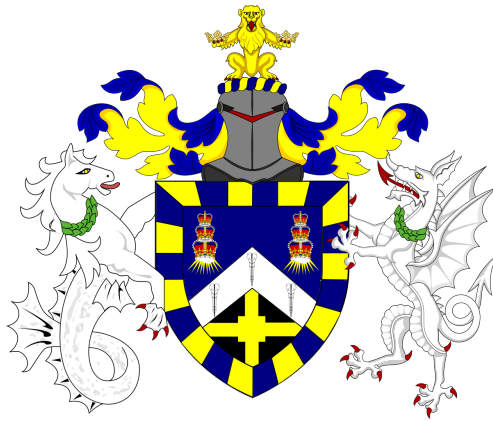


Applications of Machine Learning to the Monopole & Exotics Detector at the Large Hadron Collider



Lewis Richard Millward

Supervisor: Prof. Jonathan Hays

Submitted in partial fulfilment of the requirements
of the Degree of Doctor of Philosophy

School of Physics and Astronomy

Queen Mary University of London

Submission date 18th Dec 2022

Statement of originality

I Lewis Richard Millward, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: 18th Dec 2022

Abstract

MoEDAL is the Monopole and Exotics Detector at the Large Hadron Collider. The Moedal Experiment uses Passive Nuclear Track Detector foils (NTDs) to look for magnetic monopoles, and other heavily ionising exotic particles at the Large Hadron Collider (LHC). Heavy particle radiation backgrounds at the Large Hadron Collider make image analysis of these NTD foils non-trivial compared to NTD image analysis under lower background conditions such as medical ion beam calibration or nuclear dosimetry. This thesis looks at multichannel and multidimensional Convolutional Neural Network (CNN) and Fully Convolutional Neural Network (FCN) based image recognition for identifying anomalous heavily ionising particle (HIP) etch pits within calibration NTD foils that have been exposed to both a calibration signal (heavy ion beam), and real LHC background exposure, serving as detector research and development for future MoEDAL NTD analyses.

Image data was collected with Directed-Bright/Dark-Field illumination, parametrised at multiple off-axis illumination angles. Angular control of the light intensity distribution was achieved via a paired Fresnel lens and LED array. Information about the 3D structure of the etch pits is contained in these parametrised images which may assist in their identification and classification beyond what is possible in a simple 2D image.

Convolutional Neural Network etch pit classifiers were trained using Xe, and Pb ion data with differing levels of LHC background exposure. An ensemble approach of combining classifiers trained on different objects, and data-channels is shown to improve classification performance. Transfer learning was used to generate Fully Convolutional Neural Networks for identifying HIP etch-pit candidates from wide area foil scan images. The performance of the FCN algorithm is evaluated using a novel MoEDAL R&D foil stack, in order to obtain blinded estimates of the signal acceptance and false prediction rate of an ML based NTD analysis. Additionally a method for pixel to pixel alignment of NTD foil scans is demonstrated that can be used for the training of U-Net FCN architectures.

Acknowledgements

I would like to thank my very patient supervisor Professor Jon Hays, and secondary supervisor Professor Adrian Bevan for their guidance and advice over the years. Thank you all staff and students in the PPRC, in particular Tom Charman, Tom Stevenson, and Rodrigo Gamboa.

I would like to thank Jame Pinfold, and the MoEDAL collaboration. Especially Laura Patrizii, Vincent Togo, and Attanu Malik, and all the NTD analysis team at INF Bologna for providing the etched NTD foils used in this research.

Thank you to Aman Upadhay, Kanik Palodhi, and Joydeep Chatterjee for collaborating with me on the upcoming MoEDAL ML paper.

List of abbreviations

HIP	Heavily Ionising Particle
MoEDAL	Monopole and exotic detector at the LHC
CNN	Convolutional Neural Network
FCN	Fully Convolutional Neural Network
Conv	Convolutional
WP	Working Point, classification decision boundary
NTD	Nuclear Track Detector
RI	Range In
RO	Range Out
R&D	Research and Development
SMP	Stable Massive Particle
VEV	Vacuum Expectation Value
COM	Centre of mass
IP8	LHC Interaction point 8
HECO	Heavy Electrically Charged Object
REL	Restricted Energy Loss
WP	Working Point

Contents

1	Introduction	15
2	Monopoles and Exotics	17
2.1	The Magnetic Monopole	18
2.2	Classical Monopoles	20
2.3	Dirac Monopoles	21
2.4	Topological Monopoles	24
2.5	Electro weak monopoles	27
2.6	Properties of magnetic monopoles	28
2.6.1	Magnetic charge interaction	28
2.6.2	Estimated Monopole mass energy	29
2.6.3	Interaction of monopoles in matter	29
2.6.4	Energy profile of magnetic monopoles	31
2.7	Experimental monopole searches	32
2.7.1	Condensed matter monopole phenomena	35
2.8	Challenges of Experimental Detector Searches	35
2.9	Exotic ionising particles	38
2.9.1	Exotic candidates	39
3	The MoEDAL Experiment	41
3.1	The Large Hadron Collider	43
3.2	The MoEDAL Detector	43
3.3	Nuclear Track Detectors	46

3.3.1	Ion track formation and Restricted Energy Loss	47
3.3.2	NTD etching and analysis	47
3.4	MoEDAL NTD Arrays	49
3.4.1	MoEDAL HIP signature	51
3.4.2	NTD Analysis process	51
3.4.3	LHC backgrounds and calibration	54
4	Machine Learning	57
4.1	Machine Learning	58
4.1.1	Supervised and Unsupervised Learning	58
4.1.2	Classification and Regression	59
4.2	Neural Networks	59
4.3	CNNs: Convolutional Neural Networks	61
4.3.1	FCNS: Fully convolutional networks	62
4.4	Optimising and Training Neural Networks	63
4.4.1	Cross-Validation and Over-training	63
4.4.2	K-fold Cross-Validation	64
4.4.3	Data-Augmentation	65
4.4.4	Transfer learning	65
5	NTD Imaging study	66
5.1	Introduction	67
5.1.1	Etch pit backgrounds	67
5.1.2	Imaging for machine learning	69
5.2	NTD foil samples	71
5.2.1	Unexposed CR39, with Xe ion signal	71
5.2.2	Background free makrofol, with Pb ion signal	72
5.2.3	Makrofol with Xe ion signal and 8 months LHC exposure	72
5.2.4	Makrofol Pb ion signal, full LHC background exposure	75
5.3	Imaging methods	75
5.3.1	Imaging Equipment	79

5.3.2	Bright-field and Dark-field Microscopy	80
5.3.3	Front and reverse side imaging	82
5.3.4	Focal depth adjustments	82
5.3.5	Rotational illumination	82
5.4	Results and Analysis	83
5.4.1	Bright-field imaging	83
5.4.2	Dark-field imaging	84
5.4.3	Front and reverse surface comparisons	85
5.4.4	Focal depth adjustment	86
5.4.5	Rotational illumination	87
5.5	Final NTD scanning procedure	91
5.6	Summary and Conclusions	93
6	Xenon Ion Study	95
6.1	Introduction	96
6.2	Preliminary Study	96
6.2.1	Methodology	96
6.2.2	Observations	98
6.3	Xe Dataset preparation	99
6.3.1	Image Pre-processing	99
6.3.2	Pre-selection	101
6.3.3	Labelling	105
6.4	Optimisation methodology	107
6.4.1	Initial Results	109
6.4.2	Binary Classification	110
6.5	Ensemble	110
6.5.1	Inference on NTD foils	112
6.6	Conclusions	112
7	Pb Ion Study	117
7.1	Introduction: The Pb ion test beam stack	118

7.1.1	Experimental ML stack	118
7.2	Foil Scanning and Alignment method	122
7.2.1	Foil scanning	122
7.2.2	Global stack coordinates	123
7.2.3	Etch-pit offset correction	125
7.2.4	Summary	129
7.3	CNN classification study	135
7.3.1	Etch pit Classification dataset	135
7.3.2	Training methodology	144
7.3.3	CNN Results	148
7.3.4	Summary and Conclusions	167
7.4	Fully Convolutional Neural Network	169
7.4.1	Simple FCN	170
7.4.2	Retrained FCN	174
7.4.3	FCN final predictions	176
7.4.4	Observations	176
7.5	Evaluation	181
7.5.1	Evaluation dataset	181
7.5.2	Evaluation methodology	183
7.5.3	Results	185
7.5.4	Misclassification Errors	188
7.5.5	Uncertainties	192
7.6	Pb Ion study conclusions	194
8	Conclusions	196
8.1	Prospects for an ML based NTD analysis	198

List of Figures

2.1	Dirac solenoid	23
2.2	Stable 'Hedgehog' configuration	26
2.3	Bethe plot for magnetic monopoles	31
3.1	LHC beam schematic	44
3.2	Ion track and the etching process	45
3.3	Ionisation damage to NTD polymer bonds	46
3.4	MISSING Ion track under an electron microscope	53
3.5	A MoEDAL NTD stack	53
3.6	Ionising particle traversing an NTD stack	54
4.1	A simplified neural network with a single hidden layer.	60
5.1	HIP candidate vs uncorrelated background	68
5.2	Xe ions in CR39	73
5.3	Pb ions in Makrofol	74
5.4	Typical foil image with 8 months of LHC exposure	76
5.5	Visual symmetry between etch pit objects	77
5.6	Makrofol with full 2 year LHC background	78
5.7	Microscope scanning bed along with NTD foil and alignment jig	80
5.8	Optical diagram of a dark-field imaging setup [45]	81
5.9	Light path via fresnel lens	82
5.10	Dark field image of etch pits in the exposed Xe ion foil	85
5.11	Pb exposed makrofol at two different focal depths	87

5.12	Rotational Bright/Dark field	88
5.13	Rotational illumination of a HIP etch-pit at high magnification.	89
5.14	Rotational illumination validated in CR39	92
6.1	Image categories, shown for the first rotational illumination channel. . .	97
6.2	Before and after normalisation.	102
6.3	An example of the output of the pre-selection step. A set of x,y coordinates corresponding to etch-pits of potential interest	104
6.4	An example of the breakdown in the preselection for very high etch-pit densities.	105
6.5	Network training vs Epoch	110
6.6	Validation performance of geometric classifier ensemble	113
6.7	Ternary plot of the geometric ensemble scores	114
6.8	Confusion matrix, for geometric ensemble scores	114
6.9	Three way inference classification performed on preselected etch pits. .	115
7.1	Research NTD stack exposure scheme	120
7.2	Datum alignment images	126
7.3	Average Etch-Pit image of the rotational illumination channels	128
7.4	Average Etch-Pit image after offset correction	131
7.5	Spatial dependency over 4 quadrants of the foil.	132
7.6	Naive alignment	133
7.7	Corrected alignment	134
7.8	Image normalisation	137
7.9	High quality etch pit examples	140
7.10	Highlighted comparison of Low quality objects	141
7.11	Etch pit matching between foils	142
7.12	Scratches filtered out via CNN classifier	143
7.13	Identified fragment examples	143
7.14	Accuracy evolution during training	152
7.15	Error/Loss evolution during training	153

7.16	Signal/Background Histograms	154
7.17	Distribution of fragment scores	155
7.18	Mean classification scores in a given data channel for each etch pit. Pb Sig (blue), Bkg (red) and Fragments (green).	157
7.19	Converged performance vs unique signal examples	159
7.20	Loss convergence with reduced data quantities	160
7.21	ROC curves	162
7.22	Histograms and ROC curves for the two channel combinations	164
7.23	Dark-field classification without and with data augmentation	166
7.24	Simple FCN heatmap, for both sides of the NTD foil	172
7.25	Heatmap pixel spectrum	173
7.26	Predicted X,Y locations superimposed	175
7.27	177
7.28	178
7.29	FCN vs 'Simple' FCN	179
7.30	All channel, vs Rotational channel FCN output spectrum	180
7.31	Global comparison of predictions with known pits.	187
7.32	Correctly predicted locations, shown in the exposed foil.	189
7.33	Global comparison of incorrect predictions	190
7.34	Missed truth examples (50-75) and (275-300)	191
7.35	False prediction examples (100-125) and (150-175)	193

List of Tables

6.1	Standard CNN layer architecture used for Xe study.	109
7.1	Foil Exposure	121
7.2	Training datasets	144
7.3	Standard CNN architecture used for Pb study.	146
7.4	CNN architecture used for 3D convolution	147
7.5	Pb CNN training Hyper-parameters	148
7.6	Pb CNN validation accuracy % (fragments included in training data) . .	149
7.7	Pb CNN validation accuracy % (fragments excluded from training data)	149
7.8	False positive, False negative rates % (fragments included)	149
7.9	False positive, False negative rates % (fragments excluded)	149
7.10	Example working points for Rotational classifiers	163
7.11	False positive, False negative rates % shown for different choices of ensemble combination and working point (wp). For the ensemble produced by a numerical average of classification scores a working point of 0.35 allowed for perfect separation of signal and background, choices below this value simply allowed more false positives, while setting a higher working point simply cause false negatives to occur for real etch pits	165
7.12	FCN output head	176
7.13	Combined FCN evaluated with Eval. data only	186
7.14	Combined FCN evaluated with Training and Eval data	186
7.15	Simple (Rot)FCN evaluated with Eval. data only	186
7.16	Simple (Rot)FCN evaluated with Training and Eval data	186

7.17 Manual categorisation of the incorrect outcomes over the entire foil (for the combined FCN with an overlap of 30)	192
---	-----

Chapter 1

Introduction

It is an experimentally observed phenomena, and theoretical axiom, that magnetic charge does not exist. Nevertheless every generation of physicists has postulated the existence of particles possessing magnetic charge; The magnetic monopole. Little is known about these hypothetical magnetic monopoles, beside an anomalously strong electromagnetic interaction, and experimental limits on their mass (last established at the Tevatron CDF [[9]]). The combination of a large electromagnetic interaction, and a low velocity result in a very strong ionisation effect. The experimental signature of magnetic monopoles is shared by many other exotic particles.

MoEDAL is an LHC experiment dedicated to searching for magnetic monopoles, and other highly ionising particles. It aims to build on previous Tevatron results and provide either evidence of monopoles, or a new best exclusion limit. The main tool MoEDAL uses to search for highly ionising particles are arrays of ionisation sensitive polymer foils known as nuclear track detectors (NTDs).

This thesis describes research into the use of machine learning to automate and accelerate the search for magnetic monopoles and highly ionising particles within the MoEDAL NTD stacks. The second chapter (Monopole theory) discusses magnetic

monopoles, and heavily ionising exotic particles.

The third chapter (MoEDAL) describes the Monopole and Exotic Detector at the LHC, a special particle detector designed to look for ionisation signatures belonging to magnetic monopoles. The chapter discusses the MoEDAL detector and the principles behind the nuclear track detectors that form MoEDAL's NTD array.

Machine learning (chapter 4) gives a basic background overview of common machine learning techniques, in particular convolutional neural networks.

Imaging (chapter 5) describes the imaging and scanning of the NTD foils used in this thesis. In particular it discusses an illumination method used to resolve the 3D structure of particle etch-pits.

The Xe study (chapter 6) describes a study into machine learning based etch-pit classification. A calibrated 13GeV/A Xenon ion beam is used to simulate a highly ionising particle signal within an NTD foil along with 8 months exposure to LHC nuclear background radiation. Ensemble based methods are used to provide accurate etch-pit identification based on a small sample of training data.

The Pb study (chapter 7) describes the use of an experimental NTD array to obtain a quantitative performance estimate for machine learning based etch-pit identification. A highly ionising particle signature is generated within a complete NTD array using a 136GeV/A lead ion beam. The first foil of the array was exposed at the LHC, with the unexposed foils acting as a truth record for particle etch-pits within the exposed foil. This truth record is used to train and evaluate a fully convolutional neural network (FCN) designed to identify HIP anomalies in wide area scans.

Lastly in the conclusions I will highlight the implications of chapters 5, 6, and 8 for future ML based NTD image analysis at Moedal. The method described in chapter 7 can be replicated for any other ion species, sampling a spectrum of $\frac{Z}{\beta}$, while the ensemble techniques of chapter 6 can be used to identify the most likely hypotheses for each identified candidate.

Chapter 2

Monopoles and Exotics

This Chapter will cover the theories behind hypothetical particles known as magnetic monopoles. I will describe the three main theoretical models of magnetic monopole, with their shared properties. An overview of experimental search methods, and the associated difficulties will be given. The chapter will also describe other exotic ionising particles with a similar detection signature, that may be detected by monopole searches.

2.1 The Magnetic Monopole

The non-existence of magnetic monopoles has remained a feature of electromagnetism as it has evolved through relativistic electromagnetism, quantum electrodynamics, through to the modern Electroweak theory of the standard model. It is not until we extend to grand unified theories beyond the standard model that monopoles emerge as a necessary requirement of the theory.

There are multiple motivations for postulating the existence of magnetic monopoles. The most basic justification is restoring the symmetry between electric and magnetic charge, which almost exists yet is broken in Maxwell's equations. The question of why that asymmetry exists is a fundamental question of physics in and of itself. Dirac showed that the existence of a magnetic monopole would link the quantisation of angular momentum to a necessary quantisation of electric charge [17]. This property holds for monopoles when they are extended to Electroweak unification, Beyond Standard Model (BSM), and to Grand Unified Theories (GUTs). Rather than merely being allowed, they emerge as a stronger and stronger prediction as the energy scale of the theory increases. Many grand unified theories require the existence of magnetic monopoles, and presence of monopoles in the early universe provides an explanation for matter antimatter asymmetry. In short the existence of magnetic monopoles would solve multiple fundamental questions in particle physics.

Scientists have postulated the existence of magnetic monopoles since the discovery of magnetism. There are several theoretical models that are of particular importance. These models explain unknown questions in physics, illustrate core physical principles, and constrain monopole properties such as mass and charge. These key models are summarised below;

- Dirac monopoles - Represent the minimal QED conceptualisation of a monopole as a point particle possessing magnetic charge. There are no theoretical constraints or

bounds on the mass of Dirac monopoles.

- Topological monopoles - Which emerge as topological soliton solutions of various theories. In particular they are predicted by most grand unified theories, an example would be string theory which always has a finite mass monopole. Most GUT monopoles are predicted to form at energies closer to the Planck mass than current or future collider energies. The T'Hooft Polyakov monopole is the canonical example emerging from $SO(3)$ gauge theory.
- Electroweak monopoles - They represent a generalisation of the Dirac monopole compatible with electroweak unification. The main example is the Cho Maison monopole. Many theories of electroweak monopoles lie close to the TeV scale, which makes them of particular interest to experimental LHC searches.

In many electroweak (and beyond) models of monopoles they no longer resemble traditional particles, rather they are topological particles, solitons in fundamental fields such as the Higgs field [10][19]. These topological solitons can form thermally [46] once the local energy scale undergoes a phase transition that restores the symmetry spontaneously broken by the Higgs mechanism.

The existence of magnetic monopoles would change the topology of the universe. The standard model of particle physics is framed as a perturbative expansion around the non-zero vacuum state expectation of the Higgs field. In a monopole (and other exotic solitons) free universe the only boundary conditions to the smooth Higgs field lie at the edge of the known universe. A universe with monopoles possesses additional boundary conditions surrounding the topological defect where the monopole occurs. The invariance of physics to our choice of mathematical perturbation expansion requires that phases in the standard model be unchanged by topological windings around this defect in the smooth perturbative model [44]. Topological winding invariance replicates the Dirac proof of quantised electromagnetic charge, changing the predicted minimum magnetic

charge g_D by an integer quantity. Thus every quantum description of monopoles predicts quantised electric charge and a large minimal magnetic charge.

2.2 Classical Monopoles

In his 1861, and 1865 publications [37][38], James Clerk Maxwell unified the four experimentally known laws governing electricity, and magnetism, into a single set of equations. Where ε_0 is the permittivity of free space, ρ is electric charge density, J is electric current density, E is the electric field, B is the magnetic field, and μ_0 is the permeability of free space. In a foreshadowing of modern electroweak unification, the forces of electricity and magnetism were shown to be two aspects of the same phenomenon; Electro-Magnetism. The second Maxwell equation (2.2) known as Gauss's law of magnetism, states that the total magnetic flux through any Gaussian surface is zero. No magnetic charges (monopoles) exist, rather magnetism stems from dipole moments. This principle is familiar to us with simple ferro-magnets, which exist as paired north-poles and south-poles.

$$\nabla \cdot E = \frac{\rho}{\varepsilon_0} \quad (2.1)$$

$$\nabla \cdot B = 0 \quad (2.2)$$

$$\nabla \times E = -\frac{\partial B}{\partial t} \quad (2.3)$$

$$\nabla \times B = \mu_0 \left(J + \varepsilon_0 \frac{\partial E}{\partial t} \right) \quad (2.4)$$

There is no reason or justification for why magnetic charges don't exist, other than that they have never been observed experimentally. In the absence of electric charges Maxwell's equations become symmetric between electricity E and magnetism B , The equations decouple into separate wave equations from which we derive the wave propagation of

light in a vacuum (2.5).

$$\nabla^2 E - \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2} = 0 \quad (2.5)$$

$$\nabla^2 B - \mu_0 \varepsilon_0 \frac{\partial^2 B}{\partial t^2} = 0 \quad (2.6)$$

Maxwell's equations can be generalised to instead include magnetic charge ρ_M and current J_M . The generalised Maxwell's equations result in a modified form of the Lorentz force that would be experienced by a magnetic monopole (2.9). Of particular interest is the behaviour of a monopole in a constant magnetic field (common in particle detectors) where the force acting on a magnetic charge q_m is constant and is decoupled from both the direction and the magnitude of the velocity vector $v = \beta c$.

$$\nabla \cdot B = \rho_M \quad (2.7)$$

$$\nabla \times E = -\frac{\partial B}{\partial t} - J_M \quad (2.8)$$

$$F = q_e \left(E + \frac{v}{c} \times B \right) + q_m \left(B - \frac{v}{c} \times E \right) \quad (2.9)$$

2.3 Dirac Monopoles

Dirac's equation[16] (2.10), and the associated Lagrangian \mathcal{L}_{Dirac} (2.11), attempted to provide a Lorentz invariant description of the quantum mechanical behaviour of massive spin 1/2 particles via a spinor field ψ . The free-field solutions of the Dirac equation predicted the existence of anti-particles, along with fermions having an intrinsic spin angular momentum. Electromagnetism is introduced to Dirac's formulation, through the substitution of ∂_μ for the covariant derivative D_μ (2.12), which is invariant under U(1) phase transformations of the spinor ψ . In the theory of Quantum-Electro-Dynamics (QED), ψ and A become the operators of quantum fields associated with fermions and

photons respectively, and the electromagnetic tensor $F_{\mu\nu}$ is defined via Eqn.2.14. Using perturbation theory Dirac's QED successfully predicted many EM phenomena, including the gyromagnetic ratio of the electron from first principles, and the precise value of the electron g-factor g_e (2.15)

$$0 = (i\hbar\gamma^\mu\partial_\mu - m)\psi \quad (2.10)$$

$$\mathcal{L}_{Dirac} = \bar{\psi}(i\hbar\gamma^\mu\partial_\mu - m)\psi \quad (2.11)$$

$$D_\mu \equiv \partial_\mu + ieA_\mu \quad (2.12)$$

$$A_\mu \rightarrow A'_\mu \equiv A_\mu - \frac{1}{e}\partial_\mu\alpha(x) \quad (2.13)$$

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu \quad (2.14)$$

$$g_e = 2 \left(1 + \frac{\alpha}{2\pi} \right) = 2.002319... \quad (2.15)$$

Dirac's formulation of QED cannot be used to intrinsically describe magnetic monopoles. The definition of the magnetic field in terms of the potential A rules out point sources of magnetic charge (2.16). This did not prevent Dirac from proposing the existence of magnetic monopoles in order to explain the integer quantisation of electric charge. In his 1931 paper [17] Dirac hypothesised a one dimensional theoretical construct known as a Dirac string (fig2.1). The Dirac string represents a solenoid that is both infinitely long and infinitely thin. Magnetic field lines or flux Φ , would enter one end of the solenoid (resembling a negative divergence of the magnetic field, ie, a magnetic anti-pole $-g_m$) and exit via the other end of the solenoid (resembling a source of magnetic field lines, or magnetic pole g_m). Magnetic field lines within the infinitesimal width of the solenoid would have no observable physical influence. In this formulation a Dirac string would be indistinguishable from an infinitely separated monopole anti-monopole pair.

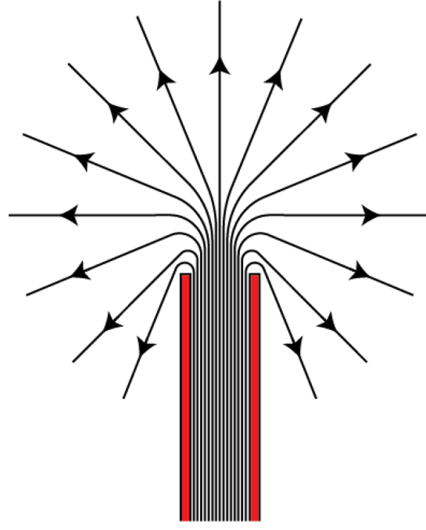


Figure 2.1: Dirac solenoid

[[52]] Magnetic flux lines are shown passing through the solenoid and emerging. at the limits of infinitesimal distance and infinitesimally narrow solenoid the field is identical to a magnetic point charge or monopole.

$$\begin{aligned}
 B &\equiv \nabla \times A \\
 \therefore \nabla \cdot B &= \nabla \cdot (\nabla \times A) = 0
 \end{aligned}
 \tag{2.16}$$

A key feature of Dirac's theory of magnetic monopoles is that for the Dirac string to be physically unobservable the phase shift $\Delta\theta$ in the wave function of any charged particle passing either side of the theoretical solenoid must be restricted to integer \mathbb{Z} multiples of 2π . This requires that electric charge must be quantised in integer units of e (The Dirac quantization condition), which matches what is experimentally observed.

$$\Delta\theta = qg \in \mathbb{Z}2\pi
 \tag{2.17}$$

This equation also implies a minimum quanta of magnetic charge g_D/q_m , which is noted to be much larger than the equivalent electric charge $g_D = 68.5e$. The high value of g_D has several implications;

- The QED interaction of magnetic monopoles becomes strongly coupled, and is inadequately described by low order perturbation theory. This makes it hard to obtain accurate quantitative predictions of their behaviour.
- At a minimum, the strength of magnetic interactions will scale with $\left(\frac{ng_D}{e}\right)^2 = 68.5^2 = 4692.25$. This means magnetic monopoles will have an abnormally strong ionising interaction with matter compared to electrically charged particles.

The Dirac string formulation of monopoles postulates monopoles as point-like particles with quantised magnetic charge. The theory does not provide any constraint on monopole mass, nor is it gauge invariant. The core predictions arising from the theory are a quantised magnetic charge g_D and electric charge g , arising from the topology introduced by the Dirac string. These core predictions will arise out of other magnetic monopole theories due to similar topological considerations.

2.4 Topological Monopoles

In the Standard model Electroweak theory, electromagnetism and weak nuclear interactions emerge out of the spontaneous symmetry breaking of the $SU(2) \times U(1)$ invariant Higgs potential. Prior to the Glashow-Salam-Weinberg (GSW) Electroweak theory of the standard model, many theorists proposed alternate gauge theories. Once such theory was the now-superceded Georgi-Glashow (GG) $SO(3)$ model [22]. A feature of the Georgi-Glashow model is that it allows for stable non-vacuum solutions for the Higgs field known as ‘Topological solitons’. These solitons exhibit mass energy, and magnetic charge, de facto appearing as massive stable magnetic monopoles. The Georgi-Glashow

$SO(3)$ model was disproved in favour of the GSW model. Many theories attempt to unify the electroweak force, with the strong nuclear force, these are called Grand Unified Theories (GUT). In these theories the $SU(2) \times U(1)$ electroweak force and the $SU(3)$ strong nuclear force emerge as a result of the spontaneous symmetry breaking of a higher order gauge symmetry. T'Hooft and Polyakov demonstrated in 1974 [27][42] that any GUT theory where $U(1)$ electromagnetism is produced by a broken gauge symmetry would possess monopole-like soliton solutions, due to the topological considerations of homotopy theory.

The Georgi-Glashow model has a three component real Higgs field with a non-abelian internal $SO(3)$ gauge symmetry. Similar to the standard model the zero field solution is not the lowest energy state. The model possesses a degenerate set of non-zero vacuum solutions for the Higgs field. Namely the vacuum solutions lie on the surface of a sphere. Once the $SO(3)$ symmetry is spontaneously broken by the choice of Higgs field the theory still possesses a $U(1)$ symmetry corresponding to rotations about the axis of the Higgs field. This $U(1)$ gauge symmetry gives rise to familiar electromagnetism.

Theorists Gerard T'Hooft, and Alexander Polyakov showed that a stable configuration for the Higgs field in the $SO(3)$ model was the so called 'hedgehog-configuration' 2.2. In the hedgehog configuration the internal $SO(3)$ alignment of the Higgs field corresponds to its direction from the origin in 3-D space. At the origin the Higgs field vanishes, and full unbroken $SO(3)$ symmetry is restored. At infinity the Higgs field is approximately uniform, corresponding to the typical vacuum solution. The T'Hooft Polyakov hedgehog configuration possesses three important properties outlined below;

- The configuration is topologically stable, and can't be changed by smooth deformation.
- The zero-potential at the origin possesses energy in excess of the vacuum energy

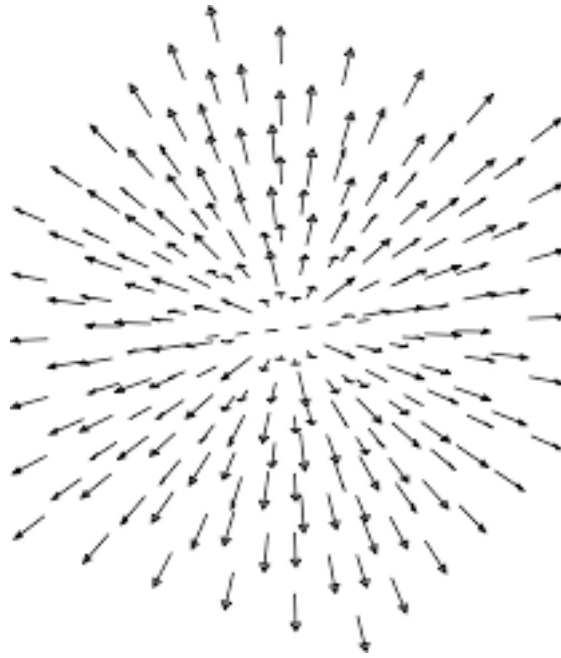


Figure 2.2: Stable 'Hedgehog' configuration

The arrows indicate the direction and strength of the Higgs field vector. In the hedgehog configuration the Higgs field vector lies parallel to the displacement from the origin \hat{r} . At the origin the Higgs field goes to zero. The configuration cannot be deformed into a uniform field without creating a local discontinuity

due to the non-zero vacuum expectation value (v.e.v). This static energy effectively acts as a rest mass.

- The configuration represents a divergence of magnetic flux equivalent to a magnetic charge of $g = \frac{4\pi}{q} = 2g_D$

The combination of these properties means that this topologically stable field configuration effectively behaves like a massive particle, possessing a magnetic charge, ie, a magnetic monopole.

2.5 Electro weak monopoles

It is commonly assumed that the $SU(2)_L \times U(1)_Y/U(1)_{EM}$ topology of the standard model does not allow for finite energy monopoles to exist as topological soliton solutions. In 1997 Cho and Maison [11] demonstrated that there was a gauge independent alternative abelian decomposition of the GSW standard model. This decomposition results in a non-abelian $SU(2)$ component with a topology that allows monopole like soliton solutions (akin to the the Georgi-Glashow model), and a non-trivial abelian $U(1)_Y$ topology allowing point-like Dirac monopoles. These Cho-Maison monopoles (and dyons [30]) represent the electroweak generalisation of the Dirac monopole. The periodicity of $U(1)$ subgroup of $SU(2)$ is 4π rather than 2π resulting in the minimal electroweak magnetic charge being $\frac{4\pi}{e} = 2g_D$. Due to possessing a point-like singularity at the origin the classical energy of these standard model monopoles is infinite.

However minor modifications to the standard model, such as the approach of Cho, Kimm, Yoon (CKY)[29], in modifying the kinetic hypercharge term of the standard model Lagrangian (which approximated to 1 everywhere far from the monopole), or applying quantum corrections (namely renormalising the $U(1)_Y$ coupling) make it possible to regularise this singularity and obtain finite energy monopoles [35]. Depending on the approach used, and what interactions are considered we obtain mass estimates for

electroweak monopoles of around 4 to 10 TeV. The central point of the estimates lies at approximately 7.2 TeV, with many models predicting a monopole mass close to this point [18]. A Bogomol'nyi-Prasad-Sommerfield (BPS) limit may be calculated for the lower bound of a CKY Cho-Maison monopole, and is found to be $M \geq 2.37$ TeV [2].

2.6 Properties of magnetic monopoles

While not exhaustive the above outline covered the main theoretical models, with other models representing variations on these principles. In experimental monopole searches these models and their derivatives are united by the following common properties; They possess magnetic charge, they have a large mass, and due to quirks of the magnetic charge interaction they are very heavily ionising. This section will discuss these properties.

2.6.1 Magnetic charge interaction

By definition monopoles possess magnetic charge, for the models above magnetic charge will occur in integer multiples of the Dirac charge $g_D = \frac{hc}{2e}$. On the level of Feynman diagrams (and perturbative expansions thereof) magnetic charge does not differ from electric charge in its interactions. In a practical context monopoles differ wildly from standard QED and electroweak interactions. The fine structure constant $\alpha \approx \frac{1}{137}$ mediates the strength of electric charge coupling between photons and electrically charged matter. Since α is small, electromagnetic interactions in QED and electroweak can be expanded in perturbative orders of α . Magnetic monopoles and other magnetically charged particles possess a strong coupling strength $\alpha_g = \frac{g_D^2}{hc} \approx 34.25$ which even at its lowest possible value is highly non-perturbative. This makes accurate predictive calculations about magnetic charge interactions extremely difficult. One approach, the Drell-Yann approximation, is to scale electrically charged processes by the magnetic charge coupling $\left(\frac{ng_D}{e}\right)^2 = n^2 68.5^2$ where n is the magnetic charge number of the minimal magnetic charge ($n=1$ for dirac,

$n=2$ for electroweak), this provides an order of magnitude estimate. Other effective Drell-Yann models abstract higher order processes into an effective coupling strength.

2.6.2 Estimated Monopole mass energy

Dirac monopoles do not have an associated mass prediction. The vanishing Higgs field at the core of topological monopoles possesses vacuum energy that acts as the de facto monopole mass. The T'Hooft-Polyakov monopole occurs for any theory where $U(1)$ emerges from a higher order gauge symmetry unifying the strong and electroweak forces. In the $SO(3)$ Georgi Glashow model, magnetic monopoles would have occurred after electroweak unification energies ≈ 100 GeV. However for most GUT's the unification energy $\approx 10^{15}$ GeV is far higher than typical colliders.

Electroweak Cho Maison monopoles have predicted energies between a lowest limit of 2.37 TeV [2] and an upper limit of around 10 TeV [29][30]. Since most EW mass estimates cluster around 7.2 TeV, Electroweak monopole anti-monopole pair production may be observable at the large hadron collider. [18]

Past accelerator searches at Fermilab have established a mass limit up to 850 GeV [4]. The last direct experimental collider search at the Tevatron CDF in 2006 [9] set an estimated mass limit of $m_M > 360$ GeV. I will take the approach of the MoEDAL collaboration and assume that the a potential monopole mass remains a free parameter of TeV scale.

2.6.3 Interaction of monopoles in matter

The interaction of electrically charged ionising particles in a physical medium is described by the Bethe equation (2.18)[5], sometimes modified to correct for QED and density corrections. Here $K = 4\pi N_A r_e^2 m_e c^2$. r_e is the classical electron radius, N_A is Avogadro's number. I is the mean excitation energy of the medium, approximately Z . z is the

electric charge of the ionising particle, while Z is the atomic number of the material being ionised. The Bethe equation represents the average mean energy loss rate for charged particles as a function of relativistic velocity $\beta\gamma$. In the relativistic limit where $\beta \approx 1$ the Bethe equation reduces to the simplified Bethe equation. The minimum ionisation point occurs at approximately $\beta\gamma \approx 3$, ie, energy exceeding 3 times the rest mass of the particle. Most particles produced in a relativistic collider environment at a primary or secondary level possess energy much greater than their rest mass. Relativistic charged particles with $\beta\gamma > 3$ are referred to as 'Minimum ionising particles'. These particles ionise inefficiently for their energy.

$$-\left\langle \frac{dE}{dx} \right\rangle = \frac{z^2}{\beta^2} \frac{KZ}{A} \left\{ \frac{1}{2} \ln \left\{ \frac{2m_e \beta^2 \gamma^2}{I} \right\} - \beta^2 \right\} \quad (2.18)$$

The strong electromagnetic coupling of magnetically charged particles modifies their ionisation behaviour in two distinct ways; First the electric charge dependence is replaced by the much stronger magnetic charge $g_D \approx 68.5e$. Secondly the magnetic force exerted by the monopole (eq.2.9) scales with β , cancelling out the inverse β^2 dependence of the formula. We may substitute in the effective electric charge of the monopole (2.19) into the Bethe equation to get the Bethe equation for monopoles (2.20).

$$z_{eff} = \frac{ng_D\beta}{e} \quad (2.19)$$

$$-\left\langle \frac{dE}{dx} \right\rangle = \left(\frac{ng_D}{e} \right)^2 \frac{KZ}{A} \left\{ \frac{1}{2} \ln \left\{ \frac{2m_e \beta^2 \gamma^2}{I} \right\} - \beta^2 \right\} \quad (2.20)$$

This changes the behaviour significantly in two main regimes. Relativistic poles $\beta \approx 1$ with a single Dirac charge ($n = 1$) will have an ionisation strength $68.5^2 \approx 4700$ times that of relativistic standard model particles. The very high ionisation rate doesn't change significantly as the monopole loses energy (Fig. 2.3). A monopole in matter is expected to lose energy rapidly and consistently until it enters the non relativistic regime $\beta < 0.1$

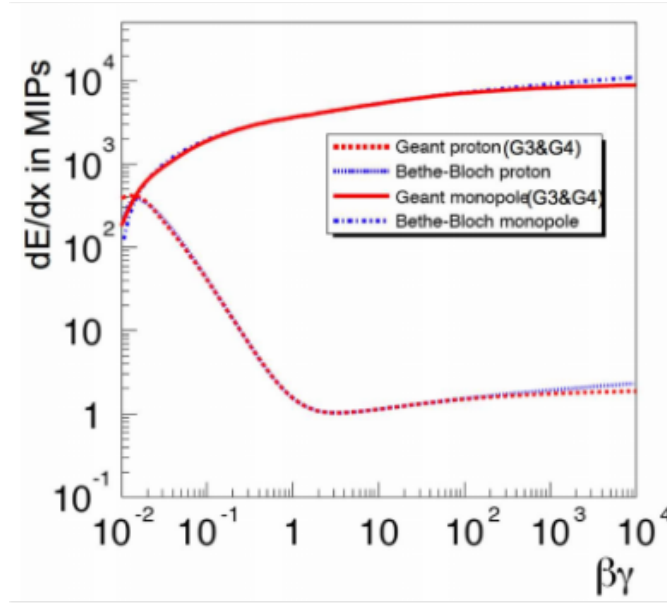


Figure 2.3: Bethe plot for magnetic monopoles

Bethe plot showing the ionisation energy loss (in air) of magnetic monopoles vs protons as a function of $\beta\gamma$, relative to the minimal ionisation point of the proton. Courtesy of the MoEDAL collaboration.

where it exhibits further deviance from standard model behaviour, and the energy loss may be approximated via (2.21).

$$-\frac{dE}{dx} \approx K\beta \quad (2.21)$$

Here the ionisation of a magnetic monopole falls as it loses velocity, while standard model particles reach their peak ionisation at low velocities. This is distinct difference that can be exploited in experimental searches.

2.6.4 Energy profile of magnetic monopoles

Many magnetic monopoles possess predicted formation energies in the TeV scale or beyond. Even monopoles formed in 14TeV LHC collisions will contain most of their energy as rest mass. The relatively low kinetic energy may place them significantly below the relativistic $c \approx 1$ realm of typical particle physics. The very short stopping distance means that even the slightest interaction with non-sensitive detector material

and infrastructure can dramatically change the kinematic behaviour in a way that crosses multiple descriptive models and regimes [41]. These are outlined below;

- Fast Poles - near relativistic speeds that ionise rapidly, with much larger radiative energy losses than standard particles.
- Slow Poles - with $0.01 < \beta < 0.0001$ lose energy primarily through electronic excitations of the material they pass through, and secondly through nuclear recoil collisions. Energy loss is extremely high in ferromagnetic material.
- Very slow, and Trapped monopoles - When a monopole has lost sufficient energy it can no longer excite atoms and loses energy through elastic collisions. One explanation of why we haven't observed any magnetic monopoles is that they are simply trapped. This can occur due to electrostatic image forces in electrical conductors, or through the formation of bound states via nuclear magnetic dipole interactions.

Despite the rapid energy loss, and de-acceleration to low β regimes monopoles still possess enormous kinetic energy and momenta resulting in minimal deflection in material.

2.7 Experimental monopole searches

There are two possible sources of observable magnetic monopoles, either production at high energy colliders, or the detection of cosmic remnant monopoles. In the latter case monopoles would have been produced thermally in the early universe, or via the Kibble mechanism. Due to their stability such monopoles should still be present, and may even be dark-matter candidates.

In accelerators the mass energy of monopoles that may be produced is limited by the centre of mass collision energy $2M \leq \sqrt{s}$. The LHC where $\sqrt{s} = 14\text{TeV}$ may produce many possible types of electroweak monopole. The main monopole M anti-monopole \bar{M}

production mechanisms are via electromagnetic processes, Drell Yann production and photon fusion. Uniquely monopoles may also form as topological solitons during the instantaneously high thermal fluctuations of the Higgs field during collisions [10]. Because of the non-perturbative nature of monopole interactions there are significant systematic uncertainties in estimating monopole production cross-sections. Approximating monopole production via scaling of the the leptonic Drell Yann cross-section (2.22) we obtain (2.23).

$$\sigma(qq \rightarrow l^+l^-) = \frac{Q^2}{3} \left(\frac{4\pi\alpha^2}{3s^2} \right) \quad (2.22)$$

$$\sigma(qq \rightarrow M\overline{M}) = \left(\frac{ng_D}{e} \right)^2 \times \sigma(qq \rightarrow l^+l^-) \quad (2.23)$$

There are two possible search strategies;

Direct search methods look to directly detect and identify magnetic monopoles via their anomalous magnetic interaction.

- SQUIDs (Superconducting Quantum Interference Devices) rely on the permanent magnetic pole of a monopole to induce a permanent current (2.24) in a superconducting ring. SQUID experiments have been deployed both terrestrially and in space. The most famous example is the 'valentines day' event' observed by Cabrera in 1982 [8].

$$I = \frac{\mu_o q_M}{L} \quad (2.24)$$

- Nuclear Track Detectors (NTDs) look for anomalous ionisation in layers of ionisation sensitive detector medium. NTDs have been deployed at most colliders to date, and represent half of all experimental monopole searches. To date no hypothesised monopole events have been observed at colliders. An unexplained monopole hypothesis event did occur during a high altitude balloon experiment in 1975 [43]. NTDs have also been employed in high altitude arrays such as the SLIM experiment

[24]. NTDs will be covered extensively later in this document.

- Due to their magnetic charge monopoles will exhibit anomalous curvature in the magnetic fields found in most collider experiments. Notably the field will act on the particle parallel to its direction of motion rather than being limited to the perpendicular plane.

Indirect search methods which don't attempt to identify the monopole directly, but look for indirect evidence of its existence.

- Topological monopoles may catalyse nucleon decay. The core of a topological monopole possesses unbroken gauge symmetry (assuming higher order gauge symmetry). The potential barriers that prevent nucleon decay are removed if symmetry is restored, thus a proton directly contacting this core is predicted to decay rapidly. Experiments such as MACRO [23] have looked for monopole catalysed nucleon decays.
- Due to their extreme ionisation and large radiative energy losses many experiments have attempted to look for the Cherenkov radiation that would accompany a cosmic monopole falling to earth. A recent example is the IceCube experiment in Antarctica [14]. Like the NTD based ionisation search this would produce a much higher signature than a minimally ionising particle generated through a cosmic ray shower.
- The monopoles predicted by most GUTs must be compatible with current observations and theories regarding early universe cosmology and inflation. The Parker limit represents a bound on cosmic monopole density [36][7].
- Lastly it may be possible to observe magnetic monopoles or virtual monopoles through reconstructing their decay to standard model particles detectable via

general purpose detectors, or their impact on the cross section of well known standard model processes.

2.7.1 Condensed matter monopole phenomena

Monopole like topological solitons have been identified and observed in condensed matter systems [47]. These are pseudo-particles and different from fundamental magnetic monopoles, although they emerge from the same mathematical principles.

2.8 Challenges of Experimental Detector Searches

Collider based searches fall into two distinct categories. Direct passive detector searches, such as those using NTDs, and direct and indirect active detector searches. The MoEDAL experiment discussed in the next section represents a continuation of direct passive collider searches only at higher LHC energy scales, taking over from previous Tevatron searches. Active detector searches have also taken place at ATLAS [13][48], some of the considerations of an active detector search are given below;

Indirect search reconstruction

The very large theoretical uncertainties surrounding magnetic monopoles makes it hard to obtain accurate predictions about possible decay mechanisms or how they would modify the cross-sections of existing processes. The basic approximation is to use Drell-Yann electromagnetic scattering with the electric charge replaced with magnetic charge, and adjust the allowed phase space of final decay states. While this gives us a useful order of magnitude estimate it contradicts with one of our few known certainties about magnetic monopoles, namely their strongly non-linear electromagnetic interaction. A further complication is that magnetic monopoles may possess inherent topological stability meaning they are unable to decay except through interaction with a corresponding

anti-monopole.

Pixel vertex interactions

Detectors at the LHC have to be radiation hard to resist the large showers of secondary and tertiary nuclear spallation radiation created indirectly through LHC collisions. Most detector subsystems are optimised for sensitivity to singular electric charge particles travelling near the speed of light. Many systems do not possess the dynamic range needed to accurately measure ionisation events 4700 times larger than optimal range. These large ionisation events are likely to saturate multiple pixels. While this may act as a distinctive sign of a heavily ionising particle, it will hinder the accuracy with which we could measure and reconstruct such a particle, and increase pixel dead-time.

Bunch crossing and event resolution

Large Hadron Collider proton bunch crossings occur every 25 nanoseconds, Detector subsystems are timed to be sensitive to the time-of-flight arrival of near light speed particles arriving from the interaction point after each bunch crossing. Particles slower than $\beta \approx 0.5$ are likely to be associated with the wrong collision event, or smeared between two events. Due to high TeV mass and ionisation energy loss, monopoles will have a low velocity when they enter detector systems. The sensitivity of ATLAS and CMS to magnetic monopoles is primarily limited by velocity.

Triggering

The raw data collection rate of the LHC's main electronic detectors greatly exceeds their capacity to offload and store this information. To handle this excess data most detectors employ a hierarchical series of event triggers that decide which events to keep. Most of the detailed and computationally intensive event reconstruction must be performed away from the detector environment where electronics can be shielded from the ionising radiation

near the interaction point. The initial trigger electronics responsible for reducing the event rate to manageable levels must operate within 0.5 microseconds with limited raw information from the detector. The majority of LHC events must be thrown away, and depending on the electronics used (embedded, vs FPGA vs more programmable) the trigger must be decided on before data collection.

Ferromagnetic trapping

Magnetically charged particles such as monopoles have an extremely short stopping distance in ferromagnetic and conductive media. Contact with any detector infrastructure may cause them to lose energy very rapidly into the medium in a way that doesn't replicate a typical hadronic spallation shower. Energy is dissipated in the movement of free electrons displaced by the moving magnetic field of the monopole.

Conclusion

Each of the above mentioned points represents a potential difficulty for electronic detector sensitivity to magnetic monopoles. The difficulty is compounded by the interaction of these factors. Possible signs of a magnetic monopole are subtle differences in their physical behaviour, that possess a large similarity with extremely common spallation radiation backgrounds, or mimic detector glitches. The theoretical and phenomenological parameter space for magnetic monopoles is large and difficult to anticipate in advance. The likelihood of a possible monopole event being thrown away is fairly high, while much of the information needed for an accurate reconstruction is likely to be lost. An observation of a hypothesised magnetic monopole would face the same scientific challenges faced by the observations in 1975, 1982, and 1985 namely;

- Demonstrating it wasn't an electronic glitch.
- Proving no standard model process could have caused the event.

- Allowing properties such as mass or magnetic charge to be reconstructed in a way that puts constraints on different monopole theories.

2.9 Exotic ionising particles

There are many other highly ionising particles that possess many of the same practical difficulties as detecting monopoles citeAcharya2014, particularly stable massive particles (SMPs). Most particle detectors rely on complicated reconstruction chains to identify known particle hypotheses from raw detector outputs. They rely on accurately identifying known particles, such as high energy muons in muon chambers, or large quantities of missing transverse momentum (signifying high energy neutrinos), which act as 'signatures' indicating decay from high mass fundamental particles, and triggering instructions to preserve that event record. Stable particles will not decay over the 25ns timespan of a typical event window, if they decay at all. Additionally TeV scale particles will not be travelling close to the speed of light, leading to a non-trivial time of flight between event origin and detector instrumentation. Together these factors drastically reduce the trigger efficiency of general purpose detectors looking for SMPs.

The low velocity of SMPs also means they operate in the highly ionising regime of the Bethe equation (2.18), behaving similarly to nuclear spallation ions, with a rapid energy loss that encourages further ionisation. Where SMPs differ from standard model particles is in possessing large energy and momentum, even at low velocities. This means they will ionise for longer, and further into the material, until their kinetic energy is depleted. While this means they have a distinct ionisation profile vs LHC background ions, it also means they are prone to being absorbed by detector material, especially dense hadronic calorimeters (which rely on imprecise sampling). Some SMP candidates carry multiple electric charges, further increasing their ionising potential. Concerns over channel saturation are valid for Exotic SMPs as well.

2.9.1 Exotic candidates

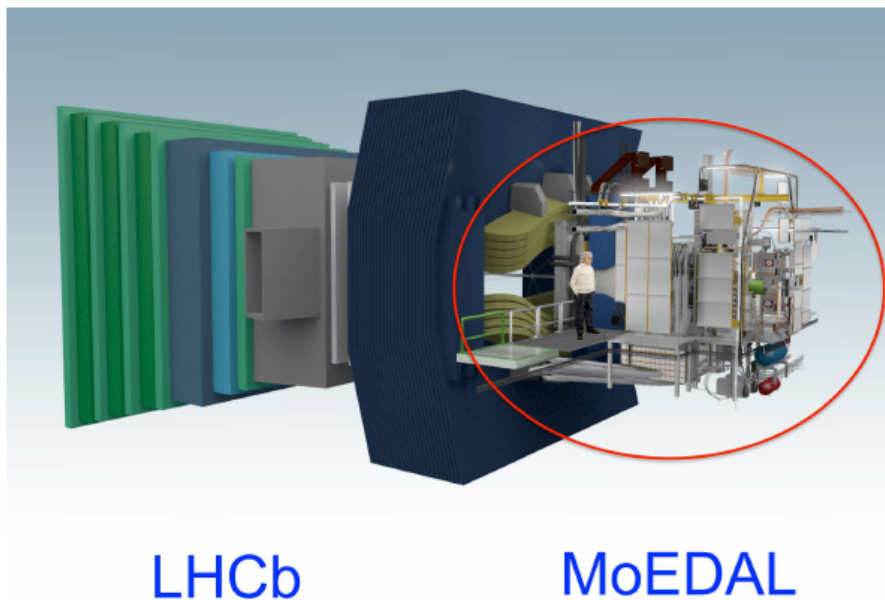
There are multiple theoretical particles that will produce an SMP or HIP signature

- Dyons. Which are a generalisation of magnetic monopoles, possessing both electric and magnetic charge. Also fundamental skyrmions, which are topological solitons of the magnetic field (as opposed to the Higgs field).
- Sparticles, the super symmetric partners to standard model particles predicted by SUSY theories. Conservation of symmetry charge quantum numbers would make them stable in the absence of available anti-partners to annihilate with. They would also possess a very large rest mass in the TeV range. In the Minimum Super-symmetric Standard Model (MSSM) stops, staus, charginos, and gluinos may all be SMPs. R-hadrons may also be detectable.
- Heavier quarks, and Leptons, that would extend our existing 3 generation standard model (where leptons and quarks are arranged in 3 generations of otherwise identical particles of increasing mass). Leptoquarks, interacting both hadronically and leptonically, emerge from technicolour extensions of the standard model. Prior searches have placed their mass limit above 1 TeV.
- Multi charged particles and High electric charge objects (HECOs) that possess multiple electric charges. Higgs triplet models predict a double charged version of the Higgs boson. Depending on model parameters these particles may be pseduo-stable, in addition to being massive, slow, and multiply charged. Pentaquark hadrons and other forms of quark matter [26] also provide the possibility of heavy objects with multiple electric charges.
- Black hole remnants. Mini black holes produced in LHC collisions [32], whose mass and charge is the aggregate of the particles that formed the black hole.

Magnetic monopoles and exotic HIPs produced at the LHC will be very rare, and much slower than most collider particles. While general purpose detectors are theoretically capable of detecting such particles practical detection is limited by velocity and triggering. Reconstruction is likely to be incomplete, limiting what we could learn from such a detection. Large systematic uncertainties make indirect statistical reconstruction an unfavourable search method. The MoEDAL experiment aims to address these complications with a direct search approach that does not rely on triggering or event reconstruction, and is optimised for detecting low velocity heavily ionising particles. This experiment will be described in the next chapter.

Chapter 3

The MoEDAL Experiment



"A major concern at the LHC has been whether it can satisfy the conditions to produce a monopole. It can and MoEDAL has the best chance to find it."
- Prof Y.M.Cho

The CERN Large Hadron Collider (LHC) is the largest particle accelerator built to date. Its centre of mass collision energy of 14TeV is high enough to produce TeV scale monopoles, and exotics. The LHC currently represents the most likely source of man-made magnetic monopoles. There have been monopole search experiments [9][4] with every generation of particle colliders, establishing higher and higher exclusion limits on the possible mass of magnetic monopoles. For the reasons mentioned in the previous chapter (§2.8) most of these searches have relied on nuclear track detectors (NTDs). The Monopole and Exotics Detector At the LHC (MoEDAL) is a dedicated heavily ionising particle (HIP) detector explicitly designed to search for magnetic monopoles.

This chapter will describe the LHC, and the MoEDAL experiment, along with its sub-detectors. A general overview of the nuclear track detectors and their operating principles will be given. There will be an extended discussion of the NTD arrays used in MoEDAL, and the unique considerations that apply to using nuclear track detectors at the LHC.

3.1 The Large Hadron Collider

The Large Hadron Collider is a hadronic particle accelerator based at CERN in Geneva, Switzerland. The LHC is located in the 27km circular tunnel formerly used by the LEP2 experiment. The LHC is (primarily) a proton-proton collider. The LHC operating schedule alternates between gathering collision data (runs), and long periods of shutdown for repairs, maintenance, and upgrades. During $p - p$ collisions the LHC has a 14TeV maximum centre of mass energy (CoM), with each beam having an energy of 7 TeV. Collisions are generated at 4 interaction points (IP) where the main experiments, namely ATLAS; CMS; ALICE; and LHCb; are located 3.1.

- Atlas and CMS are general purpose detectors covering a 4π solid angle. Primarily intended to find the Higgs boson.
- The ALICE detector is a special detector designed to probe QCD using heavy ion collisions.
- LHCb is designed to study CP violation in b-quark systems. The b-hadrons are heavily boosted, hence LHCb can use an asymmetric design that only covers a 2π solid angle around the interaction point. The remaining 2π angle is used for the physically smaller MoEDAL experiment.

3.2 The MoEDAL Detector

MoEDAL stands for Monopole and Exotics Detector at the Large Hadron Collider, and is the 6th LHC experiment. The design focuses on a direct search for magnetic monopoles, and other electromagnetically anomalous particles. MoEDAL's primary sub-components consist of passive detector material placed around the LHCb Interaction Point 8 (IP8) in the VELO cavern, covering a 2π solid angle opposite the LHCb detector. Using passive detectors MoEDAL avoids the triggering issues (§2.8) associated with exotics searches

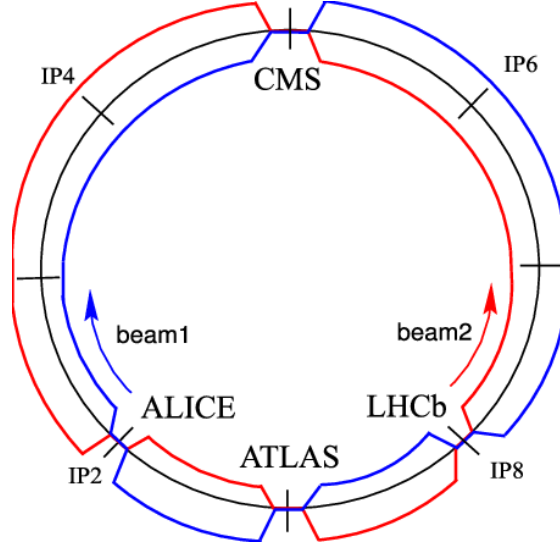


Figure 3.1: Schematic showing the location of the LHC experiments and interaction points (IPs). At the detector IPs the particle beams cross each-other and detectable interactions occur.

which use active detectors. Moreover the final signal is designed to exclude every possible standard model background. The passive components are sensitive during all parts of an LHC run, including ramping up and down, and potentially during ALICE heavy ion runs (currently heavy ion collisions don't occur at IP8). There are two primary sub-detector assemblies in which we look for direct evidence for the passage or presence of a monopole, stable massive particle (SMP), heavy charged object (HECO), or other exotic particles:

- Nuclear track detector (NTD) arrays, which consist of layers of ionisation-sensitive polymer foil encased in aluminium housings. Ionising particles passing through a foil create an ionisation trail of residual molecular damage through the polymer. The ionisation trail is known as a latent (ion) track. By exposing the foil to a chemical etchant that dissolves the polymer, it is possible to turn the latent track into a microscopically visible structure known as an etch-pit. The trajectory and ionising energy loss of the particle may be obtained from the etch-pit geometry.
- The magnetic monopole trappers (MMTs)[?][15], large aluminium blocks which

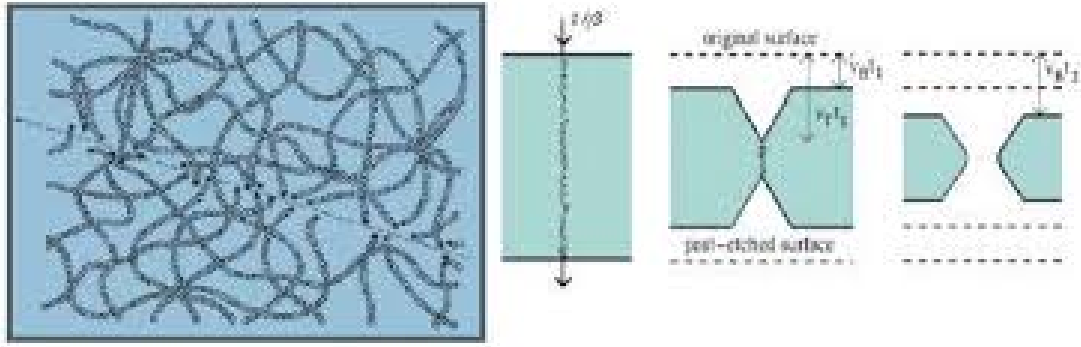


Figure 3.2: Ion track and the etching process

exploit monopoles susceptibility to ferromagnetic trapping. The large restricted energy loss (REL) of monopoles in matter is amplified in conductors and ferromagnets. Electrostatic image forces are able to trap a magnetically charged particle with low enough velocity. Material that contains a trapped monopole will induce a permanent current through a super-conducting quantum interference device (SQUID)(§2.7). SQUIDs can be calibrated by superimposing magnetic dipoles end-to-end until the net separation between the final end poles becomes macroscopic. Each pole will then behave locally like an isolated magnetic monopole (akin to the principles behind the Dirac string).

- There is also a third sub-detector called MAPP (MoEDAL Apparatus for Penetrating Particles), that is physically separate from the main detector and heavily shielded. In contrast to the others, it is optimised to search for weakly interacting millicharged or neutral particles. MAPP utilises plastic scintillator bars, each monitored by a low-noise photo-multiplier tube.

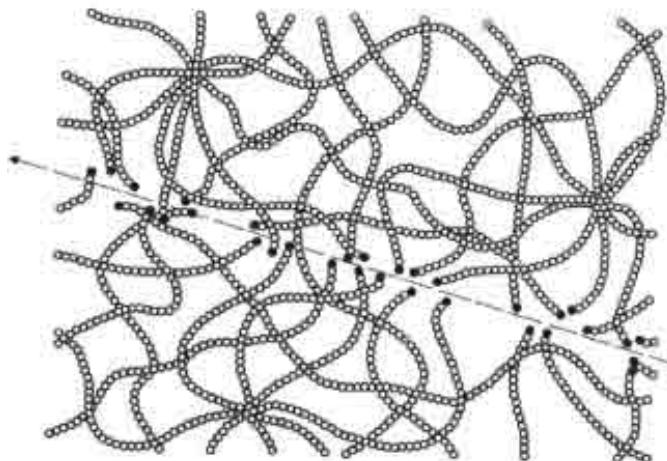


Figure 3.3: Ionisation damage to NTD polymer bonds

3.3 Nuclear Track Detectors

Nuclear track detectors are an effective tool for studying ionising particles. They provide accurate measurements of ionisation energy loss. They are particularly useful when studying spectra of ionising particles and can often identify the incident ion. They have advantages in being cost-effective and durable. Common applications include radiation environment dosimetry, and calibration of particle beam based medical therapies. NTD calibration is achieved through choice of polymer, foil thickness, and the parameters of the chemical etching process. NTD foils are well suited to automatic scanning and analysis. It is possible to forgo etching entirely and directly examine the ion tracks under an electron microscope. This provides exceptional trajectory resolution at the cost of vastly increased effort and expense. An intermediate imaging technique, less complex than electron microscopy, but more involved than regular microscopy, is time-lapse etching where a sample is imaged continually during the etching process detailing the etch-pit formation at each stage.

3.3.1 Ion track formation and Restricted Energy Loss

The ionisation damage caused to the polymer structure of NTD plastic is localised to a nanoscopic region, moreover it is closely linked to the restricted energy loss (REL) of the ionising particle. The Bethe equation (§2.18) describes the expected mean energy loss rate of an ionising particle in matter, with energy lost through individual collisions with a discrete transfer of energy T . Higher velocities produce inefficient ionisation. Different mediums have different energy thresholds for ion track formation; if ionisation energy transfers are too low then the polymer bonds won't be broken, while higher energy transfers (above a threshold T_{cut}) reduce the density of molecular damage to the point where the track will not be detectable. In the electronic excitation regime ($\beta < 0.01$) of the Bethe equation (§2.18) the restricted energy loss corresponds to the total energy loss of the particle. The restricted energy loss is defined as the Bethe-Bloch energy loss restricted only to energy transfers $T < T_{cut}$. Minimally ionising high energy particles will not show up in nuclear track detectors, while they might have the occasional hard scattering in the detectors this will not result in the consistent efficient ionisation needed to leave a detectable ion track. REL can be related to $\left(\frac{Z}{\beta}\right)$ via the Bethe equation. For an ion species with a known nuclear charge this can be used to measure the velocity β .

3.3.2 NTD etching and analysis

Etch-pits form when nuclear track detectors are exposed to a chemical etchant. Etch-pits form because the polymer etches away faster along the latent ion track than in undamaged bulk medium. The 'etching velocity' V_T along the latent track, which is the physical speed at which the NTD material is etched away is proportional to the energy loss of the ionising particle. The bulk etch rate V_B of the material where there are no ion tracks, and V_T , determine the geometry of the resulting etch pit. If the density of broken bonds is too low to impact the etch-rate along the track then $V_T < V_B$ and no etch-pit will be formed. For a single particle the pit will be conical, with the cone-base determined by the

restricted energy loss (REL), provided that the etching is calibrated. If the etching is not calibrated then etch-pit formation will eventually plateau and REL will no longer scale linearly with pit size. The displacement of the point of the cone and the ellipticity of its base can be used to determine the trajectory (θ, ϕ) of the particle (where θ is the angle relative to the normal vector of the foil plane, and ϕ is the polar angle within the foil plane). For a latent ion track traversing an NTD foil, there are three possible scenarios for etch-pit formation:

- **Ranging in** occurs when a particle only loses sufficient energy to begin efficient ionising within a foil. This will produce a single etch-pit that appears to exit the foil.
- **Etch-pit pairs** form where a particle both entered and exited the foil, assuming the particle was maximally ionising at both points, with the cone of each pit being aligned to one another. The difference in size ($\approx \text{REL}$) between the entry and exit pits effectively gives us the gradient of the Bethe equation with respect to β , further aiding in identifying the particle.
- **Ranging out** occurs when the particle loses sufficient energy to cease ionising within that foil. Only a single entry etch-pit will be present, as the particle won't have enough energy to leave an exit pit.

By combining NTD foils into layered 'stacks' it is possible to chart ionisation stopping power as a function of β , as a particle enters a stack and slows down, losing velocity. Any particle entering a given medium will have its own distinctive stopping power curve, however all standard model particles follow the same general shape. The stopping power of magnetic monopoles has a unique β dependence due to the magnetic charge interaction (§2.6.3) resulting in a stopping power curve completely distinct from all standard model particles. For this reason, NTD foil stacks have been a cornerstone of most magnetic

monopole searches as the evolution of ionisation behaviour for a magnetic monopole through an NTD stack has no standard model backgrounds.

3.4 MoEDAL NTD Arrays

The MoEDAL NTD arrays are composed of modular NTD stacks laid out in an approximately cubic shape around the walls and ceiling of the LHCb VELO (Vertex LOcator) cavern. The exact array geometry is modified between LHC runs to accommodate changes to LHCb detector infrastructure. The stacks cover an area of 25m^2 and each individual stack measures 25cm^2 arranged in groups of six encased in modular housings measuring $75 \times 50\text{cm}$. The housings are made from commercially pure aluminium 1mm thick. The housings help shield the detectors from the highest levels of LHC spallation radiation, filtering out weakly penetrating ions. The plates also allow for a degree of ferromagnetic trapping similar to the MMTs. In between LHC runs the arrays are removed for analysis and replaced.

MoEDAL NTD stacks contain up to 9 layers of NTD detector foils. Most foils are 0.5mm thick, with 0.2mm foils at the front back, and middle of a stack. The stacks contain a mix of low threshold and high threshold NTD foils. Stack arrangement changes between runs, either to optimise for different run conditions, or based on feedback from experience. In addition to the regular NTD stacks there are the VHCC (Very High Charge Catcher) arrays which operate on the same principles but optimised for HECO searches. The polymers used in the NTD arrays are;

- **CR39** - Low threshold highly sensitive detector material with a threshold of $\frac{Z}{\beta} = 5$. The CR39 layers are optimised to allow accurate measurement and reconstruction of exotic candidate tracks after they have been found.
- **Makrofol** - high threshold detector material placed at the front of the stack where the density of the nuclear spallation background is highest. The Makrofol stacks

are used to identify etch-pit anomalies corresponding to candidates.

The REL thresholds of the polymers render the detectors totally insensitive to minimally ionising particles. As a consequence there are only two types of particle capable of producing an ionisation track in a MoEDAL nuclear track detector:

- **Heavily ionising standard model particles** - These include heavy nuclear ions with singular or multiple ion charge travelling with a low velocity and typically low momenta. There are two main sources of such spallation ions at the LHC; secondary and tertiary interactions of the particles coming from the primary interaction point, and beam collisions with residual beam-pipe gas atoms.
- **Heavily ionising exotic particles (HIPs)** - Including magnetically charged particles, stable massive particles (SMPs) such as stops and staus, multi-charged objects including multi-charged SMPs. These particles will come from the primary interaction vertex, and will possess enormous kinetic energy and momenta.

It can be observed that there is a gap in physical behaviour between low energy highly ionising composite nuclei, and high energy highly ionising fundamental particles. The MoEDAL NTD stacks exploit this gap. The thickness of the MoEDAL foils is chosen so that most nuclear ions capable of penetrating the aluminium housing, with a restricted energy loss high enough (β low enough) to range in and leave an ion track, will range out due to energy losses before leaving the foil. Within an etched foil most nuclear ions will only produce a single etch-pit, corresponding to ranging in or ranging out (on the back, and front surfaces respectively). Where a few nuclear ions are able to produce a through-going ion track, the size of the etch-pit pair will be highly asymmetrical (reflecting the gradient of the Bragg peak). Only high energy heavily ionising particles are able to sustain the energy losses associated with ion track formation throughout multiple foils, producing (with etching) symmetrically sized etch-pit pairs in each foil they pass through.

3.4.1 MoEDAL HIP signature

It is possible for nuclear ions to mimic an exotic ionising particle in a single NTD foil layer, it is not possible for them to mimic their behaviour through an entire detector stack. NTDs can allow for extremely accurate track reconstruction, and if calibrated correctly they allow for very accurate measurements of relativistic energy loss. The first definitive sign of an exotic HIP would be a series of co-linear entry and exit etch-pits or ion tracks suffering none of the deflection we might expect from low momentum spallation nuclei. The second indication would be a substantially greater range in the material, likely penetrating many or all the layers, while spallation ions will range out very rapidly through a combination of low energy and high energy loss. Thirdly an exotic candidate track would point back to the primary interaction vertex, which can be reconstructed to within 1cm. Lastly for a magnetically charged particle we would observe a non-standard model change in the ionisation rate through the foils. Nuclear ions would see a spike in their ionisation energy losses as they came to a halt, while a monopole or dyon would see a slow decline.

3.4.2 NTD Analysis process

The nominal MoEDAL NTD analysis chain is described below;

- The NTD stacks will be removed from the VELO cavern at the end of each run, when radiation levels make them safe to access and handle. The exposed NTD stacks will each have three alignment holes precision drilled through the stack ($50\mu m$ tolerance). The holes are 1.9mm in diameter. These can be used to reconstruct the track of detectable objects as §7.2.2 will demonstrate.
- The first foil in the stack, closest to the interaction point, undergoes a strong etching to reveal the ionisation etch-pits. The etched foil is scanned, and manually examined twice under low magnification with stereo microscope to look for corresponding

anomalous etch-pits on both sides of the foil that could indicate a HIP candidate. While laborious, this process allows for immediate investigation of candidates with the capacity to change focus, magnification, and illumination. An experienced NTD expert can reject anomalies that could be due to complicated spallation processes, defects in the foil, effects of the etching or combinations of the above.

- If a candidate is found it will be examined in detail at high magnification (2040X). Presuming the HIP hypothesis is not rejected, the geometry of the front and back etch pits will be measured. This is used calculate the REL, an angle of incidence (θ, ϕ) , for each etch-pit. An agreement to within 20% between front and back is required. The expected location of the ion track in down stream foils can be estimated.
- The last foil of the stack will be etched under soft etching conditions. Ethyl alcohol added to the etchant increases the REL detection threshold, reducing the visibility of background particles. Soft etching allows for higher quality cone measurements. The foil analysis is repeated in an area around the hypothesis candidate site. If the HIP hypothesis is confirmed again then the middle foil is etched and the process repeated again.
- At this stage we will have a strong candidate track, with 6 coinciding etch-pits, each with an anomalously large ionisation energy loss. Careful etching and cone measurements of the remaining foils would provide the measurements of trajectory and REL to invalidate other hypotheses. Finally nuclear track detectors represent an analogue medium that is not tied to the etching process. For an event that strongly represented an exotic discovery the un-etched foils would be analysed under electron microscope. Every nuclear ion track in the vicinity of the candidate track would be painstakingly reconstructed and matched to a specific atomic mass and ion charge.

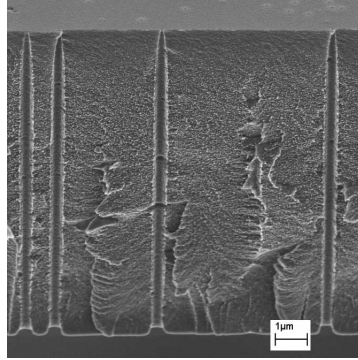


Figure 3.4: Close up cross-section of an ion track imaged with an electron microscope

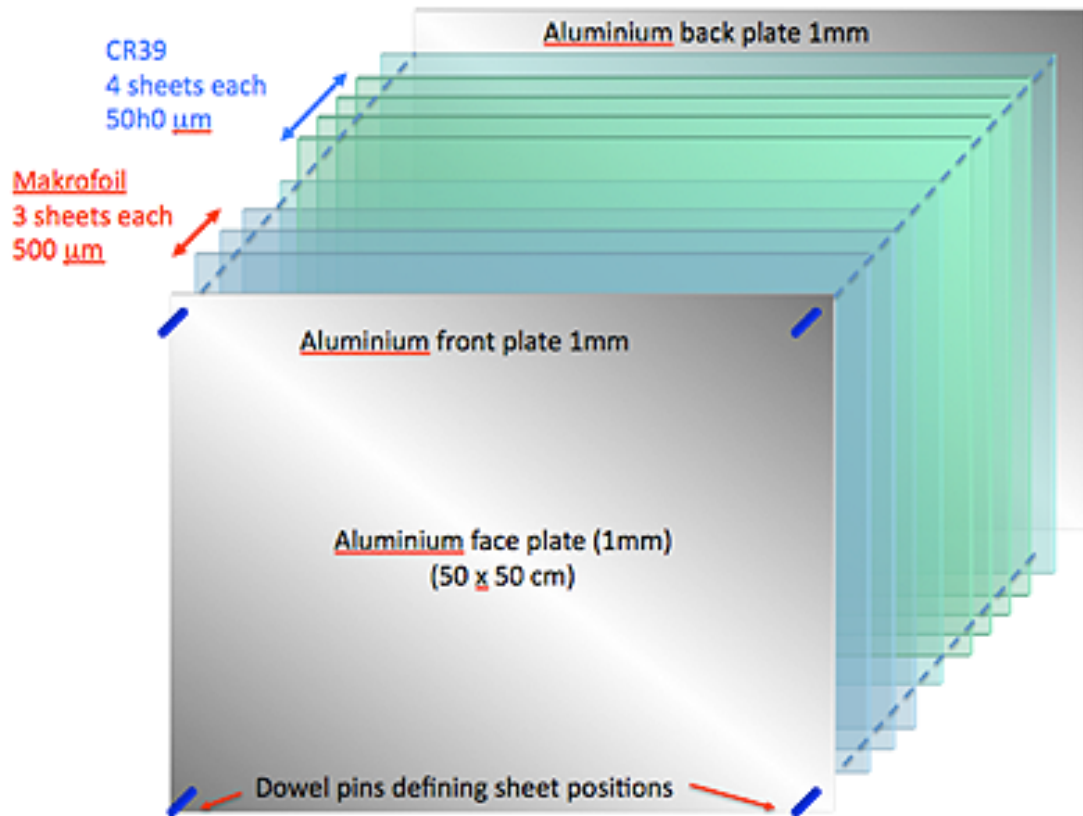


Figure 3.5: Different stack compositions are used on different LHC runs, and for different detector regions such as the VHCCs

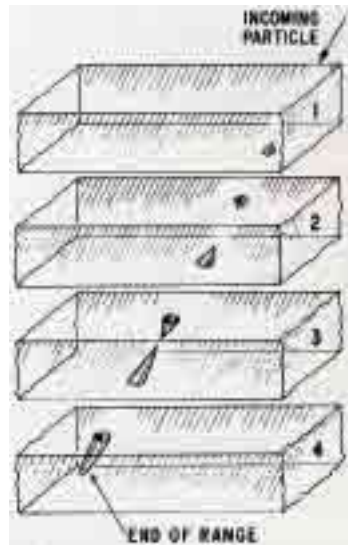


Figure 3.6: Note: MoEDAL NTD stacks are designed so that most background particles will range out within a single foil layer

3.4.3 LHC backgrounds and calibration

The LHC is the highest luminosity particle collider ever constructed, with a huge background radiation flux of highly ionising, minimally ionising, and theoretically non-ionising particles. This poses unique challenges in terms of NTD analysis, especially with the long 2 year exposure time of each LHC run.

LHC background conditions

The main NTD detector background consists of low velocity nuclear spallation ions. The numerical density of this background falls exponentially with distance in the detector medium, with a typical stack being 3.6mm thick. The front layers of each stack are exposed to large doses of ionising radiation, both from gamma rays and ionising heavy particles. The large flux of nuclear ions creates a large visual background for NTD image analysis. In the foils with highest exposure we no longer observe singular isolated etch pits, instead the pits are extremely clustered. Overlapping or densely packed ion tracks frequently merge non-linearly under etching, and etch-pit geometry only approximately

resembles the idealised cone. In addition to the visual impact, the density of ion tracks in LHC exposed NTD detectors is such that it modifies the effective bulk etch-rate V_B of the detector material compared to an unexposed foil. At ion track densities of 100 million per cm^2 the ionisation damage to the foil is so great that it will disintegrate completely under etching. While the gamma radiation does not leave an ion track in the detector it will change the chemical properties of the polymer and alter its detector response compared to an un-irradiated NTD foil, again impacting the effective bulk etch-rate.

At the first order the response of the detectors (in terms of etch-pit size and appearance to REL) must be recalibrated to account for LHC conditions. A second order effect will be local variations in background density, and thus etch-rate. This also results in local variations in NTD foil thickness meaning that the focal depth required to optimally image the etch-pits will vary across a MoEDAL exposed NTD foil, as compared to unexposed foils.

Calibration

NTDs are calibrated by exposing them to heavy ion beams, using a range of ions and energies. The goal of calibration is to match the Restricted Energy Loss (REL), through polymer ionisation, expected of a magnetic monopole. In practice this means selecting suitable etching conditions and measuring the etching response of the detector foils under appropriate LHC background conditions. The heavy ion test beams can be used to match the apparent signal of an exotic HIP within a given foil layer, at a well defined cross-section. Different ions are used to simulate different parts of the potential HIP energy spectrum, lead ions may be used to simulate more energetic monopoles. For a calibrated detector the response (ie, area of an etch-pit) is linearly proportional to the REL, which is governed by $\frac{z}{\beta}$, and related to the bulk etch-rate. By exposing the NTD to the test beam ions (with a known $\frac{z}{\beta}$), and plotting a histogram of the cone area of identifiable etch-pits, one can identify the peaks corresponding to known ions and

their fragments. This is used to establish the exact relation between $\text{REL}(\frac{z}{\beta})$, etch-pit area, and $\frac{z}{\beta}$. The effective bulk etch-rate V_B can be calculated and used to identify the $\frac{z}{\beta}$ of unknown particles. This calibration process also provides an insight into the visual appearance of heavily ionising particle (HIP) etch-pits in heavy LHC background conditions. The non-linear response of the NTD detectors to different etching conditions and material thresholds means the practical appearance of these etch-pits can vary wildly.

Summary

Analysis of the first foil layers is important because many exotic particle candidates other than fast monopoles might only be seen in the early layers. Finding and locating candidates early also allows us to calibrate subsequent etchings to the linear detector response regime of the hypothesised particle we are trying to reconstruct. Adding higher concentrations of ethyl alcohol can raise the effective REL threshold of the foil, eliminating certain etch-pit backgrounds. Done poorly this can reduce the visibility of the signal we are trying to find.

The remainder of this thesis focuses on machine learning techniques for identifying and tagging prospective HIP candidates in this first foil layer where backgrounds are highest, and the appearance of the signal we are looking for is most uncertain. The very high background densities, and variability in possible signal due to theory parameters, etching conditions, and imaging have meant this has been an expert human task. It is hoped that an effective method for tagging HIP candidates and anomalies will speed up MoEDAL NTD analysis, allowing attention to be focused on the most promising candidates.

Chapter 4

Machine Learning

Machine Learning (ML) is increasingly used to automate tasks that call for a human level of insight and understanding. In High Energy Physics machine learning methods have been applied to event selection and particle identification, delivering greater sensitivity and discriminating power than manually derived selection criteria. This thesis is based on applying machine learning to the MoEDAL NTD image analysis, specifically Image recognition based on Convolutional Neural Networks. The aim is to accelerate and automate the preselection of etch-pits corresponding to Heavily Ionising Particle (HIP) candidates. This is complicated by the requirement for very high signal efficiency across a range of possible HIP signatures. Even for an LHC experiment MoEDAL features an extremely high background density due to the passive detector design, that records an entire run's worth of ionisation data simultaneously. Any ML method adopted for MoEDAL must also possess an extremely low false positive rate, and be robust to the clustering effects of this background.

4.1 Machine Learning

Machine learning typically resembles high dimensional function fitting (4.1). The inputs x , outputs y , and parameters θ , can be multi-dimensional and may represent tens, thousands or millions of degrees of freedom. During the ‘Learning’ process, we iterate over θ to find ‘optimal’ parameters that minimise or maximise a given performance metric. Relative advances in the last decade come from the scale, ie number of parameters allowing a very high degree of fitting power, and effective back-propagation algorithms for optimisation which has allowed machine learning to be applied problems of greater semantic complexity.

$$y = f(x|\theta) \tag{4.1}$$

4.1.1 Supervised and Unsupervised Learning

There are two main approaches to machine learning; Supervised, and Unsupervised. In a supervised learning problem data comes in the form of a set of labelled pairs $\{X, Y\}$ of input x and output y examples. The task is to learn an approximate mapping between X and Y , from the training samples that will hold for new unseen examples. In a typical image recognition task x would be an image and y would be a vectorised set of classification labels. An adjustable parametrised mapping known as a model is proposed, which dimensionally matches the required inputs and output. During the optimisation phase (‘Training’ or ‘Learning’) the parameters of the model are adjusted to minimise a Loss function (4.2), which expresses the mathematical distance between the models predictions and the true labels.

$$\mathcal{L}(f(X_i|\theta), Y_{\text{truth}}) \tag{4.2}$$

In unsupervised learning, we have raw data without labels. The problem is typically framed as finding an optimal decomposition of the data. Unsupervised learning has the

advantage that it scales very well with more raw data, since labelling and label accuracy is typically a bottle neck in supervised learning.

4.1.2 Classification and Regression

Machine learning is typically divided into classification tasks and regression tasks. In classification tasks the aim is to map the input to a set of discrete output labels. Regression typically features a continuous output. Many classification tasks map the input data to a probability or likelihood of belonging to various categories, with a final step casting this continuous regression type output to discrete classes.

4.2 Neural Networks

Neural networks are a form of biologically inspired machine learning model [21], consisting of a network of virtual neurons and synapses connecting a set of input neurons (x) to a set of output neurons (y). Inspired by synaptic connections in the living brains the model adjusts by altering the 'weight' given to different synaptic connections (Fig:4.1).

The practical strength of Neural networks comes from grouping neurons into vector 'layers' $x^{(i)}$, where i represents the i 'th layer. The layers allow the model to be expressed, and decomposed into an alternating sequence of two mathematical operations that determine the value of the next layer in the network $x^{(n+1)}$; Linear transformation of $x^{(n)}$ by a matrix $W^{(n)}$ that represents the weights connecting one layer's neurons to the next. Followed by an element-wise transformation by a non-linear layer specific 'activation' function $a^{(n)}$.

$$x^{(n+1)} = a^{(n)}(W^{(n)}x^{(n)} + b^{(n)}) \quad (4.3)$$

This decomposition allows for efficient parallel computing, especially via GPUs which

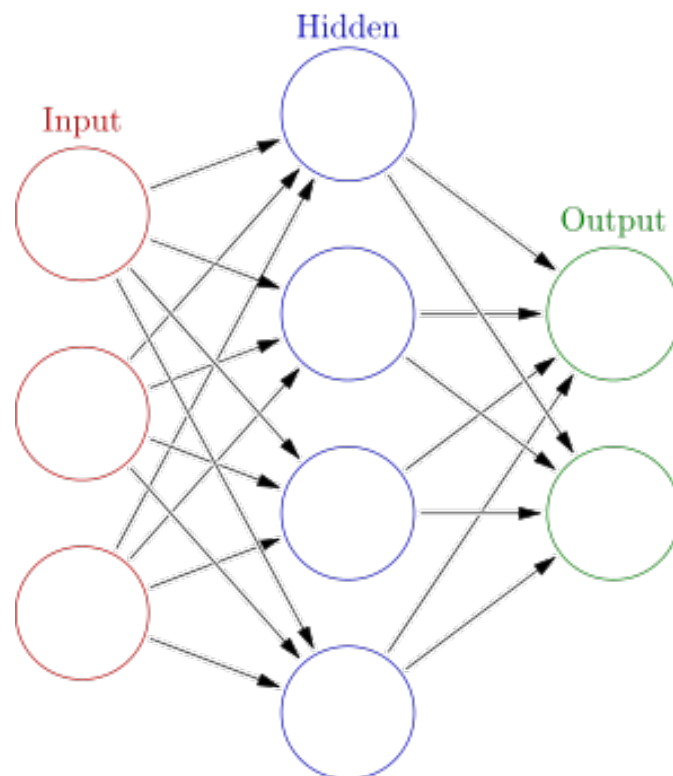


Figure 4.1: A simplified neural network with a single hidden layer. Each neuron is connected to every neuron in the next layer, and the preceding layer. The arrows represent the weights connecting each neuron to those in the following layer. [25]

utilise thousands of low speed processing cores. Given sufficient neurons and weights a neural network can approximate any mapping from $\mathbb{R}^n \Rightarrow \mathbb{R}^m$ to an arbitrary degree of accuracy. This is known as the Universal Approximation Theorem [28]. Human input comes in choosing the 'architecture' ie, the size and type of each layer, and number of layers, and 'hyper-parameters' eg, choice of optimiser, how many iterations over the dataset and activation function type. Networks, or layers of a network where every neuron is connected to every other neuron in the subsequent and previous layers are referred to as dense (or fully connected). An important factor affecting practical learning performance is the shape of the data, or representation $x^{(i)}$ as it passes through each layer of the network. For dense networks the (flat) shape is simply the number of neurons in each layer, for more complicated networks the shape is treated as an N-dimensional array.

4.3 CNNs: Convolutional Neural Networks

A convolutional kernel filter is a method in image processing and image recognition. An image is combined with a filter to produce a transformed image, or feature map. The pixel intensity of the feature map represents how closely the equivalent region of the input image matches the pattern of the filter. This is equivalent to the mathematical convolution of a function $f(x)$ with a Kernel $g(x - x')$, if we treat the pixelated image as a discrete representation of the continuous distribution $f(x)$. Kernel filters can be used for either pattern recognition or image transformation.

$$(f * g)(x) = \int f(x')g(x - x')dx' \quad (4.4)$$

Convolutional Neural Networks combine Neural networks with convolutional filters. Instead of a fixed kernel CNNs are free to optimise their own set of filters. At each point and locality in the input the CNN kernel acts like a mini dense neural network. An

alternative view would be that each CNN filter represents the convolution of an image with a smaller dense neural network. By passing the output of one convolution layer to another convolution layer as input, CNNs are able to form a hierarchy of recognised structures and patterns, starting from basic features like edges and curves, up to complex objects such as faces or kittens. CNN's typically feature a repeating two layer structure known as a Conv-Pool [51] operation where convolution operations (layers) are paired with pooling layers;

- **The convolution operation (Conv)** convolves the input along its spatial dimensions (e.g., x, y) with a set of filters F producing a higher dimensional output.
- **The pooling operation (Pool)** compresses the spatial information of the convolution output in a lossy manner. For example 2x2 Max-pooling takes the highest value in each 2x2 square of the original input.

In each Conv-pool step the spatial dimensionality of the representation $x^{(i)}$ is reduced, while the semantic dimensionality (or 'feature depth') is increased. By reducing the overall size of the representation we hope to force the network to learn the most useful features that represent the input image.

4.3.1 FCNS: Fully convolutional networks

CNNs are typically used to map images to non-spatial outputs such as categories or numbers. Fully Convolutional Networks [34](FCNs) preserve spatial dimensions in their output, mapping images to images. This may be accomplished by foregoing some or all of the pooling layers, preventing a reduction in spatial resolution. Alternatively FCNs can make use of 'Upsampling' layers which function like convolution layers in reverse. The input representation is convolved with upsampling kernels, and the output summed over the input features. This typically reduces the feature depth, and increases the spatial

width of the representation. FCNs of particular interest are U-Net architectures [49]. U-nets architectures produce outputs that have the same spatial shape as the input, which makes them useful for image transformation and segmentation tasks. U-Nets can also be used as auto-encoders (networks where the input data X is also the output target $X, Y = X$).

4.4 Optimising and Training Neural Networks

Machine learning models are typically trained and executed using software libraries (Machine Learning Frameworks) that provide a standardised infrastructure for parameter optimisation and data handling. The frameworks used in this thesis consist of Tensor-flow [1] in conjunction with Keras [12]. During training, data is presented to the model in batches, the loss function is calculated for each batch and the model weights/parameters are adjusted by an optimisation algorithm before the next batch. Optimisation algorithms normally rely on gradient descent and back propagation to calculate the weight adjustments needed to minimise the loss function. Training typically iterates over the entire dataset multiple times, with either a set number of repetitions ('epochs') or until performance is deemed to have either converged or over-trained.

4.4.1 Cross-Validation and Over-training

Prior to optimisation the input data $\{X, Y\}$ is split into testing $\{X_{Test}, Y_{Test}\}$ and training $\{X_{Train}, Y_{Train}\}$ subsets. The test data provides an independent statistical estimate for the model's performance after (and during) training. A divergence between training performance and testing performance means over-training has happened, meaning the model is optimising for the features of individual training examples rather than learning generalisable rules. While methods (such as dropout) exist to reduce over-training it is pragmatic to cease further training epochs if over-training has occurred. Over-training

is heavily dependant on the number of data examples N relative to the number of parameters θ in the model. An excess of parameters or a shortage of data encourages fitting to the specific training data.

Optimisation may be repeated multiple (R) times with different randomised initial weights, producing different models ('learners') each time. Having a set of models provides an estimate for the average performance of the model architecture and hyper-parameters, and indicates if the learning process is stable and repeatable. To obtain a final result the models may be averaged, the best performing model may be selected, or a new model may be trained via the same method using all available data.

Commonly hyper-parameters will be adjusted by the experimenter (or via algorithm) in order to optimise test performance. Best practice involves withholding an additional validation dataset. This dataset provides an independent estimate of the final performance vs unseen data, and guards against bias in selecting the model that best performs against the test data.

4.4.2 K-fold Cross-Validation

Generic K-fold Cross-Validation is a data validation scheme where a given metric is evaluated over the data multiple times (K), each time withholding a subset of the data known as a fold (K^i). K-folding provides a more accurate estimator for the metric that accounts for variability due to the choice of data used to evaluate the metric.

In the machine learning context [39] the dataset is split into 'K' subsets known as folds, which are rearranged into 'K' test train pairs. The i 'th test set comprises the i 'th data-fold, whilst the remaining ($K-1$) folds form the i 'th training set. A separate model is trained and evaluated for each test-train fold. K-folding allows all available training data to be used, while providing an ensemble of models (and performance estimates) that accounts for variations in the training and evaluation data that may otherwise lead to spuriously good or bad results.

K-folding is often combined with repetition in a scheme known as nested K-folding. Learning over each fold is repeated R times producing $R \times K$ learners. This gives a very good indication of the performance and stability of the network and hyper-parameters.

4.4.3 Data-Augmentation

Data Augmentation [?, dataaug] is a technique used to improve the performance and robustness of a model by increasing the amount of effective data N' relative to model parameters θ . Input data $\{X, Y\}$ is duplicated, with the duplicate data being transformed in such a way that the output target is invariant $\{X', Y\}$. For example, a cats vs dogs CNN classifier should be invariant to horizontal reflections, doubling the effective data available. A fruit classifier should be rotationally symmetric, allowing each data example to be included again with any arbitrary rotation angle.

4.4.4 Transfer learning

Transfer learning [6] is a useful technique for reducing the amount of training (in terms of time and data) required to optimise a model. In transfer learning new layers are combined with the pre-trained layers of an existing network. Existing parameters θ_o are held fixed, and only θ' parameters associated with new layers are adjusted. Transfer learning can produce more robust networks, less prone to over-training by reducing the effective number of parameters θ' relative to training examples, while allowing for complicated models with many parameters $\theta_o + \theta'$

Chapter 5

NTD Imaging study

The purpose of this thesis is to find machine learning techniques for identifying heavily ionising particle (HIP) candidates within NTD foils exposed to the full LHC spallation ion background. To do so we use several research and development foils that have been exposed to various heavy ion calibration beams (Xe and Pb) in addition to exposure at the LHC as part of MoEDAL NTD arrays.

This chapter describes work done to find suitable imaging parameters and scanning techniques that I subsequently used for data collection in the Xenon ion and Pb ion studies described in Chapters 6 and 7. This chapter will start by introducing MoEDAL NTD image analysis and the features of test beam ion track etch pits, and LHC background etch pits. The specific NTD foils used in this research thesis will be described. I will discuss the imaging and scanning of these NTD foils. Finally I will discuss an imaging method I identified in the course of this research, which assists the visual recognition of anomalous etch pits. This has applications to both human and ML based NTD analysis, especially in wide area images and scans.

5.1 Introduction

Nuclear track detectors are an established method for analysing the ionisation behaviour of particles. Particles with an ionisation energy loss large enough to leave an ion track in a nuclear track detector will produce one of the following etch pit patterns discussed in §3.3.2 when the NTD foil is etched;

Ionisation track	Etch pit
Ranging in	Exit pits
Ranging out	Entry pits
Through going ionisation	Entry-exit pairs

The NTD foils of the MoEDAL experiment are calibrated so that only particles with an anomalously large and consistent ionisation will produce ion track etch pits on both sides of a foil continuing in downstream foils. The foil thickness and threshold is chosen so that most standard model backgrounds will range out within one foil.

The objective of the MoEDAL NTD analysis is to identify peak ionisation etch pits occurring on both sides of an NTD foil layer (Through-going), or to look for anomalously large ionisation etch pits within a foil layer. Indicators of either type of signal will be referred to as 'Candidates' or 'Hip candidates'. The signal hypothesis can be verified or tested through further detailed investigation over subsequent foil layers, to see if such candidate etch-pits form a track which could correspond to a magnetic monopole or exotic HIP hypothesis.

5.1.1 Etch pit backgrounds

There is no background process that can produce a HIP signature in an NTD stack, thus MoEDAL is described as being background free. However within a single etched NTD foil there are many physical and visual backgrounds caused by high levels of LHC spallation radiation (§3.4.3) that complicate candidate identification. These are outlined below;

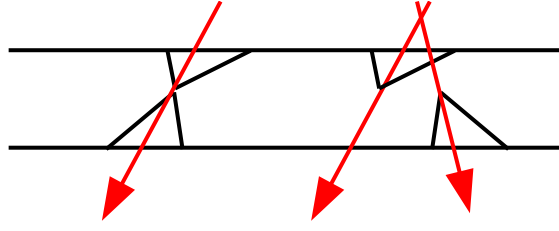


Figure 5.1: HIP candidate vs uncorrelated background

A through going entry-exit etch pit pair formed from a single highly ionising particle track (left). At LHC background density entry and exit etch pits from uncorrelated ion tracks will frequently mimic this entry exit pattern.

Through going standard model background particles will occur rarely. As entry exit pairs they warrant investigation. As etch pit area scales with REL we can expect a large size asymmetry between the entry and exit pit. This indicates a large Bethe gradient of energy loss vs distance, along with its direction. Exotic HIPs possessing high energy and momenta experience little change in ionisation energy loss despite large energy losses. Their entry and exit pits demonstrate a high level of symmetry.

Uncorrelated entry and exit pits corresponding to the uncorrelated background particles ranging in, and ranging out. At high etch pit densities they will frequently overlap. This closely mimics through going ionisation pairs (fig 5.1). Higher order information such as orientation, symmetry of the etch pits, or the presence of a light tube may be needed to make an ID in these circumstances.

Etch pit clusters of exit pits or entry pits act as a background in multiple ways. The chemical etching process can cause etch pits on the same surface to merge non-linearly. Several smaller etch pits can thus become a larger one, indicating a larger ionisation. Imaging LHC exposed foils presents more challenges than unexposed foils and its not always possible to distinguish entry pits and exit pits in 2D images. Clusters can visually resemble a entry exit pairs, and may come from a common spallation process. Candidate etch pits will also be impacted by clustering and merger with background pits, and at

high background densities our task becomes identifying clusters with candidates from those without.

Non-etchpit objects will also contribute a visual background to NTD analysis. An example will be scratches on the surface of the NTD foil that may be exacerbated by etching, along with the possibility of particles of dust being present. In the degraded foil conditions produced by LHC exposure there may be transient visual effects caused by inconsistencies in illumination or focus.

5.1.2 Imaging for machine learning

Nuclear track detector image analysis in an LHC context is an expert task, requiring an understanding of the chemical etching process, and experience of what processes caused by LHC spallation background could form an alternate hypothesis. Prioritising the most promising candidate locations via ML based candidate tagging would reduce the burden on skilled human NTD analysers, who represent a limited labour pool with the requisite LHC knowledge and NTD analysis experience.

The research in this thesis has been carried out using several Research and Development foil samples along with a complete Research and development NTD stack. The physics signature of an exotic HIP is replicated outside of the LHC using calibrated heavy ion beams. These foils are exposed to the LHC as part of the MoEDAL array, generating real LHC spallation background ion tracks. This is the same approach used in the calibration of the main MoEDAL NTD image analysis (§3.4.3). These NTD foils are described in the next section.

A design feature of the MoEDAL experiment is that after etching there is no constraint on the imaging methodology used, the NTD foils are a physical analogue medium that may be imaged repeatedly by different analysers using different hardware, to an arbitrary degree of detail. This study looks into different imaging options for machine learning

based NTD analysis at the LHC. Desirable features and requirements are listed below;

- **Maximise the appearance and identifiability of through-going etch pits**

Ideally we want to allow for time efficient and reliable human ID in order to produce a labelled truth record for the training and evaluation of various machine learning algorithms. This replicates many of the challenges of human NTD analysis, with the distinction that the calibration beam provides a guaranteed signal with a guaranteed cross-section.

- **Scaleability** The MoEDAL NTD foils cover 300m^2 , and each foil can contain many millions of etch pits. Any imaging method has to be capable of imaging large numbers of etch pits simultaneously and allow the rapid rejection of standard and trivial backgrounds. Consideration must be given to the resource limits of data storage, scanning time, and image processing time.

- **Suitable for automated data collection and analysis** During manual NTD scanning and analysis it is possible to immediately adjust magnification, focal depth, and illumination to investigate promising candidates, or account for local foil conditions. Immediate 3D investigation is also possible. We want to identify an imaging process suitable for automated scanning which collects enough information that analysis can proceed from non-interactive scan data alone.

- **Consistency** The optical appearance of etch-pits can change rapidly based on imaging parameters, etching parameters, and background conditions. This has limited the application of ML image recognition at MoEDAL to date. Machine learning ultimately functions as sophisticated pattern matching. We want to reduce the impact of these systematic variations and present a simple and consistent pattern, while minimising the appearance of non etch pit visual backgrounds and objects.

5.2 NTD foil samples

This section describes the ion beam Research and Development NTD foils ordered by complexity of background conditions. Foils were exposed at the NASA Jet Propulsion Laboratory (JPL) with a beam incident at 15deg from normal, focused on the centre of the foils but with ions present throughout. Initial research used a 13GeV/A Xenon ion beam. Later work used an NTD stack exposed to a 156GeV/A Lead ion beam. The beam ions are highly ionising TeV scale particles (with an artificial source). All NTD foils were manufactured by TASL and provided by the MoEDAL experiment. Etching was carried out by the NTD analysis team at INF Bologna. Typical Moedal NTD foils are 250mm x 250mm, however the R&D foils have been cut down to 100mm x 100mm NTD.

5.2.1 Unexposed CR39, with Xe ion signal

CR39 foils are used by The MoEDAL experiment for calibrated precision measurements of candidate etch-pits. This foil has been exposed to a 13GeV/A Xenon ion calibration beam. With no LHC background the HIP/Xenon ion track represents an ideal HIP candidate signature identifiable by a matched pair of entry and exit pits, on the front and back surface of the foil respectively (Figs 5.2a,5.2b). The entry and exit etch-pits are visually symmetric, and of a similar size, indicating minimal change in the rate of ionising energy loss. Adjusting the focal plane via Z-axis motion of the lens allows us to focus on either the entry pits on the front surface of the foil or the exit pits on the back surface. This provides a quick and reliable distinction between entry and exit events for low etch pit densities. The corresponding pit in the entry exit pair can still be seen, although out of focus. The displacement $(\delta x, \delta y)$ of the entry and exit pits, and the known thickness of the foil (δz) , can be used to identify the trajectory of the ionising particle. Under dark field illumination (fig. 5.2c) see §5.3.2 one can make out a faint visible structure

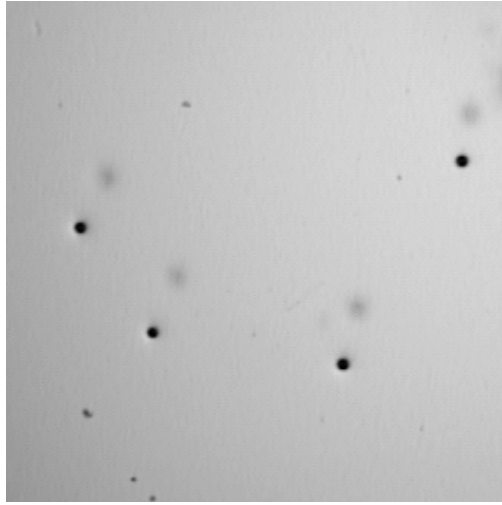
connecting the front and back surface etch-pits. This is where the conical points of the front and back surface etch-pits approach along their shared ion track. These structures are sometimes referred to as light tubes. The presence of a light tube is a key indication that both etch pits come from the same through going ion track. Imaged in isolation either etch pit is representative of ranging in or ranging out. With no background and a known ion beam trajectory it is trivial to match entry and exit pits.

5.2.2 Background free makrofol, with Pb ion signal

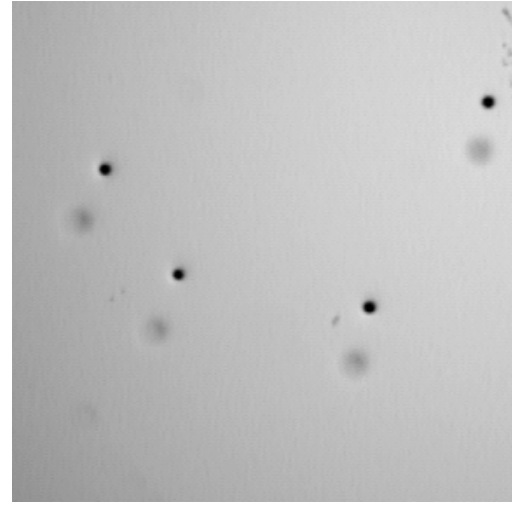
The Pb ion study described in Chapter 7 focused on a dedicated NTD research stack of five makrofol foils exposed to a Pb ion beam. Four of these foils with no subsequent LHC exposure are used to accurately reconstruct the trajectories of the Pb ion tracks. Etch pits left by the 156GeV/A Pb ion beam are distinct and readily identifiable. In the thinner foil the entry and exit pits overlap too much to readily distinguish between them (fig 5.3). Imaging the front and back surface separately via focal plane adjustments is also harder. In addition to the Pb ion etch pits there is also a higher incidence of scratches and other visual imperfections. Despite having had no LHC exposure we can still identify several smaller non-Pb etch pits in multiple images. Some of these will come from Pb ion fragmentations as well as spallation from the test beam exposure. In a 2D image these fragmentations are particularly hard to separate from through going Pb ion tracks (especially since each ion fragment may also form through going ion tracks).

5.2.3 Makrofol with Xe ion signal and 8 months LHC exposure

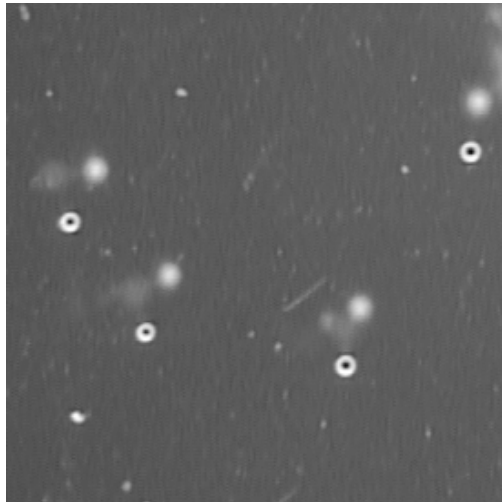
The research in Chapter 6 took place using a makrofol NTD foil exposed to the same 13GeV/A Xenon ion beam as the CR39 foil. This calibration beam represents a lower energy monopole or exotic HIP. The foil was exposed to the LHC as part of the MoEDAL NTD array for 8 months (1/3rd of a full 2 year run) generating a dense pattern of LHC spallation ion etch-pits. Background density varies significantly throughout the foil.



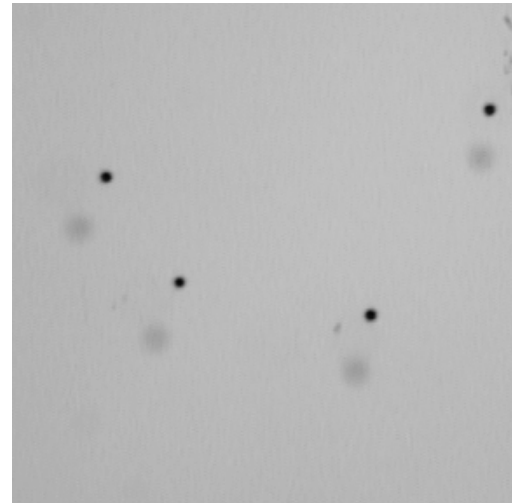
(a) Front surface focus



(b) Back surface focus



(c) Dark field illumination



(d) as (b) except illuminated through the lens

Figure 5.2: Xe ions in CR39

In the front and back surface focus images (a,b) the out of focus exit/entry pits are visible. Illumination is provided via backlighting. There is a consistent displacement between the entry and exit etch-pits, and both are of a similar size. In the dark field image (c) a faint light-tube is most visible in the lower right etch-pit. Through lens illumination (d) is poor for illuminating objects through the foil.

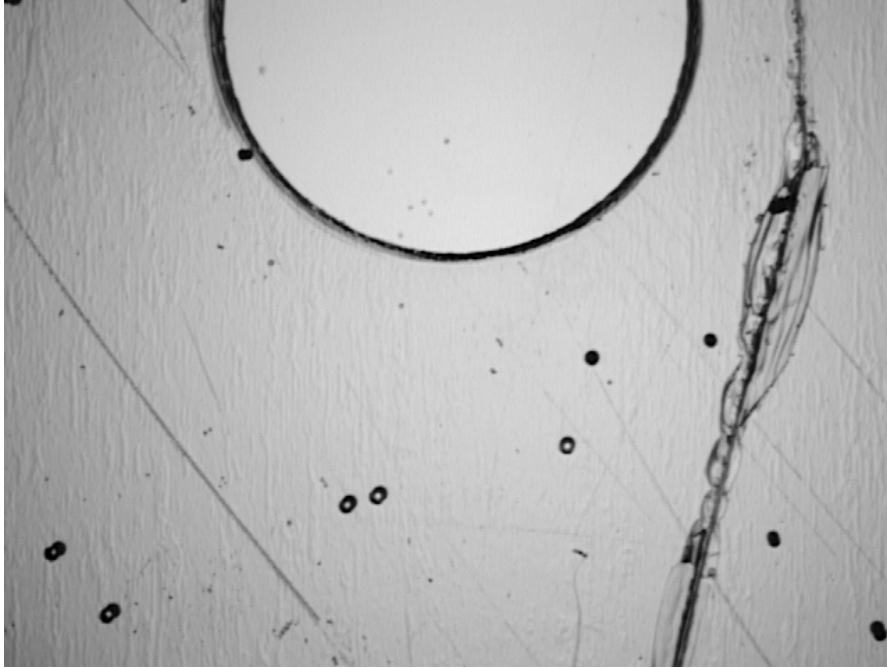


Figure 5.3: Pb ions in Makrofol

In this image we can identify ten Pb ion etch pits (not counting entry and exit separately). The entry and exit pits overlap. The foil is more susceptible to scratches (eg, the large straight lines of varying depth and damage) which can form a visual background. The large circle is a stack alignment pinhole (1.9mm compared to the $50\mu m$ etch pits).

Fluctuations in foil thickness (3.4.3) mean not all etch pits in an image can be captured in sharp focus by a single choice of focal plane. The 13GeV/A Xe etch pits are much less distinct than the 156GeV/A Pb etch pits. There is a strong visual symmetry between ranging in/out and through going pairs, and the signal etch pits aren't immediately discernable from the LHC background etch-pits (fig 5.5). In 2D bright field images focal depth cannot be relied on to distinguish between etch pits on the front and back surface of the foil (ie, separate ranging out events from ranging in). Lastly clustering issues become much more extreme in regions of higher background density. In the Xenon ion study there are no aligned and unexposed foils with which to reconstruct the Xe ion trajectories. Thus it is necessary to be able to manually identify and label Xe etch pits in a time efficient manner.

5.2.4 Makrofol Pb ion signal, full LHC background exposure

The final makrofol NTD foil formed the front foil of the NTD research stack exposed to the Pb ion beam and subsequently exposed for a full 2 year LHC run. This represents the highest density of spallation background etch pits that can be expected in MoEDAL NTDs. At this level of exposure the entire foil surface is covered in etch pits which form an almost continuous sea of background etch pits on both sides of the foil. Due to the larger ionisation the Pb etch pits are in many cases visible, however Pb fragments are all but indistinguishable against the background (fig 5.6).

5.3 Imaging methods

This section details the available imaging equipment, and the different imaging techniques that have been investigated in this study. Here the focus will be on providing a technical overview, while analysis and discussion will be reserved for the next section. It is commonly accepted that a simple 2D image doesn't provide enough information to make

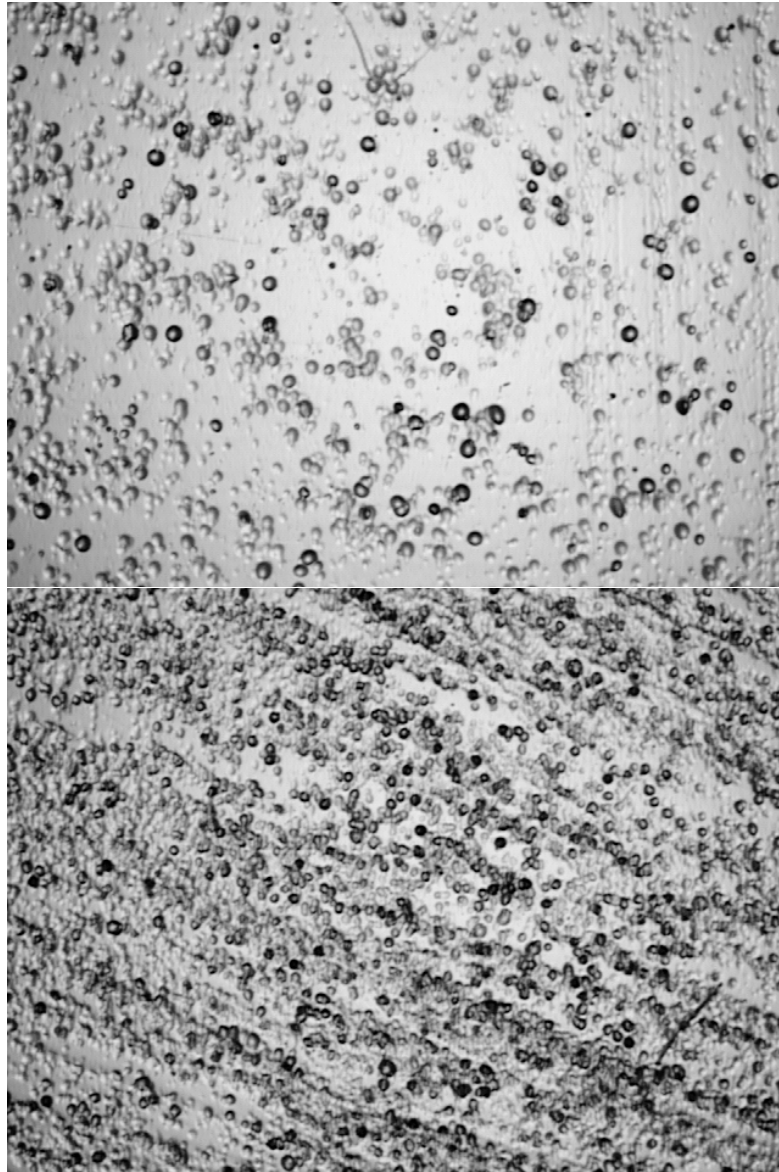
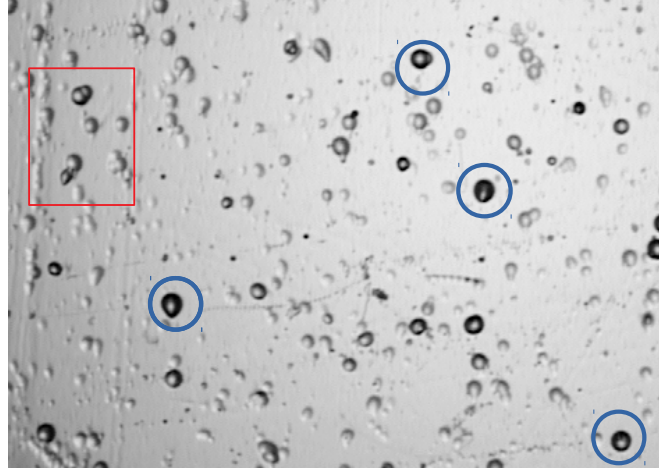
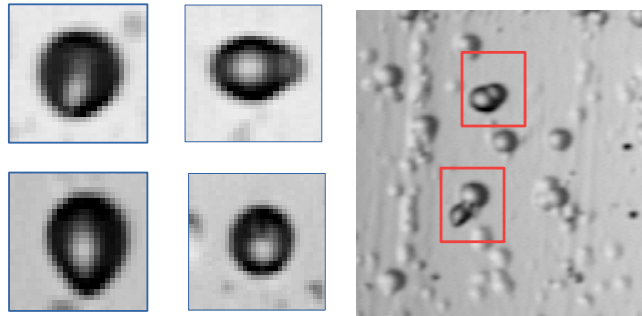


Figure 5.4: Typical foil image with 8 months of LHC exposure

The top image is typical for a region of the makrofol NTD foil after 8 months of LHC exposure. The background density is very high and clustering is frequent. From the image alone it is not possible to distinguish ranging in vs ranging out, let alone to identify Xe ion etch pits. The bottom image shows a region of higher density spallation radiation where there is almost no region of the foil without an etch pit. The impact of the background on overall image intensity is noticeable.



(a) Selection of etch pit objects in LHC exposed makrofol

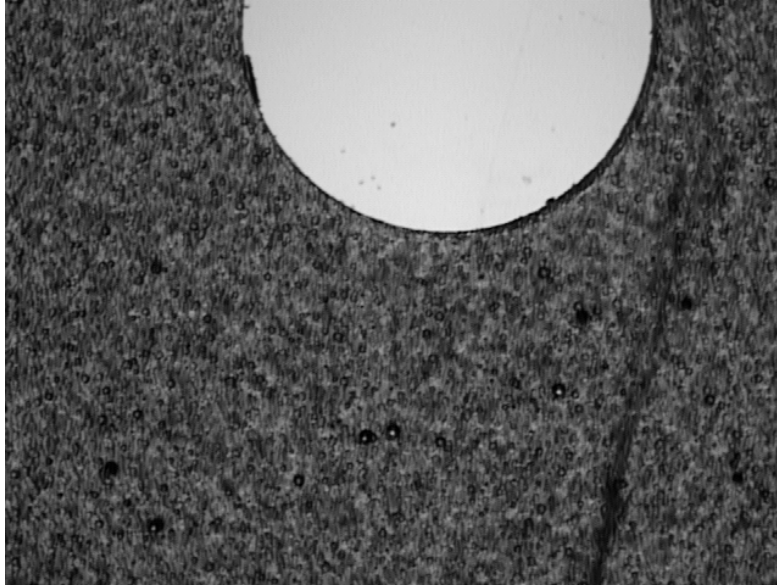


(b) Close up

Figure 5.5: Visual symmetry between etch pit objects

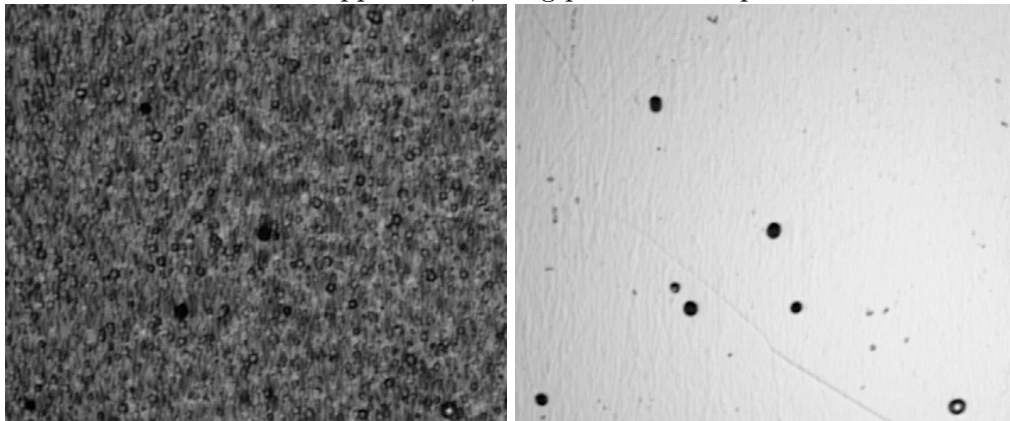
Blue circles highlight four large etch pit objects representing large energy depositions.

All objects look similar in the 2D image. Counter clockwise from the top left they represent; Two smaller etch pits that have merged, A large ranging out 'entry pit', A through going pair with a large entry pit and tiny exit pit, indicating a background hypothesis, and lastly a pair of equally sized etch pits with a light tube between them, indicating a true Xe ion etch pit. In the red square the bottom set of pits represents a through going pair while the top is simply clustered background.



(a) LHC exposed makrofol around the alignment pinhole

With the full LHC exposure every part of the surface contains an etch pit cluster. The highly ionising Pb ion etch pits are still visible, although several background etch pits mimic their appearance, being possible false positives.



(b) Aligned LHC exposed and unexposed images

We can notice that the smaller etch pits are barely visible in the LHC exposed image. (in the case of the smallest one it is possibly a fragment that had not ranged in when traversing the exposed foil)

Figure 5.6: Makrofol with full 2 year LHC background

a positive etch pit identification in LHC conditions [40] so an emphasis is placed on finding an imaging technique to supplement the image data.

5.3.1 Imaging Equipment

The NTD foils used in this research were scanned and imaged at Queen Mary's PPRC lab using an OGP flash 200 digital microscope. Its lowest optical zoom level corresponds to a digital image of 640x480 pixels, corresponding $5.6\mu\text{m}$ per pixel, or a real area of just over $3\text{mm} \times 2\text{mm}$. The relative difference between the highest and lowest magnification is x12. Unlike some of the scanning equipment available to the MoEDAL experiment it does not allow for continuous scanning, although it does allow for automated routines to be programmed. An aluminium jig is used for initial foil alignment. The glass scanning bed measures $200\text{mm} \times 200\text{mm}$ and may be illuminated in any combination of the following ways;

- Illuminated through the lens, with light reflecting off the foil forming a bright-field image with a central bias.
- Illuminated via backlight from underneath the microscope bed and through the transparent NTD foil. This forms an even bright-field image.
- An annular ring of LED lights, or halo surrounding the lens at an angle $> 45^\circ$ from the normal axis of the foil plane. Surfaces parallel to the plane of the foil cause this light to be reflected away from the imaging lens. This forms a dark field image and acts to enhance the appearance of edges.
- Via a controllable LED array. Light from the array passes through a Fresnel lens which focuses it onto the focal point of the lens. The LED array and Fresnel lens combination may be used to control the angular light distribution within a 45° cone of the imaging location.

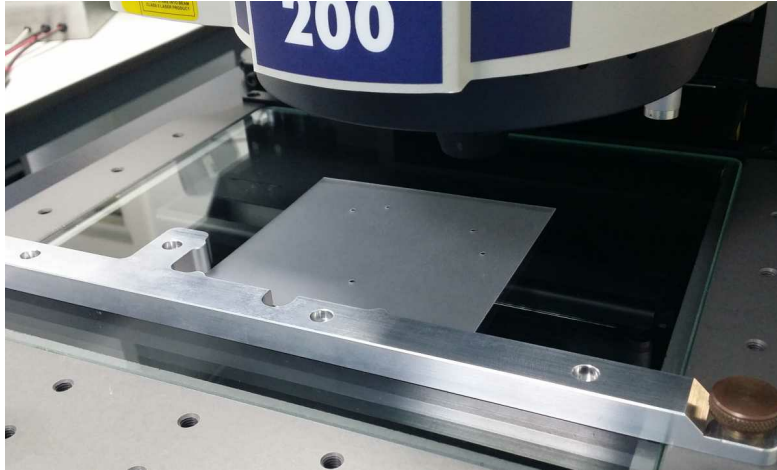


Figure 5.7: Microscope scanning bed along with NTD foil and alignment jig

5.3.2 Bright-field and Dark-field Microscopy

In Bright-field illumination, a large illumination field is directed normal to the sample being imaged. Light reaching the lens either passes through the sample (backlighting), or is reflected by the sample (through lens illumination). Dark areas correspond to where light is attenuated, either absorbed in the specimen or otherwise scattered away from the imaging lens, while most of the directly transmitted light is collected by the imaging lens (hence 'Bright' field of vision).

Dark-field Microscopy represents the reverse of bright field light scattering, the sample is illuminated so that directly transmitted light does not reach the lens (fig 5.8), resulting in a characteristic dark image field (hence 'Dark' field). Features that normally scatter light away from the lens are responsible for scattering it into the lens under Dark-Field-Illumination, while the drastically lower overall intensity means these features can be imaged with greater detail. Applications of Dark field Illumination include industrial quality control in translucent material, where fractures and imperfections in the material create areas where light can scatter into the lens.

NTD images were collected via all of the illumination methods with various levels of light intensity.

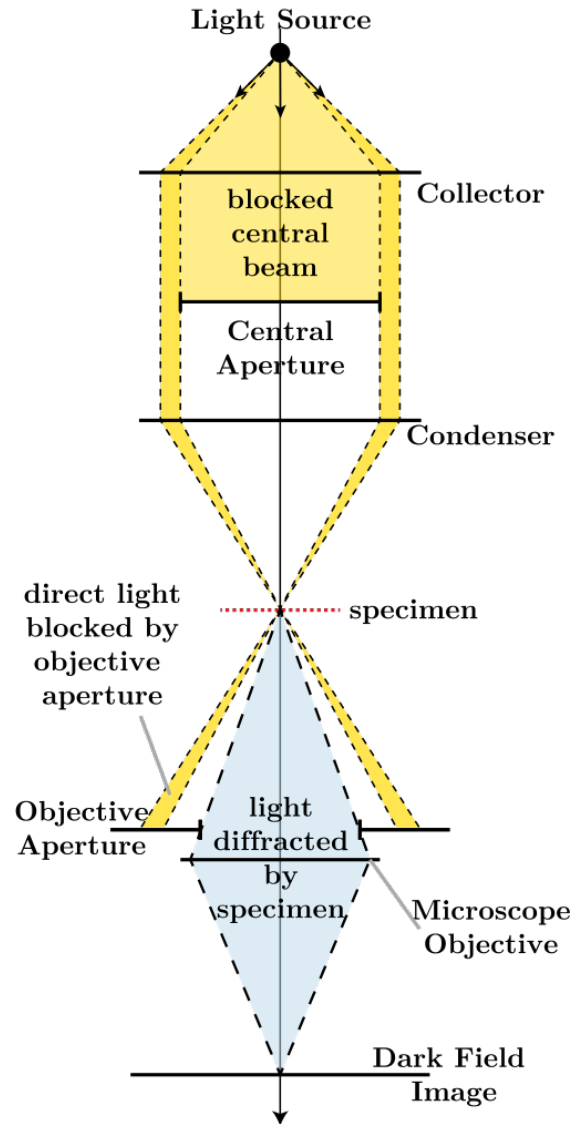


Figure 5.8: Optical diagram of a dark-field imaging setup [45]

Directly transmitted light is not captured by the imaging lens. Only light diffracted (or reflected in our NTD imaging microscope) by features in the specimen enters the lens.

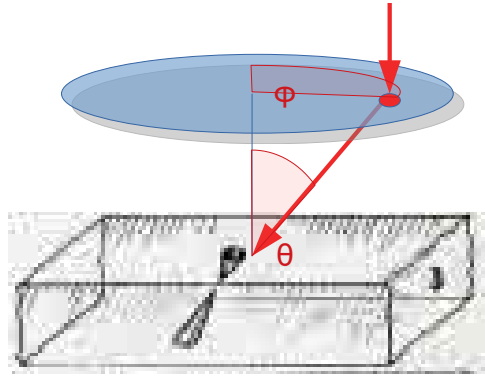


Figure 5.9: Light path via fresnel lens

Light from a given position (r, ϕ) in the LED array is refracted by the Fresnel lens onto the focal point of the lens. The light thus appears to come from an angle (θ, ϕ) .

5.3.3 Front and reverse side imaging

The front surface of the NTD foil, and the reverse surface may be imaged separately and compared afterwards to look for simultaneous occurrence of etch pits on both sides of the NTD foil. The limitation on this method is the precision to which the two images can be realigned afterwards. This method does not allow for immediate investigation of a candidate etch pit without physically inverting the foil and re identifying the area.

5.3.4 Focal depth adjustments

Another method for imaging the front and reverse surface of the NTD foil is to adjust the focal depth of the microscope in the z-axis to focus on either surface independently. There is no XY motion of either the lens or foil, so images taken from a given location are automatically aligned. A variation on this technique is to progressively image the foil at different focal depths. This is reliant on a relatively unexposed transparent NTD foil.

5.3.5 Rotational illumination

In this subsection I will describe an illumination method I refer to as rotational illumination. This uses the LED array and Fresnel lens described in §5.3.1. Different segments of

the LED array may be turned on or off independently, providing approximate control of the angular light intensity pattern falling on the etch pits, in terms of θ and ϕ (fig 5.9).

When illumination is restricted to the outermost LED segments, the specimen is evenly illuminated at a 45deg angle in θ , replicating the appearance of dark-field imaging, along with its edge enhancing effects. Introducing a strong directional component (by restricting illumination to certain angles in ϕ) indicates which edge of an object is being illuminated, and thus which side of the foil that object appears on.

Since the LEDs can be controlled electronically it is possible to rapidly image the foil from many different illumination angles without any mechanical motion that would disrupt the alignment of the focal plane and the foil. By illuminating the etch-pits from multiple directions it is possible to build up a reasonably confident hypothesis of the 3D structure.

I experimented with multiple illumination sequences for collecting multiple images at a location. The standard approach I have adopted is to capture a sequence of 8 images at a given location, each illuminated by one the outermost 45° segments of the LED array, incrementing (or rotating) the illuminated segment by 45° between images.

5.4 Results and Analysis

This section evaluates and comments on the imaging methods and options presented in the previous section. This is used to select a final foil scanning methodology described in the next section.

5.4.1 Bright-field imaging

The standard approach that I have used for imaging the NTD foils has been to collect bright-field images, illuminated via backlight, at the lowest level of optical zoom, in a focal plane midway through the NTD foil. This imaging method was the best way to

navigate the foil and repeatedly find a specific location or object (precisely because all background objects were visible). For this reason it is the primary baseline imaging method used for scanning in Chapters 6 and 7, and all other images collected have a matching backlit reference image taken at the same location. Backlighting was found to deliver a sharper image quality vs through lens illumination. Under these conditions all objects within the foil can be seen clearly without the need to adjust focus. The visual appearance of etch pits imaged this way is insensitive to variations in focal depth or light intensity. Overall this imaging method is simple and consistent, making it easy to compare between foils.

The disadvantage of this method is that, in the makrofol Xe foil, backlighting does nothing to aid etch pit identification. Even separating entry and exit pits is non-trivial. For the significantly more ionising Pb ions backlighting aids in identification in some cases where the etch pit has a white dot in the centre. This is due to the remaining foil material being very thin or non-existent where the cones of the Pb ion pits meet. This renders some Pb etch pits trivial to identify but cannot be relied upon as a general principle.

5.4.2 Dark-field imaging

I found dark field imaging effective for highlighting the edges of the etch pits giving a strong indication of the surface gradient. Although entry and exit pits still look similar, the visual symmetry between entry and exit pits is sometimes broken, with one side being brighter or in better focus than the other (fig 5.10). There are exceptions to this rule. It is possible for these variations to come about through focussing inconsistencies, which are unavoidable when imaging micro-scale objects. The dark-field illumination also exaggerates the appearance of scratches and smaller etch pits not of interest.

In regions with higher background density, with many clustered objects in the field of view, and visual imperfections in the foil, the images are noisy and hard to interpret.

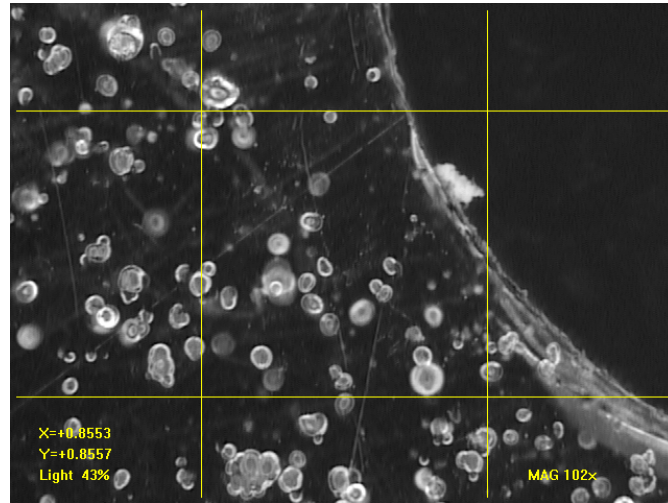


Figure 5.10: Dark field image of etch pits in the exposed Xe ion foil
The edges of etch pits are clear and highlighted. Ranging out etchs pits on the front surface appear brighter.

The 'Dark' field no longer appears as dark, nor exhibits consistently good contrast due to the large light leakage into the foil.

A method for suppressing the appearance of scratches, imperfections, and non-etch pit objects in the polymer is to add a small component of backlit illumination to saturate these objects. At high magnification the dark-field halo images are sensitive to perturbations in focus. Although they can be harder for a human to interpret, the dark-field halo images were found to deliver similar yet complimentary classification results to the backlit images in machine learning tasks.

5.4.3 Front and reverse surface comparisons

While objectively very simple in theory this method was difficult to apply in practice. During real time analysis it requires physically flipping the foil. Re-identifying the original area of investigation is tricky, particularly if there are few common features of interest. Comparing scans from two different surfaces requires context shifting between two different images, which can be difficult when images might contain a thousand or

more objects. Object ID still presents the same challenge as before in the less exposed foils where the same object may be visible on both sides. The foil alignment methodology introduced later in §7.2.2 allows for a pixel to pixel feature alignment of NTD foil scans. This makes the comparison more effective, particularly if the reverse images are digitally flipped so that objects visible on both sides have the same orientation. Matching between front and reverse side images was the primary method of etch pit quality control used for the unexposed Pb ion etch pits in chapter 7. This was done by automatically identifying Pb ion etch pits within either foil scan, then looking for coincidences within a tolerance. In LHC exposed foils such coincidences occur too frequently to be useful, however if the objects of interest are filtered to a lower density then this method becomes useful again.

5.4.4 Focal depth adjustment

In the unexposed CR39 foil, focal depth adjustments are the easiest way of imaging the entry and exit pits of the Xe ions. In the thinner unexposed Pb foils both entry and exit pits are essentially inseparable and each appears regardless of focal depth choices. In the LHC exposed NTD foils precision adjustment is required to focus on etch pits on one surface or another. In many cases this adjustment is insufficient to simultaneously image all of the etch pits within an image, let alone the entire foil. Object appearance (under various illuminations) can vary drastically as they go in and out of focus, and this is exacerbated at higher resolutions (fig 5.11). In the heavy clustered Pb foil foreground, objects obstruct clear viewing of the reverse side of the foil.

For candidate objects of interest, it is possible to observe some form of structure in partial focus the entire way through the foil, when moving the focal plane in the Z axis from entry to exit pit. This technique is useful for the inspection of single candidates, although it requires comparison to a sequence of images to distinguish from out of focus background objects. For scans of the entire foil an additional dimension of motion significantly increases overall scanning time, regardless of how the XYZ motion is

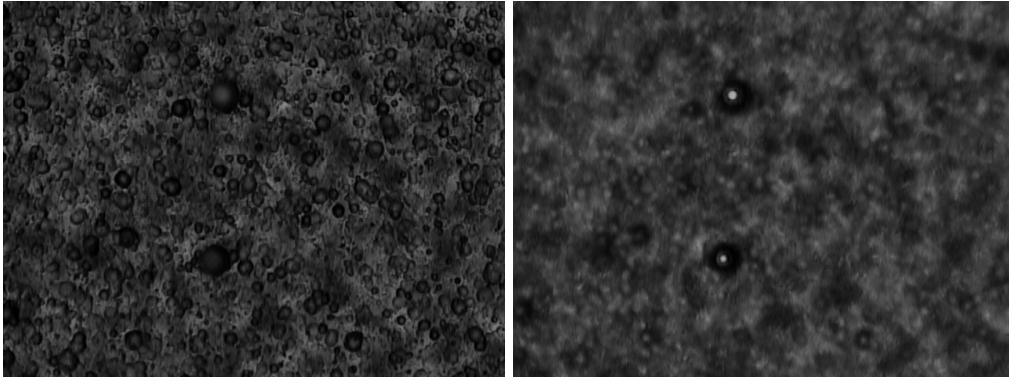


Figure 5.11: Pb exposed makrofol at two different focal depths
We can observe the visual appearance of objects in the image changes significantly with a small variation in the focal depth

decomposed.

5.4.5 Rotational illumination

Rotational illumination was found to be the most effective imaging method investigated. It enhanced the identifiability of all types of etch pit, including HIP candidates. Rather than a 2D image in (x, y) coordinates the rotational illumination forms a 3D image in (x, y, ϕ) . In a given (x, y) location this represents the scattered light intensity as a function of illumination angle ϕ , indirectly revealing information about the 3D surface structure scattering the light. This was found to be much less sensitive to variations in magnification, focal depth, or the overall light intensity of the image. Etch-pits can be identified on both sides of the NTD foil, removing the requirement to align two separate scans. Front surface and reverse surface etch pits can be distinguished from a single 2D image, simply by noting which side of the etch pit reflects the most light (eg, illumination coming in from the right side of an image will be reflected from the left side of 'bowl' shaped front surface pits, and the right side of the 'dome' formed by reverse surface etch pits). More complicated objects may be figured out by viewing more images in the sequence.

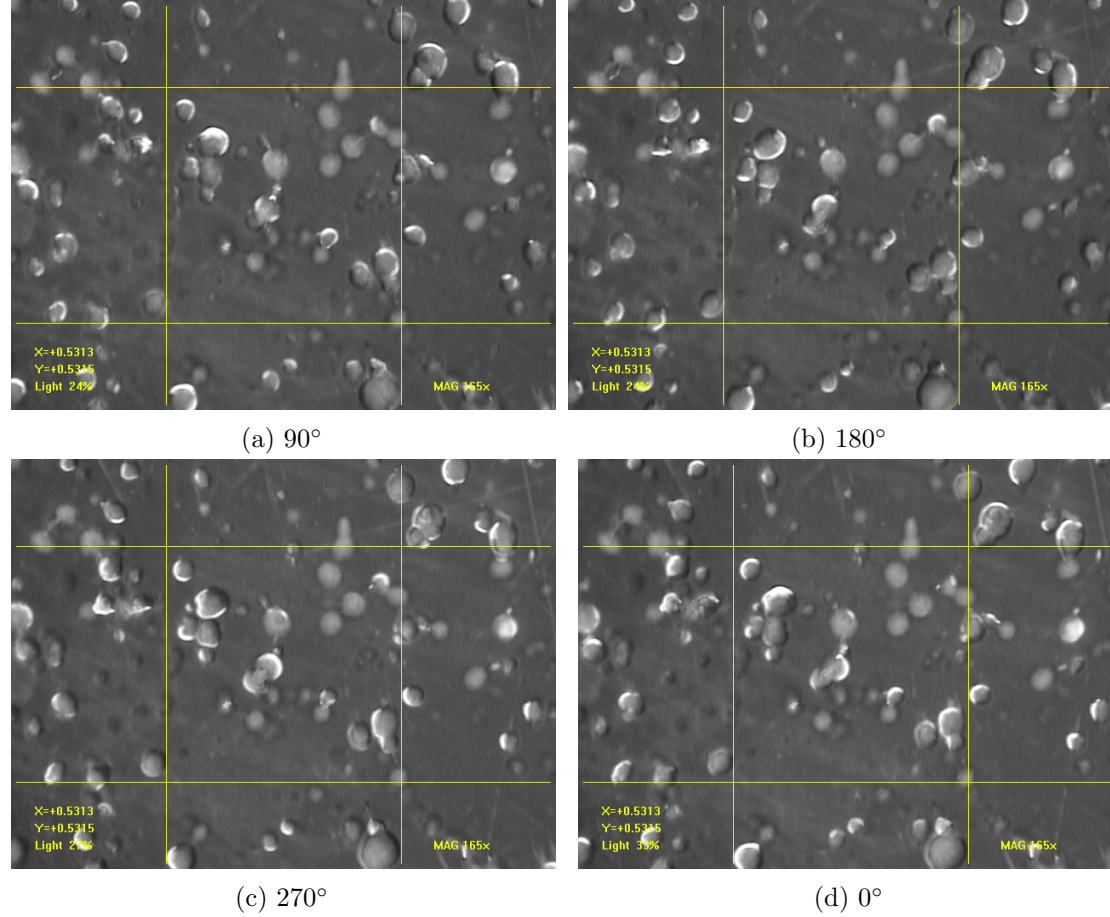


Figure 5.12: Rotational Bright/Dark field

The same Etch-pits are illuminated from 4 different angles. Of particular note is the central etch-pit. This appears as a two etch-pit structure. For various angles of illumination light is scattered from opposite surfaces, indicating each etch-pit is on the opposite side of the NTD foil. This implies this is a through going pair of etch pits. In image a) this object shows a disproportionately low light return. I hypothesis light is being lost down the light tube connecting the etch-pits and is not scattering into the lens. In contrast the cluster of etch-pits in the top left corner of the central image box scatter light from the same orientation. This indicates they occur on the same surface and are simply an overlapping cluster of background etch pits.

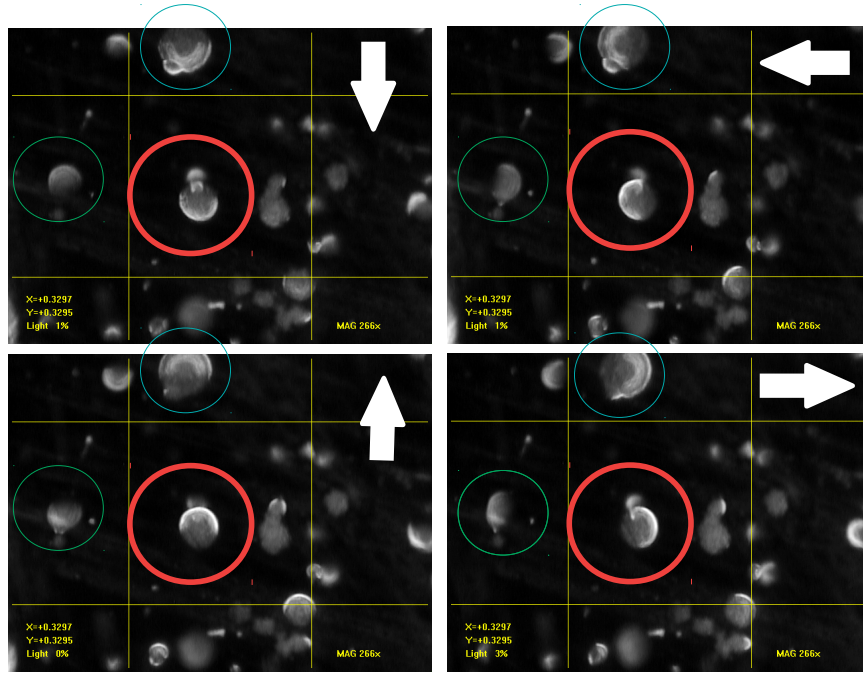


Figure 5.13: Rotational illumination of a HIP etch-pit at high magnification. Here we can observe a through going pair of etch pits (Red circle). The white arrow indicates the direction of illumination in a 180 degree arc. The edge of the pit is illuminated on opposite sides. Connecting structure is observable between the two pits (light tube). Although this indicates a highly ionising particle we can also observe that the exit etch pit is substantially smaller, indicating that the particle is starting to range out.

Varying the illumination angle in ϕ parallel to the plane of the NTD foil was found to add more useful information than varying it in θ , the angle of incidence perpendicular to the NTD foil. While it is possible to vary both this produces significantly more scan data. Images taken with the largest value of θ (just under 45°) were the easiest to interpret, likewise 45° incrementing segments in ϕ produced fewer ambiguities than 90° increments.

A benefit of restricting imaging to a fixed value in θ is that the images in ϕ form a closed sequence, possessing a continuous rotational symmetry (providing the index of the initial image is changed). Firstly this opens up multiple options for data augmentation §4.4.3, secondly it works very effectively with animation;

Rotation animation The 8 rotational illumination channels can be combined into a single looping animated image (effectively casting the ϕ dimension to time), allowing the three dimensional data can be analysed as a two dimensional moving image. In the animated images the illuminated edges appears to rotate about the etch-pits in a simple periodic rotational pattern. It is possible to visually scan the image with a wide field of view, taking advantage of our natural senses which are attuned to motion detection.

Etch pits on the same surface move in-phase with each other, while those on the opposite surface move in anti-phase. Objects that are not etch pits, and don't possess any significant 3D structure do not display any rotational pattern. Viewed this way complicated clustered background structures become simple co-rotating pits, with matched phase. While pits that have merged with each-other demonstrate rotation about a common etch-pit boundary.

Anomalies in this pattern are readily visible. Through-going etch pits appearing on both foil surfaces display a distinctive figure of 8 dipole pattern. Light tubes are particularly sensitive to illumination angle, especially when aligned with the tube. Depending on their geometry light will either appear to vanish down the tube, or show a spike in intensity as light is scattered out from the tube structure.

Scanning Rotational illumination provided the greatest assistance in identifying etch pits. Therefore it was the method used for data collection in chapters 6 and 7. Structures can be identified via their pattern in (x, y, ϕ) , this places less requirement on resolution in x and y . Thus images were collected at the lowest magnification, with a focal plane midway through the NTD foil to image objects on both sides, this also provides the greatest margin of error when imaging degraded LHC exposed foils. The patterns in (x, y, ϕ) result from the structure of the foil surface and are mostly invariant under illumination changes. A small element of backlit illumination (Accounting for 8% of main illumination via trial and error) was included to assist in saturating some of the non etch pit noise present in the images.

5.5 Final NTD scanning procedure

In order to collect data it was desirable to define a method for automatic data collection used for scanning the NTD foils. This procedure was used in chapter 6 and 7, with only a slight increase in illumination intensity to account for the higher etch pit background in the LHC exposed Pb ion foil.

- Data collection was automated by programming the scope to collect data via a routine. Image data was collected from a series of (x, y) coordinates arranged in a regular grid.
- The scope is set to a focal depth mid-way through the NTD foil. This minimises the optical impact of foil surface variations, and ensures image features have a consistent appearance throughout the foil scan.
- The scope is set to the lowest level of optical magnification. This allows a larger field of view imaging many etch pits simultaneously. Scanning time is limited by the number of (x, y) imaging locations N . Using lower magnification images allows for

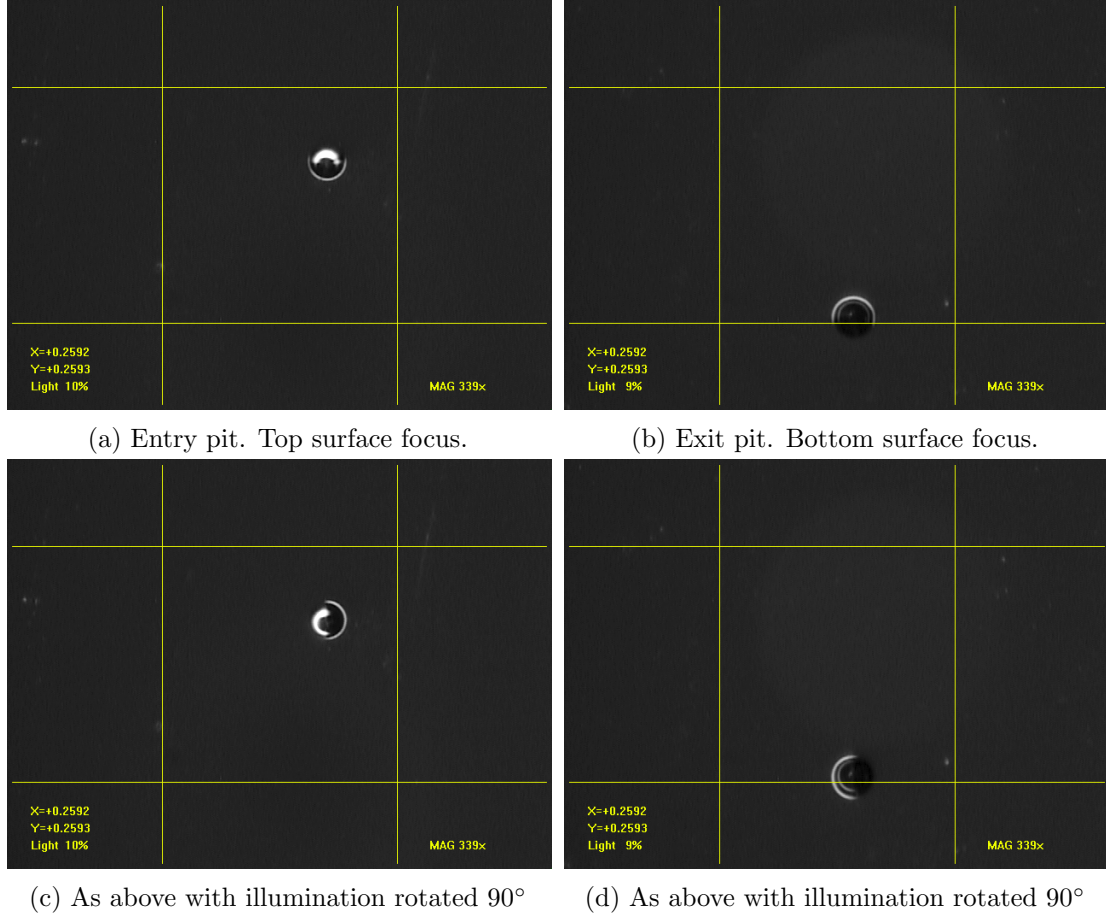


Figure 5.14: Rotational illumination validated in CR39

The left hand images represent the top surface etch pit, and the right hand side represents the bottom surface etch pit, relative to illumination coming from above.

Between the top row and bottom row the illumination angle has been rotated 90° . Illumination consists of a hemisphere of illumination covering 180° in ϕ and coming from approximately 45° in θ . The edge illumination appears on exactly opposite sides for the entry and exit pits respectively. The central light return is probably due to light being fully reflected from within the steeper etch pit cones.

faster scanning and offsets the additional data required by taking multiple images at each location.

- At each x,y location the scope rapidly cycles through different illumination conditions, collecting an image with each. This is controlled entirely through electronic switching with no motion. focus and focal depth is not adjusted for different illumination conditions. This minimises the time spent at each imaging point, and ensures different illumination channels can be compared directly.
- For the LHC exposed foils ten images are collected at each point. These include a backlit bright-field image, a dark-field image, and 8 rotational images. For the unexposed foils only the backlit field image is collected.

This imaging procedure is simple and robust. Minimal human involvement is required to initiate scanning. All that is required to image a different NTD foil is to replace the foil on the scanning bed. Collecting data across multiple illumination channels adds negligibly to the scanning time. Scanning the 100mm×100mm R&D foils takes 45 minutes per side, which due to the lower magnification is the fastest scanning time achievable with the available equipment. This would imply approximately 3000 hours to scan the first foil of every NTD stack in MoEDAL NTD array.

5.6 Summary and Conclusions

This chapter has looked into the imaging requirements for machine learning based NTD image analysis. In foils with very high levels of LHC background radiation exposure (ionising, and gamma) there are local variations in foil thickness produced by differences in the local bulk etching rate. This produces local variations in optimal imaging parameters. The density of LHC background etch pits presents challenging imaging conditions. The Xe ion and Pb fragment etch pits are almost indistinguishable from these background

etch pits, especially in standard backlit images.

The purpose of this study was to find an imaging method that aids the identification of HIP candidate objects. Based on the observations of this study rotational illumination was used as the main method of data collection, supplemented by backlit and dark field images. This method is primarily reliant on light scattering in the x,y image field as a function of the angle of incidence of the illumination ϕ . This allows all motion to be restricted to an (x,y) plane parallel to the image bed, allowing the focal plane of the optical set up to image many etch-pits simultaneously. This made it particularly suitable for bulk data collection. It also has the benefit of allowing for rapid analysis when combined with animation.

Chapter 6

Xenon Ion Study

This chapter will focus on the methodology and results of a research and development study into using CNN based image recognition to identify anomalously ionising particle etch-pits produced by a xenon ion test beam in a sample of makrofol NTD foil. The NTD foil was subsequently exposed at IP8 as part of a MoEDAL NTD array, acquiring a dose of LHC background ion tracks, and etched at INF Bologna. The NTD foil was only exposed for 8 months, compared to a typical 2 year MoEDAL exposure, resulting in a background etch-pit density only a third as high as might be expected.

The Xe ion used feature a lower $\frac{Z}{\beta}$ than the Pb (lead) ions used in the second Lead ion study. This makes the Xe ion etch pits harder to distinguish from common LHC ionising background particle etch pits. I will introduce the data gathering and pre-processing that went into this study, in particular a convolutional kernel based pre-selection step. Initial investigation focused on a naive signal vs background classification approach. Performance using this approach was found to plateau (after optimisation). Much better performance could be obtained in classifying background objects using the same architecture. An ensemble method was adopted that improved overall signal classification accuracy.

6.1 Introduction

The Xe ion foil data consisted of 480*640 pixel images covering 8 channels corresponding to 8 different illumination angles in ϕ , along with backlit and darkfield images. These were taken at lowest optical resolution of the scope, at 500 (5 in the preliminary) x,y spatial positions.

6.2 Preliminary Study

Prior to the main study described in this chapter a brief preliminary classification study was undertaken. CNNs were trained to identify 'Candidate' etch pits corresponding the the HIP calibration beam, from a generic background including other pits, empty areas of foil, and non-pit entities. The quantitative results were inconclusive so the description will be brief, however the qualitative observations motivated the design choices of the main Xe ion study.

6.2.1 Methodology

The large images were split into smaller 'patches' or bins. This technique was used in the previous MoEDAL ML image analysis using CR39 foil images. Each patch is 28x28 pixels (technically 28x28x8 due to having 8 frames / channels in each image). This closely matches the typical scale of etch-pits and their surrounding context features without having to be centred in the image. 28x28 also matches the image size of the MNIST dataset, a dataset of 10000 black and white handwriting samples of the numerical digits 0-9, commonly used as a reference dataset in image recognition tasks.

A supervised learning approach was adopted, and each cropped image sorted into one of three categories, as shown in fig 6.1;

- Images deemed to contain an etch-pit of interest consistent with a Xe ion track, exhibiting HIP properties that would warrant further investigation in this foil layer

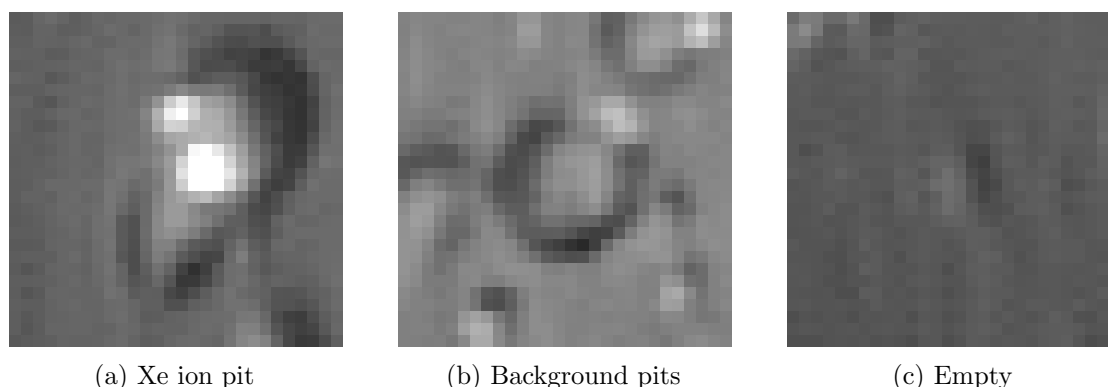


Figure 6.1: Image categories, shown for the first rotational illumination channel. In image (a) light is scattered from opposite sides of the entry and exit pits. There are multiple etch pits in (b) but none likely to be Xe ions.

or the next.

- Images with etch-pits, but no sign of HIP candidate.
- 'uninteresting' areas of foil containing no etch-pits or objects of interest.

The images were sorted according to their compressed file size which is approximately related to their information density. This enabled the dataset to be split into half. One half contained the images with few objects of interest. eg, empty space, single pits. The other half contained images with Xe pits, multiple pits and clustered background. The interesting images were then manually labelled, producing 33 Xenon ion examples vs 448 background etch pit examples.

Convolutional Neural networks with two, three, and four convolution layers were investigated. The networks used all compress the image representation down to 64 features. These are then used in a final dense layer to form the final classification. The Inputs to this study are three-dimensional greyscale images formed by the rotational illumination data. These were treated as 2D multi-channel images with each rotational illumination angle effectively being a 'colour' with a separate convolution performed in each.

6.2.2 Observations

The training runs were unstable, with a large run-run variability, and no smooth convergence. Of the 3000 cropped images there were only 50 signal examples. With such small statistics there aren't enough training examples to learn a consistent underlying rule. The sharp imbalance between signal and background also introduces larger batch to batch deviation in the metrics used to assess accuracy, and thus adjust the network, during training.

When the networks did train they exhibited the phenomena known as over-training. This is where the training results don't generalise when tested on a new unseen data. This happens when the network has enough free parameters that it can learn to recognise explicit examples in the training set, and the 'correct' label. Here the number of free parameters available to the network greatly exceeds the number of signal examples. Heuristically an accepted rule for image classification is at least 500 examples for each class. An attempt was made to use data augmentation to compensate for the low number of signal examples. A 90 deg rotation was applied to amplify the dataset by a factor 4, but did not yield a significant improvement. Nor did applying higher weighting to the signal examples.

The preliminary investigation demonstrated the need for a higher quantity of labelled signal data than could be provided by the patch and label method. The small signal ($N < 30$) statistics produced very unstable fluctuations in training and accuracy, nor can accuracy be reliably measured without a sufficient test dataset. I hypothesise that data quality played a role. Cropping may have removed key semantic features of some signal and background examples at the boundary of patches. Ideally each patch would be centred on an etch pit. With so few signal examples it would also be prudent to maximise the training value of each image by explicitly pre-processing the images, versus allowing the pre-processing to be learnt and performed by the network. Evidence of this is the acceptable performance of similar CNNs used with similarly low quantities of data eg,

fig. ?? trained in Chapter 7.

Compared to the multi-class etch pit labelling used later in this chapter the patch method itself is rapid, especially with binary categorising into patches with or without signal. The problems originated with the small patch size used creating a low frequency of signal patches. Significantly larger patches with many etch-pits would allow a more rapid spatial coverage of the foil, and improve the statistical balance of signal to background image examples. This would be a good option for future MoEDAL ML research.

6.3 Xe Dataset preparation

Based on the preliminary study I looked into ways of improving the quantity and quality of the Xe CNN dataset. A larger dataset of Xe foil etch pits was created in three distinct stages;

- **Pre-processing** applied to the overall 640x480(x8) images to denoise them and improve contrast and clarity, in addition to removing illumination based systematic factors.
- **Pre-selection** (automatic) to identify etch-pit locations (x,y) of interest within the large field of view images, and created a set of n 28x28(x8) reduced/cropped images of each point of interest, centred upon the relevant etch-pit.
- **Labelling** (manual) of the preselected of 28x28(x8) etch-pit examples into one of four track event object categories.

These stages are described in more detail below;

6.3.1 Image Pre-processing

Typical pre-processing in machine learning is to either 'Normalise' (Eqn. 6.1) or 'Standardise' (Eqn. 6.2) the input data x into a $[0, 1]$ or $[-1, 1]$ range, that better matches the

scale of the activation functions used in the neural network. This avoids saturating their response with very high or low values, normalisation can also be to factor out certain systematic biases affecting the dataset as a whole.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (6.1)$$

$$x' = \frac{x - \mu}{\sigma} \quad (6.2)$$

In the standardisation approach each input data example is normalised to the mean data example of the overall dataset μ and scaled relative to the standard deviation of the data set. When working with image data this applies to each pixel value, and amounts to subtracting the average image of the dataset.

I used a variation on the usual standardisation scheme in this analysis. Each rotational image channel is normalised from $[0,256]$ to a $[0,1]$ pixel intensity scheme. The directional illumination means we can calculate the average image for the normalisation over the 8 illumination channels. This is a local average of all the rotational illumination channels (effectively removing the directionality of the illumination). Subtracting this local average from each image channel results in a $[-1,1]$ normalisation where the featureless background is defined as zero for all images. Pixel values after normalisation represent the relative deviation from the local average, see Eqn. 6.3.

$$x'_i = x_i - \frac{1}{8} \sum_j x_j \quad (6.3)$$

Here x_i represents the i 'th channel of the $640 \times 480 \times 8$ image x

This yields several advantages.

- It suppresses 'static' features that appear in all image channels, reducing non-etch-

pit background, eg, post etching scratches on the polymer, thickness fluctuations due to local variations in bulk etch rate, and shallow etch pits indicating minimal ionisation are all suppressed resulting in a less-visually cluttered image.

- Image whitening is effectively performed for free. Images are scaled to their own local intensity average, eliminating image to image variation or bias. eg, intensity fluctuations due to the number of etch-pits in the field of view that alter the visual appearance of etch pits.
- Systematic bias within an image due to the central illumination of the image is also removed. While barely perceptible on an image by image basis it is a consistent bias over the entire image dataset.
- Lastly the key patterns produced by etch pit features become relative excesses or deficits of light intensity, with less dependency on specific values.

6.3.2 Pre-selection

In the preliminary study interesting HIP etch-pit pairs (signal) were heavily outnumbered by regions of background. This required a large number of images to be manually labelled to identify a small number of Xe ion signal examples.

In order to accelerate the labelling process an automated pre-selection step was used to find likely HIP candidate locations. The main element is a convolutional Kernel filter matching the general pattern of an entry and exit etch pit pair within the same area. The exact algorithm consists of the following;

- Convolutional filter in 3 dimensions, designed to look for the rotational pattern of front and reverse surface etch-pits. (Effectively two spirals in the 3D feature space 8 pixels across)

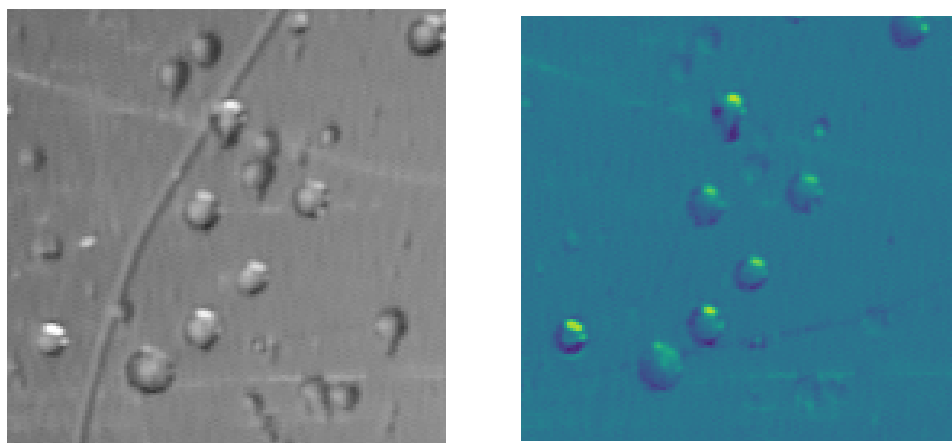


Figure 6.2: Before and after normalisation.

On the left is an example of the original greyscale image. On the right is a heatmap image. The blue background corresponds to zero. Yellow represents positive values, and dark blue represents negative values. The large scratch on the polymer no longer appears in this image, nor do many small features. The effective etch-pit density has been reduced to approximately 7.

- Image thresholding is applied, requiring a local normalised intensity excess of 15% in the normalised data channels. To kill off small details below the scale of an etch pit.
- An edge detection filter, requiring a 15% excess in the difference between the Dark-field and Bright field image channels. This works effectively because the dark-field and bright field images are approximate inverses of each-other.
- Binary Opening and Closing. Also known as erosion and dilation operations are used on all filters, including the final filter to remove small scale structure, and to decompose tightly overlapping objects [50].

Applying these filters in sequence to the image produces a binary image segmentation mask, the operations are performed on the image array data directly via numpy. This mask can be de-clustered, segmented into discrete object regions, and the centre of each region calculated to obtain a pit coordinate. Cropped images are taken centred on each of these points of interest and labelled to form the final dataset. These filters have the desired effect of ignoring most simple etch pits corresponding to ranging in and ranging out.

The advantage of this method is that the image segmentation map used to identify points of interest could be replaced by any other form of image segmentation map, such as output of a fully convolutional neural network which produces a pixel by pixel classification. This approach could be used to allow one CNN to bootstrap the creation and training of an improved successor. The goal of this work was to accelerate the creation of a training dataset for a CNN based HIP classifier. In this context there is far less constraint on signal acceptance and background rejection than required for analysis. A problem with the pre-selection is that it breaks down at very high background densities. This occurs when entry and exit etch-pit densities become high enough that coincidences are very common, and background acceptance becomes too high. This is particularly

Figure 6.3: An example of the output of the pre-selection step. A set of x,y coordinates corresponding to etch-pits of potential interest

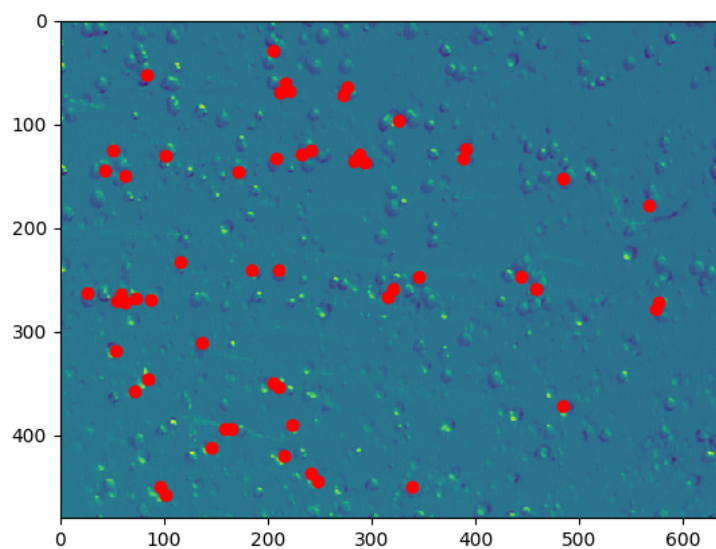
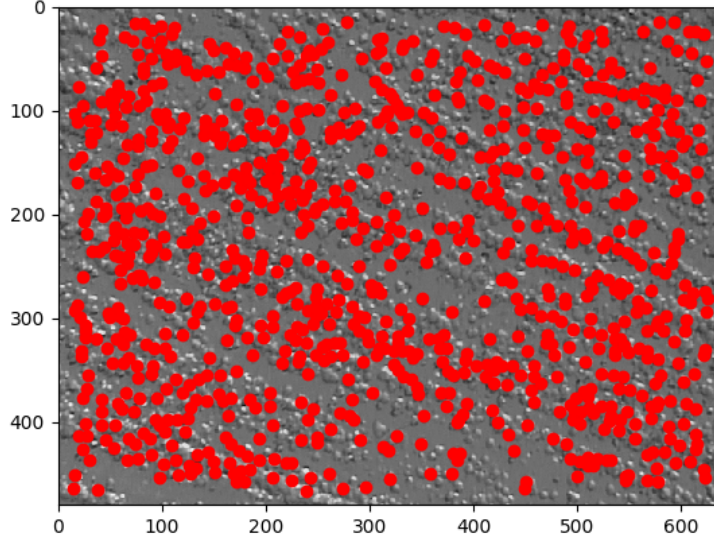


Figure 6.4: An example of the breakdown in the preselection for very high etch-pit densities.

With this many etch pits the proportion of background etch pits matching the basic entry exit pattern by coincidence becomes a significant fraction of all etch pits.



exacerbated by the weak form of de-clustering used. To account for this I excluded images where the number of preselected etch pits exceeds a threshold (30). A better long term option would be to parametrise the choice of filter preselection algorithm by background density.

6.3.3 Labelling

The cropped 28x28(x8) etch-pit examples identified in the pre-selection step were manually labelled into one of four categories to create a labelled dataset for supervised learning;

- Candidate (Signal) Etch-Pits; Which I judged to be an overlapping pair of entry and exit pits, going through the foil. For a detailed analysis we would confirm the particle trajectory implied by the geometry of each etch pit, and Restricted Energy

Loss matched. However for preliminary training this is not necessary, and may be an over-complication.

- Entry and exit pits; Corresponding to ranging in (RI) and ranging out (RO) pits occurring on the bottom surface and top surface of the foil as viewed through the microscope. As they already caused a hit on the preselection they typically correspond with large ionisations and possibly with a small asymmetric partner. Since they can be distinguished easily via rotational illumination they are labelled separately.
- Miscellaneous; Junk objects that passed the initial preselection filter. Pits much smaller than we would expect from a Heavily Ionising Particle, Surface imperfections or damage on the foil not matching a pit like structure. Pits or structure defying simple or quick explanation. Due to the pre-selection step this represented a trivially small number of examples that were filtered out of the dataset.

Additional sub-classification of examples was not considered. This is because subdividing the classes would weaken the statistical power of the learner, by reducing the effective number of training examples for each class. Due to the density of pits on the foil there are multiple images where other etch pits are visible in the same image. Examples are categorised by the largest, most central etch-pit. This means the training images include realistic non-linear background. No attempt was made at this stage to create regression labels, such as etch- pit size (corresponding to REL), or to measure pit clustering (eg N pits in each image), or etch-pit separation (for candidates).

To validate my ability to correctly identify Heavy ion beam etch-pits I visited the main NTD analysis facility at INF-Bologna. After sorting the data classes I double checked to identify any 'odd ones out' within a particular class of object.

6.4 Optimisation methodology

Hyper-parameters The nominal hyper-parameters were settled upon through a combination of systematic investigation, and trial and error [3][33].

- Convolution and MaxPooling [20] layer, a 3x3 kernel was used in the convolutional layer, mapping to 32 features, followed by a 2x2 MaxPooling layer.
- Epochs, represent the number of complete iterations over the training dataset. For most purposes a value of 75 was used, this appears to be sufficient for most networks to train to within a few % of maximum stable accuracy.
- The Adam optimiser [31] was used, with a learning rate of 0.0002, optimising on Binary Cross-entropy, with a batch size of 50. The lower learn rate helps guarantee a steadier learning process.

Network Depth

Networks were evaluated with two, three, four, and five convolution layers paired with a max-pooling step. Four layers were found to yield optimum results, yielding on average a 5% improvement, over a three level network, a single layer network did not yield any result better than guessing. This may be expected since a regular convolution filter was already used in the pre-selection stage, thus a single layer perception would not yield any improvement over this pre-selection step. There was a slight proportional increase in training time for each conv-layer, as the initial conv and final fully connected layers are much smaller. Five layers represents the maximum practical depth due to the spatial compression of the max-pool layer. The small initial resolution (28x28) of the pits is completely covered by five x2 MaxPool operations, and with five layers we see reduced network performance. This should not be surprising since five layers exceeds the feature size of an etch pit and includes redundant information in the image.

Convolution kernel Width

The kernel width represents the spatial area of pixel correlation considered in each convolution step. A wider kernel width looks at pixel correlations over a wider area, but at the expense of more parameters in the network. This represents an overtraining risk unless this additional information improves classification ability. A 2x2 kernel did not lead to efficient training, likely as there is too little information contained within for the network to learn from. Kernels larger than 3x3, eg, 5x5 can cover a large section of an etch-pit image sample. Using larger kernels made it too easy for the network to memorise particular images in the dataset, leading to a rapid overtraining situation.

Optimiser and Loss

The optimiser refers to the algorithm used to update the network weights to minimise or maximise the loss function. Different common optimisers were tried; 'Adam', 'adamax', 'adadelata', and SGD (stochastic gradient descent). SGD did not give steady learning performance, adamax and adadelata could be used, but Adam performed consistently better in terms of speed of convergence. Since the choice of optimiser only impacts the training stage it was decided to stick with Adam [31]. The loss function is the function we try to minimise (or maximise). In classification tasks it is chosen to represent or reflect the misclassification error. Typically the mathematical distance between the classification scores and their true classification labels for the training sample. Different standard loss functions were tried; Binary accuracy, Categorical accuracy, Mean squared error, and Kullback-Leibler divergence. Binary and categorical accuracy performed best, whilst mean squared error sometimes led to training problems.

The objective of the second study was to produce a preliminary HIP candidate tagger capable of identifying Heavy Ionising particle etch-pits in the presence of dense heavy ion backgrounds.

Layer	Feature depth	Activation
Conv (3x3)	32	Tanh
MaxPool (2x2)		
Conv (3x3)	32	Tanh
MaxPool (2x2)		
Conv (3x3)	32	Tanh
Conv (3x3)		
MaxPool (2x2)	32	Tanh
Global MaxPool (2D)		
Dense	22	ReLU
Dense	2	Softmax

Table 6.1: Standard CNN layer architecture used for Xe study.

6.4.1 Initial Results

The data samples used consist of 800 bottom surface etch pits, 196 top surface etch pits, and 259 HIP candidate etch pits. A sub-sample of background pits was taken, so that the number of background samples matched the number of signal samples. 30% of the Data was reserved as a testing set. Initial experimentation struggled to improve Xe vs background classification network performance beyond 70-80%. While over-training was assumed to be the cause, further study showed the initial diagnosis of over-training to be false. The variation comes from difficulties in distinguishing signal from some background classes. While the network can achieve good accuracy and results in distinguishing any two classes (especially when evenly balanced) it struggles when distinct sub-classes are present. Eg, if top pits are included as background, it will correctly identify signal events, and bottom events, but will confuse top events for signal (evidenced by $< 50\%$ confusion rate) or resort to a random guessing strategy. This Min-Max strategy is viable because top pits constitute a small fraction of overall pits, however it limits the overall performance of the network.

The same behaviour can be observed if other labels are substituted. Eg, A network trained to distinguish Top pits and Signal pits, will struggle in identifying Bottom surface pits. This suggests a key feature the network searches for is the presence of structures

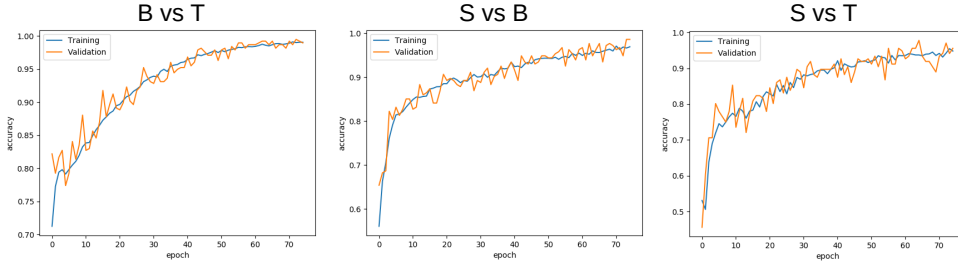


Figure 6.5: Network training vs Epoch

When trained on specific binary classifications the networks performed more accurately, converging on a 95% accuracy. When trained in this fashion no overtraining was observed.

on the opposite side of the foil to its primary background class. This is sensitive to the proportion of different pit categories included in the training data.

6.4.2 Binary Classification

In contrast to training against a physically representative background sample, training in simple A vs B classification tasks with pure samples of A and B respectively were able to train and converge easily to an approximate $\approx 95\%$ accuracy with no overtraining divergence (Fig.6.5). Classifiers were trained in Xe vs RI discrimination (S/B), Xe vs RO (S/T), and RI vs RO (B/T). The problem is that pure backgrounds aren't physically realistic. Foils contain a variety of different etch-pit phenomena at once, especially at high levels of LHC exposure. Classifiers trained only on a single background type have an unacceptable misclassification rate when performing classification on other background types.

6.5 Ensemble

The weakness of a single classifier or learner can be overcome using Ensemble (Committee) methods. Committee methods such as BDTs have often been used in particle physics

with many different approaches, eg BDTs to improve the performance and robustness of weak learners. The output of multiple classifiers (learners) are combined to produce an aggregate output that is more accurate and robust than its component parts.

Inspired by the good CNN performance in classifying between well defined backgrounds I sought to combine the classifiers resulting from each combination of Xe vs Range in and Range out. I adopted a geometric combination of classifier output scores, where the overall ensemble 'signalness' score is the weighted product of the 'signal' score of the signal vs ranging out classifier and the signal vs ranging in classifier. A similar score is calculated for each Ranging in, and Ranging out, and the scores are normalised. To achieve a high 'signalness' score requires a strong output from both signal classifiers. The final object prediction is taken as the highest scoring category.

$$P(S) = N \times P(S|S \cup RO) \times P(S|S \cup RI)$$

$$P(S) \equiv \text{Probability of being signal} \sim \text{signal score}$$

$$N \equiv \text{Normalisation such that probabilities sum to 1}$$

$$P(S|S \cup RO) \equiv \text{Probability of being signal given either signal or ranging out}$$

$$\sim \text{Signal score of classifier trained only on signal or ranging out events}$$

Via this method multi-class classification can be performed using a combination of two-class CNN classifiers trained on just two object categories. In this context this is beneficial as the two class CNNs delivered more accurate results, that were more consistent over training repetitions, than classifying against a background containing multiple etch-pit types. This ensemble approach can also be expanded to additional etch-pit object categories without requiring a full retraining of all the classifiers involved. The final predictions are also explainable, as the final signal score is proportional to a weighted sequence of classifications against every alternate background hypothesis. An

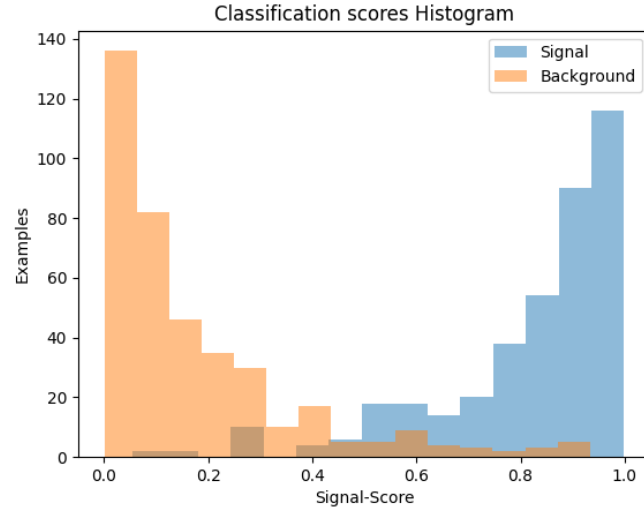
additional 200 Xe signal examples, and 100 RI, and RO examples were collected and used to validate the ensemble performance. The Signal and background separation is shown in Fig.6.6a, along with the ROC curve in Fig.6.6b. A working point of 0.5 results in a validation signal acceptance of 92.9% and a background rejection of 92.1%. This is a substantially improved result compared to a classification trained against all backgrounds. Using the highest class score as our class ID yields a slightly different result given in Fig.6.8, which is equivalent to a signal acceptance of 94.5%, with a higher false positive rate with 12.5% of preselected RIs and 8% of preselected ROs becoming false positives. In Fig. 6.7 we can see what is happening in practical terms, the more ambiguous signal examples overlap heavily with either the RI or RO pits but not both, making it easier to distinguish signal pits from background pits in the higher dimensional space provided by the additional RI vs RO classification.

6.5.1 Inference on NTD foils

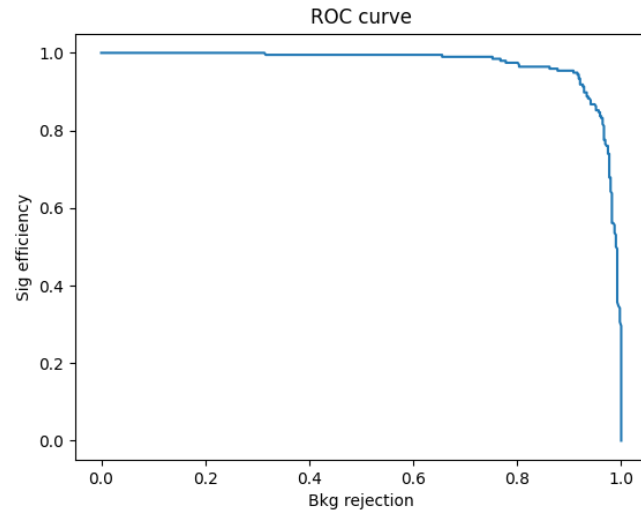
By running the preselection algorithm on regions of foil, and applying the geometric ensemble classifier to the selected locations, it is possible to generate predicted candidate locations in unseen foil images. This was applied across multiple NTD images of different etch pit densities. While this performs well for low to medium etch-pit densities the ensemble classification breaks down in regions of higher density, similar to the preselection algorithm, triggering excessive false positives throughout the image.

6.6 Conclusions

The adoption of a preselection technique for filtering data proved highly beneficial. Pre-selecting only objects of interest lowered the data-input of an NTD foil scan by an order of 1000, from 50gb for 10 image channels at lowest resolution to 50mb of selected objects. This drastically reduced resource requirements in terms of data storage and processing.



(a) Signal vs background for geometric ensemble
Xe vs All backgrounds separation histogram



(b) ROC curve for the geometric ensemble

Figure 6.6: Validation performance of geometric classifier ensemble
The signal vs background separation achieved, and the ROC curve obtained using the previously withheld validation data.

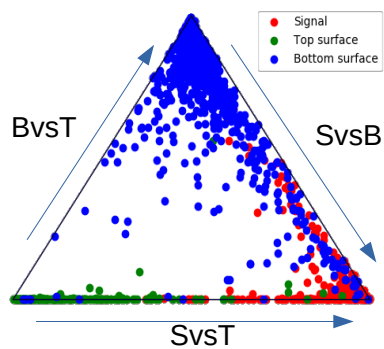


Figure 6.7: Ternary plot of the geometric ensemble scores

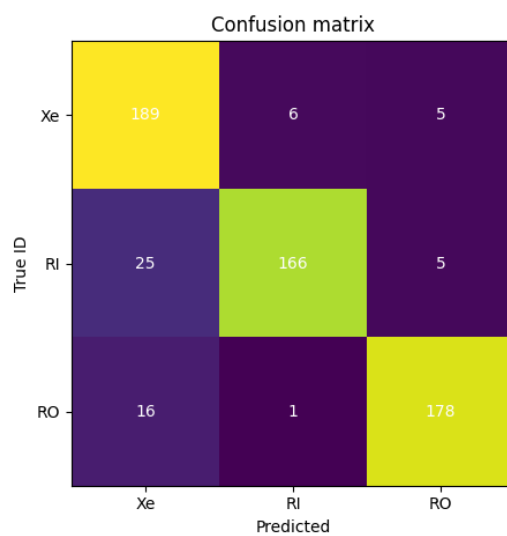


Figure 6.8: Confusion matrix, for geometric ensemble scores
 Adopting the highest likelihood as the predicted ID gives the following confusion breakdown.

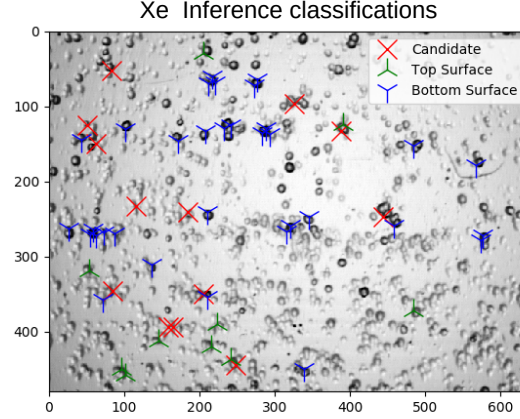


Figure 6.9: Three way inference classification performed on preselected etch pits.

The approximate data scaling of the MoEDAL experiment is given in Eq.6.4. Assuming a minimum $0.5\text{Tb}m^{-2}$ for a 2D image at lowest usable magnification. N_C is the number of data channels collected, whether via focal adjustments or different illumination channels. Etch % is the fraction of the $300m^2$ NTD array that is etched and imaged in the first pass of the analysis (approx 10%). This represents a minimum of 30Tb scaling with the magnification and number of image channels necessary to make an identification. Without this, and assuming a 10 pixel minimum etch pit resolution, a minimal 2D scan would take 0.5Tb per m^2 .

$$\text{Data} = 2 \times \left(\frac{0.5\text{Tb}}{m^2} \right) \times \text{Mag}^2 \times N_C \times 300m^2 \times \text{Etch \%} \quad (6.4)$$

Applying some form of preselection to the wider MoEDAL NTD analysis would shrink this substantially, allowing analysis of the preselected objects to be done on a laptop or desktop computer. As NTD scanning is non-destructive this simply represents a search prioritisation rather than an absolute loss of data.

The Geometric ensemble performed well on low density regions of foil, outperforming the sum of its parts, and giving a better result than a simple Xe vs background classifier. However its break down at higher etch pit densities means it is unsuitable for analysis of

fully exposed LHC foils. The main bottleneck to this research was labelling the selected etch pit images, hindered by the larger fraction of RI events, and the ambiguity of many etch pits. In the Xe foil there is no way of establishing a definite truth identity for each etch-pit. Instead human judgement must establish a best guess. Because the Xe ions represent a subtle signal, with their low energy of $13\text{GeV}/A$ closely resembling the LHC background, human labelling reflects a substantial statistical and systematic error in our truth that will impact both training and evaluation.

Chapter 7

Pb Ion Study

The principle of the Pb ion study is to obtain a quantitative evaluation of CNN based anomalous etch-pit detection in the context of a full 2 years of LHC background exposure. The previous Xe ion study was limited in the following ways. Firstly the foil was only exposed to the LHC for a period of 8 months vs the 24 month exposure of a typical MoEDAL NTD stack, meaning it failed to fully capture the difficulties encountered at higher LHC background levels. Secondly due to being a single NTD foil it was not possible to confirm with certainty the true identity of HIP candidate etch-pits beyond human inspection and judgement. This introduced a strong systematic human bias. The study outlined in this chapter aims to correct these flaws.

This chapter focuses on a special R&D test stack. In the introduction I will describe the NTD test stack, its design, exposure, and etching. A coordinate system was constructed to allow for post-scanning alignment of NTD images. A CNN classification study was performed for the Pb ions in higher background foil. An FCN algorithm was developed for the analysis of wide area images. The last section will focus on the quantitative evaluation.

7.1 Introduction: The Pb ion test beam stack

The aim of the Pb ion study was to apply machine learning based HIP etch pit detection to MoEDAL foils with the full 2 years LHC exposure. This has the effect of causing every area of the NTD foil to feature some form of etch-pit structure on the front and reverse surface. This represents a domain transition. Many etch pits cease to be discrete objects, and every portion of the surface fulfils the criteria of having an exit pit and entry pit, causing the preselection used in §6.3.2 to break down. This motivated a move from discrete classification, in particular replacing the fixed kernel for a machine learning based approach.

During the research with the Xenon ion foils a consistent bottleneck was the time consuming effort of validating and confirming the identity of the Xe HIP candidates. This was a barrier to gathering a large signal dataset for either training or validation purposes, and balancing the two meant a trade off between the accuracy that could be trained to, vs the precision to which it could be measured. The human labelling also introduces a strong unknown systematic bias, where an error on behalf of the human labeller is copied through to the algorithm and any measurement of its performance. The success of the ensemble method lay in utilising the plentiful background etch pits. In the dense background of the Pb ion foils the uncertainty and delay in identification would be larger.

7.1.1 Experimental ML stack

Experience with the Xe ion foils motivated the concept of an experimental R&D stack used as a heavy ion target (Pb), with some of the foils then exposed to the LHC, and others remaining unexposed (fig 7.1). The R&D stack fulfils two functions;

Firstly the locations of Pb etch pits in the unexposed foils are correlated with the location of HIP etch-pits in the LHC exposed foil. These known locations can be used as

training examples for supervised machine learning. Secondly we can test the performance of machine learning algorithms in finding HIP etch pits in the exposed foil by investigating the predicted locations in the unexposed foil (where we should find a down stream ion track). This allows us to make estimations of the entire stack sensitivity, signal efficiency rates and false positive rate. Through a subdivision of the 'horizontal' stack area (vs 'vertical' foil depth) it is possible to ensure the second evaluation process is blinded from the first training process. This gives us an unbiased measurement of the performance of our algorithms.

The work described in this chapter proceeded in four stages;

- First it was necessary to find a way to align the foils, in order to recreate the Pb ion trajectories.
- The second stage involved replicating the CNN classification study with the Pb etch pits.
- The third stage was adapting this into a fully convolutional algorithm suitable for analysing and finding candidates within wide area images.
- Lastly final performance evaluation proceeded using an entirely separate region of foil. Known etch pit locations were compared to predicted locations identified in the unexposed foil and used to obtain a false positive and false negative rate.

The remainder of this section will proceed with a more detailed discussion of the NTD stack.

R&D stack design and exposure The stack consists of five layers of makrofol, representing the initial makrofol layers of a complete Moedal NTD stack. Foils within the stack are held in place using alignment dowel pins passing through all the foils in the stack. The location of the pin holes allow us to recreate the position of foils in the stack.

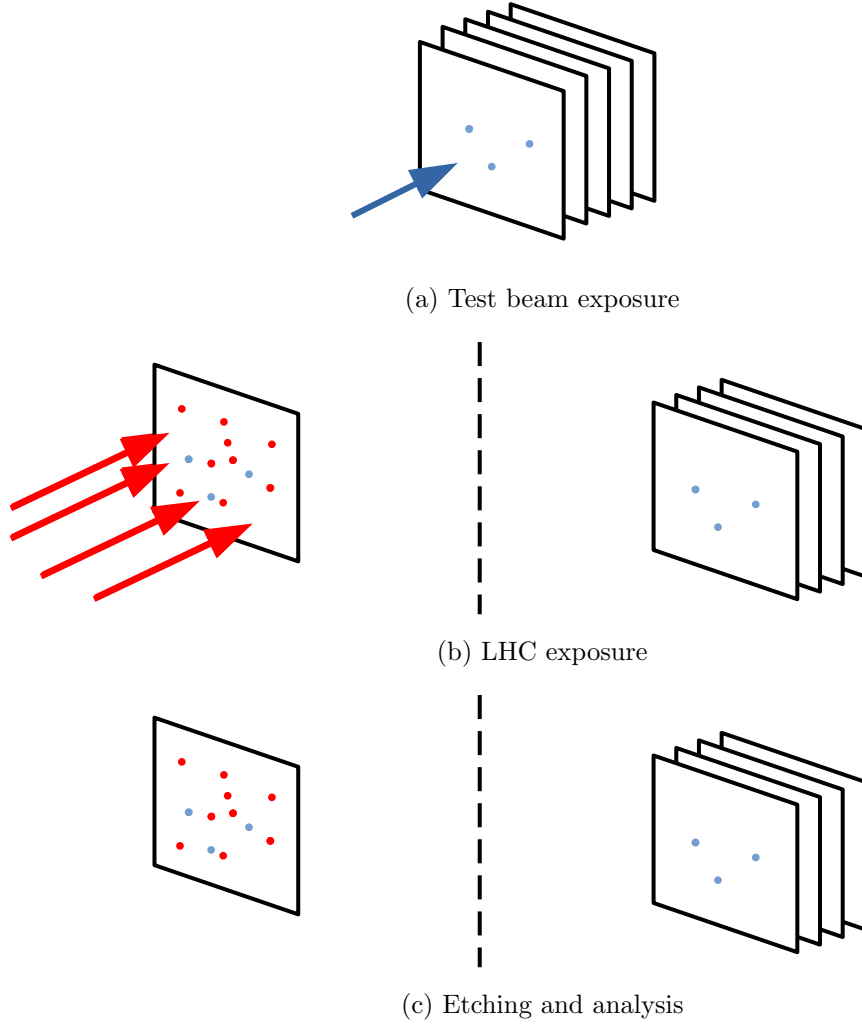


Figure 7.1: Research NTD stack exposure scheme

The entire aligned stack is exposed to the Pb ion test beam. Ions penetrate the stack leaving ion tracks in all foils. The stack is divided in two. The first foil is exposed to the LHC leaving additional spallation ion tracks. After etching ion tracks turn into etch pits. The pattern of Pb etch pits in the unexposed foils matches that in the exposed foil and can be used to identify the Pb ion tracks.

Foil	Pb	LHC
1	156GeV/A	2 years
2	156GeV/A	
3	156GeV/A	
4	156GeV/A	
5	156GeV/A	

Table 7.1: Foil Exposure

The assembled stack, including the aluminium front plate was exposed to a calibrated Pb ion beam of 156GeV/A. These lead ions act as a stand in for a particularly high energy HIP originating from the interaction point. Pb ions from the beam penetrate through the stack, leaving Heavily Ionising Particle (HIP) ion-tracks through each foil they pass through. When etched these tracks will produce corresponding HIP etch-pits at the entry and exit locations in each foil. Due to the momentum of the beam particles there will be minimal deflection in the NTD foil material allowing the particle to be considered as a single linear track traversing the stack, consistent with our predicted monopole signal.

After Ion beam exposure the stack was disassembled and the last 4 foils in the stack were removed. Due to the low etch-pit background these four foils should be sufficient to recreate the the Pb tracks. The first NTD foil along with the aluminium stack housing were then placed in the MoEDAL NTD array for two years during LHC run 2. Thus this foil has received a full run of LHC background radiation exposure as a normal moedal NTD stack. Indeed it is a normal NTD stack except for being smaller in cross-sectional area, and in having the last foils removed prior to LHC exposure.

7.2 Foil Scanning and Alignment method

An essential requirement of this study is the ability to compare potential or expected etch pit locations between the exposed and unexposed foils. This requires accurate alignment of the NTD foil scans. The method I have adopted is to define a system of global coordinates defined relative to the datum alignment pinholes drilled through the NTD stack. This allows for a more accurate reconstruction and tracking of candidate objects. In the following subsections I describe the scanning of the NTD foils, the definition of the global coordinate system and the procedure for converting into these coordinates, the last subsection covers the offset correction needed to accommodate for the angle of incidence of the ion beam.

Accounting for all these factors allows for a pixel to pixel comparison of etch-pit features between images taken from different foils or different scans of the same foil taken from the reverse side. Not only does this allow for accurate identification of objects in the exposed foil for training and evaluation purposes but it also allows for the direct training of U-Net style algorithms using aligned images.

7.2.1 Foil scanning

Etched foils within the stack were scanned and imaged using the following procedure, based on the method described in §5.5 also used in the Xe ion study;

- The foil to be imaged is placed on the microscope bed, with an aluminium jig used for initial approximate alignment.
- The NTD foils all have a set of three alignment pinholes that were drilled through them whilst part of a complete aligned stack. Labelling them A,B,C image scans were taken to approximately cover the rectangular area between A,B,C. These alignment pinholes are taken as datum points for a left handed Cartesian coordinate system. An on-screen visual target is used to identify the centre of the top left

datum alignment pinhole A (top right when scanning the reverse side of the foil). Taking A as the origin, the line AB forms the approximate x axis, with y axis being perpendicular to AB, pointing towards the bottom alignment hole (C).

- Images are collected at 500 locations arranged as a 20×25 (x, y) grid using an automated routine. The images are spaced so that there are 14 pixels of overlap between each image. At each location 10 images were collected at a resolution of 640×480 pixels using backlit illumination, dark-field illumination, and 8 different angles of rotational illumination described in §5.3.5. Illumination parameters were changed slightly from the Xe study to account for the different opacity of the Pb foils. At each location the set of 10 images represent one multi-channel image with backlit, dark-field, and 3D rotational image channels. The location of each image is stored as metadata.
- After each foil is scanned it is flipped over and scanned again from the reverse side using the same pinhole as the origin. The coordinate system is flipped to right hand Cartesian.

The scanned images are whitened and normalised. Images taken on the reverse side of the foil are digitally flipped to preserve the relative orientation of the etch pits compared to front side. The rotational image channels are permuted to account for this flipping.

7.2.2 Global stack coordinates

In the context of the NTD foil stack it is useful to define three important coordinate systems;

Local - A coordinate system defined relative to the images in the format (i, x, y) where i is the index of the image, and x, y represent pixel coordinates within the image, relative to an LHS Cartesian coordinate system with the origin in the top left corner of the image.

Scanning - A coordinate system (x_s, y_s) defined relative to the microscope scanning bed. This information is contained in the metadata for each image.

Global - An coordinate system of the form (x_g, y_g) parallel to the x y plane of the aligned stack and NTD foils, such that an ion beam normal to the stack would produce etch-pits with the same global coordinates in each foil of the stack.

Any object or feature visible in the scanned images intrinsically possesses a local coordinate. The aim is to convert local coordinates into global coordinates defined relative to the datum ABC so that objects in different scans can be compared.

Datum alignment and coordinate transformation Visual inspection of the scanned images is used to obtain the local position of each datum alignment hole. The diameter of each alignment pinhole is approximately 1.9mm, which corresponds to approximately 165-173 pixels in the lowest magnification images. A concentric circular target (inner radii 165 pixels, outer radii 173 pixels) is aligned with the edges of the alignment pinhole and its location adjusted till the edge was aligned within bounds as seen in fig 7.2. The error on this process is estimated at approximately 4 pixels (half the size of a typical etch-pit at this resolution). The location of each image relative to the microscope scanning bed can be extracted from the scan metadata. In combination with the image scale it is possible to assign each image pixel an x,y coordinate relative to the scan. Thus the local datum point coordinates (i, x, y) can be transformed into physical scanning coordinates (x_s, y_s) . The datum point coordinates can be converted from scanning coordinates into global coordinates via simple linear transformations, namely;

- Translation; such that datum point A becomes the origin by construction.
- Rotation; by a correction angle θ_c , such that datum point B lies on the x axis.
- Reflection; For scans taken on the reverse side of a foil, and where reverse side images are digitally flipped.

- Scaling; to convert between coordinates in physical SI millimetre units and pixel values. For practical reasons global coordinates are represented in pixels.

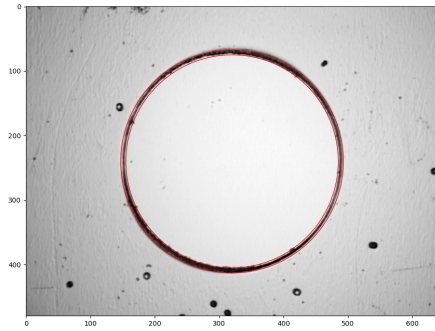
The local position of the datum points A and B determine the necessary translation and rotation angle θ_c for each scan such that A is the origin and B intercepts the x_g axis by construction. Overall this process defines a method for converting local image coordinates (i, x, y) into global coordinates (x_g, y_g) and vice versa that may be used to compare etch-pit coordinates in one foil to etch-pit coordinates in another (operating in reverse etch pits are assigned to the nearest image location). This process can be verified to first order by converting the location of the datum alignment holes in one scan to that of another. There is a 4 pixel error in the reconstruction of datum B. The following sources of error will contribute;

- Error in the location of the alignment pinhole
- Positional error of the microscope
- Mechanical error in the drilling of the alignment pinhole, and in the assumed flatness of the foil (which cannot be assumed on a microscopic scale)

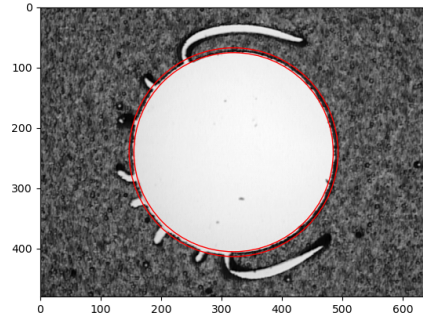
While this positional error scales with distance from the origin it is within the typical 8 pixel size of an etch-pit at this resolution. It is also within the 28×28 etch-pit neighbourhood used previously in the Xe study. Overall this enables approximate pixel to pixel level comparison between two different foil scans, whether taken from different foils or opposite sides of the same foil.

7.2.3 Etch-pit offset correction

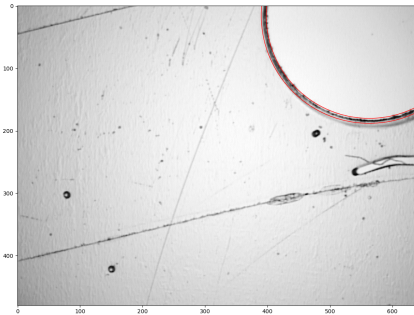
The global coordinate system makes it possible to compare coordinates between foils. For a normally incident Pb ion beam the global position of a Pb ion etch pit in one foil scan would correspond to the global position of subsequent etch pits along the track in



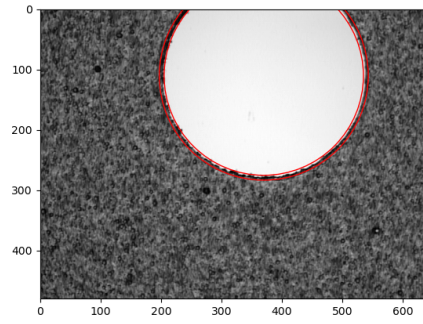
(a) Unexposed, datum A



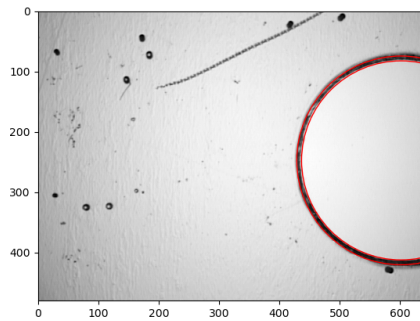
(b) Exposed, datum B



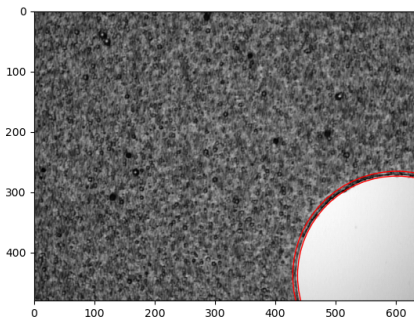
(c) Unexposed, datum B



(d) Exposed, datum B



(e) Unexposed, datum C



(f) Exposed, datum C

Figure 7.2: Datum alignment images

The first, second, and third datum alignment holes for the first unexposed and exposed foils respectively. The same procedure was followed for the other foils and the reverse side scans. The scans cover the same overall area, however on this scale each individual image will not necessarily match. After alignment a pixel to pixel correspondence between features can be established.

other foils. Because the experimental NTD stack was exposed to a Pb ion beam at a 15° angle of incidence, there will be a consistent offset of etch-pit locations from their expected position (had the angle of incidence been 0°).

To adjust for the angle of incidence I apply a fixed translation or offset to the etch-pit locations when converting between the exposed and unexposed foils. This is applicable for all the Pb ions, and allows for a more accurate reconstruction for etch-pit locations. Accounting for this offset also makes it possible to digitally align images to achieve pixel to pixel correspondence of etch pits in one foil to another. For analysis foils the position of the foil within the NTD stack relative to the IP determines the hypothetical angle of incidence of any HIP candidates.

Offset calibration For the research foils bulk properties of the Pb ion etch pits were used to calibrate the exact offset required. Pb etch pits were identified in the unexposed NTD foils (detailed subsequently in §7.3.1). Their approximate position in the exposed foil was calculated using the global coordinate system. As the rotational illumination channels capture light scattered of the edges of of an etch-pit we would expect a higher overall pixel intensity at these edges, with an approximately circular shape. Averaging the cropped images taken from the approximate locations of the Pb etch-pits over all the etch pits and all rotational imaging angles we can see in fig 7.3 a clear Gaussian peak where the known HIP etch-pits display a slight pixel intensity excess over the background.

Looking at the X and Y projections we can locate the peak intensity and calculate the pixel correction required to centre the Gaussian. We can also identify the approximate size of etch-pit features, namely a half-width of approximately 14 pixels at the lowest magnification setting ($\approx 80\mu m$), with most pits being around 10 pixels. We can now map from the centre of etch-pits in the unexposed foil to the centre of etch-pits in the exposed foil (fig 7.4). This process is repeated for the reverse side. Lastly we can verify that there is not a spatial variation in the correction by dividing the foil into 4 quadrants ABCD. If

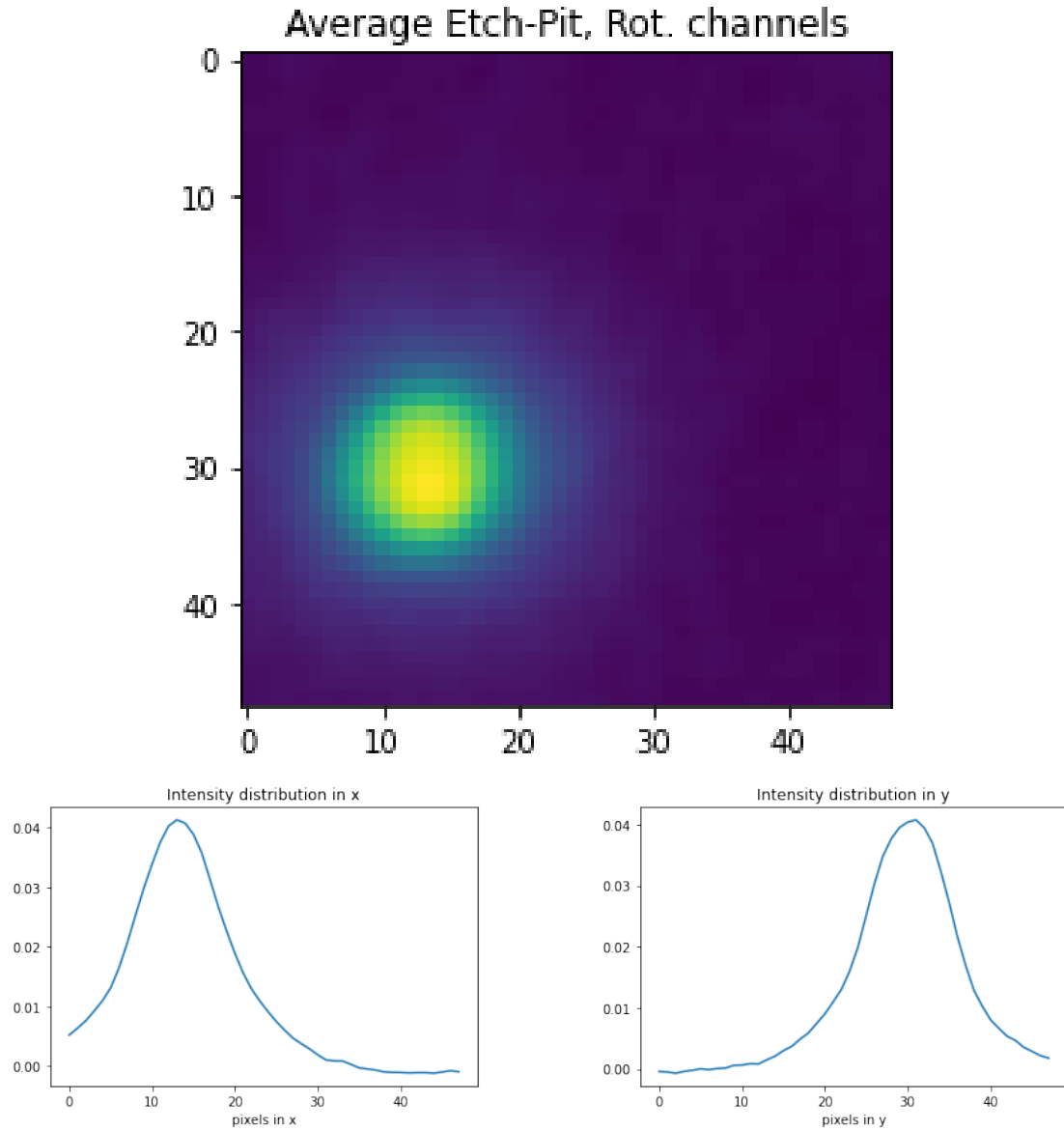


Figure 7.3: Average Etch-Pit image of the rotational illumination channels (averaged over 8 channels, and all etchpits). We can see that the locations derived from the unexposed foils are approximately correct. Shown with a 48x48 pixel region centred on the expected location with X and Y projections, used to derive the required correction.

there was a significant spatial variability we would expect the correction to change with distance from the origin. The observed variation (fig 7.5) is within 1 or 2 pixels.

7.2.4 Summary

In this section i have outlined a method for aligning etch pits and their features between two separate foils or scans. Converting etch pit coordinates forms a three step process;

- Local image coordinates (i, x, y) in one foil are transformed into global coordinates (x_g, x_y) using the transformation defined for that foil
- A fixed offset is applied to the global coordinates accounting for the positional shift between foils due to the beam angle of incidence.
- The global coordinates are transformed into the local coordinates of the other foil.

This allows for coordinates and objects to be compared between foils. Alternatively the images themselves can be cropped or shifted so that features match on a pixel to pixel level, this has multiple applications for simultaneous analysis of both foils (an example of such aligned images can be seen in Fig 5.6b). An illustration of the coordinate transformation process is shown. Etch pit locations (fig 7.6a) are identified using the Gaussian Laplace algorithm §7.3.1. Fig 7.6b shows the unexposed positions naively superimposed on the exposed images. While they appear correct in this image (taken near the origin) they will diverge substantially at the furthest regions as seen in the datum images (fig 7.2). Expected etch pit locations (blue) in fig 7.7a obtained by converting the unexposed locations into global coordinates and back (with no beam correction) delivers much closer results. This alignment holds at the extremities of the foil. Only the consistent effect of the beam angle prevents an exact match between expected locations and their actual positions. When the beam offset is corrected for (fig 7.7b) the etch pit locations can be matched between the unexposed and exposed foil to a precision smaller than an typical etch pit. This alignment means exposed etch pits can be reliably found

from the unexposed locations. Conversely candidates locations in the exposed foil can be precisely tracked in the downstream foil.

This methodology can be replicated for any test beam stack. The beam offset calibration process absorbs any fixed error in coordinate alignment. For MoEDAL analysis foils there is discussion of replacing the mechanical alignment pins with a laser marking. Some of these marks could be made at a fixed angle. This would allow for a similar data driven calibration between the known angle and the observed displacement between the marks. The foil alignment offset could then be adjusted to an origin hypothesis.

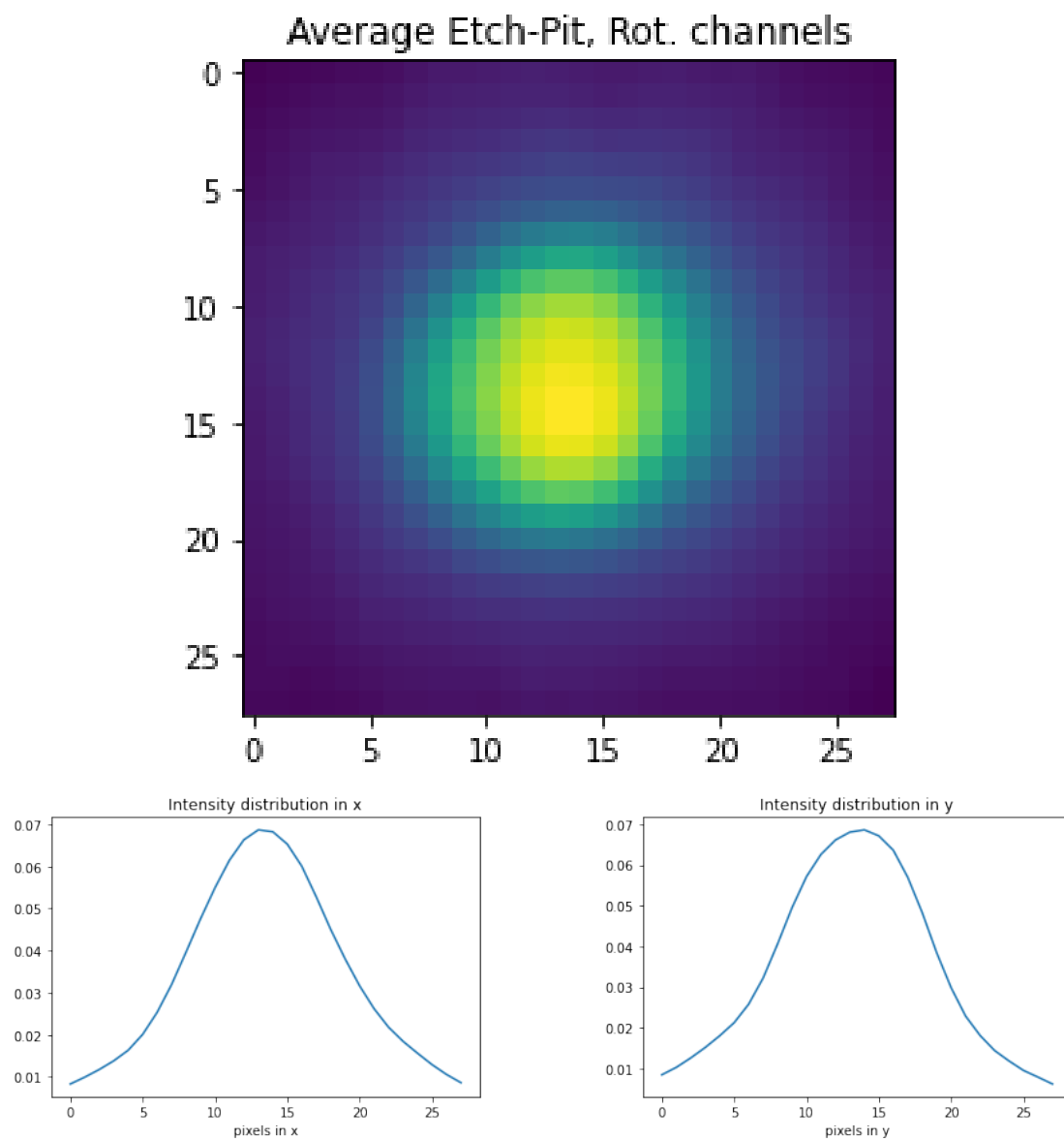


Figure 7.4: Average Etch-Pit image after offset correction

After applying the beam offset correction the average position of etch pits in the exposed foil (given by the gaussian) matches their expected position derived from the unexposed foil (given by the image window). The image window shown is also the smaller 28x28 pixels I use for pit classification images.

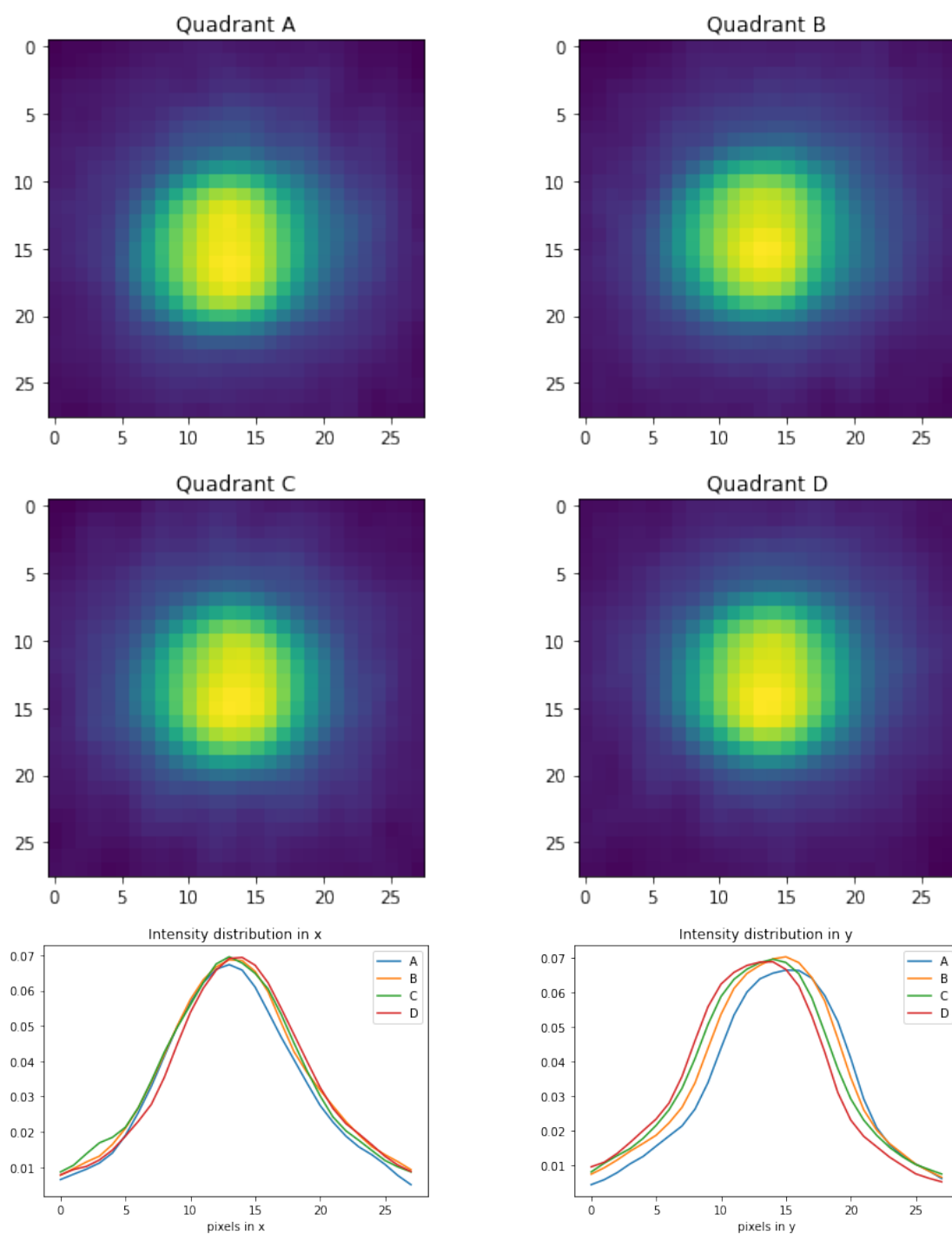
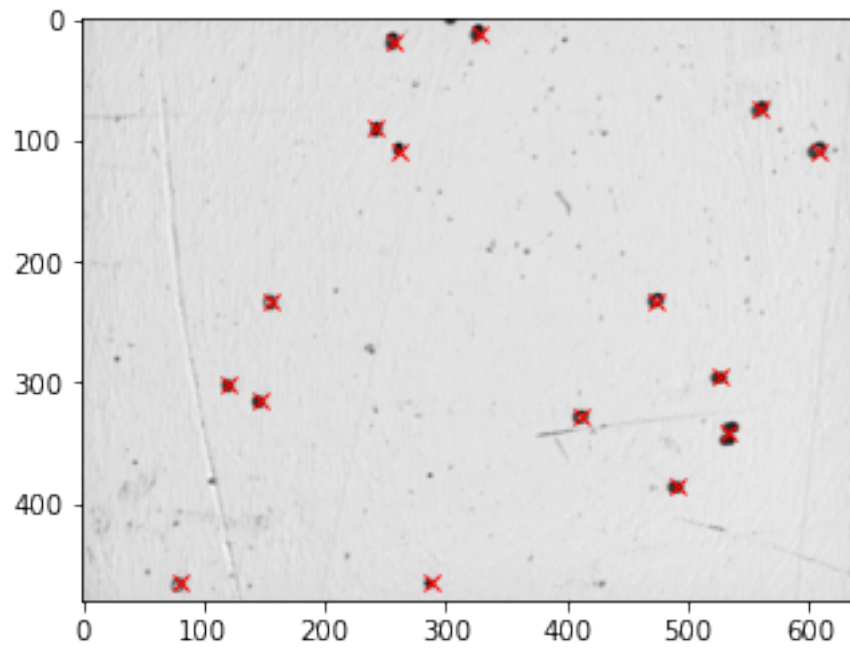
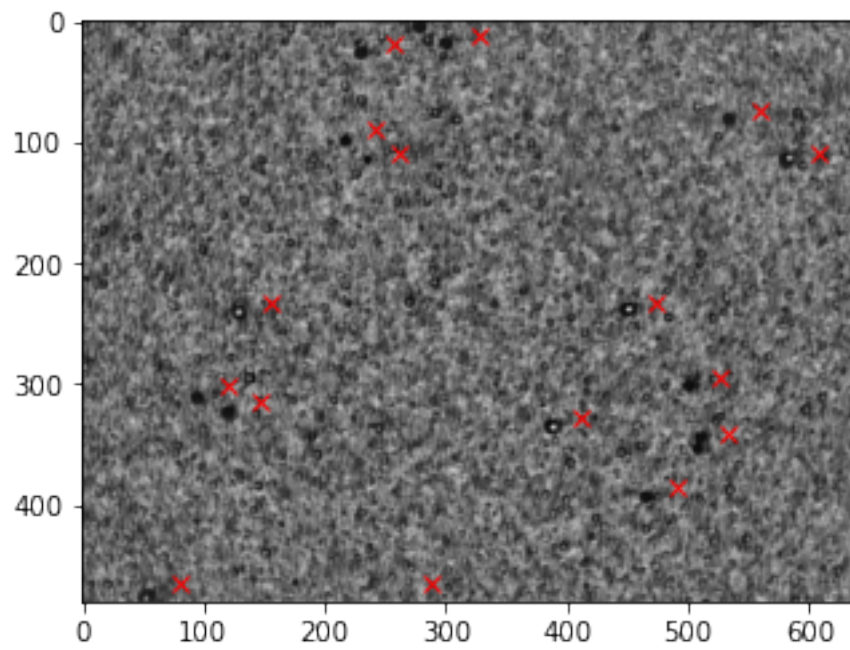


Figure 7.5: Spatial dependency over 4 quadrants of the foil.
 The average image in appears centrally, with no significant positional variation between each quadrant. indicating the alignment procedure works consistently throughout the foil.

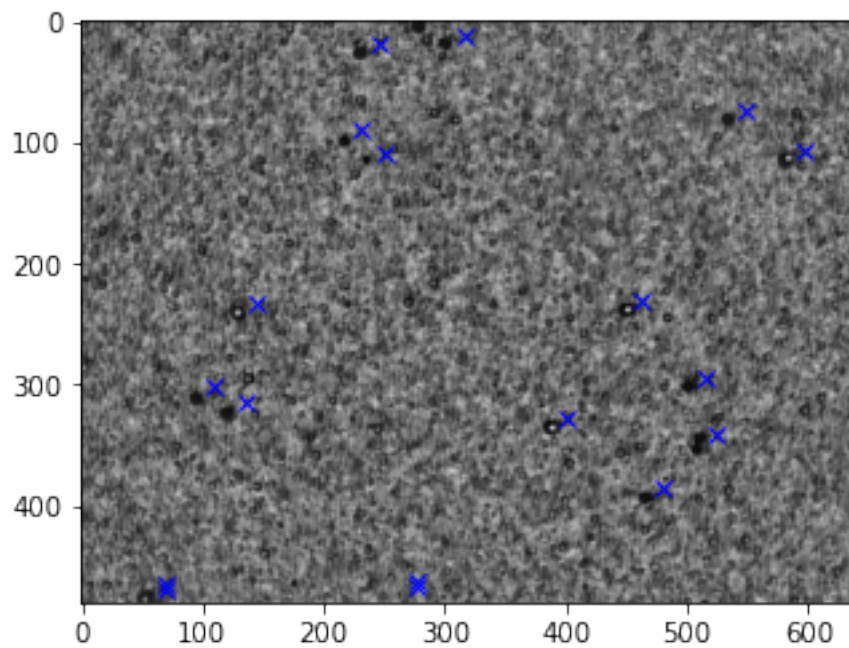


(a) Unexposed foil locations, found via Gaussian laplace

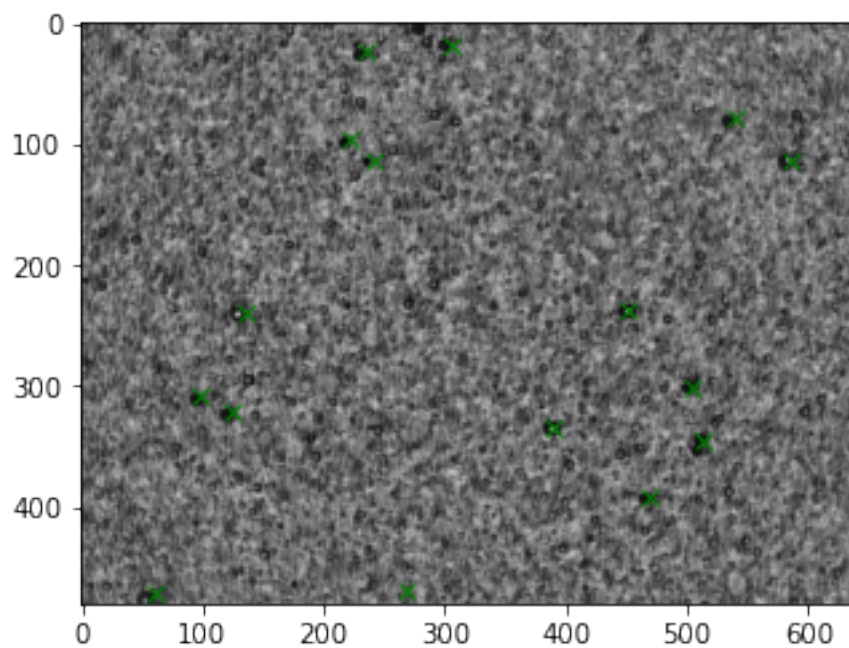


(b) Unexposed locations (local) superimposed on exposed foil

Figure 7.6: Naive alignment



(a) Unexposed locations converted using global coordinates



(b) Location corrected to account for beam angle of incidence

Figure 7.7: Corrected alignment

7.3 CNN classification study

The eventual aim is to be able to identify regions of interest within the NTD stack using a regression algorithm over wide area images. Nevertheless the investigative starting point is to look into CNN etch pit classification which is the focus of this section. The approach adopted is as follows;

- 1) Establish a baseline performance in classification tasks.

Classification provides an easier environment for quantitative evaluation. It is easier to both train and evaluate numerically (as an image either contains signal features or it doesn't). The baseline classification performance represents an estimate of the upper limit of performance we might expect. CNN classification can also be used to optimise hyper-parameters and architecture

- 2) Use this as the basis for inference based location prediction, covered in the next section. The methods used for classification can act as a starting point when approaching location prediction, which due to its more open ended nature can be harder to optimise for. This is essentially a transfer learning approach (see §4.4.4) that aims to leverage the results of the narrow etch pit classification task to aid in wide area regression.

7.3.1 Etch pit Classification dataset

This subsection details the automated method used to obtain a set of known Pb ion locations to act as 'Truth labels' used to train the CNN models. This method was also used to collect true etch pit locations used for the final evaluation.

Data preparation

The scanning data was divided approximately 50:50 into a training region and an evaluation region. The first 250 scanned images form the training region, to be used for

training and optimising the CNN based etch-pit identification. The last 250 images were held as a blinded evaluation region, ensuring the data used to obtain final performance measurements in §7.5 is unbiased. Additional exclusion cuts were applied to remove small regions where exposed images (1-250) and unexposed images (251-500) overlapped in global coordinates, this ensured that all training etch-pits appear only in images 1-250, and all evaluation etch-pits appear only in images 251-500. Also regions on the edge where simultaneous exposed and unexposed image data wasn't available were excluded. These exclusions were applied by discounting objects (such as predicted or true etch pits) with global coordinates in these regions from the final analysis.

For each foil, and each illumination data channel the average 640x480 image was calculated for the entire training set. This average image demonstrates the systematic bias of the microscope within the image in each illumination channel. Both the training and evaluation image sets were normalised by subtracting the average image (normalised over the number of images, and smoothed to remove etch pit level features). The result is improved contrast and a more consistent image. This was applied to both the training and evaluation image set (using the average image determined from the training set).

Images taken on the reverse side of the NTD foil were digitally flipped. For rotational images the order of the channels were permuted to preserve the significance of various etch pit patterns.

Identification of Pb ion tracks

Positive identification of Pb ion tracks within the unexposed NTD foils was obtained by first identifying probable Pb Etch-pits within the images taken from each scan of the NTD foil. Within the unexposed NTD foils Etch-pits associated with Pb ion tracks appear as distinct dark round objects in the backlit images, approximately 8 pixels in width (micro size / measurement). Pb etch pits may or may not appear with a bright central region indicating light transmission.

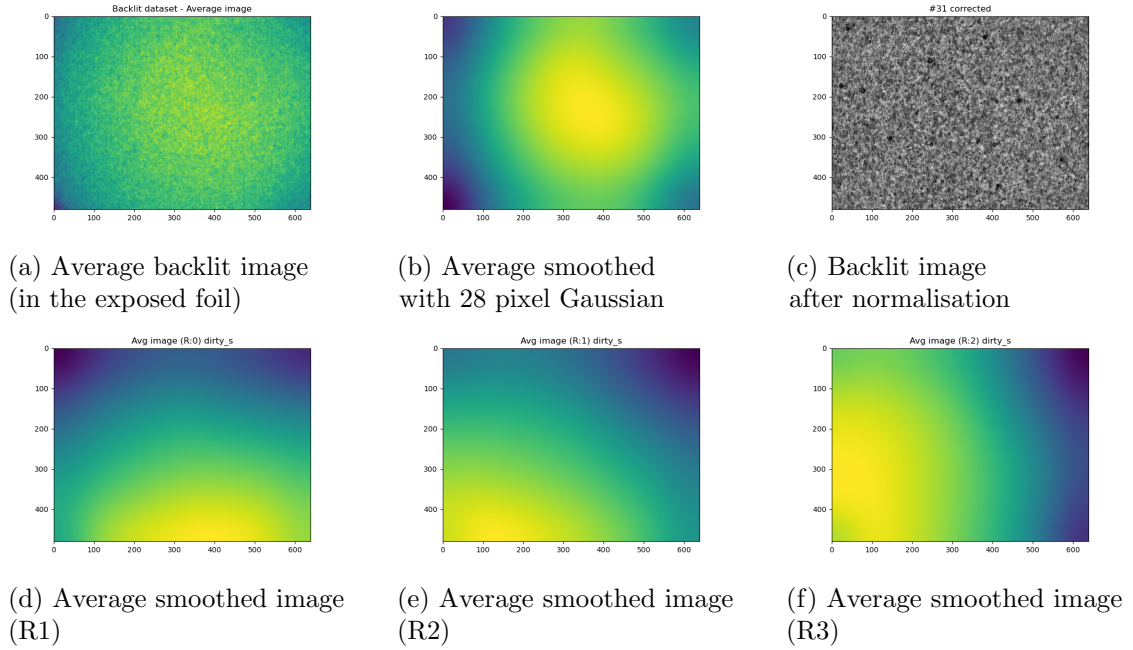


Figure 7.8: Image normalisation

The average image (a) is calculated for each illumination channel, and smoothed with a 28 pixel Gaussian to remove etch pit noise (b). This represents the systematic bias in pixel intensity within the image present with each choice of illumination method. The original images can be renormalised by subtracting this systematic bias (c). Images (d), (e), and (f) show the systematic bias of the first three rotational imaging channels which moves clockwise around the foil with the direction of the incoming light.

Gaussian laplace algorithm Etch pit detection in the unexposed foils is achieved via an algorithm utilising repeated application of the Gaussian Laplace filter (via SciPy.ndimage [50]), and a threshold cut. The Laplace filter represents the second order derivative of an image, acting as an edge detection algorithm. The Laplacian is typically combined with a Gaussian smoothing element to reduce noise sensitivity;

- Gaussian Laplace operator, with a width of $\sigma = 8$ pixels
- Threshold cut, discarding regions below the threshold
- Applying a second Gaussian Laplace operator with the same width.
- A second threshold cut. The centre of each Pb etch-pit will typically be a distinct separate region in the image segmentation map, the regions in each image are labelled and the etch-pit's location taken as the centre of each region.

The Gaussian Laplace algorithm is successful in identifying Pb ion etch-pits within a single unexposed image as demonstrated in fig 7.6a. A smaller or larger Gaussian width can be used to detect smaller or larger etch pit like objects. The 8 pixel width parameter was chosen to avoid picking up on smaller non-Pb etch pit noise. This algorithm is robust to different illumination conditions as it depends on changes in intensity rather than absolute value. However it will occasionally misidentify a surface scratch or other malformed region of foil as an etch-pit.

Etch pit matching between foils Potential Pb ion tracks were identified by matching potential etch pit identifications between both sides of the first unexposed foil, and the second unexposed foil. Possible Pb ion tracks were parametrised into high and low quality groups according to the following scheme;

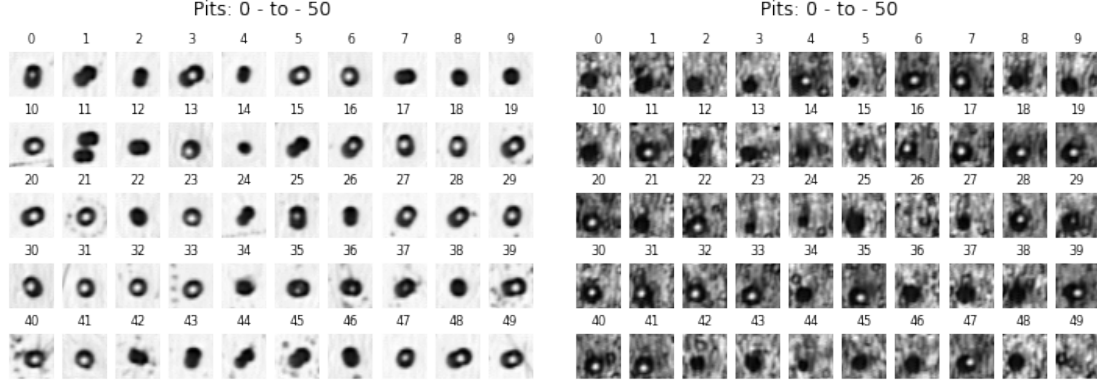
- High quality through-going etch-pits, requiring a positive etch-pit identification on both sides of the first unexposed foil, and in the second unexposed foil, with

subsequent etch-pits identified within a 12 pixel tolerance of their expected location in global stack coordinates. (High likelihood of being from a Pb ion track).

- Low quality through-going etch-pits, requiring a positive etch-pit identification on both sides of the first unexposed foil as above, but without a requirement of an etch-pit in the second unexposed foil.

Pits meeting the high quality criteria were present on both sides of all four foils, thus only data from the first two unexposed foils was required to categorise through-going etch-pits. The 12 pixel tolerance factor is based on the size of etch pit objects, and potential alignment error, while being small enough to avoid mismatching. Track ID is invariant under small changes to this tolerance value. Objects only appearing on one side of a foil, or only one foil can be rejected as not being Pb ions. A likely source besides non-etch pit noise is fragmentations from the Pb ion beam. While etch pit tracks in the low quality category are not necessarily Pb ions they are still of significant interest as they may represent through going ionisations in the exposed foil. There is no guarantee that these will have left any ion track or identifiable etch-pits in the LHC exposed foil as they didn't leave a signature in the second unexposed foil.

Etch pit quality cuts Scratches and other items of non-etch-pit image noise detected by the Gaussian Laplace algorithm were filtered out using a simple 2D CNN classifier trained on 28 manually selected examples of scratches (and 28 genuine Pb etch-pit images). The architecture of this classifier is the same as that used in §7.3.2 for generic 2D classification in the exposed foils. Due to the simplicity of image classification in the unexposed NTD foils it is trivial to train to a high degree of accuracy with a very small training sample. This was used in inference to filter out scratches from both training and evaluation pit datasets before manually validating the results to remove a small number of errors. No scratches were mistakenly identified as etch pits. While matching between foils is largely sufficient for identifying Pb ion tracks this method is useful for recovering



(a) High quality etch pit examples in unexposed foil (b) Corresponding etch pits in exposed foil

Figure 7.9: High quality etch pit examples

Etch pits meeting the high quality criteria. When their location is known they can be quite distinct within the exposed foil.

etch pit from the low quality object group.

Fragments and weak ionisations Another object category was identified after initial research as being persistently misclassified. These etch pits met the high quality criteria, passing through all the unexposed NTD foils, yet did not appear to be Pb ions. These etch-pits correspond to weak ionisations as indicated by their etch-pit size less than half that of a Pb ion (since the size of an etch-pit is proportional to the REL restricted energy loss). Since they pass through multiple NTD foils indicating a high energy and momenta we can infer that they came from the Pb ion test beam; this would indicate they are other ion species originating from fragmentation of the high energy Pb ions, often appearing as pairs, indicating a fragmentation of the Pb ion into two smaller fragments.

While fragments and weak ionisations traversing the NTD stack meet the definition of through-going ion tracks they will not leave the same signature in the exposed NTD foil as a Pb ion. We don't know they ranged-in and began ionising above the threshold of etch-pit formation before or after the LHC exposed NTD foil. And their smaller etch pits might not be practically identifiable vs the numerous background etch-pits caused

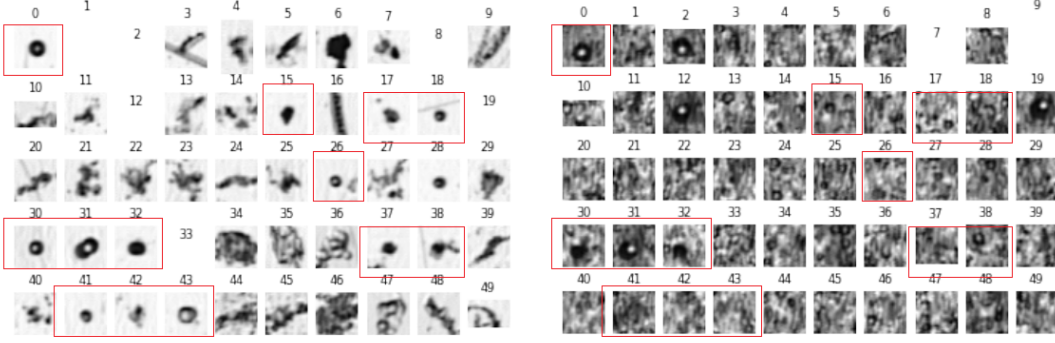


Figure 7.10: Highlighted comparison of Low quality objects

We can see many objects are large scratches or ambiguous, but we can also identify several obvious etch pits. Coming from weaker ionisations they may only just have ranged in, and might be too subtle to identify in the exposed foil.

by many spallation ions and fragments at the LHC.

These track events were filtered, in the same way as the scratches, by training a simple 2D classifier acting in the unexposed foils. Accounting for 104 etch pits they will still be used in this research as we would like to maintain sensitivity to such events. However they will not be included in the main Pb ion dataset.

Training datasets, and validation split

After requiring matching, and filtering out scratches, and fragments, the tracks identified within the training region were taken to form a discrete classification dataset. The exposed locations were calculated using the foil alignment §7.2.2. The area around each etch-pit was cropped with a margin of 14 pixels from the images in each illumination channel. This produces an image of 28x28 pixels for the backlit channel, and halo channels, and 28x28x8 pixels for the rotational illumination channels. Alternately this may be thought of as a single 28x28x10 image.

Of 1397 Pb tracks meeting the high quality criteria, this was reduced to 1249 after accounting for exclusions. This produces a dataset of 1249 Pb signal examples. 249 examples were separated off to act as a validation dataset. The remaining 1000 training

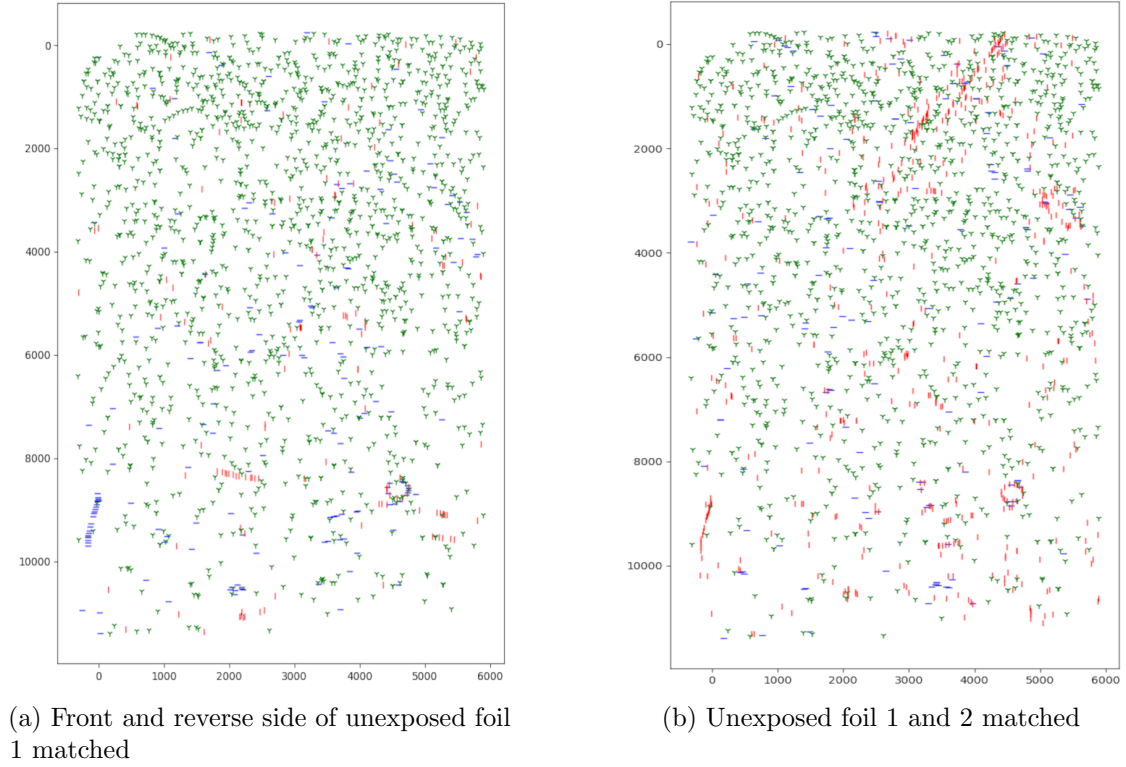


Figure 7.11: Etch pit matching between foils

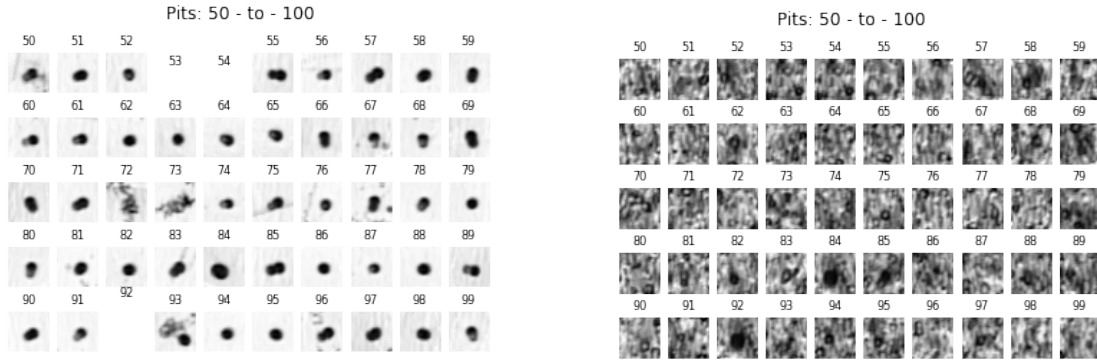
Etch pit like objects detected using Gaussian Laplace are required to match with objects in down stream scans to within 12 pixels. In the first plot green highlights detected Pb etch pits matched on both sides. Red indicates objects only appearing on the front surface of the foil, blue indicates objects only occurring on the reverse. In the second confirmed objects from foil 1 are matched with objects on the surface of foil 2. Green indicates a match, Red indicated foil 2 only, and blue indicates foil 1 only. We can observe a set of false positives around the confusing alignment pin hole. In the lower left we can observe matching lines produced by a large anomalous scratch that appears on the front of both clean foils (possibly left by some item of foil handling equipment during etching such as a tweezer)



(a) Predicted etch pit examples

(b) Predicted scratches

Figure 7.12: Scratches filtered out via CNN classifier



(a) Fragments in unexposed foil

(b) Fragments in the exposed foil

Figure 7.13: Identified fragment examples

In the unexposed foil we can observe that the fragment etch pits are all smaller than the etch pit size of a typical Pb ion, indicating a smaller value of $\frac{Z}{\beta}$. In the exposed foil many of these fragments are undetectable to the naked eye among LHC background.

Total folded dataset	1000	500	250	200
Training fold	800	400	200	160
Test fold	200	100	50	40
Validation	249	249	249	249

Table 7.2: Training datasets

examples were separated into 4 different overlapping subsets of 1000, 500, 250, and 200 examples so that achievable performance could be benchmarked with different quantities of training data. These small, medium, and large sized datasets were further subdivided according to a K folding scheme (K=5) §4.4.2 for fold sizes of 40, 100, and 200 respectively. Thus giving an 80:20 split between training and testing data for each subset.

For clarity it should thus be noted that there are three levels of testing or validation data present in this chapter. Each K-fold training dataset has its own testing fold used to evaluate the performance of the learner trained on that fold. As this data will be used to tune optimisation a separate validation set is used to obtain final performance estimates for CNN classification. The blinded evaluation dataset, taken solely from the blinded evaluation region of the foil is used to evaluate performance of the full FCN algorithm introduced later, and not used to evaluate CNN classification performance.

'Background Etch-pits' Representative background etch-pits were identified by randomly sampling from regions of the foil stack that did not contain either high or low quality tracks. Each (sub)dataset in table 7.2 was paired with an evenly balanced set of background etch pits. Thus doubling the overall size of each dataset.

7.3.2 Training methodology

The networks used in the Pb study were trained using a standardised process, mostly using a common set of architectures and hyper-parameters as detailed below;

K-fold training process CNN classifiers were trained on the K-folded datasets using the following process; for a given task, and set of hyper-parameters 5 network instances would be trained from scratch, each using one fold as a test set, and the remaining K-1 (4) folds as training data. The application of K-folding to the input data provides an estimator for the variance in performance (and average performance) resulting from different choices of input data, for a given architecture. Each training run was repeated between 3 and 4 times to ensure that the time to converge (in epochs), average loss/accuracy converged upon, and the spread of results between K-folds was consistent and repeatable, with no significant dispersion from the average result. Thus repetition r is representative of the training evolution that may be expected of repetition $r + 1$. This follows the principle of nested K-folding (§4.4.2) where results would be averaged over both folds and repetitions). Here only the final repetition is kept, once training runs are shown to be repeatable, consistent, and representative of one-another. If results were shown to not be repeatable then modifications were made to input parameters until they became consistent.

Where K-folding hasn't been used for a dataset repetition alone is used instead.

Network architecture Most investigatory work used a common network architecture, similar to that used for the Xenon ion study. This architecture can accept either 2D or 3D input, with the additional spatial depth being absorbed into the feature depth. The main convolutional layers alternate between convolution operations and max pooling operations, before being condensed down. Variations of this architecture and others were investigated and optimised before settling on this choice. The tanh activation is retained so that the network can act with either a [0,1] or [-1,1] input range. No effort is made to either narrow or widen the feature depth, as no problems were observed with either convergence or over-training. The kernel size was increased compared to the Xe networks, and the final output changed to a single value sigmoid, this improved compatibility with the fully convolutional output of §7.4 while delivering identical results in terms of

accuracy and stability.

Layer	Feature depth	Activation
Conv (4x4)	32	Tanh
MaxPool (2x2)		
Conv (4x4)	32	Tanh
MaxPool (2x2)		
Conv (4x4)	32	Tanh
Maxpool (2x2)		
Conv (4x4)	32	Tanh
Dense	22	ReLU
Dense	1	Sigmoid
Global MaxPool		

Table 7.3: Standard CNN architecture used for Pb study.

This architecture performed consistently and robustly with no adjustments required in the Pb, and Xe foils, in the classification of different types of etch pit object, and across multiple data channels. While elements of this architecture could be considered sub-optimal, the properties of generalisability and robustness are worth the trade-off especially as input data was by far the biggest determinant of final performance.

In the rotational illumination channels comparisons were made between a 2D convolution in (x, y) , using a 3 dimensional kernel and using a full 3D convolution operation in (x, y, ϕ) with a smaller 3 dimensional kernel. 3D convolution was explored using the architecture show in table 7.4, and variations thereof. While models using a 3D convolution converged rapidly they weren't able to converge on the same level of performance as the more typical 2D convolution.

Hyper-parameters, and over-training Most networks were trained using the following set of standard hyper-parameters. Similarly to the Xenon ion investigation different values and combinations were tried and tested, however changes to the standard set were not adopted. Normally the common concerns when optimising machine learning

Layer	Feature depth	Activation
Conv3D (3x3x3)	32	ReLU
MaxPool (3x3x3)		
Conv3D (3x3x3)	32	ReLU
MaxPool (3x3x3)		
Dense	22	ReLU
Dense	1	Sigmoid

Table 7.4: CNN architecture used for 3D convolution

tasks are the speed of convergence, and the risk of introducing Over-training (4.4.1). Over-training occurs when a network learns to pick up on specific features of individual training examples, rather than general signal patterns, over-optimising to identifying those specific training examples at the expense of generalisable performance. While there are multiple approaches to reduce over-training the most direct approach is typically to stop the learning process early, thus the number of epochs training runs for is typically a significant hyper-parameter in the overall network optimisation process.

In the case of the MoEDAL etch-pit classifiers over-training is hardly ever observed (Figure 7.14). I would hypothesise this is because each network is relatively small, with few parameters to optimise, while the signal represents a simple yet distinct pattern. Performance converges rapidly on an average result, consistent between different data folds, with no loss or gain in test performance achieved by allowing the learning process to continue longer.

For this reason typically important hyper-parameters; batch size, training epochs, and the number of training steps in each, are relatively unimportant vs the final accuracy or loss performance achievable, which is limited by the network architecture and the patterns available in the training data. The Adam optimiser once again proved most effective. Most networks were allowed to run for 60 epoch, with most converging in 10-20 depending on task.

Hyper-parameters	
Optimiser	Adam with a learning rate of 0.0002 ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 0$)
Batch size	50
Epochs	60 'epochs' of 5 batches per epoch (each being 250 examples total)
Loss	Binary cross-entropy

Table 7.5: Pb CNN training Hyper-parameters

Epochs normally represent a run over the entire dataset. Here the approximate size of the validation dataset (249 examples) is used as a standard measurement for comparison.

7.3.3 CNN Results

This section highlights the performance of the classifiers trained on each data channel (backlit, dark-field and rotational). Results are obtained using the validation dataset of 249 Pb examples, and 249 background examples, unseen during training. The only exceptions to this are the predicted scores for the fragments (Figure 7.17), which use the entire set of Pb fragments due to low total numbers, and the validation metrics in the training history which use the test sets of each fold. Where appropriate the results are averaged over the 5 k-fold models.

CNN performance tables

Table 7.6, and Table 7.7 shows the final CNN performance (measured as classification accuracy) achieved in each data channel with fragments included or excluded from the training data respectively. The results are shown for each fold, along with the average and standard deviation of all 5 folds. All models achieve a high level of accuracy. The results are robust and consistent, with little variation between folds and repetitions. Classifiers trained on dark-field and rotational illumination data consistently outperforming backlit illumination.

The average false positive, and negative rates are shown in Tables 7.9 and 7.8. Rotational illumination trained without fragments has a particularly low false positive rate, which is the most important metric in the context of MoEDAL. This may be

Channel	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average	Std Dev
Backlit	97.29	96.45	97.08	97.50	97.29	97.12	0.36
Darkfield	99.79	99.79	99.16	99.38	99.58	99.54	0.24
Rotational	99.79	99.58	99.58	99.38	99.58	99.58	0.13

Table 7.6: Pb CNN validation accuracy % (fragments included in training data)

Channel	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average	Std Dev
Backlit	97.50	98.33	97.71	98.13	97.71	97.88	0.31
Darkfield	99.79	99.79	100	99.79	99.17	99.71	0.28
Rotational	99.38	98.96	99.38	98.75	99.17	99.12	0.24

Table 7.7: Pb CNN validation accuracy % (fragments excluded from training data)

expected as it will be very hard for background etch pits to mimic the distinct features of HIP candidates in the 3D rotational illumination data.

Channel	False positive rate	False negative rate
Backlit	2.41	3.33
Darkfield	0.50	0.42
Rotational	0.67	0.167

Table 7.8: False positive, False negative rates % (fragments included) 0.5 working point.

Channel	False positive rate	False negative rate
Backlit	1.25	3.00
Darkfield	0.33	0.25
Rotational	0.17	1.58

Table 7.9: False positive, False negative rates % (fragments excluded) 0.5 working point.

Training history

Figure 7.14 shows the accuracy evolution during training of the 5 k-fold models when they have been trained with and without Pb fragments included as part of the Pb signal training data. In all cases the classifiers train quickly and converge within 15 epochs, plateauing around a similar performance level, with no evidence of over-training. Training

is a robust and highly repeatable process, delivering a similar result each time irrespective of the data samples used. Including Pb fragments as part of the training and test data doesn't impact the final validation performance, although they negatively impact the stability of training accuracy. Figure 7.17 shows most fragments effectively appear as background samples hence their inclusion with the main Pb signal data adds irrecoverable errors into each batch of training data without any benefit to robustness or over-training. Figure 7.15 demonstrates similar results for the loss metric (binary cross entropy).

Histograms

The histograms in Figure 7.16 show the score distribution, of the Pb signal and LHC background etch pits. For each data channel the results are amalgamated for every classifier in the k-fold ensemble. There is a distinct separation of signal and background in all cases. There is minimal overlap of the two distributions. A small set of signal examples overlap with the background peak in the bin nearest to 0.0. This means fine-tuning the score threshold between signal and background would do almost nothing to affect the performance and accuracy of the classifiers. The false negatives (missed Pb examples) are effectively irrecoverable, for that classifier at least. When comparing classifiers trained with fragments excluded from the training data, versus those where Pb fragments were included as Pb signal examples, we can see in the left hand plots of Figure 7.16 that training without fragments in the signal training data gives a sharper peak. Figure 7.17 shows the distribution of fragment scores for each ensemble of classifiers. The effective distribution is the same regardless of training methodology. Most Pb fragments peak at 0.0, and aren't distinguishable from LHC background, this is especially noticeable with the rotational illumination classifiers. The conclusion for a single classifier is that a small fraction of Pb etch pits, and a large fraction of Pb fragments, are indistinguishable from background etch pits and irrecoverable.

2D Histograms

Figure 7.18 compares the output of different classifiers for the same etch pit. Each data point represents an etch pit with three (average) classification scores assigned to it, one for each data channel. Pb etch pits are shown in blue, while background etch pits are shown in red, and fragments in green. These results are shown projected onto 2D scatter plots. Each axis represents the mean output of the classifiers acting on a different data channel (backlit, dark-field (halo), and rotational). This demonstrates the degree of correlation between the results obtained in different data channels. If classification

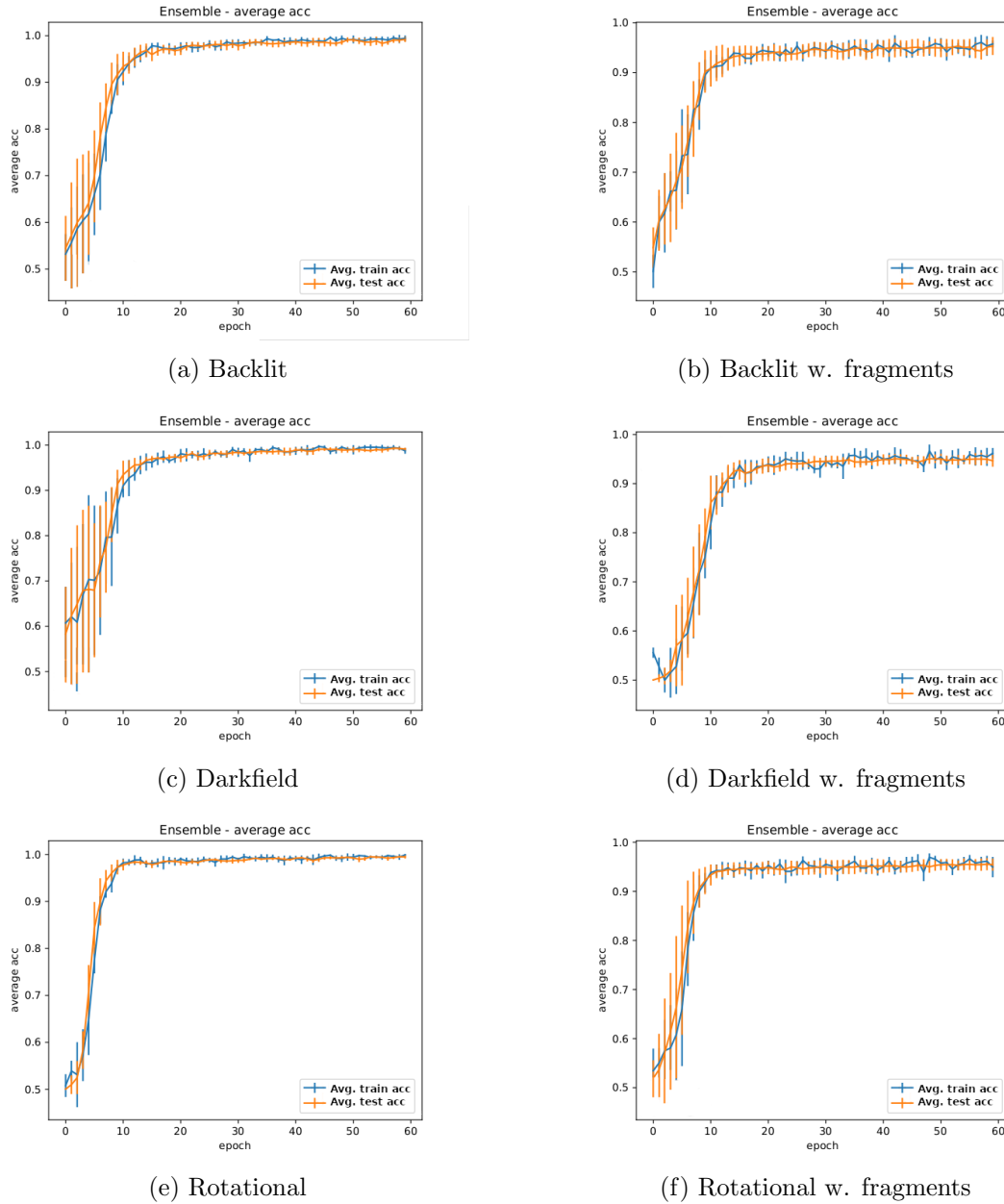


Figure 7.14: Accuracy evolution during training

Shown as an average training accuracy (blue), and average test accuracy (orange) of 5 k-fold models, and their spread. Left hand plots show results without fragments in the signal sample, as might be expected including fragments produces a greater variation the the converged result. The wide variation in the first 10 training epochs reflects how quickly the models learn basic etch pit features, which ones are learnt first by each model produces greater initial variation. The consistency of training in the rotational channel suggests all models are relying on the basic rotational pattern discussed in Chapter 5.

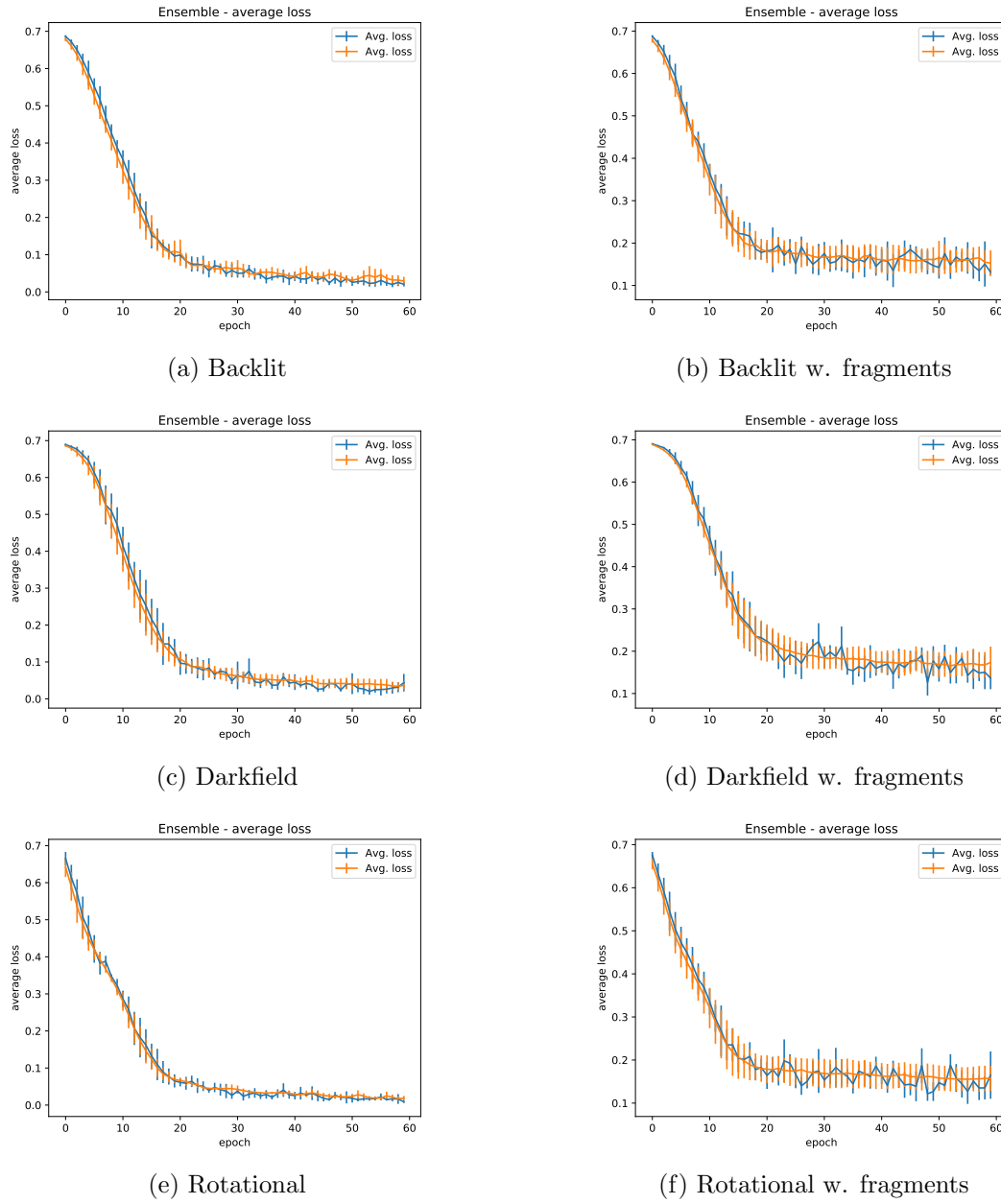


Figure 7.15: Error/Loss evolution during training

Shown as an average training loss (blue), and average test loss (orange) of 5 k-fold models, and their spread. Left hand plots show results without fragments in the signal sample. Including fragments produces a greater variability in the results.

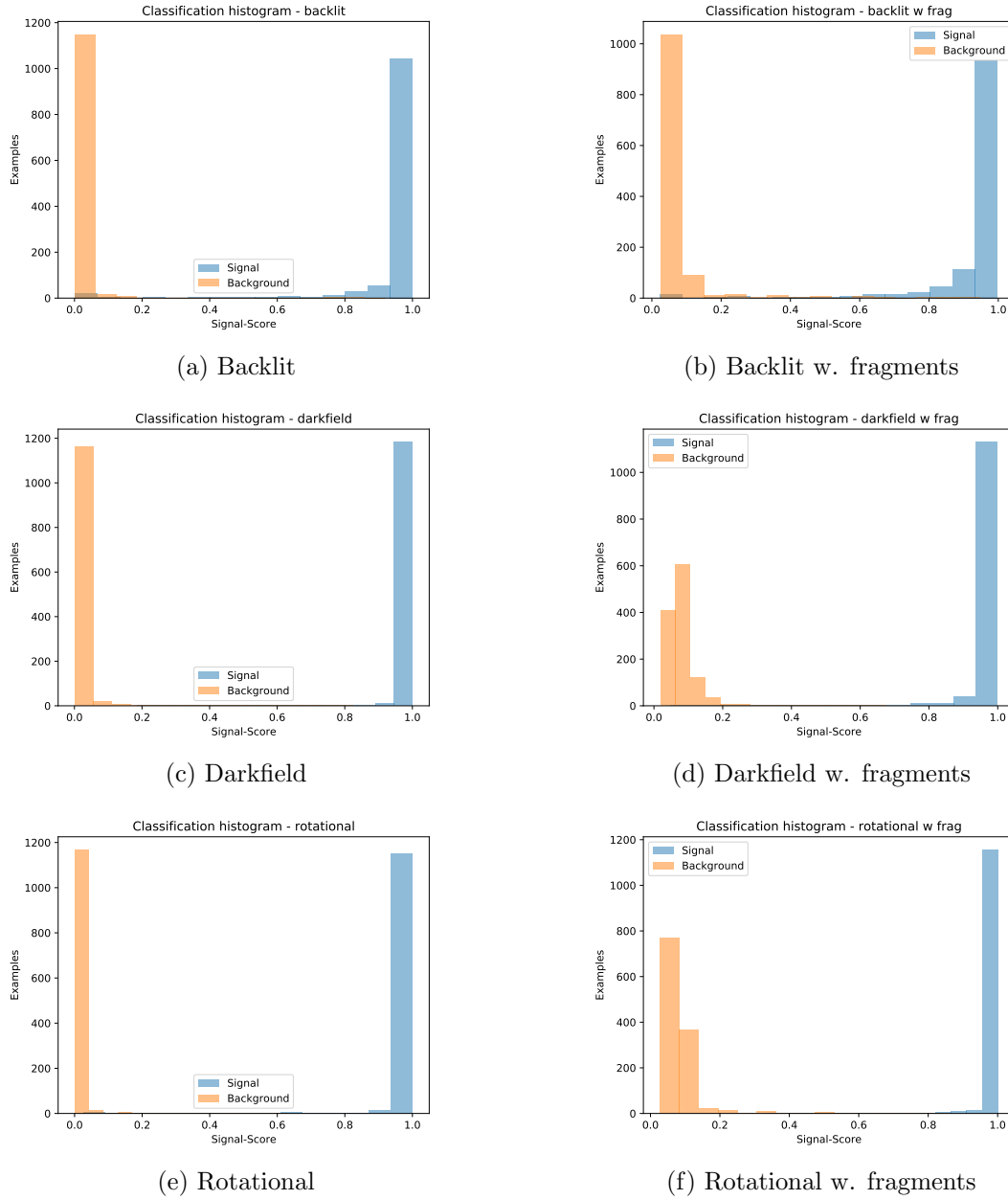
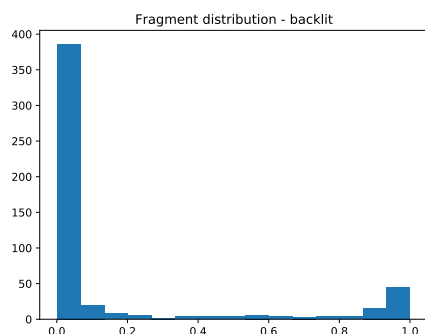
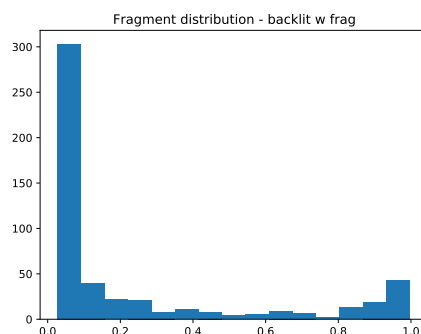


Figure 7.16: Signal/Background Histograms

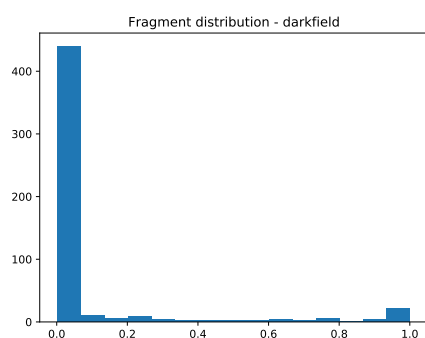
Signal and background score distributions for each classifier, amalgamated for each k-folded ensemble. In b) d) and f) the fragments were treated as signal examples for training purposes. The signal histograms are for pure non-fragment signal examples only, fragments are shown in Fig 7.17.



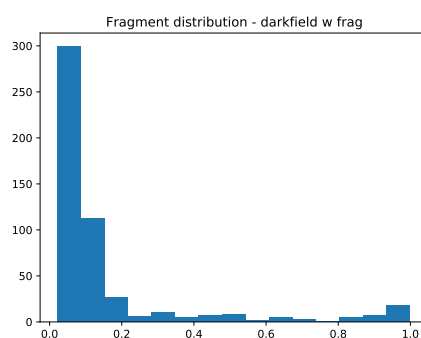
(a) Backlit



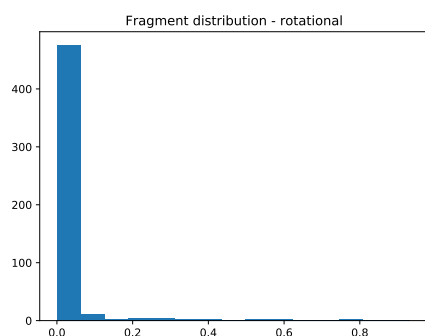
(b) Backlit w. fragments



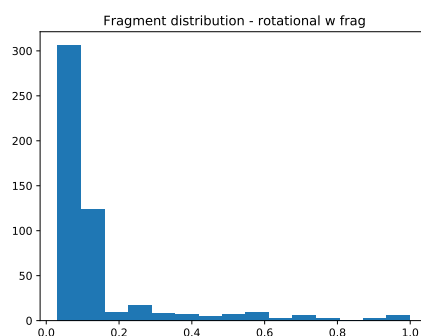
(c) Darkfield



(d) Darkfield w. fragments



(e) Rotational



(f) Rotational w. fragments

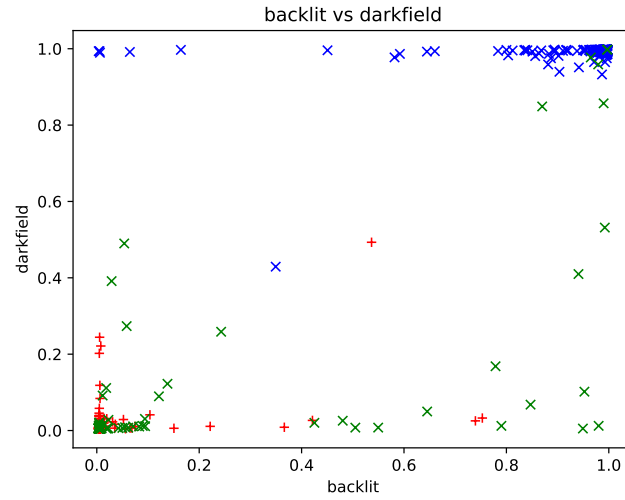
Figure 7.17: Distribution of fragment scores
Fragment score distributions for each classifier, amalgamated for each k-folded ensemble.

performance was perfectly correlated between the data channels we would expect the scores to be arrayed along a line between (0,0) and (1,1).

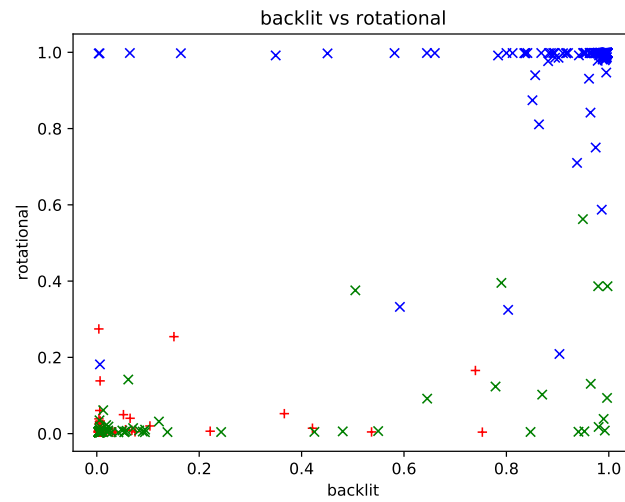
As expected there are clusters at (0,0) and (1,1), etch pits which look 'signal-like' or 'background-like' under any illumination. Most outliers appear at the edges, and few towards the centre (0.5,0.5). This indicates that ambiguous or misclassified examples are uncorrelated between the different classifiers in different data channels. Etch pits that are ambiguous or misclassified under one set of illumination conditions may be accurately classified using another. This is a strong argument for using multiple illumination conditions as they are capable of complementing each-other. False negative etch pits rejected by one classifier may be recovered through analysis with another in a different data channel.

In Fig. 7.18a Pb etch pits are most readily identified by the dark field classifier with a long tail of outliers present under backlighting. The reverse is true of the background etch pits. This suggests signal acceptance and background rejection are better optimised by dark-field and backlit illumination respectively. Backlighting is better at recovering fragment examples. Fig. 7.18b shows a similar effect with the rotational illumination demonstrating sharper signal resolution with less capacity for recovering fragments. Tails are present for both classifiers. Lastly Fig.7.18c demonstrate the best 2D separation of signal and background, a simple average (diagonal decision boundary) would yield complete separation.

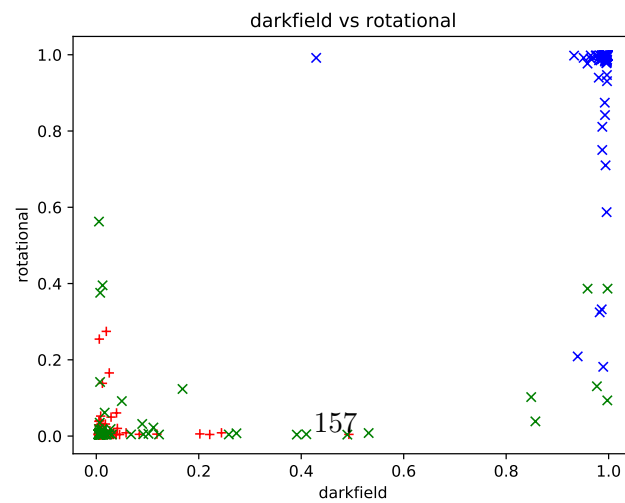
How best to utilise and combine multiple results from different classifiers is a different question. A simple approach could be to average or combine the scores, or a geometric approach might define a region of interest in higher dimensional classifier space, such as a sphere around (1,1,1), or a background exclusion around (0,0,0). In the context of the MoEDAL experiment we are not limited by statistical reconstruction, and could eventually examine all high scoring etch pits, prioritising those that score highly in all data channels first.



(a) Backlit vs Darkfield classification scores



(b) Backlit vs Rotational



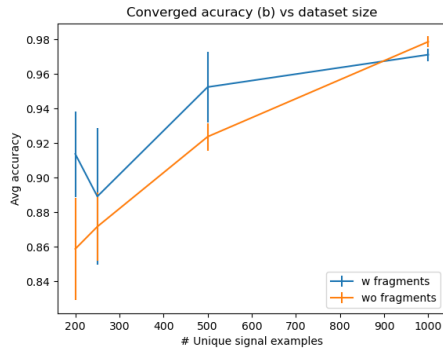
(c) Darkfield vs Rotational

Figure 7.18: Mean classification scores in a given data channel for each etch pit. Pb Sig (blue), Bkg (red) and Fragments (green).

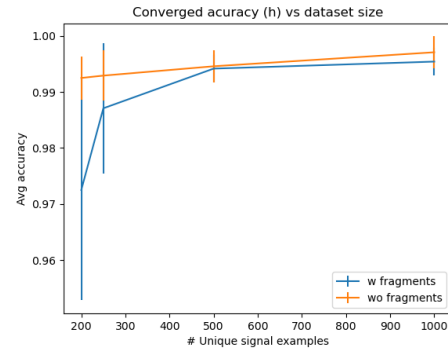
Effect of data quantity

These figures show the stable performance achieved by the CNN classifiers after training with 1000, 500, 250, and 200 training examples respectively. Shown are the cases with fragments included and excluded from the training data. Increasing the quantity of training data yields improved performance, although with diminishing returns. Larger quantities of training data improve the consistency of the results, with a smaller spread between folds and repetitions. 1000 unique signal examples (which isn't many compared to many machine learning tasks) provides enough data for accurate classification with additional data unlikely to yield massive improvement. Robust classifiers may be trained on the dark-field and rotational illumination channels with even fewer signal examples, especially when fragments are excluded, delivering 98% performance. This suggests the HIP pattern in these channels is very distinct and easy to learn, as shown in Figure 7.20 c) and d). Even with very low numbers of training examples (eg, $N = 50$) the networks can still produce useful yet reliable results. This is despite each network having up to 50,000 parameters.

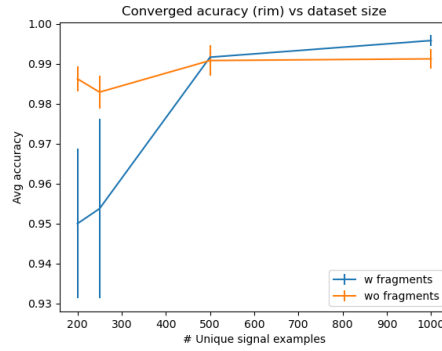
Figure 7.20 highlights the loss convergence of the models during training, and illustrates what causes a breakdown in performance at low data quantities. The backlit classifiers (a) appear to start over-training after 15 epochs (with $N=200$, or 160 examples per training fold). Detailed inspection of the folds reveal that this is only the case for a few models on some data folds such as fold 2 (b) and fold 3, distorting the overall average (repetition shows that this is not limited to just these data-folds). Rather than an even degradation of performance with less data, most models still produce 'good' results of $> 90\%$, however the likelihood of a particular model over-training or even failing to train becomes significantly higher. There is more variety of appearance among backlit Pb etch pit images, making it harder to cover all possibilities. Backlit trained without fragments ($N=200$) was the worst performing set of models on the validation data (Figure 7.19), yet showed no issues evaluated on the test folds.



(a) Backlit



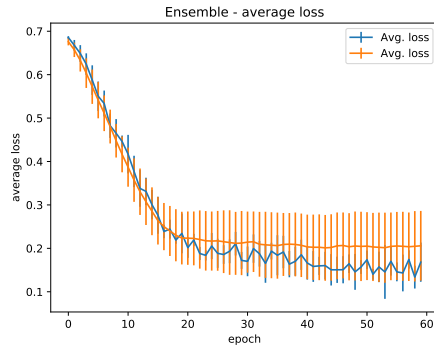
(b) Dark-field



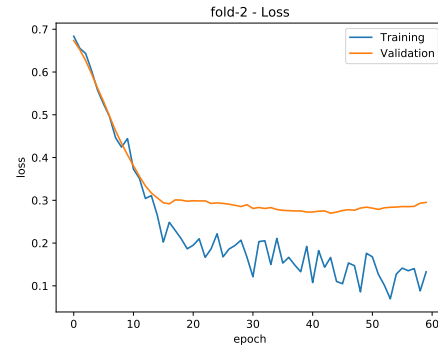
(c) Rotational

Figure 7.19: Converged performance vs unique signal examples

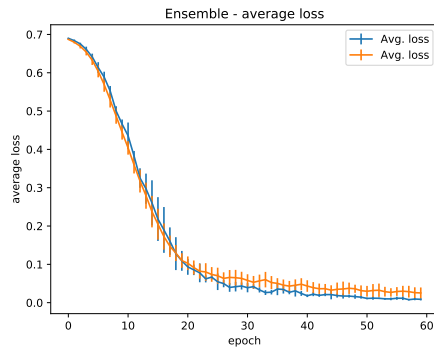
The data points represent the average validation accuracy and standard deviation achieved for the five K-fold models at the end of the training run. Classifiers trained on the dark-field and backlit channels produce good performance with fewer data examples, suggesting the key features of a HIP candidate are more visible under these illumination conditions. This is especially pronounced when fragments are excluded from the training data.



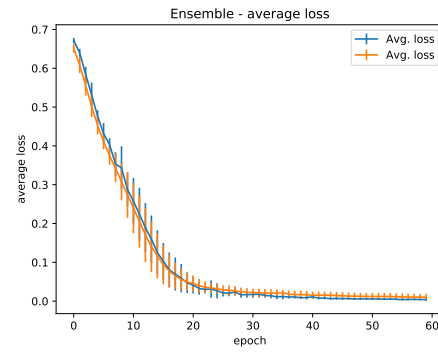
(a) Backlit N=200



(b) Backlit N=200 fold-2 model only



(c) Darkfield N=200



(d) Rotational N=200

Figure 7.20: Loss convergence with reduced data quantities

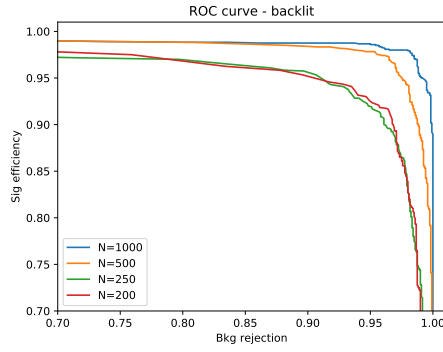
Loss metric during training is shown for Backlit with fragments a), and Darkfield c) and Rotational d) without fragments trained using 200 signal examples. a) b) c) show as an average and standard deviation for all 5 models in the ensemble. b) shows the loss evolution for a single model in the backlit channel, forming an outlier to the average shown in a), with such low data quantities outliers such as this become far more frequent.

ROC curves

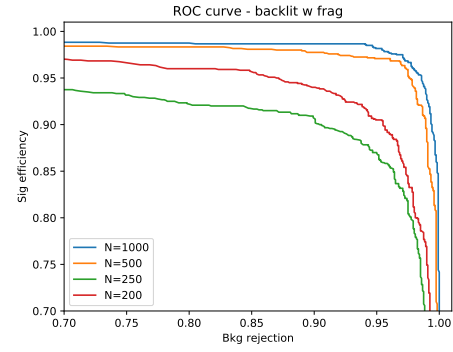
The ROC curve highlights the potential signal acceptance or background rejection rate possible with different levels of threshold or working-point, typically representing a trade off between the two. Results are amalgamated for all 5 models on each K-folded dataset. Figure 7.21 shows the respective curves for each set of models, for each channel, and size of dataset. A sharp rectangular shape can be observed for the dark-field (c), and rotational illumination (e), also true for backlit with large values of N . There are asymptotic regions of the curve that maximise either signal efficiency or background rejection with little trade-off to the other, and a relatively small 'curved' region where performance may be fine-tuned. This is what we observed in Figure 7.16. Adjusting the working point within this region away from a 0.5 classification threshold does little to impact overall results. Instead of trying to optimise within this region I would instead advocate adopting two working points optimising signal efficiency and background rejection respectively. This would divide etch-pits into three regions;

- High signal purity. Rejecting early all background, while keeping most signal.
- Low signal purity. Mostly containing background under real world conditions
- Background rejection. A region almost entirely of background, where any signal is likely irrecoverable.

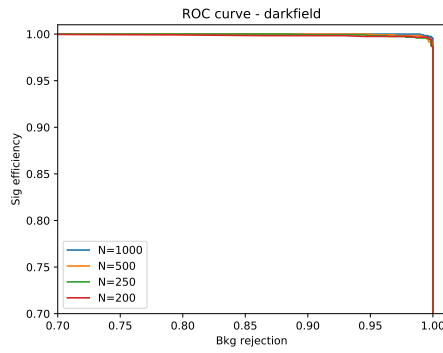
As MoEDAL relies on a direct reconstruction of a monopole or exotic track, as opposed to an indirect reconstruction, we aren't statistically limited. We aren't required to 'throw away' any etch pits. A natural strategy would be to prioritise high purity etch pits first, given the very high levels of background rejection possible. A subsequent focus on low signal purity etch pits would cover the systematic risk of genuine exotic HIPs scoring poorly on classifiers trained using REL matched heavy ions. Table 7.10 illustrates several possible working points, including high and low examples.



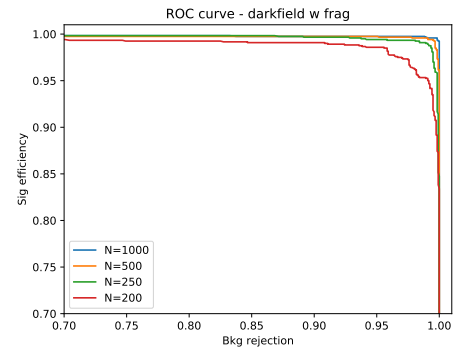
(a) Backlit



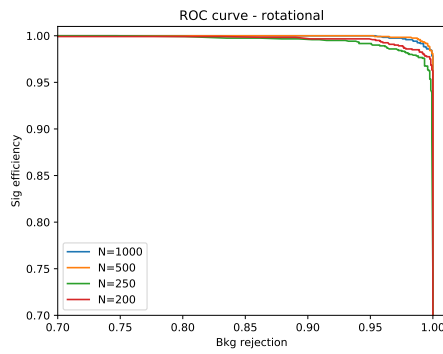
(b) Backlit w fragments



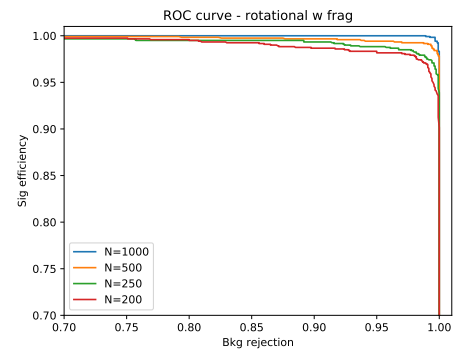
(c) Darkfield



(d) Darkfield w fragments



(e) Rotational



(f) Rotational w fragments

Figure 7.21: ROC curves

Working point	0.1	0.5	0.9
Bkg. rejection	98.7	99.8	100
Pb sig. acceptance	99.4	98.4	96.3
Pb frag. acceptance	6.02	2.14	0.19

Table 7.10: Example working points for Rotational classifiers (N=1000), no fragments in training.

CNN classifiers: Average ensembles

As established earlier in this section the misclassification errors of the CNNs acting on different data channels are largely uncorrelated. Classification performance can be vastly improved upon by combining the results of different classifiers. Here I will outline two possible static committee ensembles of classification scores; A geometric average of the mean classification scores for each data channel, and a numerical average. A third option would be using a secondary neural network to combine the scores. However since the first two simple methods are already capable of achieving 100% background rejection on the validation data this has not been attempted.

$$S_{geo} = S_{backlit} \times S_{darkfield} \times S_{rotational} \quad (7.1)$$

$$S_{avg} = \frac{1}{3} (S_{backlit} + S_{darkfield} + S_{rotational}) \quad (7.2)$$

The corresponding signal and background histograms for these combinations, and corresponding ROC curves are shown in Figure 7.22. As many background etch pits score very low ($\sim 0.01\%$) on most classifiers the geometric combination of three different classifiers produces an extremely low score for most background objects (technically 15, as the score for each channel is the mean of all 5 folds). The geometric ensemble can thus produce very high levels of either signal purity, or background purity, the latter being safe to reject. Scoring etch pits based on the numerical average over all the classifiers produces superior separation of signal and background, as erroneous results are smoothed

out, in fact the signal and background distributions are entirely separate.

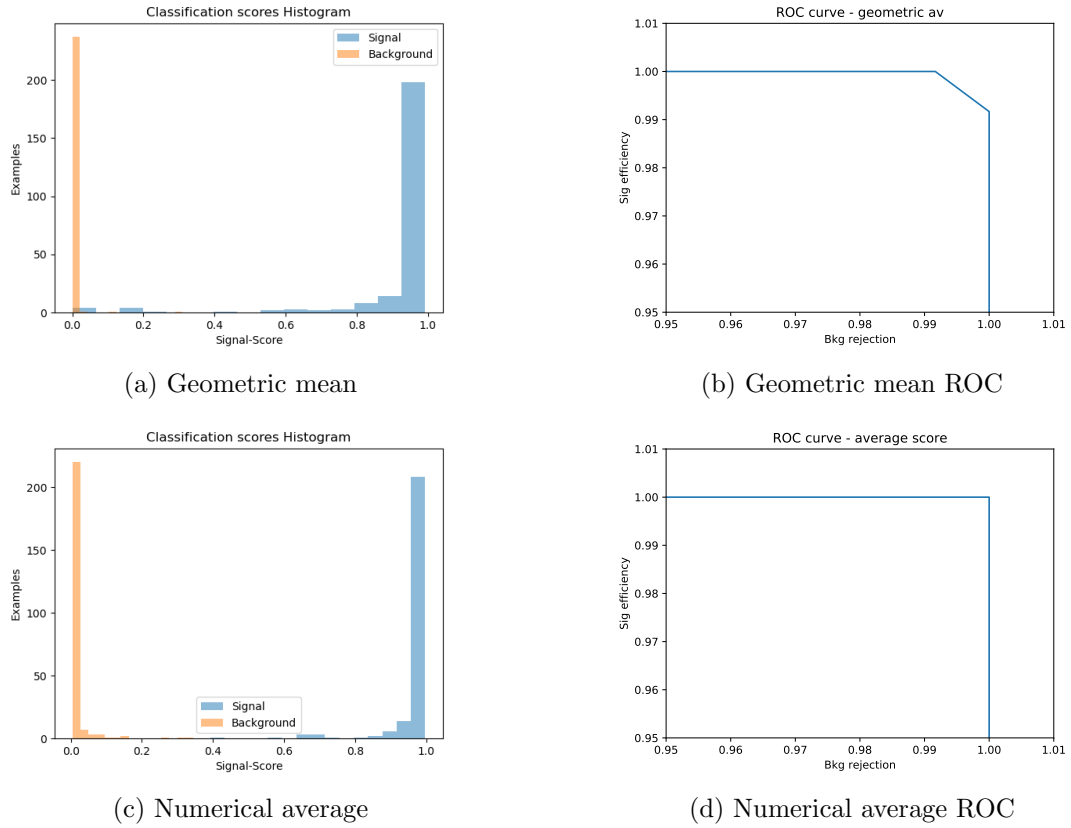


Figure 7.22: Histograms and ROC curves for the two channel combinations. The background peak in a) had to be widened by a factor of 100 to appear visible.

Table 7.11 shows the levels of background rejection (false positive rate), and signal acceptance possible with different working points for the combined ensembles. In both cases it is easy to achieve 100% background rejection whilst maintaining signal acceptance near or above 99%. For the numerical average ensemble it is even possible to find a working point that achieves perfect separation of signal and background. Obviously these results are close to the margin of error of our validation dataset.

Combination	False positive rate	False negative rate
Geometric Avg. (wp = 0.05)	0.0	1.20 (3)
Geometric Avg. (wp = 0.1)	0.0	1.61 (4)
Geometric Avg. (wp = 0.5)	0.0	4.02 (10)
Numerical Avg (wp = 0.1).	2.40 (6)	0.0
Numerical Avg (wp = 0.3).	0.80 (2)	0.0
Numerical Avg (wp = 0.35).	0.0	0.0
Numerical Avg (wp = 0.5).	0.0	0.4 (1)

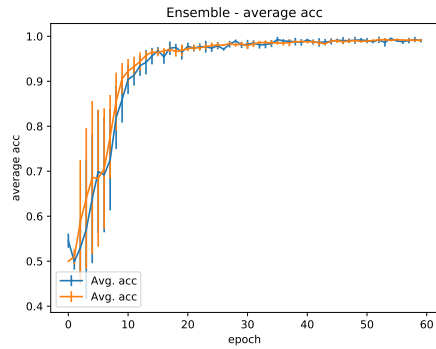
Table 7.11: False positive, False negative rates % shown for different choices of ensemble combination and working point (wp). For the ensemble produced by a numerical average of classification scores a working point of 0.35 allowed for perfect separation of signal and background, choices below this value simply allowed more false positives, while setting a higher working point simply cause false negatives to occur for real etch pits

Data augmentation

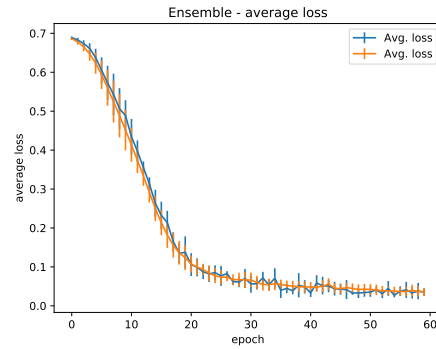
Data augmentation was investigated as a means to increase the effective data quantity (§4.4.3). Etch pits within each fold were duplicated for every 90 degree rotation (for the rotational imaging, channels were permuted to maintain relative angle of illumination) increasing N_{eff} by a factor 4. Surprisingly neither the final converged accuracy or loss improved across any channel as a result of the additional data. The additional effective data didn't improve the time taken for the networks to train either. The impact of data augmentation on the dark-field channel is shown in Figure 7.23 similar results were obtained for all other channels.

We should expect the increase in the effective data to improve performance. Data augmentation works under the assumption that the classification is symmetric under the augmentation being applied. One conclusion is that this assumption is not valid. This would imply that the 3D angle of incidence of the test beam forms a substantial element of signal etch pit classification.

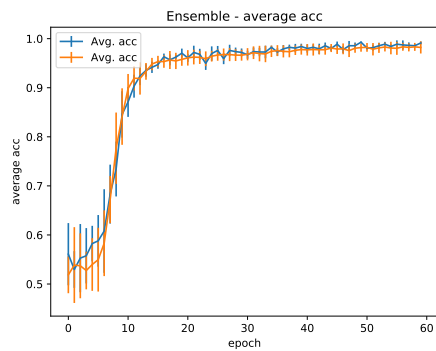
This implies classification performance will be biased depending on the orientation of the hypothesis particle. If we want to factor out angle of incidence, for an explicitly angle of incidence agnostic classifier then we should train using rotational data augmentation



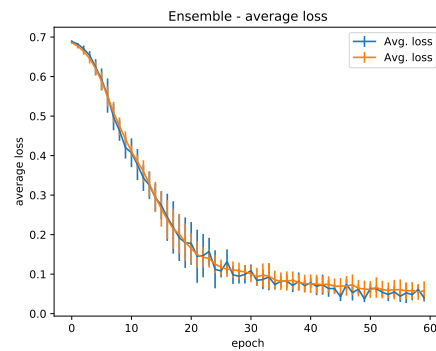
(a) Avg. Accuracy, no data augmentation



(b) Avg. Loss, no data augmentation



(c) Avg. Accuracy, with data augmentation



(d) Avg. Loss, with data augmentation

Figure 7.23: Dark-field classification without and with data augmentation

There is no performance improvement resulting from rotational data augmentation

to remove this dependency .

Angle of incidence detector

The earlier result inspired an effort to produce an angle of incidence detector trained using data augmentation. Existing Pb etch pit data was duplicated by rotating it by a 90° angle an integer number of times. The ordering of the rotational illumination channels was permuted so that initial illumination angle is unchanged. This replicates the Pb beam coming in from four different angles. The CNN classifier was trained to detect the rotation integer. The only feature breaking the rotational symmetry of the foil is the angle of incidence of the ion beam (along with a weaker IP effect for the LHC

background), thus this acts to guess the angle of incidence of the Pb ion.

This was done for Pb etch pits in both the exposed and unexposed foils. In the unexposed foils it was trivial for the network to accurately reconstruct the incidence angle ϕ of the Pb test beam. In LHC exposed foils angle of incidence detection was much less effective. It was only possible predict the angle of incidence with a 90 deg margin of error either side. This suggests a lot of directional information is lost in the presence of LHC background pits. Being able to detect the angle of incidence from an etch pit alone can still significantly cut down on false positive track candidates. Every etch pit belonging to a candidate track (whether entry or exit) would be required to align. Even with angle of incidence detection limited to a 180deg arc this would reduce the number of false positives by a factor of 2 for each additional candidate etch pit along the candidate track.

7.3.4 Summary and Conclusions

The conclusions of the Pb CNN classification study are intended to apply more generally to CNN based MoEDAL NTD analysis with the full LHC background, where the potential signal will be much weaker. The results are summarised as follows;

The fundamental conclusion is that overall the performance of CNN based etch pit classifiers is good. CNNs are capable of accurate ($> 97\%$) classification performance even with restricted quantities of training data, with optimisation being a robust and repeatable procedure. Typically significant factors such as architecture, and hyper-parameters have little impact on the final results.

Performance is relatively binary, when degraded due to a lack of data (eg, Backlit $N=200$), the transition to poor performance is rapid and unpredictable. I hypothesise that the main determinant of MoEDAL CNN performance is the presence of detectable signal features in the data to hand ('key features'). When these are available CNNs will perform well regardless of architecture or training method, even with very low quantities

of data.

The 'key features' are most visible under darkfield and rotational illumination. The corresponding classifiers are more performant than the backlit CNNs by a factor of 2.5, achieving near 99% accuracy on Pb ions, but are nearly unable to detect smaller Pb fragments, suggesting those key features are lost against the LHC background. The key features of various signal and background categories are uncorrelated between data channels as observed in 7.18, which is a very strong argument for using at least two different illumination methods for MoEDAL data collection. The inability to recover fragments suggests a potential systematic error when trying to identify HIPS outside of the trained REL range.

There is minimal capacity to adjust or improve classifier performance by changing the classification threshold. Most classifications are simply unambiguous consistent with the key feature hypothesis. This means we would have little scope to adjust the sensitivity of our candidate search. Where performance can be improved is via combination with other classifiers. Actual NTD analysis involves a miniscule signal, and an overwhelmingly large background. Combining together multiple accurate and robust classifiers over different data channels allows for extremely high classification accuracy.

Data augmentation did not improve performance. It did however reveal that the hypothesised angle of incidence biases classification performance. Future MoEDAL research should be aware of this, and should take steps to account for it. For example, image data from different parts of the NTD array should be rotated (eg, in 90deg increments) so that the hypothetical angle of incidence ϕ of a particle from the interaction point most closely matches the angle of incidence of the ion beam used to produce the ML training data of any ML classifier being used. Alternately data augmentation (rotational) could explicitly remove this bias by including 'augmented' data representing all angle of incidence in the training data. In the latter case a secondary angle of incidence predictor could be used to constrain potential candidate tracks.

7.4 Fully Convolutional Neural Network

In this section I approach our ultimate goal: Identifying predicted etch-pit locations from a set of large area NTD foil images, which have too many etch-pits to classify discretely. There are multiple methods that could be used;

The Xenon Ion study used a fixed convolutional kernel, utilising Prior knowledge of etch-pit structure. The output was filtered and clustered to identify a selection of discrete points within an image that best matched the known pattern. This produced an unacceptable number of possible matches in regions of higher background. At low background densities this method efficiently pruned the data, and turned a regression problem into a simpler classification problem. However any examples missed by the preselection kernel were effectively lost.

Another method would be to divide a large area image into many small images (or patches) with some overlap. A CNN classifier allows for a significant degree of translational symmetry, and could likely duplicate success in identifying HIP etch pits in 28×28 images into similar performance with 100×100 pixel images. This would be a weak learning task, identifying the presence of a weak or subtle anomaly.

A type of Convolutional neural network architecture particularly suited to wide area image analysis are fully convolutional neural networks (FCNS). FCNs preserve spatial information in their output rather than the common approach of compressing down to a flat output. The output of a Fully convolutional neural network (FCN) is similar to a convolutional kernel filter, turning an image type input to an image like output.

Parallel MoEDAL ML research undertaken by Kanik Palodhi, Joydeep Chatterjee, and Attanu Malkik, uses a deconvolutional-convolutional method (DCC) using the R&D stack NTD image data I gathered for this chapter. Research carried out by Aman Upadhyay, supervised by myself and my own supervisor Prof. Jon Hays, uses the aligned

NTD stack data to build and train a U-Net FCN architecture from scratch. LHC exposed foil images and unexposed foil images are used as the paired input and target respectively.

This thesis focuses on using transfer learning to construct an FCN by adapting the existing convolutional layers of the CNN classifiers trained in §7.3. This has the advantage that they have already been trained and validated using the etch pit classification dataset. The following sections will describe the construction of two FCN models;

- A simple conversion of a CNN to an FCN, without any additional training of the network. This works because of minor changes to the design of the CNN network architecture (discussed in §7.3).
- An FCN constructed using the convolution layers from multiple CNN classifiers, with additional layers combining them into a single output. This network is then retrained with the CNN convolution layers held fixed. This essentially functions as an ensemble of classifiers with an FCN output.

I will also discuss how discrete etch pit predictions may be obtained from the FCN output. Qualitative observations will be discussed in this section. The evaluation method I describe in the next chapter is used to perform quantitative performance measurements (§7.5) of the FCN algorithm. A paper in preparation uses this method to analyse and compare the results of the three algorithms (DCC,U-net FCN, and Transfer FCN)

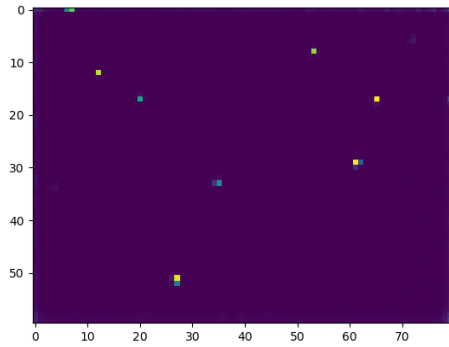
7.4.1 Simple FCN

The CNN classifiers used in the previous section (Tab.7.3) can be converted into FCNs with minimal modification. Removing the global max pooling operation allows the learned CNN layer weights to be reused as an FCN (this retroactively motivated the change in CNN architecture). The output of the CNN is a classification score or likelihood of the $28 \times 28 \times C$ input image containing a Pb etch pit. As the max pooling operation simply selects the highest value of its inputs we can summarise that the (4,4) output

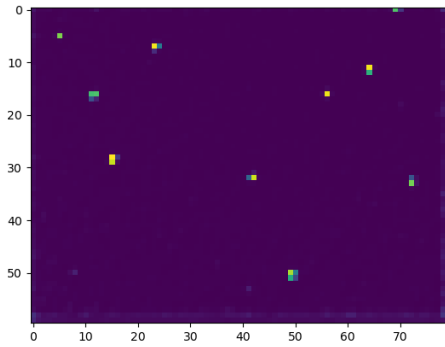
of the preceding layer is proportional to the likelihood of each section of the (28,28) input containing a Pb like etch pit. The output of the FCN is a heatmap with hotspots indicating potential Pb etch-pits, An example of the output is visible in Fig.7.24. As the FCN network still contains three pooling layers (thus technically not fully convolutional) the FCN output compresses each 8×8 pixel region of the input image to a single pixel corresponding to the approximate 8 pixel size of an etch pit. A full size $640 \times 480 \times C$ image results in an 80×60 output. No additional training is required to generate an FCN via this method. The original CNN can be trained and validated very quickly and robustly on the significantly smaller CNN training dataset, which is a factor of 1000 times smaller than the equivalent full dataset.

Heatmap analysis Simple FCNs and their heat-map output have been generated for each classifier used in §7.3. Comparing heat-maps between both data channels and folds shows that they differ substantially in the background floor, along with the peak intensity and width of Pb etch-pits. While the channel to channel differences are expected the difference between folds are not, within this variation some networks are more significant outliers.

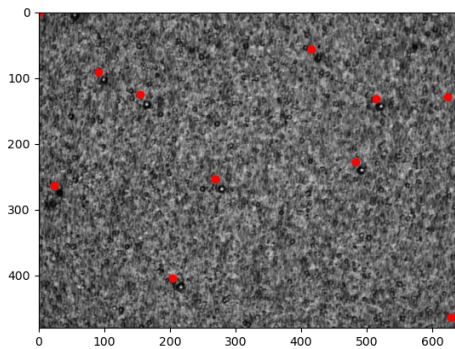
This is an interesting result as each of the original CNN networks delivered consistent classification performance. This suggests that while they converged on the same classification decisions they likely focused on different features of the input images. Looking at the spectra of heat-map outputs (Fig.7.25) we can observe a signal large spike at 1.0. Unlike the CNN classifiers many predictions fall in the range $[0.0, 1.0]$, meaning that there is a large scope to fine-tune the working point used to select final predictions. However the variation between the outputs of the different models mean that a different working point would likely be needed for each.



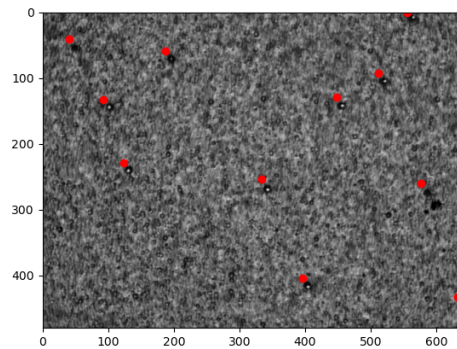
(a) FCN heatmap



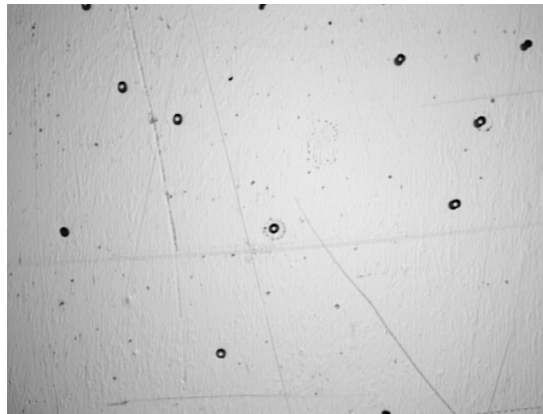
(b) FCN heatmap (reverse)



(c) NTD foil, with heatmap predictions



(d) NTD foil, with predictions (reverse)



(e) Corresponding region of clean foil

Figure 7.24: Simple FCN heatmap, for both sides of the NTD foil
 Images (a) and (b) show the heatmap output of the simple FCN for the rotational data channels. Pixel intensity indicates the likelihood of finding a Pb etch pit. (c) and (d) show the exposed NTD foils with the predictions overlaid. There is a (14,14) pixel displacement from the kernel. Some predicted etch pits near the threshold are nearly invisible in the heatmap. An interesting example is the top left pit in (b), it is a fragment. It is detected on the reverse side, but doesn't meet the threshold on the front side.

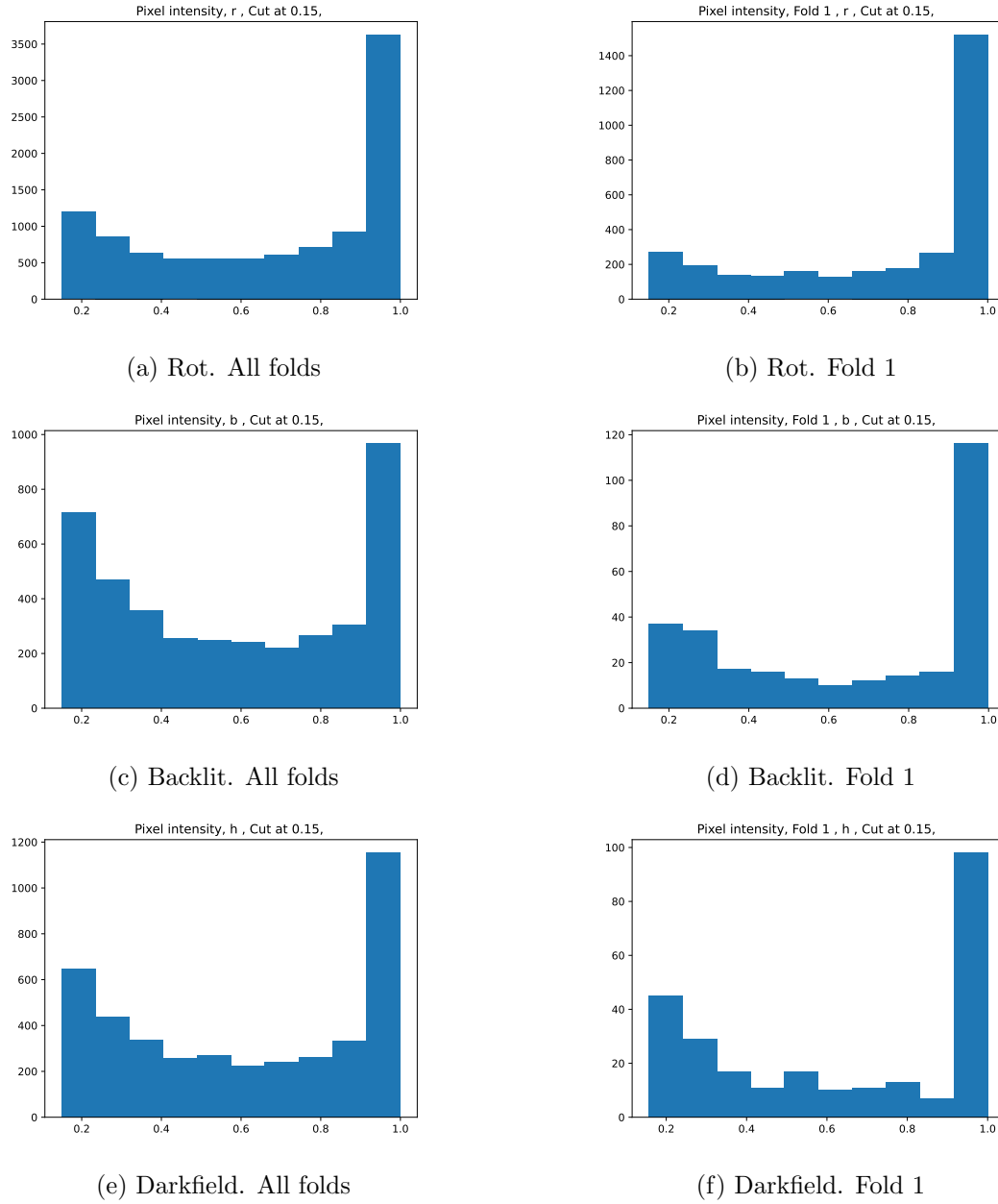


Figure 7.25: Heatmap pixel spectrum

The pixel spectrum of the FCN outputs, shown for the rotational, backlit, and dark-field models without fragments. Fold 1 results show a large divergence. A threshold has been applied at 0.15 to remove the background peak.

Etch pit predictions Predicted etch-pit locations are obtained by applying a threshold cut to the heat-map at the desired working point, obtaining an image segmentation map from which the etch pit locations are derived, as done in §6.3.2 and §7.3.1. The choice of threshold can be used to tune the sensitivity of the FCN.

There is a consistent alignment shift of (14,14) pixels. This is due to the compressed representation, and the prediction being marked at the top corner of the convolution kernel. There is the potential for systematic bias due to edge effects near the boundaries of the image since the overall convolution kernel of the FCN is larger than the size of an etch-pit. These issues can be seen in Fig.7.26. While problematic this doesn't warrant concern as these effects can be trivially mitigated in production by either stitching images together, or ensuring suitable overlap between images.

Observations Some CNN networks are able to yield fairly accurate results as FCNs, this will be covered in 7.5, generally having a low false positive rate, albeit with a high false negative rate. The training run variability of the CNN derived FCNs render them unsuitable for MoEDAL image analysis due to the inconsistency in their output impacting repeatability. An advantage they have is that their graduated output makes them easier to tune. As the CNNs are trained simply on whether the image contains an etch pit this method produces a poor spatial resolution. Computationally these simple FCNS are trivially quick to implement and run. Due to the spatial compression their output is smaller by a factor of $8^2 = 64$ for the 2D backlit images, and $8^3 = 512$ for the 3D rotational images.

7.4.2 Retrained FCN

The previous simple FCN architecture involved no additional network training. By adding an additional set of network layers forming a new output head (Table 7.12) it is possible to combine the convolutional layers of the various CNN classifiers to achieve a

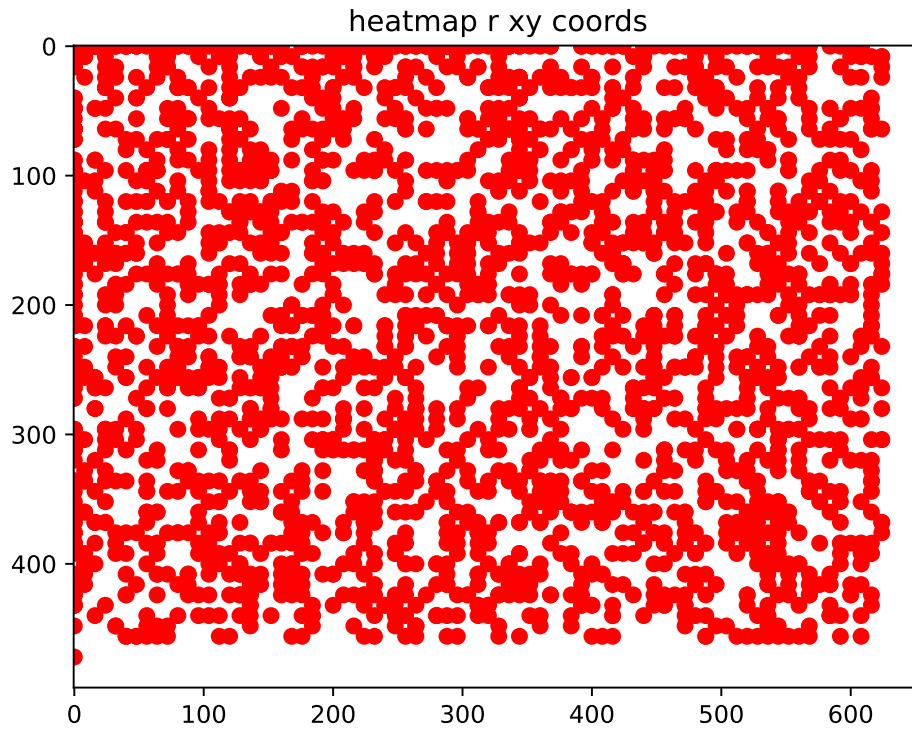


Figure 7.26: Predicted X,Y locations superimposed

We can observe an excess of predictions at the leading edges of the images, and an absence at the trailing edges. The former can be removed by eliminating predictions touching either $x = 0$ or $y = 0$, the latter is corrected by shifting the predicted coordinates to the centre of the kernel.

desired output. While additional training is required to optimise the new layer weights, transferring the trained weights from the existing models means the amount of training required is minimal.

FCNs were constructed combining all 5 CNNs across the rotational, backlit, and dark-field channels. The final FCN model used in the evaluation section of this thesis, combines all 15 CNNs in all three data channels.

The final FCN output head is trained using a set of target images produced from the known etch pit coordinates, within the training set. In these images each pixel has a

Layer	Feature depth	Activation
Concatenated pre-trained Conv layers		
Dense	10	ReLU
Conv (2x2)	30	ReLU
Dense	20	ReLU
Dense	1	ReLU

Table 7.12: FCN output head

An additional output head containing approx. 2000 parameters which can mix and combine any number of pre-trained convolution layers.

value of 1 if it contains a known etch pit and 0 otherwise. The FCN was trained over 20 epochs of the entire dataset. Examples of the input, target, and output are shown in Fig.7.27 and Fig.7.28.

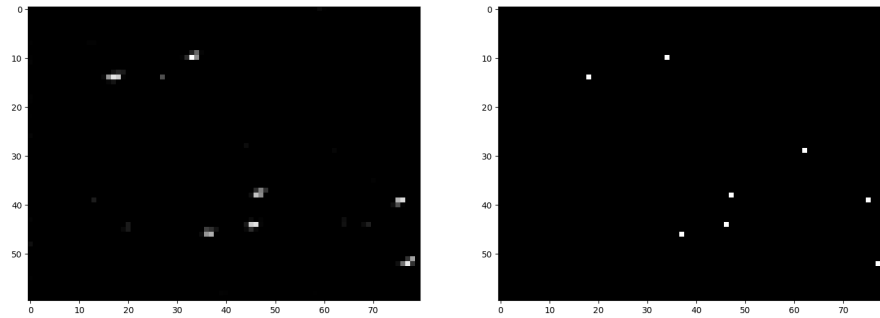
7.4.3 FCN final predictions

The retrained FCNs all take NTD foil images as input, and output predictive heatmaps which reflect the 'signal-likeness' or 'interestingness' of various regions of foil, with pixel intensity representing the likelihood of finding a HIP candidate at that area. The overall likelihood distribution of the final FCN can be seen in Fig.7.30. Applying a threshold cut to the heatmaps reduces them to a discrete set of predicted locations.

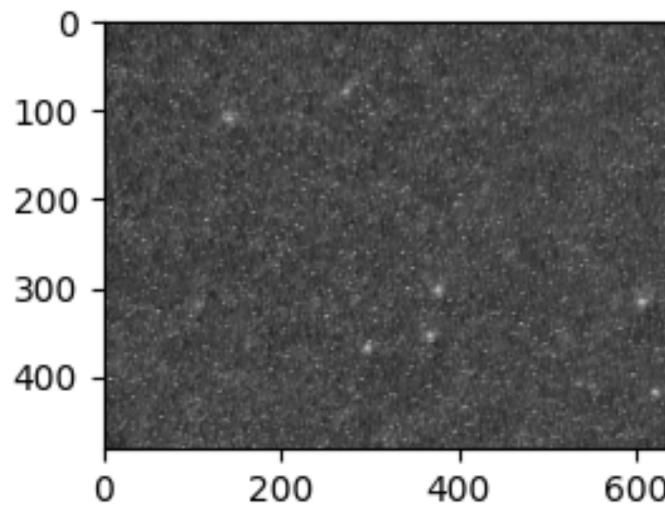
7.4.4 Observations

An advantage of the transfer learning approach to the FCN is that while the final model can combine hundreds of thousands of parameters, only a few thousand have to be trained. The FCN network can also be decomposed into a small nested FCN (equivalent to the output head) acting on the combined outputs of the pre-learnt CNN derived FCNs. While mathematically identical in theory, in practice this is far more computationally efficient in terms of time and memory, as the nested FCN can be run on the stored output of the simple CNN derived FCNs (which in turn can be generated independently of each-other). This would make it trivial to use in an embedded application.

Slide 27 fcnn vs target



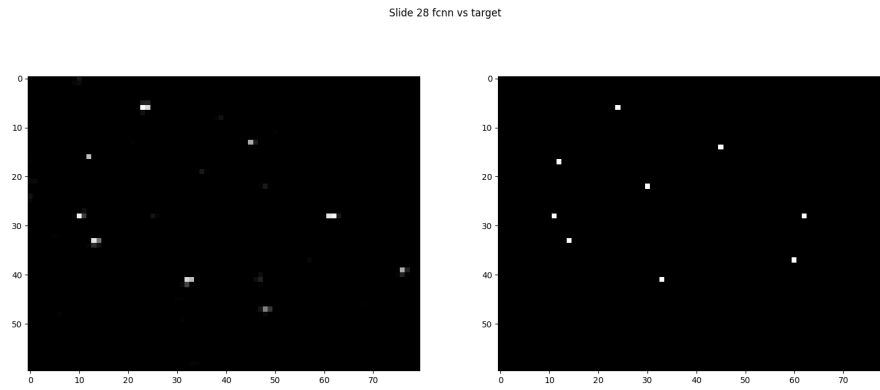
(a) FCN output and training target



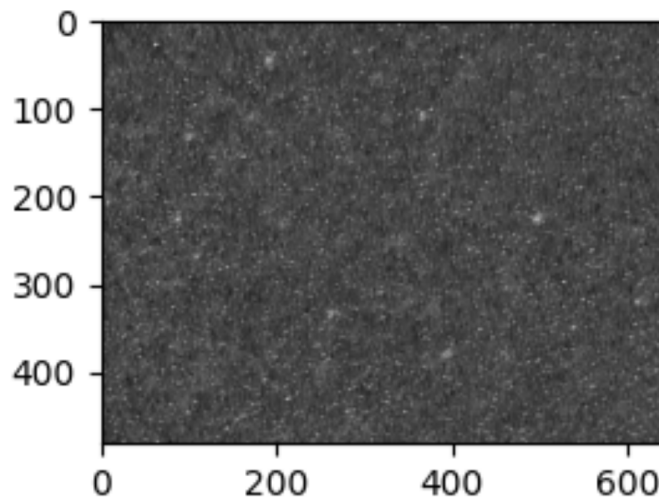
(b) Rotational channel

Figure 7.27

The FCN output is shown along with the target input, and one channel of the input image data. The probability around each etchpit is more distributed than in the heatmaps produced by a single model.



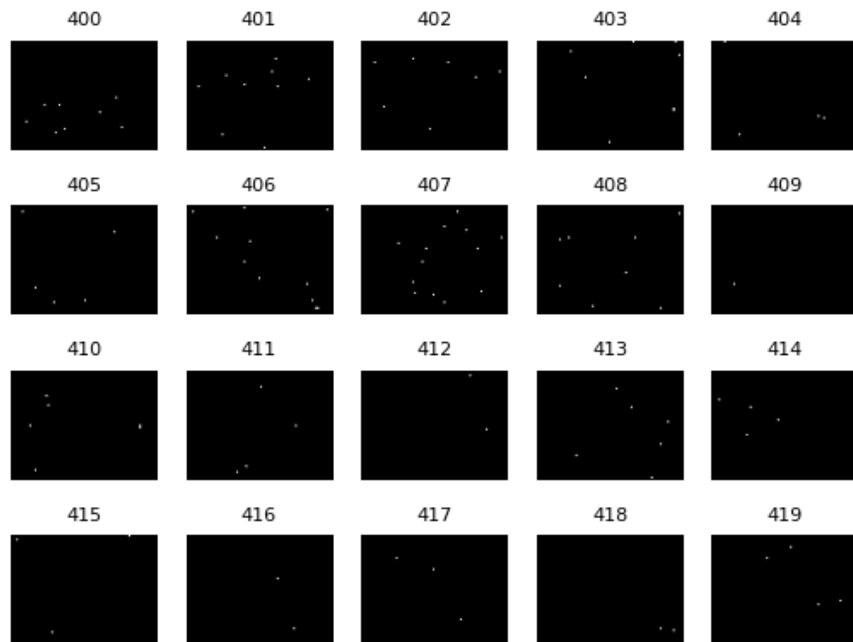
(a) FCN output and training target



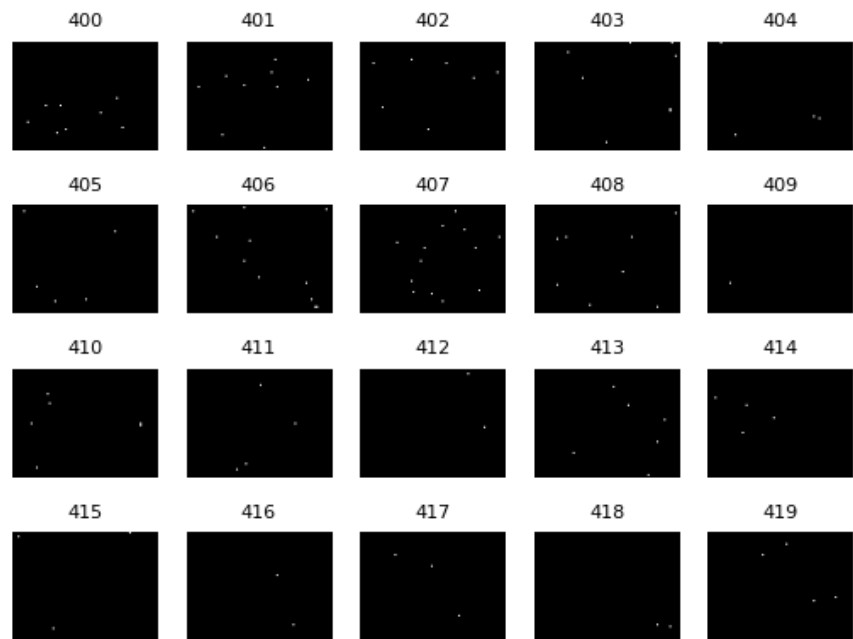
(b) Rotational channel

Figure 7.28

The FCN output is shown along with the target input, and one channel of the input image data.



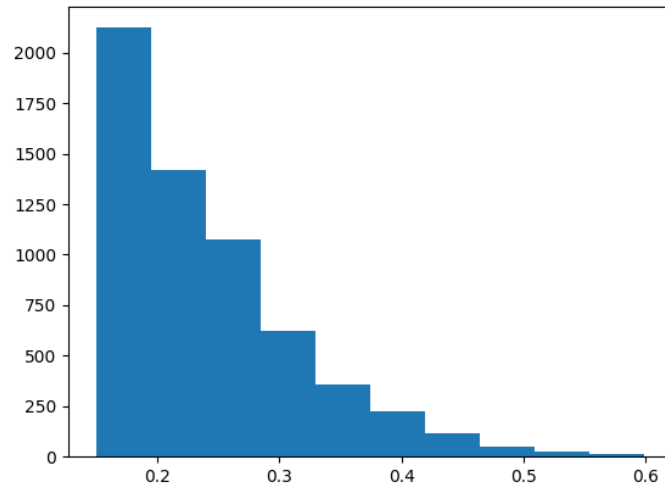
(a) FCN outputs



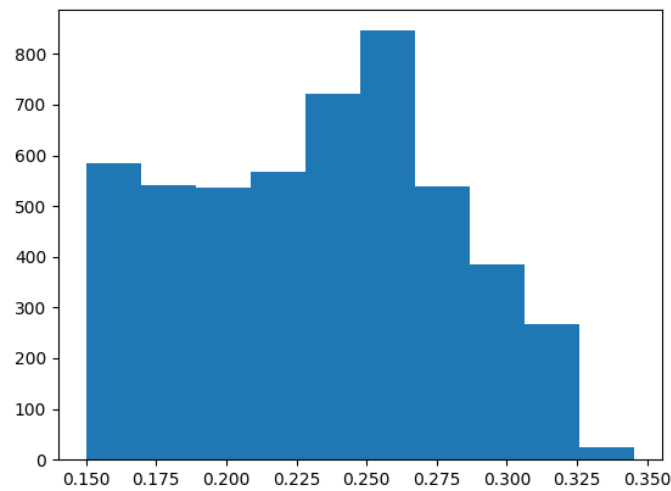
(b) 'Simple' FCN outputs

Figure 7.29: FCN vs 'Simple' FCN

The retrained FCN can be used to alter output into a desired a desired format, this makes it useful for combining or comparing data channels. Here there is little practical difference in the outputs (and no difference in the output format) between the Simple FCN and the retrained FCN, except the latter could be readily retrained to include additional data channels or networks.



(a) 'All Channel' FCN spectrum



(b) 'Rot.' FCN spectrum

Figure 7.30: All channel, vs Rotational channel FCN output spectrum

The different FCN output spectra between the All Channel, and Rotational Channel FCN. Similar to the simple FCN the sensitivity may be tuned through choice of working point. Unlike the simple FCN there is no sharp signal like peak, as the likelihood around each etchpit is more spread out. This means the signal sensitivity of the combined channel FCN can be adjusted and fine-tuned more easily

7.5 Evaluation

This section provides quantitative performance measurements of the final combined FCN, and a simple FCN derived from a rotational channel CNN. This is used to estimate the HIP signal acceptance and the false positive rate of an FCN based NTD analysis.

7.5.1 Evaluation dataset

The evaluation dataset was prepared using the blinded 'evaluation' half of the NTD foil. This region of foil and image data taken from it have not been used in any part of the investigations in §7.3 or §7.4. All training and threshold optimisation on the FCNs were performed using the training half of the foil.

Etch-pits within the evaluation region were identified using the same algorithm used to prepare the CNN training dataset in §7.3.1. Potential Pb ion tracks were categorised by quality, graded according to how many etch-pits are detected in subsequent NTD foils. Scratches in the foil, and possible optical anomalies are removed using a 2D CNN that was trained for this purpose using data from the training set. A similar approach was used to identify Pb ions that have fragmented. These fragmentations represent a possible source of systematic error as we don't know how clearly they may or may not appear in the unexposed foil.

The evaluation region is defined as the following;

- Images 251-500 of the LHC-exposed NTD images. Each image consists of backlit, dark-field, and rotational darkfield illumination channels (10 total). These are used to generate predicted etch-pit locations using the algorithm.
- Images 251-500 of the unexposed NTD images. Each image consists solely of the backlit image channel. These are used to identify the location of the real Pb ion tracks which form a truth record to compare against the algorithm's predictions.

The evaluation dataset is taken to be the mutual overlap of LHC-exposed images 251-500 and unexposed images 251-500, within global coordinates. Because the foil scans do not perfectly align, the following regions are excluded from the evaluation dataset:

- Any region within the exposed (unexposed) NTD images that does not overlap with a corresponding region in the unexposed (exposed) NTD images. It is not possible to verify the identity of Etch-pits within this region. To achieve this a rectangular boundary is applied in global coordinates.
- Any part of the unexposed evaluation NTD images where the aligned location in the exposed NTD foil appears in the first 250 images which were used for training purposes. This eliminates all potential pits where the unexposed location occurs in images (251-500) and the exposed location occurs in (1-250), this is to eliminate potential human error, and to avoid over-counting

Additionally the following areas are excluded;

- The area surrounding the alignment pinholes are excluded due to the abnormal geometry around the edge of the hole which does not occur in the bulk of the NTD foil.
- Areas of damaged foil or anomaly. Given such areas occur even in the unexposed NTD foils it is natural to expect them in the exposed NTD foil. Such regions are identifiable by visual inspection. Or via anomalies present in global coordinate view.
- Areas in the margin of each exposed image, equivalent to 14 pixels from the edge, removing possible edge effects. This corresponds to 28x28 pixel image size used for CNN training, and also excludes the possibility of double counting etch-pits where images overlap (Edge effects are not an issue of concern in this experiment, as the

imaging size of 640x480 is a limitation of the equipment used in this R&D study which does not apply to analysis scanning hardware).

These exclusion regions are applied in practice by filtering out etch pit coordinate locations that fail these exclusion criteria. This reduces the number of evaluation truth pits from 1277 to 1194.

7.5.2 Evaluation methodology

Using the global coordinates of the foil stack, and the known offset due to the angle of incidence of the ion beam, it is possible to compare the predicted etch-pit locations to the known etch-pit locations.

Inference Predictions Predicted Etch-pit locations were generated by running the final FCN algorithm (§7.4) on scan data from both sides of the exposed NTD foil. This produces a set of FCN output images for each side. A threshold value of 0.08 was applied to separate signal and background regions and obtain a set of discrete predicted locations. The same method was used to obtain a set of predictions for the simple FCN, based on a rotational channel CNN, with the exception of a 0.5 threshold being used, these results are included as a comparison.

Etch-pit comparison True Pb ion etch pit locations (derived from the unexposed foils) and the Predicted etch pit locations (from the LHC exposed foil) are both converted into the global coordinate system introduced in §7.2.2. The true locations are shifted to account for the beam offset, and the predictions undergo a (14,14) shift so the predicted coordinate aligns with the centre of the kernel. Locations within the exclusion regions are removed. A direct comparison between our predictions and our truth record is made within the global coordinates. True Pb etch-pits are checked for an overlapping predicted Pb etch-pit occurring within an allowable tolerance of 14 pixels (approx 2 pit

widths). Results are also given for a larger value of 30 pixels. A tolerance value of 12 was found (via the training dataset) to account for errors in alignment, particle deflection, or other position uncertainties without causing unnecessary matches between uncorrelated etch-pits. This was increased to 16 for the evaluation matching to account for variability in the 'found' location of predicted etch pits. No significant change in numerical results is observed if this value is changed. Following the overlap matching each location fits one of four scenarios;

- **True positives.** Pb ion tracks with a correct nearby HIP candidate prediction.
- **False positives.** Predicted HIP candidate locations with no matching ion track in the unexposed foils.
- **False negatives.** True Pb ion tracks with no matching HIP candidate prediction
- **True predictions.** Predicted HIP candidates with a true Pb ion track in the vicinity.

No consideration is given to the True negative scenario, where no Pb ion track is present in the unexposed foil and no prediction is generated in the exposed foils. At this level of LHC exposure background particle etch pits appear everywhere in the exposed foil and the correct behaviour is that they should be dismissed trivially. Another point to note is that there is not a 1:1 correspondence between True positives and True predictions as would be the case in a discrete classification task. It is possible for the FCN heat-map signatures of two closely occurring Pb ions to become irresolvable when converted to a discrete location prediction. This clustering scenario is inapplicable to real monopoles and massive exotic particles, and if any true HIP region of interest has been tagged then the algorithm has been successful.

7.5.3 Results

The results of the overlap matching are given for the Combined FCN in (Tab.7.14 & 7.13) while the simple FCN results are shown in (Tab.7.16 & 7.15). The simple FCN outperforms the combined FCN in terms of false positive rate, this might be the result of more classifiers in the combined FCN being more likely to pick up on a feature of an etch-pit. This is supported by the higher signal efficiency ϵ of the combined FCN. The consistency between the blind evaluation data and the full dataset is encouraging, this suggests the method is robust and hasn't been influenced by optimisation bias. Similarly the consistency between different choices of overlap tolerance demonstrates it is not a large factor.

Predicted locations are shown for the multichannel combined FCN (Fig.7.31a), and for the rotational simple FCN (Fig.7.31b). These are shown for the entire foil.

We can estimate the HIP signal acceptance (for a given $REL \approx \frac{Z}{\beta}$ and angle of incidence θ matching the test beam), and the false positive density of the FCN analysis from Tab.7.16 & 7.15. As the background LHC ion flux, and the test beam Pb ion flux are two independent physical processes these results form two separate measurements.

The signal acceptance, or fraction of HIPs we would expect the algorithm to find is simply the fraction of true Pb pits with positive truth predictions. From the evaluation results we can estimate the signal acceptance of the combined FCN as being between 82.8 – 86.5%, quoting the best and worst result. While the simple models give a range of 80.3 – 83.9%.

$$A|_{REL,\theta} = \frac{N_{\text{Found}}}{N_{\text{Incident}}} = \frac{1033}{1194} = 86.5\% \quad (7.3)$$

The false positive rate, we would expect emerging from LHC background is given in Eq.7.4. this is the number of false positives B , identified by the FCN, in area of foil with

	True/Positive	False/Negative	Total	Correct rate (%)
Predicted	938	112	1050	89.3
True Pb	1019	175	1194	85.3
Predicted	941	109	1050	89.3
True Pb	1033	161	1194	86.5

Table 7.13: Combined FCN evaluated with Eval. data only

Overlap tolerance of < 14 (first 2 rows) and < 30 (last 2 rows), showing No. Correct predictions, Correctly predicted true pits, False predictions, and true pits missed, using a threshold/working point of 0.08.

	True/Positive	False/Negative	Total	Correct rate (%)
Predicted	1904	248	2152	88.5
True Pb	2074	432	2506	82.8
Predicted	1913	239	2152	88.9
True Pb	2103	403	2506	83.9

Table 7.14: Combined FCN evaluated with Training and Eval data

Overlap tolerance of < 14 (first 2 rows) and < 30 (last 2 rows), showing No. Correct predictions, Correctly predicted true pits, False predictions, and true pits missed, using a threshold/working point of 0.08.

	True/Positive	False/Negative	Total	Correct rate (%)
Predicted	902	70	972	92.8
True Pb	980	214	1194	82.1
Predicted	912	60	972	93.8
True Pb	1002	192	1194	83.9

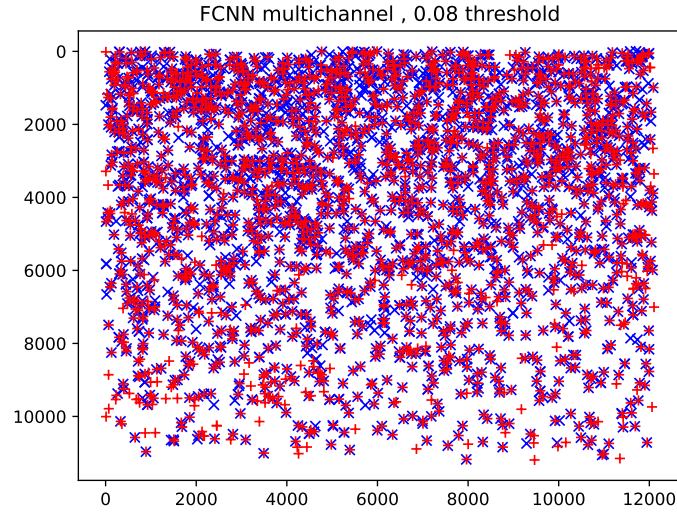
Table 7.15: Simple (Rot)FCN evaluated with Eval. data only

Overlap tolerance of < 14 (first 2 rows) and < 30 (last 2 rows), showing No. Correct predictions, Correctly predicted true pits, False predictions, and true pits missed, using a threshold/working point of 0.5.

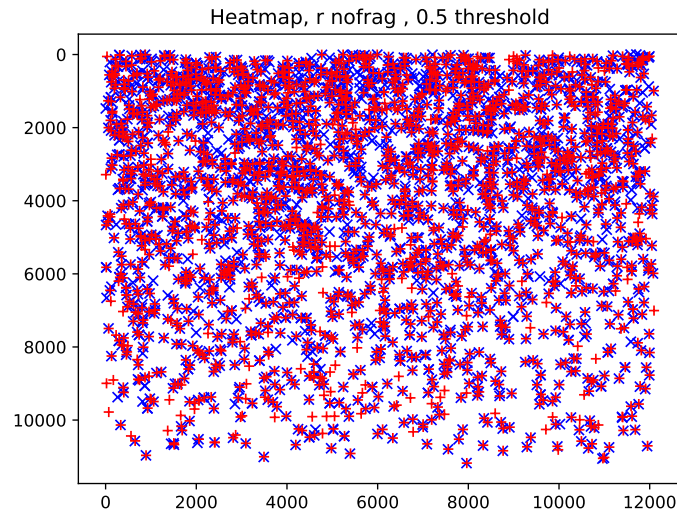
	True/Positive	False/Negative	Total	Correct rate (%)
Predicted	1844	152	1996	92.4
True Pb	2013	493	2506	80.3
Predicted	1860	136	1996	93.2
True Pb	2049	457	2506	81.8

Table 7.16: Simple (Rot)FCN evaluated with Training and Eval data

Overlap tolerance of < 14 (first 2 rows) and < 30 (last 2 rows), showing No. Correct predictions, Correctly predicted true pits, False predictions, and true pits missed, using a threshold/working point of 0.5.



(a) Combined FCN predictions (red +) vs true pits (blue x)



(b) Simple (Rotational) FCN predictions (red +) vs true pits (blue x)

Figure 7.31: Global comparison of predictions with known pits. We can observe a greater ion density towards the top of the foil The FCN output is shown along with the target input, and one channel of the input image data.

2 year run exposure. A value of $5.62 - 5.83 \text{ cm}^{-2}$ is obtained with the combined FCN algorithm. The corresponding figure of the simple FCN is $3.20 - 3.58 \text{ cm}^{-2}$. The simpler model delivers a lower false positive rate.

$$B = \frac{N_{fp}}{Area} = \frac{239 - 248}{(12110 \text{ pix}) \times (11200 \text{ pix}) \times (5.6 \mu\text{m})^2} = 5.62 - 5.83 \text{ cm}^{-2} \quad (7.4)$$

7.5.4 Misclassification Errors

The overall global positions of incorrect predictions is shown for both the simple, and combined FCNs in Fig.7.33. It can be seen that there is correlation between the missed etch-pits of each classifier. Many occur in the top left corner where the ion density is largest. As this section will show that many of the missed 'signal' examples are fragmentations and weak ionisations. By their nature Fragmentations may produce multiple 'missed' etch pits (such as 287 in Fig.7.35). There is no a pattern of near misses visible in global coordinates, suggesting misses aren't caused by the overlap tolerance.

False negatives The 403 missed missed signal examples, from the combined FCN with a tolerance of 30 were inspected manually. Examples are shown in 7.35. Errors due to edge effects or anomalous regions of the foil were noted. Of the remainder the vast majority (322) of misclassified signal were fragments and weak ionisations. Due to their smaller ionisation they are less detectable in the exposed foil and around half are below the practical detection threshold in the exposed foil. Removing the anomalies and approximating a practical detector 50% acceptance for the fragments we can estimate a corrected acceptance of 92.3% (vs 83.9%) for (stable/non-fragmenting) Pb like HIP objects.

$$A_{Cor} = \frac{N_{Found}}{(N_{Cor} - 50\%N_{Weak/Frag})} \approx \frac{2103}{2437 - 161} \approx 92.3\% \quad (7.5)$$

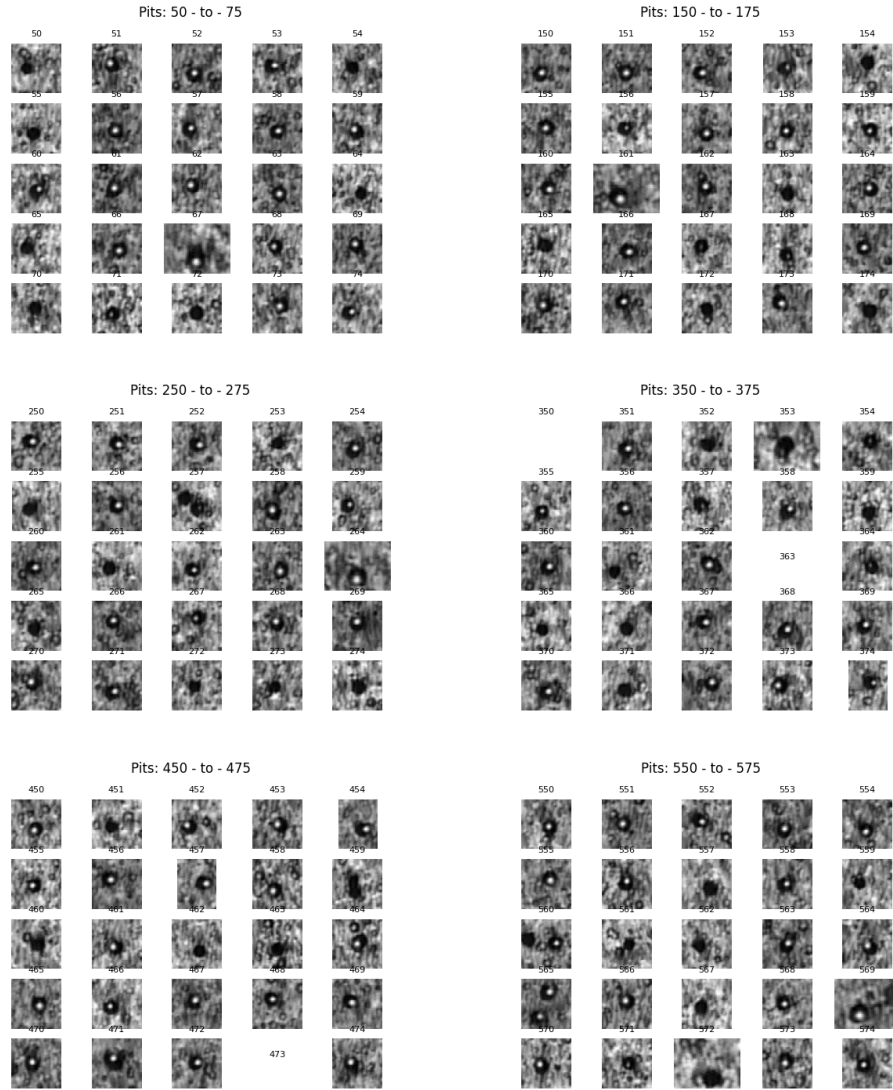
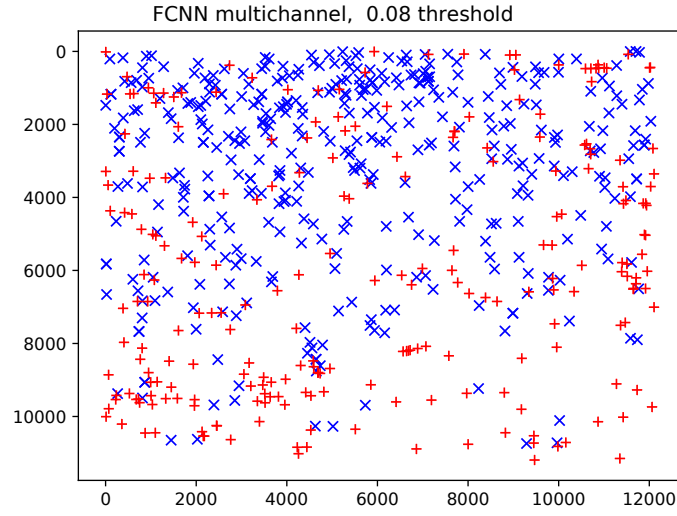
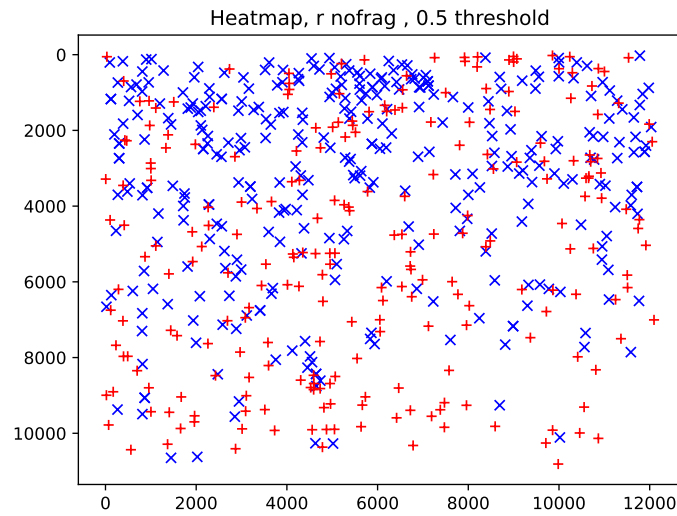


Figure 7.32: Correctly predicted locations, shown in the exposed foil.



(a) False positives (red +) and negative truth (blue x), for the combined FCN.



(b) False positives (red +) and negative truth (blue x), for the simple (Rotational) FCN.

Figure 7.33: Global comparison of incorrect predictions
Consistent with Fig. 7.31a & 7.31b we can see a greater density of misclassified etch pits concentrated in the top left of the foil. Otherwise there are few major patterns

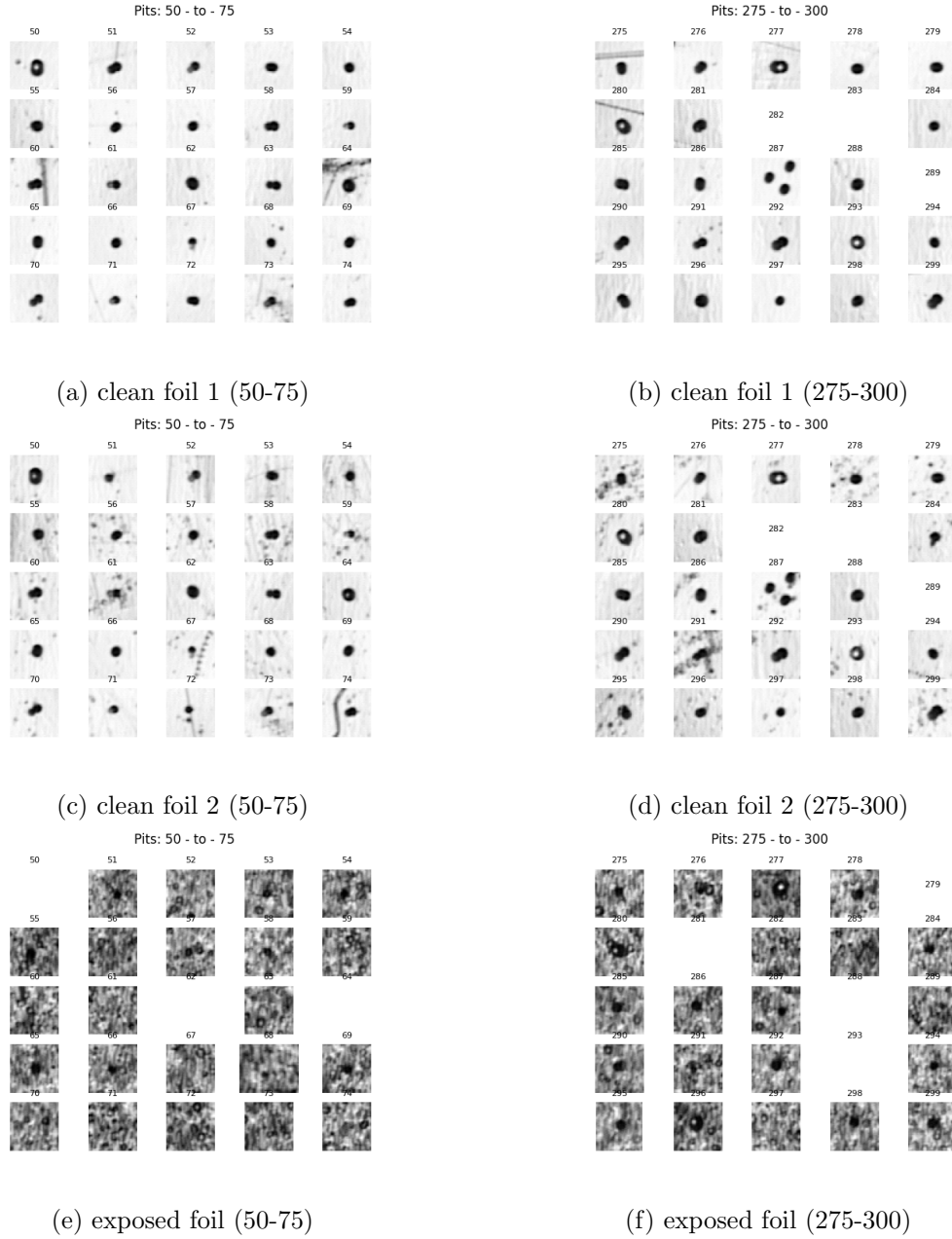


Figure 7.34: Missed truth examples (50-75) and (275-300)

In this set all examples appear in both unexposed foils meeting the definition of through-going HIP objects. However they are nearly all weak ionisations. Fragmentation can be seen in multiple examples, especially 287. Little evidence is visible in the exposed foils (e,f). Images are extracted with a margin of 20 pixels, where an image is missing it indicates it is cut off by the edge of the source image, suggesting they may have been lost due to edge effects.

403	Total
64	Edge effects
5	Anomalies
12	Clear misses
322	Ambiguous

(a) Missed 'Truth' pits

239	Total
66	Edge effects
9	Anomalies
10	Single foil hit
3	Penetrating
151	Remaining

(b) False positives

Table 7.17: Manual categorisation of the incorrect outcomes over the entire foil (for the combined FCN with an overlap of 30)

False positives Performing a similar analysis for the false positives shows a similar number of edge effects. The false positives all show anomalously busy areas of the unexposed foils, indicating a correlation which suggests there might be minor effects from the test beam exposure. There are several detections of real penetrating ions, with most having penetrated through to the first clean foil but not the second. Foil anomalies are part of the practical challenge of NTD image analysis. Removing the edge effects, and the genuine penetrating etch pits which would still warrant further investigation as part of NTD image analysis, modifies our false positive rate by a factor of 67% leading to an adjusted rate of 3.76cm^{-2} .

7.5.5 Uncertainties

The following sources of error impact this analysis, or the ability to generalise its result.

Errors in predicted position, and truth alignment The positional resolution of the FCN is worsened by the reduced dimension of its output. This is exacerbated by the capacity of multiple predictions to become one region after threshold, shifting the predicted coordinate by a pixel or more of the FCN output (≈ 1 pit in the NTD images) potentially causing the centre of the predicted region to be outside of overlap range with the corresponding truth pit. Error in the datum alignment, and deviations from the calibrated beam angle of incidence may result in deviations in the truth locations.

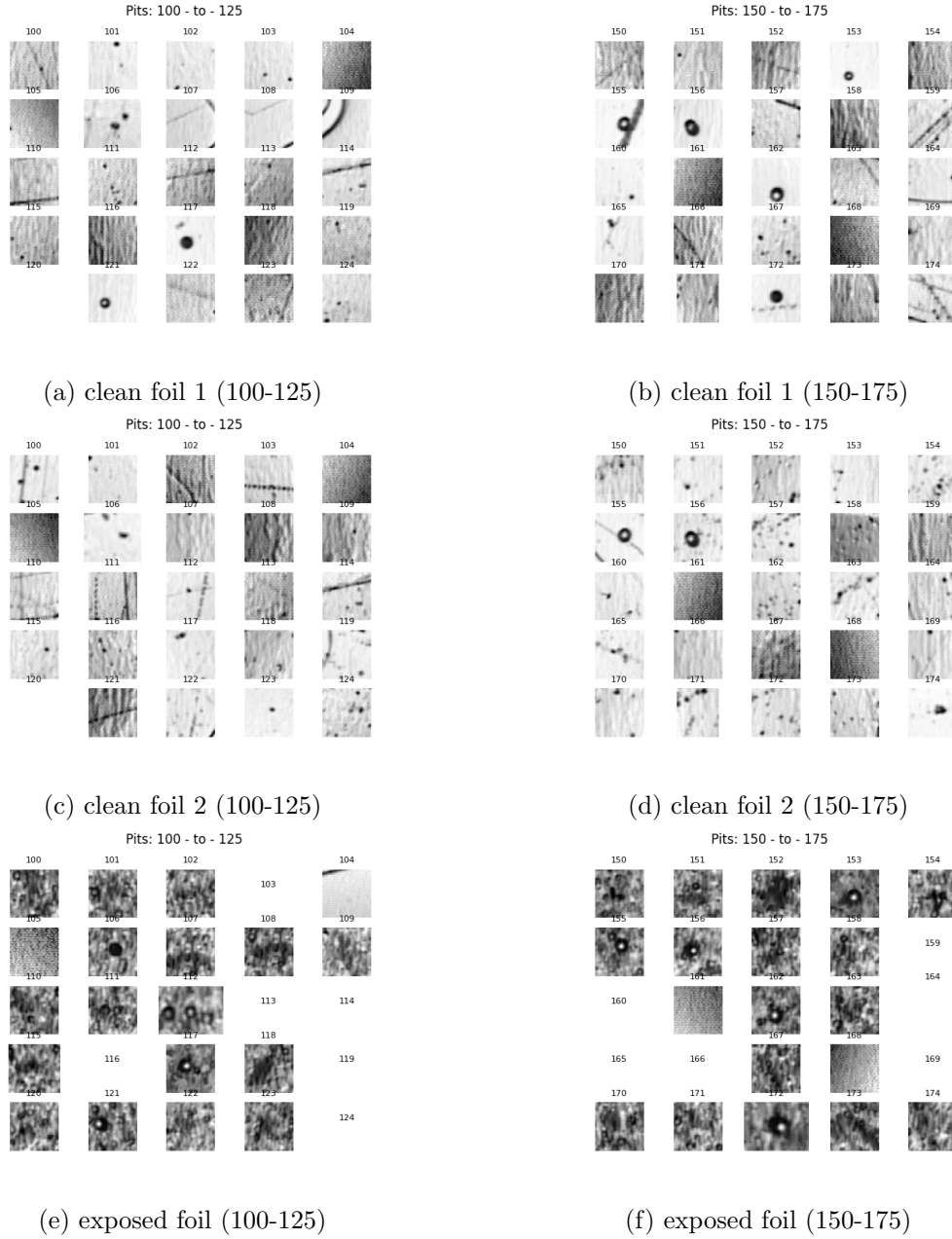


Figure 7.35: False prediction examples (100-125) and (150-175)

In (e) this we can observe several anomalous regions of image at 104 and 105. There are distinct pits in 106 and 112, while 117, 121, 153, 162 and 172 are penetrating tracks, although only for a single foil. The unexposed images show anomalies which could suggest that some event correlated between the foils is triggering a prediction by the networks. 156 and 155 are etch pits visible in all foils, but probably undetected in the unexposed foils by the Gaussian Laplace algorithm due their small size and weak ionisation.

Mitigating these errors by doubling the overlap tolerance produces a +1% average improvement in the accuracy rate, without an appreciable impact on efficiency.

Pb ion fragmentation The significant fraction of Pb ion fragmentations causes an error in the signal acceptance estimate by reducing the effective number of detectable Pb ions. We don't know if the fragments of a particular ion should be considered a HIP candidate, or if such fragments are detectable among the spallation background in the exposed foil. This is the largest uncertainty in the experiment shifting the corrected acceptance (eq.7.3) by $a \pm 5\%$ depending on how they are handled.

Performance at other REL values It is unknown how well either algorithm would handle HIP candidate etch pits with an REL other than that of the $156\text{GeV}/A$ Pb ion beam they were trained to identify. A significant change may change the defacto appearance of such a pit in LHC exposed foils.

Surface anomalies Anomalies in the NTD foil cause both false positives and false negatives. This accounts for 0.002% of false negatives which has no meaningful impact on the signal acceptance. Anomalies are responsible for false positives at a rate of 0.2cm^{-2} .

7.6 Pb Ion study conclusions

This study has shown that a Fully Convolutional Neural Network is FCN is effective for identifying the majority of Pb etch pits, even within the background etch-pit conditions of a full 2 year LHC run. The estimated signal acceptance (Eq.7.3) of the Pb FCN, implies we could lose 7.6% of penetrating HIP candidates that may have been identified by human analysis. As a passive experiment this doesn't represent an irrecoverable event loss.

Designed as a direct search experiment (as opposed to indirect search experiments)

MoEDAL has no statistical background. Detailed investigation of any HIP track candidate can be used to confirm or reject all alternate hypotheses. Instead, the experimental limit is the practical limits of performing detailed investigation on every identified candidate track. A major consideration is the etching of any subsequent foils downstream of an identified HIP candidate etch-pit. Ideally we would wish to calibrate such an etching so that the estimated REL of our hypothesised HIP candidate falls into the linear response portion of the dynamic range of the NTD detector. Calibrating the etching of downstream foils in a stack to a single candidate hypothesis requires having only one potential candidate pit for each front facing foil of a 25x25cm NTD stack. The false positive rate of 5.62cm^{-2} or 3.72cm^{-2} suggests an order of 2000~3000 false positive identifications for a full $25\times 25\text{cm}^{-2}$ MoEDAL NTD foil. For context this number is almost the same as the number of Pb etch-pits analysed in the R&D stack.

$$(25\text{cm})^2 \times (5.6\text{pix}\mu\text{m}^{-1})^2 = (50,000)^2 \times (28 \times 28) = 2.5 \times 10^9 \text{ Pit images} \quad (7.6)$$

While the level of false positives due to LHC background generated by the FCN algorithm is too high to allow for a completely automated NTD analysis, the automated process of generating point coordinate predictions is much faster than human analysis. The FCN can produce predicted candidate locations and extract etch pit data within seconds, without any cognitive overhead in comparing and contrasting multiple images. For a given scan, assuming the same resolution used in this analysis, the area of a regular foil is equivalent to 2.5 billion 28×28 pit images (Eq.7.6). Reducing this to a set of 2500 discrete pit locations reduces the analysis by a factor of a million in terms of time and data. Analysing etch pit data as a set of cropped and centred images (such as Fig.7.32) is much faster and less mentally than analysing scan data directly.

Chapter 8

Conclusions

Machine Learning methods adopted for Nuclear Track Detector (NTD) analysis at MoEDAL have to combine a very high background rejection, with a high signal acceptance. The density of the trillions of LHC background etch-pits accross $300m^2$ of NTD foil, means the criteria normally used to search for heavily ionising particles (HIPs), namely entry and exit etch-pit pair occurs everywhere in the foil.

The 3D structured illumination demonstrated in §5.3.5 breaks the visual symmetry between uncorrelated background etch pit, and potential HIP candidate pits. Applying machine learning to this data can identify subtle features that would be missed in simple 2D images. This imaging method is suitable for the rapid scanning of NTD foils as it can be performed at lower resolution with less precision in the scanning setup. Collecting data using multiple illumination sources allowed ML classifiers to trained on different data channels of the same images. Misclassification errors were found to be uncorrelated between these classifiers allowing them to be combined to deliver vastly superior performance to single data channel classifier §7.3.3. The preprocessing methods used in §6.3.1 and §7.3.1 were effective for removing noise and systematic illumination bias from the scan data, which is beneficial for both automated and visual analysis.

Analysing and classifying the etch pit background hypotheses can improve the overall

robustness and accuracy of etch pit classification. The ensemble methods, and transfer learning techniques featured in this thesis can be readily extended with additional object categories, classifiers, or data channels. This makes them highly suited to future MoEDAL analyses featuring more ion species in training and calibration data.

The Pb lead ion R&D stack has been an effective tool for NTD machine learning research. The alignment of the foils allowed accurate tracking of and reconstruction of HIPs through the NTD stack. This alignment made it possible to automate the extraction and creation of ML datasets, and evaluation of ML algorithms. Prior to this the manual labelling required to create a suitably large supervised etch-pit dataset bottlenecked the Xe study, and previous MoEDAL ML research. The pixel to pixel alignment method can allow future MoEDAL ML algorithms to train directly from the aligned foil data. This has already led to an accurate U-Net FCN algorithm.

Results from the $13\text{GeV}/A$ Xe and $136\text{GeV}/A$ Pb ion studies have shown that relatively few examples are needed to reliably train accurate and robust ML classifiers on the NTD data. The same generic CNN architecture can perform many different etch-pit classification tasks. Pb fragmentation was found to be the largest source of errors, and must be considered in future studies using test beam ions. The semi-automatic training/evaluation regime of the Pb study makes it possible to reuse architectures and pipelines to retrain for other object categories and types. Repeating the exercise with multiple calibration ions would allow the development of CNN/FCNs covering whole hypothetical monopole/exotic REL spectrum. As etch pit formation occurs after ion track formation it would make little difference if test beam exposure occurs before or after LHC exposure. Unetched foils from the Run-2 MoEDAL NTD array (along with new unexposed foils) could be exposed to a variety of test beam ions, replicating the setup of the Pb ion study while removing the long delay needed for LHC exposure.

8.1 Prospects for an ML based NTD analysis

The false positive rate of Machine Learning algorithms, and the signal efficiency required has been the main barrier to the adoption of ML methods for NTD analysis in the context of LHC/MoEDAL. The FCN algorithm can be expected to result in ~ 2500 false positives per foil. While this performance cannot entirely replace human NTD analysis, running over all foils in the MoEDAL NTD array would reduce the overall analysis to approximately several million candidate locations, and a few gigabytes of data, becoming a very tractable problem. For context the algorithm could run over the entire dataset in a few hours on a workstation which is significantly less than the many months required to manually scan the foils. The high signal acceptance (Eq.7.5) of the FCN gives us a significant chance of finding any HIPs (With REL similar to $156\text{GeV}/A$ Pb) present in the scanned NTDs. Altering the NTD analysis procedure to perform the initial etching on multiple foils would expand the potential for an ML analysis even further as the analysis would only have to focus on candidates occurring on both sides of multiple foils. This would reduce the number false positive tracks (as opposed to false positive pits) to a trivially small number that could be verified easily.

The speed of FCN inference means that a Machine Learning NTD Analysis is a low risk, high reward activity that could be conducted in real time, during or after scanning, with the predicted candidate data made accessible to the MoEDAL collaboration. Since Human analysis of the NTD Array can take months there would be plenty of time for NTD Analysis, or Machine Learning groups to examine this data. Further classification and inspection would reduce the of number of candidate locations even further, producing priority locations for detailed microscope examination. Even an unsuccessful Monopole search would provide a useful dataset of unusual anomalies and interesting background etch-pits for future ML analysis research.

Bibliography

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from [tensorflow.org](https://www.tensorflow.org).
- [2] Petr Beneš and Filip Blaschke. Electroweak magnetic monopole: The lower mass bound. In *Proceedings of Science*, volume 390, 2021.
- [3] Bengio Bergstra. Random search for hyper-parameter optimisation, 2012.
- [4] M. Bertani, G. Giacomelli, M. R. Mondardini, B. Pal, L. Patrizii, F. Predieri, P. Serra-Lugaresi, G. Sini, M. Spurio, V. Togo, and S. Zucchelli. Search for magnetic monopoles at the tevatron collider. *EPL*, 12(7), 1990.
- [5] H. Bethe. Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. *Annalen der Physik*, 397(3), 1930.
- [6] Fulgosi Bozinovski. The influence of pattern similarity and transfer learning upon the training of a base perceptron b2. *Proceedings of Symposium Informatica*, 1976.
- [7] B. Cabrera, M. L. Cherry, Donald E. Groom, and J. Musser. Magnetic monopoles at the Parker limit: New prospects for large scale detectors. In *Snowmass '86 Summer Study on the Physics of the Superconducting Supercollider*, pages 665–672, 1986.

BIBLIOGRAPHY

- [8] Blas Cabrera. First results from a superconductive detector for moving magnetic monopoles. *Physical Review Letters*, 48(20), 1982.
- [9] CDF Collaboration. Direct search for dirac magnetic monopoles in pp collisions at $s=1.96\text{TeV}$. *Physical Review Letters*, 96(20), 2006.
- [10] Y. M. Cho. Physical implications of electroweak monopole, 2019.
- [11] Y. M. Cho and D. Maison. Monopole configuration in Weinberg-Salam model. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 391(3-4), 1997.
- [12] Francois Chollet et al. Keras, 2015.
- [13] ATLAS collaboration. 124(3), jan 2020.
- [14] IceCube collaboration. Searches for relativistic magnetic monopoles in IceCube. *The European Physical Journal C*, 76(3), mar 2016.
- [15] MoEDAL collaboration. Magnetic monopole search with the full moedal trapping detector in 13 tev pp collisions interpreted in photon-fusion and drell-yan production. *Phys. Rev. Lett.*, 123:021802, Jul 2019.
- [16] P. Dirac. A Quantum theory of the electron. 1928.
- [17] P. Dirac. Quantized Singularities in the Electromagnetic Field. 1931.
- [18] John Ellis, Nick E. Mavromatos, and Tevong You. The price of an electroweak monopole. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 756, 2016.
- [19] John Ellis, Nick E. Mavromatos, and Tevong You. Light-by-light scattering constraint on born-infeld theory. *Phys. Rev. Lett.*, 118:261802, Jun 2017.

BIBLIOGRAPHY

- [20] J. Nagi et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition, 2011.
- [21] Kunihiko Fukushima. Neocognitron. *Scholarpedia*, 2(1):1717, January 2007.
- [22] Howard Georgi and Sheldon L. Glashow. Unified weak and electromagnetic interactions without neutral currents. *Physical Review Letters*, 28(22), 1972.
- [23] G. Giacomelli and A. Margiotta. The macro experiment at gran sasso.
- [24] M. Giorgini. Search for massive rare particles with the slim experiment. In *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications - Proceedings of the 10th Conference*, 2008.
- [25] Glosser. CC BY-SA 3.0 [<https://creativecommons.org/licenses/by-sa/3.0/>], via Wikimedia Commons, 2013.
- [26] Bob Holdom, Jing Ren, and Chen Zhang. Quark matter may not be strange. *Phys. Rev. Lett.*, 120:222001, May 2018.
- [27] G. t. Hooft. Magnetic monopoles in unified gauge theories. *Nuclear Physics, Section B*, 79(2), 1974.
- [28] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [29] Kyoungtae Kimm, J. H. Yoon, and Y. M. Cho. Finite energy electroweak dyon. *European Physical Journal C*, 75(2), 2015.
- [30] Kyoungtae Kimm, J. H. Yoon, S. H. Oh, and Y. M. Cho. Mass of the electroweak monopole. *Modern Physics Letters A*, 31(9), 2016.

BIBLIOGRAPHY

- [31] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
- [32] Ben Koch, Marcus Bleicher, and Horst Stöcker. Black holes at LHC? *Journal of Physics G: Nuclear and Particle Physics*, 34(8):S535–S542, jul 2007.
- [33] Haffner LeCun, Bengio. Gradient -based learning applied to document recognition, 1998.
- [34] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2014.
- [35] Nick E Mavromatos and Sarben Sarkar. Regularized kalb-ramond magnetic monopole with finite energy. *Phys. Rev. D*, 97:125010, Jun 2018.
- [36] Nick E Mavromatos and Sarben Sarkar. Regularized kalb-ramond magnetic monopole with finite energy. *Phys. Rev. D*, 97:125010, Jun 2018.
- [37] J. C. Maxwell. On physical lines of force. 1861.
- [38] J.C Maxwell. A dynamical theory of the electromagnetic field. 1865.
- [39] Garriga Ojala. Permutation tests for studying classifier performance. *Journal of Machine Learning Research*, 2010.
- [40] L. Patrizii. No Title. *Private correspondence with the MoEDAL NTD analysis group at INF Bologna*, 2019.
- [41] L Patrizii and M Spurio. Status of Searches for Magnetic Monopoles. *Annual Review of Nuclear and Particle Science*, 65(1):279–302, 2015.
- [42] A. M. Polyakov. Particle spectrum in the quantum field theory. *JETP Lett.*, 1974.

- [43] P. B. Price, E. K. Shirk, W. Z. Osborne, and L. S. Pinsky. Evidence for detection of a moving magnetic monopole. *Physical Review Letters*, 35(8), 1975.
- [44] M. Quirós, J. Ramírez Mittelbrunn, and E. Rodríguez. On the topological meaning of magnetic charge quantization. *Journal of Mathematical Physics*, 23(5):873–877, 1982.
- [45] R0oland. CC BY-SA 3.0 (<https://creativecommons.org/licenses/by-sa/3.0/>), via Wikimedia Commons, 2012.
- [46] A. RAJANTIE. FORMATION OF MAGNETIC MONOPOLES IN HOT GAUGE THEORIES. In *Strong and Electroweak Matter 2002*, jun 2003.
- [47] M. W. Ray, E. Ruokokoski, S. Kandel, M. Möttönen, and D. S. Hall. Observation of dirac monopoles in a synthetic magnetic field. *Nature*, 505(7485), 2014.
- [48] A. De Roeck, A. Katre, P. Mermod, D. Milstead, and T. Sloan. Sensitivity of LHC experiments to exotic highly ionising particles. *The European Physical Journal C*, 72(4), apr 2012.
- [49] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 2015.
- [50] P. J. Verveer. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [51] Juyang Weng, Narendra Ahuja, and Thomas S. Huang. Learning recognition and segmentation of 3-d objects from 2-d images. *1993 (4th) International Conference on Computer Vision*, pages 121–128, 1993.
- [52] Tom Whyntie. Dirac’s magnetic monopole (figure). 2016.