of the form $f(x) = mx + b$, and the set of model parameters $\Theta = \{m, b\}$. The function f is fixed throughout, but the training process will aim to find the optimum values of m, b that most accurately describe the relationship between the given data.

This is done by minimising a "loss function" $\mathcal{L} : \{\mathbf{x}, y, y_p|\Theta\} \to \mathbb{R}^+$ between true (y), and predicted ($y_p \equiv f(\mathbf{x})$) outputs.

Ultimately, the choice of ML model essentially amounts to a choice of f and $\mathcal{L}$, along with a certain way to generate/adjust $\{\theta_m\}$ at each iteration of the training process that will converge.

6.1.1 Over/Underfitting

Overfitting can occur when a model has been trained to correspond too closely to the training set and therefore is unable to predict future observations accurately. That is, the model contains more parameters than can be justified by the data. It corresponds to having accidentally fit a model to predict the variation of the measurements in the dataset as if they were part of the model.

Similarly, underfitting occurs when the model used to describe the data is too simple to accurately represent the underlying relationship.

An example of over and underfitting can be seen in Figure 6.1 where data generated from a sine curve with some additional Gaussian noise has been fitted with multiple different polynomials of varying degree.

In the case of the overfitting example the model resembles the data set too closely and is trying to predict the Gaussian noise in the samples; in the example of underfitting the linear fit to the data is too simple to accurately predict the sample data. However, when selecting a polynomial of degree 5 the model is able to predict/approximate the data well.

Finding evidence of overfitting is generally simple, when the accuracy of the training data and the testing data starts to diverge it's likely the model is starting

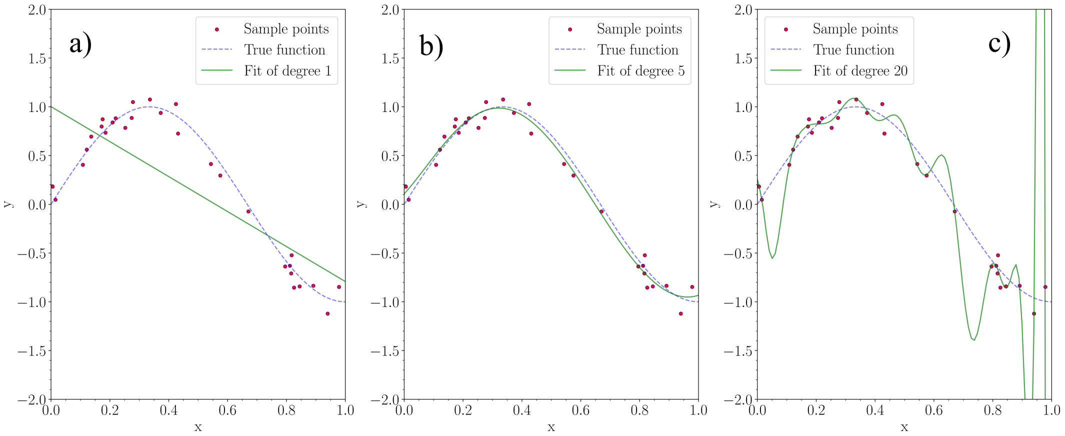


Figure 6.1: Example of over/underfitting showing how model hyperparameters can either cause the model to be underfit, overfit or just right. The points are generated from a sine curve with a small amount of Gaussian noise added representing noise in a given measurement. Figure a) shows the result of applying a linear fit to the data, leading to underfitting; b) shows a polynomial of degree 5, which fits the data well; and c) shows the result of a polynomial of degree 20 which overfits the data.

to overfit to the training data. Signs of underfitting can be less clear but in general models with a high average bias (difference between model prediction and target) are likely underfit.

6.1.2 Curse of Dimensionality

The *curse of dimensionality* refers to various phenomena that arise when analysing data in high-dimensional phase spaces that do not occur in low-dimensional spaces. Generally it refers to issues that arise when the number of samples is small relative to the dimensionality of the samples. While also not necessarily well-defined in definition or consequences as they can vary a lot, one very relevant example for ML use is that of various distance metrics as the number of dimensions increase. The Euclidean (L_2) norm ($||\mathbf{a} \cdot \mathbf{b}||_2 = \sqrt{\sum_{i=1}^{m}(a_i - b_i)^2}$), or Manhattan ($L_1$) distance ($||\mathbf{a} \cdot \mathbf{b}||_1 = \sum_{i=1}^{m}(|a_i - b_i|)^1$, for example, are frequently used in ML to measure the "distance" between sample points in high-dimensional phase space. Distance metrics are essential for many types of models such as k-nearest neighbour (Sec. 6.2.1), or the loss functions in regressions models (Sec. 6.3). However,

---

[1] Both of which are special cases of the more general Minkowski Distance: $(\sum_{i=1}^{m}(|a_i - b_i|)^p)^{1/p}$ for $p = 1, 2$ respectively.

as both of these distance metrics break down at high dimensions [171, 172] there is incentive to keep the number of variables limited when possible. Suggestions of using L_k norms with $k < 1$ have been put forward to combat this issue [172] but it has recently (2020) been shown that these also have issues at higher dimensions [173]. As such to prevent this issue entirely, or any others that come as a result of high-dimensionality data, some form of dimensionality reduction is often performed prior to training.

Further examples of the curse in action include data becoming sparse, computational methods becoming exponentially slower, and the amount of samples required increasing.

6.1.3 Dimension Reduction

Dimensionality reduction is when data is transformed from a high dimensional space to a lower one in such a way that the low dimensional space retains the meaningful properties of the original space. An intuitive example of this in action is reducing (x, y) coordinates simply to r, which for some datasets may be enough to retain properties of the original position. Generally some information is lost in this process, but the key is to reduce the data in such a way that all meaningful information is retained. Converting (x, y) to r is an example of *feature extraction* whereas simply dropping variables is known as *feature selection*, both of which are valid techniques.

6.1.4 Normalisation

Normalisation can have a lot of meanings, however in the context of machine learning it almost exclusively refers to adjusting the values of the inputs to match a different, usually (but not necessarily) common scale.

One often used example of data normalisation is called *standardization* which converts a value measured x, to a *standard score* z, which represents the distance (in units of σ) between the measured value x, and the mean of the population via the formula

$$z = \frac{x - \mu}{\sigma} \tag{6.2}$$

where μ and σ are the mean and standard deviation of the population respectively. In cases where neither of these values are known the z-score is estimated by taking the sample mean, and sample standard deviation instead.

For normalising feature variables in ML applications a specific form of normalisation is deployed known as *feature scaling* which is used to scale the range of the feature variables and is part of the data preprocessing step.

Min-max scaling is one of the most popular and simple methods of normalisation which is usually used to scale the features to the range $[-1, 1]$ or $[0, 1]$, though it can be generalised to any interval. The general formula for scaling the variable x to the interval $[a, b]$ is given by

$$x\prime = a + \frac{(x - \min\{x\})(b - a)}{\max\{x\} - \min\{x\}} \tag{6.3}$$

In machine learning methods where distance metrics in phase-space are deployed such as k-nearest neighbour (Sec. 6.2.1) and neural-nets (Sec. 6.4) data normalisation is an essential step in data preprocessing, as any variable that has a range significantly larger than another will dominate the resulting distance metric.

6.2 Classification Models

We have discussed generally the training process, and some important features of data pre-processing, now we look at the specific algorithms in place for training some models. It has been mentioned how the choice of model is ultimately a choice of f, $\mathcal{L}$ and Θ, and in regression models these are much more well-defined. However, in many classification models these objects are more abstract and the training process follows a more algorithmic rather than analytic approach to training. This section describes a few of the more commonly used classification models and their training algorithms.

In the case of classification models, the target variable (y) is discrete and non-continuous (i.e. $y \in \{0, 1, ..., n\}$ where n is the number of classes, and $n = 1$ for a binary classifier).

6.2.1 k-Nearest Neighbour

The k-nearest neighbour (k-NN) algorithm is a form of supervised learning developed in 1951 [174] and subsequently improved upon throughout the following decades [175].

While technically it can be used for both classification and regression it is far more popular as a method of classification due to its consistency and the limit that can be imposed on the error rate with respect to the Bayes error rate [175][2].

k-NN works simply by predicting the label of a given value by considering the labels of the values "closest" to it in the training data set. Sometimes this is limited by only looking at the k nearest neighbours, or only at neighbours within a certain "distance" d. It relies on the assumption that similar classes will sit closer to each other in feature space, creating clusters that can be compared to the point to be classified.

To formalise this consider the training set D to be a set of pairs,

$$D = \{(\mathbf{x}_0, y_0), (\mathbf{x}_1, y_1), ..., (\mathbf{x}_k, y_k)\} \subseteq \mathbb{R}^m \times \{0, 1\}, \tag{6.4}$$

so that y_i is the label of $\mathbf{x}_i$ (in this case a binary classifier). Then given some distance metric $||\cdot||$ on our feature space ($\mathbb{R}^m$) and a point we would like to classify $\mathbf{x}'$ we can reorder the training set in order of distance from $\mathbf{x}'$. I.e. we can consider $D' = \{(\mathbf{x}_a, y_a), ..., (\mathbf{x}_b, y_b)\})$ to be an ordered set such that $||\mathbf{x}_a \cdot \mathbf{x}'|| \leq ... \leq ||\mathbf{x}_b \cdot \mathbf{x}'||$.

Then, the classification of the point $\mathbf{x}'$ is simply either the majority vote of the first k values in D', or the majority vote of all neighbours in D' such that the norm is less than some positive distance d.

In theory any vector norm on $\mathbb{R}^m$ is valid, however the most commonly used vector norms are the Euclidean distance (L_2 norm) or Manhattan distance (L_1 norm).

It's notable that this method requires recalculating the distance between every point in D and the new point $\mathbf{x}'$ when classifying a new point, meaning it can be computationally inefficient, especially when one has many data points, or high dimensional data.

---

[2] The Bayes error rate is the minimum achievable error rate given the distribution of the data.

For this reason, as well as the curse of dimensionality in the context of the Euclidean norm (discussed in Sec. 6.1.2), dimension reduction is usually performed prior to the k-NN algorithm when the feature space has high dimensionality or large sample dataset [176].

Further, since k-NN relies on distance metrics to make predictions, normalisation of data is very important, especially when the features represent different physical units or come in vastly different scales, something very common with our chosen variables.

6.2.2 Decision Trees

Decision trees are flowchart-like structures, resembling trees, used to predict an outcome. The tree is made up of nodes, branches, and leafs where each node denotes a feature of the dataset, each branch a decision (or a cut), and each leaf a prediction. The topmost node in the tree is known as the *root* node. It learns to partition the tree recursively based on the *separation* of the feature values. Their easy visualisation and interpretability makes them desirable for many uses, especially where *explainability* is important.

To train an individual tree first consider the dataset D to be as in Equation 6.4. Let $\mathbf{x}_0 = \{a_0, b_0, c_0\}$ be our input vector in feature space consisting of three unique features a, b, c. In this specific example $m = 3$ as the input consists of three variables/features. Then consider the set D_a to be a sorted set of one dimensional pairs, using only one of the three available features: $\{(a_0, y_0), (a_1, y_1), ..., (a_k, y_k)\}$ such that $a_0 \leq a_1 \leq ... \leq a_k$. We are able to split D_a into two separate subsets, using any point as the boundary (usually the mean of two adjacent points), essentially placing a cut on the variable a, and by doing so the two resulting subsets D_L, and D_R[3] will have differing amounts of samples from class 0 ($y = 0$) and from class 1 ($y = 1$). The amount these two subsets differ in this regard is called the *separation*, and a good cut on a well performing variable will have a higher separation index than one that doesn't. There are multiple ways to place an absolute value on this separation but the GINI index is the most common for

---

[3] Here $D_L = \{(a_0, y_0), ..., (a_l, y_l)\}$ and $D_R = \{(a_{l+1}, y_{l+1}), ..., (a_k, y_k)\}$ for some $0 \leq l \leq k$ so that $D_x = D_L \cup D_R$.

machine learning methods. The GINI index of a cut is defined as $p(1-p)$ where p is the *purity* of one of the sets after the cut $p = |\{D_L \mid y = 0\}|/k$. This process can be tested with every feature (a, b, or c) in the phase space, and with every cut in the sorted set, resulting in a final cut on a specific variable in the input space that provides the best separation of the two populations. This cut is then chosen as a branch for the decision tree, and a new node is created repeating the same process, this time only with entries in D that passed (or failed) the cut.

If a cut results in a pure node (i.e. one where all the elements belong to the same class) the process is stopped here and this node becomes a decision leaf. This whole process repeats until all nodes are leaf nodes or a sufficient stopping condition is reached.

An example of a decision tree, designed to classify animals as either birds or dogs is shown in Figure 6.2.

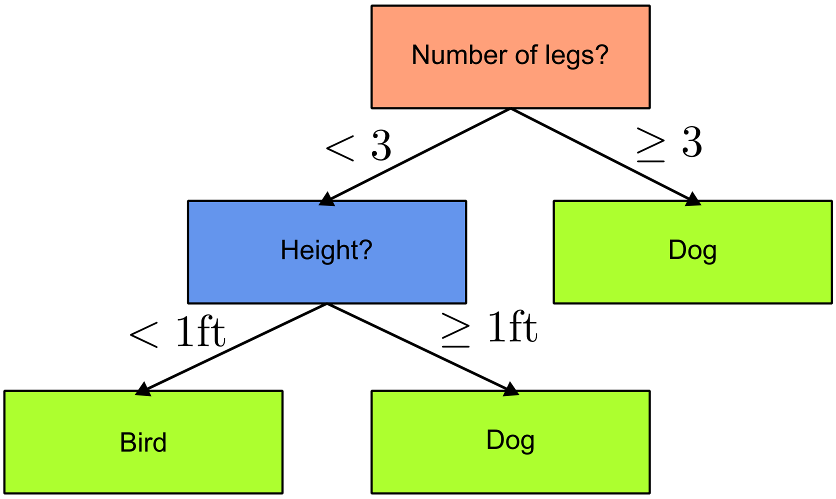


Figure 6.2: An example of a decision tree designed to classify animals into one of two classes: bird or dog, based on the height and number of legs. The root node is in pink, the decision nodes/leafs are in green, and a normal node can be seen in blue. The arrows, along with the corresponding cut value are the branches. It is notable how explainable this model is, we can see the decisions the model is making and understand them, something not possible with more abstract models, however this sometimes comes at a price as one can also note that any two legged bird larger than 1ft in height will be incorrectly classified as a dog in this configuration. Here the maximum depth of the model is two, as it contains two layers.

In practice, doing this until all nodes are 100% pure will result in overfitting, so instead usually there is a limit to the depth of a tree, and to the purity required for a node to become a leaf. Thus, instead of deploying a single giant tree with every possible decision inside, generally multiple of these smaller trees are deployed together, in a process known as *ensemble learning*.

6.2.3 Ensemble Learning

On its own a single decision tree is known as a *weak learner*, i.e. generally it is only slightly better than guessing the outcome randomly. A *strong learner* however, is one that is "well-correlated"[4] to the true classification problem.

In 1989 Kearns and Valiant posed the question "Can a set of weak learners become a strong learner?" [177]. This question was answered in 1990 with a "yes" from Schapire [178], who went on to develop one of the more popular ensemble learning methods used today: AdaBoost (or adaptive boosting) [179] which earned him the Gödel prize in 2003.

Another form of ensemble learning is known as *bagging* (**b**ootstrap **agg**regat**ing**) which was introduced by Breiman in 1996 [180] where new trees are trained only with access to a subset of the data (or even of features [181]), resulting in more variety of the trees and overall a more robust model.

There are many forms of ensemble learning other than bagging or boosting, however, both of these are very common, and the resulting models, random forests and boosted decision trees respectively, are some of the most frequently used ML models available today.

6.2.4 Random Forests

Random forests are a form of ensemble learning method where multiple decision trees are combined to create a collection of trees, or a *forest* which is ultimately more robust than the individual trees themselves.

Consider a single tree to be a function (denoted f_t) that maps the feature space to the classification label, i.e. $f_t : \mathbb{R}^m \to \{-1, 1\}$, then a forest is a collection of these trees in the form:

$$F(\mathbf{x}) = \sum_{t=0}^{T} \alpha_t f_t(\mathbf{x}) \tag{6.5}$$

where T is the total number of trees in the forest, and α_t is the *weight* of tree t.

---

[4] This label is somewhat arbitrary, and the true definition of "well-correlated" depends on the given problem and true limit of accuracy possible.

Generally, to prevent each tree from being identical, boosting or bagging is deployed to ensure variety. Otherwise, by following the same separation algorithm outlined in Sec. 6.2.2 we would obtain identical trees at every iteration. In bagging each of the trees is trained on a different subset of the training data set (either via "random sample with replacement", or via "bootstrapping"), then combined (or aggregated) with a unique weight (α) for a final "bagged" model, $F(\mathbf{x})$.

There are multiple ways to decide on the weights, α_t, however most generally $\alpha_t = 1/T$ is used, essentially giving each tree the same weight and an equal say in the overall classification.

6.2.5 Boosted Decision Trees

Boosted decision trees (BDTs) are an ensemble method that can be used incongruently with bagging, and can rely on not only weighting each tree, but also by weighting training samples.

AdaBoost is one such boosting algorithm which works by weighting training samples at each iteration of the training process by whether they were correctly classified by the previous iteration or not. By giving these "harder to classify" events more weight, later trees are incentivised to classify these samples correctly, over the "easy to classify" events, resulting in a model that is, in general, much more accurate.

That is to say, subsequent trees are trained using a slightly modified event sample in which the weights of previously incorrectly classified events are multiplied by a common *boost weight*, denoted α. The weight is a direct result of the misclassification rate (f_{err}): $\alpha = \frac{1-f_{\text{err}}}{f_{\text{err}}}$. The rate of learning is controlled by something known as the AdaBoost beta: $\alpha \mapsto \alpha^{\beta}$.

Further to this the trees themselves are also weighted by how accurate they were (also based on the number of weighted samples it correctly classified). While AdaBoost is generally presented for binary classifiers, it can be expanded to multiclass problems. The resulting classifier is then defined as

$$F(\mathbf{x}) = \frac{1}{T}\sum_{t=0}^{T} \ln \alpha_t f_t(\mathbf{x}) \tag{6.6}$$

It's important to note here that although the weak learners $f_t(\mathbf{x})$ only output a single class prediction -1, or 1, the ensemble method will output a prediction $y_p \in [-1, 1]$ due to the way the weak learners are combined with the weights. The sign of which is generally considered the prediction, and the absolute value the confidence in said prediction. It is possible however to set a cut value somewhere else in the interval (other than the default at 0) to reduce false positives, doing so however has potential to increase false negatives and choosing an optimum cut in this interval is discussed further in Sec. 8.6.

6.2.6 Support Vector Machines

Support vector machines (SVMs) are models that aim to linearly separate two classes in phase space. There are likely many hyperplanes that do this and so the one that maximises the distance from it to the nearest point on either side is known as the *maximum-margin hyperplane*.

By constructing multiple hyperplanes SVMs are able to be multi-class classifiers, and by deploying the *kernel trick* (discussed in Sec. 6.3.3) they are able to create non-linear hyperplanes which helps when data is not linearly separable.

More formally the linear hyperplane can be expressed as: $\mathbf{w}^T\mathbf{x} - b = 0$ and the one that separates the data best is chosen for the SVM. Assuming the data is separable (i.e. a boundary exists such that on either side are the two different classes) in the feature space, SVMs can be quite accurate, however their need to evaluate distance metrics between each training sample and multiple hyperplanes can make it extremely slow, especially when the training data contains a lot of samples.

6.3 Regression Models

In this section, we turn to regression models, focusing on how different choices for the decision function f, the loss function $\mathcal{L}$, and the parameter set Θ influence the model's performance and applicability. Several widely used regression models are presented, highlighting their individual characteristics and applications.

6.3.1 Linear Regression

Linear regression (LR) is a model which predicts the linear response between the feature and target variables. The most basic case (where both feature and target are scalars) is known as simple linear regression, and results in a straight line in two dimensions, giving the following decision function:

$$y_p \equiv f(x) = mx + b. \tag{6.7}$$

In this configuration the set of parameters Θ is simply $\{m, b\}$ and they are generally determined by minimising the mean-squared-error (MSE) loss function:

$$\mathcal{L} = \underset{\{m,b\}}{\arg\min} \sum_k (f(x_k) - y_k)^2 \tag{6.8}$$

where the sum is over each member of the training data set D.

However, in the extension to multiple/vector valued features ($x = \mathbf{x}$) which is known as multiple linear regression, or multivariable linear regression we have:

$$y_p \equiv f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{6.9}$$

where $\mathbf{w}^T$ is a vector of coefficients known as the “weights”, and b is the bias term. Both of these are adjusted during the fitting procedure to minimise the same loss function as in Equation 6.8.

Linear regression offers the advantages of simplicity, interpretability, and computational efficiency, making it a good starting point for regression analysis; and, when the data is indeed described by a linear model it works effectively. However, this strength is also its limitation as without the assumption of linearity the model will not capture the data well. Further, linear regression models can be sensitive to outliers, and have the potential for overfitting with high-dimensional data. Generally though its inability to capture non-linear patterns in the data is what limits its applicability in more intricate real-world scenarios.

6.3.2 Ridge Regression

Ridge regression (RR) is essentially identical to linear regression, and the decision function is the same, but the loss function contains an extra regularisation term ($\frac{\lambda}{2}||\mathbf{w}^T||^2$) which combines the L2-norm of the weights with a scalar (known as the "strength" of regularisation, denoted λ). This scalar value governs the relative importance of the regularisation term as compared to the sum-of-squares error. The resulting loss function can be seen in equation 6.10:

$$\mathcal{L} = \underset{\{\mathbf{w},b\}}{\arg\min} \sum_k (f(\mathbf{x}_k) - y_k)^2 + \frac{\lambda}{2}||\mathbf{w}^T||^2 \tag{6.10}$$

This regularisation term acts as a penalty for high weights, meaning the fitting process will try to keep the weights as small as possible. This can help increase the overall robustness of the model in comparison to linear regression [182], however, the same limitations apply here as with standard linear regression.

6.3.3 Kernel Ridge Regression

One thing to note about the two previous models is that while they make work for linear relationships in a given dataset, if the relationship between the feature and target is not linear, they will not provide good predictions (see for example Figure 6.1a).

Kernel ridge regression (KRR) differs from standard ridge regression in the decision function, whereby it utilises a transformation of the input vector to a higher dimensional feature space, allowing for non-linear relationships in the feature phase-space.

The decision function in KRR is given by

$$y_p \equiv f(\mathbf{x}) = \mathbf{w}^T k(\mathbf{x}) + b \tag{6.11}$$

where $k(x)$ is the transformation, potentially to a higher dimensional space, which described the data better than the previous linear relationship. Note that linear regression is just a form of KRR where $k(x) = x$.

This mapping ($k : \mathbf{x} \rightarrow \mathbf{x}'$) may be computationally difficult to calculate so to circumvent this issue, something known as the "kernel trick" is applied. During the fitting process, only the dot product between samples is actually required ($k(\mathbf{x}_0) \cdot k(\mathbf{x}_1)$). So if we let $K(\mathbf{x}_i, \mathbf{x}_j) = k(\mathbf{x}_i) \cdot k(\mathbf{x}_i)$ and set K as something computationally simple, we are able to calculate the weights required for Equation 6.11 without needing to perform the implicit mapping k. The function K is known as the "kernel function", and many choices for K exist depending on the given problem, such as:

- Linear (this collapses to linear regression given the same loss function):

$$K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$$

- Polynomial (once again this collapses to linear regression for $n = 1$ and $r = 0$):

$$K(\mathbf{x}, \mathbf{y}) = \left(\mathbf{x}^T \mathbf{y} + r\right)^n, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, r \geq 0, n \geq 1.$$

- Radial basis function (RBF):

$$K(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \sigma > 0$$

Following the methods in [183] using the representer theorem [184] the decision function in Equation 6.11 can be equivalently written as:

$$y_i \equiv f(\mathbf{x}) = \sum_j \alpha_j K\left(\mathbf{x}_j, \mathbf{x}\right) \equiv \alpha_j k\left(\mathbf{x}_j\right) \cdot k(\mathbf{x}) \tag{6.12}$$

Here the sum over j is over the training set, and α_j are the Lagrange multipliers used during fitting, essentially the weights.

Interestingly, while the goal of most ML methods is to find the optimum set of model parameters $\Theta = \{\theta_m\}$ s.t. the loss is minimised based upon previously seen training example, KRR differs in the sense that it is a form of "instance-based" or "memory-based" learning. It "memorises" certain samples, attributes weights to them and constructs its prediction on the training samples themselves. This sometimes means that the complexity of KRR models can grow with the amount of data, however, many computation tricks are deployed to limit this issue [183].

The same loss function, measuring the mean squared distance between each sample with a regularisation term, is utilised as in ridge regression (Equation 6.10).

A good (and classic) example of the usefulness of this transformation to higher dimensions is to return to classification, and to consider 2D points inside and outside a given circle as belonging to two classes needing to be linearly separated. Trying to linearly separate points in a circle with a straight line will not work, however using a kernel function of $k : (x, y) \rightarrow (x, y, x^2 + y^2)$ transforms these 2D points to 3-dimensional space, where they are now able to be linearly separated by a hyperplane. A visualisation of this example can be seen in Figure 6.3.

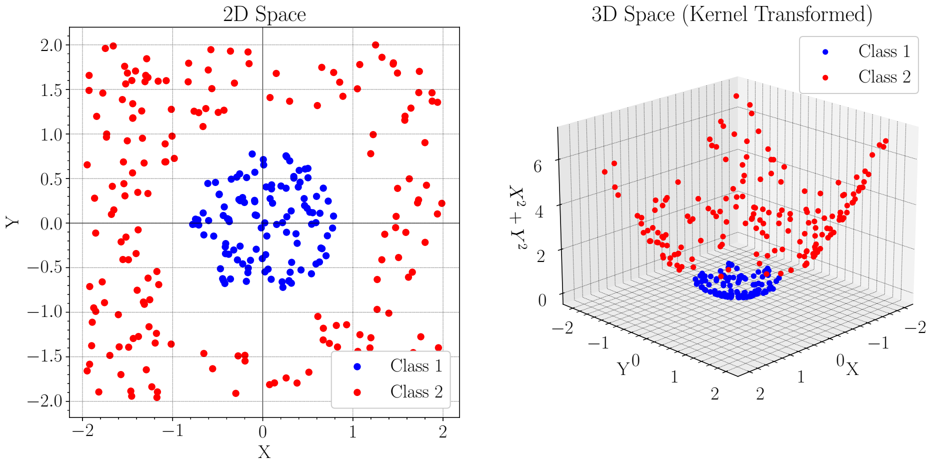


Figure 6.3: An example of the kernel trick transforming 2D, non-linearly-separable data to a higher dimension, where it is now able to be separated using a three-dimensional linear hyperplane.

This same concept is applied in both regression or classification to allow non-linearities to be represented by the model which cannot occur in linear models. However, as KRR utilised this additional function, and "memorises" training samples this can make them more computationally intensive to train and use. Regardless KRRs are still a powerful tool for various regression tasks.

6.3.4 Support Vector Regression

Support vector regression (SVR), proposed in 1997 by Vapnik et al. [185] support vector machines (SVMs) work similarly to KRR, differing only in the loss function, where instead of the standard ridge loss, SVR introduces what is known as an "epsilon-insensitive loss". This new loss function includes a $\max\left\{|y_p - y| - \epsilon, 0\right\}$ term, which doesn't punish predictions deemed "close enough" to the target the distance of which is defined by ϵ. This amounts to minimising the loss function subject to the condition $|y_p - y| > \epsilon$ for some $\epsilon > 0$. The complete loss function (keeping still the regularisation term from Sec. 6.3.2) is:

$$\mathcal{L} = \arg\min_{\Theta} \sum_j \max\left\{|y_p - y| - \epsilon, 0\right\} + \frac{\lambda}{2}||\mathbf{w}^T||^2 \tag{6.13}$$

The corresponding decision function is defined as:

$$y_p = \sum_j (\alpha_j - \alpha_j')K\left(\mathbf{x}_j, \mathbf{x}\right) + b \tag{6.14}$$

This time the sum over j is not over the entire training set; but over the subset of samples selected for inclusion in the loss function. These samples are known as the "support vectors", and again α_j, α_j' are the Lagrange multipliers used during fitting.

It is important to note here that both kernel methods require a selection of kernel and some hyperparameter tuning, the choice of which can often feel arbitrary and random. Either extensive knowledge of the problem and data is required, or an exhaustive search on the hyperparameter space to find optimum solutions. There is no closed-form solution for selecting the "correct" hyperparameters.

6.4 Neural Networks

Artificial neural networks (ANNs), or just "neural networks" (NN), are computational systems that draw inspiration from the intricate networks found in biological brains and can be used for both classification and regression purposes. These networks consist of interconnected nodes known as artificial neurons, which serve as simplified models of the neurons in our brains. Similar to biological synapses, connections between neurons enable the transmission of signals throughout the network. Each artificial neuron receives various input signals, processes them, and can then transmit signals to other connected neurons further downstream the network. The strength of these connections (represented by weights) adjusts as the network learns from data. Non-linear functions applied to the sum of inputs determine the output of each neuron.

There are many kinds of ANN, but the one most commonly utilised is the multi-layer perceptron (MLP).

6.4.1 Multi-Layer Perceptron

An MLP is a fully-connected forward-directed graph; consisting of *nodes* (or neurons) and *edges* (or synapses). It is a specific architecture for an ANN, of which there are many options.

In general, MLPs have at least three *layers*: an input layer (the same size as the input $\mathbf{x}$); one or more hidden layers of varying number of nodes; and finally an output layer (the same size as the output y). An example diagram of an MLP with three layers and a single output can be seen in Figure 6.4.

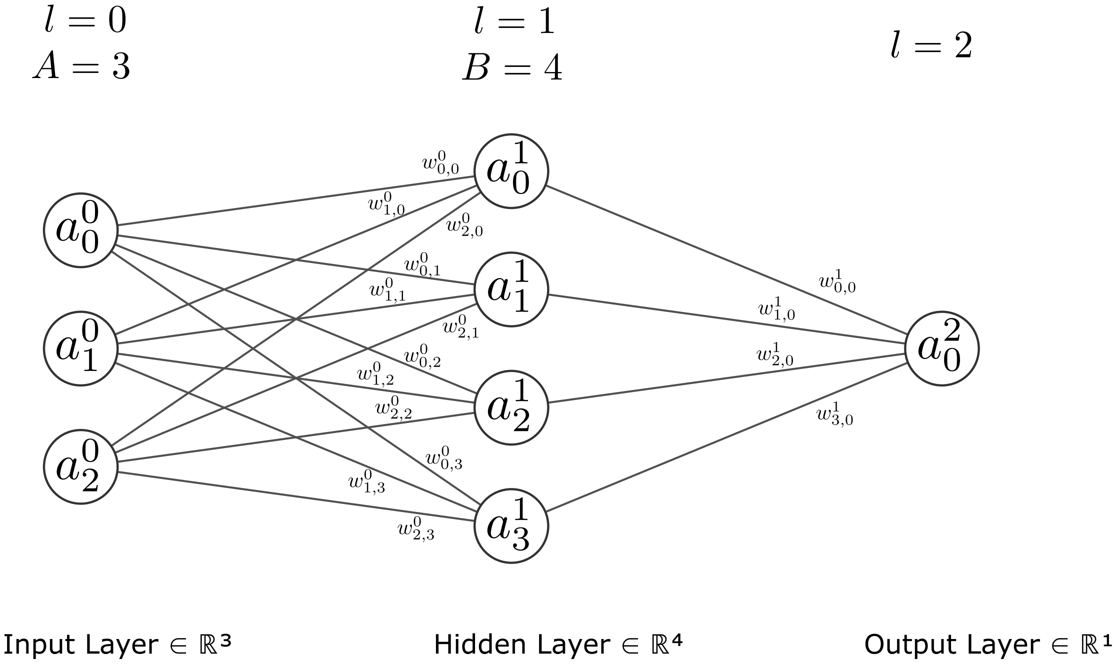


Figure 6.4: Example schematic of a 3-layer MLP with layer sizes (3,4,1). Each node's value/output is denoted a_j^l where $a_j^0 = x_j$ is the input/feature vector. Here l refers to the *layer* number (from 0-2) and A, B represent the number of nodes in layers 0, 1 respectively. In this case the output is a scalar and therefore the size of the final layer is fixed at 1, representing our prediction.

In this schematic a_i^l represents the output of the i-th neuron in layer l, and $w_{i,j}^l$ represents the weight of the edge connecting neuron i in layer l with neuron j in layer $l + 1$.

For each node in the network there also exists a *bias* term b_i^l (not displayed in Figure 6.4) such that the net input of neuron i in layer l can be described as:

$$z_i^l = \sum_j w_{i,j}^{l-1} a_j^{l-1} + b_i^{l-1}. \tag{6.15}$$

To obtain the final *output* or *activation value* of that node (a_i^l) we then apply an *activation function* (denoted $g(x)$) to this input:

$$a_i^l = g(z_i^l) = g\left(\sum_j w_{i,j}^{l-1} a_j^{l-1} + b_i^{l-1}\right) \tag{6.16}$$

In a fully-connected graph each node in a previous layer is connected to every node in the next layer; and the choice for activation function is similar to that of the kernel where it should be non-linear to reflect non-linearities in the data. Common choices for this function include the sigmoid function:

$$g(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} = 1 - g(-x), \tag{6.17}$$

tanh activation:

$$g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{6.18}$$

or the rectified linear unit (ReLU) function:

$$g(x) = x^+ = \max(0, x). \tag{6.19}$$

Generally the tanh function is preferred over sigmoid as it seems to improve performance but it does not solve the “vanishing gradient problem” like the ReLU function does.

A MLP with at least one hidden layer is known as a “universal approximator”, and it can be shown via the universal approximation theorem [186] to represent any function given a sufficient number of layers and nodes.

By starting with the input vector and moving forward through the network, calculating the output of each node at each layer eventually the model reaches the output value which is the final prediction of the model.

So, for example, the overall decision function of the model shown in Figure 6.4

using activation function g, can be described as:

$$y_{p,i} = g\left(\sum_{m}^{B} w^1_{i,m} \cdot g\left(\sum_{j}^{A} w^0_{m,j} x_j + b^0_m \right) + b^1_i \right). \tag{6.20}$$

More generally, we can define the decision function for a generic MLP with a total of L layers (including input and output layers) by incorporating each b^l_i into the function g^l such that $g^l(x) = g(x + \mathbf{b}^l)$, and by denoting the matrix of weights at layer L with $\mathbf{W}^L$ as:

$$y_p \equiv f(\mathbf{x}) = g^{L-2}(\mathbf{W}^{L-2} g^{L-3}(\mathbf{W}^{L-3} ... g^0(\mathbf{W}^0 \mathbf{x})...)). \tag{6.21}$$

While MLPs do still have a loss function, usually, it is referred to as a “cost” function (denoted C, allowing L to be free to describe the number of layers), and the MLP converges to optimum values of $\mathbf{W}^L$ and $\mathbf{b}^L$ through a process known as *backpropagation.*

6.4.2 Backpropagation

In the backpropagation process, the loss and activation specifics are unimportant; as long as their derivatives can be evaluated efficiently, the process will converge. These choices will ultimately affect the accuracy of the final model and are considered further hyperparameters of the model, but it does not change the training process, and presented here is a general description of the process for which this choice is independent.

For a given cost function C acting on our training set $D = \{(\mathbf{x}_k, y_k)\}$ (which contains a series of input-output pairs) the total cost of the model, given weights and biases in the sets $W_S = \{w^l_{i,j}\} = \{\mathbf{W}^L\}$, and $B_S = \{b^l_m\} = \{\mathbf{b}^L\}$, is

$$C(D; W_S, B_S) = \frac{1}{k} \sum_k \lambda(\mathbf{y}_k, f(\mathbf{x}_k)) \tag{6.22}$$

where f describes the output of the model (as per Equation 6.21) and λ is a distance metric between inputs and outputs (the mean squared error, $\lambda(\mathbf{u}, \mathbf{v}) = \sum_i (u_i - v_i)^2$ is most commonly used for this). The sum in Equation 6.22 is (in theory) over all k training samples however later in this section we will see why this is not always the case.

After evaluating the value of the cost function a process known as *gradient decent* is utilised to shift the weights and biases. By taking the partial derivative of C with respect to the weights then we are able to find the direction of the step required by the weights that minimise the total value of the cost function.

That is to say by calculating $\frac{\partial C}{\partial w^l_{j,k}}$ and $\frac{\partial C}{\partial b^l_j}$ for each $w^l_{j,k} \in W_S$ and each $b^l_j \in B_S$ we are able to construct a vector, known as the gradient of C:

$$\nabla C = \begin{bmatrix} \frac{\partial C}{\partial w^0_{0,0}} \\ \dots \\ \frac{\partial C}{\partial b^L_L} \end{bmatrix}$$

and take a corresponding minimising "step" in phase space until we reach a global minimum. There is much discussion about the effect of saddle points in this space [187, 188] (or whether they even exist [189]) but in general this process will converge to some minimum value given enough time, or *epochs* as they are commonly known.

In order to calculate ∇C it is necessary to first calculate the value of f for each training sample k (i.e. $f(\mathbf{x}_k)$). This is known as the *forward-propagation* step as we take the inputs $\mathbf{x}_k$ and feed them forward through the network, calculating the response of each node as we go until we get the final response of the model.

Each individual component of ∇C can be computed using the chain rule on Equation 6.21, but doing this for each weight individually will be incredibly inefficient, especially with large models that can often have thousands of weights. Instead, backpropagation efficiently computes the gradient by avoiding duplicate calculations and only calculating the gradient of each layer, denoted δ^L, and avoiding this full calculation entirely.

By using the chain rule, it can be shown δ^{L-1} is dependent only on δ^L, and therefore the values can be saved at each layer, ready to calculate the derivatives for the previous layer. Therefore in backpropagation, we work our way backwards through the network from the last layer to the first, each time using the last derivative calculations via the chain rule to obtain the derivatives for the current layer and populating ∇C.

The derivation and formal definition of this process is omitted for brevity; however the key point in this process is that as the only way a weight in layer L affects the loss is through its effect on the next layer. Further, as this effect is linear, δ^L is the only data one needs to compute the gradients of the weights at layer $L-1$. Further details about this process can be found in, for example, [190–192].

In this way, the backpropagation algorithm allows us to efficiently calculate the gradient with respect to each weight and to avoid duplicate calculations.

We then shift the whole phase space of weights and biases by ∇C, whole layers at a time, and repeat the process until convergence.

This shift of weights and biases is known as a "gradient descent" step, but a true gradient descent step occurs only when using every sample in the training set to calculate the true "most efficient" step to take that minimises C. However, even with the shortcuts of backpropagation, this can still be a slow process, especially in models with many weights, or if we have a large training sample. As such, if we instead take the sum over k in Equation 6.22 to be over some subset of the training data, we are still able to take a good (though not perfect) step in the correct direction. This is known as a "stochastic gradient descent" step and the whole process of using smaller "batches" of samples is known as *batching*.

Commonly used cost functions include the mean squared error for regression (described previously), and binary cross entropy for classification:

$$BCE(\mathbf{x}, y) = -\frac{1}{N}\sum_{i=0}^{N} y_i \log(f(\mathbf{x}_i)) + (1-y_i)\log(1-f(\mathbf{x}_i)) \tag{6.23}$$

where f is a probability of a specific class.

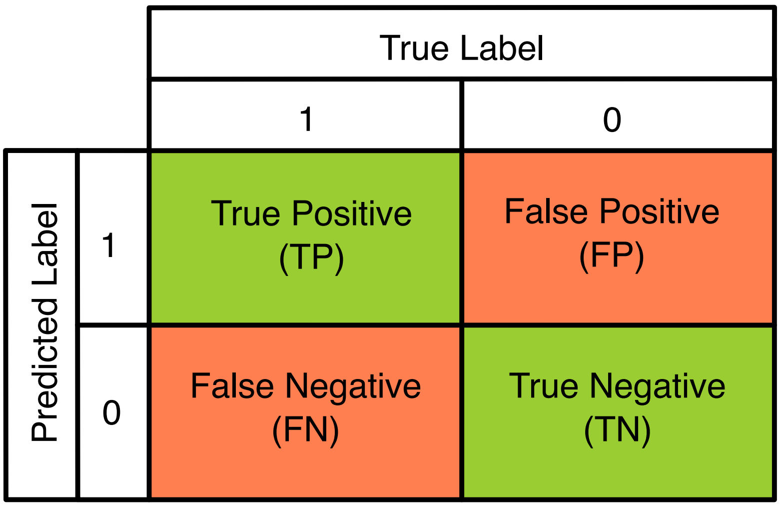
6.5 Metrics to Evaluate Machine Learning Algorithms

Understanding different types of models is only part of the process; it's equally important to employ the right evaluation metrics to quantify their performance. The choice of metric depends on the type of model (classification or regression) since their outputs differ in nature, requiring distinct functions to assess their performance accurately.

6.5.1 Classification Metrics

To evaluate the performance of classification models, the following metrics are commonly used:

1. **The confusion matrix** is essentially a table layout that allows one to visualise the performance of a classifier. Usually used in a binary classifier to compare type I errors (false positives) with type II errors (false negatives) but can be expanded to multi-class models also. An example of a confusion matrix for a binary classifier can be seen in Figure 6.5.



		True Label	
		1	0
Predicted Label	1	True Positive (TP)	False Positive (FP)
	0	False Negative (FN)	True Negative (TN)

Figure 6.5: Example of a 2 × 2 confusion matrix. Generally type I errors refer to FP, and type II to FN (although often they are switched arbitrarily).

This metric is often especially important in medical applications, where false negatives can be more dangerous than false positives.

2. **Accuracy** simply counts as a percentage how many test samples were correctly classified $\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$

3. **Recall**, sensitivity, or true positive rate is $\frac{\text{TP}}{\text{TP+FN}}$

4. **Specificity**, or true negative rate is $\frac{\text{TN}}{\text{FP+TN}}$

5. **Precision** is $\frac{\text{TP}}{\text{TP+FP}}$

6. **F1 score** is the harmonic mean of the precision p, and the recall r: $F1 = \frac{2}{(1/p)+(1/r)}$

7. **Receiver operating characteristic** curve, or ROC curve is used to illustrate the performance of a binary classifier that has varying threshold values such as BDTs or random forests. The ROC curve is a plot of the true positive rate against the false positive rate for each threshold value available.

8. **Area under curve** (AUC), usually of the ROC curve is a way to quantify the total performance of the model. In general models with a higher AUC will perform better, but the shape is not accounted for in this metric. For example, a model may perform better than another in one area of the ROC curve, but worse in another area, resulting in a lower AUC despite performing better at the specific region of interest.

6.5.2 Regression Metrics

Regression models are able to be more numeric in their metrics owing to the continuous nature of the outputs. Some common metrics include:

1. **Mean absolute error** between predictions and outputs:

$$\frac{1}{N}\sum_{j=1}^{N}|y_j - f(\mathbf{x}_j)| \tag{6.24}$$

2. **Mean squared error** between predictions and outputs:

$$\frac{1}{N}\sum_{j=1}^{N}(y_j - f(\mathbf{x}_j))^2 \tag{6.25}$$

6.6 Conclusion

The basis for many machine learning algorithms to be used for the remainder of the thesis have been presented, as well as a few methods for evaluating the performance of the resulting models.

Chapter

7

Machine Learning Reconstruction of Distorted IPM Beam Profiles

During a six-month placement the author spent time at the CERN Beam Instrumentation (BI) group working on using machine learning to reconstruct beam profiles obtained with an ionisation profile monitor (IPM).

The principle of an IPM is quite simple: when a beam passes through a gas chamber it will ionise the gas creating electron-ion pairs. By applying electric and magnetic fields to direct these particles downward towards a detector, it is possible to measure a profile of the given beam.

However, it is possible that the profile becomes distorted before being recorded by the IPM, and this is especially common with high-brightness beams with a large space charge effect.

This chapter briefly examines the underlying mechanisms responsible for these distortions before exploring the application of machine learning algorithms to reconstruct the true beam profile from distorted measurements. While good results are obtained through multiple tests conducted on simulated data, demonstrating the effectiveness of these algorithms, and reproducing results from similar previous studies (see [193–195]); the first applications of these reconstruction methods on data collected from a currently operational IPM, the CERN PS-BGI-82, are promising but not conclusive, likely owing to the problems occurring in the dataset itself. These findings suggest that machine learning algorithms could prove highly valuable for beam reconstruction in IPMs, but further data is required to demonstrate their effectiveness in reality.

7.1 Introduction and Motivation

Using IPMs to measure the beam profile in storage rings is a highly valuable diagnostic, as it is (assuming the density and volume of gas is tuned correctly) *non-destructive*, meaning the profiles can be obtained for the beam in real time, while it is being used for physics. Further, the capability to monitor the beam profile in real-time is an essential aspect of cancer treatment in medical accelerators [196, 197].

The profile that is recorded on the detectors can often vary from the true profile of the beam due to multiple different compounding mechanisms, and while the application of guiding electric and magnetic fields are able to suppress this somewhat, in situations where there are space, cost, or time restrictions the ability to forego these hardware solutions would be preferable. These requirements have motivated the search for reconstruction methods that either do not require a magnetic field, or allow the field to be minimised as much as possible.

7.2 IPMs

IPMs have been widely used in particle accelerators since their invention in the late 60s [198, 199], and are still important tools for modern and future accelerators [200, 201]. They work by passing the beam through a small chamber in which a residual gas is present, causing the ionisation of the gas where the beam passes and results in a non-destructive measurement of the beam profile. Due to the simplicity of their design, and the one dimensional nature of the resulting profiles they can be measured extremely fast, allowing beam profiles to be measured per turn or even per bunch.

A schematic diagram of an IPM can be seen in Figure 7.1, showing the beam (which is travelling towards the page), the position sensitive detectors (usually either an MCP or a pixel detector) and the direction of the electric and magnetic field.

The most commonly used IPMs can be distinguished by their use or omission of a magnetic guiding field [202]. Generally IPMs without need of a magnetic field are favoured due to their smaller size, lower cost, and simpler design. However, with the high brightness beams such as those at CERN, the space charge effects (one of the mechanisms of profile distortion, to be discussed in Sec. 7.3) can often cause large distortions of the profile, and as such generally a magnet it required in their design. The main goal of an IPM is generally to determine the *beam width*, defined as the standard deviation of the one dimensional profile obtained

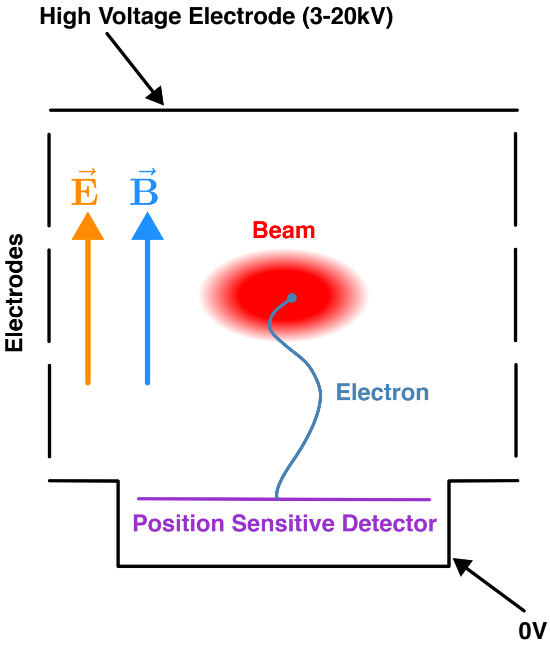


Figure 7.1: Schematic diagram of a typical IPM. The position sensitive detector could be an MCP or a pixel detector/wire array.

by the device, denoted σ. However, the lack of magnetic field can often cause the recorded beam width σ_m to differ from the true beam width σ_0.

7.3 Profile Distortion

The equations of transverse motion for the electrons generated during the ionisation process are generally well-defined. Given an electron of mass m, and charge q in a vertical field of $E_y = V_0/D$, the vertical and horizontal movement $(\mathbf{x}(t) = (x(t), y(t)))$ can be described by the following equation [202]:

$$\mathbf{x}''(t) = \begin{bmatrix} f_x(x,y,t)x \\ \frac{q}{m}E_y + f_y(x,y,t)y \end{bmatrix} \tag{7.1}$$

where

$$f_{(x,y)}(x,y,t) = -\left(\frac{q}{m}\right) \cdot \frac{\partial U(x,y,t)}{\partial (x,y)}, \tag{7.2}$$

and U is the space-charge potential of the bunch. This value is proportional to beam current $J(t)$ and depends on the beam density distribution which typically scales proportionally to the radius squared (r^2) until $r \gg a$ for some a, where it then scales with $\ln(r)$.

As can be seen from Equation 7.2, the more significant the space-charge potential of the beam, the larger the resulting perturbation in x, y will be. Therefore, the electron motion is defined by three main aspects.

1. *Space charge interaction* - The interaction of the electric field generated by the bunch and the electron may alter their trajectory significantly, distorting the measured profile/width. This is especially evident in high-intensity and high-energy beams, and often a large magnetic field is required to overcome this motion.

2. *Ionisation Momenta* - The magnetic guiding field influences the initial momenta of electrons upon ionisation. As such, any initial momentum along the axis translates to a transverse displacement of the electrons, while transverse momentum corresponds to a longitudinal displacement. This effect is more prevalent for beams with lower energy or for IPMs with a weak guiding magnetic field [195].

3. *Gyromotion* - The electrons' motion under the influence of a guiding magnetic field is defined by a helix. The radius of this helix is defined by both the intensity/charge of the beam and the strength of the magnetic field. In cases of a particularly large radius, this can result in a similarly large displacement of the products of the resulting profile.

7.4 Profile Corrections

While the approach to profile correction in this chapter will solely focus on ML methods there do exist many other forms of profile correction discussed in the following sections.

7.4.1 Analytical Solutions

In order to correct the distortions that occur as a result of these mechanisms it is generally required to simulate the beam conditions, compare them to measured results, and define a fit function that relates σ^m to σ_0. [199, 203–205] outline this process for the IPMs used in the Fermilab booster, and the Brookhaven AGS. The formulas derived from these simulations are

$$\sigma_m = \sqrt{\sigma_0^2 + C_1 \frac{N^2}{E_0^2 \sigma_0}}, \tag{7.3}$$

$$\sigma_0 = C_2 + C_3 \sigma_m + C_4 N, \tag{7.4}$$

$$\sigma_m = \sigma_0 + C_6 \frac{N^{1.025}}{\sigma_0^{1.65}} \left(1 + 1.5 R^{1.45}\right)^{-0.28}, \tag{7.5}$$

$$\sigma_m = \sigma_0 + C_7 \frac{N}{\sigma_0^{0.615}} + C_8 \frac{N^2}{\sigma_0^{3.45}} \tag{7.6}$$

(from [199, 203–205] respectively).

In these equations, N is the number of particles in the bunch (generally in units of 10^{12} protons), $E_0 = V_0/D$ is the guiding electric field due to a voltage V_0 over a given distance D, R is the aspect ratio of the beam ($\sigma_{0,y}/\sigma_{0,x}$), and C_i are various fit parameters derived from available simulations and measurements.

Despite the good agreement between these approximations and measured data the unsubstantiated physical reasons for many of the variables used has created a general call for more sophisticated reconstruction algorithms in recent years [202].

Further, such approximations generally only exist for IPMs without a magnetic guiding field, and only a few cases exist for those with a magnetic guiding field [206–209]. In high intensity beams such as those at CERN, where the space charge of the bunch itself can cause major distortions to the recorded profile, the use of a magnetic field is generally considered crucial to minimise profile distortions.

7.4.2 Magnetic Field Corrections

In most IPMs, the magnetic guiding field is strong enough to reduce profile distortions to a minimum; however, in some instances, the space charge interaction can create electric fields up to a few MV/m (compared to the typical extraction field of $\sim 50 - 300$kV/m). Fields of this magnitude mean that initially the electrons are trapped within the space-charge region of the beam due to its dominant field overpowering that of the guiding field near the centre of the bunch. The electrons are then forced onto significantly different trajectories, resulting in a substantial increase in the gyroradii discussed in Sec. 7.3. Determining the suitable strength for the magnetic guiding field to effectively suppress this increase in gyroradii is non-trivial, as these effects depend on various parameters of the particle beam which can be hard to quantify. In theory, it is possible to simply increase the magnetic guiding field until gyroradii are bounded to sufficient levels (many such examples exist of relating this bound to the magnetic field required [195]), however, in some scenarios, to ensure this upper bound is sufficiently low the use of a very large magnetic fields would be required [206]. Acquiring magnetic fields on the order of those needed to ensure sufficiently bounded gyroradii can be costly, and require significant space within already congested rings, making them undesirable. As technology progresses, it is reasonably safe to anticipate increased applications for high-energy and high-brightness beams, (especially with plans for the *future circular collider* (FCC) moving forward [210]), challenging the effectiveness of current magnetic IPM designs.

7.4.3 Other Solutions

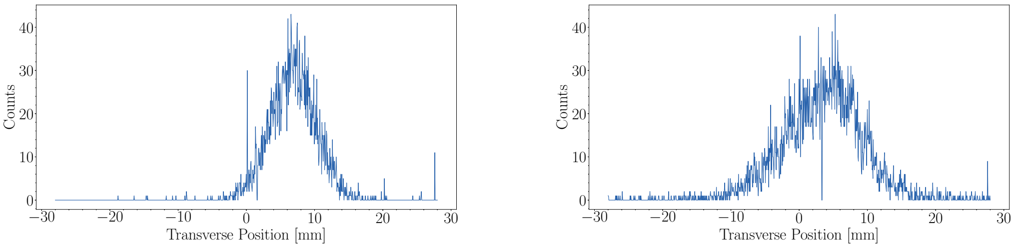
Further attempts to correct for this effect that do not rely on stronger magnets or arbitrary fit functions have been proposed, such as the “electron sieve”, which would aim to filter the electrons before they reach the detector and selectively remove particles with too large of a perturbation. Simulation results showed good potential; however, one of the main pulls of finding these alternate methods is for ease of use and compact use of space. The electron sieve was found to be too large and complicated to implement into existing machines.

Other approaches include parametric curve fitting or utilising more well-defined analytic functions, which could not be generalised or justified strongly.

More recently, however, attempts to create some form of an inverse mapping based on specific beam parameters using look-up tables [211] or supervised learning based on simulated data [193–195] have been attempted. However, both of these rely on well-understood simulations, which are not always readily available.

7.5 CERN PS-BGI-82

In CERN on the PS, a pair of beam gas ionisation (BGI) IPMs have been installed and operational since 2021 [212]. Both use an array of 4 Timepix3 [213] detectors for their position sensitive detection, and are deployed orthogonally to one another, allowing for both x and y transverse profiles to be recorded throughout the beam cycle. Both use electric and magnetic fields to direct the electrons to the detectors, the magnetic field coming from a 0.2T triplet dipole magnet. All data in this section comes from the horizontal BGI, known as the PS-BGI-82. By operating the magnet at full power, and at lower currents we are able to obtain profiles with varying amounts of distortion. An example of a profile recorded by the PS-BGI-82 with the magnet operating at full strength can be seen in Figure 7.2a, and at about 15% the maximum (26A/0.03T) can be seen in Figure 7.2b.



(a) An example profile at 0.2T

(b) An example profile at 0.03T

Figure 7.2: Example profiles from the PS-BGI with the directing magnet operating at different output fields.

It can be seen from these profiles that the reduction of the magnetic field corresponds to a broadening of the recorded profile, demonstrating the point spread that electron would experience inside an IPM [214].

The readout is designed to allow for the number of particles per profile to be tuned, up to a maximum of 1024 profiles per cycle. For example, if during one cycle of a beam in the PS, the ionisation cross-section and the chosen residual gas results in a total of 450,000 detected electrons, then setting the particles per profile (ppp) to 5000 would result in 90 profiles recorded out of the total possible 1024.

7.6 Using Simulated Data

To first test the validity of using ML to reconstruct profiles in the PS-BGI-82 a round of simulations was performed, mimicking both the true and distorted profiles to allow for a complete fully labelled dataset to provide to ML algorithms. Simulations are generally cleaner and much simpler than real datasets, and as such it is expected that our models are able to work accurately on simulated data before considering applications on real data. This form of reconstruction has been shown to work successfully before, albeit in beams with slightly different parameters than those of the PS [193, 194, 215].

7.6.1 The Dataset

The simulations were performed using Virtual-IPM (VIPM), which is an open-source package explicitly developed for bunch charge particle tracking in IPM devices. The package has been tested with various beams and benchmarked to show good agreement [216]. The modular nature of the package allows users to tune each part of the hardware and beam to more accurately represent the expected results.

Table 7.1 shows the beam parameters used for the first round of simulations, initially tuned to those typically found in the PS at extraction [217].

The position of all 3,000,000 simulated particles was recorded upon generation (ionisation) and also as they hit the detector region. The data is then binned into 50μm bins giving a complete training set of “real” and “distorted” profiles for training and reconstruction tests.

7.6.2 Reconstruction of Simulated Profiles

This set of training data was fed into each of the ML methods described in Sec. 6.3, both to predict the original beam width σ_0 and to reconstruct the full original profile $\mathbf{y}$ (except for SVMs which were only used to predict σ_0). The models were trained using the Python machine learning package scikit-learn [218]. Tensorflow [219] was used to verify some models and gave similar results, though they are

Variable	Value Simulated
Beam Particles	Protons
Energy	25.445GeV
Beam Width	850-950μm
Guiding Magnetic Field Strength	0.01T
Bunch population	1.2-1.3e11
Beam Height	$800\mu m$
Beam Length	1.0-1.1ns
Electric Guiding Field	286kV/m
Time step	0.0005ns
Total Particles simulated per profile (ppp)	3,000,000
Total Profiles simulated	19,956
Bin size	$50\mu m$
Particles per μm	$\sim$ 3,157-3,529

Table 7.1: Beam parameters simulated in VIPM for our application. We set the magnetic field to 0.01T so that it is not sufficiently strong to remove the distortions of the profiles, intentionally giving us something to correct.

not displayed here. For each of the models, the following hyperparameters were used: KRR: $d = 2$, $\alpha = 10^{-3}$; SVR: $\gamma = 10^{-4}$, $C = 10^3$, $\epsilon = 10^{-8}$; MLP: 4 hidden layers of sizes: (200, 170, 140, 110), ReLU activation, with a batch size of 8. Detailed descriptions of these hyperparameters can be found in the respective software packages.

The accuracy for each ML method and both the case of predicting width and full profiles of our simulated beam are displayed in Table 7.2. The mean-squared error in the case of predicting beam width is $\sum_k (y_p - \sigma_0)^2$, where y_p is the result of the model, and for predicting full profiles, we first convert the full reconstructed profile ($\mathbf{y}$) to σ_p by taking the standard deviation of the profile once binned in x. This allows both approaches to be compared directly, and while there are further benefits to reconstructing the full profile instead of just the width, if doing so results in a significant decrease in the ability to predict σ_0, then reconstructing the full profiles will not be valuable. Table 7.2 shows the residuals, errors, and correlations (between true and predicted values) for each of the models trained.

These results show that for the beam simulated linear regression was the best-suited model, and with it the original beam width of the distorted profile can be simulated to within $\sim 1\mu$m or $\sim 0.12\%$ of the true beam width. Pearson's

Method	Residual Mean (μm)	Residual Sigma (μm)	NMSE	Pearson's R^2
Linear Regression (predicting σ_0)	0.7722	0.585	0.001 124	0.999 438
Linear Regression (predicting $\mathbf{y}$)	0.7884	0.601	0.001 176	0.999 422
KRR (predicting σ_0)	0.7628	0.579	0.001 098	0.999 451
KRR (predicting $\mathbf{y}$)	0.7722	0.582	0.001 120	0.999 451
SVR (predicting σ_0)	0.8773	0.664	0.001 449	0.999 275
MLP (predicting σ_0)	0.8098	0.608	0.001 228	0.999 442
MLP (predicting $\mathbf{y}$)	0.8099	0.613	0.001 236	0.999 409

Table 7.2: Resulting metrics for each model configuration, both in predicting σ_0 directly, or predicting the full profile $\mathbf{y}$ and then using its standard deviation as y_p. These results suggest that ML is capable of predicting true beam width to within $\sim 1\mu$m or $\sim 0.12\%$ on the simulated beam.

R^2 coefficient shows a strong correlation between prediction and target, all above 99.9%.

However, when predicting the full beam profile, it is possible to obtain more than just the width of the beam, potentially allowing for more diagnostic ability. Table 7.2 shows that using these profiles to in turn predict σ_0 is similarly (though slightly less) accurate as predicting σ_0 directly.

Given that measurements and reconstructions are binned, the process amounts to sampling from an underlying probability density function (PDF), and as such the χ^2 measure is not the best way to compare predicted and true profiles. Instead, we must turn to other statistical methods designed to directly compare histograms drawn from an unknown distribution with expected values of zero as both the χ^2 goodness of fit and histogram distance measure severely punish a predicted value of anything other than 0 when the expected value is 0 (in fact, it returns ∞). Outlined in [220] are many alternative methods for measuring the distance between two histograms. We will use, for two histograms of length n given by h_1, and h_2, the following four methods of measuring the accuracy of our resulting profiles, each of which probes a different aspect of the accuracy of the prediction:

- Manhattan Distance:

$$D_M = \sum_j |h_1(j) - h_2(j)|. \tag{7.7}$$

- Euclidean Distance:

$$D_{E2} = \sqrt{\sum_j (h_1(j) - h_2(j))^2}. \tag{7.8}$$

- Kolmogorov-Smirnov Divergence:

$$D_{KS} = \max_j |h_1(j) - h_2(j)|. \tag{7.9}$$

- Pearson's Correlation Coefficient:

$$D_P = \frac{\Sigma_j \left(h_1(j) - \frac{1}{n}\right)\left(h_2(j) - \frac{1}{n}\right)}{\sqrt{\Sigma_j \left(h_1(j) - \frac{1}{n}\right)^2 \Sigma_j \left(h_2(j) - \frac{1}{n}\right)^2}}. \tag{7.10}$$

These metrics, comparing the true and predicted beam profiles are displayed in Table 7.3. For comparison, these metrics for a set of 2613 profiles which were simulated with the magnet at 0.2T, suppressing the majority (but not all) of the distortions is included as a benchmark.

Method	$D_M(\mu)$	$D_{E2}(\mu)$	$D_{KS}(\mu)$	$D_P(\mu)$
0.2T Magnet	100.2189	13.1476	4.4141	0.999 987
Linear Regression	91.1116	11.9112	3.9997	0.999 989
KRR	89.5337	11.7404	3.9513	0.999 989
MLP	89.2321	11.6961	3.9379	0.999 989

Table 7.3: Distance metrics between true and predicted profiles for each ML method. In this case, the MLP was the best at reconstructing the full original profiles even surpassing the accuracy of the 0.2T magnet.

This table suggests that using linear regression to reconstruct the profiles with our simulated data is just as accurate, if not more so, than using a 0.2T magnet. An example of one such reconstruction using each method, and compared to both the distorted and original profile, can be found in Figure 7.3 where it can be seen that on the scale of the distortions, the difference between the original profile and the various reconstructions is negligible.

The study was performed with many other beam parameters and magnet strengths all of which gave similar results and degrees of accuracy. In general, regardless of the beam setup, using data generated in VIPM to predict the original beam profiles or width gave accurate models. For example, the same pipeline was run on a beam with the parameters found in Table 7.4 and the results obtained (which can be found in Table 7.5 and Table 7.6 respectively) show similar results to the first beam simulated.

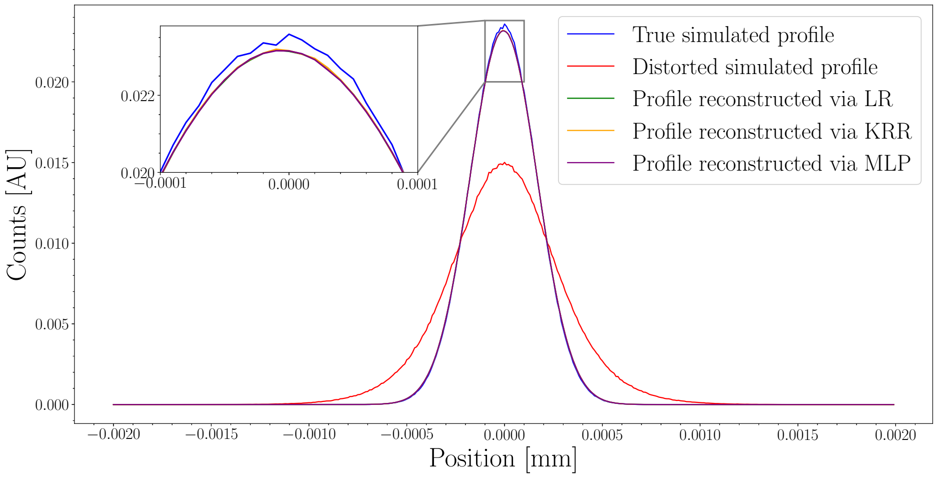


Figure 7.3: Example of a simulated profile, its distorted profile, and the reconstruction generated from each ML method. Here the reconstructed profiles are overlaid and do not show much difference.

Variable	Value Simulated
Beam Particles	Protons
Beam Width	3000-5500μm
Guiding Magnetic Field Strength	0.0-0.03T
Total Profiles simulated	10,560
Bin size	$55\mu m$
Particles per μm	∼545-1000

Table 7.4: Table showing beam parameters simulated in a second round of VIPM simulations. Here the magnet varies from 0-0.03T and the beam width is wider, at 3-5.5mm. Parameters not included in this table are the same as those found in Table 7.1.

Method	Residual Mean (μm)	Residual Sigma (μm)	NMSE	Pearson's R^2
Linear Regression (predicting σ_0)	131.3219	83.4145	0.046 93	0.976 310
Linear Regression (predicting $\mathbf{y}$)	132.6185	84.8920	0.048 08	0.975 774
KRR (predicting σ_0)	77.8011	49.2558	0.016 44	0.991 804
KRR (predicting $\mathbf{y}$)	78.4334	49.6476	0.016 71	0.991 695
SVR (predicting σ_0)	46.7781	55.5777	0.010 23	0.994 893
MLP (predicting σ_0)	55.7733	44.2230	0.013 41	0.995 435
MLP (predicting $\mathbf{y}$)	67.6684	51.8170	0.020 15	0.994 811

Table 7.5: Resulting scores for each model after the second round of simulations. This time the best models are accurate to within $\sim 70\mu m$, or 1.3-2.3% of the original beam width.

In this second batch of simulations the data was binned into 55μm bins to better match the acquisition system of the Timepix3 detectors used by the CERN PS-BGI-82, and the number of particles per profile was reduced. However, varia-

Method	$D_M(\mu)$	$D_{E2}(\mu)$	$D_{KS}(\mu)$	$D_P(\mu)$
0.2T Magnet (95 profiles)	175.8747	11.1870	2.1382	0.999 948
Linear Regression	601.2416	34.2943	4.9063	0.999 580
KRR	389.6027	22.7574	3.6003	0.999 817
MLP	337.8616	19.8724	3.3012	0.999 853

Table 7.6: Distance metrics between true and predicted profiles for each ML method after the second round of simulations. Again the MLP is the best performer of the models, however in this case the 0.2T magnet has overtaken all models.

tions in beam shape are not considered, as only Gaussian beams were simulated.

One variable the models were found to be sensitive to is the number of particles per μm. Having profiles with a lower number of total particles introduces a form of stochastic noise to the resulting profiles, something that machine learning has been shown to struggle with [221]. When simulating 3 million particles, the resulting profiles are very smooth, and this is not an issue; however, training on profiles with 5000 particles in total introduced a significant drop in the resulting accuracies.

Results similar to these were obtained in [194], where the authors attempted to predict the beam width of a simulation of a typical LHC beam, and [195], where the full profile was predicted, both with similar levels of accuracy as the results presented here.

Ideally these models would be used to predict the true profiles of real data recorded by the PS-BGI-82 however, both of these simulations were not calibrated well to the true beam in the PS, and as such, the model was unable to accurately predict the original beam width of recorded data when trained on these simulated datasets. Instead, training on the real data itself yielded slightly better results.

7.7 Using Recorded Data

Measurements were performed on the horizontal PS-BGI-82 on the 5th of May 2022 over a period of 8 hours when a total of 171,540 profiles were recorded, each with ∼5000 counts per profile, while the magnet ranged from 0.2-0.01T. A summary of the chosen currents and the number of profiles recorded can be seen in Table 7.7.

Current	Magnetic Field	Number of Profiles Recorded
176.0	0.2T	44370
88.0	0.1T	29070
44.0	0.05T	24480
35.0	0.04T	32850
26.0	0.03T	35550
18.0	0.02T	4950
9.0	0.01T	270

Table 7.7: Chosen currents and the number of profiles recorded at each current for the day of operation.

Unfortunately, as the PS-BGI-82 is the only horizontal IPM currently installed in the PS, obtaining profiles at lower magnetic field strengths and recording the true profile during ionisation simultaneously is impossible, meaning this dataset is not fully labelled. In order to obtain a more ideal dataset one would need two horizontal BGIs or another instrument capable of recording the beam width at a similar rate to that of the BGI.

Instead, by exploiting the stability of the beam while cycling and measuring the profile with and without the magnet it is possible to roughly match profiles to the true width by their place within the *cycle*. As the beam circulates in the ring, many profiles can be collected, each time the beam orbits the ring is known as a *turn*, and the entire duration the beam is in the ring is known as a cycle. This does make it impossible to match individual profiles to their true width; however, by taking an average of each cycle's width at 0.2T, we can find the mean beam width of that profile in relation to its position in the cycle and map each profile to that average expected true width. This converts our true regression problem into a semi-classification problem, instead now mapping the distorted profile to the width we would expect given its position in the cycle. However, in each case we still trained regression models for this problem. The training data is shuffled, and information on the profile's place in the cycle is not fed to the model so we are still attempting to build a model that can reconstruct beam widths of any recorded profile. This dataset is clearly imperfect, many assumptions about the beam being delivered are made; but this dataset is limited by available equipment. Figure 7.4 shows the mean beam width at all 90 positions in the cycle, as well as the standard deviation of the sample for each position; two example profiles with the magnet at 0.2 and 0.03T are shown for reference.

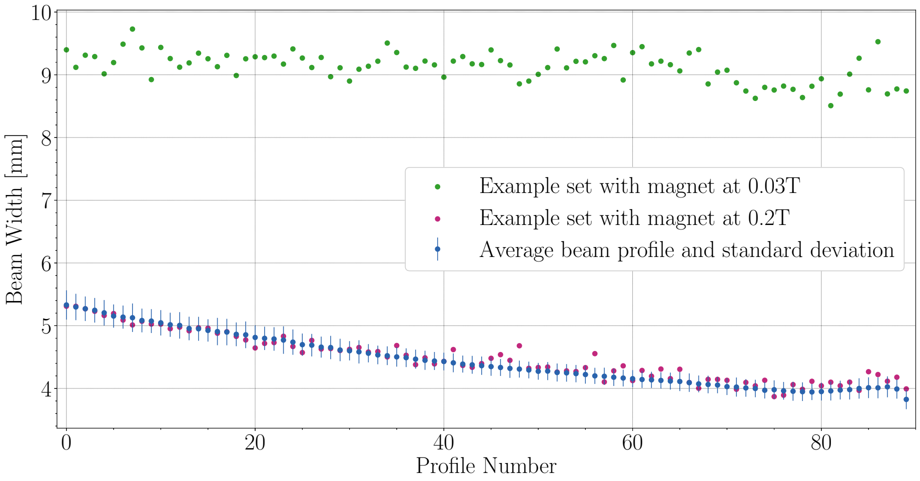


Figure 7.4: Plot of the mean and standard deviation band of the beam width for the given profile in the cycle. An example of the profile widths throughout one super cycle with the magnet operating at 0.2T and 0.03T are shown in purple and green respectively.

The mean beam widths are the target data and for each point, we have a full profile recorded. In this example we would pair distorted profiles with the matching width expected from the cycle and feed that data into the ML training process.

As mentioned, this is an imperfect dataset, acting only as a proof of concept, not a fully complete novel reconstruction method. The profiles are binned in 1024 55μm bins matching the 4 Timepix3 detectors installed on the BGI and as in the second batch of simulations.

The results of training models in this way can be seen in Table 7.8.

Method	Residual Mean (μm)	Residual Sigma (μm)	NMSE	Pearson's R^2
Linear Regression	31.939	27.162	0.0102	0.994866
KRR	29.037	26.070	0.0089	0.995585
SVR	21.467	22.718	0.0057	0.997152
MLP	22.794	15.062	0.0043	0.999017

Table 7.8: Resulting metrics for each model configuration predicting σ_0 using data recorded from the CERN PS-BGI-82.

Again we see good agreement between predictions and labels, however in this case the labels for this dataset are not exact. By looking at the %RMS (also known as the coefficient of variation - and defined as μ/σ) we aim to quantify exactly how much they differ, and can see that over the 90 cycles this value ranges from 1.53-2.30% and has a mean value of 1.82%.

The residuals reported here are slightly smaller than those recorded in the comparative Table 7.5, likely owing to the averaging method used to create the target values, resulting in a lower spread of final predictions. Logically this makes sense, and we might assume that given a more accurate dataset, this reconstruction method would reproduce results more similar to those found in Table 7.5.

Figure 7.5 shows each reconstruction method's target and predicted beam width. The banding in the x-axis shows the discrete nature of the targets thanks to the average method used to generate labels.

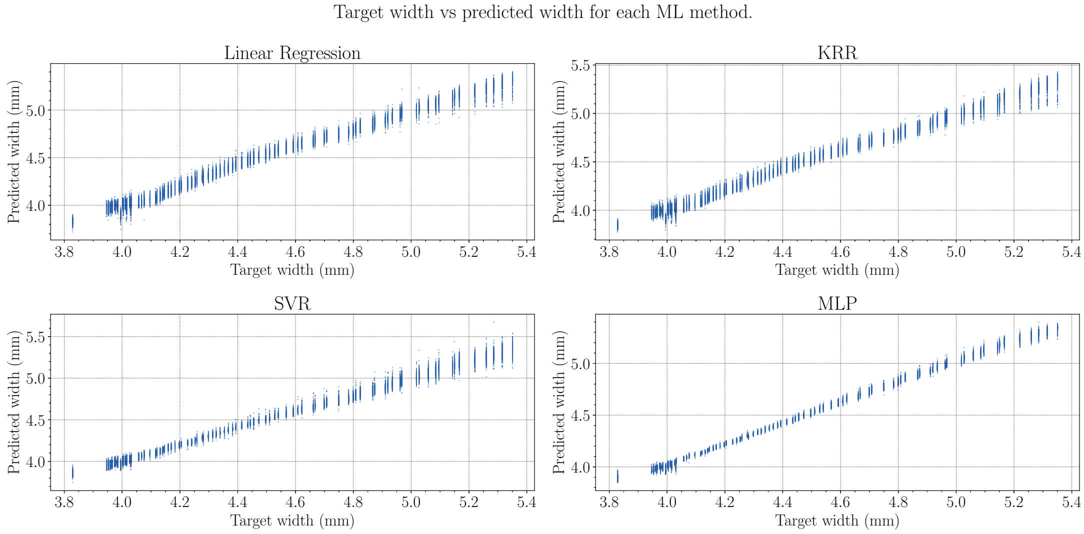


Figure 7.5: Plot of target vs predicted width for the various models trained. The bands in x shows how the targets are discretised by the averaging method.

Applying models trained on simulated data to this dataset did not yield good results. This is possibly due to the stochastic noise effect (so far, all simulated profiles contained about 3 million particles, while these real ones only contain 5000), which the simulated models had not seen before, or potentially due to the physical differences between simulation and reality, specifically the varying shape of the beams. In an attempt to correct for this one final round of simulations was run, specifically tuned to better match these profiles recorded from the BGI.

7.7.1 Tuning Simulations

After collecting data from the PS-BGI-82, simulations can be tuned to better match the data recorded.

After scanning the parameter space it was found that the beam parameters reported in Table 7.9 were the best at reproducing the profiles seen in the PS-BGI-82.

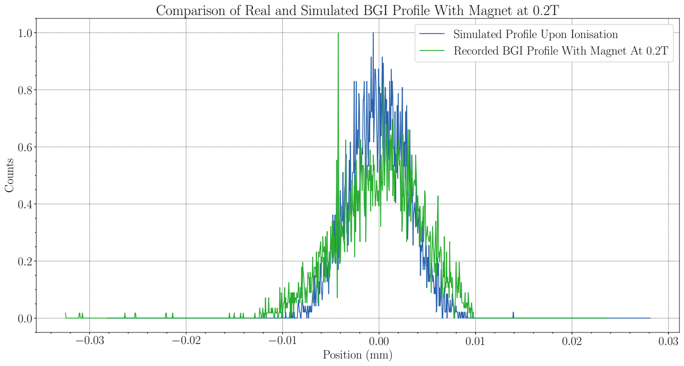
Variable	Value Simulated
Energy	20GeV [217]
Beam Width	3000-5500μm
Guiding Magnetic Field Strength	0.02-0.03T
Bunch population	8e12 [217]
Total Particles simulated per profile (ppp)	5,000
Beam Height	$4000\mu m$
Beam Length	2ns
Total Profiles simulated	10,000
Particles per μm	$\sim$1.6

Table 7.9: Table showing beam parameters simulated in VIPM for the third and final round of simulations, tuned to better match the profiles recorded in the PS-BGI-82. Parameters not included in this table are the same as those found in Table 7.1. The most noticeable difference is the particles per μm, which is greatly reduced in these simulations.

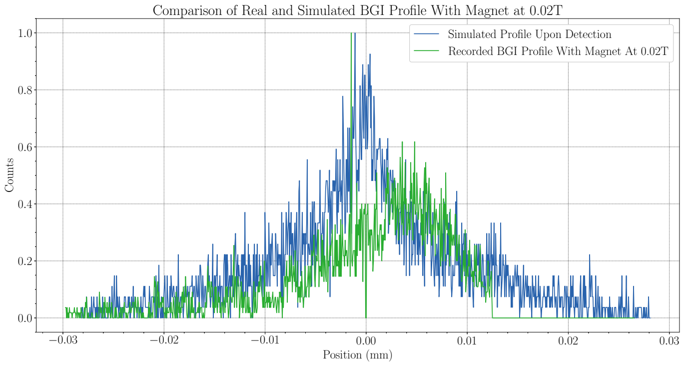
Plots comparing this simulated beam with the one recorded by the PS-BGI-82 at 0.2T and 0.02T can be found in Figure 7.6a, and Figure 7.6b respectively.

The results, when training and testing on this final simulated dataset for each of the different ML methods, are summarised in Table 7.10 and Table 7.11 respectively.

There is a considerable drop in the accuracy of the models after this change, likely as a result of the stochastic noise introduced into the profiles by dropping the number of particles per profile. Further, attempting to apply the models trained with this final simulated dataset to the real profiles recorded showed low accuracy. Not having a truly accurate dataset with which to test these models on makes it difficult to truly grade the precision accurately. As such, this is left as a project for when a better dataset can be obtained.



(a) An example of a profile recorded from the PS-BGI-82 with the magnet operating at 0.2T, compared to a simulated profile.



(b) An example of a profile recorded from the PS-BGI-82 with the magnet operating at 0.02T, compared to a simulated profile.

Figure 7.6: Figure comparing the simulated beam with the real recorded beam after tuning the simulation to match the recorded data better. In these plots the BGI profiles have been shifted to match the mean of the simulated profile ($x \to x - \mu$), and counts are normalised to the range $[0, 1]$. This image shows good agreement between simulation and reality, especially regarding width. However, we also note that the maximum height of the profiles varies; the simulation generally shows more counts in the centre of the profile, while the real data shows more counts in the fringes. We also see another challenge of collecting data using the PS-BGI-82 that is not discussed in Figure 7.6b, where the last quarter of the profile is not being read out due to a heating issue on one of the 4 Timepix3 detectors in the readout system. Profiles like this need to be discarded prior to the training process.

7.8 Improvements

As mentioned, this method of averaging the beam width over the cycle and using that as a target variable (shown in Figure 7.4) is not ideal. It would be preferable to have two operating BGIs capable of running at varying magnetic fields, or at

Method	Residual Mean (μm)	Residual Sigma (μm)	NMSE	Pearson's R^2
LR (predicting σ_0)	490.39	374.70	0.735	0.519
LR (predicting $\mathbf{y}$)	493.55	369.67	0.734	0.517
KRR (predicting σ_0)	358.32	284.21	0.404	0.760
KRR (predicting $\mathbf{y}$)	367.53	284.01	0.416	0.751
SVR (predicting σ_0)	278.10	275.79	0.296	0.840
MLP (predicting σ_0)	174.18	185.56	0.125	0.933
MLP (predicting $\mathbf{y}$)	177.13	199.12	0.137	0.928

Table 7.10: Resulting scores for each model after the third and final round of simulations, specifically tuned to match the beam seen in the PS-BGI-82. The best models now have a mean residual of $\sim 175\mu m$ or $\sim 3-6\%$ of the simulated beam width.

Method	$D_M(\mu)$	$D_{E2}(\mu)$	$D_{KS}(\mu)$	$D_P(\mu)$
0.2T Magnet (1000 profiles)	5080.0945	332.5328	63.7164	0.956 580
Linear Regression	13 813.3001	1074.0703	318.4035	0.812 246
KRR	11 595.8992	864.8491	234.9934	0.860 495
MLP	4045.8963	263.2566	52.9423	0.973 416

Table 7.11: Distance metrics between true and predicted profiles for each ML method after the third and final round of simulations, specifically tuned to match the beam seen in the PS-BGI-82. This time, only an MLP could surpass the accuracy of the magnet.

the very least another device capable of measuring the beam width in the PS. This setup would allow the collection of a training dataset much more similar to the simulated data, upon which the results were generally quite accurate.

Further, it would be advantageous to spend more time simulating the conditions found in the PS-BGI using VIPM and to better generalise the models to the problem. It has been shown that using simulated (or *synthetic*) data to train models that apply to real-world problems can be a viable option [222], however in general, simulations excel when one can simulate the behaviours of the system in question with high fidelity and map the behaviour to the output for prediction; ML excels in the case where we lack a quality simulation but have a lot of data to create a general "black-box" solution. Usually, we either have one or the other but not both. Newer so-called "hybrid" solutions that attempt to merge both approaches do exist [223] but are not always applicable and were not explored due to time constraints.

Another desirable trait would be to simulate different shapes of beams to feed into the models. All the beams simulated were Gaussian in shape, and as such, the models had not been trained to reconstruct beams with different shapes which might significantly drop their accuracy when applied to real data, as the beam is not always Gaussian.

7.9 Conclusion

We have been able to reproduce results shown by Vilsmeier et al. [195, 216] and demonstrate that for simulated beams of varying size, ML methods are generally quite capable of reconstructing the original beam profile and width from distorted profiles. We also took the first steps to applying these models on a real IPM, the CERN PS-BGI-82. Although the hardware limits the data set we were able to obtain; a proof-of-concept solution was deployed by mapping each profile to an expected beam width resulting from the stability of the beam and the profile's placement within the cycle. The %RMS of the targets in this case is around 1.5-2.3%; and despite this, the models show good levels of prediction ability, showing promise for future applications. While these models are not ones that should be deployed in their current state, the potential of the models trained in this way should act as a proof of concept that the ideas previously shown to work on simulated beams could translate over to real data if a more ideal dataset can be obtained. To continue this study, it is necessary to collect a dataset where both the true and distorted profiles are known. Doing so would require another IPM or profile width monitor to be installed into the PS. Alternatively, any other ring that already contains multiple beam width monitors is an ideal candidate to continue this study.

Chapter 8 Machine Learning for ALPHA

The methods described in Chapter 6 and specifically Sec. 6.3 could be useful to the ALPHA experiment in various ways. For example, the reconstruction algorithms (described in Chapter 4) can be slow, and where there is a need for real time information during operation this can be costly. The reconstruction algorithm for the ALPHA-g detector can be up to 20 times slower than real time making dynamic debugging during operation difficult, often forcing people to wait many hours prior to seeing the results of a given experiment. This can be particularly tedious during parameter sweeps. However, it may be possible to use a regression model capable of predicting the vertex position from raw detector outputs to "skip" the reconstruction step entirely and instead use this temporary result to tune on, only using the fully reconstructed results when necessary. While attempts at that have been made, nothing is yet accurate enough to be presented here.

Instead, we turn to the more vital use for ML at ALPHA: background event rejection. As mentioned previously the ALPHA detectors are sensitive to the passage of charged particles through their volume, however, only events resulting from antimatter annihilations are of interest, and other events arising from various sources are not useful for physics. There are a number of background processes that can look similar to antihydrogen events all of which can obscure results; the largest and most troublesome source of background events by far are cosmic ray events, which are due to the constant flux of atmospheric muons.

8.1 Cosmic Ray Background

Cosmic rays are high energy particles originating from outside of Earth. As these rays impact the Earth's atmosphere they can generate a shower of secondary particles, some of which reach the surface (the majority are deflected by the atmosphere or the Earth's magnetic field).

When one of these charged particles or secondaries passes through the detectors of the ALPHA experiment, we are, without further analysis, unable to distinguish between these events and the antihydrogen events the detectors were designed to reconstruct. In experiments where sometimes < 50 antihydrogen atoms are available at a time, these background events can result in errors on the physics results, or at least reduce the statistical strength attributed to them.

The flux of these rays on the earth's surface is quite well understood: they are predominantly muons, have a mean energy of 4GeV, a mean rate of $\sim 1/\text{cm}^2/60s$, and the angle of their incident is represented by a $\cos^2$ distribution [144]. As the ALPHA detectors lay on the surface, have minimal shielding, and are relatively large this mean rate represents a potentially huge source of background for the ALPHA experiment. Further, since these events generally result in two tracks passing directly through the detector (which is perfectly valid for the reconstruction algorithm to process), a vertex will often be generated despite there being no true annihilation vertex.

Running the detectors during periods where no antiparticles in the machine (known as a *cosmic run*) makes it possible to run the same detection and reconstruction algorithms described in Chapter 4, giving information about the rates, angle of incident, and topology of the cosmic background flux. The preferred choice of background rejection deployed by ALPHA is some form of multi-variate analysis (MVA), either rectangular cuts, or an ML binary classifier. The rectangular cuts used for general online analysis within the SVD are described in detail in [153] and will not be the further discussed in this thesis. Instead, ML models designed to improve on these cuts for offline analysis will be explored.

8.1.1 Other Background Sources

As mentioned, there are other sources of background that result in trigger and readout of the detector systems. It is important to note that in this context *background* refers to any event causing the triggering and readout of the detector, that does not come as a result of the specific experiment intended, which can include other antiparticles. *Signal* refers to signals causing trigger and readout coming as a direct result of the experiments conducted on the specific antimatter sample. Examples of other sources of background include:

1. Detector noise causing a trigger readout. Tuning hardware triggers, and thresholds may prevent these types of background from occurring; however

in general these events seem rare, and do not generally result in a vertex.

2. $\bar{\mathrm{H}}$ annihilating on gas. This is discussed in more detail in Sec. 9.1.1, and can become an issue for specific experiments that cause large volumes of gas to be evaporated from the electrode surface where they are frozen (i.e. microwave exposure).

3. Other beam operations at CERN such as injection into the AD, into ELENA, or into ALPHA itself. This source can be easily cleared entirely by cutting out all time windows where one of these events occurred for which we receive various trigger signals.

In general however, the cosmic ray flux is dominant and as such the ML models described in this chapter are specifically tailored to address this problem alone.

8.2 Selection Variables

Many different features or *selection variables* for distinguishing between signal and background events have been trialled, and ultimately used, in both these MVA methods. Most of which have been specifically designed to probe the topological differences between antimatter annihilations, and cosmic rays (for example the aforementioned angle of indecent, or the *time of flight* discussed in Sec. 4.4.2) resulting in a group of robust and accurate selection variables to classify events on.

Any cosmic rays that just graze the detector will not produce a vertex and can be rejected quickly. However, when a cosmic ray passes through the centre of the detector a vertex is often reconstructed. Fortunately, these distinctions in their topology are often noticeable enough to classify events on an event-by-event basis with $\sim 96\%$ accuracy.

The variables chosen for this analysis differ slightly based on the experiment (and therefore detector), however the following variables are applicable to both detectors:

1. The number of hits: N_{hits}, either on the SVD for ALPHA-2, or on the pads for ALPHA-g.

2. The number of reconstructed/candidate tracks: N_{CT}.
3. The number of used/good tracks: N_{GT}. (For the distinction of candidate and good tracks, see Sec. 4.2.1).
4. The ϕ component of the generalized sphericity tensor [224] [225] eigenvector: $\vec{\mathbf{v}}_{\phi}$. The generalized sphericity tensor represents the relative distribution of particles in momentum space, and is often used to try to select events with certain topologies. It is claimed that an event shape observable like transverse sphericity is able to discriminate between jet-like events (cosmics passing through the detector) and events that originate from within the detector [226]. Detailed description of the calculation of this tensor can be found in Appendix C.
5. The z component of the generalized sphericity tensor eigenvector: $\vec{\mathbf{v}}_{z}$.
6. The combination of the two largest eigenvalues of the generalized sphericity tensor: $\sqrt{\lambda_1 + \lambda_2}$.
7. The minimum radius of all helices: r_{min}.
8. The mean radius of all helices: r_{mean}.
9. The minimum pitch of all helices: λ_{min}.
10. The mean pitch of all helices: λ_{mean}. (For details about radius and pitch of helices in canonical form see Appendix A).
11. An integer sum of the sense of curvature of all tracks in the event: $c_{\pm}$. ($+1$ for right curved tracks, and -1 for left curved tracks.)

We also define the following variables specific to the ALPHA-2 models:

1. The squared residual δ. For events with more than two helices we take each pair of helices and fit a straight line to the six hits. After we have found the set of six hits with the highest correlation coefficient, we then find the sum of the residuals squared, where here the residual is defined as the perpendicular distance from the point to the line. This final residual is known as the

squared residual and denoted δ. While technically possible in ALPHA-g the number of spacepoints involved in each helix made this variable error-prone and therefore was abandoned for better performing variables.

2. The mean distance of closest approach between pairs of tracks: D. (Calculated from Equation 4.8). Again, though technically possible in ALPHA-g, this variable was left out for better performing variables.

3. The r position of the resulting vertex. Both this and the ϕ position of the vertex was left out of the ALPHA-g models to ensure no z bias was introduced before these variables could be studied in more detail.

4. A 72×4 matrix: Σ_r representing the sum of signals recorded in each ASIC (4) of each module (72).

These 14 high level variables (excluding Σ_r) are the standard for ML classification in ALPHA-2 and have been used in various combinations to create models for the detectors before. Each of these high level variables require the full reconstruction algorithm to be computed for each event. The additional low level variable, Σ_r, is the raw detector output representing the signal on all 288 ASICs during an event. Use of this variable is limited only to NN architecture, and will not be used in the standard classification models used to analyse physics results. However, the fact that this input would skip entirely the reconstruction step makes it a desirable input if an accurate model can be constructed using these variables. Details of using this variable to build a model can be found in Chapter 10.

Finally, for ALPHA-g we include the following variables, generally only applicable to the TPC and BV:

1. The vertex status: v. An integer describing an overall quantitative value of the quality of the vertex fit. The value is based on specific markers such as the error on the fits on the individual helices, the number of used helices etc.

2. The number of space points: N_{sp}. This is the number of space points reconstructed from the time drift walk out of the hits on the pads (see Sec. 4.4.3).

3. The number of wire hits: N_{AW}. This is the number of hits recorded on the anode field wires surrounding the detector (see Sec. 4.4.3).

4. The mean number of points per track: μ_{ppt}. This is the mean number of space points used to construct the tracks used for vertexing.
5. The mean χ_r^2 of all tracks used for vertexing: μ_{χ}. χ_r^2 is defined in Equation A.4.
6. The number of BV bars hit in the event: N_{bars}.
7. The number of ADCs registering hits in the event: N_{ADC}.
8. The number of end hits in the event: N_{ends}. This is defined as individual BV bar ends registering a hit (as they are read out from both ends, and not every event causes signal on both ends).
9. The number of clusters of two or more consecutive bars in the BV, when reading out each end of the bar separately: $N_{\text{half}-\text{bar}-\text{clusters}}$.

The resulting twenty high-level variables are used for classification in ALPHA-g data during the 2022 experimental run. Due to time constraints the BV calibration was incomplete at the time of the 2022 experimental run, and as such the time of flight variables were not used in the resulting model. While this section will focus on the model used in [17] the BV calibration has since been completed, and the improvements to the resulting model are good but not dramatic.

8.3 Training Data

As discussed in Sec. 6.1 in order to perform any supervised ML training we require a set of events, each labelled as either signal or background. To obtain a sample of events for the background data set we operate the detector and collect data when there are no antiparticles in the machine. With no antiparticles present during these runs this dataset will consist of a 100% pure sample of events not occurring as a result of any antihydrogen annihilation, but as a result of cosmic rays or detector noise.

Conversely, the signal sample was collected during the mixing (Sec. 3.12) phase of the experimental cycle, filtered to only include periods of high-rate particle detection. By ensuring the detector is saturated, we are able to maximise the number of events in this sample resulting due to antihydrogen annihilation, and minimise those events as a result of background processes. However, it is inevitable that some of the events in this sample are background events.

In ALPHA-2, given the cosmic trigger rate of 9.603(50)Hz, and that the typical rate seen during these high rate windows is 8495(6)Hz we estimate the cosmic contamination to be 0.1130(6)%.

For ALPHA-g, the cosmic trigger rate is around six times higher at 58.95(6)Hz, and during these high rate windows we see a trigger rate of 39470(40)Hz and therefore estimate the cosmic contamination to be 0.1493(2)%.

Studies of the effect of intentionally adding cosmic events to the signal sample are shown in Sec. 8.7.2, suggesting BDTs are not particularly sensitive to this.

8.4 Training a Binary Classifier for ALPHA-2

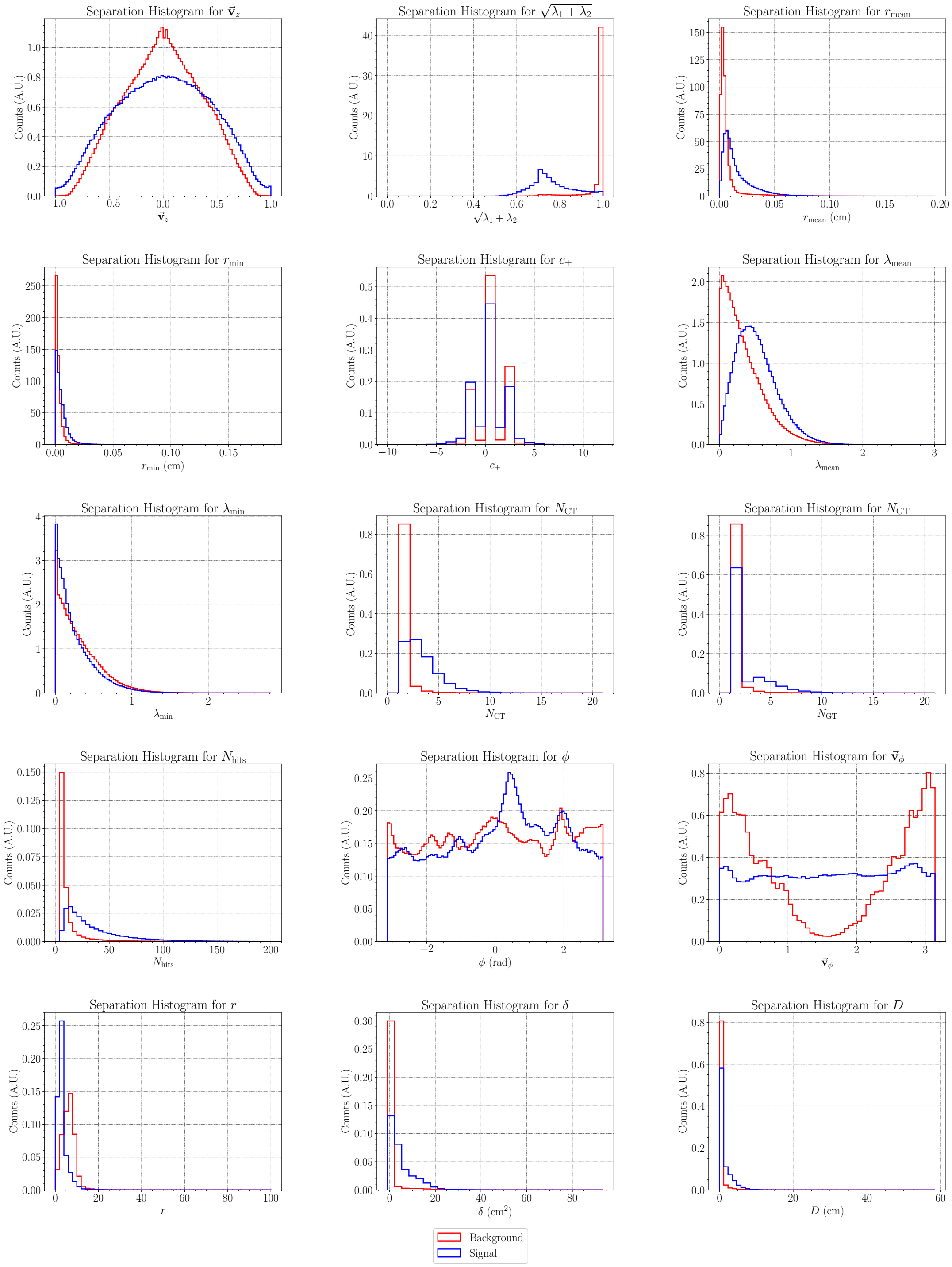
The previously described ML methods and data can be used to train a model for binary classification designed to classify events into signal or background. Understanding the distribution of variables before beginning the training step is important in order to verify and interpret the model in its context, as such we begin with a detailed look at the variables, their distribution and correlations.

8.4.1 A Look at the Selection Variables

Looking at the *separation histogram* of a variable can give insight into whether a chosen variable will have any power in discriminating between signal and background. By plotting the two datasets on the same axis we are able to see how similarly distributed the samples are with respect to the chosen variable. Variables whose histogram shows two very distinct distributions will have a better performance than one where background and signal are similarly distributed.

The separation histograms for all one-dimensional selection variables used in the ALPHA-2 MVA are shown in Figure 8.1.

Another important property of a chosen variable is its correlation or “bias” towards variables that the classifier must be independent of, for example z, and in some cases r. Scatter plots between each variable and both z and r can be seen



Separation Histogram for $\vec{\mathbf{v}}_z$

Separation Histogram for $\sqrt{\lambda_1 + \lambda_2}$

Separation Histogram for r_{mean}

Separation Histogram for r_{min}

Separation Histogram for $c_{\pm}$

Separation Histogram for λ_{mean}

Separation Histogram for λ_{min}

Separation Histogram for N_{CT}

Separation Histogram for N_{GT}

Separation Histogram for N_{hits}

Separation Histogram for ϕ

Separation Histogram for $\vec{\mathbf{v}}_\phi$

Separation Histogram for r

Separation Histogram for δ

Separation Histogram for D

Figure 8.1: The separation histograms for each of fourteen selection variables described in Sec. 8.2 and ϕ. All histograms are normalised. These histograms show that variables such as $\sqrt{\lambda_1 + \lambda_2}$ are good at discriminating between signal and background, while variables such as λ_{min} are less powerful (though still show some separation).

in Figure 8.2.

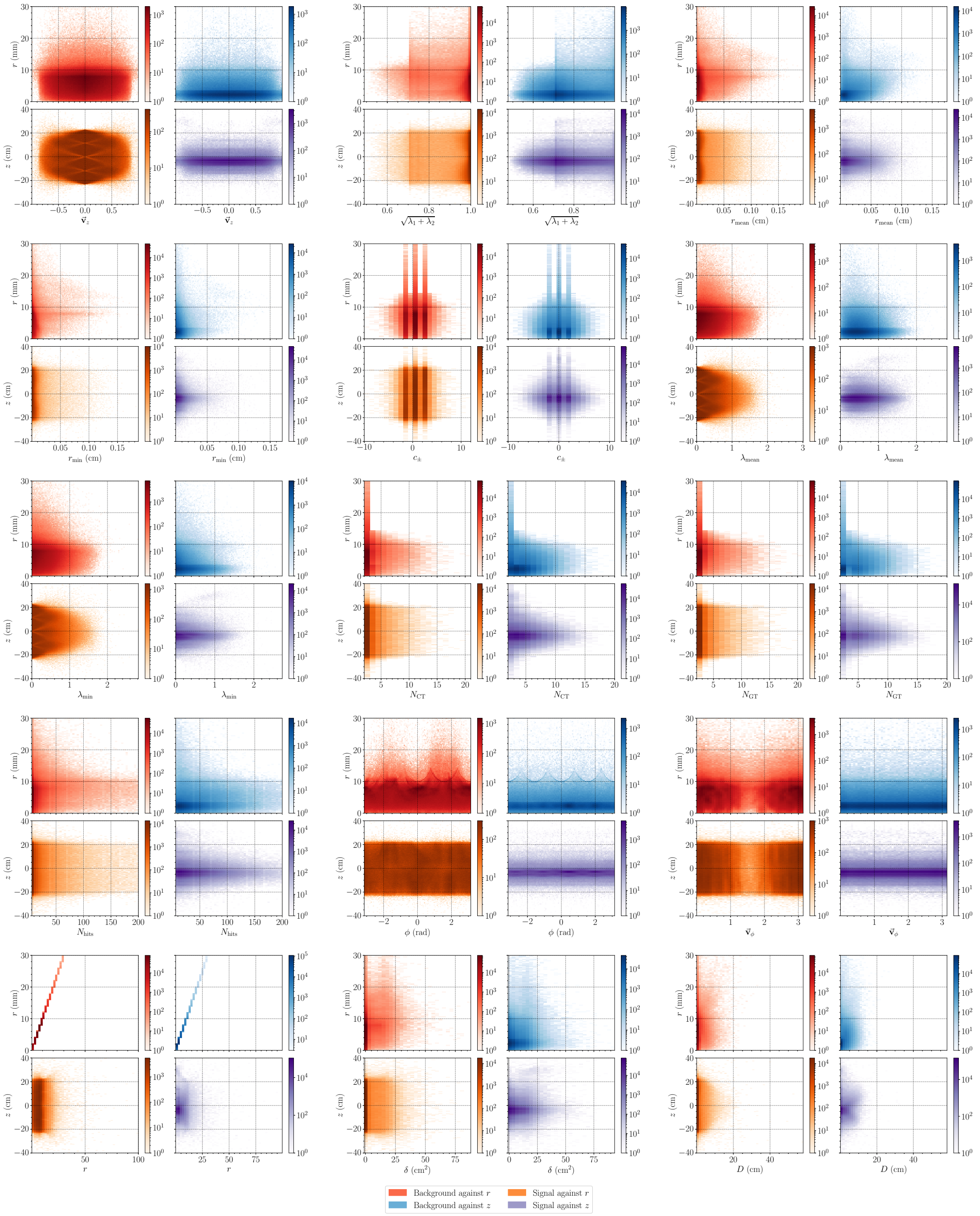


Figure 8.2: 2D histograms of each selection variable against z and r for both signal and background. None of these variables show any significant correlation with z or r. Cuts on r in the range $[0, 30]$, and on z in the range $[-40, 40]$ are applied to improve legibility (by removing large outliers) while retaining the majority (99.975% on average) of samples.

We can quantify these observations by evaluating the *separation* of each variable, and the Pearson correlation coefficient (PCC - denoted ρ[1]) between r, or z and the signal, or background samples. The *separation* $\langle S^2 \rangle$ of a variable v is defined as [227]

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(v_S - v_B)^2}{v_S + v_B} dy \tag{8.1}$$

where y_S and y_B are the PDFs of the variable y in the signal and background samples respectively. The separation is 0 for identical signal and background shapes, and 1 for shapes with no overlap. In cases where the true PDF is unknown the integral in Equation 8.1 can be approximated numerically using histograms of the variable's distribution such as those found in Figure 8.1. A table showing the separation, and both the signal and background samples correlation with z and r are shown in Table 8.1.

Variable	ρ^r_B	ρ^r_S	ρ^z_B	ρ^z_S	$\langle S^2 \rangle$
$\sqrt{\lambda_1 + \lambda_2}$	-0.157	0.151	0.023	-0.001	0.766
δ	0.208	0.155	-0.001	0.039	0.527
N_{hits}	0.131	0.171	-0.005	0.015	0.467
N_{CT}	0.047	-0.002	0.001	0.008	0.454
r	1.000	1.000	0.029	0.049	0.352
r_{mean}	0.136	0.287	-0.020	0.037	0.287
$\vec{\mathbf{v}}_\phi$	-0.012	0.001	0.006	-0.001	0.145
D	0.110	0.373	0.010	0.073	0.135
λ_{mean}	0.030	0.016	-0.010	0.047	0.128
N_{GT}	0.065	0.176	0.001	0.018	0.114
r_{min}	0.025	0.091	-0.027	0.015	0.098
$c_\pm$	0.048	0.008	0.151	0.001	0.051
$\vec{\mathbf{v}}_z$	0.006	-0.002	0.002	0.001	0.025
λ_{min}	-0.000	-0.011	-0.010	0.042	0.013
ϕ	0.052	-0.018	-0.013	0.009	0.009

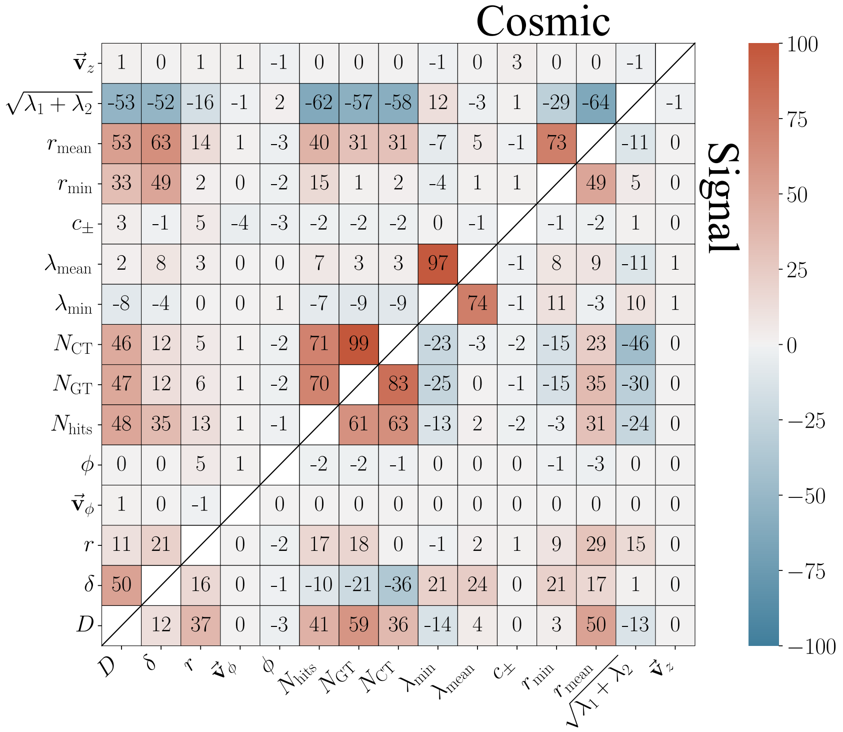
Table 8.1: Table displaying the correlation of each variable with r, and z in both the signal and background samples; and the separation of each variable. $\rho^{r(z)}_{B(S)}$ denotes the correlation of the background (signal) sample with respect to r (z). Table is sorted in descending order of the separation.

The variable Σ_r is multidimensional and therefore not displayed in either plots or table. While this variable has been used in the training of a specific model described in Chapter 10 it will not be discussed in this chapter.

---

[1] Generally, ρ is usually used to denote the correlation of the population, and r is used for samples; however in order to avoid confusion with the selection variables r, r_{min} etc. ρ will be used throughout this chapter.

Further, though not necessarily vital, is a look at the correlation between each variable, this "*correlation matrix*" is good for ensuring that variables do not probe the same aspect of the event. Variables that are entirely orthogonal are preferred. Plots of the correlation matrices for both the signal and background data samples can be found in Figure 8.3.



	D	δ	r	$\vec{\mathbf{v}}_\phi$	ϕ	N_{hits}	N_{GT}	N_{CT}	λ_{min}	λ_{mean}	$c_\pm$	r_{min}	r_{mean}	$\sqrt{\lambda_1+\lambda_2}$	$\vec{\mathbf{v}}_z$
$\vec{\mathbf{v}}_z$	1	0	1	1	-1	0	0	0	-1	0	3	0	0	-1	
$\sqrt{\lambda_1+\lambda_2}$	-53	-52	-16	-1	2	-62	-57	-58	12	-3	1	-29	-64		-1
r_{mean}	53	63	14	1	-3	40	31	31	-7	5	-1	73		-11	0
r_{min}	33	49	2	0	-2	15	1	2	-4	1	1		49	5	0
$c_\pm$	3	-1	5	-4	-3	-2	-2	-2	0	-1		-1	-2	1	0
λ_{mean}	2	8	3	0	0	7	3	3	97		-1	8	9	-11	1
λ_{min}	-8	-4	0	0	1	-7	-9	-9		74	-1	11	-3	10	1
N_{CT}	46	12	5	1	-2	71	99		-23	-3	-2	-15	23	-46	0
N_{GT}	47	12	6	1	-2	70		83	-25	0	-1	-15	35	-30	0
N_{hits}	48	35	13	1	-1		61	63	-13	2	-2	-3	31	-24	0
ϕ	0	0	5	1		-2	-2	-1	0	0	0	-1	-3	0	0
$\vec{\mathbf{v}}_\phi$	1	0	-1		0	0	0	0	0	0	0	0	0	0	0
r	11	21		0	-2	17	18	0	-1	2	1	9	29	15	0
δ	50		16	0	-1	-10	-21	-36	21	24	0	21	17	1	0
D		12	37	0	-3	41	59	36	-14	4	0	3	50	-13	0

Figure 8.3: Correlation matrices between all fourteen high level variables, separated into signal (right) and background (left) datasets.

The plot shows that the chosen variables are generally completely orthogonal, and no two variables are too strongly correlated. The highest correlation exists between N_{CT} and N_{GT}, with 83% and 99% correlation in signal, and background samples respectively. Though this is quite high, and a case could be made for excluding this variable, the slight difference between signal and background does hold some discriminating power and for this reason it was kept. All data was collected using the `a2dumper.exe` (described in Appendix B), a program built to perform the reconstruction algorithm described in Sec. 4.2.1 while saving specific pre-defined variables into a `.root` file (utilising the tree/branch structure that is standard for ROOT I/O [167]) for analysis and model training.

8.4.2 Model Architectures

In order to ensure the best possible model is trained, many different binary classifier architectures were trialled. All models were either trained using ROOT's TMVA package [227], or Python's `scikit-learn` [218] and a table of each trained model's details can be found in Table 8.4.2.

Table 8.2: Table describing the models trained.

Model Type	Model Name	Package	Variables	Description
BDT	BDT_WithR	ROOT::TMVA	All 14 high level variables.	Uses 850 trees, adaptive boost, a maximum depth of 3, a min node size of 2.5%.
KNN	KNN	scikit-learn	All 14 high level variables	Using Euclidean distance with k=20.
SVM	SVM_Linear	scikit-learn	All 14 high level variables	Linear kernel (see Sec. 6.3.3).
SVM	SVM_Poly	scikit-learn	All 14 high level variables	Polynomial kernel of degree 2 (see Sec. 6.3.3).
SVM	SVM_RBF	scikit-learn	All 14 high level variables	RBF kernel (see Sec. 6.3.3).
BDT	BDT_NoR	ROOT::TMVA	All 14 high level variables, except r	Uses 850 trees, adaptive boost, a maximum depth of 3, a min node size of 2.5%.

The parameters of the models themselves (such as number of trees in a BDT, distance metric in KNN) are known as *hyperparameters* and, where computationally feasible, a grid search was performed over these variables to select the best performing hyperparameters for the given model. Only the grid search of the resulting best performer is displayed within this chapter, in Sec. 8.7.1.

In all models the data is normalised to $[-1, 1]$ prior to training, and events which do not reconstruct a vertex are excluded.

8.4.3 Results

Using the various evaluation metrics described in Sec. 6.5.1 we are able to quantitatively compare the performance of each model against one another. Table 8.3 presents the accuracy, F1 score, area under ROC curve, and signal efficiency at 1% background for each of the models described in Table 8.4.2.

Further, the ROC curve for each of the models can be seen in Figure 8.4.

These results show that for classification purposes the BDT, utilising all available high level variables (BDT_WithR) is the best performer.

Model	Accuracy	F1 Score	AUC (ROC)	Signal Efficiency at 1% Background
BDT_WithR	0.9623	0.9469	0.9943	0.8894
KNN	0.9505	0.9296	0.9862	0.7915
Linear SVM	0.9459	0.9218	0.9776	0.7055
Polynomial SVM	0.9526	0.9321	0.9853	0.8039
RBF SVM	0.9541	0.9343	0.9865	0.8228
BDT_NoR	0.9534	0.9345	0.9912	0.8035

Table 8.3: Metrics for each trained model. Details of each of these metrics can be found in Sec. 6.5.1.

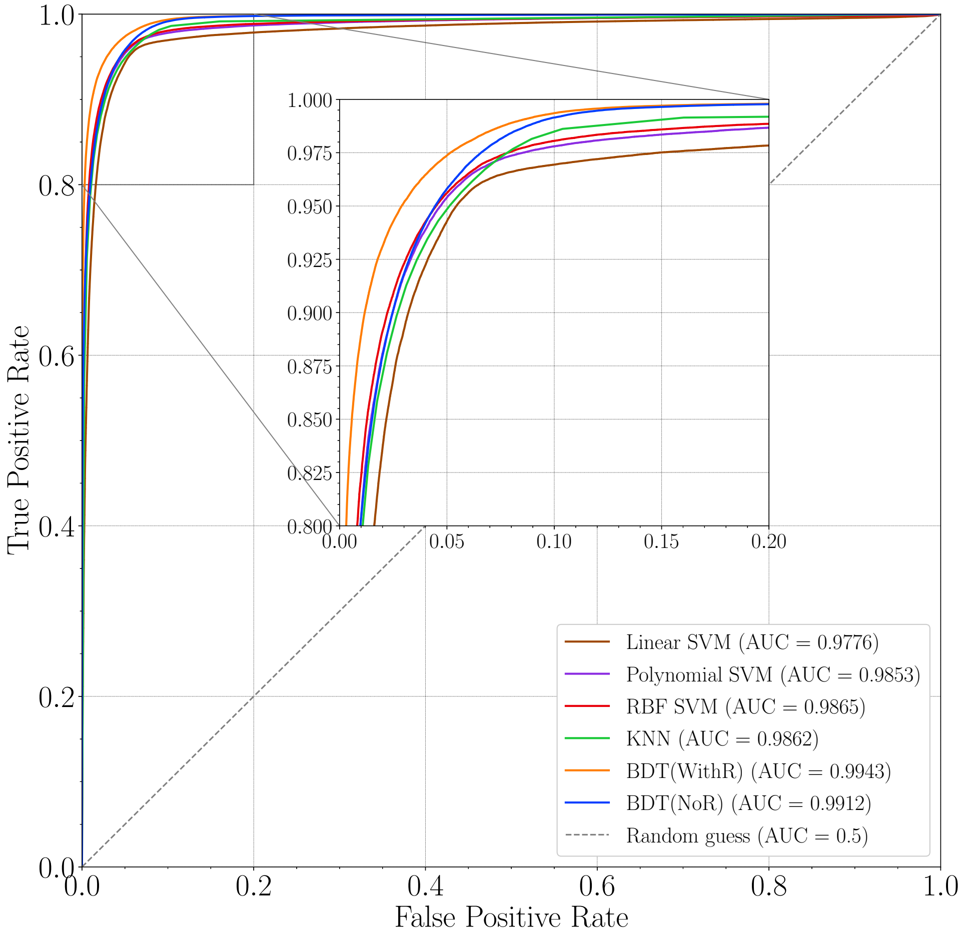
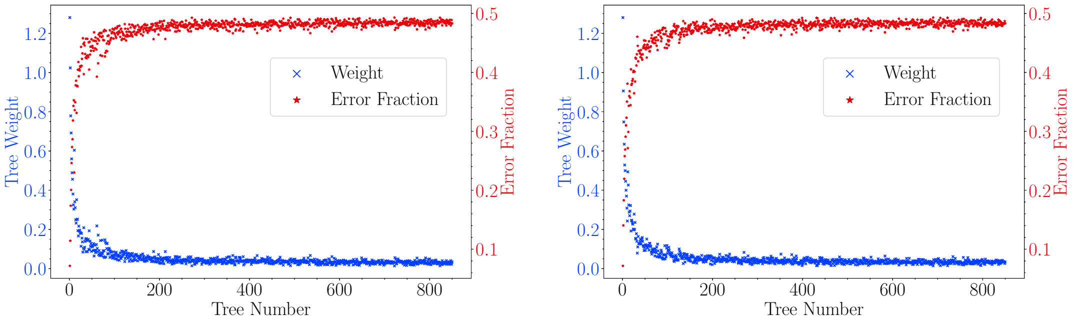


Figure 8.4: ROC curves for each model described in Table 8.4.2.

While the model BDT_NoR is able to keep up with BDT_WithR initially, as the false positive rate (and therefore background rate) decreases, its performance starts to decrease rapidly. This shows that while possible, choosing to exclude

r as a variable does come at a cost. For each of the individual trees that make up the BDTs we can plot the trees final weight, and the error fraction, found in Figure 8.5a and Figure 8.5b for WithR and NoR respectively. These are sometimes referred to as the *control plots* and show the effect of AdaBoost.



(a) The tree weight and error fraction for each of the 850 individual trees making up the WithR model.

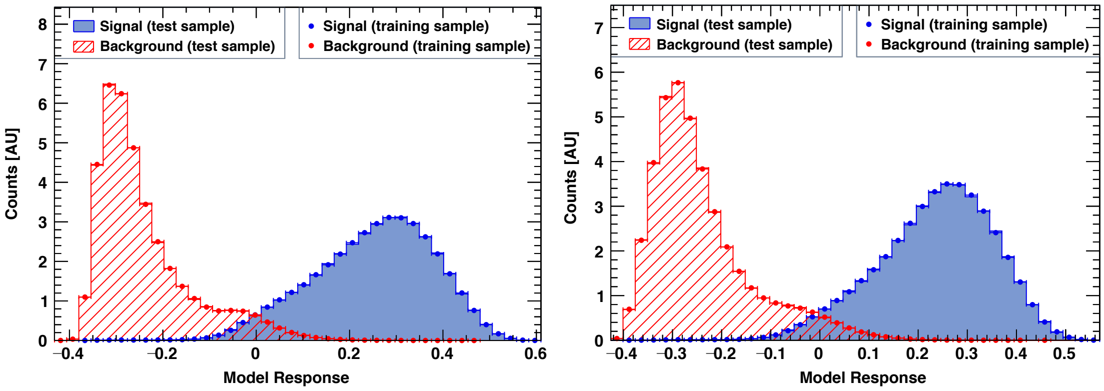
(b) The tree weight and error fraction for each of the 850 individual trees making up the NoR model.

Figure 8.5: Control plots for the BDT_WithR and BDT_NoR models. We can see here that the first tree is roughly 93% accurate alone. However, as each iteration of AdaBoost re-weights the training samples, more weight is put on the events that are "harder" to classify. This results in the error fraction approaching 50% and the tree weight approaching 0. We also see the curve flattening out, which is a suggestion no more trees are needed or else overfitting could occur, and to potentially see if reducing the number of trees is possible.

Further a histogram of the response of both the BDT_WithR and BDT_NoR models can be seen in Figure 8.6a and Figure 8.6b respectively.

In these plots the training and test data set is superimposed to check for overfitting (these two histograms will diverge if the model is over-fit to the training set), and they show good general separation between the populations suggesting they are good performers and show no signs of overfitting.

Overall the models trained using the described methods show a good degree of accuracy and no signs of overfitting or bias. These models have been internally commissioned, and are to be used in the analysis of the multiple papers ALPHA is planning to submit from the 2023/2024 experimental season.



(a) Histogram of WithR model response on both the training and test dataset. These two histograms show good separation, suggesting a strong model.

(b) Histogram of NoR model response on both the training and test dataset. These two histograms show decent separation, though slightly worse than WithR.

Figure 8.6: BDT responses for both the WithR and NoR models.

8.5 Training a Binary Classifier for ALPHA-g

The ALPHA-g experiment requires a binary classifier for signal and background, much like ALPHA-2, and this chapter describes the model that was trained for the 2022 physics run, and initially introduced in [17]. This model marks the first time an ML model designed to operate on the newly commissioned ALPHA-g TPC has been trained and used; and, though the model was used and reviewed for publication [17], additional details beyond those found in the paper are presented here.

8.5.1 A Look at the Selection Variables

As described in Sec. 8.4.1, it is important to look at the separation histograms and correlations of the selection variables prior to training. Of particular importance to models trained for ALPHA-g is the elimination of bias in relation to the detector axis, z. The results presented in [17] depend upon the general z direction of the vertices, and it is therefore important not to introduce a bias in the z direction during the MVA stage of the analysis.

Figure 8.7 shows the separation histograms of all twenty variables used for the model in [17], and Figure 8.8 shows the correlation of these variables with respect to r and z separated into signal and background.

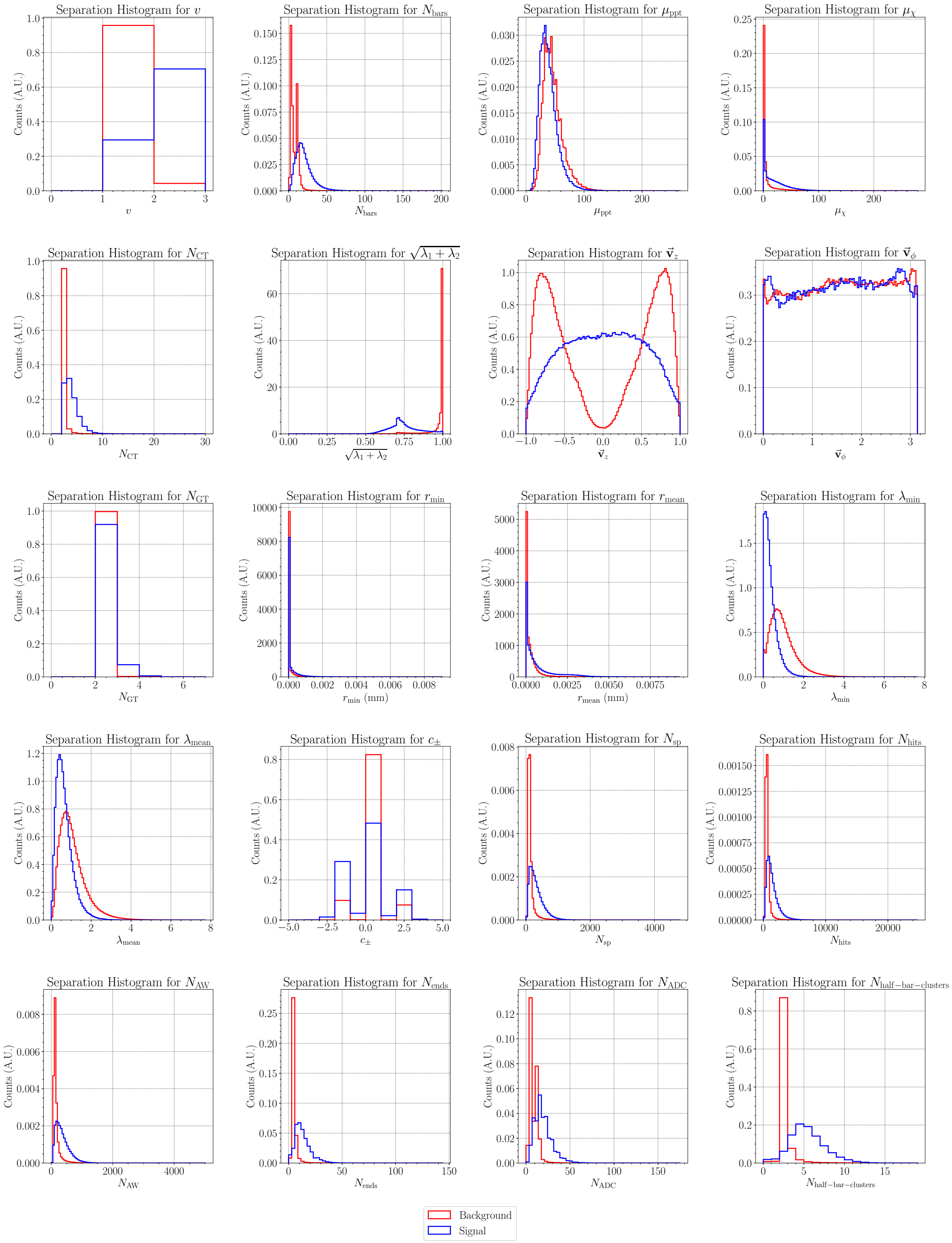


Figure 8.7: The separation histograms for each of twenty selection variables. All histograms are normalised. These histograms show that variables such as $\sqrt{\lambda_1 + \lambda_2}$ are good at discriminating between signal and background, while variables such as λ_{min} are less powerful (though still show some separation).

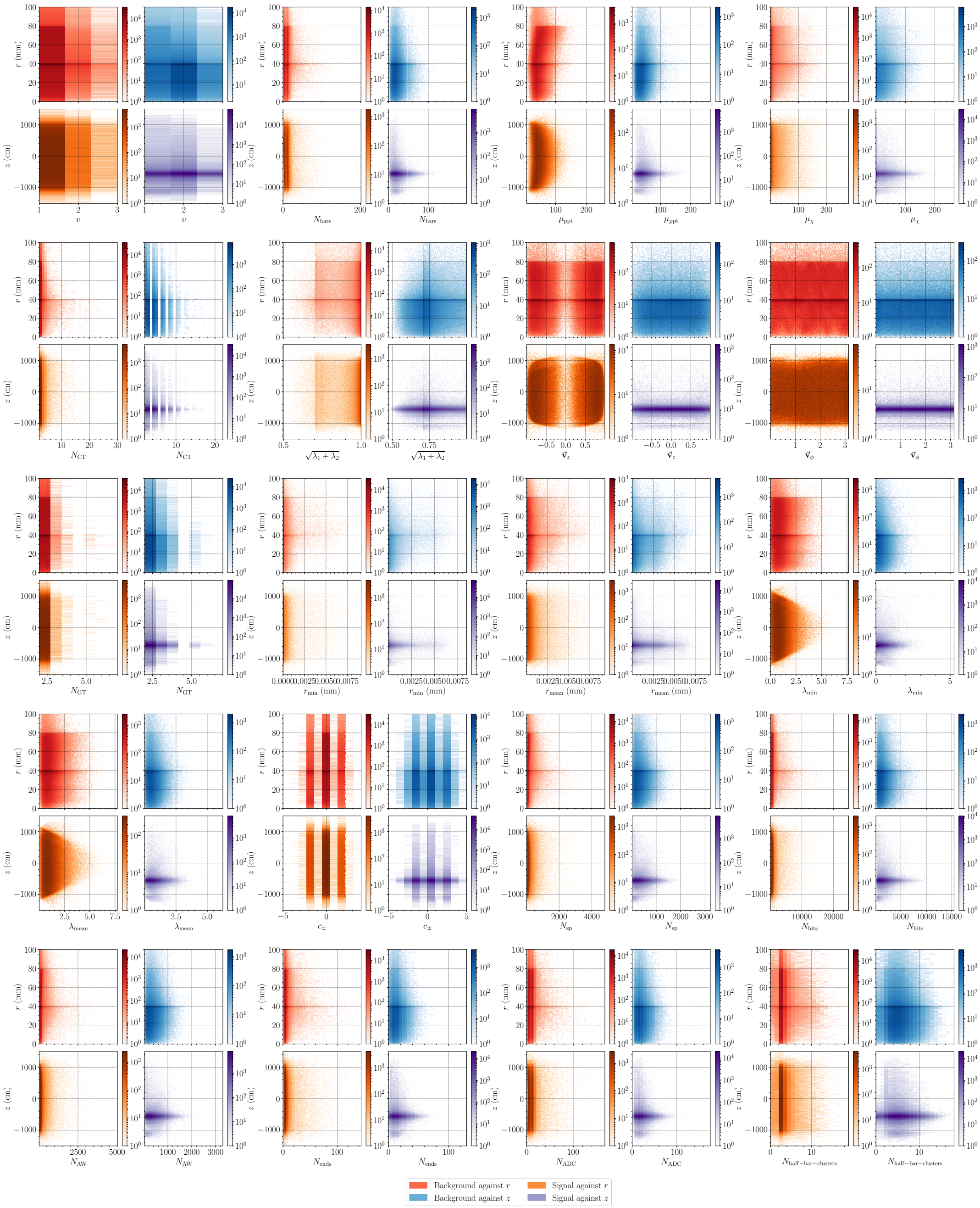


Figure 8.8: 2D histograms of each selection variable against z and r for both signal and background. No variables show significant correlation with z or r. Cuts on r in the range $[0, 100]$, and on z in the range $[-1500, 1500]$ are applied to improve legibility (by removing large outliers) while retaining the majority (99.158% on average) of samples.

The table quantifying these correlations and separations can be found in Table 8.4.

Variable	ρ_B^r	ρ_S^r	ρ_B^z	ρ_S^z	$\langle S^2 \rangle$
$\sqrt{\lambda_1 + \lambda_2}$	-0.003	0.003	-0.014	0.026	0.735
$N_{\text{half-bar-clusters}}$	-0.001	-0.003	-0.012	-0.030	0.691
N_{ends}	-0.001	-0.002	-0.005	-0.013	0.616
N_{bars}	0.001	0.000	-0.042	-0.032	0.496
N_{ADC}	0.001	0.000	-0.012	-0.018	0.489
N_{CT}	-0.001	-0.001	0.010	0.019	0.470
v	-0.001	-0.003	0.009	-0.004	0.469
N_{AW}	-0.001	0.000	0.032	0.026	0.379
N_{hits}	-0.002	-0.001	0.030	0.033	0.342
N_{sp}	-0.001	0.000	0.031	0.030	0.342
λ_{min}	-0.007	-0.006	-0.002	0.056	0.325
μ_χ	-0.001	0.001	0.008	0.002	0.195
$\vec{\mathbf{v}}_z$	-0.000	-0.000	0.031	-0.001	0.190
$c_\pm$	0.002	-0.001	-0.013	-0.004	0.139
λ_{mean}	-0.006	-0.006	0.002	0.039	0.134
r_{mean}	-0.003	-0.003	0.005	0.013	0.099
μ_{ppt}	-0.002	0.001	0.036	0.042	0.063
N_{GT}	-0.000	-0.002	0.004	0.000	0.037
r_{min}	-0.001	-0.001	0.005	0.021	0.037
$\vec{\mathbf{v}}_\phi$	0.000	0.004	0.011	0.001	0.001

Table 8.4: The correlation of each variable with r, and z in both the signal and background samples; and the separation of each variable. $\rho_{B(S)}^{r(z)}$ denotes the correlation of the background (signal) sample with respect to r (z). The table is sorted in descending order of the separation.

Further Figure 8.9 shows the cross correlation between variables, showing a good range of orthogonal variables, unsurprisingly N_{sp}, N_{hits}, N_{AW} are strongly correlated with one another, one or two could be dropped in future models for ALPHA-g, though this should be validated first.

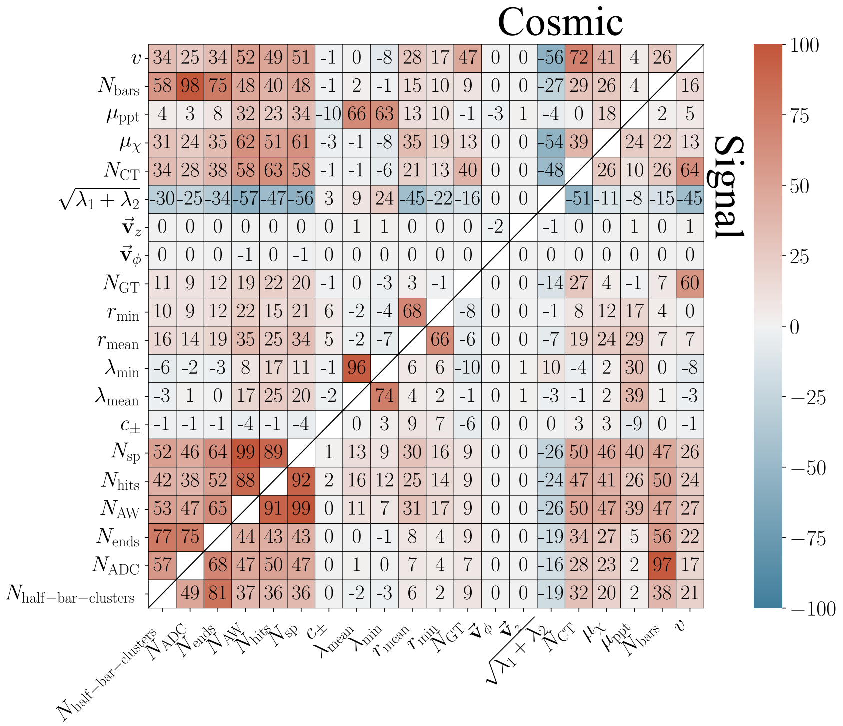
8.5.2 Model Architectures

The models trained for ALPHA-g mimic those trained for ALPHA-2, and a table of the models can be seen in Table 8.5.2.

Table 8.5: Table describing the models trained.

Model Type	Model Name	Package	Variables	Description

Continued on next page



	$N_{\text{half-bar-clusters}}$	N_{ADC}	N_{ends}	N_{AW}	N_{hits}	N_{sp}	$c_\pm$	λ_{mean}	λ_{min}	r_{mean}	r_{min}	N_{GT}	$\vec{\mathbf{v}}_\phi$	$\vec{\mathbf{v}}_z$	$\sqrt{\lambda_1+\lambda_2}$	N_{CT}	μ_χ	μ_{ppt}	N_{bars}	v
v	34	25	34	52	49	51	-1	0	-8	28	17	47	0	0	-56	72	41	4	26	
N_{bars}	58	98	75	48	40	48	-1	2	-1	15	10	9	0	0	-27	29	26	4		16
μ_{ppt}	4	3	8	32	23	34	-10	66	63	13	10	-1	-3	1	-4	0	18		2	5
μ_χ	31	24	35	62	51	61	-3	-1	-8	35	19	13	0	0	-54	39		24	22	13
N_{CT}	34	28	38	58	63	58	-1	-1	-6	21	13	40	0	0	-48		26	10	26	64
$\sqrt{\lambda_1+\lambda_2}$	-30	-25	-34	-57	-47	-56	3	9	24	-45	-22	-16	0	0		-51	-11	-8	-15	-45
$\vec{\mathbf{v}}_z$	0	0	0	0	0	0	0	1	1	0	0	0	-2		-1	0	0	1	0	1
$\vec{\mathbf{v}}_\phi$	0	0	0	-1	0	-1	0	0	0	0	0	0		0	0	0	0	0	0	0
N_{GT}	11	9	12	19	22	20	-1	0	-3	3	-1		0	0	-14	27	4	-1	7	60
r_{min}	10	9	12	22	15	21	6	-2	-4	68		-8	0	0	-1	8	12	17	4	0
r_{mean}	16	14	19	35	25	34	5	-2	-7		66	-6	0	0	-7	19	24	29	7	7
λ_{min}	-6	-2	-3	8	17	11	-1	96		6	6	-10	0	1	10	-4	2	30	0	-8
λ_{mean}	-3	1	0	17	25	20	-2		74	4	2	-1	0	1	-3	-1	2	39	1	-3
$c_\pm$	-1	-1	-1	-4	-1	-4		0	3	9	7	-6	0	0	0	3	3	-9	0	-1
N_{sp}	52	46	64	99	89		1	13	9	30	16	9	0	0	-26	50	46	40	47	26
N_{hits}	42	38	52	88		92	2	16	12	25	14	9	0	0	-24	47	41	26	50	24
N_{AW}	53	47	65		91	99	0	11	7	31	17	9	0	0	-26	50	47	39	47	27
N_{ends}	77	75		44	43	43	0	0	-1	8	4	9	0	0	-19	34	27	5	56	22
N_{ADC}	57		68	47	50	47	0	1	0	7	4	7	0	0	-16	28	23	2	97	17
$N_{\text{half-bar-clusters}}$		49	81	37	36	36	0	-2	-3	6	2	9	0	0	-19	32	20	2	38	21

Figure 8.9: Correlation matrices between all twenty high level variables, separated into background (left) and signal (right) datasets.

Table 8.5: Table describing the models trained. (Continued)

BDT	BDT	ROOT::TMVA	All 20 high level variables.	Uses 850 trees, adaptive boost, a maximum depth of 3, a min node size of 2.5%.
KNN	KNN	scikit-learn	All 20 high level variables	Using Euclidean distance with k=20.
SVM	SVM_Linear	scikit-learn	All 20 high level variables	Linear kernel (see Sec. 6.3.3).
SVM	SVM_Poly	scikit-learn	All 20 high level variables	Polynomial kernel of degree 2 (see Sec. 6.3.3).
SVM	SVM_RBF	scikit-learn	All 20 high level variables	RBF kernel (see Sec. 6.3.3).

In this case all models used all twenty available variables.

8.5.3 Results

Using the previously described evaluation metrics (Sec. 6.5.1) we are able to quantitatively compare the performance of each model against one another. Table 8.6 presents the accuracy, F1 score, area under ROC curve, and signal efficiency at 1% background for each of the models described in Table 8.5.2.

The ROC curves themselves can be seen in Figure 8.10.

The BDT is the best performer here similarly to ALPHA-2, outperforming all

Model	Accuracy	F1 Score	AUC (ROC)	Signal Efficiency at 1% Background
BDT	0.9591	0.9465	0.9912	0.8329
KNN	0.9558	0.9420	0.9865	0.7603
Linear SVM	0.9442	0.9261	0.9808	0.6907
Polynomial SVM	0.9513	0.9362	0.9865	0.7633
RBF SVM	0.9534	0.9390	0.9877	0.7817

Table 8.6: Various metrics for each model trained. Details of each of these metrics can be found in Sec. 6.5.1.

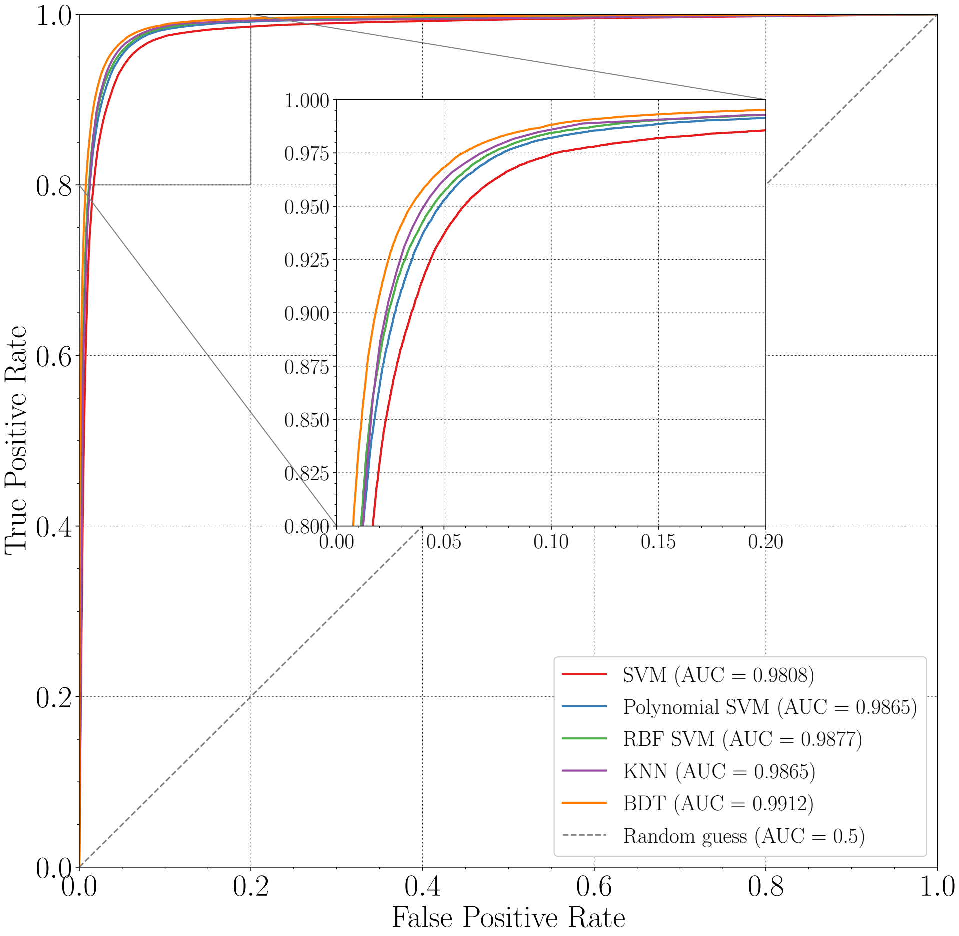
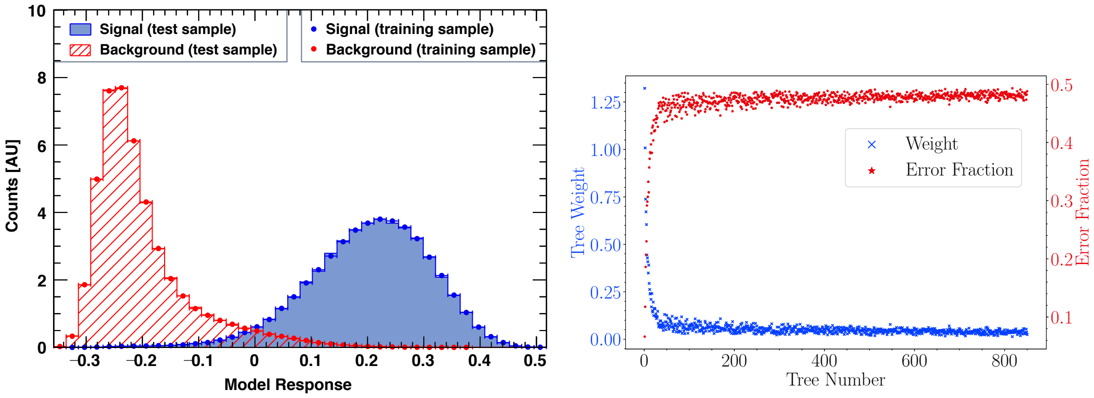


Figure 8.10: ROC curves for each model described in Table 8.5.2.

other models in all regions of the ROC curve (high signal, and low background regions). Figure 8.11a shows the BDT response for the ALPHA-g BDT model, which shows the familiar separated histograms corresponding to an accurate model, and a good agreement in these distributions between training and test data suggest the model was not overfit. The control plots for this model can be seen in Figure 8.11b.



(a) Histogram of the ALPHA-g model's response on both the training and test dataset. These two histograms show good separation, suggesting a strong model.

(b) The tree weight and error fraction for each of the 850 individual trees making up the resulting BDT in the ALPHA-g model.

Figure 8.11: The control plot, and BDT response histogram for the ALPHA-g model.

These results suggest the model is accurate, and capable of filtering events effectively. There is no evidence the model was over fit to the training data or has any bias towards a particular z direction.

8.6 Cuts Placement

Given a model with a variable signal efficiency and background rate, (i.e. any model which has an associated ROC curve), the question of where to place the cut on the curve is as important as choosing the model itself. The signal efficiency should be maximised while keeping the background rate as low as possible, however in some cases a lower background is more important, whereas in others a higher signal is more important. In order to formalise the choice of cut, to eliminate bias in its selection, and to state its sensitivity, an appropriate *figure-of-merit* (FOM) is chosen to be maximised, resulting in the final cut for analysis.

Classically, this is done via *hypothesis testing*, a default hypothesis (*null hypothesis*, denoted H_0), should be accepted or rejected in favour of a new theory H_f (f generally indicates the free parameters of the theory, i.e. laser/microwave frequency).

Consider a simple counting experiment, where we record n counts in a given window and would like to know if the number of counts are a result of the background or some process, such as atom ionisation.

The null hypothesis in this case is that the number of observed counts, n, is consistent with the Poisson distribution of mean B, while H_f is the hypothesis that the number of counts in this window is a Poisson with a mean larger than B, instead $B + S_f$ where S_f is the contribution of signal events as a result of the experiment.

Assuming we know B (by taking a background measurement) we then have

$$p(n|H_0) = \frac{e^{-B}B^n}{n!} \tag{8.2}$$

$$p(n|H_f) = \frac{e^{-(B+S_f)}(B+S_f)^n}{n!} \tag{8.3}$$

and we would like to know the minimum number of counts we need (n_{min}) in order to accept H_f over H_0, and the associated *significance*, α, of said test where α is the probability of falsely rejecting H_0 when it is true.

Given that the number of signal and background events are functions of the chosen cut c, that is $S = S(c)$, and $B = B(c)$, then the optimisation of the cut is generally done with one of the following two FOMs:

$$\frac{S(c)}{\sqrt{B(c)}} \tag{8.4}$$

also known as the *discovery* FOM, and

$$\frac{S(c)}{\sqrt{B(c)+S(c)}} \tag{8.5}$$

also known as the *measurement* FOM. These equations are considered “significance-like” [228] as they do not exactly measure the significance but more so describe the relative uncertainty.

It can also be seen that though Equation 8.4 is independent of the cross-section of the process (unlike Equation 8.5) it breaks down as $B \to 0$ (note that it would favour a cut that accepts 0.1 signal events to 10^{-5} background over 10 signal to 1 background).

As such, following Punzi [228] we can derive a different FOM (known now as the *Punzi FOM*) which is independent of these issues.

By making a Gaussian approximation of the Poisson distribution, we see that the optimum cut is found by maximising the equation

$$\frac{\epsilon(c)}{a/2 + \sqrt{B(c)}} \tag{8.6}$$

where $\epsilon(c)$ is the signal efficiency of the cut c, and a is the number of sigma corresponding to a two-sided Gaussian test of significance α. This FOM is independent of the cross-section of the process, and does not break down as $B \to 0$ making it generally a more practical choice for cut selection than the others.

In cases where the cross-section or number of signal events, S, can be estimated from either well-informed simulations, or other detector channels; and where there is significant signal expected over background: optimising either Equation 8.4 or Equation 8.5 will suffice, however it is generally preferable to use Equation 8.6.

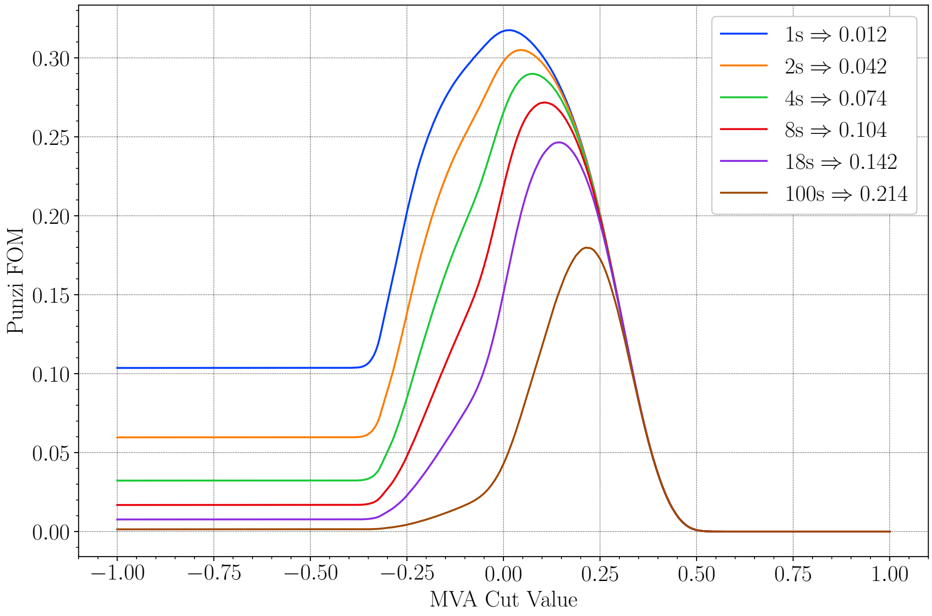
A plot of the Punzi FOM for various bin lengths (and therefore background counts), along with the corresponding optimum cut can be seen in Figure 8.12a. A plot of the discovery FOM, used for the ALPHA-g cut, can be seen Figure 8.12b.

8.7 Systematic Studies

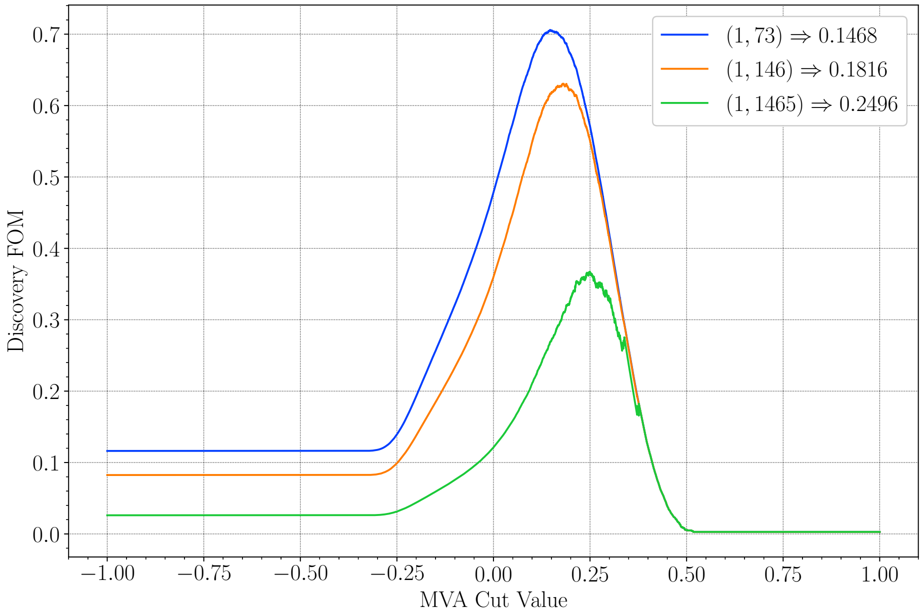
Many important systematic studies are required to ensure the robustness of our chosen model to data analysis. As the BDT was found to be the best performer we will focus on only that model, and since these studies should equally be applicable to both ALPHA-2 and ALPHA-g this section will only describe the ALPHA-2 model.

8.7.1 Grid Search

The hyperparameters of a model can greatly effect the resulting performance of the model, as discussed in Chapter 6.



(a) Punzi FOM for various bin widths in the ALPHA-2 BDT_WithR model, the corresponding optimum cut value is also shown.



(b) Discovery FOM for various numbers of signal and background denoted (S, B) in the ALPHA-g BDT model, the corresponding optimum cut value is also shown. These three models are referred to as “Ideal”, “Low Background”, and “Ultra-Low Background” and make up the three cuts used day-to-day in ALPHA-g. “Ideal” was chosen for the paper, optimising for 1 count of signal in a period of 8 seconds.

Figure 8.12: Various FOMs for both ALPHA-2 and ALPHA-g models.

As such a *grid-search* is often deployed where many different models are trained, varying the hyperparameters of the model and converging to a final optimum choice. The parameters swept and resulting signal efficiency at 1% background for the ALPHA-2 BDT are displayed in Table 8.7.

Variable	Range Tested	Results
Number of Trees	[300,500,**850**,1000]	[0.882, 0.8867, **0.8894**, 0.8902]
Minimum Node Size	[1%, **2.5%**, 5%, 10%]	[0.8912, **0.8894**, 0.8855, 0.8763]
Maximum Node Depth	[2,**3**,5,10]	[0.8801, **0.8894**, 0.894, 0.8947]
Ada Boost Beta	[0.2, **0.5**, 0.6, 0.9]	[0.8865, **0.8894**, 0.889, 0.8817]
Bagged Fraction	[0.2, **0.5**, 0.6, 0.9]	[0.885, **0.8894**, 0.8895, 0.8889]

Table 8.7: Variables used while searching for optimum hyperparameters in ALPHA-2. Bold entries indicate the default value from where the sweeps were performed. A true grid search would try all combinations of these parameters, but the above grid search alone would require 1024 models and as such linear sweeps over individual parameters were favoured instead. The results column shows the signal efficiency at 1% background for each of the models.

These results suggest the model hyperparameters are at a stable plateau, and improvements on the default model are marginal. They all stem from minimal modifications and pose the risk of overtraining (more trees, smaller nodes); therefore, we decide to remain with the default model.

8.7.2 Introducing Cosmic Contamination

We are able to quantify the effect of our roughly 0.11% cosmic contamination within the dataset by intentionally introducing cosmic contamination. To measure this we intentionally trained a BDT with certain amounts of incorrectly labelled data to mimic the effect of the cosmic contamination. We limited the number of signal training events to 300,000 and intentionally changed the percentage of this sample to be cosmic. A cosmic introduction of 50%, corresponds to 150,000 true signal events, and 150,000 true background events being combined into a single dataset and labelled signal prior to training. We used a further 300,000 signal events for testing, and 570,000 cosmic events for both the training and testing background datasets. The model accuracy, and signal efficiency at 1% background as cosmic contamination is gradually increased can be seen in Figure 8.13

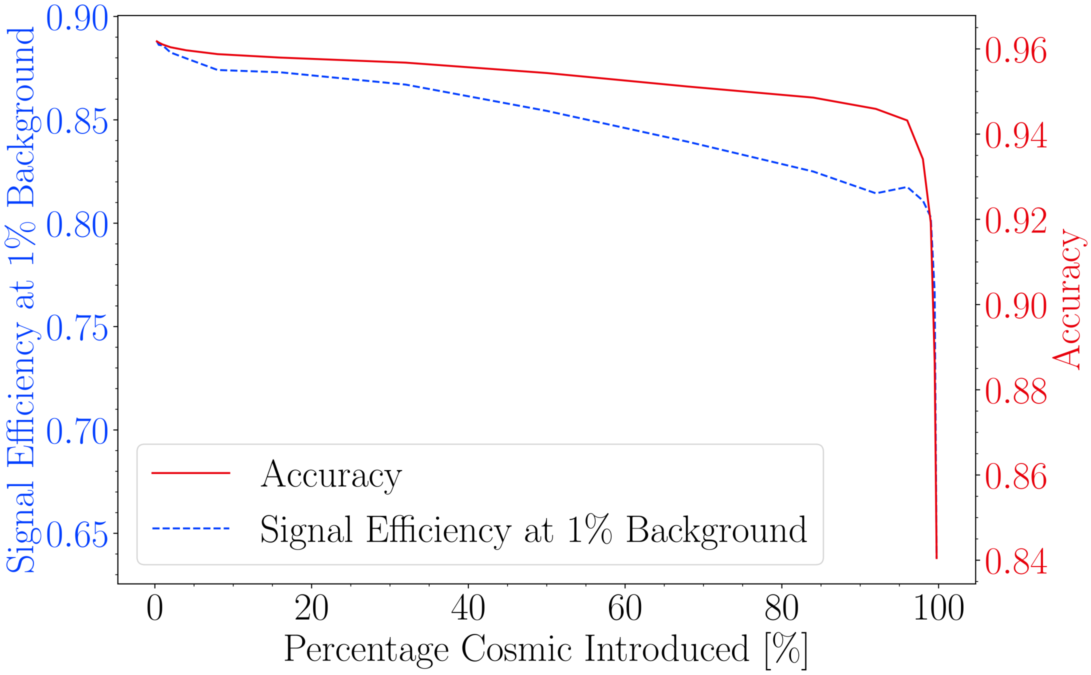


Figure 8.13: The effect of intentionally introducing cosmic data mislabelled as signal to the training step. Interestingly the models performed poorly on the mislabelled training data, but well on the testing data. Suggesting the model was able to ignore the labels and find good cuts regardless.

These results suggest that the BDT is not particularly sensitive to contamination until the curve approaches very low numbers of signal events as compared to background. This is potentially due to how far away the events are in feature space, meaning that despite the mislabelled data and poor performance on test data, it is still able to find meaningful relationships/cuts in the data.

8.8 Conclusion

In conclusion it has been shown that BDTs are the most accurate performers for the classification of background and signal events for the ALPHA experiment given the high level variables generally used. The first model trained on data from the ALPHA-g detector, used to generate results in [17] was presented, and it's lack of z bias and overfitting was shown. The model is $\sim 96\%$ accurate, and has a signal efficiency of $\sim 83\%$ when the background rate is reduced to 1%. The models trained for the 2023/2024 experimental run on the SVD have been presented, offering the first in-depth look at the ML procedures at ALPHA. We have tested the parameter space for trained models, and it showed the default choices are at, or near, a local peak suggesting there is not much room for improvement with hyperparameter scans alone. Finally, it has been attempted to quantify the effect of cosmic contamination in classification models, and they are shown to be quite robust against this, likely owing to how far away events sit in parameter space.

Chapter
9

Analysis & Results

Once the best performing model has been found, and the optimum cut chosen using the methods in Sec. 8.6, the final step in the detection and reconstruction chain is to analyse the data with the chosen model, applying the cut and rejecting events that do not pass.

This chapter will show the performance of the best performing ALPHA-2 and ALPHA-g models applied to a final validation data set containing data taken from the SVD during the 2023-2024 experimental runs, showing the difference in the cosmic pass cut rate, and the signal efficiency before and after its application; and the TPC during the 2022 experimental run.

9.1 Validating Models

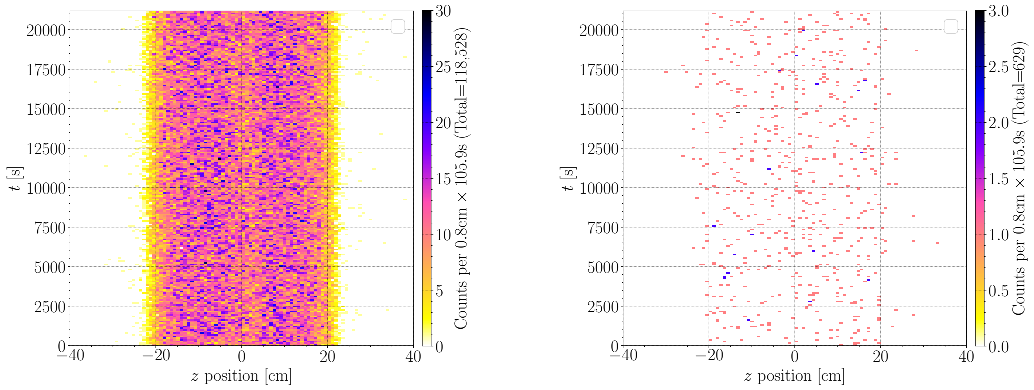
In order to provide an unbiased evaluation of the final model we use a set of data the model has not seen during the training and testing process. This final set is known as the validation set.

9.1.1 ALPHA-2 Model(s)

The best performing model found for use in ALPHA-2 was the boosted decision tree including r, known as BDT_WithR. There is a purpose to the model without r, and this will be discussed and presented briefly in Sec. 9.1.1. The BDT_WithR model however is to be used to analyse results where r sensitivity is not important, i.e. where we do not expect significant background from annihilations on residual gas. The two most important properties of a given MVA is the background rejection efficiency (and subsequent cosmic pass cut rate), and the signal efficiency. To ensure we do not introduce any z bias to the model we look at the z distribution between data with and without the cut.

Background Rejection

The SVD has a cosmic trigger rate of 9.603(50)Hz, roughly 57% of which are reconstructed to vertices, giving a cosmic vertex rate of 5.467(16)Hz. The BDT model trained in Sec. 8.4 (labelled BDT_WithR) operating at a cut value of 0.142 - chosen by optimising the Punzi FOM (Equation 8.6) for 18s windows[1] - is able to reduce this background rate to 29.68(12)mHz. This new rate is over 180 times lower than without the ML model. A plot comparing the $z-t$ distributions of all vertices reconstructed during a single 5hr53m cosmic run with and without the MVA cut is shown in Figure 9.1.



(a) 21187s of cosmic data taken with the SVD. The total number of reconstructed vertices is 118,528 - giving a cosmic rate of 5.59(2)Hz.

(b) 21187s of cosmic data taken with the SVD filtered to only include events that pass the machine learning cut. The total number of cosmic events that pass the cut is 629 - giving a cosmic rate of 29.68(12)mHz. This corresponds to a rejection of 99.469% of cosmic events.

Figure 9.1: A 2D histogram of a) all reconstructed vertices; and b) all reconstructed vertices that passed the ML cut during 5h53m of pure cosmic data. Each bin is 0.8cm wide in z and $\sim$ 106s long in t.

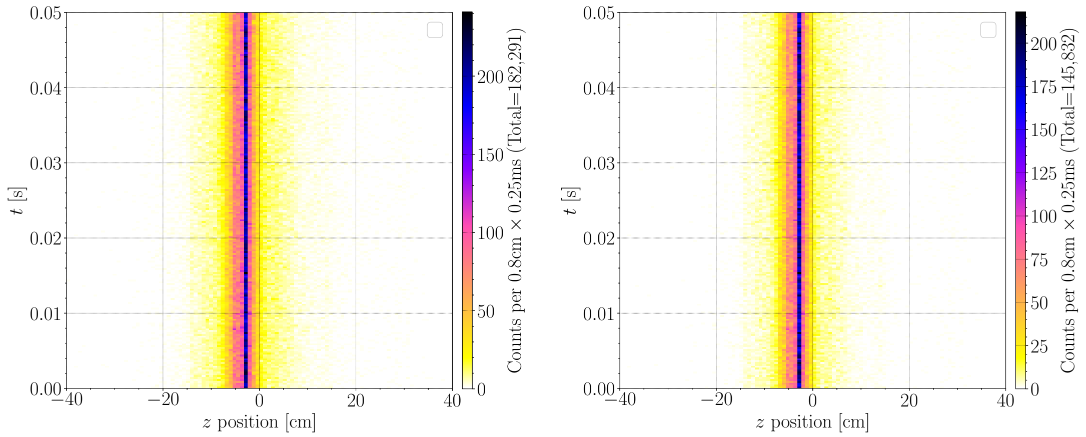
This reduction in cosmic rate corresponds to a successful rejection of 99.469% of cosmic events and is a significant improvement on the detector performance. However, while this cosmic background suppression is important for improving the performance of the SVD, any model that reduces the cosmic background, must maintain high signal efficiency to be usable. The trade-off between signal efficiency and background rejection is a heavily discussed topic in commissioning ALPHA models.

---

[1] 18s windows were used for two experiments this year and as such this is chosen as the cut value for the validation set in this chapter.

Signal Efficiency

The signal efficiency of the SVD with no cuts is roughly 88%, however, this would correspond with a cosmic pass cut rate of 5.467(16)Hz, which is too high. The data sample collected for training ML models (and described in Sec. 8.3) is close to pure. By using data generated in the same fashion we can evaluate the signal efficiency of the chosen cut. Again using the BDT_WithR model operating at a cut value of 0.142 we find that of the 182,291 vertices reconstructed during 347.8s of these saturated windows, a total of 145,832 vertices pass the MVA cut. This gives a signal efficiency of 79.8(3)% compared to total vertices. A plot comparing the $z - t$ distributions of all vertices reconstructed during the 347.8s of windows where the detector is sufficiently saturated both with and without the MVA cut can be seen in Figure 9.2.



(a) 347.8s of close to pure antimatter annihilation data taken with the SVD. The total number of reconstructed vertices is 182,291.

(b) 347.8s of close to pure antimatter annihilation data taken with the SVD filtered to only include events that pass the ML cut. The total number of signal events that pass the cut is 145,832 - giving a signal efficiency of 79.8(3)%.

Figure 9.2: A 2D histogram of a) all reconstructed vertices; and b) all reconstructed vertices that passed the ML cut in 347.8s of near-pure signal. Each bin is 0.8cm wide in z and $\sim$ 0.25ms long in t.

The reduction of the background by over 180 times was deemed more desirable for the ALPHA experiment than maintaining roughly 10% of the signal and this model marks one of the best performing models produced for the SVD to date. For comparison, the standard pass cuts without the ML (described in [153]) have a cosmic pass cut of 58(2)mHz, and a signal efficiency of 76.1(3) i.e. roughly double the background for less signal. Further the cuts are fixed rectangular cuts, meaning that one cannot change dynamically the signal efficiency and cosmic acceptance rate by moving the threshold along the ROC curve in order to tailor cuts to the chosen measurement.

z-Bias

To ensure there is no z-bias introduced by applying the MVA we can compare histograms of the z distribution of both cosmic and signal data before and after applying the MVA cut. Figure 9.3a shows the z distribution of cosmic events for all vertex events, and those events that pass the MVA cut overlaid. Figure 9.3b shows the same distribution for the signal dataset. Further, Figure 9.3c and Figure 9.3d show Q-Q plots of both these distributions.

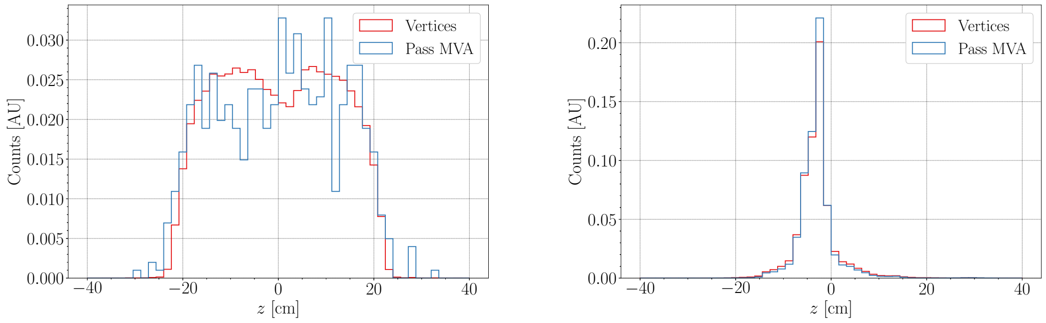
By looking at these plots we can see there is no significant change to the z distributions as a result of applying this ML cut. Coupled with Table 8.1 and Figure 8.2 we can conclude that no z bias is introduced by applying the model. While this is not as vital in ALPHA-2 as compared to ALPHA-g due to the nature of the gravity experiment (which relies on z positions), it is still important to deploy a model that is not biased, especially since the signal data set has a strong z bias as can be seen from Figure 9.3b.

r-Bias

Much like the previous section we can also compare the changes in the r distributions as a result of applying the cut.

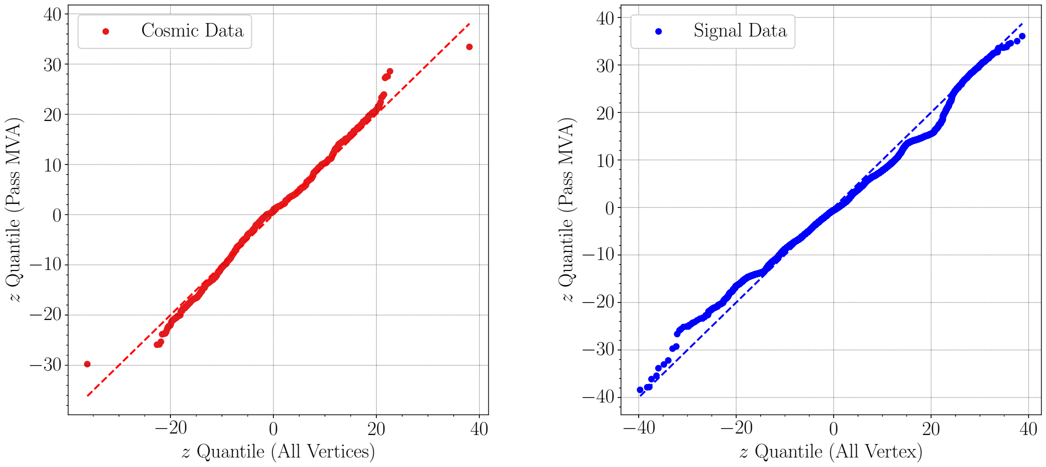
Figure 9.4a shows the r distribution for cosmic data before and after the MVA cut, and Figure 9.4b shows the same distributions for the signal dataset. Q-Q plots for these distributions can be found in Figure 9.4c and Figure 9.4d.

The plots show there is a significant difference between the r distributions, which makes sense since the BDT_WithR model uses r as one of its selection variables. This difference in distributions is not necessarily a bad thing, r can be a powerful selection variable; and the reason for that makes intuitive sense. As the r coordinate of signal events will be on the trap wall (roughly at $r = 2.5$cm), while



(a) z distributions of all vertices, and those that pass the MVA cut for cosmic data sample.

(b) z distributions of all vertices, and those that pass the MVA cut for signal data sample.

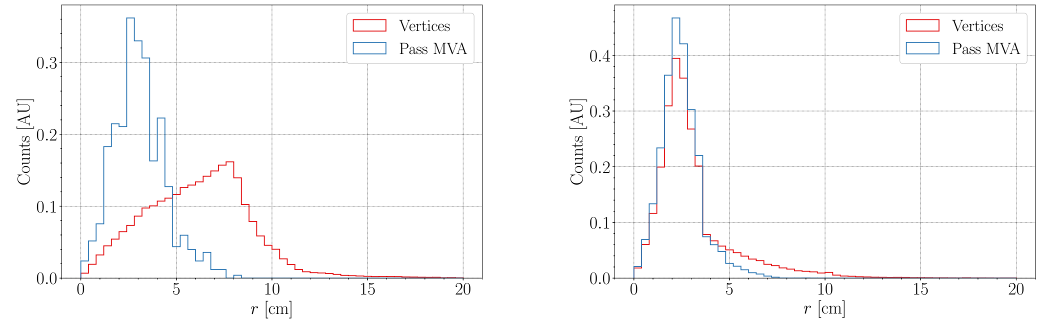


(c) Q-Q plot of z distributions before and after cut, on cosmic data ($R^2 = 0.9912$).

(d) Q-Q plot of z distributions before and after cut, on signal data ($R^2 = 0.9795$).

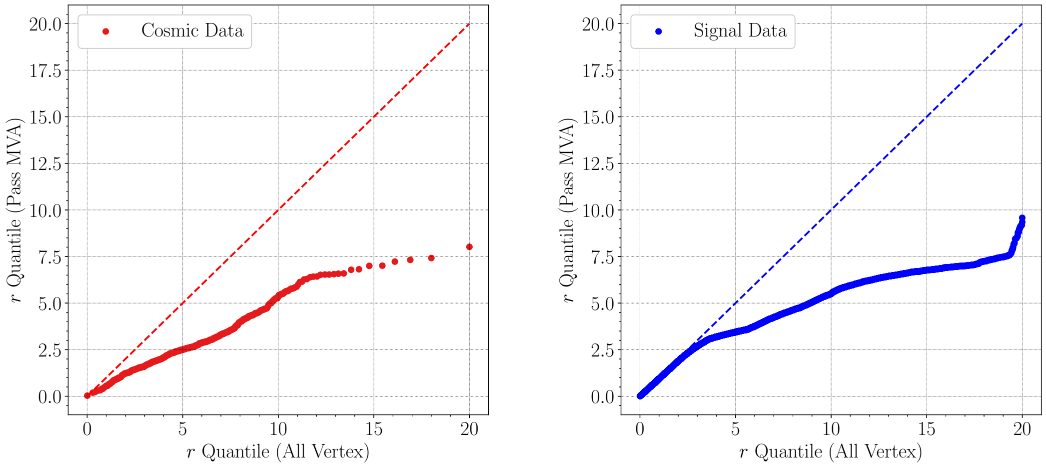
Figure 9.3: Histograms comparing the z-distributions of samples of cosmic and signal data, both before and after applying the MVA cut. Below those sit Q-Q plots for each pair of distributions.

cosmic events could come from anywhere in the trap therefore by removing events with $r > 2.5$ we can potentially filter some cosmic events (detector reconstruction smearing effects notwithstanding). However, by adding this variable to the model we lose the ability to discriminate this ourselves and control the variable exactly. This is important in experiments where we want to place strict cuts on r, and want to better understand the r dependence of our results. One such experiment uses microwaves to probe the antihydrogen hyperfine splitting. When microwaves are injected into the trap, they cause heating on the electrode surfaces. This can cause ice which has formed on the wall due to bad vacuum to melt, and be ejected



(a) r distributions of all vertices, and those that pass the MVA cut for cosmic data sample.

(b) r distributions of all vertices, and those that pass the MVA cut for signal data sample.



(c) Q-Q plot of r distributions before and after cut, on cosmic data ($R^2 = -0.3976$).

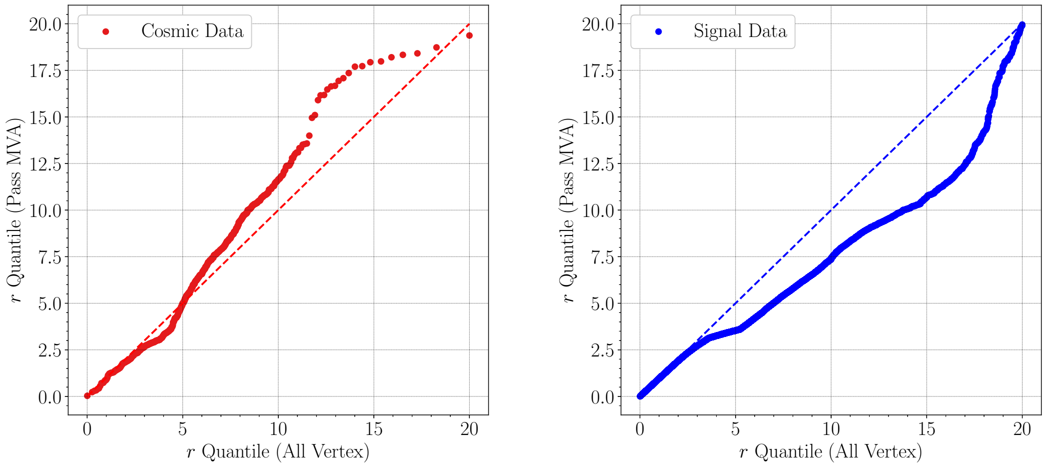
(d) Q-Q plot of r distributions before and after cut, on signal data ($R^2 = 0.5615$).

Figure 9.4: Histograms comparing the z-distributions of samples of cosmic and signal data, both before and after applying the MVA cut, and Q-Q plots comparing these distributions.

into the trap temporarily, causing annihilations of the trapped $\bar{\mathrm{H}}$ on these gas molecules. These events are still antimatter annihilations, and their fingerprints will be generally similar to events in the signal sample of our training data, meaning the ML models cannot discriminate these events correctly. In terms of the given experiment these events are considered background, and as such having the model control the variable r is undesirable. Instead, it would be preferable to apply a rectangular r cut after the ML classification, which is designed to maximise signal-to-background based on this condition and other experimental results (i.e. where we see annihilations when the microwaves are on but off-resonance).

Removing r-Bias

This is where the BDT_NoR model would be useful, allowing us to discriminate events to the best of our ability without using r. The r cut, specifically tuned to the vacuum conditions of the experiment to maximise signal over background (by including annihilations on residual gas) can then be applied afterwards. The results in performance of the "BDT_NoR" model are displayed in Table 9.1. However, looking at Q-Q plots (Figure 9.5), we again see some r bias.



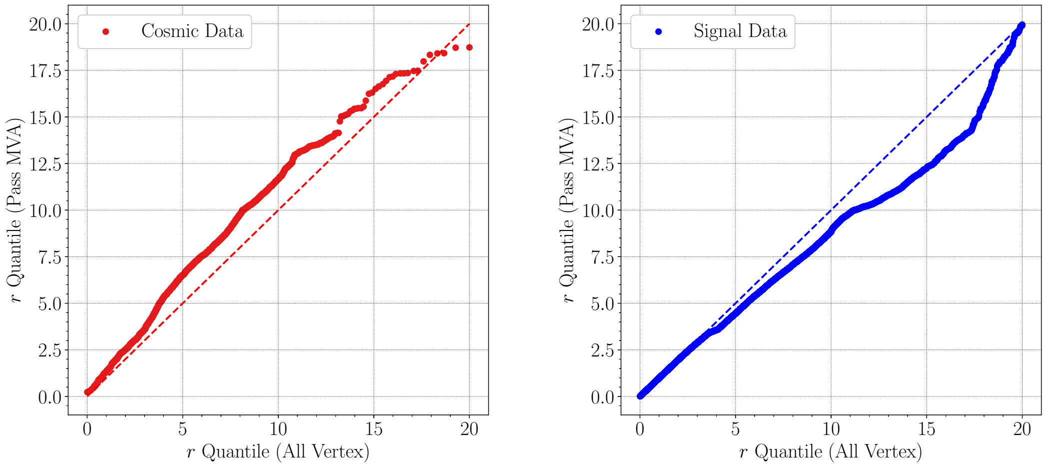
(a) Q-Q plot of r distributions before and after cut, on cosmic data ($R^2 = 0.8299$).

(b) Q-Q plot of r distributions before and after cut, on signal data ($R^2 = 0.8301$).

Figure 9.5: Q-Q plot of the r-distributions of samples of cosmic and signal data, both before and after applying the NoR MVA cut. Here we see despite removing the variable r itself, the model still has some bias in terms of r.

Returning to Figure 8.3 we see that a few additional variables have a significant correlation with r. By removing these variables and training a new model using only the 7 least correlated variables with respect to r we find the AUC of the ROC curve drops to 0.974, and the signal at 1% efficiency drops to 0.504. However, looking at the new Q-Q plots generated as a result of testing this new model on a validation set (Figure 9.6) we see that the r bias is now less significant, though still not completely independent as it was of z.

It can be argued that complete independence of r is not the goal (especially since there is a genuine dependence of r in our distribution) and that the idea of removing r from the model was not to entirely remove bias, it was simply to



(a) Q-Q plot of r distributions before and after the ML cut on cosmic data ($R^2 = 0.7511$).

(b) Q-Q plot of r distributions before and after the ML cut on signal data ($R^2 = 0.9503$).

Figure 9.6: Q-Q plots comparing the r-distributions of samples of cosmic and signal data, both before and after applying the NoRBias MVA cut.

allow a pure rectangular cut to be applied and not give control of this powerful variable to the "black box" of the ML model. Regardless of final implementation, these results suggest that simply removing r from the model might not be enough, and gaining complete independence from r is not as easy as previously expected, coming at a significant reduction in accuracy (though the rectangular cut may make back some of this lost performance). A table presenting the cosmic pass cut rate and signal efficiency of the NoR model, this new model (denoted NoRBias) with the reduced variable selection, and both with various rectangular r cuts can be seen in Table 9.1.

Though some of these efficiencies may be low, in the era of stacks of 15,000+ antihydrogen atoms per experiment, one of the benefits is allowing for cuts that decrease signal efficiency significantly. This benefit is already being exploited by experimental procedures, where cooling techniques such as adiabatic expansion (which can lose up to 50% of the sample) reduce the sample significantly in exchange for other benefits, such as sub mK samples of $\bar{\mathrm{H}}$. There is no reason these benefits could not also be applied to the selection of cuts used at ALPHA.

Model	r-cut	Background Rate (mHz)	Signal Efficiency (%)	$\epsilon/\sqrt{B}$
NoR	$\forall r$	37(1)	73.9(3)	12.2(2)
NoR	$r \in [1, 4]$	9.8(7)	62.1(2)	19.8(7)
NoR	$r \in [2, 3]$	4.3(5)	30.6(1)	14.8(9)
NoR	$r \in [2.2, 2.7]$	1.7(3)	16.4(1)	13(1)
NoRBias	$\forall r$	113(2)	66.1(2)	6.22(6)
NoRBias	$r \in [1, 4]$	16.5(9)	51.7(2)	12.7(4)
NoRBias	$r \in [2, 3]$	5.9(5)	25.1(1)	10.3(4)
NoRBias	$r \in [2.2, 2.7]$	3.3(4)	13.33(9)	7.3(7)

Table 9.1: Background rate and signal efficiency of both models trained to eliminate r bias from the model. By applying a rectangular cut after the MVA we are able to obtain better performance and still maintain clear understanding of the r distribution of vertex events. The cuts in r here are arbitrarily chosen to contain the radius of the electrode wall ($r = 2.2275$) within some interval. A true analysis of the space of this cut would be required before settling on a final optimum cut. The $\epsilon/\sqrt{B}$ FOM is included to provide a method of comparing the models to one another and is related to the Punzi FOM (see Equation 8.6). In all cases the NoR model outperforms the NoRBias model, and in all cases some type of r cut improves performance.

9.1.2 ALPHA-g Model

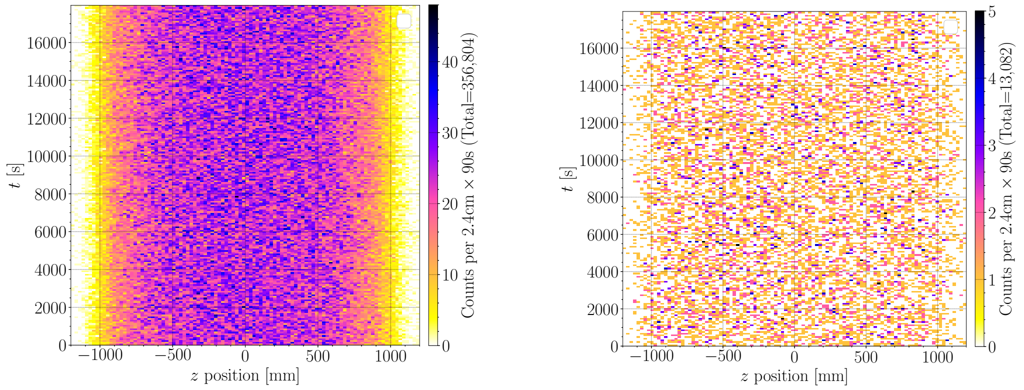
Finally, we present the same results for the ALPHA-g BDT.

Background Rejection

The TPC has a cosmic trigger rate of 58.95(6)Hz, roughly 34% of which are reconstructed to vertices, giving a cosmic vertex rate of 19.86(3)Hz. The BDT model trained in Sec. 8.4 labelled (BDT) operating at a cut value of 0.1468 - chosen by optimising the measurement FOM (Equation 8.5) with 1 signal and 73 background events[2] - is able to reduce this background rate to 0.728(6)Hz. This is considerably higher than ALPHA-2, and only corresponds to a 27 times reduction in cosmic rate, or a successful rejection of 96.3% of cosmic events. A plot comparing the $z - t$ distributions of all vertices reconstructed during a single 4h59m cosmic run with and without the ML cut is shown in Figure 9.7.

---

[2] These values were used to optimise the cut in [17] and come from vertex data without a cut to estimate the signal.



(a) 17968s of cosmic data taken with the TPC. The total number of reconstructed vertices is 356,804 - giving a cosmic rate of 19.86(3)Hz.

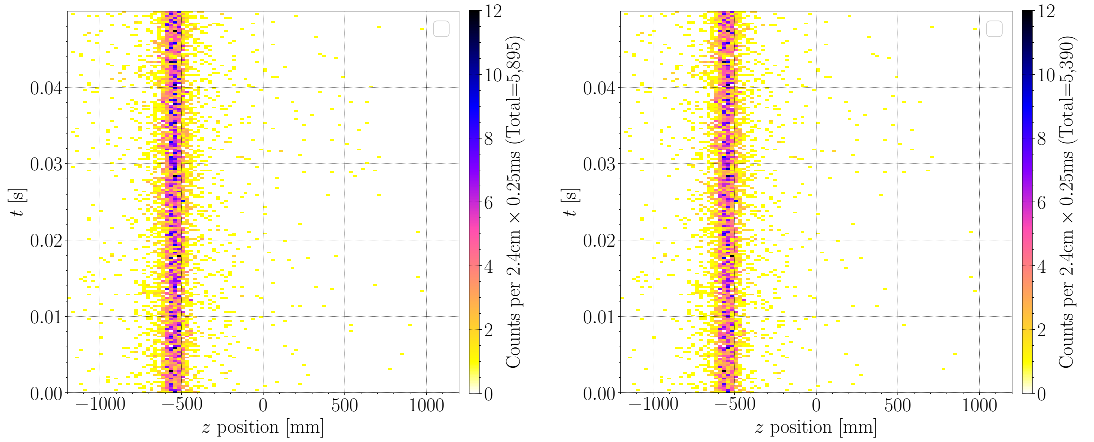
(b) 17968s of cosmic data taken with the TPC filtered to only include events that pass the machine learning cut. The total number of cosmic events that pass the cut is 13,082 - giving a cosmic rate of 0.728(6)Hz. This corresponds to a rejection of 96.3% of cosmic events.

Figure 9.7: A 2D histogram of a) all reconstructed vertices; and b) all reconstructed vertices that passed the ML cut in 4h59m of pure cosmic data. Each bin is 2.4cm wide in z and ~ 80s long in t.

Signal Efficiency

The signal efficiency of the TPC with no cuts is roughly 55% corresponding to the full cosmic pass cut rate of 19.86(3)Hz. The challenges of using a TPC become apparent at this point as it generally has a lower vertex efficiency rate, a much higher background rate, and, due to the fact that antihydrogen production in this new trap is not well refined yet, much fewer atoms to work with. These may be a result of the infancy of this detector, and the performance of the SVD is not only a result of the hardware but years of software development which is not yet present in the TPC. Regardless this makes the ML, and specifically the signal efficiency of the chosen cut crucial.

Again using the near pure training data sample collected for ML (and described in Sec. 8.3) we can evaluate the signal efficiency of the chosen cut. We find that of the 5,895 vertices reconstructed during 28s of these saturated windows, a total of 5,390 vertices pass the MVA cut. This gives a signal efficiency of 91.4(1.7)% compared to total vertices. A plot comparing the $z - t$ distributions of all vertices reconstructed during the 28s of windows where the detector is sufficiently saturated both with and without the MVA cut can be seen in Figure 9.8.



(a) 28s of close to pure antimatter annihilations taken with the TPC. The total number of reconstructed vertices is 5,895.

(b) 28s of close to pure antimatter annihilations taken with the TPC filtered to only include events that pass the ML cut. The total number of signal events that pass the cut is 5,390 - giving a signal efficiency of 91.4(1.7)%.

Figure 9.8: A 2D histogram of a) all reconstructed vertices; and b) all reconstructed vertices that passed the ML cut in 28s of near-pure signal. Each bin is 2.4cm wide in z and ~ 0.25ms long in t.

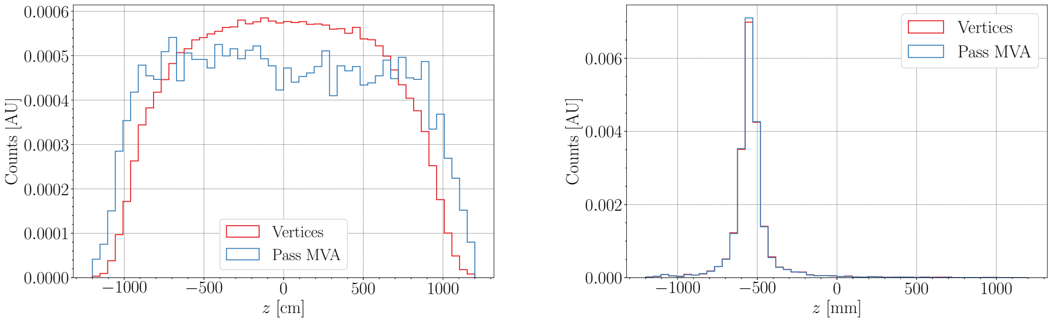
z-Bias

Again we can compare histograms of the z distributions of vertices before and after applying the MVA cut to ensure no z-bias is introduced by applying the MVA.

This metric is extremely important for ALPHA-g, the entire result hinges on determining the z position of resulting annihilations, any model that introduces a bias in this regard cannot be used. This is especially risky considering all training data comes from one region of the trap while cosmic events can come from anywhere. Any model that contains z will quickly learn to accept events in the "mixing" region of the trap, and reject all others.

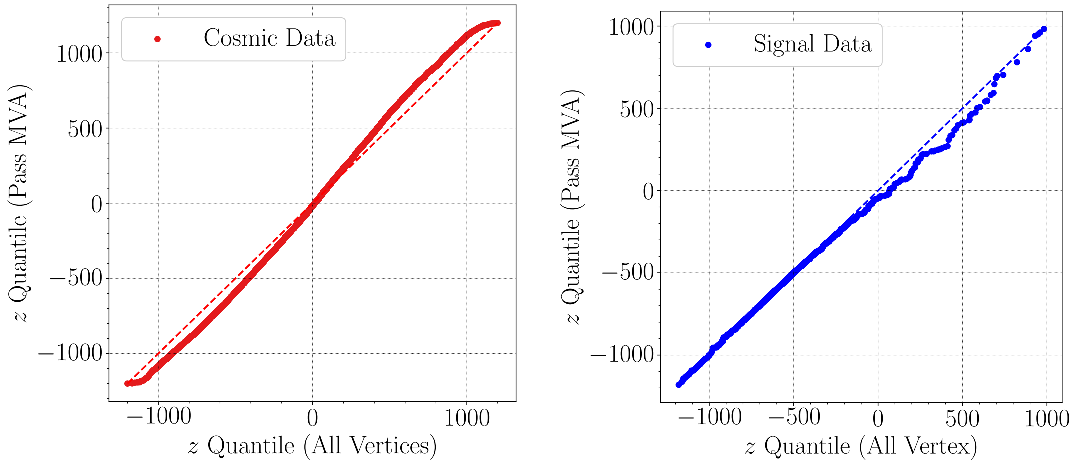
Figure 9.9a shows the z distribution of cosmic data for all vertex events, and only those events that pass the MVA cut overlaid; and Figure 9.9b shows the same distributions for the signal dataset. Further, Figure 9.9c shows the Q-Q plots of the cosmic distributions, and Figure 9.9d the signal Q-Q.

By looking at these plots we can see there is no significant change to the z distributions as a result of applying this cut. Coupled with Table 8.4 and



(a) z distributions of all vertices, and those that pass the MVA cut for cosmic data sample.

(b) z distributions of all vertices, and those that pass the MVA cut for signal data sample. Here the size of the detector and lack of data at the positive z end of the detector causes a drift, but in the region of interest the distributions match well.



(c) Q-Q plot of z distributions before and after cut, on cosmic data ($R^2 = 0.9965$).

(d) Q-Q plot of z distributions before and after cut, on signal data ($R^2 = 0.9923$).

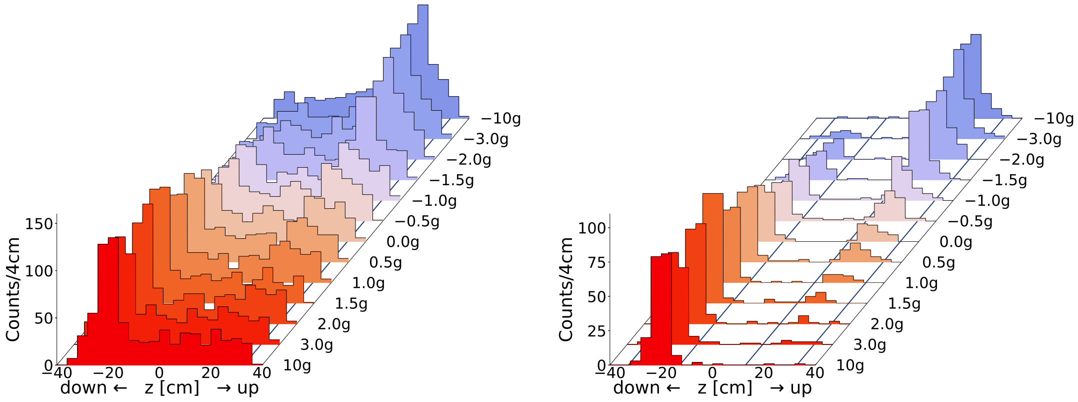
Figure 9.9: Histograms comparing the z-distributions of samples of cosmic and signal data, both before and after applying the MVA cut. Below those sit Q-Q plots for each pair of distributions.

Figure 8.9 we can conclude there is no z bias introduced through the application of this model.

9.1.3 Analysing Published Results

Gravity Measurement

In principle the measurement scheme designated to ALPHA-g is quite simple. Once the $\bar{\mathrm{H}}$ is trapped, the mirror coils at each end of the trap can be lowered, allowing the $\bar{\mathrm{H}}$ to escape axially. The effect of gravity should translate into a difference between the number of atoms escaping down vs the number escaping up. Further, by biasing the difference between the currents of the mirror coils that define the well as they ramp down to such a value that is comparable to g we are able to either amplify, cancel out, or reverse the effect of gravity on the sample of $\bar{\mathrm{H}}$. In fact, by performing this experiment with a range of biases we are able to determine exactly where the bias cancels out the effect of gravity (found to be $0.75 \pm 0.25g^3$ [17]). The experiment can be repeated at multiple biases to improve statistics. A histogram of the results from this experiment organised into biases (and displayed in units of g) both with and without the ML can be seen in Figure 9.10.



(a) Escape histograms with no cut.

(b) Escape histograms with machine learning cut.

Figure 9.10: The raw event z-distributions or "escape histograms" for each of the experimental bias values, including the $\pm 10g$ calibration runs.

---

[3] The error of this result comes mostly from two sources: 1. The temperature distribution of the $\bar{H}$ is not well know, which adds to uncertainty in the supporting simulations, and 2. The uncertainty in the off axis magnetic fields during the magnet ramp down which are hard to measure dynamically.

A further possible advantage to the ML would be allowing for slower measurements, which gives more control of the magnetic fields (a major source of error in the result), but results in considerably more background. Having an ML cut capable of shifting the optimum cut would allow for measurements like these to be trailed again, potentially giving a more accurate result for the value of $\bar{g}$. During the 2022 run a few of these "slow ramp" runs were performed (the ramp occurs over 130s instead of 20s) but not enough were performed to obtain a full measurement. Preliminary results [229] suggest that the value obtained during these runs is much closer to the expected value of $\frac{g}{\bar{g}} = 1$ however without a full systematic analysis of the ramp, and a full range of biases the analysis is not complete. It should be noted however that without the application of ML in these 130s ramps, which contain over six times as many background events, these experiments would be impossible.

Chapter 10 Deep Learning at ALPHA

Deep learning using raw detector signals is an attractive idea at ALPHA. Using these signals would allow classification to skip the reconstruction process entirely. This has the potential to be more accurate than even the BDTs, or could at least work in cooperation to augment the BDT results and improve the overall reliability and accuracy of results. As such this chapter briefly explores the idea of using raw signals from the SVD ASICs to classify events without reconstruction.

10.1 The Data

The data used for this application is a sum of all signals received per ASIC, and, though the data was saved per channel (each ASIC reads 128 channels) they were converted prior to training to match this format. The result is a vector of length 288, containing a value ranging anywhere from 0 to 16,000 with means of 50-100 and RMS $\sim$ 200. To ensure that the models do not learn that certain regions of the trap correspond to signal and other to background (due to mixing in one electrode in the trap) a new dataset was introduced that has z values from other regions of the trap. "Slow release" experiments (described in Sec. 5.2.1) were performed in various regions throughout the trap to ensure a wide range of z values in the signal sample. Unfortunately this sample is not as pure as the mixing data because the same amount of $\bar{p}$ are released as in the mixing windows over 200s instead of 2s. It is possible to release these $\bar{p}$ over a shorter time, saturating the detector and obtaining more pure data. However, in order to perform the counting experiment from Sec. 5.2.1 it is important that the detector is not saturated meaning these experiments cannot be performed in tandem and a dedicated experiment must be performed to obtain this saturated dataset for machine learning. Due to constraints on time with the machine itself these experiments could not be performed and as such the unsaturated data was used. In order to make sure the cosmic contamination stays low we used only events that passed the standard rectangular cuts sometimes used in ALPHA-2 [153]. While this does reduce the cosmic contamination, it potentially only allows the "easy-to-classify" events into the dataset. The choice of using this cut is a trade off between allowing cosmic contamination, or eliminating more difficult to classify events. The additional cuts were applied for this dataset, as the pure dataset which includes all events will also be used alongside it. This ensures we have a good mix of events from all parts of the trap, with a range of “difficulty”, and limited cosmic contamination.

The combined data set contains 1,036,144 signal events from mixing data 2/3 of which was used for training; 1,567,938 events from the slow release experiments 1/3 of which was for training; and 2,894,001 cosmic events 2/3 of which were for training. The hyperparameter scans used 300k signal and 300k background events for training (and the same for testing), and the final model (Sec. 10.4) used 800k signal and 800k cosmic events for training (and the same for testing).

10.2 Architecture

While a convolutional neural network (CNN) would be an interesting approach (and these have been shown to work well on the ALPHA-g detector, looking at pad outputs only) the SVD contains three concentric layers of panels, meaning it is not possible to simply flatten the data into a 2D tensor and apply a convolution. Relationships between layers are important, especially between panels that are directly behind one another. Further, there is cylindrical symmetry in the detector, and any convolution would have to be able to fold over the edge of the 2D data, connecting nearby strips with those on the other side of the 2D image. For this reason a convolution was not attempted and only deep neural networks/MLPs were implemented. Nevertheless, convolutional neural networks that are able to handle these spatial issues could have great potential and should be investigated in the future.

All training was performed with ROOT::TMVA package [227], and though Python has many options for deep learning it was unable to load the samples into memory with the hardware used, and as such only ROOT was capable of handling the large files and variables. Python packages such as `uproot` can handle this issue better, being specifically designed to stream ROOT data into Python ML libraries, and this was trailed for deep learning in ALPHA-g where it showed promising results, however this has yet to be implemented for ALPHA-2.

10.3 Tuning the Model

We started by training a series of fully connected MLPs, and scanning the learning rate to see the effect this had on the model. The models had six layers with the following number of nodes: 576,288,72,72,50,50, each using ReLU activation and a batch normalisation layer [230] siting between each layer. We used L1 regularisation with a staling factor of $1e^{-4}$ to reduce weight sizes, binary cross entropy loss (Equation 6.23) and Xavier uniform [231] to initialise the model weights. We used a batch size of 256, and Adam optimiser to compute the stepping. A Gaussian transformation is applied to the variables prior to training. These “default” values were chosen as they were previously found to give good results in less systematic feasibility tests.

The learning rates (η) chosen to test consist of 0.1, 0.05, 0.015, 0.005, 0.001, 0.0005, and 0.0001. A plot showing the ROC AUC and the signal efficiency at 1% background for the training and test dataset for each of the different learning rates can be seen in Figure 10.1.

Here we see that the lower learning rates performs better (according to the AUC), however we notice the model begins to overtrain at around $\eta = 0.0010$ as the training and test accuracy begin to diverge. Based on this plot a learning rate of 0.0005 to 0.0010 was deemed acceptable.

After this we performed a scan of the batch size (S_b) while keeping all other parameters the same (learning rate fixed at 0.0010 based on the previous study). The batch sizes trailed include: 8, 32, 128, 256, 512, 1024, 2048, 4096. The resulting AUC and signal efficiencies at 1% background can be seen in Figure 10.2.

The AUC increases as the batch size gets larger (up to 4096 - see Figure 10.2), however the risk of overtraining increases concurrently, though the test accuracy still improves until 2048.

Further, as the batch size increases the training time per epoch decreased exponentially; however, the limiting factor in training time is performing the Gaussian transformation on the dataset, and, as most training sessions terminated at ~ 300

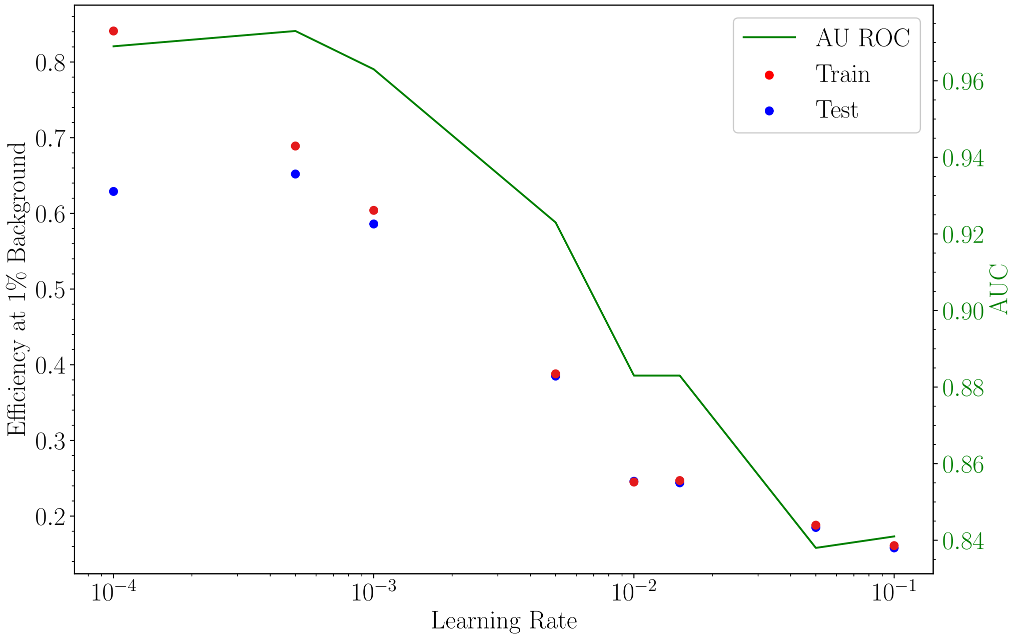


Figure 10.1: AUC, and signal efficiency at 1% background of both the training and test dataset while scanning the learning rate of the model. All points use the axis on the left aside from the AUC which uses the green axis.

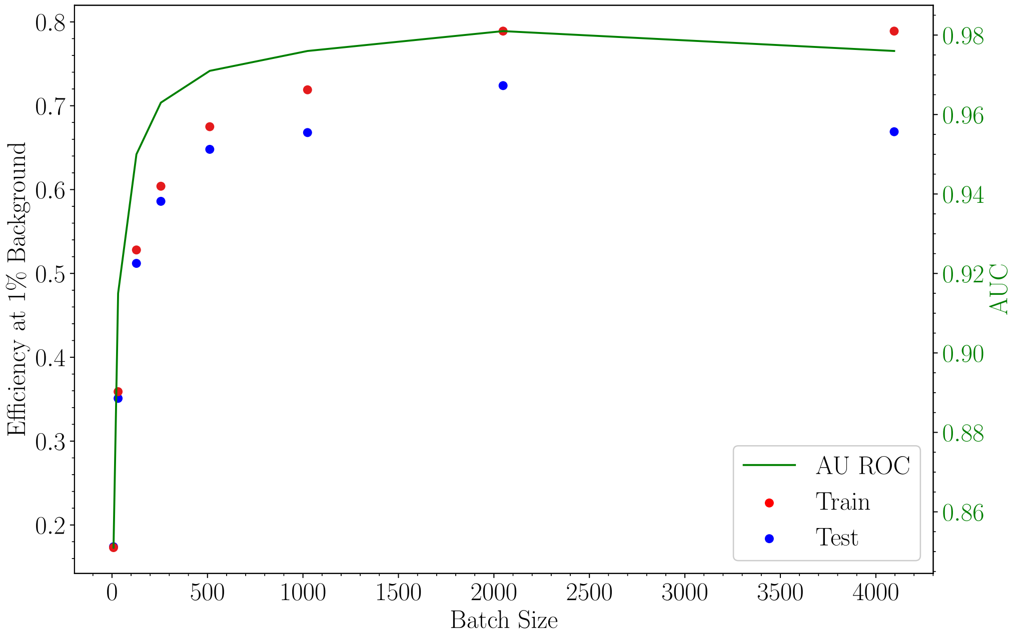


Figure 10.2: AUC, and signal efficiency at 1% background of both the training and test dataset while scanning the batch sizes used for training. All points use the axis on the left aside from the AUC which uses the green axis.

epochs after 150 epochs without improvement the speed of convergence did not influence the choice much. Regardless, we conclude a batch size of 1024-2048 to be acceptable.

Next we decided to try different model *shapes*[1] attempting to reduce the size of the model if possible. We performed this scan at two places, once with a learning rate at 0.0005, and a batch size of 2048 which aimed to give maximum accuracy, and once with a learning rate of 0.001 and a batch size of 1024 aimed at minimising overtraining. A table of the results of these models sorted in order of the signal efficiency at 1% (SEA-1) of the $\eta = 0.001$, $S_b = 1024$ models can be seen in Table 10.1.

	$\eta = 0.001$, $S_b = 1024$			$\eta = 0.0005$, $S_b = 2048$		
Shape	AUC	**SEA-1**	μ	AUC	SEA-1	μ
288,72,50	0.976	**0.675**	0.043	0.963	0.587	0.138
576,288,72,72,50,50	0.976	**0.668**	0.051	0.971	0.639	0.145
576,288,72,72,50	0.976	**0.665**	0.050	0.970	0.633	0.137
288,72,72,50,50	0.975	**0.664**	0.043	0.964	0.598	0.127
288,288,72,72,50,50	0.975	**0.661**	0.049	0.965	0.594	0.141
576,576,288,288,72,72,50,50	0.972	**0.634**	0.048	0.967	0.598	0.156
72,72,72,72,72,72,72	0.965	**0.603**	0.042	0.965	0.606	0.107
72,50	0.962	**0.585**	0.038	0.964	0.594	0.073
50,50,50,50,50,50,50	0.958	**0.573**	0.024	0.961	0.573	0.075
72	0.942	**0.450**	0.037	0.944	0.460	0.054
0	0.849	**0.167**	0.004	0.849	0.168	0.003

Table 10.1: AUC, signal efficiency at 1% background (SEA-1) for test samples, and the difference between these two (denoted μ) for each of the model shapes trailed. The model is sorted in order of SEA-1 of the $\eta = 0.001$, $S_b = 1024$ models. The final model takes the 288 inputs and directs them immediately to the output later, hence the shape is denoted with "0". Bold typeface denotes the column in which the table is sorted by.

This table shows that $\eta = 0.0005$, $S_b = 2048$ increased overstraining considerably given the same model shape (indicated by the size of μ), and cannot match the performance of the (288,72,50) model with $\eta = 0.001$ and $S_b = 1024$. This model has the added benefit of fewer synapses than other previously trained models and as such from here we consider this the default model.

We also tried training models with *momentum*, which allows the previously

---

[1] Here model *shape* refers to number and size of the layers within the model.

calculated gradient decent step to contribute to the current step, meaning the algorithm can continue to move in the same direction (in the parameter space) unless a significant push in another is applied. This can help avoid plateaus and allow the algorithm to glide more efficiently towards the global minimum. The strength of this momentum (denoted p) is zero for standard batch gradient descent and a momentum of 1 would ignore all current steps and only use the previously calculated gradient at each epoch, which means only the first gradient decent step would be used for the entire training process. Adam optimiser (which is used throughout this chapter) implements its own form of momentum and this additional momentum parameter was not expected to help; and after running a few different values the results did not show much change in the overall accuracy of the model, hence it was kept at 0. This potentially is more powerful when not using Adam optimiser, but other optimisers were found to be less accurate overall and as such this was not tested.

The final hyperparameter we decided to sweep was the strength, and type of regularisation used. Currently, we have been using L1 normalisation and a strength of $1e^{-4}$. We tried $1e^{-3}$ and $1e^{-5}$ with L1 normalisation, surrounding our current value; and $1e^{-3}$, $1e^{-4}$, and $1e^{-5}$ with L2 normalisation. The results are presented in Table 10.2 where we conclude that the current values are optimum.

Finally, to ensure no additional improvements could be made to the model, some additional parameters were tested.

We tried using sigmoid (Equation 6.17) and tanh (Equation 6.18) activations for the layers instead of ReLU; removing regularisation; using Xavier initialisation; and other data transformations (the default model uses only a Gaussian transformation). Using a tanh activation layer increases the test SEA-1 by 1.3% but no other significant improvements could be made. The final parameters of the model then with this final change can be found in Table 10.3.

10.4 Final Model

Now that the optimum parameters for the model have been determined we can proceed with the final model training, using a larger number of samples. We

Norm	Strength	AUC	Signal Efficiency at 1% Background		μ
			Test	Train	
L1	$1e^{-4}$	**0.976**	**0.675**	**0.718**	**0.04300**
L2	$1e^{-4}$	0.967	0.617	0.661	0.04400
L1	$1e^{-3}$	0.945	0.500	0.510	0.01000
L2	$1e^{-3}$	0.939	0.466	0.481	0.01500
L1	$1e^{-5}$	0.931	0.425	0.518	0.09300
L2	$1e^{-5}$	0.929	0.412	0.488	0.07600

Table 10.2: AUC, SEA-1 for training and test samples, and the difference between these two (denoted μ) for each of the models trailed. The model is sorted in order of test efficiency. We notice L1 norm outperforming L2 in each case, and that the current strength value of $1e^{-4}$ is optimum. We also note that increasing the strength of this regularisation decreases overtraining, but also decreases accuracy on testing set. However, if the results of this model do not perform well on validation data, this is a potential place to decrease overtraining and increase generalisation of the model. Bold typeface indicates the best model from the previous step.

Parameter	Value
Shape	Three layered MLP: (288,72,50)
Activation	tanh
Regularisation	L1
Strength	$1e^{-4}$
Learning Rate	0.001
Batch Size	1024
Weight Initialisation	Xavier Uniform
Optimiser	Adam

Table 10.3: The final model parameters after all sweeps. The model had a ROC AUC: 0.976, and a SEA-1 on the test (train) sample of 0.746 (0.688).

increased the number of test and training samples to 800k events each (this seems to be the limit with the hardware used, attempting to use more samples results in a stall at the Gaussian transformation stage as we were not able to load more than this into memory to perform the Gaussian transformation). This increase in samples both improved the performance and reduced overtraining. The final accuracy, F1 score, AUC, and signal efficiency at 1% on test (train) data was 0.964, 0.966, 0.980, and 0.723 (0.739) respectively. A plot of the training process showing train and test loss per epoch for the first 60 epochs (more were performed but the process never improved after epoch 33) can be seen in Figure 10.3.

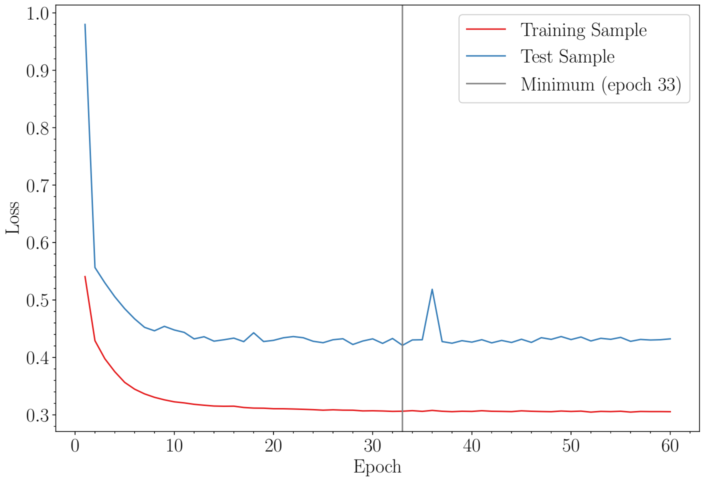


Figure 10.3: Train and test loss per epoch for the MLP. The vertical line represents the stopping point.

The ROC curve for this final model can be seen in Figure 10.4 along with the ROC curve for BDT_WithR and BDT_NoR for comparison.

While this MLP performs well, we were unable to outperform the standard BDT trees used at ALPHA, nor the model excluding r. However, as the current logic on reading out events is entirely random, during high rate windows a model such as this which can work on the raw signals could improve the signal-to-noise ratio of the detector. Alternatively, this model could be used in tandem with the BDTs to improve their accuracy, or used to classify events prior to reconstruction. This model, paired with a regression model capable of reconstructing z from the same signal could be a very powerful tool for ALPHA, allowing for instant vertex position and classification. The final cosmic pass cut rate, and signal efficiency of this model when optimising Punzi FOM (Equation 8.6) for 18s windows is 1.363(8)Hz and 54.3(2)% respectively; a considerable drop from the BDTs but a good first step in the direction of deep learning at ALPHA.

10.5 Conclusion

The deep learning models were not able to outperform the BDTs but do show promising initial results. The problem is well suited to a CNN, but the challenges

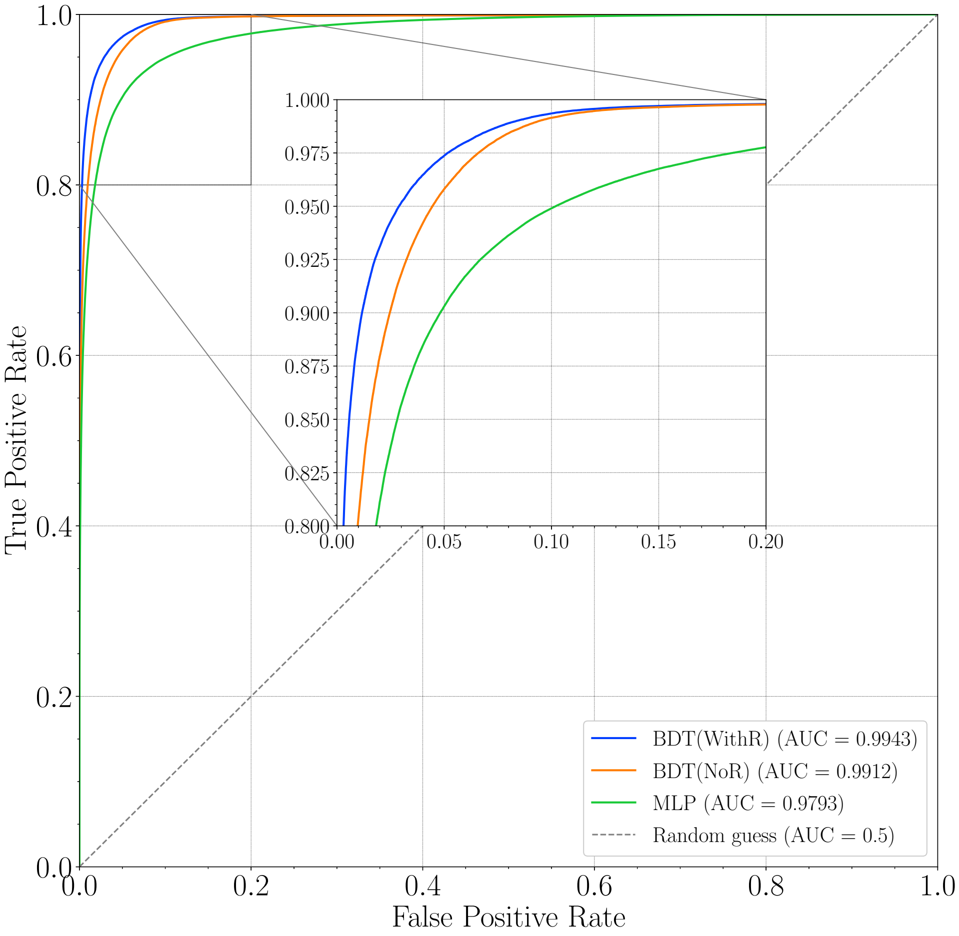


Figure 10.4: ROC curve for the MLP described in Table 10.3 along with the ROC curve for BDT_WithR and BDT_NoR models presented in Sec. 8.4.3.

involved with the data makes this architecture difficult to apply at this point, and calls for additional research. These models were trained with some of the lowest level variables possible in the SVD, making them extremely desirable and it's possible that combining this result with the BDT would improve the classification at ALPHA considerably. Further, the combination of a model like this, with a regression model capable of predicting z position from the same input would provide instant results for the ALPHA experiment.

The model trained here is likely small enough to be installed on an FPGA and could provide the basis for the first ML trigger used in the ALPHA experiment, potentially increasing the signal-to-noise ratio of the detector as a whole.

Chapter 11 Conclusion & Future Work

11.1 Conclusion

By performing Monte Carlo simulations, we have been able to predict the efficiencies of specific panels throughout the experiment (see Chapter 5), allowing us to trace precisely when losses are incurred in our antihydrogen production cycle. These values could help improve transfer efficiencies throughout the experiment and, ultimately, antihydrogen production rates. When antihydrogen is as rare a substance as it is, any improvements that increase efficiency and production rates are crucial. We have performed specific experiments to verify these simulations, the results of which have been presented, and generally show good agreement.

We have reviewed the currently deployed ML techniques used by ALPHA to classify events, and improved the efficiency and performance of the already well established models used in the SVD. Further, the first ever ML model capable of classifying events in the TPC has been presented and has already proven essential in performing the first direct measurement of the effect of gravity on the motion of antimatter, where $\bar{g}$ was found to be $0.75 \pm 0.25g$ [17]. Without this model, the signal-to-noise ratio reduces the significance of our results considerably, and overall, this model contributes to a substantial improvement in the performance of the TPC for all past and future results. The first in-depth analysis into the z and r dependence of the models trained has been presented, which is a necessary study for both detectors. The results suggest that classifying events without using r as a variable is feasible, which could be critical to improving the significance of certain measurements where vacuum conditions can cause annihilations, though it does not come without drawbacks.

We have also demonstrated the first attempts at using raw detector signals to classify events in the ALPHA experiment precluding the need for event reconstruction. Development in this vein would allow for instant event classification, and potentially an ML based trigger for the ALPHA experiment, further improving the signal-to-noise ratio, and therefore overall performance of the detectors. Using these models in collaboration with BDTs may be the way forward for classification in the ALPHA experiment and could potentially mark the first significant increase in the performance of the SVD for a while.

We can look at the ROC curve of all the models trained so far, in units more familiar to the ALPHA experiment: cosmic pass cut rate, and signal trigger efficiency to get a better understanding of each models' performance. The resulting ROC curves showing the ALPHA-2 MLP, WithR, NoR, NoR with a rectangular cut of $r < 4$ models; and the ALPHA-g BDT can be seen in Figure 11.1.

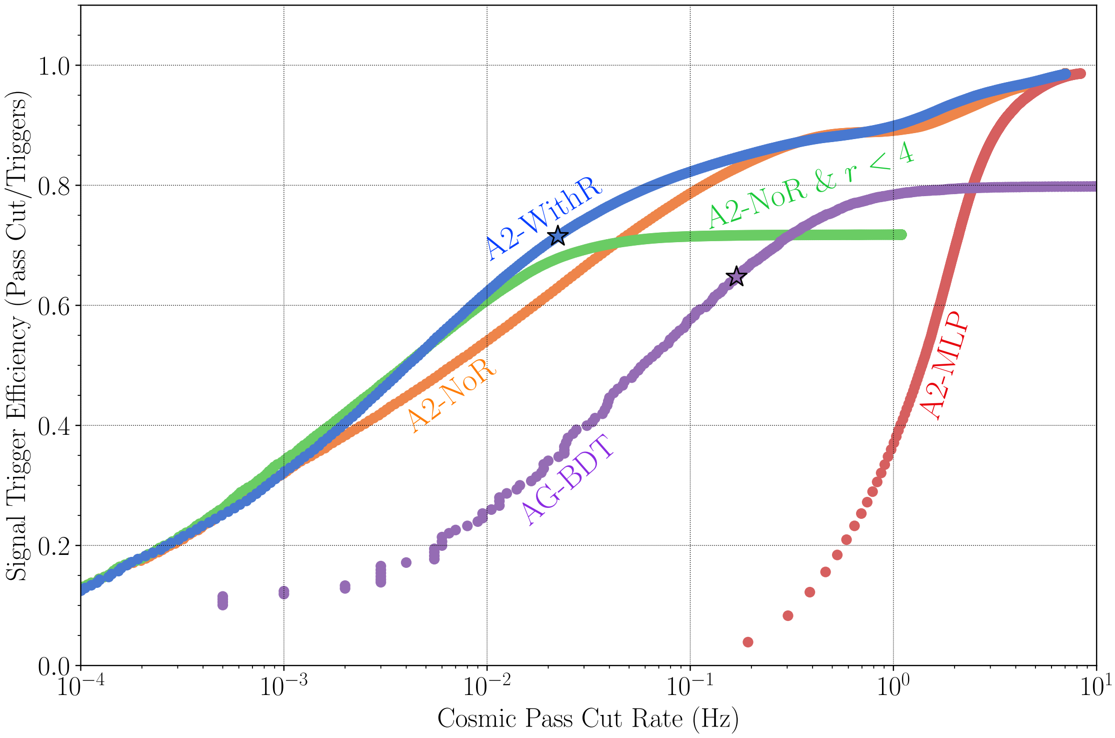


Figure 11.1: ROC curve for five models presented throughout this thesis in units of cosmic pass cut rate and signal trigger efficiency. The stars represent the cut values used for the analysis in Chapter 9, and, in the case of the ALPHA-g BDT, in [17].

Beyond ALPHA, ML methods aimed at beam profile reconstruction in a beam profile monitor have been trailed, recreating previously published results (on slightly different beam conditions) and applying these techniques to real data from an operational profile monitor at CERN for the first time. The data available limits the conclusions to be drawn from this test. However, the results were promising and suggested further data collection would be beneficial. As beams become brighter and more intense, it is reasonable to assume that future profile monitors will need more sophisticated methods of profile reconstruction than simply installing new, larger magnets. This first exploration of ML on real data points to a possible solution to this problem. If a method of confirming the validity of these results, along with some way to train models for these devices prior-to or upon installation, could be established, the material cost saved would be significant, and may become the standard for beam profile reconstruction in the future.

At the time of writing, the last year and a half (2023 and 2024 until September 1st) were spent on measurements in ALPHA-2, such as the hyperfine structure, the 1S-2S transition frequency, the 2S-4P frequency, and the 2S-2P Lamb shift transition. New techniques for cooling and measuring the temperature of antihydrogen populations have been developed, and improvements to antihydrogen production rates and laser cooling provide consistently large and cold samples of antihydrogen for all of these experiments. The results of these experiments are currently being analysed and will be described in various papers. The rest of the 2024 run is dedicated to ALPHA-g and commissioning the top trap for antihydrogen production, with the goal of obtaining a more accurate gravity measurement firmly in sight.

The ML event classification will be crucial to all of these results, reducing background considerably from the default rate of the detectors. The results in this thesis show that while the new TPC may not be able to match the performance of the SVD yet, the first steps in event classification in the TPC have been made, and there is potential for improvements in the near future. Similarly, though it's unlikely the deep learning models could ever match the performance of the BDTs alone, there could be a place for them at ALPHA, either working in cooperation with the BDTs or working to achieve a slightly different goal such as the trigger logic itself. The first evidence that this may work has been demonstrated, and there are many avenues of investigation open toward achieving this goal.

Ultimately, these results have helped improve the performance of both detectors used at ALPHA and will continue to do so in the coming years. The models presented here will allow for higher precision on measurements performed by ALPHA with increasing significance, ensuring our understanding of antimatter and, therefore, the universe, continues to develop despite the challenges that come with it.

11.2 Future Work

There are several potential directions to continue the work presented in this thesis, ranging from further data collection to improving simulations and models:

- The simulations used for calibrating panel efficiencies throughout the ALPHA experiment (described in Chapter 5) would benefit from adding the cosmic background radiation. We see a potential discrepancy in the CT simulations where the SiPM_AND channel may see more counts than expected, while all channels with a logical OR seem to match experimental data well. This could be a result of cosmic background rays triggering the detector readout erroneously, and it would be important to verify this.
- Further, experiments could be performed to verify the simulated results from the interconnect panels, namely tuning the half-dumps in the ALPHA-g trap and comparing the signal on the TPC to those simulated on the interconnect, much like the experiments performed on the DSAT stick (Sec. 5.2.3).
- In order to truly measure the effectiveness of using machine learning to reconstruct the profiles recorded from the PS-BGI-82 (as in Chapter 7), a better dataset is required. Another profile monitor must be installed in the same ring as the PS-BGI-82 to collect this data. As this seems unlikely to happen in the near future, any other ring with two profile monitors (one of which is an IPM) is an ideal target for collecting this data and verifying the methods used in this thesis.
- More time could also be spent trying to match simulations to the real data taken from the PS-BGI-82 and train models on the simulated data instead of training on real data. Although this did not show much promise in the limited initial trials, it would be beneficial not to require two IPMs per ring in order to train these models.
- Now that the TOF calibration for the ALPHA-g BV has been completed, and the variables are available to be used, a new MVA needs to be trained and analysed using the methods described in this thesis. An attempt at this

has been made, resulting in a model with ~ 0.0155Hz cosmic pass-cut rate and a signal efficiency of $\sim 61.8\%$ when using the same FOM as used in the gravity paper [17]. However, these results need to be verified and properly evaluated before being used in any results for the 2024 physics run. The ROC curve of this model compared with the ROC curve for the standard ALPHA-g model can be seen in Figure 11.2. In this plot the ROC curve units have been converted to cosmic pass cut rate, and signal trigger efficiency via evaluation on a validation set.

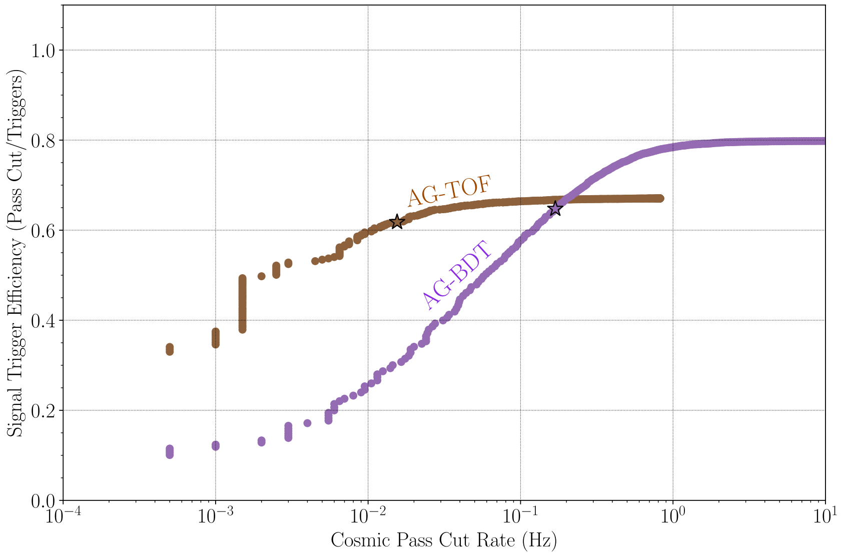


Figure 11.2: ROC curve for the ALPHA-g BDT and new ALPHA-g model with TOF variables.

The results show significant improvement on the previously used model; however, improvements to the reconstruction algorithm also contribute to this increase in performance (as well as to the reduction in signal efficiency). Further, statistics limit the exact nature of the curve at low background. More testing on this new model is required, as well as quantifying the improvement resulting from the changes to the reconstruction algorithm over the inclusion of TOF variables in the model.

- Data from the other half of the ALPHA-g detector would be instrumental in ensuring the independence of these models to z in all regions of the trap. As of 2024, the “top [PM] trap” has been installed, and it is now possible to trap particles in this region and, in theory, perform mixing to collect this data. Performing this test is a primary goal for the remainder of the 2024 physics run, meaning this data could be available before the end of 2024.

- In order to improve the accuracy of the deep learning models in ALPHA (described in Chapter 10), it would be beneficial to collect a more pure dataset from all regions of the ALPHA-2 trap; this could be done by speeding up the slow release experiments such that the rate particles leave the well is enough to saturate the detector.
- It is also thought that CNNs may have strong potential for deep learning in ALPHA, and once the challenges described in Chapter 10 can be overcome, this should be attempted.
- Further, attempting to train a regression model with this same set of deep learning data used for the deep neural network (see Sec. 10.1) would be an interesting study and, if successful, would allow for instant reconstruction and classification of events.
- Finally, given that all data used for machine learning is at least partially contaminated, semi-unsupervised learning methods could be trailed as these could limit the effect of cosmic contamination on the results and potentially increase BDT accuracy overall.

Bibliography

1. Hodgson, D. The First Appearance of Symmetry in the Human Lineage: Where Perception Meets Art. *Symmetry* **3,** 37–53 (2011) (cit. on p. 1).
2. Costa, G. & Fogli, G. *Symmetries and Group Theory in Particle Physics: An Introduction to Space-Time and Internal Symmetries* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012) (cit. on p. 1).
3. Noether, E. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918,** 235–257 (1918) (cit. on p. 1).
4. Dirac, P. A. M. & Fowler, R. H. The Quantum Theory of the Electron. *Proceedings of the Royal Society A* **117,** 610–624 (1928) (cit. on p. 3).
5. Dirac, P. A. M. Quantised Singularities in the Electromagnetic Field, *Proceedings of the Royal Society A* **133,** 60–72 (1931) (cit. on pp. 3, 4).
6. Anderson, C. D. The Positive Electron. *Physical Review* **43,** 491–494 (1933) (cit. on pp. 3, 5).
7. Chamberlain, O. *et al.* Observation of Antiprotons. *Physical Review* **100,** 947–950 (1955) (cit. on p. 4).
8. Chamberlain, O. *et al.* On the Observation of an Antiproton Star in Emulsion Exposed at the Bevatron. *Il Nuovo Cimento* **3,** 447–467 (1956) (cit. on p. 5).
9. Autin, B. *et al.* The Antiproton Decelerator: AD. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **391,** 210–215 (1997) (cit. on p. 4).
10. CERN RESEARCH BOARD. *Minutes of the 130th Meeting of the Research Board Held on Thursday 6 February, 1997* Geneva, Switzerland, 1997 (cit. on p. 4).
11. Amoretti, M. *et al.* Production and Detection of Cold Antihydrogen Atoms. *Nature* **419,** 456–459 (2002) (cit. on pp. 4, 12, 30, 67, 71).

12. Andresen, G. B. *et al.* Trapped Antihydrogen. *Nature* **468,** 673–676 (2010) (cit. on pp. 4, 26).

13. Andresen, G. B. *et al.* Confinement of Antihydrogen for 1,000 Seconds. *Nature Physics* **7,** 558–564 (2011) (cit. on p. 4).

14. Ahmadi, M. *et al.* Characterization of the 1S–2S Transition in Antihydrogen. *Nature* **557,** 71–75 (2018) (cit. on pp. 5, 63).

15. Ahmadi, M. *et al.* Observation of the 1S–2P Lyman-α Transition in Antihydrogen. *Nature* **561,** 211–215 (2018) (cit. on pp. 5, 70).

16. Baker, C. J. *et al.* Laser Cooling of Antihydrogen Atoms. *Nature* **592,** 35–42 (2021) (cit. on pp. 5, 70).

17. Anderson, E. K. *et al.* Observation of the Effect of Gravity on the Motion of Antimatter. *Nature* **621,** 716–722 (2023) (cit. on pp. 5, 8, 62, 65, 94, 170, 179, 190, 201, 205, 217, 218, 221).

18. Einstein, A. Grundgedanken Der Allgemeinen Relativitatstheorie Und Anwendung Dieser Theorie in Der Astronomie (Fundamental Ideas of the General Theory of Relativity and the Application of This Theory in Astronomy). *Preussische Akademie der Wissenschaften, Sitzungsberichte* **315** (1915) (cit. on p. 6).

19. Hughes, R. J. & Deutch, B. I. Electric Charges of Positrons and Antiprotons. *Physical Review Letters* **69,** 578–581 (1992) (cit. on p. 6).

20. Fee, M. S. *et al.* Measurement of the Positronium 1^3S_1–2^3S_1 Interval by Continuous-Wave Two-Photon Excitation. *Physical Review A* **48,** 192–219 (1993) (cit. on p. 6).

21. Van Dyck, R. S., Schwinberg, P. B. & Dehmelt, H. G. New High-Precision Comparison of Electron and Positron g Factors. *Physical Review Letters* **59,** 26–29 (1987) (cit. on p. 6).

22. Hori, M. *et al.* Two-Photon Laser Spectroscopy of Antiprotonic Helium and the Antiproton-to-Electron Mass Ratio. *Nature* **475,** 484–488 (2011) (cit. on p. 6).

23. Kreissl, A. *et al.* Remeasurement of the Magnetic Moment of the Antiproton. *Zeitschrift für Physik C Particles and Fields* **37,** 557–561 (1988) (cit. on p. 6).
24. Lewis, R., Smith, G. & Howe, S. Antiproton Portable Traps and Medical Applications. *Hyperfine Interactions* **109,** 155–164 (1997) (cit. on p. 6).
25. Parthey, C. G. *et al.* Improved Measurement of the Hydrogen $1S - 2S$ Transition Frequency. *Physical Review Letters* **107,** 203001 (2011) (cit. on p. 7).
26. Matveev, A. *et al.* Precision Measurement of the Hydrogen $1S{-}2S$ Frequency via a 920-Km Fiber Link. *Physical Review Letters* **110,** 230801 (2013) (cit. on p. 7).
27. Schwob, C. *et al.* Optical Frequency Measurement of the $2S - 12D$ Transitions in Hydrogen and Deuterium: Rydberg Constant and Lamb Shift Determinations. *Physical Review Letters* **82,** 4960–4963 (1999) (cit. on p. 7).
28. Essen, L. *et al.* Hydrogen Maser Work at the National Physical Laboratory. *Metrologia* **9,** 128 (1973) (cit. on p. 8).
29. Fischler, M., Lykken, J. & Roberts, T. *Direct Observation Limits on Antimatter Gravitation* 2008. arXiv: `0808.3929` (cit. on p. 8).
30. Nieto, M. M. & Goldman, T. The Arguments against “Antigravity” and the Gravitational Acceleration of Antimatter. *Physics Reports* **205,** 221–281 (1991) (cit. on p. 8).
31. Villata, M. CPT Symmetry and Antimatter Gravity in General Relativity. *Europhysics Letters* **94,** 20001 (2011) (cit. on p. 8).
32. Hajdukovic, D. S. Quantum Vacuum and Virtual Gravitational Dipoles: The Solution to the Dark Energy Problem? *Astrophysics and Space Science* **339,** 1–5 (2012) (cit. on p. 8).
33. Benoit-Lévy, A. & Chardin, G. Introducing the Dirac-Milne Universe. *Astronomy & Astrophysics* **537,** A78. arXiv: `1110.3054 [astro-ph]` (2012) (cit. on p. 8).

34. Dimopoulos, C. *et al.* Hubble Law and Acceleration Curve Emerges in a Repulsive Matter-Anti Matter Galaxies Simulations. *Astroparticle Physics* **147,** 102806 (2023) (cit. on p. 8).

35. Baur, G. *et al.* Production of Antihydrogen. *Physics Letters B* **368,** 251–258 (1996) (cit. on pp. 11, 12).

36. Blanford, G. *et al.* Observation of Atomic Antihydrogen. *Physical Review Letters* **80,** 3037–3040 (1998) (cit. on p. 11).

37. Gabrielse, G. *et al.* Background-Free Observation of Cold Antihydrogen with Field-Ionization Analysis of Its States. *Physical Review Letters* **89,** 213401 (2002) (cit. on pp. 12, 67).

38. Amoretti, M *et al.* Antihydrogen Production Temperature Dependence. *Physics Letters B* **583,** 59–67 (2004) (cit. on p. 13).

39. Gabrielse, G. *et al.* First Measurement of the Velocity of Slow Antihydrogen Atoms. *Physical Review Letters* **93,** 073401 (2004) (cit. on p. 13).

40. Amoretti, M. *et al.* Dynamics of Antiproton Cooling in a Positron Plasma during Antihydrogen Formation. *Physics Letters B* **590,** 133–142 (2004) (cit. on p. 13).

41. Fujiwara, M. C. *et al.* Temporally Controlled Modulation of Antihydrogen Production and the Temperature Scaling of Antiproton-Positron Recombination. *Physical Review Letters* **101,** 053401 (2008) (cit. on p. 13).

42. Storry, C. H. *et al.* First Laser-Controlled Antihydrogen Production. *Physical Review Letters* **93,** 263401 (2004) (cit. on pp. 13, 15).

43. Amoretti, M. *et al.* Search for Laser-Induced Formation of Antihydrogen Atoms. *Physical Review Letters* **97,** 213401 (2006) (cit. on pp. 13, 14).

44. Fujiwara, M. C. *et al.* Three-Dimensional Annihilation Imaging of Trapped Antiprotons. *Physical Review Letters* **92,** 065005 (2004) (cit. on p. 13).

45. Madsen, N. *et al.* Spatial Distribution of Cold Antihydrogen Formation. *Physical Review Letters* **94,** 033403 (2005) (cit. on p. 13).

46. Müller, A. & Wolf, A. Production of Antihydrogen by Recombination of $\bar{p}$ with e^+: What Can We Learn from Electron–Ion Collision Studies? *Hyperfine Interactions* **109,** 233–267 (1997) (cit. on p. 13).

47. Gabrielse, G. *et al.* Antihydrogen Production Using Trapped Plasmas. *Physics Letters A* **129,** 38–42 (1988) (cit. on pp. 13, 16).

48. Yousif, F. B. *et al.* Experimental Observation of Laser-Stimulated Radiative Recombination. *Physical Review Letters* **67,** 26–29 (1991) (cit. on p. 14).

49. Scrinzi, A., Elander, N. & Wolf, A. Laser Induced Recombination to Excited States of Hydrogen-like Ions. *Zeitschrift für Physik D Atoms, Molecules and Clusters* **34,** 185–194 (1995) (cit. on p. 14).

50. Blumer, P. *et al.* Positron Accumulation in the GBAR Experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1040,** 167263 (2022) (cit. on p. 15).

51. Wall, M. L., Norton, C. S. & Robicheaux, F. Two-Stage Rydberg Charge Exchange in a Strong Magnetic Field. *Physical Review A* **72,** 052702 (2005) (cit. on p. 15).

52. Robicheaux, F. Atomic Processes in Antihydrogen Experiments: A Theoretical and Computational Perspective. *Journal of Physics B: Atomic, Molecular and Optical Physics* **41,** 192001 (2008) (cit. on pp. 16, 17).

53. Glinsky, M. E. & O'Neil, T. M. Guiding Center Atoms: Three-body Recombination in a Strongly Magnetized Plasma. *Physics of Fluids B: Plasma Physics* **3,** 1279–1293 (1991) (cit. on p. 16).

54. Robicheaux, F. & Hanson, J. D. Three-Body Recombination for Protons Moving in a Strong Magnetic Field. *Physical Review A* **69,** 010701 (2004) (cit. on pp. 16, 17).

55. Robicheaux, F. Three-Body Recombination for Electrons in a Strong Magnetic Field: Magnetic Moment. *Physical Review A* **73,** 033401 (2006) (cit. on p. 17).

56. Jonsell, S. & Charlton, M. On the Formation of Trappable Antihydrogen. *New Journal of Physics* **20,** 043049 (2018) (cit. on pp. 17, 61).
57. Holzscheiter, M. H., Charlton, M. & Nieto, M. M. The Route to Ultra-Low Energy Antihydrogen. *Physics Reports* **402,** 1–101 (2004) (cit. on pp. 17, 43, 44).
58. Earnshaw, S. On the Nature of the Molecular Forces Which Regulate the Constitution of the Luminiferous Ether. *Transactions of the Cambridge Philosophical Society* **7,** 97 (1848) (cit. on p. 18).
59. Dehmelt, H. Experiments with an Isolated Subatomic Particle at Rest. *Reviews of Modern Physics* **62,** 525–530 (1990) (cit. on p. 18).
60. Kretzschmar, M. Particle Motion in a Penning Trap. *European Journal of Physics* **12,** 240 (1991) (cit. on pp. 20, 21).
61. Knoop, M., Madsen, N. & Thompson, R. C. in *Physics with Trapped Charged Particles* 1–24 (Imperial College Press, 2014) (cit. on p. 22).
62. Thompson, R. C. in *Trapped Charged Particles* 1–33 (World Scientific, 2016) (cit. on p. 22).
63. Malmberg, J. H. & deGrassie, J. S. Properties of Nonneutral Plasma. *Physical Review Letters* **35,** 577–580 (1975) (cit. on p. 22).
64. Debye, P. & Hückel, E. Zur Theorie Der Elektrolyte. I. Gefrierpunktserniedrigung Und Verwandte Erscheinungen (The Theory of Electrolytes. I. Freezing Point Depression and Related Phenomena). Trans. by Braus, M. J. *Physikalische Zeitschrift* **24,** 185–206 (1923) (cit. on p. 23).
65. Dubin, D. H. E. in *Trapped Charged Particles* 179–193 (World Scientific, 2016) (cit. on p. 23).
66. O'Neil, T. M. A Confinement Theorem for Nonneutral Plasmas. *The Physics of Fluids* **23,** 2216–2218 (1980) (cit. on p. 24).
67. Dubin, D. H. E. & O'Neil, T. M. Trapped Nonneutral Plasmas, Liquids, and Crystals (the Thermal Equilibrium States). *Reviews of Modern Physics* **71,** 87–172 (1999) (cit. on pp. 24, 55).

68. Majorana, E. Atomi orientati in campo magnetico variabile. *Il Nuovo Cimento (1924-1942)* **9,** 43–50 (1932) (cit. on p. 27).

69. Brink, D. M. & Sukumar, C. V. Majorana Spin-Flip Transitions in a Magnetic Trap. *Physical Review A* **74,** 035401 (2006) (cit. on p. 27).

70. Pérez-Ríos, J. & Sanz, A. S. How Does a Magnetic Trap Work? *American Journal of Physics* **81,** 836–843 (2013) (cit. on p. 27).

71. Bertsche, W. *et al.* A Magnetic Trap for Antihydrogen Confinement. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **566,** 746–756 (2006) (cit. on pp. 27, 29, 63).

72. Regenfus, C. *et al.* Detection of Antihydrogen Annihilations with a Cryogenic Pure-CsI Crystal Detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. Proceedings of the 3rd International Conference on New Developments in Photodetection* **504,** 343–346 (2003) (cit. on pp. 30, 71).

73. Polster, D. *et al.* Light Particle Emission Induced by Stopped Antiprotons in Nuclei: Energy Dissipation and Neutron-to-Proton Ratio. *Physical Review C* **51,** 1167–1180 (1995) (cit. on p. 31).

74. Lubiński, P. *et al.* Gold Fragmentation Induced by Stopped Antiprotons. *Physical Review C* **66,** 044616 (2002) (cit. on p. 31).

75. Klempt, E., Batty, C. & Richard, J.-M. The Antinucleon–Nucleon Interaction at Low Energy: Annihilation Dynamics. *Physics Reports* **413,** 197–317 (2005) (cit. on p. 31).

76. Ghesquière, C. *An Inclusive View on* $\bar{p}p \to n\pi$ *at Rest* in *Symposium on Antinucleon-Nucleon Interactions* (Liblice-Prague, Czech Republic, 1974) (cit. on pp. 31, 32).

77. Orfanidis, S. J. & Rittenberg, V. Nucleon-Antinucleon Annihilation into Pions. *Nuclear Physics B* **59,** 570–582 (1973) (cit. on pp. 31, 32).

78. Andresen, G. B. *et al.* Antiproton, Positron, and Electron Imaging with a Microchannel Plate/Phosphor Detector. *Review of Scientific Instruments* **80** (2009) (cit. on p. 36).
79. Shockley, W. The Theory of $p-n$ Junctions in Semiconductors and $p-n$ Junction Transistors. *Bell System Technical Journal* **28,** 435–489 (1949) (cit. on pp. 38, 79).
80. Eggleston, D. L. *et al.* Parallel Energy Analyzer for Pure Electron Plasma Devices. *Physics of Fluids B: Plasma Physics* **4,** 3432–3439 (1992) (cit. on p. 39).
81. Beck, B. R. *Measurement of the Magnetic and Temperature Dependence of the Electron-Electron Anisotropic Temperature Relaxation Rate* PhD thesis (University of California, San Diego, USA, 1990) (cit. on p. 41).
82. Andresen, G. B. *Evaporative Cooling of Antiprotons and Efforts to Trap Antihydrogen* PhD thesis (Aarhus University, Denmark, 2010) (cit. on p. 41).
83. Möhring, H. J. & Ranft, J. Antiproton Production from Extended Targets Using a Weighted Monte Carlo Hadron Cascade Model. *Nuclear Instruments and Methods in Physics Research* **201,** 323–327 (1982) (cit. on p. 41).
84. Hojvat, C. & van Ginneken, A. Calculation of Antiproton Yields for the Fermilab Antiproton Source. *Nuclear Instruments and Methods in Physics Research* **206,** 67–83 (1983) (cit. on p. 41).
85. Johnson, C. D. Present and Future Possibilities of Anti-Proton Production from Fixed Targets at CERN. *Hyperfine Interactions* **44** (eds Poth, H. & Wolf, A.) 21–30 (1989) (cit. on p. 41).
86. Baird, S. A. *et al. Design Study of the Antiproton Decelerator: AD* tech. rep. CERN-PS-96-043-AR (CERN, 1996) (cit. on p. 42).
87. Möhl, D. Production of Low-Energy Antiprotons. *Hyperfine Interactions* **109,** 33–41 (1997) (cit. on p. 42).
88. Charlton, M. & Humberston, J. W. *Positron Physics* (Cambridge University Press, 2000) (cit. on p. 43).

89. Murphy, T. J. & Surko, C. M. Positron Trapping in an Electrostatic Well by Inelastic Collisions with Nitrogen Molecules. *Physical Review A* **46,** 5696–5705 (1992) (cit. on p. 44).

90. Greaves, R. G., Tinkle, M. D. & Surko, C. M. Creation and Uses of Positron Plasmas. *Physics of Plasmas* **1,** 1439–1446 (1994) (cit. on p. 44).

91. Danielson, J. R. *et al.* Plasma and Trap-Based Techniques for Science with Positrons. *Reviews of Modern Physics* **87,** 247–306 (2015) (cit. on p. 44).

92. Amole, C. *et al.* The ALPHA Antihydrogen Trapping Apparatus. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **735,** 319–340 (2014) (cit. on p. 45).

93. Jørgensen, L. V. *et al.* New Source of Dense, Cryogenic Positron Plasmas. *Physical Review Letters* **95,** 025002 (2005) (cit. on p. 46).

94. Möhl, D. & Sessler, A. M. Beam Cooling: Principles and Achievements. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. International Workshop on Beam Cooling and Related Topics* **532,** 1–10 (2004) (cit. on pp. 46, 55).

95. van der Meer, S. *Stochastic Damping of Betatron Oscillations in the ISR* Geneva, 1972 (cit. on p. 46).

96. Möhl, D. *et al.* Physics and Technique of Stochastic Cooling. *Physics Reports* **58,** 73–102 (1980) (cit. on p. 46).

97. Budker, G. I. An Effective Method of Damping Particle Oscillations in Proton and Antiproton Storage Rings. *Atomic Energy* **22,** 438–440 (1967) (cit. on p. 47).

98. Poth, H. Electron Cooling: Theory, Experiment, Application. *Physics Reports* **196,** 135–297 (1990) (cit. on p. 47).

99. Hori, M. & Walz, J. Physics at CERN's Antiproton Decelerator. *Progress in Particle and Nuclear Physics* **72,** 206–253 (2013) (cit. on p. 48).

100. Gamba, D. *et al. AD/ELENA Electron Cooling Experience During and after CERNs Long Shutdown (LS2)* in *International Workshop on Beam Cooling and Related Topics* **2021** (Novosibirsk, Russia, 2021), 36–41 (cit. on pp. 48, 49).

101. Chohan, V. *et al. Extra Low ENergy Antiproton (ELENA) Ring and Its Transfer Lines: Design Report* tech. rep. CERN-2014-002 (CERN, 2014) (cit. on p. 49).

102. Friesen, T. P. *Probing Trapped Antihydrogen. In Situ Diagnostics and Observations of Quantum Transitions* PhD thesis (University of Calgary, Canada, 2014) (cit. on p. 51).

103. Fabbri, S. S. *Optimization of Antiproton Capture and Delivery for the ALPHA Antihydrogen Experiment* PhD thesis (The University of Manchester, England, 2021) (cit. on p. 51).

104. Granum, P. *Measuring the Properties of Antihydrogen* PhD thesis (Aarhus University, Denmark, 2022) (cit. on p. 52).

105. Larmor, J. LXIII. On the Theory of the Magnetic Influence on Spectra; and on the Radiation from Moving Ions. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44,** 503–512 (1897) (cit. on p. 51).

106. Brown, L. S. & Gabrielse, G. Geonium Theory: Physics of a Single Electron or Ion in a Penning Trap. *Reviews of Modern Physics* **58,** 233–311 (1986) (cit. on p. 51).

107. Spitzer, L. *Physics of Fully Ionized Gases* (Courier Corporation, 2006) (cit. on p. 52).

108. Amoretti, M. *et al.* The ATHENA Antihydrogen Apparatus. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **518,** 679–711 (2004) (cit. on p. 52).

109. Anderegg, F., Hollmann, E. M. & Driscoll, C. F. Rotating Field Confinement of Pure Electron Plasmas Using Trivelpiece-Gould Modes. *Physical Review Letters* **81,** 4875–4878 (1998) (cit. on p. 55).

110. Greaves, R. G. & Moxom, J. M. Compression of Trapped Positrons in a Single Particle Regime by a Rotating Electric Field. *Physics of Plasmas* **15,** 072304 (2008) (cit. on p. 55).

111. Huang, X.-P. *et al.* Steady-State Confinement of Non-neutral Plasmas by Rotating Electric Fields. *Physical Review Letters* **78,** 875–878 (1997) (cit. on p. 55).

112. Huang, X.-P. *et al.* Phase-Locked Rotation of Crystallized Non-neutral Plasmas by Rotating Electric Fields. *Physical Review Letters* **80,** 73–76 (1998) (cit. on p. 55).

113. Isaac, C. A. *et al.* Compression of Positron Clouds in the Independent Particle Regime. *Physical Review Letters* **107,** 033201 (2011) (cit. on p. 55).

114. Danielson, J. R. & Surko, C. M. Torque-Balanced High-Density Steady States of Single-Component Plasmas. *Physical Review Letters* **94,** 035001 (2005) (cit. on p. 55).

115. Danielson, J. R. & Surko, C. M. Radial Compression and Torque-Balanced Steady States of Single-Component Plasmas in Penning-Malmberg Trapsa). *Physics of Plasmas* **13,** 055706 (2006) (cit. on p. 55).

116. Leanhardt, A. E. *et al.* Cooling Bose-Einstein Condensates Below 500 Picokelvin. *Science* **301,** 1513–1515 (2003) (cit. on p. 58).

117. Carruth, C. *Methods for Plasma Stabilization and Control to Improve Antihydrogen Production* PhD thesis (University of California, Berkeley, 2018) (cit. on p. 59).

118. Andresen, G. B. *et al.* Evaporative Cooling of Antiprotons to Cryogenic Temperatures. *Physical Review Letters* **105,** 013003 (2010) (cit. on p. 59).

119. Butler, E. *et al.* Trapped Antihydrogen. *Hyperfine Interactions* **212,** 15–29 (2012) (cit. on pp. 59, 84).

120. Ahmadi, M. *et al.* Antihydrogen Accumulation for Fundamental Symmetry Tests. *Nature Communications* **8,** 681 (2017) (cit. on pp. 61, 68, 69).

121. Schmöger, L. *et al.* Coulomb Crystallization of Highly Charged Ions. *Science* **347,** 1233–1236 (2015) (cit. on p. 61).

122. Barrett, M. D. *et al.* Sympathetic Cooling of $^9\text{Be}^+$ and $^{24}\text{Mg}^+$ for Quantum Logic. *Physical Review A* **68,** 042302 (2003) (cit. on p. 61).

123. Sameed, M., Maxwell, D. & Madsen, N. Ion Generation and Loading of a Penning Trap Using Pulsed Laser Ablation. *New Journal of Physics* **22,** 013009 (2020) (cit. on p. 61).

124. Baker, C. J. *et al.* Sympathetic Cooling of Positrons to Cryogenic Temperatures for Antihydrogen Production. *Nature Communications* **12,** 6139 (2021) (cit. on pp. 61, 62).

125. So, C., Fajans, J. & Bertsche, W. The ALPHA-g Antihydrogen Gravity Magnet System. *IEEE Transactions on Applied Superconductivity* **30,** 1–5 (2020) (cit. on p. 64).

126. Baker, C. J. *et al.* Design and Performance of a Novel Low Energy Multispecies Beamline for an Antihydrogen Experiment. *Physical Review Accelerators and Beams* **26,** 040101 (2023) (cit. on pp. 66, 67).

127. Tajima, M. *et al.* Antiproton Beams with Low Energy Spread for Antihydrogen Production. *Journal of Instrumentation* **14,** P05009 (2019) (cit. on p. 66).

128. Enomoto, Y. *et al.* Synthesis of Cold Antihydrogen in a Cusp Trap. *Physical Review Letters* **105,** 243401 (2010) (cit. on p. 67).

129. Amole, C. *et al.* Experimental and Computational Study of the Injection of Antiprotons into a Positron Plasma for Antihydrogen Production. *Physics of Plasmas* **20,** 043510 (2013) (cit. on pp. 67, 75).

130. So, C. *Antiproton and Positron Dynamics in Antihydrogen Production* PhD thesis (University of California, Berkeley, 2014) (cit. on p. 68).

131. Capra, A. & ALPHA collaboration. Lifetime of Magnetically Trapped Antihydrogen in ALPHA. *Hyperfine Interactions* **240,** 9 (2019) (cit. on p. 69).

132. Ashkin, A. Acceleration and Trapping of Particles by Radiation Pressure. *Physical Review Letters* **24,** 156–159 (1970) (cit. on p. 70).

133. Hänsch, T. W. & Schawlow, A. L. Cooling of Gases by Laser Radiation. *Optics Communications* **13,** 68–69 (1975) (cit. on p. 70).

134. Wineland, D. J., Drullinger, R. E. & Walls, F. L. Radiation-Pressure Cooling of Bound Resonant Absorbers. *Physical Review Letters* **40,** 1639–1642 (1978) (cit. on p. 70).

135. Neuhauser, W. *et al.* Optical-Sideband Cooling of Visible Atom Cloud Confined in Parabolic Well. *Physical Review Letters* **41,** 233–236 (1978) (cit. on p. 70).

136. Setija, I. D. *et al.* Optical Cooling of Atomic Hydrogen in a Magnetic Trap. *Physical Review Letters* **70,** 2257–2260 (1993) (cit. on p. 70).

137. Knoll, G. F. *Radiation Detection and Measurement* 2nd ed. (John Wiley & Sons, 1989) (cit. on pp. 75, 90).

138. Bertolini, G. & Coche, A. *Semiconductor Detectors* (Interscience (Wiley), 1968) (cit. on p. 75).

139. Spieler, H. *Semiconductor Detector Systems* (Oxford University Press, 2005) (cit. on p. 76).

140. Sigmund, P. *Particle Penetration and Radiation Effects Volume 2: Penetration of Atomic and Molecular Ions* (Springer Cham, 2014) (cit. on p. 82).

141. Bethe, H. A. Molière's Theory of Multiple Scattering. *Physical Review* **89,** 1256–1266 (1953) (cit. on p. 82).

142. Scott, W. T. The Theory of Small-Angle Multiple Scattering of Fast Charged Particles. *Reviews of Modern Physics* **35,** 231–313 (1963) (cit. on p. 82).

143. Motz, J. W., Olsen, H. & Koch, H. W. Electron Scattering without Atomic or Nuclear Excitation. *Reviews of Modern Physics* **36,** 881–928 (1964) (cit. on p. 82).

144. Nakamura, K., Amsler, C. & Group, P. D. Particle Physics Booklet. *Journal of Physics G: Nuclear and Particle Physics* **37,** 075021 (2010) (cit. on pp. 82, 166).

145. Burger, P. Manufacturing, Performance and New Developments in Silicon Counters in High Energy Physics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **226,** 112–116 (1984) (cit. on p. 83).

146. Hydomako, R. *Detection of Trapped Antihydrogen* PhD thesis (University of Calgary, Canada, 2011) (cit. on pp. 83, 86).

147. Llopart, X. *et al.* Timepix, a 65k Programmable Pixel Readout Chip for Arrival Time, Energy and/or Photon Counting Measurements. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **581,** 485–494 (2007) (cit. on p. 83).

148. Rossi, L. *et al. Pixel Detectors: From Fundamentals to Applications* (Springer Berlin, Heidelberg, 2006) (cit. on p. 83).

149. Haensel, S. *et al. Resolution Studies on Silicon Strip Sensors with Fine Pitch* in *International Linear Collider Workshop (LCWS08 and ILC08)* (Chicago, USA, 2009) (cit. on p. 83).

150. Krammer, M & Pernegger, H. Signal Collection and Position Reconstruction of Silicon Strip Detectors with 200μm Readout Pitch. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **397,** 232–242 (1997) (cit. on p. 83).

151. Stavitski, I. The Silicon Strip Detector with Very Large Pitch: Conceptual Design and Simulation Results. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. Proceedings of the 9th International Workshop on Vertex Detectors* **473,** 186–191 (2001) (cit. on p. 83).

152. Amole, C. *et al.* Silicon Vertex Detector Upgrade in the ALPHA Experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. Vienna Conference on Instrumentation 2013* **732,** 134–136 (2013) (cit. on p. 85).

153. Andresen, G. *et al.* Antihydrogen Annihilation Reconstruction with the ALPHA Silicon Detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **684,** 73–81 (2012) (cit. on pp. 87, 89, 101, 119, 166, 195, 207).

154. Rutherford, E. & Geiger, H. An Electrical Method of Counting the Number of α-Particles from Radio-Active Substances. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **81,** 141–161 (1908) (cit. on p. 90).

155. Almond, P. R. & Svensson, H. Ionization Chamber Dosimetry for Photon and Electron Beams. *Acta Radiologica: Therapy, Physics, Biology* **16,** 177–186 (1977) (cit. on p. 90).

156. DeWerd, L. A. & Smith, B. R. in *Radiation Therapy Dosimetry* 19–30 (CRC Press, 2021) (cit. on p. 90).

157. Kleinknecht, K. *Detectors for Particle Radiation* (Cambridge University Press, 1998) (cit. on pp. 91, 94).

158. Wilkinson, D. H. *Ionization Chambers and Counters* (New York: Cambridge Univ. Press, 1950) (cit. on p. 91).

159. Capra, A. *Testing CPT and Antigravity with Trapped Antihydrogen at ALPHA* PhD thesis (York University, Canada, 2015) (cit. on pp. 92, 94, 99, 101).

160. Olive, K. A. Review of Particle Physics. *Chinese Physics C* **38,** 090001 (2014) (cit. on p. 94).

161. Lobanov, Yu. Yu. & Zhidkov, E. P. *Programming and Mathematical Techniques in Physics* in *Proceedings of the International Conference on Programming and Mathematical Methods for Solving Physical Problems* (WORLD SCIENTIFIC, Dubna, Russia, 1994), 1–324 (cit. on p. 97).

162. Brun, R. & Rademakers, F. ROOT — An Object Oriented Data Analysis Framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. New Computing Techniques in Physics Research V* **389,** 81–86 (1997) (cit. on p. 102).

163. Hřivnáčová, I. *The Virtual Monte Carlo* in *2003 Conference for Computing in High-Energy and Nuclear Physics* (La Jolla, USA, 2003) (cit. on p. 102).

164. Agostinelli, S. *et al.* GEANT4—a Simulation Toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506,** 250–303 (2003) (cit. on p. 102).

165. Allison, J. *et al.* Geant4 Developments and Applications. *IEEE Transactions on nuclear science* **53,** 270–278 (2006) (cit. on p. 102).

166. Hřivnáčová, I. *The Virtual Geometry Model* in *24th International Conference on Computing in High Energy and Nuclear Physics* **245** (EDP Sciences, Adelaide, Australia, 2020), 02014 (cit. on p. 102).

167. Blomer, J. *et al.* *Evolution of the ROOT Tree I/O* in *24th International Conference on Computing in High Energy and Nuclear Physics* **245** (EDP Sciences, Adelaide, Australia, 2020), 02030 (cit. on pp. 102, 175).

168. Baker, C. J. *et al.* Measurements of Penning-Malmberg Trap Patch Potentials and Associated Performance Degradation. *Physical Review Research* **6,** L012008 (2024) (cit. on pp. 105, 109).

169. Hidalgo, B. & Goodman, M. Multivariate or Multivariable Regression? *American Journal of Public Health* **103,** 39–40 (2013) (cit. on p. 121).

170. Rencher, A. C. & Christensen, W. F. *Methods of Multivariate Analysis* (John Wiley & Sons, 2012) (cit. on p. 121).

171. Domingos, P. A Few Useful Things to Know about Machine Learning. *Communications of the ACM* **55,** 78–87 (2012) (cit. on p. 124).

172. Aggarwal, C. C., Hinneburg, A. & Keim, D. A. *On the Surprising Behavior of Distance Metrics in High Dimensional Space* in *Database Theory — ICDT 2001* **1973** (London, UK, 2001), 420–434 (cit. on p. 124).

173. Mirkes, E. M., Allohibi, J. & Gorban, A. Fractional Norms and Quasinorms Do Not Help to Overcome the Curse of Dimensionality. *Entropy* **22,** 1105 (2020) (cit. on p. 124).

174. Fix, E. & Hodges, J. L. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review / Revue Internationale de Statistique* **57,** 238–247 (1989) (cit. on p. 126).

175. Cover, T. & Hart, P. Nearest Neighbor Pattern Classification. *IEEE transactions on information theory* **13,** 21–27 (1967) (cit. on p. 126).

176. Beyer, K. *et al. When Is "Nearest Neighbor" Meaningful?* in *Database Theory — ICDT'99* (eds Goos, G. *et al.*) **1540** (Springer Berlin Heidelberg, Jerusalem, Israel, 1999), 217–235 (cit. on p. 127).

177. Kearns, M. & Valiant, L. Cryptographic Limitations on Learning Boolean Formulae and Finite Automata. *Journal of the ACM* **41,** 67–95 (1994) (cit. on p. 129).

178. Schapire, R. E. The Strength of Weak Learnability. *Machine Learning* **5,** 197–227 (1990) (cit. on p. 129).

179. Freund, Y. & Schapire, R. E. A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting. *Journal of computer and system sciences* **55,** 119–139 (1997) (cit. on p. 129).

180. Breiman, L. Bagging Predictors. *Machine Learning* **24,** 123–140 (1996) (cit. on p. 129).

181. Ho, T. K. *Random Decision Forests* in *Proceedings of 3rd International Conference on Document Analysis and Recognition* **1** (IEEE, Montreal, Canada, 1995), 278–282 (cit. on p. 129).

182. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer New York, New York, 2006) (cit. on p. 133).

183. Theodoridis, S. *Machine Learning: A Bayesian and Optimization Perspective* 1st ed. (Academic press, 2015) (cit. on p. 134).

184. Schölkopf, B., Herbrich, R. & Smola, A. J. *A Generalized Representer Theorem* in *14th Annual Conference on Computational Learning Theory, COLT 2001 and 5th European Conference on Computational Learning Theory, EuroCOLT 2001* (eds Goos, G. *et al.*) **2111** (Springer Berlin Heidelberg, Amsterdam, The Netherlands, 2001), 416–426 (cit. on p. 134).

185. Drucker, H. *et al. Support Vector Regression Machines* in *Proceedings of the 9th International Conference on Neural Information Processing Systems* **9** (MIT Press, Denver, Colorado, 1996), 155–161 (cit. on p. 135).

186. Hornik, K., Stinchcombe, M. & White, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* **2,** 359–366 (1989) (cit. on p. 138).

187. Goodfellow, I. J., Vinyals, O. & Saxe, A. M. *Qualitatively Characterizing Neural Network Optimization Problems* 2015. arXiv: `1412.6544` (cit. on p. 140).

188. Dauphin, Y. N. *et al. Identifying and Attacking the Saddle Point Problem in High-Dimensional Non-Convex Optimization* in *Advances in Neural Information Processing Systems 27* **27** (Montréal, Canada, 2014) (cit. on p. 140).

189. Kawaguchi, K. *Deep Learning without Poor Local Minima* in *Advances in Neural Information Processing Systems 29* **29** (Barcelona, Spain, 2016) (cit. on p. 140).

190. Dreyfus, S. The Numerical Solution of Variational Problems. *Journal of Mathematical Analysis and Applications* **5,** 30–45 (1962) (cit. on p. 141).

191. Mizutani, E., Dreyfus, S. E. & Nishio, K. *On Derivation of MLP Backpropagation from the Kelley-Bryson Optimal-Control Gradient Formula and Its Application* in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium* **2** (IEEE, Como, Italy, 2000), 167–172 (cit. on p. 141).

192. Dreyfus, S. E. Artificial Neural Networks, Back Propagation, and the Kelley-Bryson Gradient Procedure. *Journal of Guidance, Control, and Dynamics* **13,** 926–928 (1990) (cit. on p. 141).

193. Sapinski, M. *et al. Application of Machine Learning for the IPM-Based Profile Reconstruction* in *Proceedings of the 61st ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams* **61** (Daejeon, Korea, 2018), 410–415 (cit. on pp. 145, 150, 152).

194. Vilsmeier, D *et al.* Reconstructing Space-Charge Distorted IPM Profiles with Machine Learning Algorithms. *Journal of Physics: Conference Series* **1067,** 072003 (2018) (cit. on pp. 145, 150, 152, 157).

195. Vilsmeier, D., Sapinski, M. & Singh, R. Space-Charge Distortion of Transverse Profiles Measured by Electron-Based Ionization Profile Monitors and Correction Methods. *Physical Review Accelerators and Beams* **22,** 052801 (2019) (cit. on pp. 145, 148, 150, 157, 164).

196. Gwosch, K. *et al.* Non-Invasive Monitoring of Therapeutic Carbon Ion Beams in a Homogeneous Phantom by Tracking of Secondary Ions. *Physics in Medicine & Biology* **58,** 3755 (2013) (cit. on p. 146).

197. Reinhart, A. M. *et al.* Three Dimensional Reconstruction of Therapeutic Carbon Ion Beams in Phantoms Using Single Secondary Ion Tracks. *Physics in Medicine & Biology* **62,** 4884 (2017) (cit. on p. 146).

198. Fred Hornstra, Jr. & DeLuca, W. H. *Nondestructive Beam Profile Detection Systems For The Zero Gradient Synchrotron* in *6th International Conference on High-Energy Accelerators* (Cambridge, USA, 1967), 374–377 (cit. on p. 146).

199. Weisberg, H. *et al.* An Ionization Profile Monitor for the Brookhaven AGS. *IEEE Transactions on Nuclear Science* **30,** 2179–2181 (1983) (cit. on pp. 146, 148, 149).

200. Moore, R. S., Jansson, A. & Shiltsev, V. Beam Instrumentation for the Tevatron Collider. *Journal of Instrumentation* **4,** P12018 (2009) (cit. on p. 146).

201. Benedetti, F. *et al.* Design and Development of Ionization Profile Monitor for the Cryogenic Sections of the ESS Linac. *EPJ Web of Conferences* **225,** 01009 (2020) (cit. on p. 146).

202. Shiltsev, V. Space-Charge Effects in Ionization Beam Profile Monitors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **986,** 164744 (2021) (cit. on pp. 146, 147, 149).

203. Graves, W. S. Error Analysis and Correction in a Beam Profile Monitor. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **364,** 19–26 (1995) (cit. on pp. 148, 149).

204. Thern, R. E. *Space Charge Distortion in the Brookhaven Ionization Profile Monitor* in *Particle Accelerator Conference* (eds Lindstrom, E. R. & Taylor, L. S.) **870316** (Washington, USA, 1987) (cit. on pp. 148, 149).

205. Amundson, J. *et al.* Calibration of the Fermilab Booster Ionization Profile Monitor. *Physical Review Special Topics - Accelerators and Beams* **6,** 102801 (2003) (cit. on pp. 148, 149).

206. Vilsmeier, D., Dehning, B. & Sapinski, M. *Investigation of the Effect of Beam Space-charge on Electron Trajectories in Ionization Profile Monitors* in *54th ICFA Advanced Beam Dynamics Workshop on High-Intensity, High Brightness and High Power Hadron Beams* (East Lansing, USA, 2015) (cit. on pp. 149, 150).

207. Vilsmeier, D. *Profile Distortion by Beam Space-Charge in Ionization Profile Monitors* Bachelor Thesis (Universität Regensburg, Germany, 2015) (cit. on p. 149).

208. Patecki, M. *et al. Electron Tracking Simulations in the Presence of the Beam and External Fields* in *4th International Particle Accelerator Conference* (Shanghai, China, 2013) (cit. on p. 149).

209. Patecki, M. *Analysis of LHC Beam Gas Ionization Monitor Data and Simulation of the Electron Transport in the Detector* MA thesis (Warsaw U. of Tech, Poland, 2013) (cit. on p. 149).

210. Abada, A. *et al.* HE-LHC: The High-Energy Large Hadron Collider. *The European Physical Journal Special Topics* **228,** 1109–1382 (2019) (cit. on p. 150).

211. Egberts, J. *IFMIF-LIPAc Beam Diagnostics: Profiling and Loss Monitoring Systems* PhD thesis (Université Paris Sud, France, 2012) (cit. on p. 150).

212. Sandberg, H. *et al. Commissioning of Timepix3 Based Beam Gas Ionisation Profile Monitors for the CERN Proton Synchrotron* in *10th International Beam Instrumentation Conference* (Pohang, South Korea, 2021) (cit. on p. 151).

213. Poikela, T. *et al.* Timepix3: A 65K Channel Hybrid Pixel Readout Chip with Simultaneous ToA/ToT and Sparse Readout. *Journal of Instrumentation* **9,** C05013 (2014) (cit. on p. 151).

214. Giacomini, T. *et al. Ionization Profile Monitors - IPM @ GSI* in *Proceedings of the 10th European Workshop on Beam Diagnostics and Instrumentation for Particle Accelerators* (Hamburg, Germany., 2011), 419–421 (cit. on p. 151).

215. Singh, R., Sapinski, M. & Vilsmeier, D. *Simulation Supported Profile Reconstruction With Machine Learning* in *6th International Beam Instrumentation Conference* **IBIC2017** (Grand Rapids, USA, 2018) (cit. on p. 152).

216. Vilsmeier, D., Forck, P. & Sapinski, M. *A Modular Application for IPM Simulations* in *6th International Beam Instrumentation Conference* **IBIC2017** (Grand Rapids, USA, 2018) (cit. on pp. 152, 164).

217. Cappi, R *et al. High-Density and High-Intensity Beams at CERN PS* in *20th ICFA Advanced Beam Dynamics Workshop on High Intensity and High Brightness Hadron Beams* (Batavia, USA, 2002) (cit. on pp. 152, 161).

218. Pedregosa, F. *et al.* Scikit-Learn: Machine Learning in Python. *The Journal of Machine Learning Research* **12,** 2825–2830 (2011) (cit. on pp. 152, 176).

219. Abadi, M. *et al. TensorFlow: A System for Large-Scale Machine Learning* in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation* (USENIX Association, Savannah, USA, 2016) (cit. on p. 152).

220. Meshgi, K. & Ishii, S. *Expanding Histogram of Colors with Gridding to Improve Tracking Accuracy* in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)* (IEEE, Tokyo, Japan, 2015), 475–479 (cit. on p. 154).

221. Gupta, S. & Gupta, A. Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review. *Procedia Computer Science. The Fifth Information Systems International Conference, 23-24 July 2019, Surabaya, Indonesia* **161,** 466–474 (2019) (cit. on p. 157).

222. Tompson, J. *et al. Accelerating Eulerian Fluid Simulation with Convolutional Networks* in *International Conference on Machine Learning* (PMLR, 2017), 3424–3433 (cit. on p. 163).

223. von Rueden, L. *et al. Combining Machine Learning and Simulation to a Hybrid Modelling Approach: Current and Future Directions* in *Advances in Intelligent Data Analysis XVIII* (eds Berthold, M. R., Feelders, A. & Krempl, G.) (Springer International Publishing, Konstanz, Germany, 2020), 548–560 (cit. on p. 163).

224. Donoghue, J. F., Low, F. E. & Pi, S.-Y. Tensor Analysis of Hadronic Jets in Quantum Chromodynamics. *Physical Review D* **20,** 2759–2762 (1979) (cit. on p. 168).

225. Ellis, R. K., Ross, D. A. & Terrano, A. E. The Perturbative Calculation of Jet Structure in e^+e^- Annihilation. *Nuclear Physics B* **178,** 421–456 (1981) (cit. on p. 168).

226. Sas, M. & Schoppink, J. Event Shapes and Jets in e^+e^- and *pp* Collisions. *Nuclear Physics A* **1011,** 122195 (2021) (cit. on p. 168).

227. Hoecker, A. *et al. TMVA - Toolkit for Multivariate Data Analysis* 2009. arXiv: `physics/0703039` (cit. on pp. 174, 176, 208).

228. Punzi, G. *Sensitivity of Searches for New Signals and Its Optimization* 2003. arXiv: `physics/0308063` (cit. on pp. 186, 187).

229. Urioni, M. *Measurement of Earth's Gravitational Acceleration on Anti-Hydrogen with the ALPHA Experiment at CERN* PhD thesis (Università degli Studi di Brescia, Italy, 2024) (cit. on p. 206).

230. Ioffe, S. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* in *Proceedings of the 32nd International Conference on Machine Learning* **37** (Lille, France, 2015), 448–456 (cit. on p. 209).

231. Glorot, X. & Bengio, Y. *Understanding the Difficulty of Training Deep Feedforward Neural Networks* in *13th International Conference on Artificial Intelligence and Statistics* **9** (JMLR Workshop and Conference Proceedings, Sardinia, Italy, 2010), 249–256 (cit. on p. 209).

232. Avery, P. *Applied Fitting Theory IV: Formulas for Track Fitting* tech. rep. CLEO CBX 92-45 (1992) (cit. on p. 248).

233. Avery, P. *Applied Fitting Theory V: Track Fitting Using the Kalman Filter* tech. rep. CLEO Note CBX92-39 (1992) (cit. on p. 248).

234. Delgado, A. & Thaler, J. Quantum Annealing for Jet Clustering with Thrust. *Physical Review D* **106,** 094016 (2022) (cit. on p. 254).

Appendices

Appendix A

Helix Parametrisation

The trajectory of a charged particle in a magnetic field, excluding any external mechanisms to induce scattering or energy loss, is generally helical, consisting of circular motion in the (x, y) plane, combined with a constant velocity along the z axis. The helix can be described by the five parameters $\alpha = (c, \phi_0, D, \lambda, z_0)$ where c is $1/2R$ where R is the radius of curvature, ϕ_0 is the ϕ coordinate of the momentum at the $r - \phi$ point closest to the origin, D is the signed distance of closest approach, $\lambda = \cot\theta$ where θ is the polar angle measured from the positive side of the z axis, and z_0 is the z position at the point closest to the origin. These are often split into two sets, the *radial* parameters (c, ϕ_0, D) which describe the circular motion in $x{-}y$, and the *axial* parameters (λ, z_0) which describe the motion in the z axis. The resulting equations of motion (in Cartesian coordinates) is then given by [232]:

$$f_x = x_0 + \frac{p_{0x}}{a}\sin(2cs_\perp) - \frac{p_{0y}}{a}(1 - \cos(\rho s_\perp)) \tag{A.1}$$

$$f_y = y_0 + \frac{p_{0y}}{a}\sin(2cs_\perp) + \frac{p_{0x}}{a}(1 - \cos(\rho s_\perp)) \tag{A.2}$$

$$f_z = z_0 + \lambda s_\perp \tag{A.3}$$

where $(x_0, y_0, z_0) = (-D\sin\phi_0, D\cos\phi_0, z_0)$ is the point on the helix closest to the origin, (p_{0x}, p_{0y}) is the momentum vector at that point and can be expressed as $(\frac{a}{2c}\cos\phi_0, \frac{a}{2c}\sin\phi_0)$ (here $a = 0.2998Bq$ - B = magnetic field; q = particle charge), and $s_\perp$ is the arc length in the (x, y) plane from (x_0, y_0, z_0) to the point (f_x, f_y, f_z) we are trying to find.

The radial parameters of the helix are found by minimising [233]

$$\chi_r^2 \equiv \sum_{j=1}^{N} \frac{[x_j - f_x]^2}{\sigma_{x_j}^2} + \frac{[y_j - f_y]^2}{\sigma_{y_j}^2}, \tag{A.4}$$

and the axial parameters by minimising

$$\chi_x^2 \equiv \sum_{j=1}^{N} \frac{[z_j - f_z]^2}{\sigma_{z_j}^2}. \tag{A.5}$$

The sum in both these equations is over the number of points measured in the track (i.e. number of hits) and the j-th hit is given by (x_j, y_j, z_j).

The linear momentum of the particle (again assuming no scattering or energy loss mechanisms) in MeV/c is given by

$$p_x = \frac{a}{2c} \cos\phi_0 \tag{A.6}$$

$$p_y = \frac{a}{2c} \sin\phi_0 \tag{A.7}$$

$$p_z = \frac{a}{2c} \lambda \tag{A.8}$$

Appendix B

Data Dumper

A considerable amount of time has gone into the standardisation and modernisation of the pipeline for dumping data for ML analysis. This appendix describes the pipeline for both the ALPHA-2 and ALPHA-g reconstruction algorithms.

A diagram in the form of a flow chart of the full ALPHA-2 ML pipeline can be seen in Figure B.1. Though ALPHA-g is not pictured, it is synonymous.

The pipeline can be run manually (as each program can be individually run), but there is also a script designed to run the process automatically on the `htcondor` cluster at CERN. The runlist and eventlist must always be created manually, and must be done prior to using the automatic script. This automatic script allows for fast, and easy recreation of results, as well as allowing changes to be propagated and tested very quickly, without much input from the user. Especially since the runlist and eventlist are unlikely to change once settled upon.

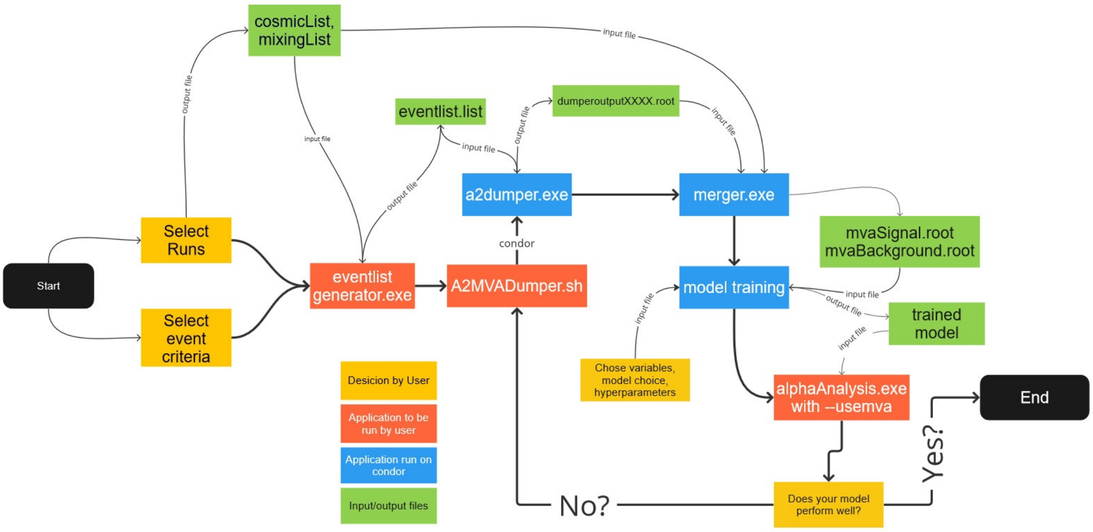


Figure B.1: Diagram of the ALPHA-2 data storing and model training pipeline.

A sketch algorithm of the model training procedure is as follows:

1. Select a list of *runs* for use, separate these into signal and background and append to a list file (`mixingList` and `cosmicList`)

2. Chose the criteria for events to be included in the dumper. Generally, for cosmic runs all events are chosen, and for mixing events only events passing the selection criteria found in `main_eventlist_generator` are used. There are options within this macro to change the minimum rate required for the detector to be considered saturated, as well as the minimal ratio between detector reads and triggers. The events can be stored using the `WriteEventList` method of the `TA2Plot` or automatically generated via the eventlist generator.

3. Run the eventlist generator (`macro_main_eventlist_generator.exe`) resulting in an eventlist (named `eventlist.list`).

4. Pass this list to the `a2dumper.exe` with your chosen run number, `XXXX`, returning the `dumperoutputXXXX.root` file. This `.root` file contains only the variables defined in the `a2dumper.o` module of the `alphasoft` repo.

5. Run the `macro_main_merger.exe` with the `mixingList` and the `-mixing` flag to return `mvaSignal.root`, which is compiled root file containing all defined signal events, separated into three `TTrees`: `trainSignal`, `testSignal`, and `validationSignal`. Each containing a third of the total events in the `eventlist.list`. Synonymously the flag `-cosmic` with the `cosmicList` will return `mvaBackground.root`.

6. Once the variables are stored in these root files the macro `A2TMVAClassification.C` can load the data from these files and train a chosen ML classifier. The macro has options to load chosen variables and train using subsets, or change model architecture and hyper parameters. This will output a model generally in `xml` format.

7. By running `alphaAnalysis.exe` with the `-usemva` flag pointing to the model the reconstruction algorithm is able to run, booking events and attributing their resulting model response. Events can later be cut by this model response once an appropriate cut is chosen.

8. This process is automatically performed on the `htcondor` supercluster at CERN if the `A2MVADumper.sh` macro is used (it requires the run lists and the event list be generated by the user).

While the ALPHA-g dumper is synonymous there are a few steps prior to the dumper as the calibration of the barrel veto requires upwards of three hours of cosmic data. As such the outline for the ALPHA-g algorithm is as follows:

1. Define the `mixingList`, `cosmicList`, and selection criteria as in ALPHA-2. Use the `ag_eventlist_generator` to generate an eventlist (`eventlist.list`).
2. For each run in the events lists save the previous twenty runs to be analysed. This number can be tuned but it was found that generally to get the three hours of cosmic data needed for barrel veto calibration twenty runs was enough.
3. Run `alphagonline.exe` on all output files. `alphagonline.exe` is a light version of the reconstruction algorithm which only reconstructs given time windows and event counts. This allows the algorithm to run faster (close to real time) in exchange for losing out on detailed vertex and track information.
4. Using the time windows from the pre-processed runs find the exact amount runs needed to get the three hours of cosmic time. Generally only 3-5 runs are needed for this.
5. Run `agana.exe` on these runs, generating full vertex, track, and BV data. `agana.exe` is the full reconstruction algorithm and is considerably slower than `alphagonline.exe` but does reconstruct all vertex and track info.
6. Run the BV calibration for each run, this will automatically look back at previous runs until it has three hours of data. The BV calibration will output a file titled `BarrelCalibrationXXXX.root` for each run `XXXX` in the original `mixingList` and `cosmicList`.

7. Proceed as from step 4 in the ALPHA-2 pipeline, instead running `agdumper.exe`, `AGTMVAClassification.C`, and `agana.exe` instead of `alphaAnalysis.exe`
8. This process is also automatically performed on the `htcondor` supercluster at CERN if the `AGMVADumper.sh` macro is used (it also requires the run lists and the event list be generated by the user prior to running).

A more detailed description of these programs can be found in the respective documentation in the `alphasoft` repo.

Appendix C

Generalized Sphericity Tensor

The generalized sphericity tensor is a measure of the shape of an event, the eigenvectors of which describe the thrust axis of the event, and how “spread out” the tracks are in space. Cosmic rays passing through the detector will be less spread out than annihilations originating from the centre of the trap, making this tensor and its associated eigenvalues and vectors a useful variable in event classification.

The tensor itself is defined as

$$S^{ab} = \frac{\sum_i p_i^a p_i^b |\vec{p}_i|^{r-2}}{\sum_i |p_i|^2} \tag{C.1}$$

where $a, b = 1, 2, 3$ corresponds to the x, y, z components of tracks momentum (defined in Equation A.6).

Originally, this was defined with $r = 2$ but taking $r = 1$ offers more desirable properties [234], and is referred to as the linearised sphericity tensor.

By diagonalising S^{ab} one can find three eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$, with $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and three eigenvectors $\vec{\mathbf{v_1}}$, $\vec{\mathbf{v_2}}$, $\vec{\mathbf{v_3}}$. The first of which corresponds to the *event axis* and is closely related (though not identical) to the *thrust*.

The *sphericity* of the event is defined as

$$S = \frac{3}{2}(\lambda_2 + \lambda_3) \tag{C.2}$$

with $S = 0$ for exactly two back-to-back tracks, and $S = 1$ for configurations with exactly equal eigenvalues (evenly spread).

The histogram of a similar variable $\sqrt{\lambda_1 + \lambda_2}$, separated for both signal and background is shown in Figure 8.1 and though the background events are not generally as low as 0, they are on average lower than signal events.

Appendix D

Measuring Scintillating Panel Detection Efficiencies

A scheme for measuring the detection efficiency for a given panel (denoted panel a) is as follows:

1. Obtain a further two panels (panels b and c) as well as the one set for measurement.
2. Place and fasten the two additional panels on either side of the measurement panel.
3. Record the counts on coincidence of both additional panels, $C_{b,c}$, and the counts on coincidence of all three panels, $C_{a,b,c}$.
4. The detection efficiency of panel a, E_a, is then defined as the number of particles passing through all three particles, compared to those not detected on panel a, i.e: $E_a = \frac{C_{a,b,c}}{C_{b,c}}$.

The detection efficiencies of all panels within the experiment were measured, and the table can be found in Table D.1.

Panel	Efficiency (%)
PMT1	28.30 ± 2.14
PMT2	58.68 ± 3.44
SiPM1	97.08 ± 0.59
SiPM2	49.29 ± 0.27
PMT3	21.47 ± 0.68
PMT4	9.00 ± 0.26
SiPM_I	77.82 ± 3.06
SiPM_J	82.16 ± 3.16
SiPM_G	82.86 ± 3.25
SiPM_H	82.22 ± 2.28

Table D.1: Measured panel efficiencies for various panels throughout the ALPHA experiment.

Appendix E

Magnet Geometry

Here we include the final geometry within both the Swansea (Figure E.1) and Carlsberg (Figure E.2) magnets after contacting both Oxford and Swansea for the design notes.

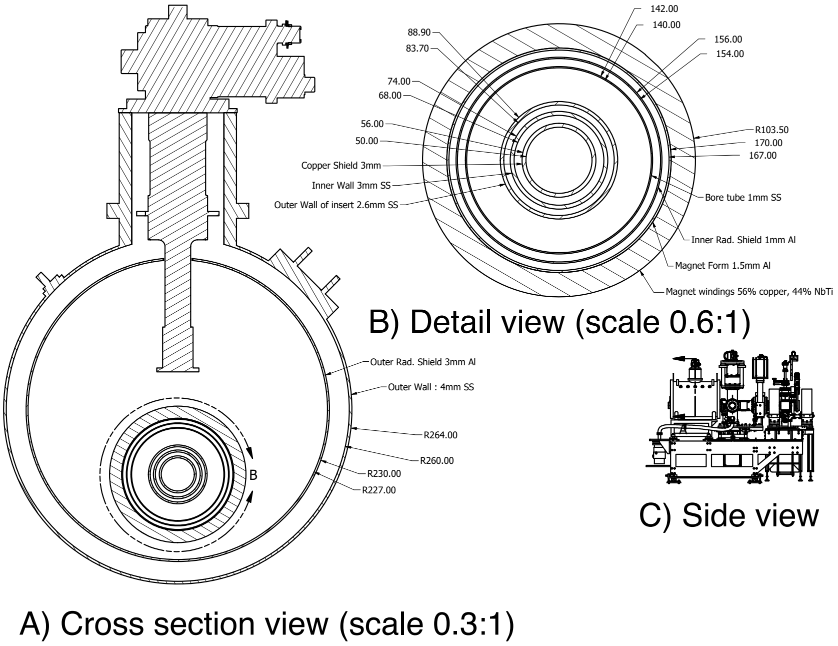


Figure E.1: Swansea magnet and its contents.

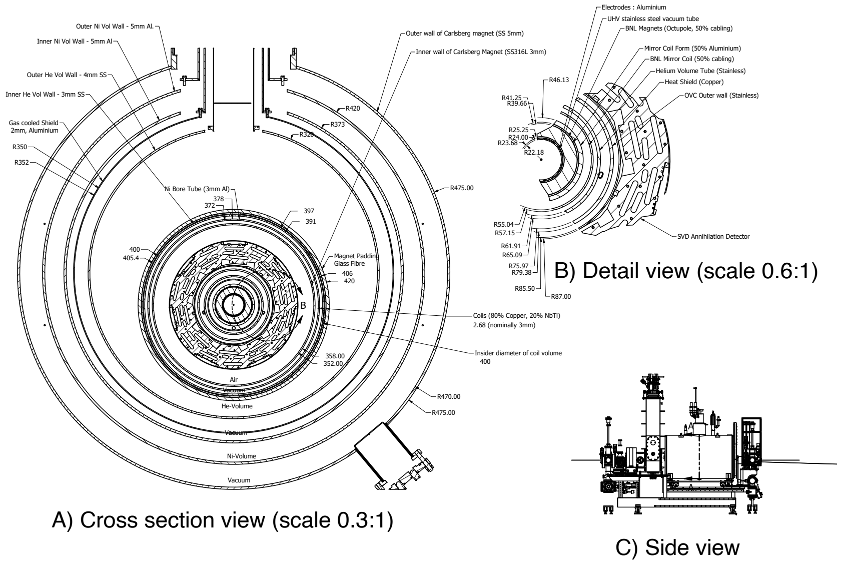


Figure E.2: Carlsberg magnet and its contents.