Complementary electron, X-ray and neutron diffraction analysis and automated characterisation of sodium-ion battery cathode materials



Andrew Bridger

University College

University of Oxford

A thesis presented for the degree of

Doctor of Philosophy

Trinity Term 2024

Table of Contents

List of Figures **x**

List of Tables **xvi**

List of Abbreviations **xvii**

1 Acknowledgements **xviii**

2 Abstract **xxii**

3 Introduction **1**

3.1 The Climate Emergency 1

3.1.1 More than a century of Climate Change 1

3.1.2 Damaging effects of a warming climate 4

3.1.2.1 Human impacts of high temperature over land . 4

3.1.2.2 Human impacts of heavy precipitation over land 6

3.1.2.3 Human impacts of drought 6

3.1.2.4 Global warming outlook 7

3.2 Net-zero 2050 7

3.2.1 Nuclear Power 9

3.2.2 Wind and Solar Power 10

3.3 Energy Storage 11

3.3.1 Storage Time-frame 13

3.3.2 Long-term storage 13

3.3.2.1 Hydrogen Storage 14

3.3.2.2 Ammonia Storage 14

3.3.3 Short-term storage 15

3.3.3.1 Pumped Storage Hydroelectricity 15
3.3.3.2 Batteries 16
3.4 Sodium-ion Batteries 17
3.4.1 How a sodium-ion battery works 19
3.4.2 Current state of Sodium-ion research 20
3.5 Thesis Outline 23

4 Layered Cathode Structures 25
4.1 Standard Phases of Na_xMO_2-type systems 25
4.1.1 Stacking Hexagonal oxygen layers 25
4.1.2 Deviation from these phases 32
4.2 The P2 phase 33
4.2.1 Sodium Mobility 34
4.2.2 Sodium Sites 35
4.2.3 Sodium Ordering - The Large Zig-Zag 36
4.2.4 Transition Metal Sites 38
4.2.5 Disorder and distortions 38
4.2.6 Morphology 39

5 Experimental Techniques 41
5.1 Introduction to structure probing techniques 41
5.1.1 Overview 41
5.2 Scattering 42
5.2.1 Coherent Scattering 43
5.2.2 Elastic Scattering 43
5.2.3 Scattering Factors and Cross sections 44
5.2.4 Electron scattering 46
5.2.5 X-ray Scattering 48

5.2.6 X-ray vs Electron atomic form factors 50
5.2.7 Neutron Scattering 50
5.3 Diffraction 52
5.3.1 Diffraction in Crystals 53
5.3.2 Bragg's Law 53
5.3.3 Ewald Sphere 53
5.3.4 Structure Factors 63
5.3.5 Multiple scattering 64
5.4 Powder X-ray Diffraction 64
5.4.1 Overview 64
5.4.2 Factors that effect pattern intensities 66
5.4.3 X-ray source: Lab vs Synchotron 67
5.5 Neutron Diffraction 68
5.5.1 Overview 69
5.6 Transmission Electron Microscopy 70
5.6.1 Overview 70
5.6.1.1 Spherical Aberration 74
5.7 Scanning Transmission Electron Microscopy 74
5.7.1 Overview 75
5.7.2 High Resolution Scanning Transmission Electron Microscopy 76
5.7.3 Probe size vs Convergence Semi-angle 77
5.8 4D-STEM 78
5.9 Spectroscopic Analysis 80
5.9.1 Energy-Dispersive X-ray Spectroscopy 81
5.9.2 Electron Energy Loss Spectroscopy 83
5.9.3 EDX vs EELS 86
5.9.3.1 Spatial Resolution 86

5.9.3.2 Energy Resolution 87
5.9.3.3 Signal Strength and Electron Dose 88
5.9.3.4 Beam Conditions 89

6 Computational Techniques **90**
6.1 Neural Networks 90
6.1.1 Basic Summary 90
6.1.2 How a model works 92
6.1.3 Input and Output data 92
6.1.3.1 Regression Model 92
6.1.3.2 Classification Model 93
6.1.4 Input to Output 94
6.1.4.1 Activation Functions 97
6.1.5 Evaluating initial performance 99
6.1.5.1 Loss Functions 99
6.1.6 Training the network 104
6.1.6.1 Gradient descent 105
6.1.6.2 Back propagation 105
6.1.7 Evaluating final performance 109
6.1.8 Improving a network 109
6.1.8.1 Hyper-parameter Tuning 109
6.2 Convolutional Neural Networks 112
6.2.1 Convolutional Layer 113
6.3 Autoencoders 115
6.3.1 Variational Autoencoder 117
6.4 Rietveld Analysis 118
6.5 Principal Component Analysis 119
6.5.1 Covariance Matrix 119

6.5.2 Non-negative Matrix Factorisation 121
6.6 Clustering 122
6.6.1 K-means clustering 122
6.6.2 Gaussian Mixture Model 124
6.6.3 DBSCAN 124
6.6.4 Structural Similarity Index 126
6.7 Simulated Annealing 127

7 Methods and Instruments 130
7.1 Equipment and Instruments 130
7.1.1 Synthesis Lab Equipment 130
7.1.1.1 Glove Box 130
7.1.1.2 Weighing scales 130
7.1.1.3 Ball Mill 130
7.1.1.4 Vacuum Oven 130
7.1.1.5 Pellet Press 131
7.1.1.6 Tube Furnace 131
7.1.2 Sample Screening Lab Equipment and Instrument 131
7.1.2.1 Transmission Electron Microscope 131
7.1.2.2 Sample Holder 131
7.1.3 ePSIC 131
7.1.3.1 E01 Electron Microscope 131
7.1.3.2 E02 Electron Microscope 132
7.1.3.3 Standard Sample Holder 132
7.1.3.4 Vacuum Transfer Holder 132
7.1.4 X-ray Diffractometer 132
7.1.4.1 I11 132
7.1.5 Neutron Diffractometer 133

7.1.5.1 HRPD 133
7.2 Synthesis 133
7.2.1 Standard Cathode Synthesis 133
7.2.2 Precursor Chemicals 135
7.2.3 Sample Synthesis 135
7.3 Sample Preparation 135
7.3.1 Powder Diffraction Capillary Loading 135
7.3.2 TEM sample loading 137
7.3.3 Experimental Samples 138
7.4 Sample Pre-screening 139

8 A Versatile, Automated Workflow for SEND data 144
8.1 Developing a SEND workflow for domain mapping 146
8.2 Workflow 149
8.2.1 The Variational Autoencoder 153
8.2.2 Latent Clustering 155
8.2.3 Boundary Estimation 155
8.2.4 Centroid estimation 157
8.2.5 Assigning data utilising convex hulls 160
8.2.6 Iterating 160
8.3 Example Datasets 161
8.3.1 Simulated dataset 161
8.3.2 Precipitates in Al alloy thin foil 164
8.3.3 Phase Distribution in P2 Cathode Material 168
8.4 Evaluating Confidence in Domain Identification 172
8.5 Comparison with Other Techniques 179
8.5.1 Direct Non-Negative Matrix Factorisation 179
8.5.2 Feature Extraction 180

8.6 Workflow Automation 187
8.6.1 ML Centroid Detection 188
8.7 Workflow Limitations and Resource Requirements 190
8.8 Conclusion 193

9 Investigation of Intra-layer Ion Orderings in Sodium Ion Cathode Materials 196

9.1 Evidence of Sodium-Ion Superstructure in Experimental Data . . 196
9.2 Development and application of structural model using an empirical potential 201
9.2.1 Development of the model and the potential 202
9.2.2 Summary of the model and the potential 203
9.2.3 Resulting insights of intra-layer ion orderings using the empirical model 206
9.2.3.1 Transition Metal Ordering 206
9.2.3.2 Corresponding Sodium Ordering 212
9.2.4 Discussion of sodium orderings produced by the empirical potential 216
9.3 Development and application of structural model using a symmetry-driven approach 217
9.3.1 Development 217
9.3.2 Summary of the implementation of the symmetry-driven approach 218
9.3.2.1 Experimental Comparison - SEND 220
9.4 Resulting insights of intra-layer ion orderings using the symmetry-driven approach 221
9.4.1 $2\sqrt{3} \times 2\sqrt{3} \times 1$ Simulation 221
9.4.1.1 SEND Comparison 222

9.4.2 Conclusion of combined results of two approaches for investigating the intra-layer ion orderings 226

10 Investigation of Inter-layer Ion Orderings in Sodium Ion Cathode Materials 229

10.1 Oxygen and Transition Metal Ordering 230
10.1.1 Oxygen Ion Ordering 230
10.1.2 Metal Faulting 232
10.1.3 Ionic orderings fit by TOPAS stacking model 235
10.1.3.1 Powder X-ray Diffraction Stacking Model . . . 235
10.1.3.2 Powder Neutron Diffraction Stacking Model . . 241
10.1.4 Discussion of results of PXRD and PND stacking models . 243
10.2 Sodium Ion Ordering 247
10.2.0.1 Sodium Faulting 247
10.2.0.2 Implementation 249
10.3 Results of intra-layer and inter-layer sodium ion ordering model . 249
10.3.1 Powder X-ray Diffraction 249
10.3.1.1 DK Model 251
10.3.1.2 TW Model 253
10.3.2 XRD-derived model for Microscopy simulation 254
10.4 Understanding why traditional approaches structure solution approaches struggle 258
10.5 Conclusions 259

11 Investigating Mn Formal Valence 261

11.1 Evidence of Mn^{3+} 262
11.1.1 Compositional Analysis 262
11.1.2 Lattice Parameters 264

11.2 Achieving atomic resolution imaging with the **c** axis in-plane . . . 269
11.2.1 Screening particles to find suitable candidates 270
11.2.2 HR-STEM Processing 274
11.2.3 Mn Oxidation State Estimation 277
11.2.3.1 White Line Intensities 277
11.2.3.2 Energy Offsets 279
11.2.3.3 Linear Combinations 281
11.2.3.4 MnEdgeNet 281
11.2.4 Results 282
11.2.4.1 Dark Field Images 282
11.2.4.2 Oxidation State Estimation 283
11.2.5 Conclusions 292

12 Conclusions and Future Work **293**
12.1 Future Work 294

Bibliography **298**

A Appendix **322**

List of Figures

3.1 The Keeling Curve 2
3.2 Land-ocean Temperature Index 3
3.3 Frequency of Extreme Weather Events 5
3.4 UK electricity generation fuel mix 8
3.5 UK Carbon Emissions from Power Stations 9
3.6 Variation in Daily Solar and Wind Energy Production 11
3.7 Variation in Annual Solar and Wind Energy Production 12
3.8 Battery Capacity Requirement Forecast 16
3.9 Elemental Abundances in the Crust, Seawater and Geographically 18
3.10 Schematic of a Sodium-ion Battery 20

4.1 Oxygen Layer Positions and Associated Interstitial Sites 26
4.2 Directionality of Transition Metal Octahedra 28
4.3 Sodium Interstitial sites generated from different oxide stackings . 29
4.4 Four Main Layered Cathode Phases 32
4.5 Diffusion Pathways 35
4.6 Large Zig-Zag Pattern 37

5.1 Bragg's Law Wave Diagram 54
5.2 Vector construction of Bragg condition 55
5.3 Top-to-tail depiction of the vector, $\mathbf{K}$ 56
5.4 Ewald Sphere Construction 60
5.5 Effect of radiation wavelength on Ewald Sphere Construction . . . 62
5.6 Powder X-ray Diffraction Experimental Setups 65
5.7 TEM and STEM Ray Diagrams 71

5.8 Graphic representing the phenomena which can result from the interaction of a sample with an electron beam. 82
5.9 Electron-beam Induced Electronic Transitions 84

6.1 Schematic representation of a Neural Network. 91
6.2 Schematic of Single Connection Calculation in a Neural Network 96
6.3 Activation Functions 98
6.4 Diagram of a Convolutional Layer in a CNN 114

7.1 TEM Finder Grid 141
7.2 Labelling Scheme for TEM Finder Grid 143

8.1 VAE Workflow Schematic 147
8.2 VAE Architecture 154
8.3 Boundary Point Estimation 157
8.4 Image: Summary of VAE workflow for a Simulated Sample 162
8.4 Caption: Summary of VAE workflow for a Simulated Sample . . . 163
8.5 Image: Summary of VAE workflow for a Sample with Precipitates on Al foil 167
8.5 Caption: Summary of VAE workflow for a Sample with Precipitates on Al foil 168
8.6 Image: Summary of VAE workflow for a Layered Metal Oxide Sample 171
8.6 Caption: Summary of VAE workflow for a Layered Metal Oxide Sample 172
8.7 Procedure for generating the Proportional Label Overlap Matrix . 175
8.8 Procedure for calculating the Statistical Consistency Maps 176
8.9 Example of Consistency Maps applied to the Al alloy dataset . . . 177
8.10 PCA evaluation 181

8.11 Comparison Metric Evaluation 182
8.12 PCA for centroid prediction 183
8.13 Image: Contemporary Approach Comparison 186
8.13 Caption: Contemporary Approach Comparison 187
8.14 ML-based Centroid Prediction 191

9.1 Evidence for Sodium Super-Lattice Ordering 198
9.1 Evidence for Sodium Super-Lattice Ordering 199
9.2 SEND Domain Mapping of Sodium Ordered Sample 199
9.3 LZZ TEM Simulation 200
9.4 Transition Metal ordering of the two layers of the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell with N33 ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$) composition. The corresponding sodium ion orderings are discussed in Section 9.2.3.2 . . 206
9.5 Transition Metal ordering of the two layers of the $10\sqrt{3}\times10\sqrt{3}\times1$ super cell with N33 ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$) composition. The corresponding sodium ion orderings are discussed in Section 9.2.3.2 . . 207
9.6 Transition Metal ordering of the two layers of the $10\sqrt{3}\times10\sqrt{3}\times1$ super cell with N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$) and N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$) compositions 208
9.7 Sodium occupancies of the two layers of the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cell with N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$) and N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$) compositions 209
9.8 Approximate charge distribution across the four metal layers of the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cells with N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$) and N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$) compositions 211
9.9 Histogram of approximate charge distribution across the four metal layers of the $10\sqrt{3}\times10\sqrt{3}\times1$ super cells with N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$) and N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$) compositions 212

9.10 Sodium ordering of the two layers of the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell with N33 ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$) composition 213
9.11 Sodium ordering of the SZZ ordering scheme compared to the LZZ ordering scheme. 214
9.12 Other potential sodium ordering of the two layers of the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell with N33 composition ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$), showing a forward biased or a backwards biased zigzag 215
9.13 TEM simulation of a potential Sodium Ordering Scheme 216
9.14 Best fitting Sodium Orderings to SEND patterns 223
9.15 Best fitting Sodium Orderings to SEND patterns 224
9.16 Summary of proposed Sodium Ordering 225
9.17 Nearest Neighbour Distances of possible Sodium Ordered Structures 227
9.18 Structural Energy predictions of possible Sodium Ordering Schemes 227

10.1 Oxygen Layer Stackings and Transition Probabilities 231
10.2 Transition Metal Stackings and Transition Probabilities 232
10.3 Transition Metal Orderings for different Ni:Mn ratios 234
10.4 Contours of the Goodness of Fit surface across the different parameterisations, P_a and P_b, of a stacking model. 236
10.5 Effect of parameters P_a and P_b on the quality of N25 PXRD fit . . 238
10.6 Effect of parameters P_a and P_b on the quality of N30 PXRD fit . . 239
10.7 Effect of parameters P_a and P_b on the quality of N33 PXRD fit . . 240
10.8 XRD fits of structural models of stoichiometric sample series . . . 241
10.9 PND fits of structural models of stoichiometric sample series . . . 242
10.10 Cartoon Schematic of Layer Ordering in Stoichiometric Sample Series 244
10.11 PXRD data for two samples of N33 composition 250

10.12Fit to the DK-N33 composition PXRD data using the disordered sodium super-structure model. 252
10.13Comparison of Probabilistic model fit to the standard SZZ and LZZ models 253
10.14Effect of disorder on PXRD sodium super-structure fits 254
10.15HR-STEM Annular Dark Field Data 256
10.16HR-STEM Dark Field Data and Simulations 257

11.1 Grid search for optimal values of S and D parameters for the PND refinement of the three stoichiometric samples 263
11.2 Compariosn of lattice paramter ratio and calculated formal Mn valence 265
11.3 Potential relationship between lattice parameters and d-orbital Energies 268
11.4 Pre-screening Sample Reference Workbook 271
11.5 Image: Pre-screening Sample Information 273
11.5 Caption: Pre-screening Sample Information 273
11.6 The summed EELS spectral intensities for each scan location in a HR-STEM experiment 275
11.7 The Dark Field image for the same scan as the EELS spectra shown in Figure 11.6 276
11.8 Routine for identifying Layers in HR-STEM 278
11.9 Identified Layers in HR-STEM 279
11.10Pre-processing routine for EELs data 280
11.11Dark Field Images of six high resolution images which were collected with simultaneous EELS 283
11.12Calibration Curves for samples in Figure 11.11 284
11.13Oxidation state estimations per layer for Sample N25-1 285

11.14Oxidation state estimations per layer for Sample N25-2 286
11.15Oxidation state estimations per layer for Sample N25-3 287
11.16Oxidation state estimations per layer for Sample N25-4 288
11.17Oxidation state estimations per layer for Sample N30-1 289
11.18Oxidation state estimations per layer for Sample N30-2 289
11.19Distribution of Estimated Oxidation States 291

A.1 Electron Dose Stability of Samples. 322
A.2 Image: Comparison of Ordered Model versus Disordered Model on XRD fits 323
A.2 Caption: Comparison of Ordered Model versus Disordered Model on XRD fits 324
A.3 Image: Comparison of Ordered Model versus Disordered Model on PND fits 325
A.3 Caption: Comparison of Ordered Model versus Disordered Model on PND fits 325

List of Tables

7.1 Information on starting materials used in solid-state synthesis. . . 135

7.2 N25, N30, and N33 sample preparation 135

7.3 Details of experimentally investigated samples, including: studied compositions, diffractometer used, detector used, scan timings and approximate sample weights. 138

8.1 Workflow Step Times 185

9.1 Symmetry-Driven Sodium Ordering Structure Table 221

10.1 N25, N30, and N33 PXRD fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF) 237

10.2 N25, N30, and N33 PXRD fit parameterisations 237

10.3 N25, N30, and N33 PND fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF) 242

10.4 N25, N30, and N33 PND fit parameterisations as well as the parameterisation for the additional P3 phase present in the N25 and N30 samples 243

11.1 N25, N30, and N33 calculated cation stoichiometry and subsequent Mn formal valence 264

A.1 N25, N30, and N33 PXRD fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF), shown for the developed ‘disordered’ models and the standard, ‘ordered’ models. 324

A.2 N25, N30, and N33 PND fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF), shown for the developed 'disordered' models and the standard, 'ordered' models. 326

1 | Acknowledgements

I am not really one for sentiment, but neither do I tend to be one for writing 300 page books on diffraction, microscopy and data analysis, so on account of the later we can make an exception to the former. In addition to my own reluctance for sentiment, the omission of these acknowledgements might give the impression that this work could have been achieved in isolation, which is grossly untrue, and so within that, there is a duty to persevere. The many contributions of the many people who have aided in the creation of this thesis deserve to be duly recognised and celebrated.

The first set of thanks goes to my academic support, foremost from my three supervisors: Mohsen Danaie, Bill David and Keith Butler. If this thesis seems like a strange amalgamation of Electron Microscopy, Structural Analysis, and Computational Chemistry, you only have these three to blame.

Mohsen was the foundation on which this project was built. He has been so supportive day-in and day-out of this DPhil, unfaltering in biweekly zoom catch-ups, always available for a chat and willing to help out with any problem - large or small. He really guided this projected experimentally, and has been extremely patient as I periodically disappeared down rabbit holes. He supported every aspect of this project, both in the research phase but also in writing up and administration. Mohsen read and gave feedback to pretty much every word of consequence which I have written over this DPhil and he was integral in getting a coherent publication, and (hopefully) a coherent thesis out of the end of it. I don't know what this project would have looked like without him, but I know with certainty that it would have been much, much worse. It was truly a privilege to have him as a supervisor and I can't overstate my appreciation enough.

I wrote this in my Part II thesis acknowledgements, but it still holds true: Professor Bill David is probably the smartest and busiest man I have ever met. His knowledge of solid state chemistry is unrivalled and it was a real pleasure to have his help guiding me through this project. Throughout this DPhil, whenever Bill and I got the opportunity to chat about how the project was progressing, I would update him and he would give his initial thoughts and suggestions for the next step. On almost every occasions his initial knee-jerk response ended up being the most pertinent insight into whatever was the problem at hand (although sometimes it took me a few hours of thinking to realise this). There is no-one who's opinion of scientific research I value more, and I am very grateful for his time and feedback in the production of this thesis.

For support in all things computational chemistry, you can't ask for a better supervisor than Keith Butler. Keith brought a wealth of knowledge to this project, and I don't think there was a single computational technique I implemented or even spoke about, that Keith was already familiar with. An example of this was for the structural energy estimation that I wanted to do for the sodium ordering, where I was think of the ways I could approach it, only for Keith to already have a working implementation of a model, which he kindly applied to all of my structures. Especially in the early stages, where I was running free with developing the SEND data processing, Keith insight into dimensionality reduction techniques was invaluable to getting the VAE workflow the point it reached. Keith also managed to provide all his support while moving institutions on a seemingly annual basis, so I am incredibly grateful for his continued support.

In addition to my supervisors I had no bounds of additional academic support from the rest of the David Group, but especially from, Tom Wood. Tom has been fantastic throughout this entire DPhil and when trying to solve a problem I don't think there is anyone I would rather work it through with. Tom has such a talent

for understanding whatever you are saying to him, no matter how incoherent it might feel, and for communicating his intelligence clearly back to you. There were certain points in this project, especially around the sodium ion orderings investigation, that if I didn't have Tom's support, I don't think I could have made any progress. I'd also like to thank all the rest of the current members of the David Group: Jonny, Matt, Emma, Zoe, Paul and Geofrey, as well as all the members over the years, for putting up with me trying to wise-crack my way through group meetings and be in the lab as little as possible (with special note to Tom and previous Part II student, Dan, who's samples I used to avoid further synthesis).

Outside of the academic circle I couldn't have asked for better support than the support I received from my soon-to-be wife, Soph (and that annoying dog that follows her around). For the last four years Soph has tolerated any and all of my trivial work woes, as well as my obsession with spending all of my free time hanging out with dudes in boats, I love you and hope you thoroughly enjoy reading the rest of this thesis.

On the topic of dudes in boats, I'd like to thank all the friends (not just the dudes) that I've made through the ridiculous pursuit of getting good at pulling on a stick. Over the 4 years of this DPhil, pretty much every friend I have made has been through the sport of rowing, and they've all helped make this such a great period of my life. There are really too many people to name, but that's the way it goes when you are in crews of 8+, so if you are a rowing friend reading this, sorry you didn't get to see your own name explicitly.

Finally, I'd like to thank my family, who must be eager to get through this so they can get stuck into the meat of research, but without you I wouldn't have had the privilege of the last 8 years in Oxford, where I will have (again, hopefully) achieved two degrees, made countless friends and found my wife, so thank you.

Anyway, enough of that, on to the chemistry....

2 | Abstract

This thesis sets out to improve the understanding and development of sodium ion battery cathode materials. The thesis focuses on layered metal oxides across a range of compositions and seeks to develop models of the atomic structure that are more expressive of the actual systems than the currently accepted approaches. In doing this multi-modal structural studies will be used, leveraging the individual advantages of Electron, X-ray and Neutron diffraction to produce a more complete model of the physical structures.

The use of cutting-edge electron diffraction techniques which produce vast quantities of data, lead to the development of a novel semi-autonomous data-processing workflow, that is widely applicable to many material systems, not just battery cathode materials. This workflow is documented and investigated across a range of use cases, as well as being applied in the subsequent cathode studies.

The electron diffraction workflow is used in conjunction with other complementary diffraction approaches, as well as pure computational studies, to develop atomic models for the series of cathode structures. The structural models developed provide a better fit to the experimental data gathered than the currently accepted models and identify a novel intra-layer sodium ion ordering scheme that has not been observed in the literature. Additionally the models express further ionic orderings, both intra-layer and inter-layer, that are supported by the experimental evidence and identify structural nuances which will effect battery performance, but are overlooked by current models.

Finally, a methodology for studying oxidation state of Mn at a near atomic resolution level is discussed, with the workflow applied to the cathode material samples and the results tying into the insights of the structural models developed in the

earlier chapters.

3 | Introduction

3.1 The Climate Emergency

The single greatest threat to the longevity of humanity and quality of life on earth, is that of the heating climate. The impacts of the Climate Emergency has reached all corners of the planet, natural disasters have become increasingly common place, clean air in populous regions has been relegated to a luxury, and without stark and immediate intervention, the situation will only continue to deteriorate. The importance of a complete societal paradigm shift to sustainable living cannot be understated. A critical component of successfully making the shift to a sustainable society is through a complete overhaul of global energy infrastructure, ceasing all energy production from fossil fuels and transferring entirely to renewable energy sources. In achieving this transition, energy storage will be a key challenge, and battery storage technology will provide an element of the solution. This thesis aims to aid the development and improvement of battery storage technologies.

3.1.1 More than a century of Climate Change

As far back as 1896, Arrhenius identified that atmospheric CO_2 levels will affect the planet's air temperature[1]. In March 1958, the American Weather Bureau's Mauna Loa Observatory first installed an atmospheric CO_2 analyser and, on the 29th of March, it collected its first measurement of $313\,\mathrm{ppm}$. Since then, it has been collecting data daily, data which forms the famous plot known as the Keeling Curve[2].

Figure 3.1 shows the Keeling Curve as of December 2023, with atmospheric con-

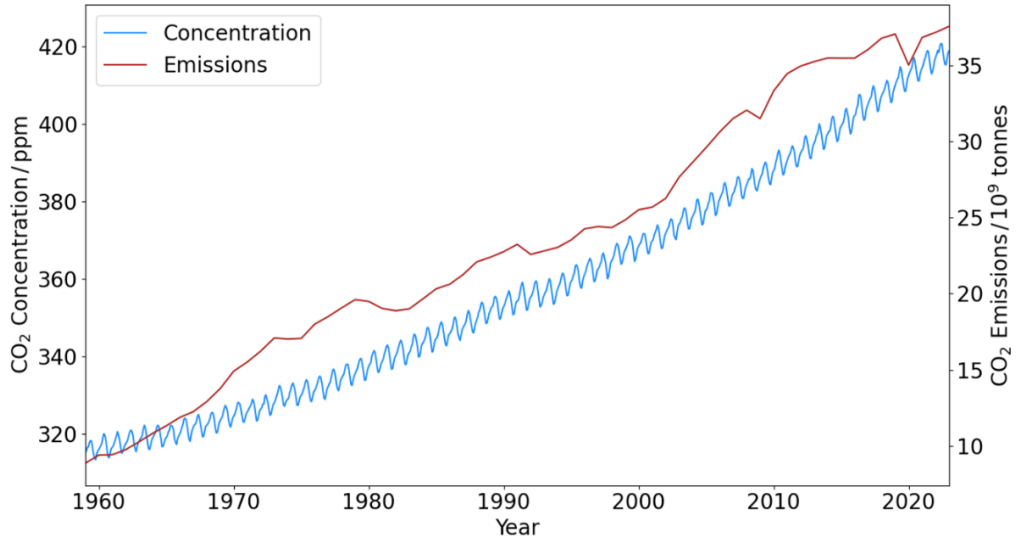


Figure 3.1: The Keeling Curve reporting the atmospheric CO_2 levels across time is shown in blue with the y-scale on the left hand side. Global CO_2 emissions across time are shown in red, with the y-scale on the right hand side.

centration of CO_2 in blue. Despite the warnings of Arrhenius' 1896 model[1], it is patently evident that over the last 66 years, the CO_2 concentration in the atmosphere has been rapidly increasing. It has been proven that anthropogenic CO_2 production has been the driving force behind this increase[3][4], with the global CO_2 emissions (given in red in Figure 3.1) showing a strikingly similar trend[5]. Alongside this increase in CO_2 concentration we also observe the corresponding increase in global surface temperature[6] as shown in Figure 3.2.

The sixth Intergovernmental Panel on Climate Change (IPCC)'s 2021 report stated directly that it is 'unequivocal that human activities have heated our climate'[3]. The document then further outlines the catastrophically detrimental effects that even modest temperature increases pose globally.

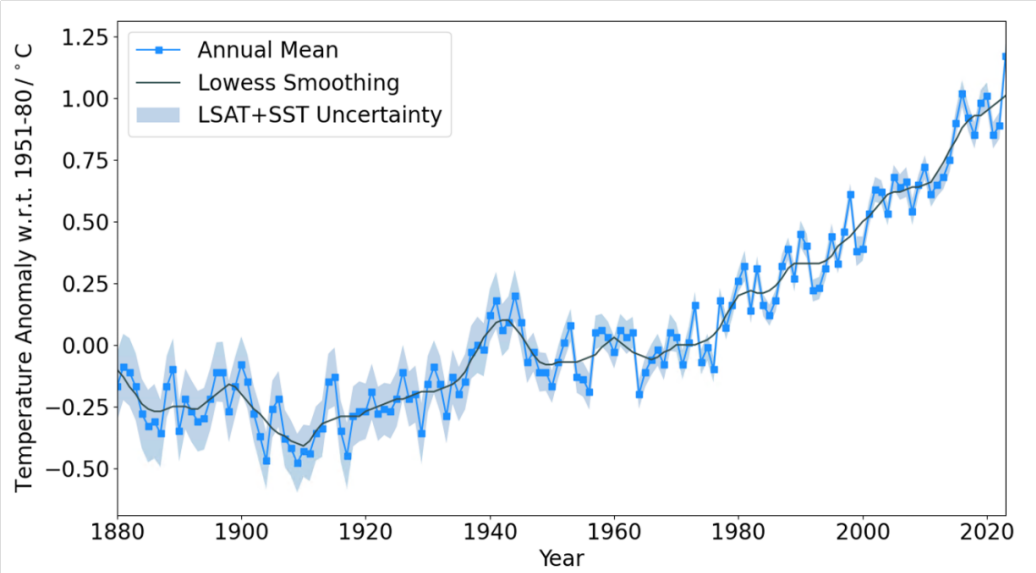


Figure 3.2: Land-ocean temperature index, 1880 to present, with base period 1951-1980. The light blue line with square markers is the global annual mean and the dark blue line is the five-year LOWESS (Local Weighting Scatterplot Smoothing) smooth. The blue shading represents the total (LSAT and SST) annual uncertainty at a 95% confidence interval

3.1.2 Damaging effects of a warming climate

Figure 3.3 is taken from the IPCC report and summarizes the link between the potential magnitude of warming and how this relates to the occurrence of different extreme weather phenomena. It shows that extreme weather events, which might have occurred once every 10 years in the historically cooler climate, will increase in both frequency and severity as global temperatures increase. The three extreme weather events documented are high temperatures over land, heavy precipitation over land, and agricultural and ecological drought. While the link of global warming to increased frequency of high temperature events is trivially intuitive, the origin of the predicted increase in both heavy precipitation and drought occurrence is perhaps less obvious. The primary justification of observing both these seemingly contradictory phenomena occurring with concurrently increased frequency is that increased global temperatures will alter atmospheric circulation patterns, such as jet streams and storm tracks. This combines with the facts that: warmer air can hold more moisture[7] (as derived from the Clausius-Clapeyron equation); increased temperatures lead to an increase in oceanic evaporation[8]; and that higher temperatures will intensify the convective processes[9]; all result in more moisture being present in a more volatile and polarised system[10]. All three of the phenomena outlined pose severe threat to human life and well-being in manners extensively, but non-exhaustively discussed below.

3.1.2.1 Human impacts of high temperature over land

Cases of extreme temperature in populated areas can have many short-term and long-term effects that range in severity and subtlety. Such high temperature events can lead directly to heat-related illnesses such as heatstroke or heat exhaustion[11], but can also exacerbate existing chronic conditions such as heart disease (as high

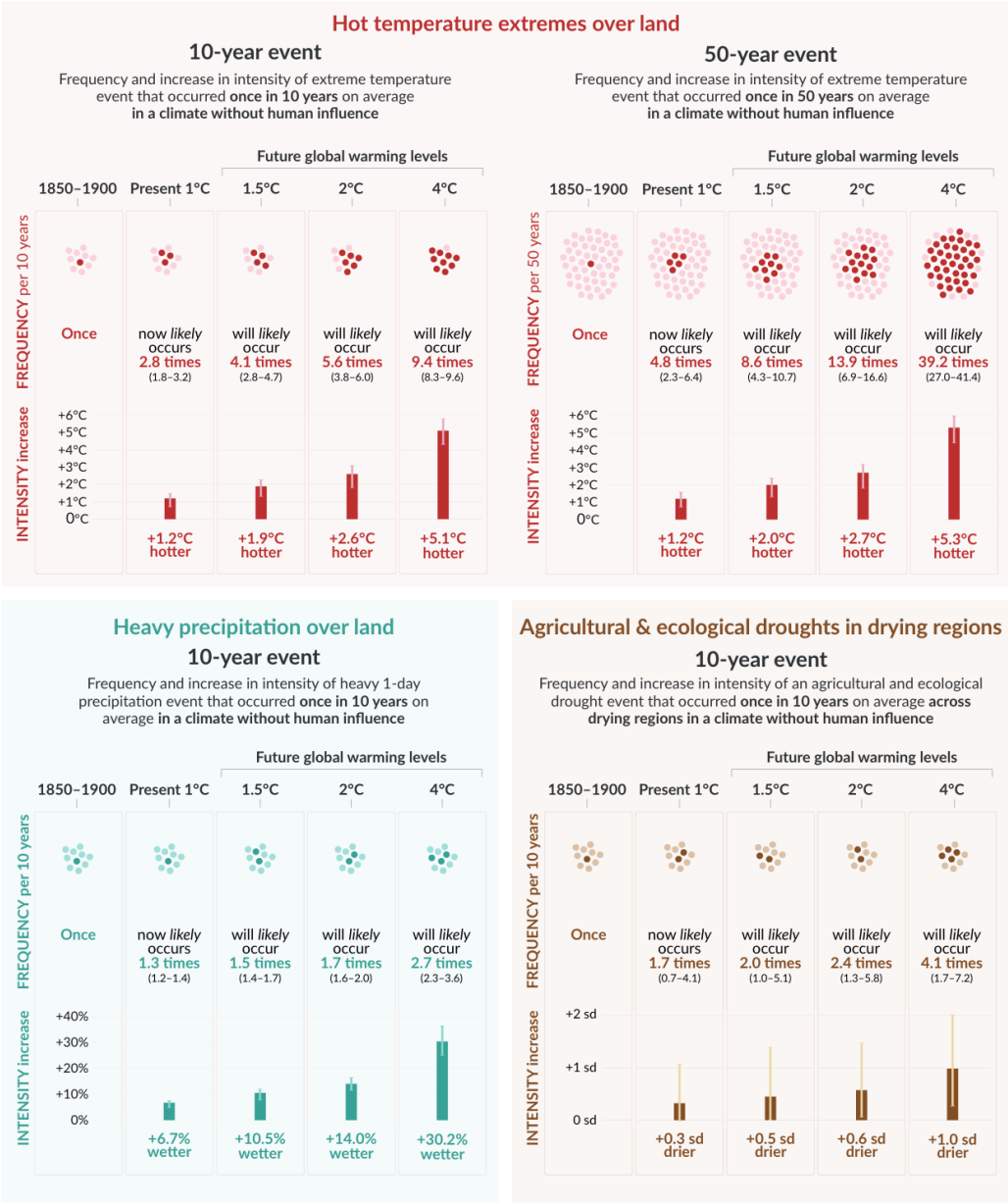


Figure 3.3: Projected changes in the intensity and frequency of hot temperature extremes over land, extreme precipitation over land, and agricultural and ecological droughts in drying regions taken from the IPCC 2021 report[3]

temperatures can lead to dehydration and electrolyte imbalance which will impact heart/ muscle function[12]), respiratory problems (as high temperatures increase air pollution and reduces air quality[13]), and diabetes (as diabetes can impair sweat production, thus preventing effective cooling and triggering heat-related illness[14]). Alongside the direct health issues, there are socioeconomic concerns, such as: a decrease in workforce productivity and earnings[15]; disruption of social services[16]; and damage to the agricultural sector with crop failures and food shortages[17].

3.1.2.2 Human impacts of heavy precipitation over land

The immediately apparent threat associated with heavy precipitation is flooding. Not only does flooding cause damage to infrastructure and property, but in extreme cases, it can lead to loss of life[18]. In addition to flooding, however, other ecological risks are present, such as landslides and mudslides, both of which can be destructive and dangerous[19]. As well, the ecological threats, heavy precipitation can cause contamination of drinking supply with waterborne diseases[20] such as cholera, typhoid, and hepatitis A. Farming can also be detrimentally effected, both from direct damage to crops from flooding, but also nutrient leaching of soils[21]. There are also a host of additional socioeconomic effects such as displacement/homelessness[22], strained emergency services [23], disruption of education[24], and mental health decline[25].

3.1.2.3 Human impacts of drought

Periods of prolonged drought can lead to very similar issues to heavy precipitation, such as crop failure[26], soil degradation[27], and strain on emergency services[28], but the primary, immediate threat is scarcity of fresh water for drinking and sanitation[29]. A scarce drinking water supply can lead to severe health

complications directly, through dehydration, but also indirectly, through conditions such as heatstroke. Additionally the dried ground can lead to an increase in dust and other particulate matter in the air, reducing air quality[30] and aggravating respiratory diseases[31]. Over the last few years drought related wildfires have been increasingly pervasive in some of the drier regions of the developed world, notably the Australian bushfires of 2019/20[32] and the California fires in 2023[33]. In the developing world, water scarcity can also lead to displacement[34], political instability[35] and conflict[36].

3.1.2.4 Global warming outlook

Simply evaluating this cursory overview of the effect of atmospheric CO_2 on surface temperature; the impact this increase will have on frequency of extreme weather phenomena, and the subsequent risks and impacts of these events, it is abundantly clear why the term 'Climate Change' has been upgraded to the 'Climate Emergency'. It is absolutely imperative that we as a society attempt to minimise the current rise in global temperatures. The 'Paris Agreement', an international treaty signed by 196 parties at the UN Climate Change Conference (COP21), sets the target of holding the increase "well below" $2\,^{\circ}\mathrm{C}$ while aiming to minimise the increase to $1.5\,^{\circ}\mathrm{C}$ above pre-industrial levels[37]. Targeting a reduction in anthropogenic CO_2 production will be an unavoidable component in meeting this goal and minimising the temperature increase.

3.2 Net-zero 2050

In June 2019, the UK became the first major economy to legislate a net-zero emission target[38]. In 2021, this was then extended to include the additional targets of running a net-zero power system[39] and to reduce emissions by 78%[40] by the

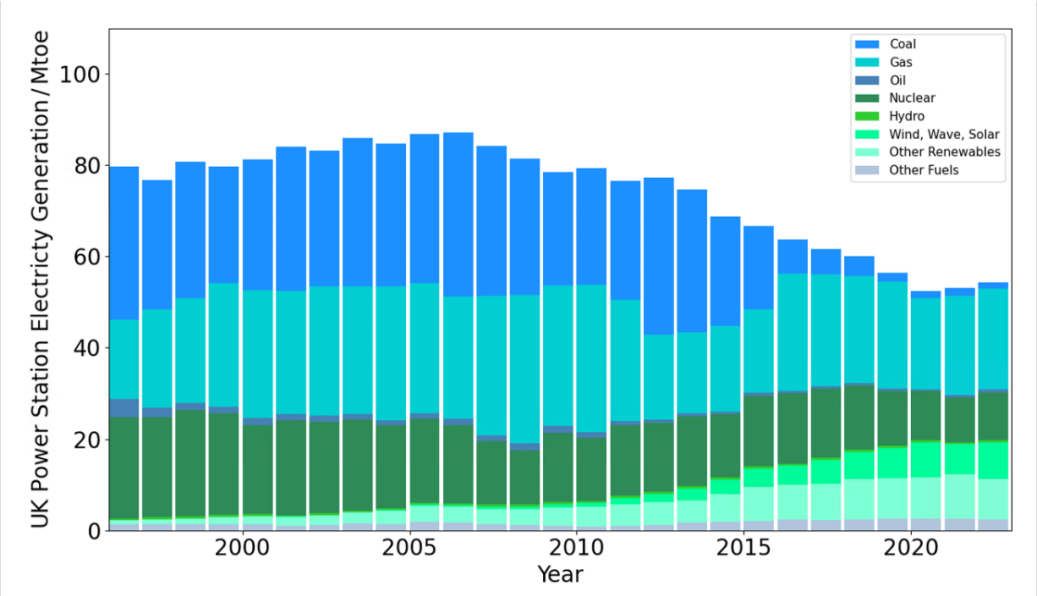


Figure 3.4: Fuel mix for UK electricity generation, 1990-2022 (Million tonnes of oil equivalent)

year 2035. While these are commendable steps to take, a report from the House of Lords published in January 2023 stated that they were 'not convinced that the necessary level of policy detail is in place to achieve these commitments'[41]. In order to achieve these goals, the UK will be required to convert national energy production from a reliance on fossil fuels to renewable energy sources, such as wind and solar as well as nuclear.

From Figures 3.4 and 3.5 it can be seen that the proportion of energy production from nuclear/renewable sources has modestly increased since 1990, while energy generation from coal has decreased substantially, with national CO_2 production reducing in step. It can also be seen that the major contribution to carbon emissions from power production is from natural gas sources. Currently, the UK grid is 40% reliant on gas energy production, and this is primarily owed to its quick response time to energy demand. This exposes a major challenge for a renewable based energy system - production intermittency. Before addressing this issue

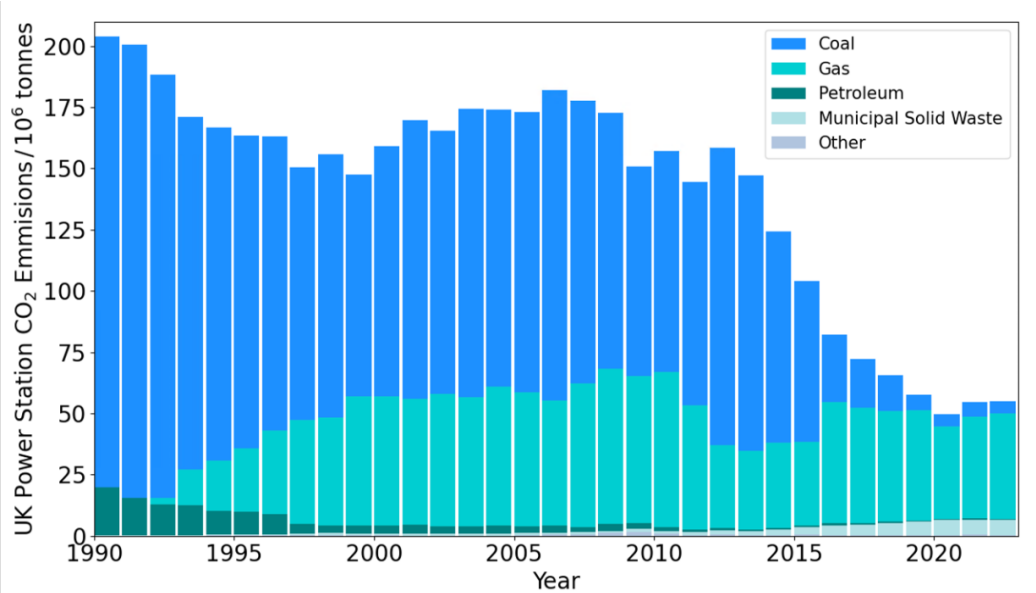


Figure 3.5: Territorial carbon dioxide emissions from power stations, UK, 1990-2022 (MtCO2)

and potential solutions, however, we will briefly outline the possible options for energy production within a renewable based energy system.

3.2.1 Nuclear Power

Nuclear power can be a polarising option for replacing fossil fuels in power production. Routine operation produces essentially zero emission products along with fast, reliable, high power outputs, making it an attractive option for supporting a national grid. Nuclear power stations can also operate with a relatively small geographic footprint, especially relative to their energy production capacity. Unfortunately, there are safety concerns over nuclear power. While it is difficult to accurately quantify the exact scope of nuclear accidents in human history, a 2016 study by Wheatley *et al.*[42] compiles a dataset of 216 accidents or incidents. Within this report they estimate that without a reform in nuclear procedures, there is a 50% chance of a 'Fukushima level' event occurring every 60-150 years. Simi-

lar concerns are posed of nuclear safety by Wu *et al.*[43] especially in the increase of nuclear reactor construction in developing countries. Atop these safety concerns are also issues surrounding storage and disposal of radioactive waste[44] and concerns that both the time and large upfront monetary costs of commissioning/decommissioning nuclear facilities[45] may prohibit it from being a ubiquitous solution.

3.2.2 Wind and Solar Power

Similar to nuclear, wind and solar power both meet the requirement of having zero carbon emission in routine energy production operation[46][47]. They can also be seen as particularly attractive options when they offer means of utilising space that might otherwise go unused (such as deserts solar-farms[48] or offshore wind-farms[49]), in addition to their relatively low operating costs, and their ability to 'democratise' energy production[50]. This idea of 'democratising' energy production, comes from the observation that the geographic regions with the highest energy production potential for particularly solar, typically align with areas of less industrially and economic development. This is due to a range of factors, (including the same climate that makes them so attractive for solar production) but the theory is: as a global shift towards carbon-free is required, it is these currently under-developed regions which should benefit from the associated economic boost.

A similar argument could conversely be touted as a disadvantage of both wind and solar: the geographic restrictions of where the strategies can be employed. Considering this more broadly, the main problem surrounding these production methods can be seen as the reliability of the energy production. This is shown both daily and seasonally in Figure 3.6 and Figure 3.7 using weather data obtained from the London Assembly[51], where the same location's energy production by

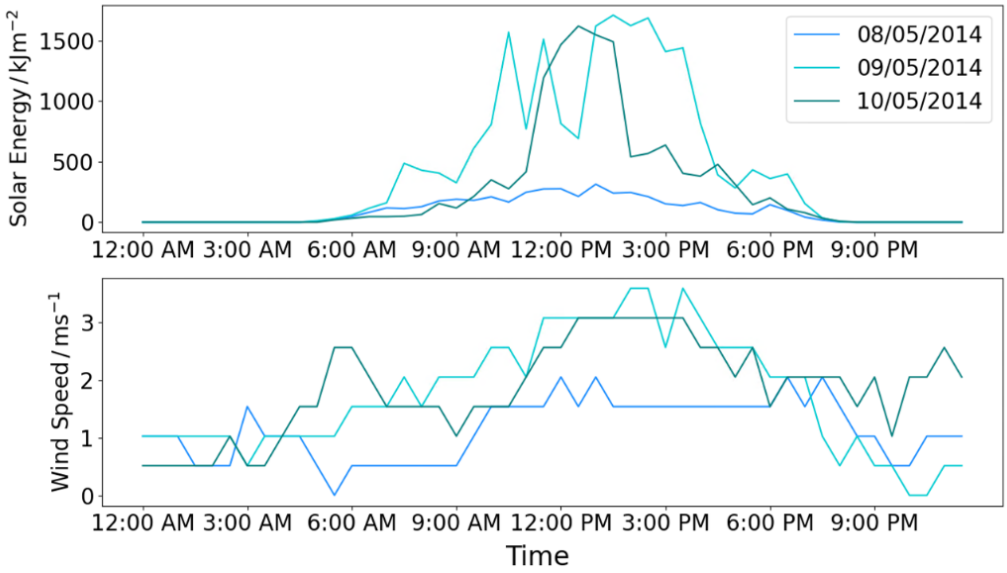


Figure 3.6: Graph showing: (top) variation in solar energy recorded over three consecutive days; and (bottom) the mean wind speed recorded over the same period.

means of wind and solar, is shown to vary significantly across hours, days and months. With such variability present in total energy production capabilities, the key challenge is matching energy demand to the inherent intermittency in supply.

3.3 Energy Storage

The most effective strategy for combating intermittency is energy storage[52][53][54]. The concept is very straightforward: to store the intermittently produced energy when it isn't needed, so that it can be used when it is needed. For this to be an effective strategy, fundamentally all that is required is that the production capacity, across the time-scale of the variability, is large enough that even when offset by losses due to the inefficiency of energy storage, it meets the energy demand. This framing gives a refreshingly optimistic outlook for a zero carbon economy, as

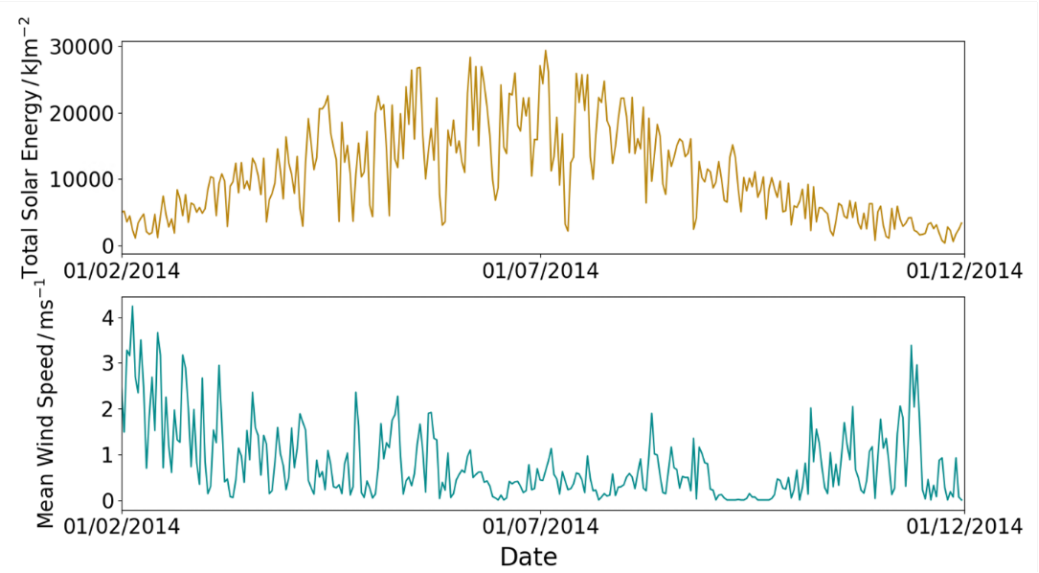


Figure 3.7: Graph showing: (top) total daily solar energy recorded over the course of a year production; and (bottom) mean daily wind speed recorded over the same period.

it requires neither instant-response emission free production, nor extremely high efficiency energy storage, just two 'good enough' combinations of *as-reliable-as-possible* energy production and *as-efficient-as-possible* energy storage. As zero-carbon energy production approaches like nuclear, wind and solar already exist, the current challenge for this zero-carbon energy production is two-fold: increasing the total production capacity of these approaches; and developing the capacity and efficiency of energy storage[55]. While both are areas that need to be targeted the focus of this thesis is on the latter: increasing the total capacity of usable, stored energy.

This mandate of 'increasing total capacity of usable, stored energy', comes itself with some intended nuance, with 'usable' referring to only the energy that can be regenerated once the efficiency of the storage has been accounted for. It is evident that 100% efficiency in energy storage is not a necessity, and the best solution is likely to weigh a balance of many factors, one of which is the efficiency[56].

When evaluating potential energy storage options, it is prudent to consider the time-frame over which the energy needs to be stored, as can affect the importance of some of these different factors.

3.3.1 Storage Time-frame

In terms of energy storage, time-frames are typically split into short-term and long-term storage[57]. Short-term storage is typically considered as seconds-days time frame, while the length of long-term storage can stretch beyond a year. In the context of renewable energy production, a simplified interpretation is that short-term storage might be used to level out the daily fluctuations of energy production and demand, whereas long-term storage would be used to meet, larger, seasonal discrepancies (like storing excess solar energy in the summer, for use in the winter, when potential for production is reduced).

These different use cases, come with different key requirements. Short-term storage will require a fast response time to meet the second by second demands of the grid. For short-term storage, long cycling life span is also a primary consideration, as these quick responses will require robustness for frequent, repeated, use[58]. In contrast, a long-term storage application might be less concerned as to the repeatability of the storage, as the energy might be stored for several months. Instead, a more pressing concern of long-term storage might be: the energy density of the storage; the stability of the energy in its stored state; and perhaps the portability of the stored energy[59][60].

3.3.2 Long-term storage

Many technologies are in development to tackle longer term energy storage, two of the most promising of these are Hydrogen Storage and Ammonia Storage.

These are in no way the breadth of the technologies being developed for long-term storage nor can these summaries cover the topics in any depth, but they have been included regardless for the sake of completeness.

3.3.2.1 Hydrogen Storage

The use of hydrogen is a widely referenced potential solution to zero-carbon, long-term energy storage. The idea is to produce hydrogen by the electrolysis of water (powered by the excess energy generated from a renewable source), store this hydrogen until the energy is required, and then burning the hydrogen to regenerate electricity and produce water as the only product. This circular hydrogen economy is referred to as 'green hydrogen'. The main challenges in the face of this technology and preventing its wide-spread application is in the energy density of the hydrogen storage, the safety of the storage and transportation of hydrogen, and the upfront cost of infrastructure required[61][62].

3.3.2.2 Ammonia Storage

Ammonia is another promising molecule for energy storage, with a few possible methods for utilising the stored energy. The most trivial approach is where the ammonia simply acts as an intermediate energy vector for hydrogen, using ammonia synthesis and cracking to bookend its use, with the main energy generation steps being unchanged from hydrogen storage. In such a case ammonia is simply the form in which the hydrogen can be stored and transported, leveraging the orders of magnitude scale up of global ammonia handling infrastructure[63]. This approach is, however, limited, as use of the produced hydrogen in a fuel cell requires a high purity of product, with a concentration of ammonia as low as $1\,\mathrm{ppm}$ (parts per million). A more nuanced application case uses the ammonia directly in the energy production, where either the cracking of pure ammonia

is used or a 'Dual-Fuel' ammonia-hydrogen mixture is utilised. Such fuels require high operating temperatures (limiting the scope of the applications) but for situations where a steady supply of energy is required for a long duration, such as freight shipping[64], these make very promising candidates. The specific gas fraction of the dual-fuel will come down to considerations of the relative safety of handling, energy storage densities, conversion efficiencies along with many other factors[65].

3.3.3 Short-term storage

For the intra-day energy storage time scale, there are many potential approaches that can be taken. Figure 3.8 shows the forecast for requirement of energy storage capacity up to 2031[66]. It shows a more than 3 fold increase in the expected global storage capacity with an even greater increase required to meet net-zero. In order to provide this energy storage capacity, it is likely that no single storage technology will be the solution. The solution will instead require all manner of technologies be used cooperatively. Two of the major classes of technologies that currently exist and will likely be utilised are Pumped Storage Hydroelectricity (PSH) and Batteries.

3.3.3.1 Pumped Storage Hydroelectricity

PSH aims to store electrical energy in the form of gravitational potential energy. This is done by using the electrical energy to do work on water, typically by pumping it up from a lower reservoir to a higher reservoir. Once done, this water can be left idly until the energy is required, at which point it is released from the higher reservoir back down into the lower reservoir. As it does so, the flow can be used to drive turbines, which in turn drive electrical generators. Such an approach is both simple and effective for meeting the needs of short term storage, as the response

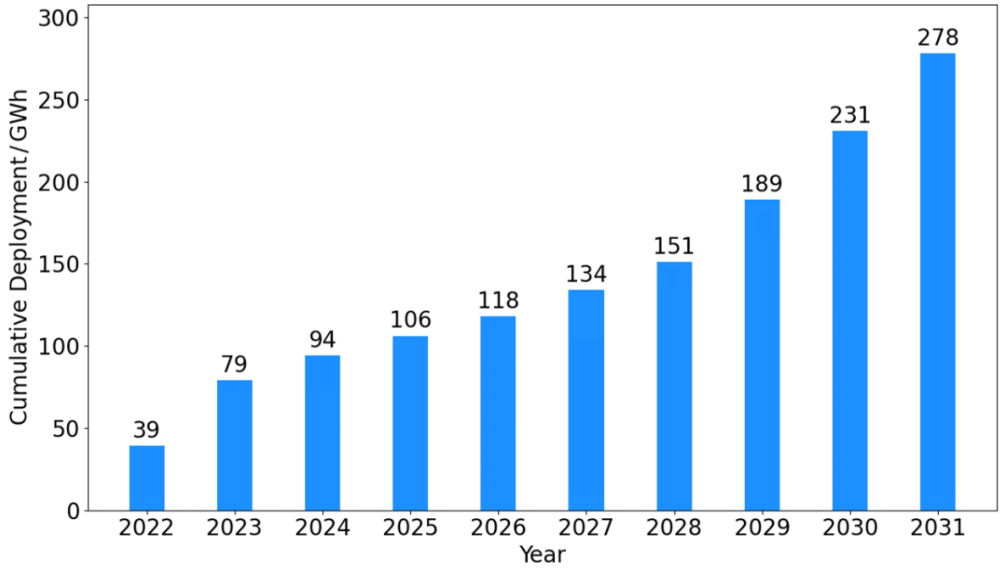


Figure 3.8: Projected cumulative deployment capacity of energy storage market worldwide in 2021, with forecast figures to 2031

time is fast and energy production can be accurately controlled by modulating the flow. The main constraints of PSH are geographical. Such stations require a large change in elevation to be effective, which is not available ubiquitously. Additionally there are ecological considerations, as these PSH stations often involve creation of dams, which can have drastic knock on effects to local ecosystems.

3.3.3.2 Batteries

The other broad category of short-term energy storage technologies is batteries. The lithium-ion battery was first developed by John Goodenough and commercialised by Sony in the 1980s[67]. Ever since the first $LiCoO_2$ battery, batteries have revolutionised all corners of modern technology and research has been ongoing to develop new, better batteries.

3.4 Sodium-ion Batteries

In many cases, the functionality of sodium-ion battery is not drastically different from the original lithium-ion battery. In its most high-level interpretation the difference is only in the substitution the Li ions for Na ions, one row down on the periodic table, along with perhaps some adjustment of the associated transition metals. The consequences of this seemingly trivial substitution, however, is the exploration of a whole range of both similar and different materials as potential battery candidates.

The desire to drive towards sodium-ion batteries from lithium-ion is due to a couple of key issues which surround the supply chain of lithium-ion cathode raw materials, as many of these make use of cobalt-based cathode materials. Figure 3.9 shows the three primary supply chain drives for the sodium-ion alternative. Figure 3.9 (a) shows the relative abundance of the Group 1 metals in the earth's crust and in seawater. Comparing Na to Li, Na is orders of magnitude more abundant than Li and thus a more appealing candidate for large scale battery production. In addition to issues with the abundance of Li, Figure 3.9 (b) and (c) also shows the global distribution of Li and Co, with far from even distributions[68][69]. A sole dependency on Li-ion technology would also be a dependency on the production capacities of these select nations. Furthermore, it is widely reported that the extraction of cobalt out of the Democratic Republic of Congo is severely unethical, employing widespread child labouring in the mining and refinement procedure[70]. In contrast, Na is more abundant within the crust and widely available from the electrolysis of sea water, providing much less scope for supply chain issues when scaling up to high global capacities.

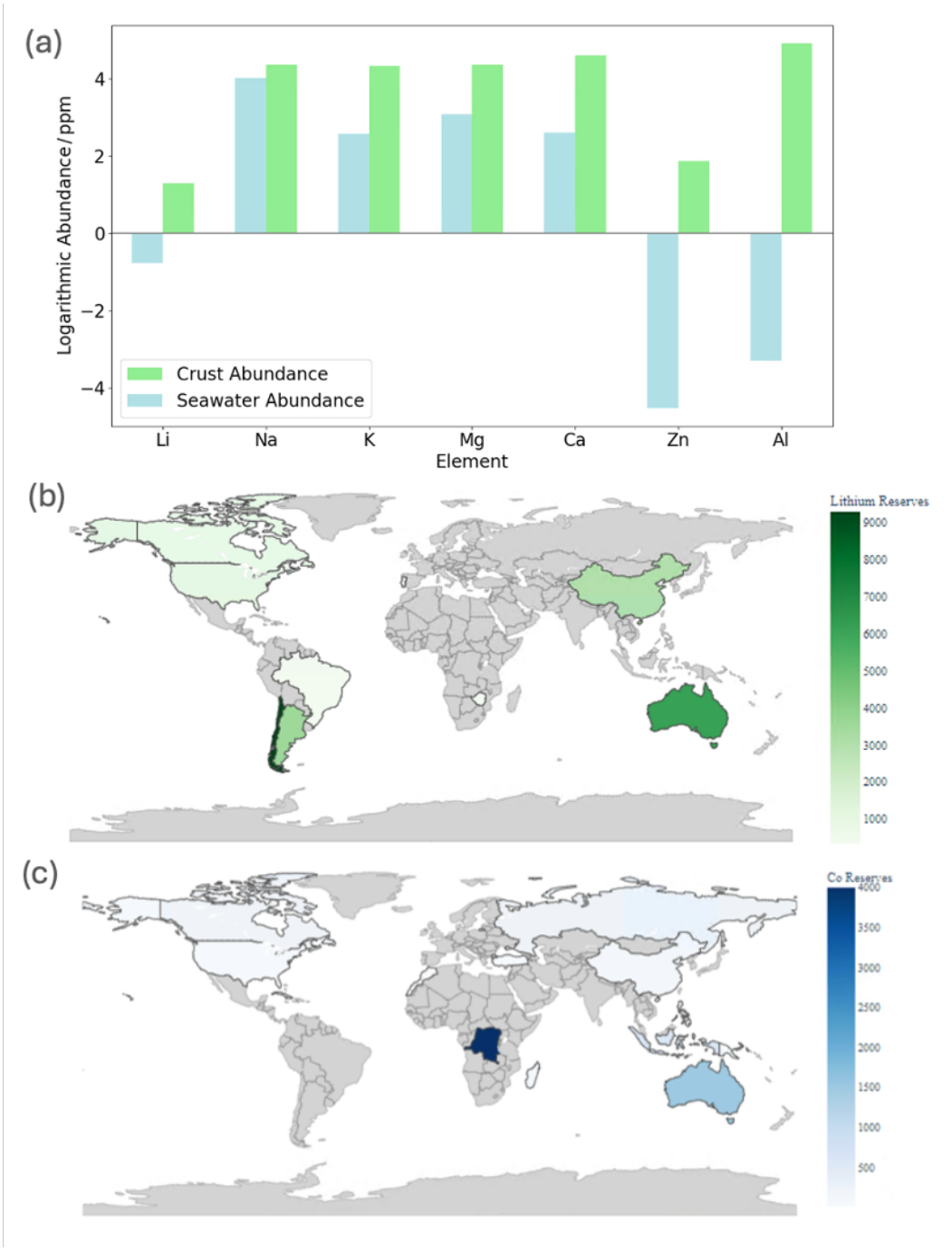


Figure 3.9: (a) Natural crust abundance of group 1 metals. (b) Map of global Lithium reserves. (c) Map of global Cobalt reserves.

3.4.1 How a sodium-ion battery works

Figure 3.10 shows a cartoon schematic of some of the key components that make up the sodium-ion battery. The schematic shows the cathode material, in this case a layered metal oxide (LMO), with the sodium ions (green) intercalated between the transition metal oxide layers (purple). At full discharge, this LMO will be in its thermodynamically most stable state, with a maximised occupancy of sodium. Upon charging, electrical energy is used to do work on the system, oxidising the redox-active transition metals. In tandem the sodium ions are extracted from between the metal-oxide layers to maintain charge balance. The resulting layered structure therefore has a reduced sodium content and is of higher energy.

When the sodium ions are extracted from the cathode, they enter solution in a highly ionic electrolyte. This electrolyte allows a net flow of ions from the cathode through the separator medium, and into the anode. Upon charging the anode undergoes a similar transformation to the cathode, moving from a low energy state to a higher energy state, however this time the low energy state is the one with very low sodium incorporation, and on charging the sodium content is increased. Typically the anode material is a 'hard carbon', a graphitic polymorph of carbon but with notably more disorder and curvature to the layers. This hard carbon material can incorporate the sodium ions by many methods including: deposition in porous cavities; as surface adsorbents; and intercalated between layers. Upon charging electrical work is done to supply electrons and reduce the anode, this in turn increases the incorporation of sodium ions to maintain charge neutrality.

Upon discharge, the reverse of the charging processes occur. The anode is oxidised and the electrons flow through the circuit to reduce the redox-active metals in the cathode. In tandem, the sodium ions flow from anode to cathode, through the electrolyte to maintain the charge balance. The flow of electrons through the

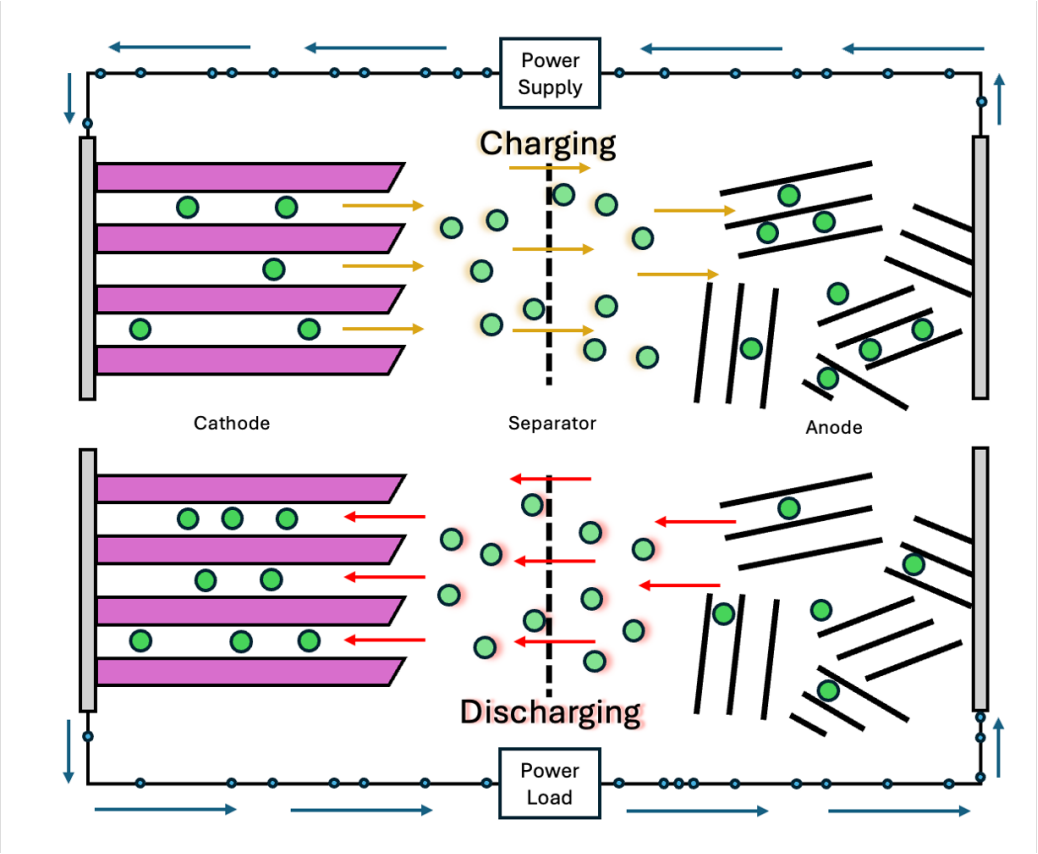


Figure 3.10: Cartoon schematic of the components of a Sodium-ion battery

circuit is the electrical energy provided by the battery.

3.4.2 Current state of Sodium-ion research

Within the overarching category of cathode research there is still much diversity in the materials and material-classes being investigated. Some example materials explored in a review by Li *et al.*[71] are: Prussian Blue (PB) and Prussian Blue Analogues (PBAs); Polyanionic-Type Cathodes; Phosphates; Fluorophosphates; and the layered metal oxides mentioned already. All these materials have various advantages and challenges associated with improving their performance but, as highlighted already, the energy storage challenge is a matter of developing global

storage capacity to as great an extent as possible. As such, development into as many different avenues and storage technologies as possible will be key, and the solution will likely be one of leveraging many different battery technologies.

As the battery is a dynamic system optimising the battery performance is not about fixing a single component, but optimising the interplay of all the components. The work of this thesis is focused upon the layered metal oxide based cathode material, and so we will outline the challenges specific to these.

Currently, for the case of a sodium-ion battery using a layered metal oxide based cathode, the bottleneck in the performance of sodium-ion battery technology, particularly compared to lithium-ion performance, is in the cycle-life of the batteries. In these systems, this cycle life limitation stems from degradation to the cathode materials, reducing the specific capacity of the cells over time.

There are many possible causes of degradation and it is inevitable that all batteries will experience some losses of capacity with successive use, but the rate of degradation is currently prohibitively high for many commercial applications. One such cause is the occurrence of phase transitions, upon cycling. These phase transitions occur when sodium is extracted. If there is a kinetically accessible lower energy structure, the layers can shift to adopt this structure. While these changes are often reversible and so do not directly hinder the functionality of the cathode, these shifts are often accompanied with volumetric changes to the structural unit cell. Such changes occurring locally within particles, will create strain, which over repeated cycling might lead to cracking but, more generally, will lead to changes in the overall morphology of the bulk cathode. What the direct consequence of the presence of such cracks would be on the electrochemical performance is not immediately obvious. One could reasonably assume that the cracking might lead to an increase in surface area through which sodium could be

extracted and improved performance. Unfortunately, however, it has been shown that the occurrence of particulate cracking is instead highly correlated with the loss of capacity and so this it is likely that this is the primary cause of the cell performance degradation[72].

While the specific cause for capacity fade is not certain, it is clear that the cathode performance, and how it evolves over time, is a function of factors across many length scales. It is clear that to fully understand the functionality of the cathode the following length scales must be considered:

- Unit-cell level atomic interaction, occurring at the angstrom level.
- Particulate level structure, occurring at the micron level.
- The statistical interactions of cathode-electrolyte in the bulk, occurring at the millimeter scale.

All these length scales contribute to a complete understanding of the material and its function as a cathode, however, all are met with challenge and even the most simple structures in pristine, as synthesised states are not fully understood.

The aim of this thesis is to develop methodologies that can address the challenges associated with modelling these structures effectively. We will combine structural studies across all these critical length scales, using a multi-modal approach, leveraging electron, x-ray and neutron diffraction techniques, in order to unify understanding of these cathode materials. Such wide-scale understanding would be invaluable in developing and improving battery performance.

3.5 Thesis Outline

The means by which this goal of achieving an improved understanding of cathode materials, will be approached is outlined within this section.

The following chapter covers the prerequisite knowledge of the layered cathode materials that will be primarily investigated within (Chapter 4). It outlines the main structural motifs associated with these materials as well as the limitations of these current models.

The subsequent two chapters then address the background theory of both the experimental techniques (Chapter 5) and the computational techniques (Chapter 6) that are either utilised or built upon within the later work. They cover the fundamentals of diffraction in crystalline systems as well as highlighting the differences between electron, x-ray and neutron diffraction, motivating the advantage of using the three approaches in tandem to extract different information. The primary resources used as reference for the experimental chapter were 'Transmission Electron Microscopy' by Williams and Carter[73], and 'Crystallography in a Nutshell' by Ripoll and Cano[74]. There were no generally overarching academic references for the Computational Methods Chapter, however some useful resources were: Grant Sanderson's video-series on Machine Learning[75] and the Sklearn/Skimage/Tensorflow documentation[76][77][78].

Chapter 7 summarises all of the experimental methodologies and specifications of instruments and equipment are also provided. Following this there are the four results chapters.

The first of these (Chapter 8) documents the development of a data processing workflow for electron diffraction data. The work presented was motivated by the overwhelming quantities of data produced by electron diffraction experiments

and the need for autonomous/semi-autonomous routines to aid in the processing and interpretation of all the data. The majority of the content of this chapter was published in the NPJ Computation Materials paper: "Versatile domain mapping of scanning electron nanobeam diffraction datasets utilising variational autoencoders"[79].

Chapter 9 then uses this workflow, along with other electron diffraction techniques, powder x-ray diffraction, neutron diffraction, and computational studies to investigate the ordering of ions present within individual layers of a standard phase of cathode material across a compositional series. This chapter addresses some of the intra-layer nuances noted in the initial Structural Chapter as limitations of current models.

This approach then extends into a study of the ordering across multiple layers within these samples (Chapter 10). This chapter highlights the utility of having multiple probing approaches, each with different advantages, to aid in the understanding of these subtle orderings.

The final results chapter (Chapter 11) further leverages the VAE workflow alongside the structural models of the compositional series developed in the previous chapter, but now focuses on investigating the electronics of the ions present, rather than their positions within the structure. This chapter emphasises the strength of having expressive structural models for gaining new understanding of the electronics of these systems and attempts to achieve this while providing a spatial component to the analysis.

The thesis then concludes by summarising the findings of the work presented and addressing the scope for furthering these investigations.

4 | Layered Cathode Structures

4.1 Standard Phases of Na_xMO_2-type systems

Within this thesis we primarily concern ourselves with the layered structures of Na_xMO_2-type systems. These structures are built up along the z-direction and consist of layers of edge-sharing MO_6 octahedra, which form $(MO_2)_n$ sheets. The Na^+ ions are intercalated between successive layers of these sheets. The coordination environment of these Na^+ ions depends on the registration of the layers of MO_6 octahedra upon one another and as a consequence of this, leads to the four aristotype phases of Na_xMO_2-type systems which, in the notation developed by Delmas[80], are referred to as P2, O2, P3 and O3.

4.1.1 Stacking Hexagonal oxygen layers

The structures of these standard phases can be understood from the stacking of hexagonal layers of oxygen ions and then populating the successive layers of interstitial sites with M^{k+} ions and Na^+ ions.

Consider stacking hexagonal oxygen ion layers upon one another: there are three register locations in the x, y plane that the next layer can occupy, these are labelled A, B and C. To generate the standard phases, we can keep stacking more of these A, B and C layers until the sequence repeats itself.

Consider the two layer case, this gives three possibilities AA, AB and AC, shown in Figure 4.1. The first stacking type: AA, gives 6 identical prismatic interstitial sites. The second and third stackings: AB and AC, both give 3 octahedral sites (blue) and 6 tetrahedral sites (green), but at different locations within the hexagonal cell.

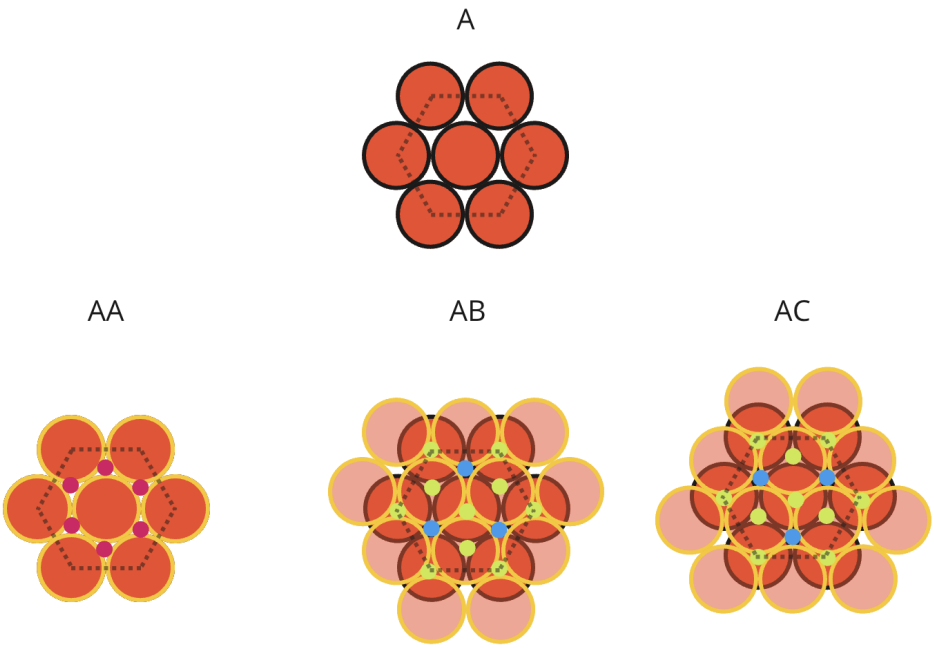


Figure 4.1: Top: Single hexagonal oxygen ion layer, the hexagonal cell is shown with a dashed hexagon. Bottom: Stacking of a second hexagonal oxygen ion layer (orange border) upon a first (black border). The different types of interstitial sites, and their respective locations in the hexagonal cell are shown with coloured dots. These sites are prismatic (pink), octahedral (blue), and tetrahedral (green)

These interstitial sites are the three main coordination environments that are encountered in these layered systems. Generally, the tetrahedral sites are too small to be stabilised by any of the ions present in these systems[80], so practically only the prismatic and octahedral sites are available to be populated.

The transition metal, M^{k+}, ions are more stable in the octahedral coordination, so require AB or AC type stacking. The result of populating all the octahedral sites of either the AB or AC oxygen stacking is the layers of edge-sharing MO_6 octahedra discussed earlier. The AB and AC stackings result in MO_6 octahedra that point in opposite directions. This is highlighted in Figure 4.2

For now, let's assume that AB and AC are equivalent, and continue stacking on AB.

The stacking possibilities of the next oxygen layer, after AB, are shown in Figure 4.3. These lead to the three layer stackings: ABA, ABB, and ABC. It can be seen that same types of sites are created in this stacking as the previous one.

The first of these stackings: ABA, gives us our first repeating stacking, as we have returned to A as the bottom layer of our next $(MO_2)_n$ sheet. This could continue to repeat ABABAB... alternating AB transition metal layers and BA sodium layers. As noted earlier, the Na^+ ions cannot stabilise the tetrahedral holes in the BA layer, so would all be required to occupy the octahedral holes. Comparing the location of the octahedral holes in the BA layer to the occupied octahedral holes of the AB layer below it can be seen that AB-type stacking would result in all the Na^+ ions and all the transition metal ions being localised in the same columns in the x, y plane. Such a configuration would produce strong coloumbic repulsion, so this is not a commonly observed stacking and it is not one of the main phases.

In contrast, the two stackings ABB and ABC can both allow for stable phases to form. ABB has the Na^+ ions in prismatic sites, ABC has the Na^+ ions in octahedral

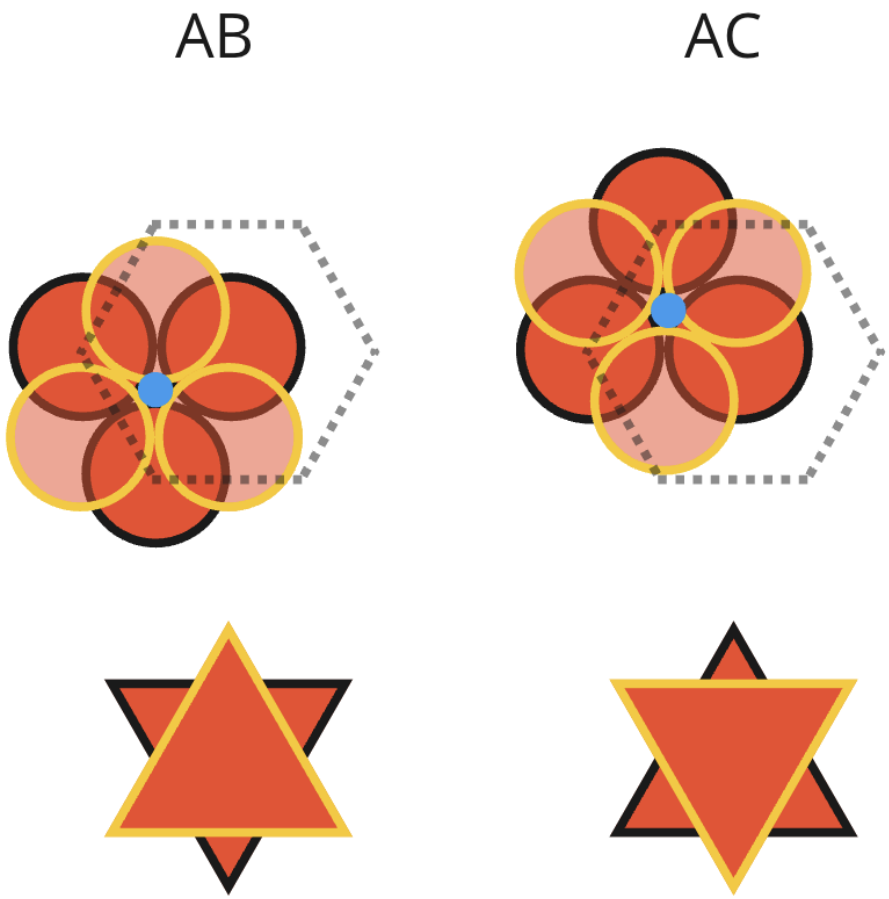


Figure 4.2: A single octahedral site for AB and AC type stacking. The directionality of the octahedra is highlighted by the cartoon below, with the top layer (orange border) pointing up for AB and down for AC.

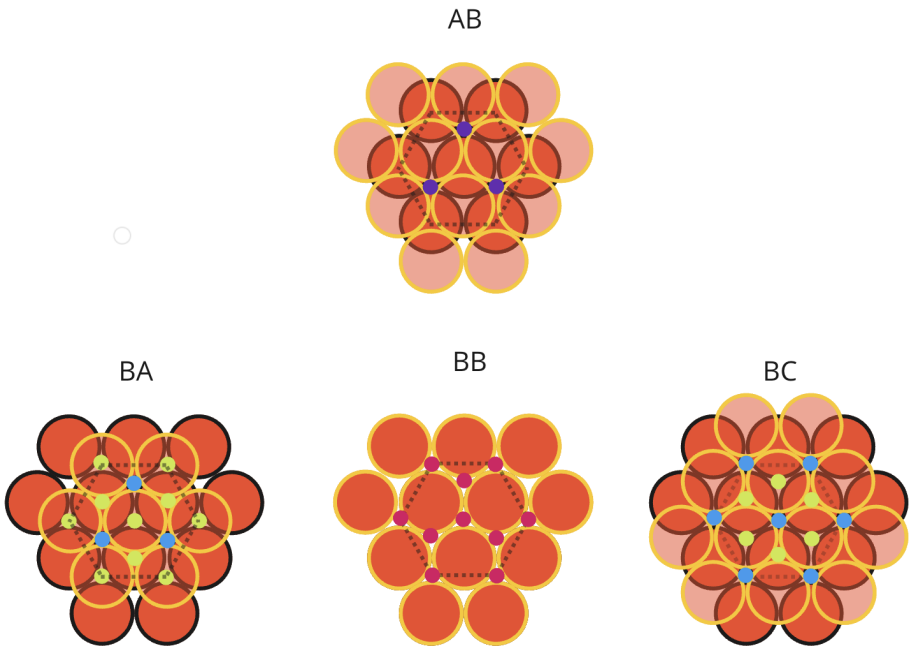


Figure 4.3: Top: Specific two layer hexagonal oxygen ion layer stacking, AB, with the M^{k+} interstitials (purple). Bottom: Stacking of a third hexagonal oxygen ion layer (orange border) upon second layer (black border). The different types of interstitial sites, and their respective locations in the hexagonal cell are shown with coloured dots. These sites are prismatic (pink), octahedral (blue), and tetrahedral (green)

sites offset from the transition metal ions.

If we take the ABB stacking, the next metal layer (again, requiring octahedral interstitials) can be ABBA or ABBC. If ABBA, the subsequent layer can give: ABBAA, ABBAB or ABBAC stacking. The ABBAA structure can now to repeat, it is characterised by an AB metal layer, BB sodium layer, BA metal layer, AA sodium layer, and is set up to continue for another AB metal layer.

This structure is denoted as the P2 structure. The 'P' is for prismatic, as the AA and BB sodium layers give the prismatic Na^+ interstitial sites, and the '2' is for the number of unique transition metal (AB and BA)/sodium (AA and BB) layers.

Neither ABBAB nor ABBAC are considered to give a standard phase, as these structures have a BB sodium layer and either an AB or AC sodium layer. Such structures will therefore contain a mixture of prismatic and octahedral sodium coordinations. It is worth explicitly noting that these stackings can exist and are not only observed but are an incredibly important class of structure for real world application as cathode materials. They are, however, considered to be mixtures of two or more of the standard phases we are currently identifying, so can be put aside for now.

Returning to ABB, if instead of ABBA, we look at the ABBC-type stacking, and again requiring that only prismatic sodium coordinations are present in the structure, this needs to be built up to ABBCCA stacking before it repeats. This structure is denoted P3, with the '3' now due to the three transition metal layers (AB, BC and CA) and the three sodium metal layers (AA, BB, CC). This gives P2 and P3 as the two standard prismatic phases.

Returning once again to the three layer case, if the first three layers are ABC, the first sodium layer has sodium interstitials in the octahedral sites. Again, the next transition metal layer requires octahedral coordination so can form ABCB or

ABCA stacking. Much like the P2 phase, ABCB can now repeat. This gives AB and CB transition metal layers and BC and BA sodium layers. This is denoted as the O2 phase, with 'O' now referring to the octahedral coordination.

Instead taking the ABCA stacking requires an additional intermediate metal and sodium layer before it can repeat, in this case repeating as ABCABC. This gives the final standard phase, which is denoted as the O3 phase.

This gives the four stackings associated to the main phases: ABBA (P2), ABCB (O2), ABBCCA (P3), and ABCABC (O3). This oxygen packing model is just a model, in the real systems the oxygens are not close packed. Deviation is especially evident in the sodium layer, where there is an extreme elongation along the z axis. The most notable result of this is that what has thus far been referred to as an octahedral sodium ion coordination is more accurately a trigonal anti-prismatic coordination environment. Regardless, the model is helpful for deriving the main phases, which are shown in Figure 4.4

The arrows in Figure 4.4 indicate the 'directionality' of the metal octahedra, as referenced in Figure 4.2. As these octahedra cannot change direction without breaking strong M−O bonds, these arrows highlight the relationship between these main phases. The P2 and O2 can interchange by gliding the $(MO_2)_n$ sheets[81], as can the P3 and O3. These glides can occur as they only require breaking the weaker Na−O bonds[82]. Conversely, the P2 cannot be distorted to form the P3, nor can the O2 to form the O3. These relationships, along with the inaccuracy of referring to the O2/O3 sodium coordinations as 'octahedral', highlight limitations of the Delmas notation. An alternative notation has been suggested: replacing 'O' with 'A' to more accurately reflect the coordination; and replacing 2/3 with -/+ to reflect the parity of subsequent MO_6 layer directions[83]. In this notation P2 would become P^-, P3 would become P^+, O2 would become A^-, and O3 would

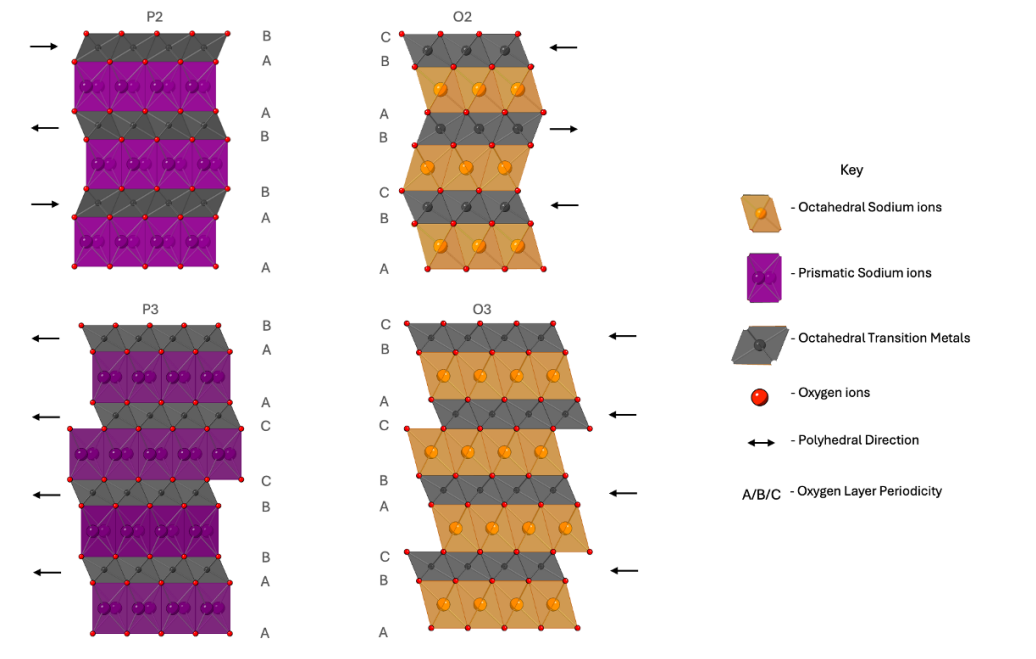


Figure 4.4: The four stackings associated to the main phases: ABBA (P2), ABCB (O2), ABBCCA (P3), and ABCABC (O3)

become A^+. This notation has the advantage of representing the coordination more appropriately and better highlighting why some phases can inter-convert, while others can't. Despite the presence of this improved notation, it is yet to be widely adopted by the literature, so we will proceed with the Delmas notation.

4.1.2 Deviation from these phases

As noted in the previous discussion, mixed phases can occur with other A/B/C stackings of the oxygen ions. When considering all the possible stackings, it is perhaps surprising that four phases can be used to even approximately describe these systems. It is true that in the real systems there is more complexity than these four phases suggest, but under the thermodynamic control of high temperature synthesis used for producing these structures, the energetic advantage of the local symmetry present in these main phases can make them quite prevalent.

This becomes less true when entering the more kinetic regime of electrochemical cycling[84], in these situations, using these small crystallographic units becomes less appropriate as long and short range disorder start to enter into the system.

Even in the pristine, post-synthesis state, these phases do only approximately describe these layered systems, Even for one of these standard phase materials, the structure has a lot of subtle nuance that can dramatically impact the efficacy of the material for its function as a cathode material[85].

In this thesis we will look, in depth, at one of these standard phases, the P2. We investigate some of the key structural nuances that are overlooked by the simple crystallographic model shown earlier and consider their relevance to battery performance. This thesis intends to highlight the amount of detail and complexity that occurs in these layered materials and lay the foundations for applying this level of rigour to future, even more complex systems.

4.2 The P2 phase

As stated, the focus of this thesis is on the P2 phase. This phase is characterised by the two repeating metal layers and sodium layers. The metal layers have AB and BA oxygen layers and the sodium layers have AA and BB oxygen layers forming the prismatic interstitial sites. The P2 phase is typically expressed using a $P6_3/mmc$ cell with a single metal site, a single oxygen site, and two partially occupied sodium sites. The functional application of P2 is as a high power cathode material[86].

4.2.1 Sodium Mobility

Neighbouring prismatic sodium sites share large rectangular faces. This allows for a low energy sodium mobility pathway straight through the middle of the rectangular face[87]. This is in contrast to the phases with octahedral coordination where, recalling Figure 4.3, neighbouring octahedral sites are joined by high energy intermediate tetrahedral sites[88][89]. Further to this, they are joined to these tetrahedral sites through triangular faces. Thus, for sodium diffusion to occur along the lowest energy trajectory, this would require diffusing approximately through the centre of these faces[90]. This leads to the trajectory having components both perpendicular and parallel to the sodium layer, rather than in-plane trajectory for the prismatic sites. It is these considerations that make the prismatic phases more suitable for high power applications, as power is quantified in Watts (W), which are Joules per second ($\mathrm{J/s}$). As lower energy sodium mobility pathways exist in the prismatic structures, these can be exploited for faster transfer of ions. As the output of electrical energy is proportional to the number of ions extracted if the same amount of ions can be extracted in a shorter period of time, the result is a higher power output. These diffusion pathways are shown schematically in Figure 4.5

The functionality trade off for the high power of the P2 is the energy density[91]. As mentioned, the output of electrical energy is proportional to the number of ions extracted, a typical P2 structure has the structural formula $Na_{2/3}MO_2$. In contrast an O3 structure typically has the structural formula $NaMO_2$. This can be understood by again using the oxygen stacking model, where the occurrence of one of the standard phases rather than another is primarily a concern of the relative energy of the systems.

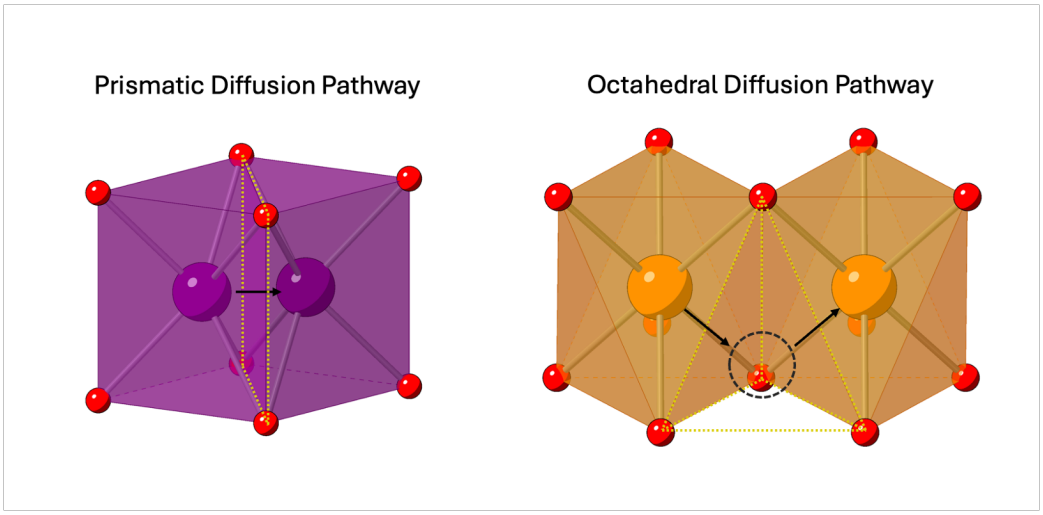


Figure 4.5: Diffusion Pathways

4.2.2 Sodium Sites

It can be observed that in the P2 structure, the transition metal sites (in the octahedral sites of oxygen stackings AB and BA) will occupy columns in the x, y plane (Figure 4.3). The result is that a prismatic sodium layer (AA or BB) between these metal layers, will no longer have 6 equivalent prismatic sites. There will be 3 sites that occupy the same atomic columns as the transition metals and 3 that do not. This is the origin of the two sites in the $P6_3/mmc$ unit cell. The result of having different sites is that those within the transition metal columns are, on average, of slightly higher energy than a similar site not within the column due to the coloumbic repulsion between the ions[92]. The total energy is a consideration of all the electronic interactions of the structure however, so depending on the configuration of other sodium ions around a site, these 'higher energy' sites may or may not be populated. It is found that when more than a two-third stoichiometeric population of sodium is present the lowest possible energy sodium configuration using a prismatic sites is higher in energy than it would be in octahedral coordination.

An understanding of this can be obtained by considering the extreme, where there is a 1:1 M:Na ratio. It can be seen that the exact locations of the three prismatic sodium ions can be identically achieved using the anti-prismatic sites of the O3 stacking. This 'octahedral' coordination will be lower energy because of the reduced coloumbic repulsion between oxygen ions of the anti-prismatic stacking compared to the prismatic.

Identifying the effects of these orderings of the sodium ions across the possible prismatic sites, and the knock on effects to structure and performance, is one of the primary focuses of this thesis.

4.2.3 Sodium Ordering - The Large Zig-Zag

The currently accepted arrangement for sodium ions in the $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ P2 structure is known as the large zig-zag (LZZ) and was first characterised by Meng et al.[93]. This structure requires a $2\sqrt{3}$ supercell of the standard $P6_3/mmc$ cell. The LZZ has 6 sodium ions per unit cell located on the 'low energy sites' surrounded by 2 sodium ions on the higher energy sites. When viewed over multiple unit cells these high energy sodium ions appear to form large zig zag structures, hence the name, but viewing of more cells makes it clear that these are more accurately described as large hexagonal units. A schematic of the LZZ structure is shown in Figure 4.6.

The large zig-zag model was found to be the lowest energy configuration for a DFT study of possible sodium ordering structures in $Na_{2/3}CoO_2$. The nature of these sodium orderings is, however, that the energy difference between different configurations is rather small. This when considered alongside the frequent development of new pseudo-potentials for DFT calculation[94] and limited direct corroboration with experimental evidence suggests that the LZZ is perhaps not

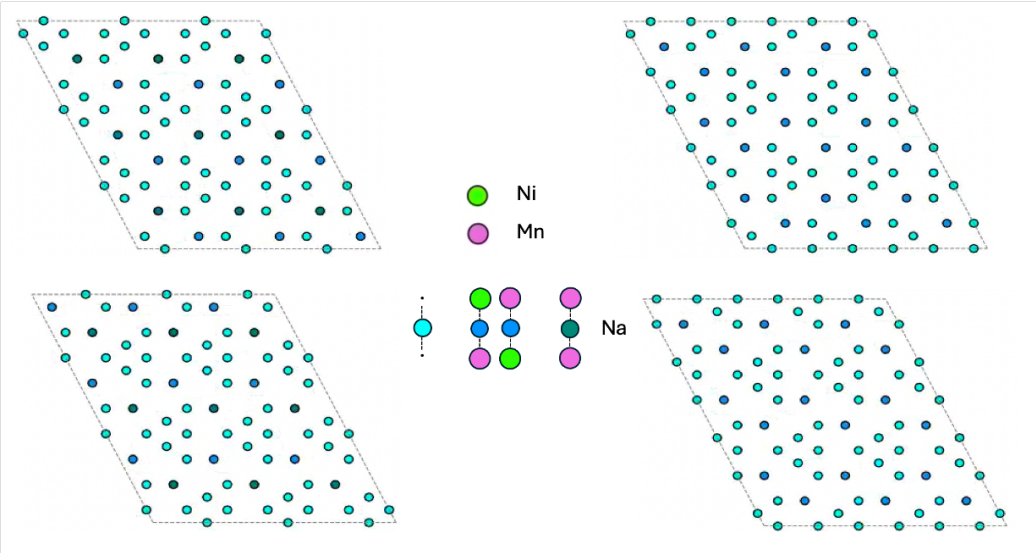


Figure 4.6: Large ZigZag sodium ordering pattern for both sodium layers in the P2 unit cell. The left and the right orderings differ by the occupancies of the transition metal sites in the intermediate layers as the left ordering contains a mixture of Mn – Na – Mn and Ni – Na – Mn sites, whereas the right only contains Ni – Na – Mn sites.

as complete a description as it could be. It is not uncommon in the literature to observe super structure diffraction peaks that have been assigned to the LZZ ordering, but it is fairly uncommon to see these peaks actually fit using an atomic model[95][96][97][92]. Work has already been done by Bassey et. al.[98] acknowledging the limitations of the LZZ and attempting to produce models that are more representative of experimental data.

Further to this, neither the LZZ nor Bassey's model attempt to discuss ordering between layers. Both the in-plane and out of plane ordering of the sodium is a topic that we will investigate at length within this thesis.

4.2.4 Transition Metal Sites

In addition to the sodium ordering, if there are a combination of transition metal ions present, it is possible for these to order as well. This ordering can occur 'in-plane', occupying a periodic arrangement of the octahedral holes within a layer, but also between layers, where the periodic arrangement in one layer has a consistent relationship to the arrangement in a neighbouring layer. A standard P2 composition that is widely explored in this thesis is $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$.

This structure has a well reported in-plane honeycomb ordering[99][85], where each of the Ni^{2+} ions are surrounded by 6 Mn^{4+}. The exact nature of the out of plane ordering is less well understood. Similarly to the stacking of oxygen layers, two successive transition metal layers can have three separate registrations for where the Ni ions are located which again can also be denoted as A, B and C. The literature reports that this $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ P2 structure orders the transition metals using an AB type stacking. This is, however, likely a simplification, as entirely consistent out of plane ordering is unlikely. This understanding has recently been expanded upon by Gabriel et. al.[100] presenting evidence of ABC-type stacking faults within these structures which is consistent with the investigation detail within this thesis.

4.2.5 Disorder and distortions

Another nuance that is present in the real structures not accounted for in the standard phase is the disorder present in the system. In the diffraction data, it is possible to observe a change in peak shape for some specific families of reflections. This is characterised by anisotropic broadening of the $(10l)$ reflections giving 'volcano-shaped' peaks. In the literature these tend to not be accounted for explicitly within the structural model and the deviation is simply noted and at-

tributed to 'stacking faults' (some deviation from the crystallographic periodicity described by the standard phase model)[101]. Pfeiffer et. al. model these peaks with the inclusion of localised P3-type stacking faults[102] and within this thesis, through combined x-ray and electron microscopy study, we obtain evidence that corroborates the presence of such P3 phases.

Regardless of their origin, the presence of these broadened peaks highlights perhaps the most limiting factor of the $P6_3/mmc$ cell: it is too ordered. These structures are so challenging as they appear to have a degree of 'ordered-disorder'. These 'volcano-peaks' are indicative of some form of disorder or distortion occurring, but occurring only for one family of reflections and in a way that is consistent from sample to sample. The limitation of the current P2 model, is that is has assumed regular stacking of hexagonal oxygen layers. We have already shown that the stacking is distorted in the z direction, with a significant elongation of the inter-layer spacing for the Na layer, it is therefore plausible that there might be distortions in the x, y direction as well. This could be x, y distortions *within* the oxygen layers, as the oxygen ions within the layers deviate from perfectly hexagonal arrangements. Alternatively the deviation could be *between* oxygen layers, as the registration of different A layers/B layers/C layers are no longer exactly the same positions in the x,y plane. The former deviations, within the layers, are already documented numerously within the literature, in the form of 'prime' phases. One such example phase is the P2', where the deviation from perfectly hexagonal oxygen layers leads to $Cmcm$ space group symmetry rather than $P6_3/mmc$[103].

4.2.6 Morphology

In addition to direct structural limitations of the standard phases, there is as a morphological component to these structures that are rarely fully explored. The models assume infinite periodic crystal lattices, but these are fundamentally not

correct. Despite the magnitude of number of unit cells in a sample being large enough that the distinction between large and infinite is not relevant, the finite size of the structures is relevant due to the dependence of cathode function on interfaces. The functionality of the materials as active cathode material requires that sodium ions be extracted and reintercalated back into these structures. This means that, in order for it to function at maximum efficacy, pathways must be in place such that a sodium ion located in the centre of a particle can reach the edge. This makes the relative position and orientation of the different layers across the sample important. Attempting to identify how layers are distributed across a particle is one of the focuses of this thesis.

5 | Experimental Techniques

5.1 Introduction to structure probing techniques

In order to better understand how cathode materials such as the P2 phase function, degrade and fail, it is important to understand their atomic structure across many different length scales. Achieving this understanding is no trivial task, but there are many physical phenomena that can be leveraged to help provide an understanding of structure. One such phenomena is the scattering of radiation.

When a propagating waves meet a non-uniform disturbance this results in a change in the propagation of the wave, in general, this process is referred to as scattering. The exact change to the propagation varies upon the nature of the wave and the nature of the disturbance. The nature of the wave is dependent on the type of radiation and its properties, such as frequency and amplitude. The nature of the disturbance, however, is a combination of the type of radiation and some specific function of the atomic structure of the sample.

As different radiation types probe different functions of atomic structure, the information available depends on the exact nature of that function. The result of this is that often different types of radiation will provide different insight into the atomic structure and a combination of techniques is thus required to overcome the bias of individual techniques and get a full understanding of the structure.

5.1.1 Overview

Three such types of radiation that can be used for scattering-based investigations are electrons, x-rays and neutrons. Initially, we will discuss the considerations of scattering that will be consistent across all three types of radiation. From there we

will then look at the individual considerations of each radiation type that lead to different information about a structure being available. With this covered we will then be able to move on to a common practical application of scattering experiments - diffraction.

5.2 Scattering

As initially stated, scattering occurs when an incident wave interacts with a disturbance that alters its resultant propagation. For this thesis we will consider the situation where the incident wave is a plane wave and the scattering can be considered as a spherical wave. A plane wave is a wave propagating in a constant, single, direction with an equation whose amplitude is not inversely dependent in the distance from the origin of the wave. The equation of a plane wave, $\psi_p(\mathbf{r}, t)$, in direction $\mathbf{r}$ at time t is given by Equation 5.1, where $\frac{1}{\sqrt{V}}$ is a normalisation factor, $\mathbf{k}$ is the wave vector and ω is the angular frequency.

$$\psi_p(\mathbf{r}, t) = \frac{1}{\sqrt{V}} \mathrm{e}^{i(\mathbf{k}.\mathbf{r} - \omega t)} \tag{5.1}$$

A spherical wave, in contrast, is a wave that propagates out from a centre at an even rate in all directions forming a spherical distribution of peaks and troughs.

A spherical wave, $\psi_s(\mathbf{r}, t)$, centred at the origin can be similarly given by Equation 5.2.

$$\psi_s(r, t) = \frac{1}{\sqrt{V}} \frac{\mathrm{e}^{i(\mathbf{k}.\mathbf{r} - \omega t)}}{r} \tag{5.2}$$

At large distances from the centre, the wave along a given vector from the centre can be approximated as a plane wave, and this approximation will be used later in

the discussion.

As well as the transition from plane wave to spherical wave, there are some other changes that can also occur as a result of scattering, such as changes in phase or loss of energy.

5.2.1 Coherent Scattering

If the resultant wave maintains a consistent phase relationship with the incident wave as a result of scattering, the scattering is referred to as 'coherent'. If, alternatively, the phase relationship between scattered wave and incident wave can take a distribution of values, the scattering is 'incoherent'. Whether scattering is coherent or incoherent is determined by the underlying process causing the scattering, and so is highly dependent on both the radiation and the scattering centre, but often it is possible for both situations to occur. This distinction between coherent and incoherent scattering has important consequences when multiple scattering events occur, as it determines how two such waves would then interact. If coherent scattering occurs, resultant waves can interfere constructively and mathematically the intensity is just a sum of the intensities of the individual waves. In contrast, when incoherently scattered waves interfere, the intensity can only be evaluated by explicitly calculating it from the resulting combined wave function.

5.2.2 Elastic Scattering

If there is no change of energy between incident and scattered radiation, then the scattering is referred to as 'elastic', otherwise, if energy is lost, it is referred to as 'inelastic'. As with coherence, the occurrence of both elastic and inelastic scattering depends on specific experimental factors and, as before, often both occur in the same experiment.

Depending on the combination of energy and phase changes, a single scattering event will be a specific combination of coherent/incoherent and elastic/inelastic. The exact type of the scattering is an important consideration in these experiments as only specific types of scattering will be suitable for specific experiments. For diffraction experiments (which this section is ultimately building to) scattering is both elastic and coherent, as such, this case will be the main focus of the further discussion.

5.2.3 Scattering Factors and Cross sections

In order to use scattering to understand structure, we need to understand how the scattering will effect the intensity of the radiation in different directions.

Stated explicitly, for the desired case of interest (a plane wave scattered as a spherical wave) we need to be able to calculate the equation for the scattered wave at a general position $\mathbf{r}$, at time t and evaluate the intensity.

A primary factor in determining the intensity at a given point is to calculate how much scattering is expected to occur, as it is to be expected that some of the incident radiation might pass a scattering centre unperturbed.

This can be quantified using the idea of a scattering cross section, which is the effective area around a scattering centre where it can will cause scattering. This scattering cross-section, along with a constant of proportionality and some radiation factors, can be used to find the amount of radiation that will be scattered in a given direction. This proportionality constant is referred to as the scattering length.

It is worth noting that in almost all situations, however, the scattering is not isotropic and there is an angular dependence to the cross section which formally requires the use of a differential cross section. For the specific case of coherent

scattering, however, the differential cross section is just the square of the scattering length. This means that for the coherent scattering used in diffraction, we can use scattering length directly, without concern of the differential cross section.

Therefore, if the incident radiation is a plane wave (of the form of Equation 5.1) given by:

$$\Psi_0 = \mathrm{e}^{i(\mathbf{k_0}.\mathbf{r_0}-\omega t)} \tag{5.3}$$

where $\mathbf{k_0}$ is the wave vector of the incident wave, $\mathbf{r_0}$ is the position of the scattering centre, ω is the angular frequency and t is the time. The resultant spherical wave at point $\mathbf{r}$ and time t is given by:

$$\Psi_\mathrm{s} = f(\mathbf{k_0}, \mathbf{k}) \frac{\mathrm{e}^{i(\mathbf{k}|\mathbf{r}-\mathbf{r_0}|-\omega t)}}{|\mathbf{r}-\mathbf{r_0}|} \tag{5.4}$$

Where $f(\mathbf{k_0}, \mathbf{k})$ is the scattering length and $\mathbf{k}$ is the scattering direction. The intensity of the scattering is then given by:

$$I_\mathrm{s} = |f(\mathbf{k_0}, \mathbf{k})|^2 \frac{1}{|\mathbf{r}-\mathbf{r_0}|^2} \tag{5.5}$$

From this, it is shown that the scattering length, $f(k_0, k)$, is the function of the sample and the radiation, that will determine the intensity of the resultant radiation at a given point, and so these are the functions that must be evaluated for each radiation type. In general, this scattering length will be the Fourier transform of the potential that causes the scattering.

5.2.4 Electron scattering

For electron scattering, it is convenient to consider the electron beam directly as a wave, with wavelength λ, using the De Broglie relationship, Equation 5.6.

$$\lambda = \frac{h}{\rho} \tag{5.6}$$

Where h is Planck's constant and ρ is the electron's momentum.

This proves a sufficient model in the case of large electron flux, which is common in such scattering experiments. In the case of single electron detection, which will be touched on later, it is more appropriate to consider the wave function as the probability distribution of individual electrons. The resulting mathematics remain largely unchanged, but this is required to account for the fact that discrete electrons will be detected.

The electron wave interacts directly with electromagnetic potentials in a coloumbic manner. The scattering length can be determined by solving the Schrödinger Equation for the electron wave in the presence of such potentials. Depending on the periodicity of the potential, the scattering length may be referred to by other names. If the potential of a single atom is calculated, the scattering length is referred to as the 'atomic form factor'.

The strength of the potential at various points in space determines the value of the atomic form factor. For neutral atoms, the potential is zero outside of the atomic radius, as the nucleus is charge balanced by the electrons. At such locations, the electron beam is unperturbed. Within the nucleus of the atom however, the potential is net-positive, as the nuclear charge is incompletely shielded by the electrons. As the beam passes closer and closer to the nucleus, a greater proportion of the

nuclear charge, Z, is generally experienced, and so the stronger the electromagnetic field. This gives an atomic form factor that tends to increase as distance from the nucleus decreases. Crucially, this means the atomic form factor of an electron beam is dependent on the electron density at a given point, as well as the nuclear charge.

From one atom to another, as the atomic number increases, electronic orbitals are generally contracted and so the shielding of the nuclear charge, at a specific radius, increases. This means that for a set distance from the nucleus, the atomic form factor will likely increase by less than the unit increase in nuclear charge. It also means, due to the contraction of orbitals, that to experience an appreciable amount of the nuclear charge, the electron wave must pass closer to the nucleus of a high Z atom than a low Z atom. As these close passes are less probable, this manifests as a sharper drop off in atomic form factor for heavy atoms than light atoms

A direct consequence of these factors is that nearby elements will have very similar atomic form factors, and thus, will display scattering behaviour that is closely comparable. It can also be determined that heavy atoms will scatter more strongly than light ones. These are the two main biases of using electron scattering to probe structure - there will be poor contrast between elements with similar atomic numbers and weak scattering from lighter elements.

A general equation for the atomic form factor for electrons, f_{el}, is given by the Mott Equation (5.7). The derivation of this equation is beyond the scope of this summary, but as mentioned earlier, the derivation entails solving the Schrödinger equation in the presence of the atomic potentials. It is given here as it will allow for comparison to the equivalent equation for x-rays later.

$$f_{\text{el}}(\Delta\mathbf{k}) = \frac{2me^2}{\hbar^2\Delta\mathbf{k}^2}\left(Z - \int_{-\infty}^{+\infty}\rho(\mathbf{r})\mathrm{e}^{-i\Delta\mathbf{k}\cdot\mathbf{r}}\mathrm{d}^3\mathbf{r}\right) \tag{5.7}$$

Where $\Delta\mathbf{k}$ is the difference between the incident and the scattered wave vector, m is the mass of the electron, $\hbar$ is the Dirac constant, and $\rho(\mathbf{r})$ is the electron density at $\mathbf{r}$.

5.2.5 X-ray Scattering

Similar to electrons, x-ray interactions are also dependent on the electron density and the atomic number. The interaction of x-rays with atoms, however, is not primarily an interaction with the atomic potential, but an interaction with the electrons of the atom.

X-rays are forms of electromagnetic radiation and thus are characterised by their influence on the electric and magnetic fields. When the electric field of an x-ray interacts with a bound electron of an atom, it can displace it. Such a displacement will produce a harmonic restoring force and as a result the electron will behave as an oscillator. The oscillation of a charged particle, such as an electron, will in turn produce electromagnetic radiation with a frequency proportional to the oscillation in a direction perpendicular to the acceleration. Assuming the x-rays are not polarised, this will give an even distribution of perpendicular directions, over all possible angles, and so for many interactions the resulting wave could be considered as a spherical wave. Specifically, the radiated field for a single interaction with one electron can be approximated by:

$$\mathbf{E} = f\frac{\mathbf{E_0}}{r} \tag{5.8}$$

where f is the scattering length, $\mathbf{E_0}$ is the electric field of the x-ray and r is the

perpendicular distance from the oscillator. In turn, the scattering length is given by:

$$f = \left(\frac{e^2}{mc^2}\right) \frac{\omega^2}{(\omega_r^2 - \omega^2 + i\beta\omega)} \tag{5.9}$$

where m is the mass of the electron, c is the speed of light, ω is the angular frequency of the x-ray, β is a dampening constant and ω_r is the resonant frequency of the oscillator. This model can be used to give some insight into the factors which will affect x-ray scattering.

As this interaction requires an electron to act as the oscillator, the atomic form factor is therefore dependent on the x-ray encountering an electron to oscillate. It is thus dependent on the electron density of the atom and by extension, its atomic number. It can also be seen that the atomic form factor will be dependent on the relative frequencies of the incident x-ray, ω, and of the resonant frequencies, ω_r, of all the electrons in the atom. These resonant frequencies will depend on the specific atomic orbital and will, in general, vary with Z.

To a first approximation, the atomic form factor will increase proportionally to the number of electrons. This means isoelectronic atoms will show very little contrast in scattering behaviour, which can be a limitation of x-ray techniques. Additionally, leveraging situations where ω is close to ω_r can allow for some additional information on electron orbitals in the structure and give insights into bonding that are not available when electrons are the probe radiation.

The more exact form of the equation for the atomic form factor for x-rays, f_{x}, can be given by another Mott Equation (5.10).

$$f_{\mathrm{x}}(\Delta\mathbf{k}) = \frac{e^2}{mc^2}\int_{-\infty}^{+\infty}\rho(\mathbf{r})e^{-i\Delta\mathbf{k}\cdot\mathbf{r}}\mathrm{d}^3\mathbf{r} \tag{5.10}$$

5.2.6 X-ray vs Electron atomic form factors

The two Mott equations 5.7 and 5.10 can be combined to give a direct relationship between the atomic form factors of electrons and x-rays:

$$f_{\mathrm{el}}(\Delta\mathbf{k}) = \frac{3.78}{\Delta\mathbf{k}^2}\left(Z - 3.54\times10^4 f_{\mathrm{x}}(\Delta\mathbf{k})\right) \tag{5.11}$$

From this equation some cursory distinctions between x-ray and electron scattering can be seen. The first is that the electron atomic form factor is approximately 10^4 times larger than the x-ray equivalent (when $\Delta\mathbf{k}$ is in the order of a few angstroms), meaning that using electrons will provide more scattering for a given radiation flux.

Additionally, the $\Delta\mathbf{k}$ dependence of each of the atomic form factors is actually slightly different due to the influence the electron density has on the underlying potentials. The result is that the electron atomic form factor falls off more rapidly with $\Delta\mathbf{k}$ than the x-ray atomic form factor does, allowing x-ray scattering to persist better at higher angles.

5.2.7 Neutron Scattering

Where the atomic form factors for x-rays and electrons were similar, and dependent on electron density and scattering direction (Equations 5.7 and 5.10), neutron scattering is different.

Neutrons are uncharged particles and their interaction can be modelled as physical

collisions with the mass in the nucleus. As this means neutrons are essentially only scattered at a point location, the position of the nucleus, the atomic form factor is essentially constant as a function of $\Delta\mathbf{k}$ and is often replaced with the 'coherent scattering length', b (this constant value can be thought of as taking the Fourier transform of a single delta function, with the result being a constant).

As with x-rays, in the case of many neutrons, a wave-type model using the De Broglie relationship (Equation 5.6) is sufficient to predict the resulting scattering, but as with the electrons, for individual neutrons this should be considered a probability function.

Modelling the interaction as direct collisions with a nucleus, one would expect the coherent scattering length to scale linearly as a function of nuclear density. This trend only holds approximately, due to the large influence that atomic resonances have on b. The effects of these atomic resonances follow no clear, simple trend relating to atomic number, which allows neutron scattering the possibility to provide large contrast between elements of very similar atomic number. This can be a key advantage of using a neutron probe as opposed to an x-ray or an electron probe.

Additionally, it is also possible to have negative scattering factors for neutrons, corresponding to an inverted phase of the scattered neutron and resulting from a resonant transition that is below the energy of the neutron. The physical origin of this resonance scattering is similar to the oscillator process of x-ray scattering, but instead of the electrons acting as the oscillators it is the nucleus itself.

5.3 Diffraction

So far we have considered a single scattering centre and discussed the information that can be gained about that scattering centre through its interaction with different types of radiation. Practically, isolated scattering centres rarely exist, and would be of limited interest, so we need to extend our view to what happens when multiple scattering centres occur near to each other and, specifically, in periodic arrangements, as seen in crystal structures.

The specific case of a plane wave being scattered elastically and coherently from the scattering centre is commonly referred to as diffraction. In the case of non-polarised incident radiation this resultant wave will be a spherical wave. A detector can be used to collect information on the intensity of the scattered wave as a function of scattering angle. The resulting intensity profile is called the 'diffraction pattern'.

When two or more scattering centres are present at a separation comparable to the wavelength of the radiation, the resulting radiation will all overlap, causing a resulting field that is the sum of all the overlapping radiation, known as 'interference'. This interference can be completely constructive, completely destructive or somewhere in between, depending how the amplitude maxima and minima line up. The interpretation of this interference is the basis of most diffraction based techniques.

The theory of diffraction presented here is applicable to all the types of radiation scattering discussed previously, however, it is important to reiterate that the limitations of a specific radiation-scattering centre interaction will be present in the diffraction also.

5.3.1 Diffraction in Crystals

When the phase difference between scattered waves is an integer wavelength, this is referred to as the Laue condition and such a regime is what is used to derive Bragg's law (Equation 5.12). Such a condition leads to constructive interference and 'peaks' in intensity in the diffraction pattern.

$$n\lambda = 2d\sin\theta \tag{5.12}$$

5.3.2 Bragg's Law

Figure 5.1 shows the typical wave diagram used to derive Bragg's law. Assuming the the two waves shown are emitted from a source a large distance away relative to the inter-layer spacing, and similarly converge at a point a large distance away, the two beams can be interpreted as essentially parallel. In such a case, the total path lengths of the the two beams will be the same, other than the two additional bold sections, on the lower line (shown in light blue). Each of these has a length of $d\sin\theta$ and so the total path difference is $2d\sin\theta$. It then follows that if this path difference is exactly an integer multiple of the wavelength (i.e. $n\lambda$) then the phase of the two waves will be unchanged and the waves will perfectly constructively interfere.

5.3.3 Ewald Sphere

A convenient approach for evaluating where constructive interference will occur, for a given crystal lattice and experimental geometry is to use the Ewald sphere construction. To derive this, we use a very similar wave diagram and line of reasoning to the derivation of Bragg's law, but using vector algebra instead of

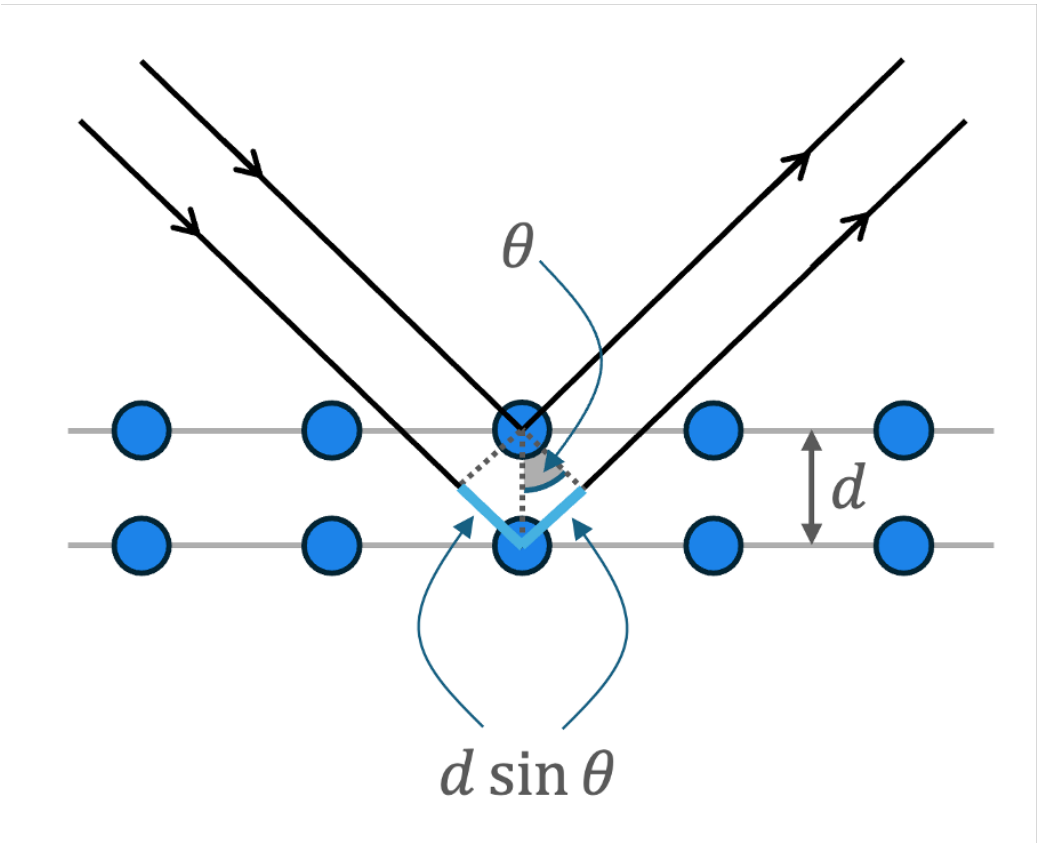


Figure 5.1: Wave diagram for the formulation of Bragg's Law

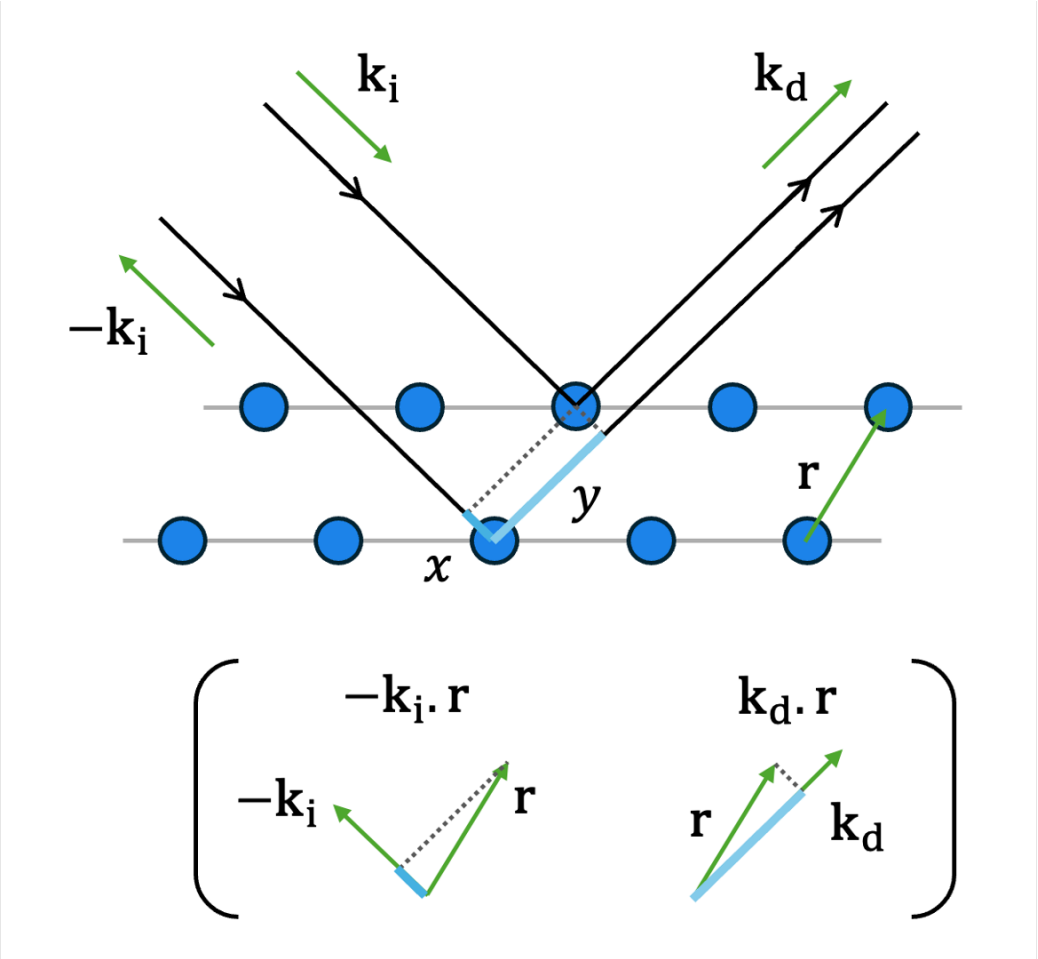


Figure 5.2: Vector construction of Bragg condition

trigonometry.

Figure 5.2 shows a generalised vector equivalent of Figure 5.1. In Figure 5.2 the spacing between the layers is now a general displacement vector $\mathbf{r}$. The direction of the incident radiation is given by $\mathbf{k_i}$ and the direction of an arbitrary scattered wave is given by $\mathbf{k_d}$. The path difference in this case is thus given by $x + y$, where

$$x = -\mathbf{k_i}.\mathbf{r} \tag{5.13}$$

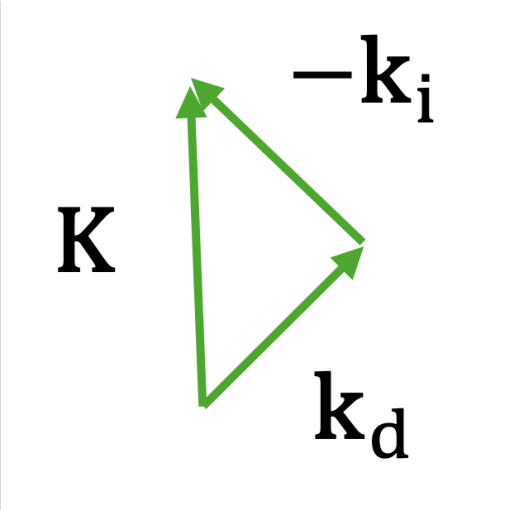


Figure 5.3: Top-to-tail depiction of the vector, $\mathbf{K}$

and

$$y = \mathbf{k_d}.\mathbf{r} \tag{5.14}$$

therefore the total path difference is given by

$$x + y = \mathbf{k_d}.\mathbf{r} - \mathbf{k_i}.\mathbf{r} \tag{5.15}$$

$$x + y = (\mathbf{k_d} - \mathbf{k_i}).\mathbf{r} \tag{5.16}$$

In Figure 5.3 it is shown that the path difference can just be written as $\mathbf{K}.\mathbf{r}$. This $\mathbf{K}$ vector is simply the displacement between the incident and scattered wave vectors, and so is just the reflection vector. As before, in order for this path difference to lead to constructive interference, we need it to be equal to $n\lambda$.

The path difference, $\mathbf{K}.\mathbf{r}$, can be used to give the phase factor of the reflected

wave. For perfect constructive interference this must be equal to 1.

$$\mathrm{e}^{-2\pi i\mathbf{K}.\mathbf{r}} = 1 \tag{5.17}$$

As $\mathbf{r}$ is a displacement vector between lattice points in the crystal, we can write this equivalently as a linear combination of some, necessarily integer, amounts n_1, n_2 and n_3 of each of the crystal unit vectors, $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$.

$$\mathbf{r} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \tag{5.18}$$

This gives the following result:

$$\mathrm{e}^{-2\pi i\mathbf{K}.\mathbf{r}} = \mathrm{e}^{-2\pi i n_1\mathbf{K}.\mathbf{a}} \times \mathrm{e}^{-2\pi i n_2\mathbf{K}.\mathbf{b}} \times \mathrm{e}^{-2\pi i n_3\mathbf{K}.\mathbf{c}} = 1 \tag{5.19}$$

For this to hold, each of these components should multiply to be 1. Using the fact that for all integers, m:

$$\mathrm{e}^{-2\pi i m} = 1 \tag{5.20}$$

We can deduce that the equation will be equal to 1 when $n_1\mathbf{K}.\mathbf{a}$, $n_2\mathbf{K}.\mathbf{b}$ and $n_3\mathbf{K}.\mathbf{c}$ are all be integers. As n_1, n_2 and n_3 are already necessarily integers, this means that $\mathbf{K}.\mathbf{a}$, $\mathbf{K}.\mathbf{b}$ and $\mathbf{K}.\mathbf{c}$ must also be integers, which we shall, just initially, denoted as α, β and γ.

We now have a condition for constructive interference: that $\mathbf{K}.\mathbf{a}$ is an integer, α; $\mathbf{K}.\mathbf{b}$ is an integer, β; and $\mathbf{K}.\mathbf{c}$ is an integer, γ. The only remaining challenge is to find for which values of $\mathbf{K}$ this condition holds.

For brevity we will skip over the mathematical derivation of the reciprocal lattice and instead just begin with the result. Much as before, for an arbitrary lattice, a general lattice point is given by necessarily integer amounts n_1, n_2 and n_3 of each of the lattice unit vectors, $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. For such a lattice, there is also a *reciprocal* lattice, for which a *reciprocal* lattice point is given by necessarily integer values h, k and l of each of the reciprocal unit vectors, $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$.

The useful property of this reciprocal lattice is that it is the inverse of the real-space lattice, such that the unit vectors are related in the following manner:

$$\mathbf{a}.\mathbf{a}^* = \mathbf{b}.\mathbf{b}^* = \mathbf{c}.\mathbf{c}^* = 1 \tag{5.21}$$

$$\mathbf{a}.\mathbf{b}^* = \mathbf{a}.\mathbf{c}^* = \mathbf{b}.\mathbf{c}^* = \mathbf{b}.\mathbf{a}^* = \mathbf{c}.\mathbf{a}^* = \mathbf{c}.\mathbf{b}^* = 0 \tag{5.22}$$

Using these equations 5.21 and 5.22, we can find the values of $\mathbf{K}$ which will satisfy our condition for constructive interference. Take an arbitrary reciprocal lattice point, $\mathbf{G}$, defined as:

$$\mathbf{G} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{5.23}$$

Recalling that the condition for constructive interference was that $\mathbf{K}.\mathbf{a}$, $\mathbf{K}.\mathbf{b}$ and $\mathbf{K}.\mathbf{c}$ must all be integers (α, β and γ.), it can be seen that when the reflection vector $\mathbf{K}$ is equal to $\mathbf{G}$:

$$\mathbf{K}.\mathbf{a} = \mathbf{G}.\mathbf{a} = h\mathbf{a}^*.\mathbf{a} + k\mathbf{b}^*.\mathbf{a} + l\mathbf{c}^*.\mathbf{a} = h \tag{5.24}$$

As $\mathbf{a}^*.\mathbf{a} = 1$, $\mathbf{b}^*.\mathbf{a} = 0$ and $\mathbf{c}^*.\mathbf{a} = 0$

Similarly:

$$\mathbf{K}.\mathbf{b} = \mathbf{G}.\mathbf{b} = k \tag{5.25}$$

and

$$\mathbf{K}.\mathbf{c} = \mathbf{G}.\mathbf{c} = l \tag{5.26}$$

As these values h, k and l correspond to the complete set of reciprocal lattice points, they are always integers, and as they so will always be appropriate values for α, β and γ. Furthermore, as the reciprocal lattice spans all possible integer values of h, k and l, the pairs of sets: h and α; k and β; l and γ; are all identical. It is thus standard for the condition for constructive interference to simply skip straight to the notation $\mathbf{K}.\mathbf{a} = h$, $\mathbf{K}.\mathbf{b} = k$ and $\mathbf{K}.\mathbf{c} = l$.

This condition tells us all the reflection vectors, $\mathbf{G}$, for which constructive interference can occur, it does not however, include the constraints of which reflections are available to be detected within the constraints of the experimental setup, this is where the Ewald sphere comes in.

Depending on the experimental setup, only some of the reflection vectors are available to be detected concurrently. This, assuming a single scattering event (per incident electron/photon/neutron), depends on two variables, the orientation of the incident radiation relative to the crystal lattice, and the wavelength of the incident radiation, λ. In reality, the quality of this assumption of a single scattering event is dependent on the sample properties, especially thickness, but that will be discussed further later on.

The construction geometry of the Ewald sphere is shown in Figure 5.4. The points show the reciprocal lattice points, which all can give rise to constructive interfer-

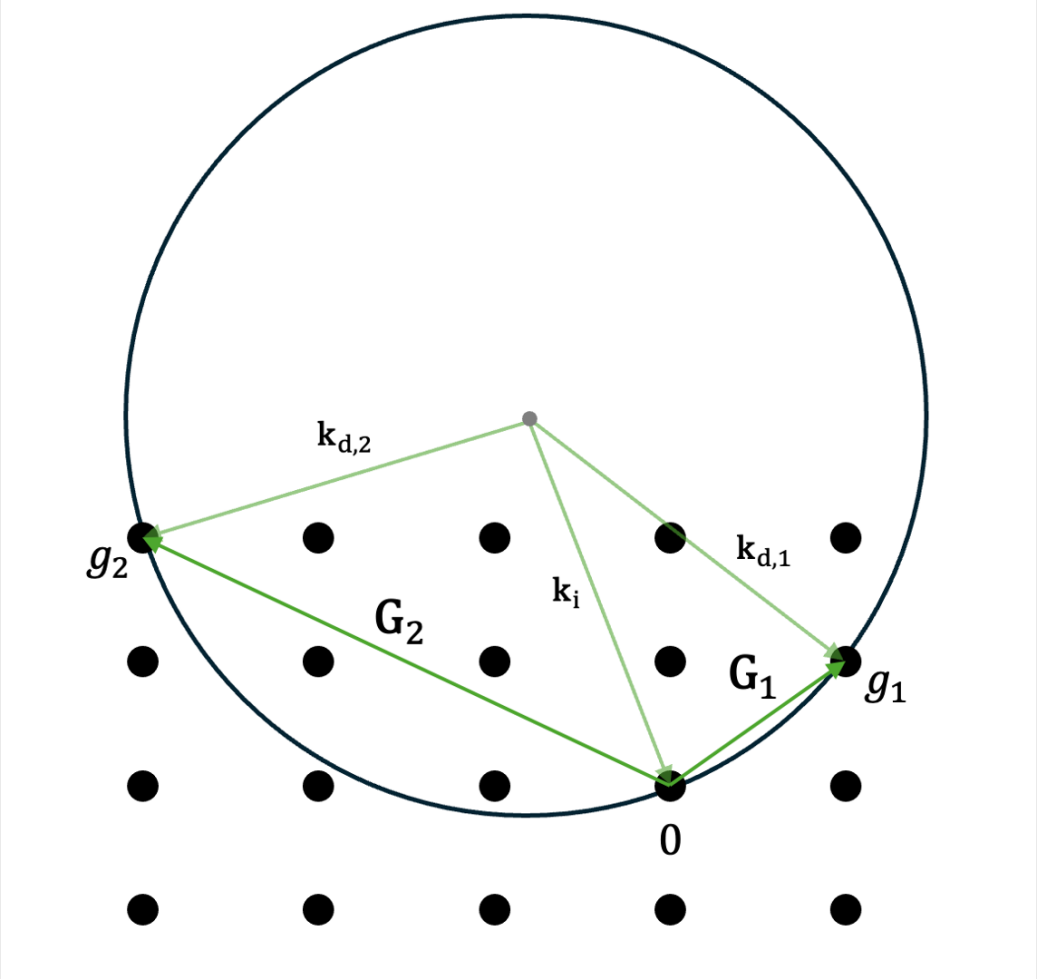


Figure 5.4: 2D representation of a typical Ewald sphere construction. Reciprocal lattice points are shown by black dots, the origin of the sphere by a grey dot and vectors are shown with green arrows.

ence. The sphere however, shows that only some of the wave vectors align with reciprocal lattice points, as the magnitude of these wave vectors are dependent on the wavelength, λ, of the incident radiation.

This construction forms a sphere as it is representing all the possible values which $\mathbf{k_d}$ could have taken back in Figure 5.2. The origin of the sphere is not at the the origin of the reciprocal lattice however, as the property of interested is not $\mathbf{k_d}$, we want to know what are the experimentally accessible $\mathbf{K}$, so as in Figure 5.3, we subtract $\mathbf{k_i}$ to displace the origin of the sphere by the vector of the incident radiation.

Now, all points on the surface of the sphere represent all experimentally accessible values of $\mathbf{K}$. Therefore, any reciprocal lattice point $\mathbf{G}$, which lies on the surface of the sphere, both fulfills the Laue condition for constructive interference, and is experimentally accessible; this is the utility of the Ewald sphere.

By changing the orientation of the crystal, the orientation of the reciprocal lattice is also changed, thus allowing different $\mathbf{G}$ vectors to intersect with the sphere's surface. Likewise, changing the value of λ will change the magnitude of the unit wave vector and so alter the radius of the Ewald sphere. The relative differences can be seen in Figure 5.5, with one of the most prominent effects being the change in the curvature of the sphere intersecting the lattice. Many diffraction experiments that aim to advance structural understanding are based on the principle of using these two degrees of freedom (orientation and radiation wavelength) to try and understand the underlying reciprocal lattice as fully as possible.

It is finally worth noting that various phenomena exist that allow reflections vectors, $\mathbf{K}$, that are not exactly equal to a reciprocal lattice vector, $\mathbf{G}$, to still be observed. These will be discussed more fully as and when it is necessary to do so, but the general cause of this occurrence is that lattices are theoretical, infinite

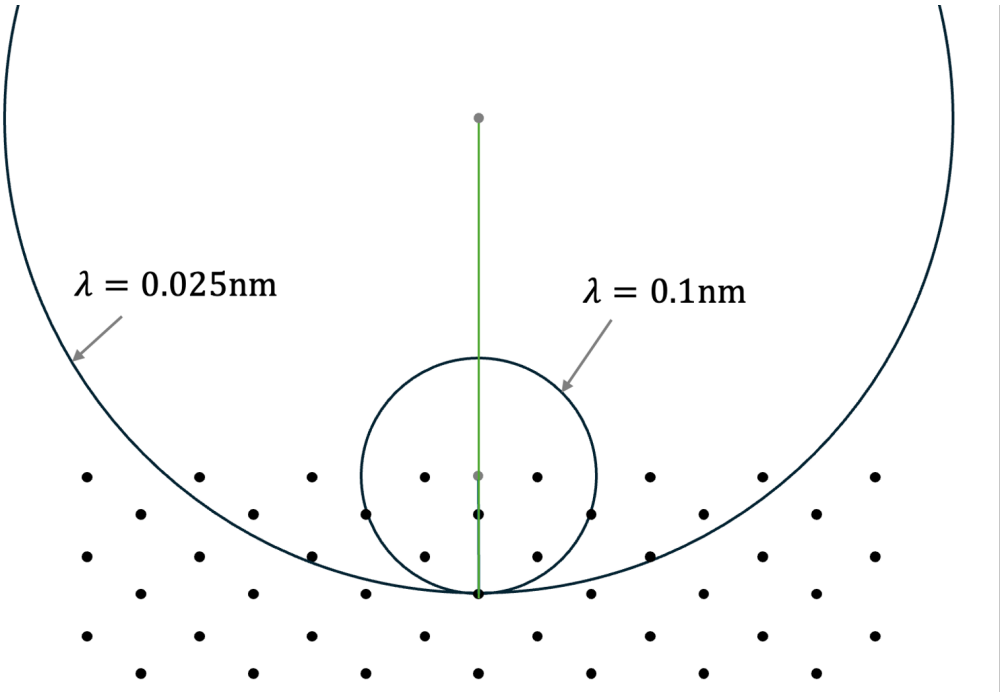


Figure 5.5: Comparison of the relative Ewald spheres of sources using $0.1\,\mathrm{nm}$ X-rays and sources using $0.025\,\mathrm{nm}$ X-rays.

systems, with infinitesimally small lattice points. As real crystals are finite, the reciprocal nature of the mathematics can essentially lead to broadening of the lattice points. In some cases this broadening can allow points to intersect the sphere, that otherwise would not.

5.3.4 Structure Factors

It is worth emphasising once more that while the Ewald sphere construction allows us to determine which reciprocal lattice points lead to constructive interference, it gives no information on the intensities of the diffracted radiation that will be detected at these points. This is because, for convenience, it is typical to deal with the lattice and the motif separately. For a full description of diffracted intensity one must account for the periodic arrangements of all the different scattering centres. As the periodicity of each type of scattering centre will all contain, at a minimum, the periodicity of the lattice, dealing with the lattice and the motif separately helps avoid repetition of calculation. The effect of the the scattering power and positions of the different scattering centres in the motif is evaluated within the structure factor. The equation for the structure factor is given in Equation 5.27, where f_j is the atomic form factor of the j-th atom and (x_j, y_j, z_j) are the fractional coordinates of the j-th atom within the motif cell.

$$F_{hkl} = \sum_{j=0}^{N} f_j e^{[-2\pi i(hx_j + ky_j + lz_j)]} \tag{5.27}$$

This structure factor equation allows quantification of diffraction intensities and the combination of this with the lattice component provides the foundation for simulating, and/or, fitting models to experimental data.

5.3.5 Multiple scattering

So far, this all assumes that each quantised unit of radiation in the incident beam undergoes only a single scattering event as it interacts with the sample, the so-called kinematic model. The quality of this assumption can be evaluated for a given sample, comparing it's scattering length to sample thickness. If the sample is much thicker than the scattering length, it then becomes probable that multiple scatterings could occur. Such situations require amendments to the theory outlined to account for the multiple scatterings and approaches to do this are 'dynamical models'. A in-depth consideration of dynamical theory is beyond the scope of this theoretical overview, but deviations from kinematic theory must be considered in certain cases.

With the background of scattering theory covered, it is now possible to outline specific experimental setups that can be used to understand aspects of atomic structure.

5.4 Powder X-ray Diffraction

5.4.1 Overview

The Ewald sphere construction provides a convenient framework for a basic understanding of how Powder X-ray Diffraction (PXRD) experiments seek to understand the structure of a sample.

Figure 5.6 shows the theoretical experimental setup for PXRD. It shows a sample with incident radiation at an angle of θ to the plane of the sample and a detector at an angle 2θ to the incident radiation. In this diagram the source and detector are swept across the arc of the circle in order to collect intensities at a range of 2θ

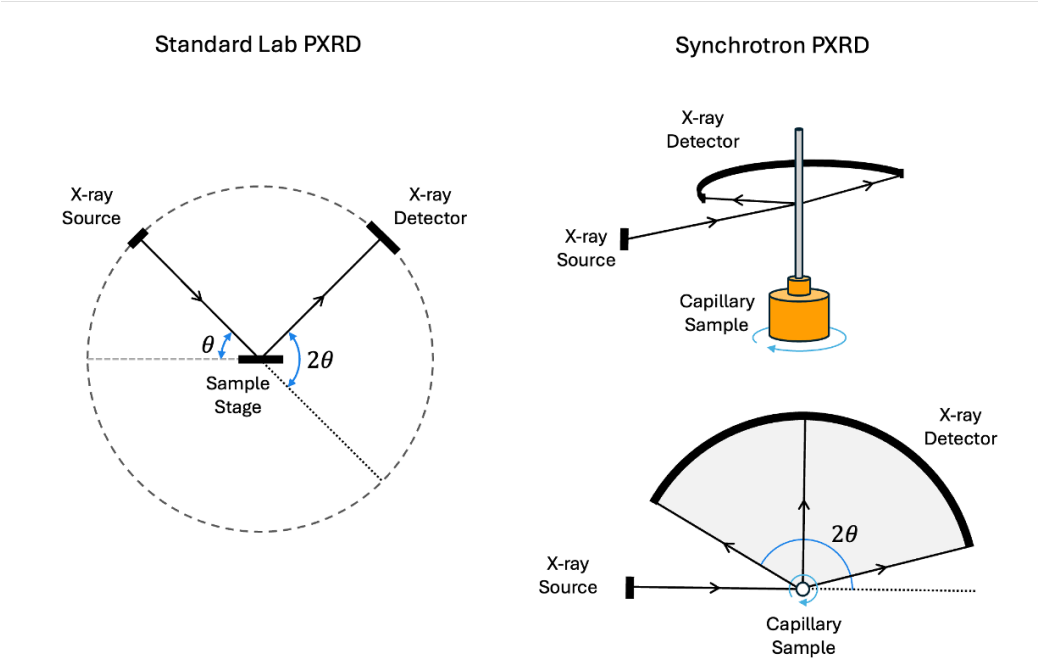


Figure 5.6: Schematic setup of both a standard lab-based Powder X-ray Diffraction experiment and that of a synchotron-based experiment.

values but in real experiments it is often more practical to achieve the same effect by rotating the sample and keeping one of the source or the detector fixed. For synchotron based techniques, where it is impractical to move either the source or the detectors, the sample is spun within the beam and an array of detectors spanning the range of 2θ is used instead.

Looking back at Figure 5.4 again, it can be seen that by rotating the orientation of the sample in the incident beam, the reciprocal lattice will be rotated relative to the Ewald sphere, allowing different reciprocal lattice vectors to intersect with the sphere surface. By doing this over a large enough range of orientations (2θ values), it is possible to intersect a comprehensive sample of reciprocal vectors and collect diffraction intensities for each of them (with the ultimate limit being the radius of the Ewald sphere and thus the wavelength of the X-rays).

With this information on the diffraction intensity at each value of 2θ, the typical procedure is to then use this information to deduce the crystal structure and the associated reciprocal lattice. This can be done qualitatively, by indexing peaks based on relative positions and thus deducing the Bravais lattice types and symmetries, or more quantitatively using simulations. The strategy is typically incorporated as part of "Rietveld Refinement", which is discussed further in the Computational Techniques chapter.

5.4.2 Factors that effect pattern intensities

In a powder x-ray diffraction experiment, there are a few factors that will modify the intensity profiles from those described in Equation 5.27. The most prominent change is multiplicity. For a given crystal structure, there are specific values of h, k and l that correspond to equivalent lattice planes. For a cubic system (where lattice parameters $a = b = c$ and $\alpha = \beta = \gamma = 90°$) there are 6 equivalent $\{200\}$ type planes {(200), (020), (002), ($\bar{2}$00), (0$\bar{2}$0), and (00$\bar{2}$)}. In contrast there are 12 individual planes in the set 110. This means that for an even sampling of orientations in the powder sample, there will be twice as many $\{110\}$ reflections as $\{200\}$ and so the intensity ratio between the (110) peak and the (200) will be twice the ratio calculated from Equation 5.27 for a single instance of each. This multiplicity ratio is inherently dealt with if all structure factors (all values of h, k and l) are calculated independently, but as indexing of patterns is usually done by attributing a peak to a single reflection of its set, the multiplicity is worth noting.

Other common factors that effect peak size and shape are: particle strain, particle size, wavelength and slit width. The specifics of exactly how these effect profile is beyond the scope of this overview, but it is worth noting them nonetheless.

5.4.3 X-ray source: Lab vs Synchotron

The quality of the x-ray source also has a major impact on the diffraction pattern and the quality of the structural insights that can be gained. Traditional lab diffractometers will generate x-rays using vacuum tube diodes. These tubes are biased at a particular voltage such that electrons are thermionically emitted from the cathode and accelerated towards the anode. The accelerated electrons interact with the anode and induce transitions in the electron orbitals. The anode is chosen based on the desired characteristic radiation. A popular choice is copper anode, targeting the K_α transition. Copper is popular owing to acceptable operating voltages and good thermal conductivity (advantageous for keeping the diode from getting too hot).

The target x-rays are not produced in isolation, however, as accelerating voltage of the tube increases, so does the radiation intensity of all the possible electronic transitions. What is then required is to pass the beam through a monochromator. Typically a monochromator is just a single crystal of fixed composition and orientation, such that its Bragg condition can be leveraged to only be satisfied for specific wavelengths. The actual sample is then placed along the path of the Bragg reflection of the monochromator. As only the specific radiation desired constructively interferes along this path, the monochromator ensures that no unwanted radiation is coherently scattered along this direction.

The main limitations of lab-based diffractometers like this are the brilliance, which is given as photons per area of emitter and normalised by the solid angle of the beam. This normalisation is included as only the beam incident on the monochromator ends up contributing to the beam incident on the sample. Vacuum-tube diodes tend to be limited both by operating voltage (which limits the total number of photons emitted) and by the spread of the beam (they are approximately spher-

ical wave sources that will be split into beams using slits, which limits the number of photons per solid angle considerably). The result of these limitations is that they provide relatively low signal to noise diffraction patterns as a result of the limited brilliance. This introduces more uncertainty into any quantification and can also result in some of the minor features being completely unresolved. This low brilliance is especially an issue for study of lighter elements in structures, which, as discussed earlier, tend to have smaller atomic structure factors, thus are less likely to cause scattering. As a result high brilliance beams are required to increase the photon flux and obtain a sufficient number of scattering events to study.

The alternative to lab-based diffractometers is to use synchotron radiation. The specifics of synchotron function are again beyond the scope of this overview, but the simplified explanation is that they accelerate high velocity electrons, which are traveling around a ring, to produce high energy x-rays. This acceleration is typically done perpendicular to the velocity of the electron, so that the radiated x-rays are all parallel in the plane of travel, resulting in a source with much less angular spread. As the electrons are at such high velocities, the amount of time to cycle the ring is short and frequency of excitation events is high, resulting in a large number of photons produced per electron in the ring. Both these factors lead to synchotron radiation having much higher brilliance.

5.5 Neutron Diffraction

Powder Neutron diffraction (PND) has many similarities to PXRD as well as some key differences.

5.5.1 Overview

The main difference between PND and PXRD is the nature of the diffracting 'radiation' as discussed in the earlier Scattering sections. PND uses accelerated neutrons as the source of radiation, leveraging the de Broglie relation (Equation 5.6). The pertinent difference to reiterate between neutrons and electrons, however, is that the neutron is an uncharged particle. This means it does not interact with the electron cloud, as X-rays and electrons do, but with the nucleus. The resultant scattering interaction is also much more consistent across the radius of the atom. Where x-rays interact with the electron density, which can vary considerably across the atomic radii, the matter density does not, it is approximately one value (the density of the nucleus) across the radius of the nucleus and zero everywhere else. The experimental consequence of this is that PND offers much better peak resolution than in PXRD.

Another key experimental difference to reiterate is that as a result of interacting with the nucleus rather than the electrons is that the scattering is no longer proportional to Z. The scattering can even vary for isotopes of the same element. This can allow for strong contrast between elements of very similar atomic numbers which can be useful for many experiments.

The final major difference in a PND experiemnt is the potential to use a 'polychromatic' beam (a spread of neutron energies in the same incident beam) and use time of flight (TOF) as a measure of the neutron energy at each reflection angle. As TOF is proportional to the wavelength using the de Broglie relation (5.6), Bragg's law can be rewritten in the following form:

$$t = 505.56Ld\sin\theta \tag{5.28}$$

where t is TOF, L is the total flight length, d is the interplanar spacing and θ is the scattering angle.

5.6 Transmission Electron Microscopy

5.6.1 Overview

One of the most fundamental application of diffraction of electrons is for Transmission Electron Microscopy (TEM). A schematic ray diagram for a standard TEM setup is shown in Figure 5.7 created using the open source ray diagram software from Tu[104].

At the very top of the TEM column is the electron source. The exact nature of this source will vary from TEM to TEM, but current state-of-the-art machines use Cold Field Emission Guns (Cold FEGs). The operating conditions can also vary based on the experiment, but these electron sources are comfortably capable of producing electrons with an accelerating voltage of $300\,\mathrm{keV}$. Such electrons have an associated wavelength of $0.001\,968\,7\,\mathrm{nm}$. This results in an Ewald sphere 10 times larger than the larger sphere in Figure 5.5. Thus the Ewald sphere for such diffraction experiments can be considered as a flat plane through the reciprocal lattice.

From the electron source the beam is focused by the condenser lenses to create a uniform beam condition on the sample surface. Depending on the setup this can be convergent beam, parallel beam or divergent beam. All of the lenses shown in a TEM schematic are magnetic lenses which allow for both powerful and precise control of focal lengths.

After the condenser lens is the condenser aperture, which is used to control the maximum angle of the illumination as only some of the beam can pass through

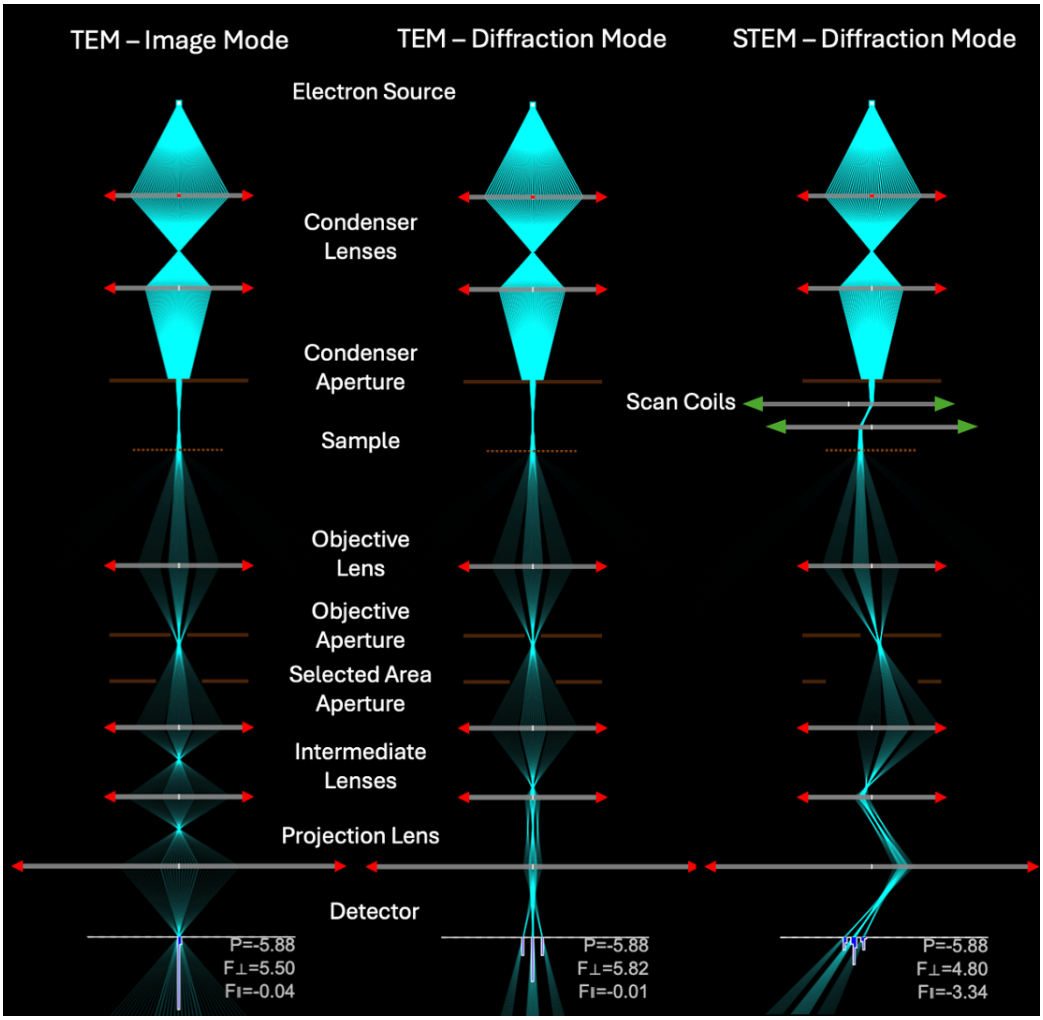


Figure 5.7: Ray diagram schematics of a TEM in both imaging and diffraction mode. The addition of tunable scan coils to deflect the beam incident on the sample forms a STEM ray diagram.

the hole of the aperture.

The beam is then passed through the sample, where it will undergo any scattering as a result of diffraction. This scattering happens in all directions but many will then pass through the objective lens. Parallel scattered beams are focused by the objective lens in the back focal plane, if these parallel beams are in-phase, constructive interference occurs and diffraction peaks are observed (from before, one can recall that these correspond to the reciprocal lattice points that are intersecting the Ewald sphere). Additionally, non-parallel beams originating from the same point on the sample also focused back to single points, where this occurs is referred to as the image plane.

The remaining lenses are then used to determine which of these planes, diffraction or image, should be focused on the screen. This is shown schematically in the first two ray diagrams of Figure 5.7, where the system is identical up to the intermediate lenses, but by changing the relative strengths of these lenses the detector changes from sampling the imaging plane to the diffraction plane.

The selected area apertures can be used to select only certain parts of these planes for imaging, for example, a small selected area aperture might be used to limit the transmitted beam, so only the electrons from a small region of the total illuminated sample, are allowed to pass through to the intermediate and projector lens. The objective aperture performs a similar function but limits beams in the back-focal plane. This means only beams associated with selected diffraction spots are allowed to pass through the final lens system and ultimately contribute to the image. This is the basis of Dark Field imaging.

The resolution of such a microscope is theoretically limited by the Rayleigh criteria, which gives the minimum separation between two point sources such that they can still be distinguished as separate points, and is given by the following

equation, for a circular aperture:

$$\theta = \frac{1.22\lambda}{b} \tag{5.29}$$

Where θ is the angle of separation between the two points, λ is the wavelength of the radiation and b is the radius of the aperture. The origin of this equation comes from the diffraction inherent to a wave passing through a circular aperture. The resulting intensity pattern (point spread function) has a maximum at the centre, and a characteristic variation with radius referred to as the 'Airy disc'. The resolution limit between two point sources is accepted as the separation when the maxima of one point spread function is aligns with the first minima of the other. The location of this minima can be derived from the first Bessel function. This is complex and beyond the scope of this thesis but the equation of the minima and is given in Equation 5.30.

$$kb \sin\theta = 3.83 \tag{5.30}$$

As $k = \frac{2\pi}{\lambda}$ this gives:

$$\sin\theta = \frac{3.83\lambda}{2\pi b} \tag{5.31}$$

Which using the small angle approximation of $\sin\theta \simeq \theta$ gives the Rayleigh criteria of Equation 5.29.

Also limiting this resolution, however, are various sources of aberration introduced by the lenses in the TEM system. These are spherical, coma, off-axial astigmatism, curvature of image field, and distortion. The main aberration of con-

cern is spherical aberration, C_s. Introduction of C_s corrected microscope was a massive breakthrough in achievable resolution.

5.6.1.1 Spherical Aberration

Spherical aberration is inherent to all lenses that make use of refraction to focus radiation. Such lenses use Snell's law:

$$\mu = \frac{\sin i}{\sin r} \tag{5.32}$$

Where μ is the ratio of the refractive indices, i is the angle of incidence and r is the angle of refraction. Most lenses assume that these rays are paraxial (make small angles to the optical axis) and thus have small values of i and r. In such cases Snell's law can be approximated with a Taylor expansion to $\frac{i}{r}$. In such a situation r varies linearly with i suggesting that all such beams will be focused to the same plane. For rays with larger angles to the optic axis, the approximation does not hold, however, and so such rays will be focused at a different length to the paraxial rays. The presence of these unfocused beams in the image plane, is the spherical aberration.

C_s correction is done by introducing a negative, non-spherical field that deflects the under-focussed rays back to the correct focal length. This is done using a series of quadrapolar and hexapolar lenses. These have been left out of Figure 5.7 for clarity and instead the lenses have been depicted as aberration-free.

5.7 Scanning Transmission Electron Microscopy

As explained in the previous section, imaging the back focal plane in a TEM will give a diffraction pattern, where the allowed reflections are given by the Ewald

sphere (which in the case of TEM is essentially a plane). The pattern will be a product of all the scattering interactions that occur along the illuminated beam path. These scattering interactions may or may not occur with a single common periodicity, relating to a consistent crystal structure in the illuminated area. This means that while TEM is a valuable technique, the interpretability can be limited by the coherence of the crystal structure within the area illuminated. This gives two extremes in TEM; either the beam is set to cover a large region of a sample, in which case the diffraction data could become difficult to interpret as if many substructures are present within that large region or the beam is set to cover a small area, in which case the structural coherence will be better but statistical sampling will be worse, meaning the structural insight gained might not be reflective of the full sample. Using a TEM is also a labour intensive process, as it requires using apertures to select regions for illumination, imaging these regions, and then swapping into diffraction mode to collect the pattern. A potential solution to these issues is to use Scanning Transmission Electron Microscopy (STEM).

5.7.1 Overview

The setup of STEM is essentially the same as the second extreme of TEM outline, in which the probe is set to cover a small region and collect diffraction data from the back focal plane. The difference with STEM is that then electromagnetic 'scan coils' are used to move the probe across the surface of the sample. A schematic example of the introduction of scan coils can be seen in Figure 5.7.

For standard STEM, the scanning probe is used to produce images where each scan location corresponds to a pixel in the image. For each of these scan locations, the intensity of the beam on a detector is used to populate the pixel value. Different detectors placed in different locations can be used to modify the contrast in the image to identify different structural effects.

Typically these detectors are large, 'monolithic' annular detectors, placed at different angles from the sample. If the detector subtends a very small angle the image is referred to as the 'Bright Field' image. If the detector subtends a large angle the image is referred to as 'Dark Field'. Typically multiple high angle detectors are used to create images simultaneously, as scattering at these different angles depends on different factors and so will highlight different structural features. At relatively small angles the dark field will display 'diffraction contrast', that is to say, differences in intensity (contrast) are dominated by the effect of diffraction. That is because at these angles the Bragg reflections will contribute the majority of the intensity.

At very high angles, Bragg reflections are limited (refer back to the atomic scattering factors dropping off more quickly for electrons), for detectors placed at these high angles, the scattering in dominated by elastic, incoherent scattering (Rutherford scattering). This high angle scattering is much more dominated by the atomic number Z, as it is essentially a function of nuclear charge. As a result, these high angle annular dark field (HAADF) images are often referred to as 'z-contrast'.

5.7.2 High Resolution Scanning Transmission Electron Microscopy

A specific application of STEM can be for high resolution lattice imaging, referred to as High Resolution Scanning Transmission Electron Microscopy (HRSTEM).

HRSTEM operates in a similar manner to a standard STEM experiment but using as small a probe size as possible. It is typical to then use an annular detectors in the back focal plane to collect the intensity information at each probe location and use this to produce the image.

From an Ewald sphere perspective, for a sample with a single underlying lattice, this approach of scanning across the sample will always intersect with the same

reciprocal lattice vectors, as translations in real space do not lead to translation in reciprocal space. What will change, significantly for such a small probe, is the structure factor (Equation 5.27), as the fractional position vector, (x_j, y_j, z_j), of the various, j, scattering centres relative to the probe do change. Thus, if the probe and step sizes are small enough, it is possible to resolve periodic information about different individual scattering centres.

As such imaging requires incredibly small probes, it is important that the beam is as well focused as possible, so this technique has only become possible with the introduction of $\mathrm{C_s}$ corrected microscopes. Additionally, well orientated and thin samples are required to maximise the in-plane spatial coherence of the atomic columns and minimise the dynamical effects of multiple scattering.

5.7.3 Probe size vs Convergence Semi-angle

A modification of the Rayleigh criteria (given by Equation 5.29) can be used to approximate the probe size, as the probe itself can be considered as an image of the point source of the beam, projected onto the sample. It is, in fact, this point-spread function (PSF) for the image of a point, that leads to the resolution limit given by the Rayleigh criteria. For a single point, the approximation of the probe size is given as:

$$d = \frac{0.61\lambda}{\alpha} \tag{5.33}$$

Where d is the diameter of the probe and α is the convergence semi-angle. This can be seen to be half the Rayleigh criteria (under the assumption that the semi-angle is small, such that the approximation $\sin\alpha = \alpha$ can be used to covert between aperture radius, b, and α).

The importance of this is that by increasing the convergence semi-angle, it is generally possible to decrease the size of the probe. In this regime the probe size is referred to as 'diffraction limited' as it is the diffraction causing the PSF that is limiting the size of the image. At a certain point, the minimum possible probe size is then achieved, as further increasing the semi-angle will lead to increased aberrations. These aberrations lead to spreading of focal lengths that increase the size of the image. This regime is said to be 'aberration limited'.

5.8 4D-STEM

Recent advancements in technology such as the development of direct electron detectors and the improvement in scan coils have allowed for the simultaneous collection of image data, as described in the STEM section and diffraction data as described in the TEM section, creating a technique referred to as 4D-STEM.

Specifically, for 4D-STEM, the beam is scanned across the sample as described for STEM. Again, each probe location can be used as the pixels of an image. The difference between 4D-STEM and traditional STEM is that instead of using annular detectors to record single intensities, the back focal plane is imaged at each probe location by a 2D detector (usually a direct electron detector to allow for fast scans while still achieving usable diffraction signal).

This experimental setup offers a promising solution to the limitations highlighted of TEM diffraction imaging: it allows use of relatively small illumination areas (which will have more structural coherency and so easier to interpret diffraction signal) but covers a large area in total, so provides more statistical significance as the data is representative of more of the total sample.

A major consideration in the setup of 4D-STEM is the convergence semi-angle.

The importance of the semi-angle is two-fold. First, as already noted, it will affect the probe size on the sample which has all the previously discussed implications about statistical sampling and structural coherency. Additionally, however, convergence semi angle also has implications on the size of the Bragg peaks in the back focal plane. If a small area of sample is illuminated (corresponding to a large convergence semi-angle), this will produce large diffraction disks in the back focal plane due to the reciprocal nature of the transformation. Operating in such a beam condition is referred to as Convergent Beam Electron Diffraction (CBED). CBED can be useful for the additional symmetry information that can present in the discs, but can be challenging to interpret in complex systems with many overlapping discs and grains, such as the structures looked at within this thesis.

The primary 4D-STEM technique used within this work is at the opposite extreme to CBED. When small convergence semi-angles are used, the probe is relatively large, but the Bragg peaks in the diffraction plane become small, well defined spots. This can be extremely useful, especially for indexing, and especially in situations where many diffraction peaks are present (peaks which might otherwise overlap with larger spot sizes). Such an operating condition is referred to as Nanobeam Electron Diffraction (NBED), in contrast to CBED, or alternatively it is known as Scanning Electron Nanobeam Diffraction (SEND).

Previously in the HR-STEM section (Section 5.7.2), we discussed the challenges of achieving a small probe, by increasing the convergence semi-angle. For SEND, it is the opposite challenge. Achieving a small convergence semi-angle ($\simeq 1\,\mathrm{mrad}$) is a non-trivial task. Generally, the convergence semi-angle can be controlled by modifying the strength of the lenses above the sample. By pushing the crossover point closer to a specific lens it passes through a smaller area of the lens before being deflected and focused, essentially reducing the angular spread of the beam. To do this, however, requires that particular lens to be operating at a much larger field

strength. For most microscopes currently available, only some lenses are strong enough to produce the fields required, and setting them to such values requires compromising other functionality such as the C_s correction.

5.9 Spectroscopic Analysis

A major advantage of operating an electron microscope in a STEM experiment is the opportunity to collect additional spectroscopic information about the sample. This leverages the fact that the interaction of the electron beam with the sample causes a host of phenomena at a variety of orientations relative to the optic axis. Some of these are depicted in Figure 5.8.

The basis for electron diffraction is that electrons undergo coherent, elastic scattering, when they interact with the atomic potentials of the sample. The 'elastic' means that no energy is lost during the interaction so that the resulting wave have the same energy as the incident wave. While plenty of these coherent, elastic scattering events occur during a STEM experiment, there are also many inelastic scattering events that occur concurrently. These result in a scattered electron with lower energy than the incident electron, and so longer wavelengths. While these might (and in fact will) be scattered in the same direction as reciprocal lattice vectors, the change in the wavelength inhibits constructive interference occurring in a way comparable to that described earlier.

For electrons to have lost energy during inelastic scattering, however, the energy must be transferred to somewhere else in the system. This offers two potential avenues for investigating the inelastic scattering process: detecting the energy change by measuring the energy lost by the electron beam after the interaction; or detecting the energy change by measuring the energy in one of the other forms it has been transformed to. These two approaches are broadly complementary and

form the foundations for Electron energy loss spectroscopy (EELS) and Energy-Dispersive X-ray Spectroscopy (EDX), respectively. While this ordering of techniques is perhaps conceptually most logical, in terms of outlining the processes of energy loss it will be most practical to discuss EDX first.

5.9.1 Energy-Dispersive X-ray Spectroscopy

One potential avenue for energy to be lost during the scattering process is by ionising core electrons in the sample. These core electrons gain some non-discrete amount of energy, equal to the difference in quantised energy levels of the excitation plus some amount of kinetic energy, and leave behind a core vacancy.

At some point in time afterwards, the excited atom will relax down to its new ground state. To do this, a reverse process occurs, where an electron in a higher energy orbital than the core vacancy, loses energy of a discrete amount, equal to the difference in quantised energy levels of the de-excitation. This energy is released as a photon of wavelength given by:

$$\lambda = \frac{hc}{E} \tag{5.34}$$

Where λ is the wavelength, h is Planck's constant, c is the speed of light in vacuum, and E is the discrete energy difference.

As the possible energies are discrete, and dependent on the specific atom, these de-excitation events will produce a discontinuous spectrum of X-ray radiation that will be characteristic of the species of the excited atoms. By analysing the spectra of X-rays collected it is possible to infer the chemical composition of the sample.

It is worth noting that quantitative compositional analysis is far more challenging than qualitative as the ratio of peak intensities is influenced by many factors. For

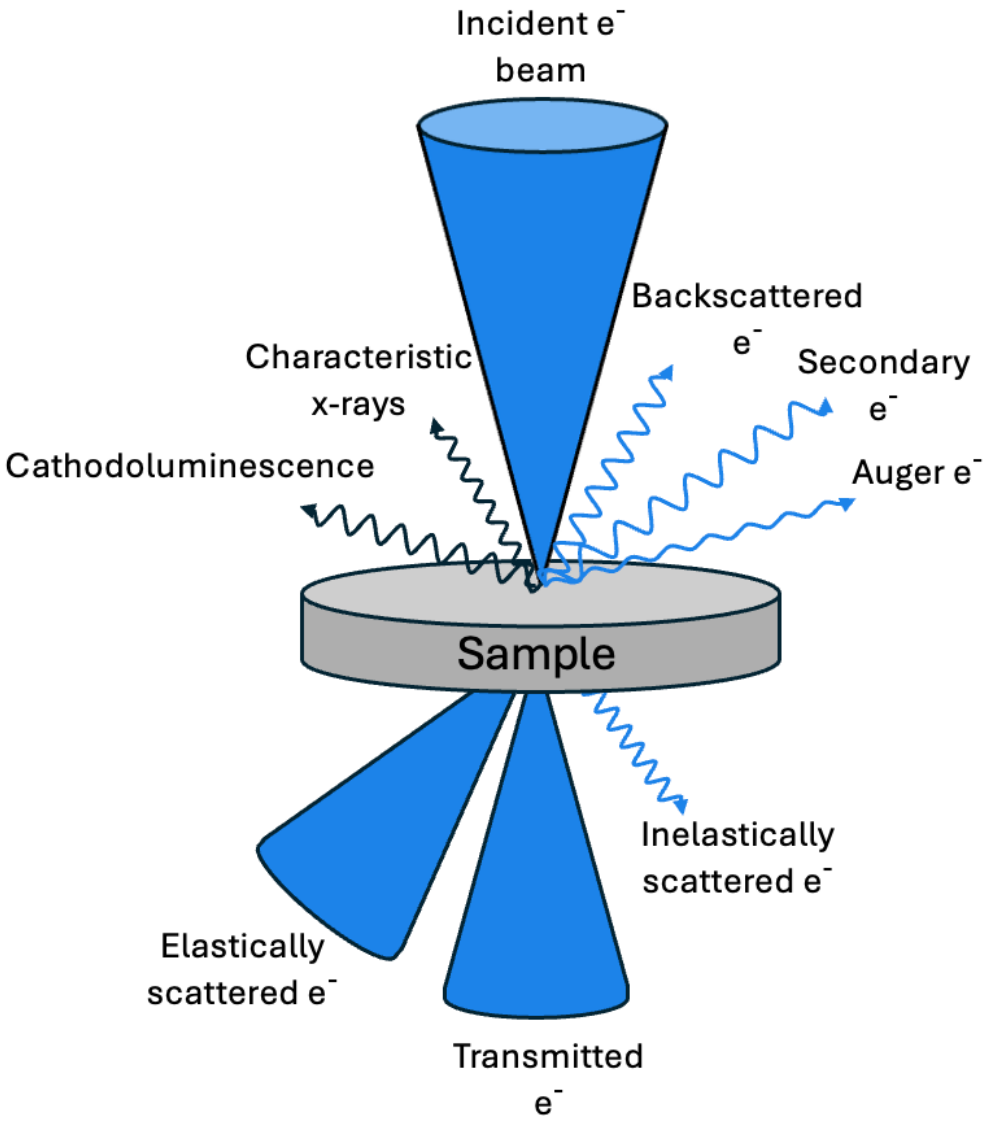


Figure 5.8: Graphic representing the phenomena which can result from the interaction of a sample with an electron beam.

example, even for one type of atom the yield of each particular one of these characteristic x-rays will be different. This is due to the fact that there may be multiple processes, direct and indirect, that can lead to specific de-excitation and these are not all equally likely. These de-excitation processes do not also need to be radiative. A summary of some of the potential direct and indirect relaxation pathways as can be seen in Figure 5.9. Actual quantitative analysis is usually done using experimentally determined ratios of crystals of known compositions.

Additionally to the discrete spectra, there is a continuous background profile known as Bremsstrahlung. Bremsstrahlung is the 'breaking energy' associated by the accelerated electrons transitioning from vacuum to the electric potential of the sample.

These x-rays are be emitted isotropically in all directions, but the nature of EDX detectors mean that only the x-rays in a limited solid angle are collected. As x-ray intensity is strongly dependent on the distance from the sample due to the inverse square law, in order to try and maximise counts, it is typical for the EDX detector to be inserted into the column to get as close to the sample as possible, and sometimes even to have multiple detectors present.

5.9.2 Electron Energy Loss Spectroscopy

Electron Energy Loss Spectroscopy (EELS) is a technique that takes advantage of the other approach described, by measuring the energy lost out of the transmitted beam.

As mentioned in the EDX section, when inelastic scattering occurs, energy is lost from the incident beam to ionise core electrons. The inelastic scattering is biased in the forward direction, so the majority of these reduced energy electrons remain along the same trajectory. Thus by taking the beam that remains on the optic axis

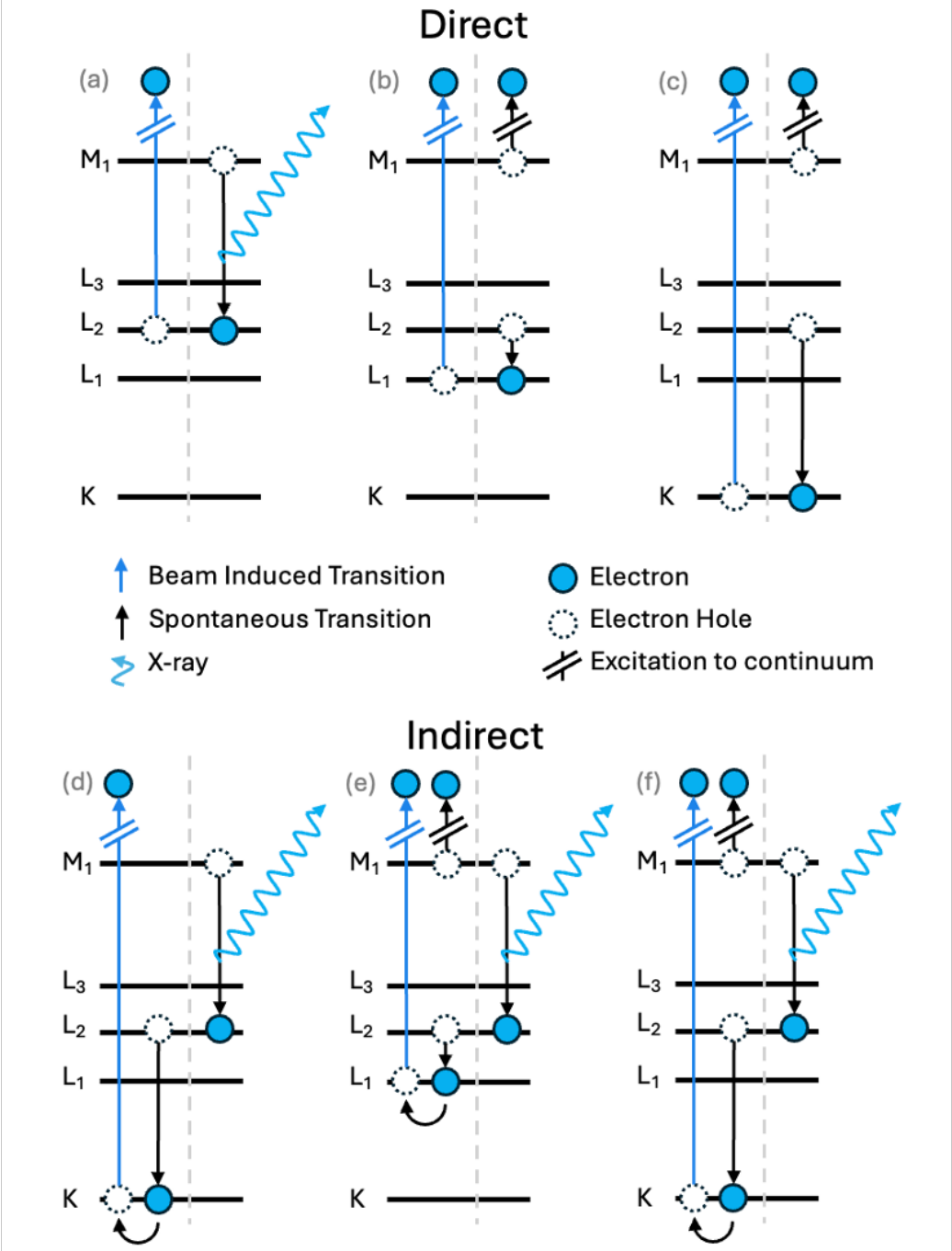


Figure 5.9: Three direct processes by which an excited ionised atom can relax. (a) Radiative transition. (b) Coster-Krönig transition. (c) Auger transition. (d) Radiative to radiative transition. (e) Coster-Krönig to radiative transition. (f) Auger to radiative transition

(potentially with the assistance of a post-sample aperture) and passing it through a magnetic prism, it is possible to spread the beam into different energy electrons based on how strongly or weakly they are deflected by the field. This essentially acts as an energy filter and allows the intensity of different electron energies in the beam to be recorded.

The amount of energy lost by the electrons in the electron beam will be characteristic of the species of atom that the beam inelastically interacted with. As the ionisation event will result in an ionised electron with some amount of kinetic energy the resulting spectra will not be discrete, but it will have a minimum value of the energy loss, and then some distribution of similar but higher energy losses present. This results in 'edges', where distinct steps in intensities occur at these minimum values. The finer structure of the subsequent profiles then contains information about the oxidation state and local environment of the atom.

As well as collecting the information on inelastically scattered electron energy losses, the so-called 'core loss', it is also common to collect the electron energies at or near the incident beam energy. The peak at the same energy as the incident beam is referred to as the 'zero-loss' peak and contains all the unscattered electrons, along with elastic and quasi-elastic scattered electrons along the optic axis. Within a few tens of electron volts of the zero-loss peak is the 'low-loss' region of the EELS spectrum. This is dominated by bulk plasmon excitations, a phenomena akin to the resonant oscillation of electrons mentioned in x-ray scattering, but primarily resulting from the the resonance of delocalised electrons in the valence band of the solid. As these plasmon resonances are properties in valence electrons, which are sensitive to many local influences like bonding and strain, thus changes in the plasmon region of the low-loss spectra can be useful for observing highly localised changes in the structure.

As samples get thicker, multiple plasmon excitations can occur, following a Poisson distribution. This leads to a primary use of the low-loss region - estimating sample thickness. This is done by using the log-ratio of the total spectra intensity compared to the zero-loss peak to approximate the occurrence of multiple excitations.

5.9.3 EDX vs EELS

While the two techniques are quite complementary, both measuring comparable phenomena occurring within the sample, they differ in terms of spatial resolution, energy resolution, signal strength, electron dose, and appropriate beam conditions. It is considerations such as these that mean the techniques are not interchangeable, different experimental requirements will favour one or the other, and there is also merit in obtaining both simultaneously.

The majority of this discussion comparing techniques will be framed in the context of HRSTEM, where the probe is small and the convergence semi-angle is large. While it is possible to collect both EDX and EELs with alternative setups, this is the situation which yields the best spatial resolution and so provides the smallest scale spectroscopic information.

5.9.3.1 Spatial Resolution

The first of these considerations is the spatial resolution. Typically EELS displays improved spatial resolution over EDX, for equivalent probes. This is due to the effects of 'beam-broadening' through the sample. Beam broadening is largely a result of the coherent, elastic scattering discussed earlier, and is a function of both sample thickness and composition. While the amount of beam broadening will not change between techniques (after all, they can be operated concurrently) the effect

of the beam broadening on the data collected does vary between the techniques.

For EDX, the effect is that as the beam broadens, a larger amount of sample is irradiated, and and part of that irradiated sample might undergo inelastic scattering to produce x-rays. These x-rays are essentially just as likely to be detected by the EDX detector as the x-rays from the 'non-broadened' area of the probe. For EELS, the interaction with the sample is the same, however, as the scattering is forward-peaked, the chance of an electron from the edge of the broadened probe scattering near enough to the optic axis to be detected by EELS, is lower than those scattered from the centre of the probe. This results in the improved spatial resolution.

5.9.3.2 Energy Resolution

The energy resolution is dependent on the individual detectors used and the exact procedure by which detection occurs. With advancements in direct electron detectors, EELS detection can now be done using a CCD-style detector similar to those used for imaging. The magnetic prism set-up can within reason, produce any desired spatial spread of electron energies so for EELs this then limits the resolution by the number and size of the pixels in the detector and the range of energy that is desired to be collected at once. In the limit of requiring very fine spatial separation for energies (collect energy at fine resolution over a large range), the diffraction limit once again becomes a factor. This can be adjusted by modifying the angular range of the incident beam allowed to pass into the detector but often requires additional lenses post-sample and will be a trade off with spatial resolution (as larger incident angles increase the effect of scattering from the edge of the broadened beam).

EDX, on the other hand, tends to use some form of semi-conductive crystal to

measure incident x-ray energies. The fact that the x-rays do not need to be spatially resolved means there is no inherent pixel size related limit on the energy resolution. The energy resolution limit on EDX is more time dependent. This is because after the crystal detects a x-ray it is converted into a voltage pulse. While the electronics do allow for discrimination of multiple voltage pulses, the accuracy of this discrimination is dependent on how many pulses occur in that time frame. If x-ray counts are too high, the detector will start rejecting new pulses to allow accurate discrimination of previous ones. This leads to detector 'dead-time'.

This can be mitigated by various strategies, like collection time and beam conditions, but such approaches will all be a compromise between accuracy of energy detection and other factors like electron-dose and spatial resolution (the beam-broadening discussed earlier is dependent on accelerating voltage).

5.9.3.3 Signal Strength and Electron Dose

For EELS, signal strength and electron dose are highly correlated factors. As the beam current is increased there is a greater electron flux through the sample, which leads to more scattering events occurring during a given collection time window. More scattering events means more detected scattering events and so a larger EELS signal is detected. A major factor in EELS is the dynamic range required of the detector, as the features are present on a decaying background function. This requires a large dynamic range to resolve features both where the background is large and where it is small.

For EDX the signal strength and electron dose are also linked, but less directly. The signal strength is the total detected x-rays, so is a function of x-ray production and detection. While increasing beam current will increase the number of x-rays produced and, in turn, the electron dose, this will not necessarily result in an

increase in detection if it just results in more dead-time. The standard approach is to collect for longer, rather than to try and overly increase x-ray production, which also leads to increased cumulative electron dose. Cumulative exposures of lower electron dose is generally regarded as less damaging to samples, especially in scanning modes where relaxation can occur between exposures.

5.9.3.4 Beam Conditions

As noted, the comparison thus far has assumed both experiments to be operating in HRSTEM mode. EDX has another advantage, separate from EELS, which is that it can also be used during a 4D-STEM experiment. As the x-rays are generated and detected independent of the post-sample beam conditions, it is possible to collect EDX signal alongside diffraction data in a 4D-STEM experiment. The primary limitation of this is having the software is in place to achieve the simultaneous collection of the data. The quality of the EDX data will also suffer relative to that collected from HR-STEM, as ideal beam conditions for 4D-STEM, especially SEND, don't necessarily line up well with the ideal beam conditions for EDX. More often than not, the exposure time for SEND is much shorter than is typically used for EDX, so x-ray counts are fairly low; however this can be modified if beam exposure is not a primary consideration. Additionally, the spatial resolution for SEND will also be fairly poor so only more general compositional trends can be observed.

6 | Computational Techniques

Alongside the experimental techniques, this thesis leverages several computational approaches to allow for processing and interpretation of data. Of these, a primary technique employed is the use of Neural Networks.

6.1 Neural Networks

A neural network is a type of machine learning model that contains many interconnected layers of nodes in a manner inspired by neurons in a human brain, hence the name 'neural network'[105].

6.1.1 Basic Summary

Neural networks are typically defined within a set layout of nodes and connections referred to as the network architecture. These nodes are simply containers of a single numerical value, known as the activation. The connections then describe which pairs of nodes are linked together and how the activation of the first in the pair should affect the activation of the second.

When first created, these networks are produced with random strengths for all of these connections. The network then undergoes a 'training' process, where the model is given examples. These examples contain both example inputs and the corresponding expected output. The model receives the inputs and propagates the values of nodes in each layer to the next through its connections, using the currently defined connection strengths to produce an output. This model output is then compared to the ideal output, and the strengths of the connections are adjusted to try and better match the network output to the ideal output.

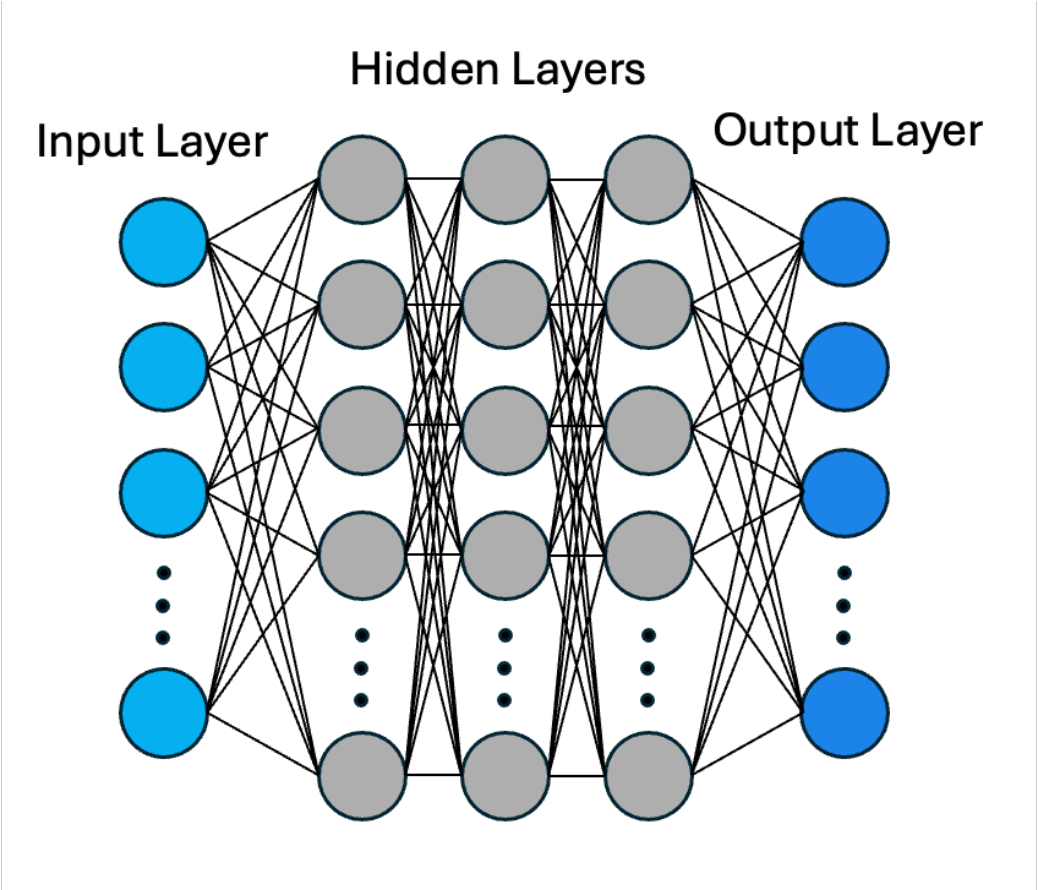


Figure 6.1: Schematic representation of a Neural Network.

This is then iterated, ideally over many, many examples, and once a desired level of precision is reached, the model is said to have been trained. If the network is well-trained such that it can be generalised, it will now be able to take new input data that it did not see in training and produce whatever the ideal output would have been for that input. This is the essence of the neural network. A basic schematic of a neural network is shown in Figure 6.1.

6.1.2 How a model works

While this basic summary provides an insight into how a network functions, it glosses over most aspects of the practical implementation of the individual steps. To understand each step better, they must be inspected with more detail.

6.1.3 Input and Output data

The first place to start with a neural network is with the input data. As the network architecture is predefined, a neural network usually requires input data of a fixed size and shape. For example, the input data might be a spectrum of absorption intensities at 10 energy values. The network would then take each of these 10 values, and each one would be given as the activation of a separate node in the first layer of the network. In principle, the values of these numbers can be whatever is required for the network, but it is typical to use floating-point numbers. It is also often common-place to apply pre-processing to normalise the range of values these input numbers can take[106].

As with the input data, the length of the output data is also fixed by the model architecture, and so care must be taken to organise training output data such that it meets this constraint. Two of the most popular broad-categories of models are regression models and classification models.

6.1.3.1 Regression Model

In the first case, a regression model, the network is being tasked with predicting a value or set of values from the input. In our spectra example, this might be an average oxidation state of an element. As this mean oxidation state would just be a numerical value that we assume can vary continuously, it can be seen how, if we made the final layer in the model architecture a single node (which itself

contains is a single activation value), we could just take the value of this node to be the output of the network. As the model trains, it can improve the accuracy of its predictions by altering the strengths of the connections such that after the input data has been propagated through the network, the final numerical value for this node that is closer to the value provided as the target output data.

It is straightforward to consider that if we wanted to predict multiple values, perhaps the oxidation states for two elements present, we would just increase the number of nodes in the final layer from one to two. While the network can, in principle, predict as many values as desired, this must be decided *before* the model training begins. Also, just because an arbitary number of outputs can be generated in principle, there is no guarantee that the network will be able to predict all these values well. Typically, more output values require larger networks, more training, and more data to achieve the best results[107].

6.1.3.2 Classification Model

The alternative broad case concerns a classification model. Here, rather than seeking to produce a continuous value or set of continuous values, the model aims to assign input data as a member of a discrete class. Again, in our spectra example, this might involve analysing the spectra and assigning it as belonging to a particular element. In this case, it is less obvious to determine what form the output should take. One typical approach is to use one-hot encoding. To do this, we again take the last layer of the network to be the output. In this layer, there should be a node for every class that the input data could possibly belong to. The target, ideal output data is then set up so that it contains a 1 for the index of the class to which it belongs, and a zero for all other classes. It is also typical to apply a function to the last layer of the network that will push values to be between 0 and 1 (this idea of activation functions will be discussed in more detail in Section 6.1.4.1).

As before, the network is trained, and the connections are adjusted such that this output layer matches the one-hot encoded vector as closely as possible. Again, it is trivial to see that if multiple classes are to be assigned, then the value can be 1 for multiple nodes. Finally, as with the classification model, this can be extended to an arbitrary number of classes, but again, this must be defined before training and is fixed by the model architecture.

6.1.4 Input to Output

Once the forms of the input and output data are defined, the next step is exploring how the model converts from one to the other. As previously outlined, models are built up by a series of layers, where each layer contains at least one node. Between these layers, there are connections. These connections are from one node (input node) in one layer to another node (output node) in the next layer. The desired form of the overall connection is a non-linear function which expresses a relationship between the activation value of the input node and the output node. Again, the value of a node is referred to as its activation. This is achieved through initially defining a linear equation for a connection into node j of a given layer, L, and is of the form:

$$z_{jk}^{(L)} = w_{jk}^{(L)} a_k^{(L-1)} + b_j^{(L)} \tag{6.1}$$

Where:

- $z_{jk}^{(L)}$ is the output of the connection between node k in the previous layer and node j in the current layer, L.
- $w_{jk}^{(L)}$ is the weight of the connection from node k to node j (and as this is a weight **into** layer L, it is given the superscript (L)).
- $a_k^{(L-1)}$ is the activation of node k in the previous layer, $L-1$.

- $b_j^{(L)}$ is a scaling factor (bias) and is the same for all the connections into node j in layer L.

The final activation, $a_j^{(L)}$, of the general node, j, in layer L, is the sum of all the connections to the nodes of the layer before, operated upon by an activation function, g. This is given by Equations 6.2 and 6.3 and is shown diagrammatically in Figure 6.2.

$$z_j^{(L)} = \sum_{k=0}^{N(L)} w_{jk}^{(L)} a_k^{(L-1)} + b_j^{(L)} \tag{6.2}$$

$$a_j^{(L)} = g(z_j^{(L)}) \tag{6.3}$$

The activation function is the key to achieving non-linearity, as without it the result of the network would simply be the sum of many linear equations, which is itself a linear equation. This would limit the network to only being able to effectively predict situations where the input and output data are related linearly, which would severely restrict the predictive power of the model. The exact form of the activation function can vary, and is discussed in more detail in Section 6.1.4.1, but for now it is just chosen to be expressly a non-linear function.

For a basic network, the layers are connected *densely*. This means that each node in a given layer has a connection to every node in the preceding layer. The values of all the weights and biases for the network are referred to as the parameters, and it is by changing these parameters that the performance of the network can be improved.

With this understanding of network connectivity and the maths that governs these connections, it is possible to understand how a network can transform input data

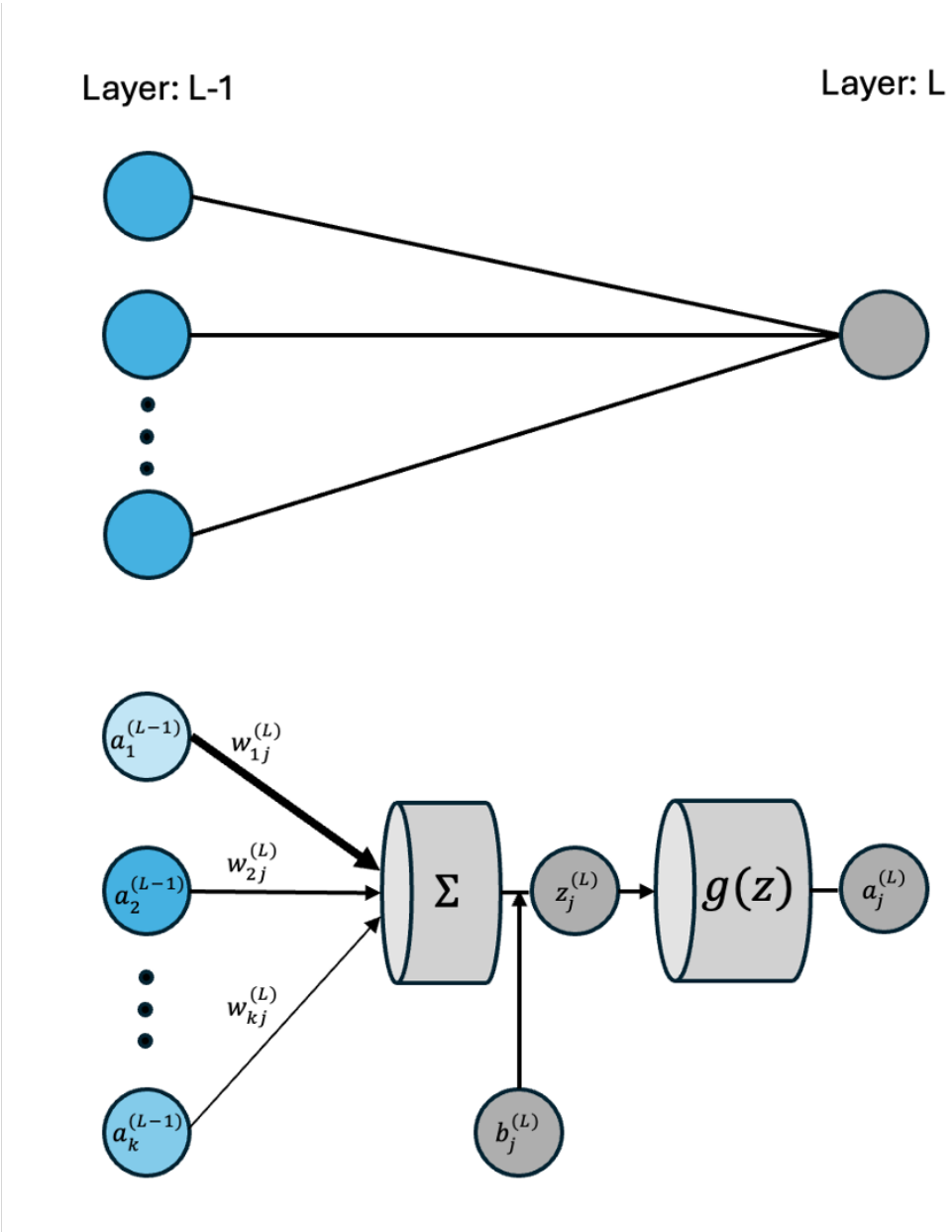


Figure 6.2: (top) Diagram showing the individual connections from one layer to the next, as depicted in Figure 6.1. (bottom) Diagram showing how the activation value of a single node is calculated from these connections.

into output values.

Explicitly, it works as follows. The first layer of the network has a node for each value in the input data (if the input data is our spectrum of 10 intensity values, then the first layer has 10 nodes). The activations of these nodes are set to their corresponding input values. These activations are then propagated through the connections to get the activations of the nodes in the next layer using Equations 6.2 and 6.3. This continues through the full network until the last layer is reached, and its activations have been calculated. As this last layer, by design, has a node for each desired output value, these activations are used as the values in the output data. As this final layer is used as the output, as mentioned earlier, it is common to use a different activation function for these nodes than for the intermediate layers, in order to aid the network in outputting values within a suitable range.

6.1.4.1 Activation Functions

As stated, an integral part of propagation through the network is the presence of activation functions to ensure non-linearity. While these activations can, in principle, take the form of a vast variety of functions[108], most networks rely on the same set of standard functions. These standard functions are: Sigmoid, tanh, ReLU, and Leaky ReLU, and are shown, with their corresponding equations, in Figure 6.3.

Inspecting the plots in Figure 6.3, the effect that each activation function has on the final activation can be seen. The Sigmoid and $\tanh$ functions behave very similarly, as do ReLU and Leaky ReLU. The Sigmoid function will take large positive input values and return an activation very close to 1 and will take large negative input values and return values very close to 0. There is only a small range of input values around 0 for which the Sigmoid function will meaningfully

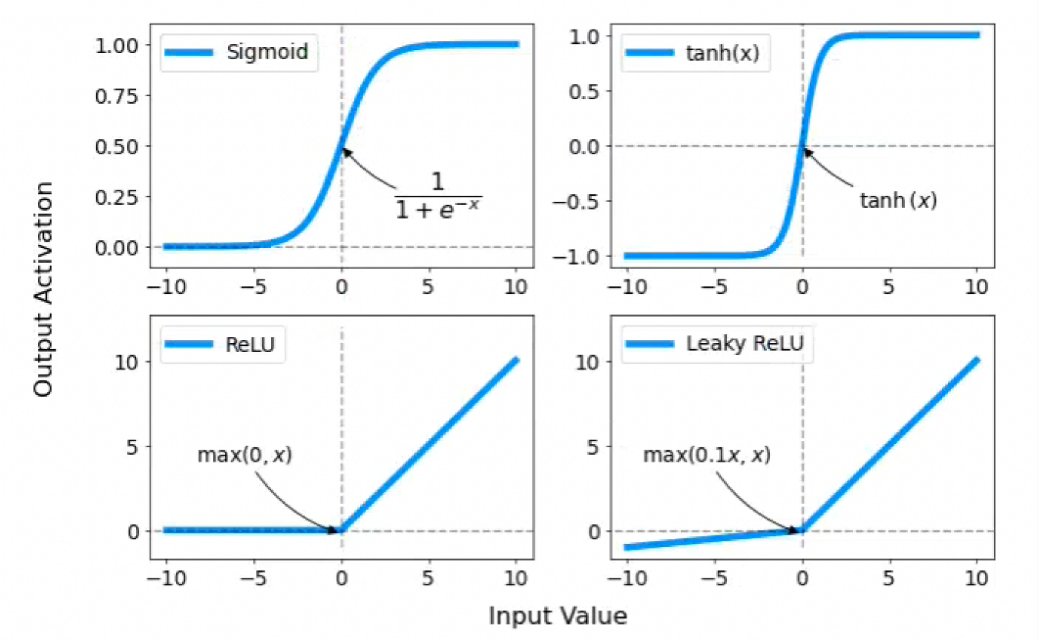


Figure 6.3: Four of the most common activation functions and their equations.

discriminate between input values. This behaviour makes it appealing for the final layer of classification models, as it 'encourages' the model to only assign high or low probabilities to each of the classes. The $\tanh$ function behaves very similarly, as the two functions are linked by the following relationship:

$$\tanh(x) = 2\sigma(2x) - 1 \tag{6.4}$$

The major difference is that whereas the sigmoid function limits the activations to the range between 0 and 1, $\tanh$ limits the activation between -1 and 1.

The ReLU, or Rectified Linear Unit, works differently and is very straightforward. It enforces that all negative values are set to zero. This simple transformation is surprisingly effective, however, as it achieves the desired effect of non-linearity while still allowing the network to express strong preference through large activations. A slight modification of the ReLU is the Leaky ReLU, which softens

the constraint and allows for negative preference to also be expressed but at a dampened intensity (in this example, one-tenth the input value).

6.1.5 Evaluating initial performance

Now that the network can convert input data to output data in a non-trivial (non-linear) manner, these connections need to be optimised. The first step in achieving this optimisation is to know how well the network is already performing the task. To do this, it is necessary to evaluate the outputs of the current network parameterisation against the ideal outputs which have been provided alongside the inputs. Before this can be done, one needs to define a 'loss function'; a function that takes the network output and the ideal output, compares them, and provides a metric for how dissimilar the two results are. The network can then evaluate performance by tracking the change in this metric, as smaller values should correlate to a closer output to the desired one.

As with most variables in constructing a neural network, there are many different loss functions that can be used[109]. The choice of loss function is often informed by the form that the output data takes; mean squared error is a common loss for regression but would be less suitable for the sparse binary data used for classification. Provided the loss chosen is appropriate to the data type, the main effect of the loss is how it directs further training, which will be discussed shortly.

6.1.5.1 Loss Functions

Before discussing how the loss is used to direct training, we will first outline some of the most common metrics used to quantify the difference between the generated outputs and their ideal counterparts. As already mentioned, one of the most common metrics is to use either mean squared error or mean absolute error,

given by Equations 6.5 and 6.6 below:

$$\text{MSE} = \sum_{i=0}^{n} \frac{(y_i - x_i)^2}{n} \tag{6.5}$$

$$\text{MAE} = \sum_{i=0}^{n} \frac{|y_i - x_i|}{n} \tag{6.6}$$

These metrics look to evaluate the total magnitude difference between the generated and ideal output data, with the primary difference between the two metrics being that MSE will place greater weighting on large differences in values, due to the square of the difference.

These two losses can often be sufficient to evaluate difference, but their main limitation comes from the independent treatment of each data point. A trivial example of this limitation is that if there is a small, consistent, systematic offset in predicted value, this will lead to quite a large final error, an error that is perhaps not representative of how close the model is to appropriately generating outputs. Various other metrics will attempt to include some consideration of how the error in one output value is related to those nearby, to try and mitigate issues such as this.

One general way of approaching this is to look at distributions of values rather than the absolute values themselves. Such a distribution based approach would be less susceptible to simple systematic errors as the overall shape of the distribution would be unchanged. This would ideally be reflected in a more appropriate loss.

This property of using distribution-based metrics also allows them to be appropriate for some other data types, such as the probability data used for classification.

Again, as with many aspects of machine learning, there are many different distribution-

based metrics that can be applied to a wide-variety of data types. One such metric is Binary Crossentropy which can be measured between two distributions, p and q, using Equation 6.7.

$$S_q(p) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(q(y)) + (1 - y_i)\log(1 - q(y)) \tag{6.7}$$

To unpack how the Binary Cross-entropy works, the first step is to define the entropy of a single distribution:

$$S(p) = -\sum_{c=1}^{C} p(y_c)\log(p(y_c)) \tag{6.8}$$

Where there are C distinct classes and $p(y_c)$ is the probability of observing the class, y_c. If all the data is of a single class then there is only one possible value of y_c, and $p(y_c)$ will equal 1. In this case the entropy is zero. Alternatively, in the binary case where there are equal amounts of the two classes, $p(y_c) = 0.5$ and the entropy has a maximum value of $\log(2)$. All other distributions of the two classes will have entropy values between these two extremes.

Ideally, to compare the distributions of the generated data to the ground-truth data we would just compute the entropy of each and compare them.

For our model, however, we do not know the distribution from which the ground-truth data is being sampled, if we did, or there would be no point trying to fit it. If we assume that the distribution of our generated data is an approximation of the true distribution it is possible to compute the 'cross-entropy' between two such distributions using:

$$S_q(p) = -\sum_c^C p(y_c) \log(q(y_c)) \tag{6.9}$$

If the approximated distribution is exactly equal to the ground-truth then $p(y_c) = q(y_c)$ and the cross entropy will be exactly the same as the entropy. Otherwise the cross entropy will be necessarily larger. The difference between the cross entropy and the entropy is named the Kullback-Leibler Divergence (or KL Divergence for short), and is given by:

$$S_q(p) - S(p) = -\sum_c^C p(y_c) \log(q(y_c)) + \sum_c^C p(y_c) \log(p(y_c)) \tag{6.10}$$

$$S_q(p) - S(p) = \sum_c^C p(y_c)(\log(p(y_c)) - \log(q(y_c))) \tag{6.11}$$

It follows that as the KL divergence is reduced, the approximated distribution tends towards the true distribution. This is the basis of using the cross entropy as a loss metric, as the actual entropy is fixed, reducing the cross entropy correspondingly reduces the KL divergence and implies a more similar distribution.

To actually compute the cross entropy for a model, each of the N data points will be used to calculate the cross entropy, so:

$$p(y_i) = \frac{1}{N} \tag{6.12}$$

and thus

$$S_q(p) = -\frac{1}{N} \sum_i^N \log(q(y_i)) \tag{6.13}$$

Computing the cross entropy for the full dataset over each of two classes separately gives:

$$S_q(p) = -\frac{1}{N}(\sum_{j}^{N_1} \log(q(y_j))) - \frac{1}{N}(\sum_{k}^{N_2} \log(q(y_k))) \tag{6.14}$$

This formulation however, is not particularly convenient as you must sum over the two classes independently, and this change of summation leads to a subtle change of variable that has been highlighted explicitly by the new indices j and k.

Before there were N samples indexed by i belonging to either class 1 or 2. Now there are two sets of samples, one set indexed by j up to N_1, that all belong to class 1, and another indexed by k up to N_2, that all belong to class 2.

Equation 6.14 can be rewritten in the form of the original Equation 6.7, so it can just be summed over N again. To do this it must be realised that for the binary distributions in Equation 6.14, $q(y_j) = 1 - q(y_k)$. This is because $q(y_j)$ is the probability of class 1 and $q(y_k)$ is the probability of class 2.

Thus we can define a general probability $q(y)$ which is equal to $q(y_j)$ in Equation 6.14. In this case, the samples belonging to class 1 have probabilities $q(y)$, but the samples belonging to class 2 now need to be given by $1 - q(y)$. This is shown by rewriting Equation 6.14 as:

$$S_q(p) = -\frac{1}{N}(\sum_{j}^{N_1} \log(q(y))) - \frac{1}{N}(\sum_{k}^{N_2} \log(1 - q(y))) \tag{6.15}$$

This is sum becomes equivalent to:

$$S_q(p) = -\frac{1}{N}(\sum_{i}^{N} a_i \log(q(y)) - b_i \log(1 - q(y))) \tag{6.16}$$

if $a_i = 1$ and $b_i = 0$ for all i belonging to class 1 and $a_i = 0$ and $b_i = 1$ for all i belonging to class 2. This is exactly the case given in Equation 6.7, with $a_i = y_i$ and $b_i = (1 - y_i)$, as $y_i = 1$ if i is in class 1 and $y_i = 0$ when i is in class 2.

These equations can be generalised and extended to the multi-class case. In such a situation it is more common to be referred to as 'Categorical cross entropy', just 'cross entropy' or 'log loss'. As it can be extended to an arbitrary number of classes it can even be used on continuous data for regression type problems. In such cases 'log loss' is probably the most appropriate term to use as it emphasises the main advantage of using the loss function in such a case - the $\log$ operation. The presence of the $\log$ can be helpful when outputs are activated by exponential based functions such as $\tanh$ or Sigmoid as it will provide larger gradients required for optimisation (this will be discussed further).

6.1.6 Training the network

Once a loss function has been chosen, the final step in producing a working network is to improve the performance to minimise this loss; this process is referred to as training. As described before, this is done by changing the values of the parameters to change how the activations of one layer are influenced by the activations of the previous layer.

It is assumed that if the model is suitably large, there exists some parameterisation for which the network produces the desired outputs. It can also be assumed that this will correspond to the minimum loss value.

The problem can then be considered as a p-dimensional minimisation problem, where p is the number of parameters of the network, as for each point in this p-dimensional space there will be a value for the total loss. All possible parameterisations of the network then form a surface in this p-dimensional space. This

surface can be referred to as the cost function. In order to find the optimal parameterisation, one simply needs to find the global minima of this surface. Finding this global minima is incredibly challenging, especially as what are considered small models have more than a million parameters. Fortunately, it is rarely the case that it is necessary to find the true global minima, so the actual target is to find a *good enough* local minimum. To find such a minimum, there are, again, many options of optimisation algorithm available, often referred to as the network's 'optimiser'. While the nuance of each optimiser will vary, almost all are built on the same two principles of gradient descent and back propagation.

6.1.6.1 Gradient descent

Gradient descent is a very common approach to finding surface minima. The general idea is that for a given point on a p-dimensional surface, the route to the best global solution can be found by taking steps where each individual step reduces the value of the surface as much as possible. A step to such a point will also have the largest negative gradient associated with it (greatest change in surface value), and so by evaluating the gradients in different directions, this step can be identified. As the minimum is approached, these gradients are likely to get smaller and smaller until, at the minimum, the gradient in all directions should be positive. At this point, the algorithm can determine it has found a minimum point. This is the essence of gradient descent.

6.1.6.2 Back propagation

The second principle, key to finding the minima, is 'back propagation'. As outlined, to find the optimal step, one needs to determine the gradient of the cost function. This gradient can be computed as the derivative of the cost function with respect to each of the parameters. The complexity of back propagation is

that the derivative of the cost function with respect to each parameter is not necessarily independent of the other parameters.

As the cost function is an evaluation of the loss between the network output and the ideal output, it is, in the first instance, dependent on the activations of the nodes in the final layer. These activations themselves are, by Equations 6.2 and 6.3, dependent on both the weights and biases of the connections into them ($\mathbf{w}^{(L)}$ and $\mathbf{b}^{(L)}$) **and** the activations of the previous layer. These activations of the previous layer are then further dependent on the weights and biases of their connections ($\mathbf{w}^{(L-1)}$ and $\mathbf{b}^{(L-1)}$) **and** the activations of the layer before that. It thus follows that if the effect of one of the parameters (a particular weight or bias in layer L) on the cost function depends on an activation in layer $L-1$, and this activation depends on the parameters described by the collections $\mathbf{w}^{(L-1)}$ and $\mathbf{b}^{(L-1)}$, then the effect of the parameters in layer L is dependent on the parameters in layer $L-1$. Therefore, to compute the derivative with respect to all parameters, it is necessary to compute the effects layer by layer, starting at the last layer and working backwards through the dependencies, hence 'back propagation'.

As the network is predefined and connections usually don't change, these derivatives can be computed using the same formula each time. The layer dependence of the derivative is in fact just a standard implementation of the chain rule from Calculus.

Stated explicitly there are three forms of chain rule equations that will determine the complete derivative of the cost function, ∇C, with respect to all the parameters of the network. These are the partial derivatives with respect to the weights, biases and activations (recalling that although activations are not parameters that can be changed directly, they are dependent on previous sets of weights and biases). These equations appear notationally complex, but are perfectly straightforward.

The partial derivative of the cost function, C, for a single training example, 0, with respect to a weight, $w_{jk}^{(L)}$, for a single connection between nodes j and k in the output layer (L) is given by Equation 6.17:

$$\frac{\partial C_0}{\partial w_{jk}^{(L)}} = \frac{\partial z_j^{(L)}}{\partial w_{jk}^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}} \tag{6.17}$$

Inspecting this chain back to front is quite intuitive, as it is simply following the dependencies set out previously. A change in the cost function at an output node, j, in output layer (L), is dependent on the activation of the node, given by $a_j^{(L)}$. This activation is determined, initially, by Equation 6.3 and so is dependent on $z_j^{(L)}$, hence the $\frac{\partial a_j^{(L)}}{\partial z_j^{(L)}}$ term. Finally, from Equation 6.2, $z_j^{(L)}$ is directly dependent on $w_{jk}^{(L)}$, hence the first $\frac{\partial z_j^{(L)}}{\partial w_{jk}^{(L)}}$.

The partial derivative for the bias, $b_j^{(L)}$, applied to node j in layer (L) can be given similarly as:

$$\frac{\partial C_0}{\partial b_j^{(L)}} = \frac{\partial z_j^{(L)}}{\partial b_j^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}} \tag{6.18}$$

The partial derivative for the activation, $a_k^{(L-1)}$, of node k in layer $(L-1)$, is slightly more complicated due to the dense connections. this node k in layer $(L-1)$ is connected to each node l in layer (L) and so will have an influence on each value $z_j^{(L)}$, and so further will influence each value $a_j^{(L)}$. As each of these $a_j^{(L)}$ can affect the cost this must be accounted for in the partial derivative. This is shown in Equation 6.19

$$\frac{\partial C_0}{\partial a_k^{(L-1)}} = \sum_{j=0}^{n_L-1} \frac{\partial z_j^{(L)}}{\partial a_k^{(L-1)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}} \tag{6.19}$$

The change is that the effect on each of the n_L nodes, j, in layer (L) is being accounted for and the total effect summed.

These expressions don't change much for layers before the output layer, the dependencies are just tracked back through longer chains of partial derivatives.

Finally, it is worth noting that this has all been for calculating ∂C_0, a single training example. ∂C will be the average of all the training examples and it is these partial derivatives that are formally required to calculate ∇C.

For computational efficiency, however, it is common not to compute the derivative of the entire cost function, ∇C. Rather, it is more common to run the training in batches and use the derivatives of these to approximate ∇C. These batches are collections of multiple inputs. The benefit of batching data is twofold: it speeds up the gradient calculation for the optimisation; and batching allows the forward and back propagation steps to be run in parallel due to the fact that the underlying transformation can be written as matrix multiplications. In practical terms, the batch size (number of inputs used in parallel) is limited by the available memory of the system, as larger batches require larger matrices, which require more memory to represent.

Using the principles of gradient descent and back propagation, it can be seen how the model can be trained and improved. The input data is passed through the network to generate output data. This output data is compared to the ideal data, the loss is calculated and the gradient of this loss. The gradients are used by the specific algorithm of the optimiser to select new values for the parameters[110]. This process is repeated until suitable parameter values have been found.

6.1.7 Evaluating final performance

Once the network has been trained, it is then common to test the network to understand how well it is performing. To do this, one usually sets aside some data, before the training starts, that the network never sees during training. Then, after training has completed, this data is fed through the network, and the loss function is used to give a more representative reflection of how well the network has learned. This is done as often networks can become very good at working with the specific data they are trained on, but do not generalise well to similar but novel inputs. When this happens, the network is referred to as being overfit. Preventing overfitting is just one way in which a fully trained network can be improved.

6.1.8 Improving a network

Reducing overfitting can be one way in which a network is altered to improve its performance. There are many strategies employed to try and mitigate overfitting. One such example is including 'dropout' layers. Dropout layers are the same as a regular layer except the connections are given a certain probability of not contributing to a successive nodes' weighting. This helps prevent individual connections from having too large an influence on the final output, something which can easily become an issue when new data is introduced after the training, that activates this connection in a way not seen in the training. Other ways of improving a network are through hyper-parameter tuning.

6.1.8.1 Hyper-parameter Tuning

The weights and biases are the parameters of a given network - the variables that can be changed in training to affect the position of the network on the cost function surface. The hyper-parameters, by extension, are the variables that can be changed

when defining the network, and these are variables which can change either the cost function surface itself or how it is traversed upon optimisation.

The first type of hyper-parameters tend to be associated with the network architecture: how many layers there are; how many nodes in each layer; the type of connections between layers. Additionally, other types of hyper-parameters can be specified for the optimiser function, depending on how the specific algorithm works, there might be a variable that controls the size of steps taken once the step direction has been calculated, usually called the learning rate.

By altering both type of hyper-parameters, ahead of the training process, the final outcome of the network can be drastically altered. This is because while two networks with different hyper-parameters should still both have the same global minimum (provided both networks are sufficiently large), the ease of finding them will depend on the exact surface and optimisation algorithm, both of which depend on the hyper-parameters.

Finding the optimal hyper-parameters is no trivial task[111]. With the vast possibilities of hyper-parameter space, it is impossible to try all sets. Even if one could test all sets of hyper-parameters, there is no guarantee that the value reached in testing is the global minimum for the training data, let alone that this is the global minimum for a generalised sample.

To expand on this, the cost function surface for a model gives the total cost for each parameterisation of that model but it is only calculated using the samples available for training. This cost function surface, therefore, is only an approximation of the complete cost function surface, one where the cost is evaluated for each parameterisation for every possible input. This complete cost function is the one for which the global minimum would provide the best generalised model for all possible inputs. As there is generally no way to sample the complete cost func-

tion, this approximation is usually a necessity, but it does mean that our choice of training sample matters if we want to compare different models' performance, as they will produce different approximations of the overall cost function surface. This is the reason for the separate test data explained in Section 6.1.7, as it allows the model to train and optimise on one approximation of the cost surface, but be evaluated upon its performance on another approximation. The best general model should perform optimally on both.

This approach can lead to an issue when comparing different hyper-parameters. As already noted, changing the hyper-parameters will change the cost function surface. The different approximations of the complete cost function surface will, in general, change differently with the hyper-parameters. Therefore, when comparing different hyper-parameters, if simply done by comparing performance on test datasets, there would be a bias to find parameterisations where the training cost function surface closely resembles the test surface. While this can happen because both are well approximating the complete cost function surface, there is already a bias in the training process to match these same two surfaces. As both are just approximations of the desired cost function surface, the risk is that this bias is more likely to find two arbitrarily similar poor approximations, rather than two similar good approximations. In such a case, the model would not generalise well to new data.

To avoid this, when hyper-parameter tuning, it is common to separate off a third 'validation' dataset. This can be used as another approximation of the complete cost function surface, but with the benefit that it has not been used to direct training or model selection before this point. This means that if a model is performing well on this validation set, it is more likely that the approximation of the cost function surface is a good one, and that the model will generalise well.

With the validation data split off, as well as the training and test data, each hyper-parameterisation can be investigated. For each hyper-parameterisation, the models can be trained on the training data and evaluated on the test data. The best performing hyper-parameterisations for on this test data can then be evaluated and compared on the validation data, and the model with the lowest cost on this validation set is selected. The means of selecting which hyper-parameterisation to investigate can vary, but methods range from sparse random selections of options, to grid search, to Bayesian optimisation techniques[112]. No ideal approach is agreed upon, and the challenge is to just find a good enough set of hyper-parameters.

Selecting the relative sizes of these training, test, and validation sets is a trade-off between approximation quality and approximation confidence. The larger the training dataset, the better an approximation of the complete cost function surface will be, and so the training will be more effective at optimising the parameters to minimise the position on this surface. Large training sets, however, will lead to smaller test and validation sets, which will in turn be worse approximations of the complete cost function surface. Thus, using these to evaluate the model performance will be less reliable. In practice, the relative sizes must be chosen based on the use case.

6.2 Convolutional Neural Networks

One specific type of network that is very popular for use with two-dimensional image data is the convolutional neural network (CNN). The CNN is, in principle, very similar to the standard neural network described above but with some additional complexity in the form of convolutional layers.

6.2.1 Convolutional Layer

The way the convolutional layer works is relatively straightforward and offers intuitive insight into why they are so popular when working with image data[113][114][115]. The convolutional layer is made up of N, $k \times k$ kernels. These kernels act like filters in traditional image processing. The function of the kernel is best explained by example.

Take a 3×3 kernel operating on a 7×7 image as shown in Figure 6.4. The 7×7 image is just a two-dimensional array of numbers as is the 3×3 kernel. It can be seen that to generate the first value of the output, the first 3×3 patch from the image is taken and element-wise multiplied with each value in the kernel. The value in the final output is then the sum of this element-wise multiplication. For the next value along in the output, the image patch would translate one column along in the image data, and the same element-wise multiplication followed by sum would take place. This repetition of the calculation across the array is the convolution.

From this basic description, it can be seen that the output value depends on the values in the kernel and the values in the image patch. By altering the values in the kernel, different features in the image patch can be selected to produce large activations in the output. It is these kernel values that are the parameters changed during the learning process. As each kernel can be seen as evaluating for a certain feature in the image patches, the use of N kernels then becomes evident, as each of these kernels can be trained to extract different features. This is shown in Figure 6.4 as the orange, cross shaped kernel gives a stronger response for the first image patch than the second. What is not shown in the figure, but is also possible, is for the kernels to take negative values, which can be used to further improve contrast of feature detection.

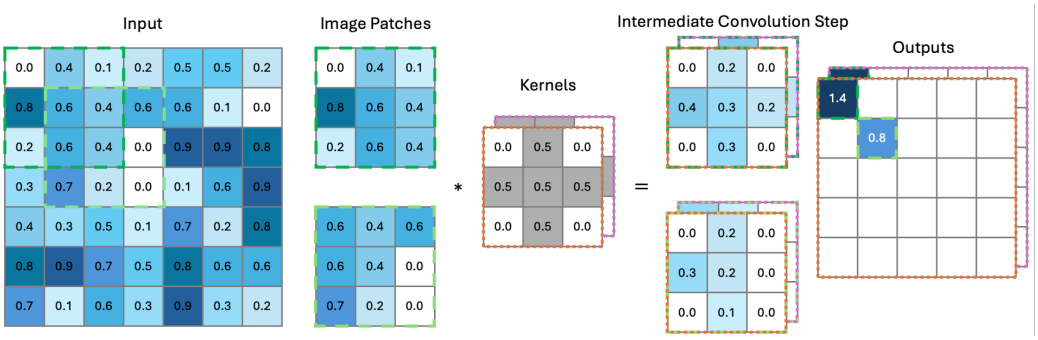


Figure 6.4: Example of how a 2D convolutional layer works. It shows how multiple 3×3 image patches can be taken from the 7×7 input (shown in the different shades of green). Each of these are then convolved with 3×3 kernels by multiplying values element-wise to produce the 3×3 intermediate arrays. The elements of the intermediate arrays are then summed to give the final value for each image patch in the output. It can be seen that each kernel (shown here in pink and orange) will produce a unique 5×5 output

The performance of these convolutional layers will also depend on the hyper-parameters of the layers. These are the kernel size, the number of kernels, the step size, and the padding. The effect of kernel size and number of kernels are straightforward. Larger kernels will allow for more specificity in the type of feature extracted, and more kernels allow for more features. The step size and the padding we have yet to cover.

In the initial example, the step size was 1, and the padding was 0. Here, the kernel stepped one position in the input image for each position in the output array. It can be seen that this step size, paired with the kernel size, results in an output of size 5. If it was desired that the output should be the same size as the input, one such way to achieve this would be to pad the input array with some additional values so that the centre of the kernel can be placed on each position of the input array. Such padding can occur using zeros, another constant value, or some symmetry-derived filling from the existing values within the input. By choosing different values of padding, step size, and kernel size, different regions can be selected, and different

sized outputs produced.

A standard architecture for a CNN will use two-dimensional image data, a few convolutional layers to extract information about image features, and then flatten the resulting output down into one long one-dimensional array. This can then be used as the input to the traditional neural network, with densely connected layers, as described earlier.

6.3 Autoencoders

As stated in the initial description of the neural network: training of a network requires both input and output data. Input data act as the initial activations, and desired output data are compared to the final activations to assess network performance. Situations where the desired result of the network can be provided directly as output training data are referred to as 'supervised learning'. In some situations, however, the actual insight that one desires to be learned by the model isn't the final output, but something else. These cases are called 'unsupervised learning'. A classic example of unsupervised learning is the autoencoder[116][117][118].

The goal of an autoencoder is to learn a compressed representation of the input data. For example, if the input data were a 256×256 grayscale image, the autoencoder might look for a *compressed* 2-dimensional description that could be used to reconstruct the original image. Such a system naturally suggests two required components: an encoder, which will take the original data and convert it down to the compressed representation; and a decoder, which will take the compressed representation and reconstruct the original data. As the form of the compressed signal is not known beforehand, this characterises this task as an unsupervised learning problem.

These two components, while straightforward in description, cannot be trained as separate networks. The encoder has input data but no ideal output data (the compressed data is unknown), and the decoder has ideal output data but no input data. The solution is perhaps obvious: to train the two as one network and then extract the two components afterwards. In the case of the combined network, there is now input and ideal output data, as the ideal output is simply the reconstructed original input data. This combined network can thus be trained to pass the input data through the intermediate layers of the network, to recreate the input data in the activations of the final layer. If this can be done successfully (with a suitably low metric for loss), then this means that all the information required to reconstruct the original image must be present in the combined information of the activations of any given layer and the parameters of the connections between all the subsequent layers.

If such an intermediate layer is, by design of the architecture, the size of the desired compressed representation, then the network up to this point can be used as the encoder. The output of this encoder is then, as required, a compressed representation of the data. The remaining layers of the network are then used as the decoder. This decoder will again, as required, take the compressed representation and reconstruct the original data. The dimensional space that the compressed representation occupies is referred to as the latent space.

Once this has been trained, the compressed representation can often be leveraged to give insight into the original input data. One such type of insight is that, as the decoder parameters are constant for all the compressed inputs, all the variance information present in the final output data (which is the same as the encoder input data) must come from the input compressed representation. Therefore, distribution trends in the data in its compressed representation can be used to indicate the distribution trends in the initial input data.

6.3.1 Variational Autoencoder

A slight modification on the standard autoencoder approach is to use a variational autoencoder[119]. This works in largely the same way, with the same approach of training the encoder and decoder as one network, but instead of directly using each activation of the encoder output as the compressed representation, instead two activations are learned for each dimension of the compressed representation. These activations are used to define a probability distribution from which the compressed representation is then sampled. The standard sampling is for a Gaussian distribution where the probability of sampling a point z in dimension x is given by Equation 6.20

$$P(z) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(\frac{z-\mu}{\sigma})^2} \tag{6.20}$$

Where μ is the mean, in dimension x, and is described by one of the learned activations, and σ is the standard deviation, in dimension x, and is the other learned activation.

This has the advantage of generalising the latent space, as regions rather than discrete points become associated with variations in the final output data and so encourages input data with similar features to be encoded near to each other, which is convenient for further interpretation.

This concludes the extent of the background on Neural Networks and Machine Learning based techniques implemented in this thesis. The remainder of the chapter will cover other computational techniques for processing and analysis that are also leveraged.

6.4 Rietveld Analysis

Rietveld refinement, in its most basic conceptualisation, is simply a technique that attempts to fully replicate the experimental data from powder diffraction experiments, such as PXRD or PND, using computer simulations[120]. It accomplishes this by utilising the physical equations that govern the experiment and taking starting parameters for the variables that change from one experiment to another (sample cell parameters, sample compositions, background profile, peak shapes, preferred orientation, etc.). The next step in the refinement process is to simulate the diffraction data one would expect to obtain from an experiment with these parameters and compare it to the actual experimental data, typically using a least-squares loss function. The refinement is then iterated, selecting new estimates for experimental parameters, simulating this new data, and recalculating the loss function.

This process is repeated until the loss function is minimised. The final fit is then evaluated by various metrics[121], such as the Weighted Profile Residual (R_{wp}, 6.21) and Goodness of Fit (χ^2, 6.22), but ultimately, it is up to the user to determine whether the fit is meaningful or whether the initial simulation parameters must be expanded or altered.

$$R_{wp} = \left(\sum_{i}^{n} \frac{w_i (Y_i^{\text{obs}} - Y_i^{\text{calc}})^2}{\sum_i^n w_i (Y_i^{\text{obs}})^2} \right)^{\frac{1}{2}} \times 100\% \tag{6.21}$$

$$\chi^2 = \sum_{i}^{n} \frac{(Y_i^{\text{obs}} - Y_i^{\text{calc}})^2}{n - p} \tag{6.22}$$

6.5 Principal Component Analysis

Principal Component Analysis (PCA) is a computational technique for data reduction[122]. The aim of PCA is to take an m-dimensional dataset and redefine these axes using linear combinations of the original axes to generate new orthogonal axes. These new axes should be selected so the first axis attempts to capture a maximised proportion of the variance within the initial dataset. Each subsequent axis then maximises its proportion of the remaining variance. By doing this, the data can be reduced by truncating the transformed data to only include the first n dimensions, utilising the fact that these will contain the majority of the variance that was present initially.

The specifics of selecting these axes can vary from algorithm to algorithm, but the most basic approach is to construct a covariance matrix and then solve for the eigenvalues and eigenvectors of this matrix to define the principal components.

6.5.1 Covariance Matrix

The covariance matrix is a symmetric square matrix, $\mathbf{C}$, where each element, C_{ij}, is the covariance between the jointly distributed, real, random variables I and J.

Covariance is a measure of how the variability in one variable is correlated to the variability in another. The covariance can be calculated as the expected value of the deviation of I from its expected value, $\mathrm{E}[I]$, multiplied by the deviation of J from its expected value, $\mathrm{E}[J]$. This is given by Equation 6.23.

$$\mathrm{cov}(I, J) = \mathrm{E}[(I - \mathrm{E}[I])(J - \mathrm{E}[J])] \tag{6.23}$$

Which can be simplified to:

$$\mathrm{cov}(I, J) = \mathrm{E}[IJ] - \mathrm{E}[I]\,\mathrm{E}[J] \tag{6.24}$$

Once the covariance matrix has been calculated, obtaining the principal components is as simple as calculating the eigenvalues, their associated eigenvectors, and then sorting the resulting eigenvectors based on the eigenvalues.

The eigenvectors are suitable for the new axes as the requirements of the new axes were that they: are orthogonal to each other; and correspond to the dimensions which span the total variance of the data. The eigenvectors are necessarily orthogonal as the covariance matrix is a Hermitian matrix (a matrix equal to its conjugate transpose) and the eigenvectors of Hermitian matrices are always orthogonal. These covariance matrices will always be Hermitian as they are always square (of size $m \times m$); they contain real values (we defined our input variables as such), so the conjugate is unchanged; and element C_{ij} is the same as C_{ji} by equation 6.24.

The second requirement of the new axes is that they each contain the maximum variance possible. An alternative framing of this would be to have axes such that there is no correlation between the variance of one axis and the variance of another (as if there was, the first axis would not be capturing the maximum variance it could, as some of its variance is 'shared' with this other variable). Such a case would result in a diagonal covariance matrix.

Mathematically, the transformation that changes the covariance matrix of the original variables to a diagonal covariance matrix can also be used to change the original variables to the eigenvectors. This transformation is given by:

$$\mathbf{S}^{-1}\mathbf{C}\mathbf{S} = \mathbf{\Lambda} \tag{6.25}$$

Where: $\mathbf{S}$ is the matrix with each column given by a different eigenvector of $\mathbf{C}$; $\mathbf{S}^{-1}$ is the inverse of this matrix $\mathbf{S}$; and $\mathbf{\Lambda}$ is a diagonal matrix where each diagonal value λ_{xx} is the eigenvalue corresponding to the eigenvector in column x of $\mathbf{S}$.

As this new diagonal matrix represents the covariance of the newly defined variables, and the covariance of any variable with itself ($\mathrm{cov}(XX)$) is just the variance of that variable (σ_x^2), then these eigenvalues must correspond to the variance of the new variables. Thus, by sorting the eigenvectors (in descending value by eigenvalue) and truncating after the desired number of components, it is possible to obtain a reduced representation of the dataset.

6.5.2 Non-negative Matrix Factorisation

Non-negative matrix factorisation (NMF) is another form of principal component analysis. Unlike the method described previously, which allows any combination of variables to create new variables, NMF requires that original components are combined only additively. This constraint is applied when it is desired that the resultant components have more physical appropriate meanings. It is often used in signal processing[123][124], where it is usually a constraint that composite signals can only be additive combinations of the basic signals. Formally, this is given as:

$$\mathbf{V} \approx \mathbf{WH} \tag{6.26}$$

where $\mathbf{V}$ is the matrix, $\mathbf{W}$ is the transformed data, and $\mathbf{H}$ is the basis vector. The constraint of NMF is that $\mathbf{V}$, $\mathbf{W}$, and $\mathbf{H}$ must all be entirely > 0. Equation 6.26 does not have a closed-form solution like the eigenvalue approach described earlier and must be solved by iteration using different algorithms.

6.6 Clustering

A fundamental data analysis technique used in this thesis is cluster analysis or clustering. Clustering, in simple terms, is grouping together data points based on some notion of similarity. The metric of similarity can have a drastic effect on which clusters are produced, as well as the granularity of the similarity, that is to say, how similar is similar enough to group data points together. Crucially, the ideal clustering is dependent on the bias of the experiment, and so the granularity of clustering can be key to obtaining useful clustering results.

There are some standard clustering algorithms that are well documented in the literature and well implemented in most data analysis software. Here, we will briefly cover three: K-means clustering[125] (perhaps the most popular clustering approach); Gaussian Mixture Model[126] (another common approach); and DB-SCAN (Density-Based Spatial Clustering of Applications with Noise)[127] (the approach primarily used in this work).

In general, the quality of a clustering algorithm is dependent on the input data and, as previously mentioned, the requirements of the clustering. Different approaches will have different pros and cons that will be addressed in each section.

6.6.1 K-means clustering

K-means clustering is an approach that strives to group data points such that the 'inertia' of each cluster is minimised. This is given by the following equation:

$$\sum_{i=0}^{n} \min_{\mu_j \in C} (||x_i - \mu_j||^2) \tag{6.27}$$

This equation strives to, for samples, i, described by a vector x_i, within a cluster

C; minimise the sum of squares difference between all the vectors and the vector describing the 'centroid' of the cluster μ_j. In general, the number of centroids is required to be predefined and then the minimisation occurs by optimising the position vectors, μ_j.

Considering this, we can determine the advantageous use cases but also the limitations of this approach. The first limitation is, straightforwardly, that the number of clusters must be appropriately predefined. Without an appropriate number of starting clusters, the approach cannot produce a satisfactory segmentation. In many cases, this might not be a trivial task and may require some trial and error or testing a range of potential values.

While the second limitation may seem trivial when stated, it is worth explicitly considering: that by optimising to minimise inertia, this necessitates that the goal clustering is one in which the final state will have the inertia minimised. In many situations, this may be the case, hence the prevalence of this technique, but it is by no means ubiquitous. In general, this assumes that the desired clusters are isotropic and convex, as only such clusters will have minimised inertia.

Another limitation is less obvious but equally important, and that is the cluster distributions. The number of samples in a cluster and the variance of the positions of such samples will both influence the calculated inertia and so the final clustering.

Finally, k-means is limited in its granularity; it is not straightforward to adjust the level of similarity that is required within clusters. The best attempt at this will be to modify the number of centroids as, in general, the more clusters there are, the smaller the population of each cluster and so the smaller the variance would be expected to be. This is not reliable, however, due to the dependence of the algorithm on cluster size, shape, and distribution. As such, adding more clusters will not necessarily increase the granularity of the clustering in a directable way

but will be biased by the fact that it is searching for the minimum inertia solution.

6.6.2 Gaussian Mixture Model

The Gaussian Mixture Model (GMM) is a logical extension of K-means clustering that initially tackles one of the major problems - anisotropy. Whereas k-means looked to minimise the inertia of clusters around a centroid, GMM simply assumes that each sample could be taken from a Gaussian distribution, centred at some given centroid, with a specific distribution in each dimension. GMM then aims to optimise the parameters of the mixture of Gaussian distributions to maximise the likelihood that the observed points would be sampled from these distributions. As the Gaussian is not limited to being equivalent in each dimension, this allows the model to fit anisotropic clusters much better than k-means. Additionally, the use of Gaussian distributions to fit the data allows for the process to be much more robust to the problems caused by uneven cluster distributions for k-means.

Despite these improvements, GMM is still limited by predefining the number of clusters and that these clusters form convex distributions. Again, somewhat obtusely, this can be extended to the general issue that as the model assumes the data can be sampled from Gaussian distributions, it only works for situations in which this is the case. Additionally, the approach also struggles if the clusters are particularly sparse, as the covariance matrices required to calculate the maximum likelihood become less well-defined. This can then pose a challenge if particularly fine grained clustering is required.

6.6.3 DBSCAN

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is the final clustering approach that we will summarise. In contrast to K-means and

GMM, this approach does not require the pre-definition of the desired number of clusters. Instead, it is a neighbour based approach that looks to classify points as: core samples; non-core samples; or outliers. The core samples are defined based on two parameters, a 'distance' and a 'minimum number of samples'. If a data point has more than the defined minimum number of samples within the specified distance it is in a sufficiently point dense region to be considered core. Further, any of these core samples that are connected together are considered to be part of the same cluster. Any point within the distance of these core samples that does not meet the minimum neighbour threshold is considered a non-core sample. This is considered to be part of the same cluster as the core sample it is connected to, but does not further propagate the cluster to other non-core samples it is connected to. A point that is not within the distance of a core sample is considered an outlier. Using this approach with the connectivity of core and non-core points, it is possible to deterministically produce a clustering of the input data.

One of the main advantages to this approach is the granularity control offered by the two thresholds as a way of influencing the level of similarity present within clusters. Additionally it is much more topologically flexible to clustering then the previous two methods as provided the connectivity is in place there are no requirements of isotropy or convexity.

The disadvantages, however, are that as these parameters have such fine grain control, setting them tends to be a manual process requiring adjustments to optimise, which can make it unsuitable for autonomous workflows. As before, an obtuse limitation is that as clustering is done by assigning core samples based on the density of points in the vector-space, the approach is highly sensitive to changes in cluster densities (approximately equivalent to cluster variance). If ideal target clusters exist with large differences in densities, there is possibly no paramterisa-

tion that can allow for desired clustering within the standard DBSCAN algorithm. Hierarchical approaches have been developed[128] to deal with is issue, but at the trade-off of more parameters and likely more user-intervention.

6.6.4 Structural Similarity Index

As alluded to at the start of this section (Section 6.6), clustering needs a kind of similarity metric in order to define clusters. For our examples, the applied metrics have been Euclidean distance, as this is a very tangible and intuitive comparison between data points.

Different metrics for comparing data points will give different evaluations of similarity, akin to the discussions of loss functions earlier, also seeking to quantify similarity. For the 2D image data used throughout this thesis, there was an additional comparison metric that proved useful: the Structural Similarity Index (SSI).

SSI is a useful metric for evaluating image similarities as it attempts to account for global factors in similarity as well as simple pixel-wise comparisons. This allows it to function more closely to human perception, which is a good benchmark for the comparison required for image data. The SSI computes three separate metrics for the images and then compares them using different cost functions, finally combining all three to give a final score. The general SSI calculation is given in Equation 6.28. The individual metrics are: luminance, with cost function l and weighting α; contrast, with cost function c and weighting β; and structure, with cost function s and weighting γ.

$$\text{SSI}(x, y) = [l(x, y)]^{\alpha}[c(x, y)]^{\beta}[s(x, y)]^{\gamma} \tag{6.28}$$

Luminance is simply the mean pixel value. Contrast is essentially the standard de-

viation of the pixel values, and structure is a pixel-wise comparison of the images, normalised for luminance and contrast. Additionally, these metrics are calculated regionally rather than globally. This is typically done by defining a window size and then calculating the SSI metric for this window convolved over each pixel. This allows the SSI to assess local similarity and addresses one of the key issues highlighted in the loss section: nearby pixels tend to be correlated, so treating images purely pixel-wise is sub-optimal.

6.7 Simulated Annealing

For some of the structure types investigated in this thesis, it is necessary to simulate atomic positions within a structure framework. A method for approaching optimisation problems like this, with such a large parameter space, is to use Simulated Annealing[129]. Simulated annealing is a method for approximating global optimisation but crucially does not guarantee finding an optimal solution. As the name suggests, the algorithm attempts to replicate the behaviour of metals annealing.

The specific implementation of simulated annealing used in this work will be explored in detail later, but we will cover the basics here for completeness.

Simulated annealing incorporates three main steps: identifying a change; evaluating the change as an 'energy' function; determining whether or not to apply the change. The algorithm aims to replicate the annealing process, which occurs when hot metals are slowly cooled. In such a situation, metallic atoms have kinetic energy proportional to temperature and distributed according to the Boltzmann distribution. Assuming the annealing occurs in a thermodynamic-control domain, a changes to the structure occur whenever the internal energy (according to the Boltzmann distribution) is greater than the energy difference between the

two states.

For implementing the approach, the most common approach for the first step, identifying a change, is to use a simple Monte Carlo approach. That is to say, select a change at random, from a corpus of possible changes. The second step is to define the energy function and evaluate the effect of the change on this energy function. Finally, all that is required is to determine whether the change should be retained. This is done by finding the energy difference; if this is a negative difference, i.e., the final state is lower energy than the initial, the change is always accepted.

If the difference is positive, the energy is used along with the Boltzmann distribution to calculate the ratio of the probabilities between the system having enough internal energy to be in the new state, i, compared to the probability of the system having enough internal energy to be in the old state, j, using Equation 6.29.

$$\frac{p_i}{p_j} = \exp\left(\frac{-\Delta E}{kT}\right) \tag{6.29}$$

Where k is a constant with suitable dimensions, which can be set to 1 for ease of calculation. When the energy change is positive, this fraction $\frac{p_i}{p_j}$ takes a maximum value of 1 and a minimum value of 0. As such, if a random sample is taken between 0 and 1 this random value can be compared to $\frac{p_i}{p_j}$ to determine if the change occurs. If the random sample is smaller than the Boltzmann ratio, the change is accepted, otherwise the state remains unchanged.

This process is then repeated, and over the course of the simulation, the temperature is gradually reduced. This means that subsequent swaps are less likely to overcome energy barriers.

This regime allows the system to make swaps that generally decrease the overall

energy function while also, initially, making exploratory changes that will result in increases in the overall energy function. This can be effective in allowing the routine to escape from local minima.

As a result of the temperature dependence of 6.29, the scheme for cooling plays a key role in the quality of the optimisation, as if there is not enough time to explore at the start, it may find itself in a local minima. In contrast, if there is not enough steps performed at lower temperatures, the system may not reach the minima, despite it being theoretically accessible.

7 | Methods and Instruments

In this chapter we will cover the methods and procedures required for conducting the experiments outlined in this thesis, including the facilities, instruments, and equipment used throughout.

7.1 Equipment and Instruments

7.1.1 Synthesis Lab Equipment

7.1.1.1 Glove Box

The glove box used was a Unilab mBraun glove box with $O_2 < 0.1\,\mathrm{ppm}$ and $H_2O < 0.1\,\mathrm{ppm}$.

7.1.1.2 Weighing scales

The scales in the glove box used to weigh out the precursor materials was a Mettler Toledo XS64 balance. This balance has a precision of $\pm 0.0005\,\mathrm{pm}$.

7.1.1.3 Ball Mill

The ball mill used was a Retsch PM100 planetary ball mill. The ball mill vessel was a $250\,\mathrm{mL}$ stainless steel ball mill container. The bearings used were $3\,\mathrm{mm}$ diameter Tungsten Carbide ball bearings.

7.1.1.4 Vacuum Oven

The vacuum oven was a ThermoFisher Scientific Lab Line Vacuum Oven.

7.1.1.5 Pellet Press

The pellet press was a Specac $25\,\mathrm{t}$ Manual Hydraulic Press and the pellets were cast in a Specac $20\,\mathrm{mm}$ Evacuable Pellet Press Dye.

7.1.1.6 Tube Furnace

The tube furnaces used were Carbolite MTF 12/38/400 tube furnaces. The pellets were fired in Carbolite Gero RCA ceramic boats.

7.1.2 Sample Screening Lab Equipment and Instrument

7.1.2.1 Transmission Electron Microscope

The microscope used for screening samples was a JEOL JEM-2100 $200\,\mathrm{keV}$ TEM. The TEM has a LaB_6 electron source, a CCD for imaging and is also equipped with an Oxford Instruments XMAX 80T EDX detector.

7.1.2.2 Sample Holder

The samples were loaded into the TEM on a JEOL Double Tilt Beryllium Holder, with a tilt range of $\pm 30^\circ$ in both x and y.

7.1.3 ePSIC

7.1.3.1 E01 Electron Microscope

E01 is fitted with an aberration corrected JEOL ARM200CF electron microscope. This has two primary operating beam energies: $80\,\mathrm{keV}$ and $200\,\mathrm{keV}$. E01 has a sub-Å resolution imaging in scanning mode (STEM) and is equipped with dual JEOL Centurio EDX detectors. The dual EDX setup provides a total collection

solid angle of $\simeq 1.5\,\mathrm{sr}$. It is also fitted with a Gatan Quantum GIF for electron energy loss spectroscopy (EELS).

7.1.3.2 E02 Electron Microscope

E01 is fitted with an aberration corrected JEOL ARM300CF electron microscope. This has multiple operating beam energies ranging from $30\,\mathrm{keV}$ and $300\,\mathrm{keV}$. E02 has sub-Å resolution imaging in scanning and transmission mode. It is equipped with a Quantum Detectors Merlin Medipix Direct Electron Detector and a Gatan OneView Camera. It is also equipped with a single Oxford Instruments XMAX 100 EDX detector.

7.1.3.3 Standard Sample Holder

The samples were loaded aerobically into the microscope on a JEOL Double Tilt Beryllium Holder, with a tilt range of $\pm 15^{\circ}$ in both x and y.

7.1.3.4 Vacuum Transfer Holder

The samples were loaded anaerobically into the microscope on a JEOL double tilt vacuum transfer holder, with a tilt range of $\pm 15^{\circ}$ in both x and y.

7.1.4 X-ray Diffractometer

7.1.4.1 I11

Powder x-ray diffraction patterns were collected on the I11 beamline at Diamond Light Source[130]. The patterns were collected using x-rays with wavelength $\simeq 0.825\,\text{Å}$. The samples were collected in capillaries and loaded onto the diffractometer by robotic arm. The patterns were collected via the use of 5 multi-analysing crystal (MAC) devices. These MAC devices are located around the

2θ-circle containing a total of 45 individual Si crystals and photo-multiplier based detectors with fast pulse shaping electronics.

7.1.5 Neutron Diffractometer

7.1.5.1 HRPD

Powder neutron diffraction patterns were collected on the HRPD beamline at the ISIS Neutron and Muon source. Detection is done using time of flight so a range of wavelengths are used between 0.5 Å and 12 Å. The patterns are collected across three detector banks. Bank 1 collects backscattered neutrons with 2θ between $158.46°$ and $176.11°$. Bank 2 collects perpendicularly scattered neutrons with 2θ between $80.41°$ and $99.59°$. Bank 3 collects low angle scattering with 2θ between $27.47°$ and $32.32°$. Patterns were collected with a 30-$130\,\mathrm{ms}$ time-of-flight window giving d-spacing ranges of: 0.65-2.60 Å for bank 1; 0.85-3.90 Å for bank 2; and 2.3-10.2 Å for bank 3.

7.2 Synthesis

The cathode samples were prepared using two standard synthetic routes. The different approaches were simply a result of when they were made and so whether a vacuum oven was available in the synthesis lab.

7.2.1 Standard Cathode Synthesis

1. The precursor materials are weighed out in weighing boats within the glove box.
2. These are then hand ground together using pestle and mortar for several minutes until homogeneous to the eye.

3. The mixture is then added to a ball mill vessel containing a 15:1 ratio of ball bearings to precursor material.
4. Acetone is then added to the mixture to create a slurry. The mass of acetone is twice the mass of the dry mixture.
5. The ball mill vessel is then clamped and sealed, and removed from the glove box.
6. The slurry is ground on the ball mill for 1 hour at 150 rpm. The direction of the mill is reversed every 15 minutes and there is a 2 minute rest before restarting.
7. The slurry is transfer to a vacuum oven set to $60\,^{\circ}\mathrm{C}$. It is evacuated and stirred every 5 minutes until the acetone appears fully evaporated. (Before the vacuum oven was available this step was: Place the slurry in a fume hood until the acetone has evaporated.)
8. The dry mixture is then pressed into pellets of approximately $2.5\,\mathrm{g}$ using the pellet press and an $18\,\mathrm{t}$ load for $30\,\mathrm{s}$
9. The pellets are transferred, under air, to the tube furnace where they are fired at the target temperature for 8 hours. The furnace is heated and cooled at a rate of $5\,^{\circ}\mathrm{C}/\mathrm{min}$.
10. The synthesised pellet is then broken and ground within the glove box using a mortar and pestle
11. The final powder is then stored in the glove box under future use.

7.2.2 Precursor Chemicals

Table 7.1: Information on starting materials used in solid-state synthesis.

Starting Material	Supplier	Purity / %
Na_2CO_3	Honeywell Fluka	99.9
$NiCO_3$	Alfa Aesar	98
MnO_2	Alfa Aesar	99.9 (metals basis))

7.2.3 Sample Synthesis

The majority of this study focuses on three stoichiometries of P2 samples: N25, N30, and N33 (so named for their target Ni composition: 0.25, 0.30, and 0.333). The mass of precursor used for each sample is given in Table 7.2.

Table 7.2: N25, N30, and N33 sample preparation

Sample	Target Stoichiometry	Mass of Reactant / g		
		Na_2CO_3	$NiCO_3$	MnO_2
N25	$Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$	1.6086	1.4528	3.1922
N30	$Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$	1.6852	1.7262	2.9500
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	1.7065	1.9112	2.7995

7.3 Sample Preparation

7.3.1 Powder Diffraction Capillary Loading

The powder x-ray diffractometers require the samples to be loaded into capillary tubes for data acquisition. The powder neutron experiment require the samples be loaded into $6\,\mathrm{mm}$ vanadium cans. The procedure for preparing a capillary is more involved and the procedure is outlined here.

1. Inside an argon glove box add desired sample to a pestle and mortar.
2. Hand grind the sample for several minutes to minimise particle size.
3. Scoop a small amount of powder into the large end of a $0.7\,\mathrm{mm}$ borosilicate capillary.
4. Using a spatula gently vibrate the tube causing the powder to fill the capillary.
5. Repeat until approximately $3\,\mathrm{cm}$ of capillary has been filled taking care to ensure it is well packed, with minimal gaps.
6. Seal the large end of the capillary with vacuum grease and remove from the glove box.
7. Apply heat to the capillary using a blow torch to melt the glass above where the sample has been filled to, permenantly sealing the capillary
8. Mount the flame-sealed capillary in $0.9\,\mathrm{mm}$ brass holders, fixing with glue.

The important considerations in the capillary loading procedure are to ensure the sample is well-packed, ensure the seal is done correctly and ensuring the capillary is mounted perpendicular to the base of the holder. The sample must be well packed to ensure there are not gaps in the sample. As the beam falls on a narrow window of the sample, any gaps might lead to poor signal acquisition. For all samples, but these air sensitive cathode materials especially, it is important to ensure the seal is formed correctly otherwise the sample might become exposed to air, altering its chemical structure. Finally, the sample must be loaded perpendicularly to the base. This is because the sample is spun in the beam and if the sample is not loaded straight this will cause artefacts in the pattern. Specifically it will cause the sample to rotate out of the path of the beam, reducing incident and therefore scattered intensity. Further, it will cause the sample to be viewed at two

different path lengths, modifying the Bragg condition slightly, which can lead to splitting of peaks.

7.3.2 TEM sample loading

Both ePSIC microscopes as well the screening TEM requires samples be loaded on to TEM grids. Standard procedure for loading cathode samples onto a grid was to dry disperse the powder, this was done as follows.

1. Inside an argon glove box add desired sample to a pestle and mortar.
2. Hand grind the sample for several minutes to minimise particle size.
3. Add small amount of ground sample to a weighing boat.
4. Tip a new TEM grid out of the holder and into the weighing boat.
5. Gently tap the weighing boat to move the TEM grid through the powder.
6. Repeat for several minutes.
7. Using reverse tweezers pick up the TEM grid at the edge and return to the holder.

The important considerations on TEM grid loading are that the powder is suitably ground, adequately and gently dispersed, and that the grid is return to the holder carefully. Many of the diffraction experiments require samples to be thin in order to collect high quality data and if the sample is not well ground it will be harder to find usable particles. Additionally, samples will be ground before preparation in a functioning cell, so grinding the sample before loading the grid makes the sample more representative of the morphology of the sample in its use case environment.

The purpose of ensuring the sample is adequately and gently dispersed is again, to maximise the opportunity to find good quality particles. If the grid is not mixed

with the powder for long enough, there will be fewer particles on the grid and so fewer particles to pick from. Also, if the grid is not handled gently at this stage it is possible to break the thin carbon film that the particles stick to, losing the particles in the process. Finally, if the grid handling is too rough, it is possible to bend the grids. Slight bending of grids can slow down the data collection, as sample height is important for focal lengths on the microscope and bent grids with exhibit a more broad range of heights. If the grid is bent too much it can also prevent the grid from being able to load into the sample holder, especially when using an anaerobic transfer holder, which has more hardware involved in fixing the grid.

7.3.3 Experimental Samples

Throughout the later sections of the thesis, there is reference made to TW and DK samples. These refer to samples, studied by me, but synthesised by Dr Thomas Wood and Daniel Kwok respectively. Synthesis was still carried out in the manner described in 7.2.1. A breakdown of these samples, with details of the experimental setup, is given in Table 7.3.

Table 7.3: Details of experimentally investigated samples, including: studied compositions, diffractometer used, detector used, scan timings and approximate sample weights.

Samples	Compositions	Diffractometer	Detector	Scan Times	≈ Mass / g
TW-XRD	N25, N30, N33	I11	PSD	3 x 10 s	0.025
TW-PND	N25, N30, N33	HRPD	90° Bank	3 hour	1.5
DK-XRD	N33	I11	MAC	2 x 15 min	0.025

7.4 Sample Pre-screening

As eluded to in the previous section, a challenge with using electron microscopy as a structure investigation technique is that there are certain requirements of the selected particles to ensure quality data is collected. The most obvious one of these is specimen thickness. As discussed in the Structural Techniques chapter, thin samples better fit the kinematic model of diffraction and can provide better quantification of diffraction intensities. Furthermore, thin samples are more likely to give homogeneous structural character along the beam direction, which will also yield more interpretable diffraction signal. As well as the requirement of thin particles there are also advantages to identifying particles which are adhered to the grid in a specific orientation. This is advantageous for two reasons. The first is time, it is much quicker to collect data on multiple particles if large changes in stage tilts are not required. In scenarios where large tilts are needed, there is the additional requirements of resetting the eucentric at each location. The second is spectroscopy. As the EDX detector in all the microscopes used are set to one side, there is usually a preferred orientation that maximises the counts detected. More importantly, there are also orientations where the detector can be partially or completely obscured. As such, it is much more reliable for collecting spectroscopic data, to have particles approximately in the desired orientation without having to tilt the stage much.

The optimal orientation of a particle depends upon what aspect of the structure is being investigated. There were two primary desired orientations for this work: perpendicular and parallel to the c-axis. Parallel to the c-axis is desirable to image the in plane ordering of the sodium ions and allowed for the study of superstructure reflections later in Chapters 9 and 10. Perpendicular to the c-axis was also studied to try and gain an improved understanding of the layer by layer vari-

ation of the structures, specifically for Mn oxidation state in Chapter 11.

The nature of the layered cathode materials is that they appear to preferentially orientate with the c-axis perpendicular to the grid (parallel to the beam direction for a neutral stage). This is due to the Na−O bonds being much weaker than M−O bonds allowing for preferential sheering between the layers. Thus, finding particles orientated with the c-axis parallel to the grid (perpendicular to the beam direction for a neutral stage) was less common.

As access to the high-quality aberration corrected microscopes was imperative to the quality of the structural investigation, but also limited, these specimen restrictions could be quite problematic. The most straight forward solution was to pre-screen the samples on a lower quality, but more readily available TEM: to identify particles ahead of time which were suitable for further investigation on the E01 or E02 aberration corrected microscopes. To do this would require a way of re-identifying particles and locations on the same grid.

The ability to re-find the same particles was also key for using a multi-modal approach to study structure, as it allowed for the same particles to be investigated on E01 and E02, leveraging the different techniques to give a more full picture of a single specimen.

To pre-screen and re-find these particles TEM finder grids were used along with annotated images taken at multiple magnification to document the particle and its neighboring environment. This is an especially helpful procedure to carry out on the uncorrected TEM as a larger field of view can be achieved without compromising beam alignment. An example composite image of a full finder grid is shown in Figure 7.1

A cartoon representation of how a particle location would be documented is shown in Figure 7.2. The particle in the top left square would be referenced as Od2-3,

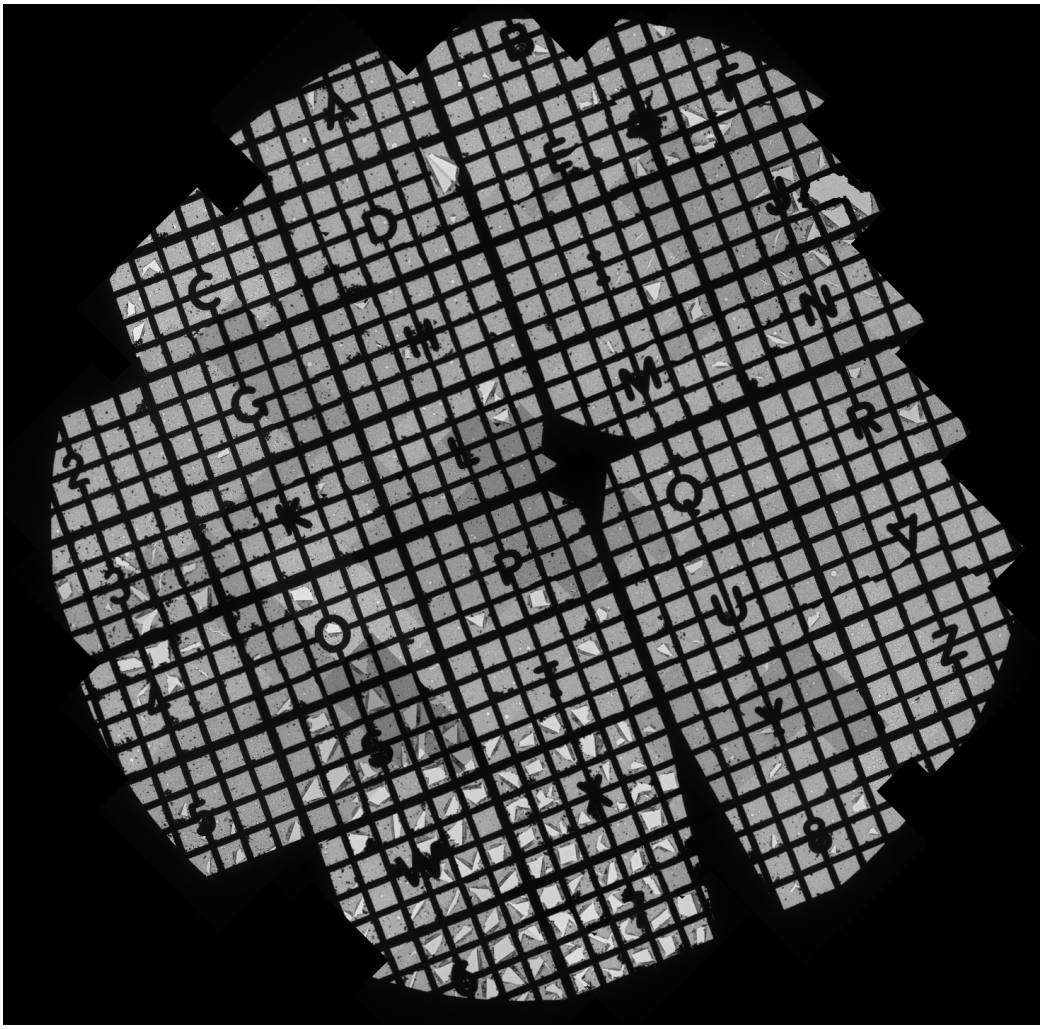


Figure 7.1: Image of an entire finder grid, showing the letters and numbers that can be used as identifying features.

and the second square would be used to re-identify Od2-1, Od2-2, Od2-3 and Od2-4. Comparing Figure 7.2 to 7.1 it can be seen how each grid square can be identified using the 'Marker, letter, number' notation and then all that is required is the annotated image of the particles within the grid bar. Multiple magnifications of these images were often helpful as the maximum field of view on the aberration-corrected microscopes is quite limited, so even identifying the large marker features could prove challenging.

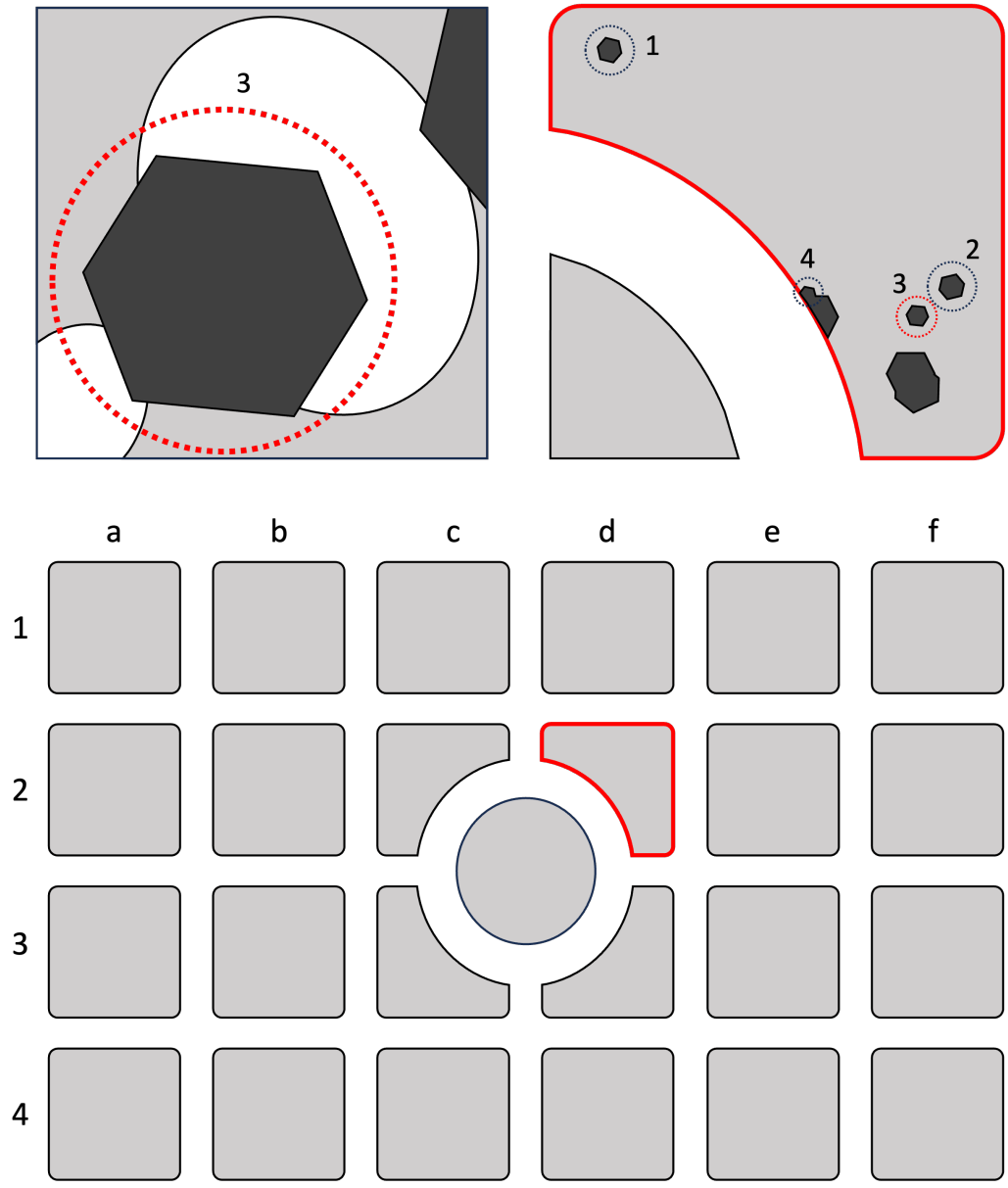


Figure 7.2: Cartoon representation of a particle that would be referenced as Od2-3 and some example information that would be required to identify it. The top two cartoon images show various annotated magnifications that would be necessary for re-finding a particle. The bottom gives a schematic of the grid layout and shows how the labelling system works.

8 | A Versatile, Automated Workflow for SEND data

As eluded to at the end of the introduction chapter (Section 3.4.2), a goal of this thesis is to drive an improved understanding of the physical properties of cathode materials by improving the fidelity with which the materials can be characterised, modelled and understood, across length scales, from nanometres to centimetres. As Scanning Electron Nanobeam Diffraction (SEND) can offer a fast and convenient approach for collecting structural information on both the nanometer and micron length scale[131], this chapter will outline utilising SEND data in a away that can help bridge the gap between the length scales of other techniques.

While a wealth of insight into the bulk structure can be obtained through techniques such as powder X-ray diffraction (PXRD)[132] and powder neutron diffraction (PND)[133], and into the immediate local structure using techniques such as High Resolution Scanning Transmission Electron Microscopy (HR-STEM)[134], it can prove challenging to unify the insight of the two approaches in a coherent manner, due to the vast differences in scale. SEND, as a technique, is becoming increasingly accessible alongside advances in fast electron detectors[135] and the technique lends itself well to microscope automation. This means it can cover areas in the micrometres length scale and, while still capturing microstructural detail not accessible via PXRD, provide statistically significant insight into the crystallographic defects and morphology of the phases present in the sample [136].

Utilising the 4D Scanning Transmission Electron Microscopy (4D-STEM) data collection approach, SEND combines this with a minimised probe convergence semi-angle to obtain distinct Bragg peaks within the diffraction patterns as discussed in Chapter 5. The advantages that SEND technique offers make it a ver-

satile approach for studying a wide variety of materials, including the cathode materials that are studied within this thesis.

The primary advantages of SEND technique are: relaxation of sample thickness requirements (relative to HR-STEM) resulting in data collection from much wider areas of sample, which in turn offers better statistics of the microstructure; and the ability to operate the microscope in a stable configuration, without requiring the frequent aberration tuning common in HR-STEM measurements, which permits for longer and more autonomous data collection.

Given the relatively large probe size, resulting from the optical setup in SEND, data collection is normally done at low to medium magnification range. This leads to a reduced electron dose subtended on the sample, making this technique particularly suitable for the study of beam sensitive systems like the layered metal oxides of cathode materials. Due to these factors, SEND data acquisition is well suited to the task of domain mapping, which is typically achieved through crystallographic orientation mapping[137, 138, 139, 140, 141].

Domain mapping, broadly speaking, is the task of identifying regions which share distinguishing properties and grouping them together. Producing these domain maps for SEND data allows for aggregation of all the diffraction information available within the entire domain into a single, cluster-averaged diffraction pattern. Working with these cluster-averaged patterns rather than the individual patterns is advantageous for studying more subtle features like superstructure and can aid pattern indexing, as the cluster-averaged patterns typically exhibit additional resolvable Bragg peaks, and improved peak definition. If the atomic structure present within the mapped domains can be identified (through direct treatment of diffraction data or combination with complementary techniques), then the statistical nature of SEND technique could provide the desired link between local

structure and the experimental results of the bulk techniques[142].

8.1 Developing a SEND workflow for domain mapping

For a given dataset, no single, ground-truth domain map exists which is inherent to that sample. Experimentally, there will always be multiple sources of variation between patterns, due to the statistical nature of the sampling. As the domain map seeks to group regions of shared properties, the returned domain map for a sample must depend on what property similarities are relevant to the specific experiment. For a given sample, it is therefore plausible to seek different domain maps, depending on what specific factors are of interest. As an illustrative example, consider a SEND dataset taken from a thin foil sample with small precipitates dispersed within it. It is plausible that one might only be interested in the locations on the foil where these precipitates have aggregated and the structural and orientation factors present in the matrix that drives precipitation in these locations. Alternatively, for the exact same dataset, one might instead be interested in the precipitate structure, and their orientation relationship with the matrix, in which case, a domain map which identifies differences in bulk/matrix lattice mean crystal structure and relative orientation would be desired. Thus, the goal of the experiment will always enforce bias onto which factors of variance are relevant and this should be considered when separating domains.

Research into automated crystallographic orientation mapping (ACOM)[143, 144] has employed SEND data collection to tackle the task of domain mapping. These techniques rely on prior knowledge of possible underlying crystal structures in order to classify the nature of individual diffraction patterns and then aggregate

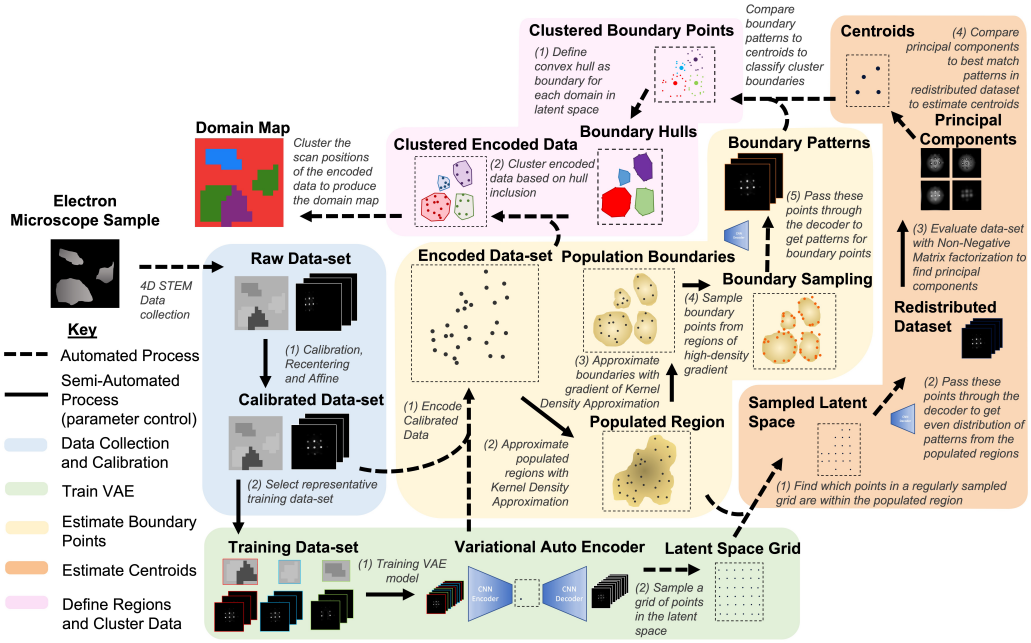


Figure 8.1: Overview of proposed workflow. The motivation of each step can be most straightforwardly understood by working backwards from the final domain map.

these into maps. For ACOM, a library of possible zone-axis orientations are generated and the individual patterns are correlated to determine the most plausible match. This approach works well when there is prior knowledge of models which can replicate all the desired variations within the experimental diffraction data collected. Unfortunately, this is frequently not the case. In many situations the simulations obtained from the models can differ from the experiment due to factors which are absent, overlooked or incorrect. Example factors such as: nuanced superstructure[145, 146]; overlapping lattice motifs[147]; non-systematic dynamical effects; or simply a lack of structural insight. Non-systematic dynamical effects usually cannot be fitted and so attempts must be made to mitigate them experimentally[148]. These simulation approaches are also unable to distinguish amorphous regions within the sample from vacuum or thick sample.

This chapter details the development and implementation of a structurally agnostic, data-driven workflow for domain mapping, which requires no prior knowledge of possible crystal structures and is equally as effective for identifying amorphous regions with little to no crystalline diffraction character. While the technique was developed with the motivating intention to apply it to data collected on the cathode materials that are the primary focus of this thesis, the generalisability of the approach makes it applicable to a much broader range of systems. To that extent, the use of the workflow as a key component in broader studies of these cathode materials, is reserved for the later chapters. Within this chapter, the direct application to those systems is somewhat diluted in service of highlighting the versatility of the technique. Additionally, techniques based on feature extraction with non-negative matrix factorisation (NMF)[149] and hierarchical k-means clustering to group similar diffraction signal[150] have recently been reported and are offered for cursory comparison.

The workflow, discussed within, leverages a Variational AutoEncoder (VAE) to

identify the variance and similarity within the diffraction data. For this workflow, the VAE is used in tandem with NMF to produce domain maps. For this application, the VAE simultaneously improves the signal-to-noise ratio and uses similarities between 2D diffraction patterns to optimally cluster regions in the low dimensional manifold of the VAE latent-space. This approach means that the factors determining the segregation are tangible and interpretable, allowing for simple alignment of the bias in the output domain map with the experimental goals. Additionally, this approach documents the production of a confidence map of the domain assignments, which means that the experimentalist can manually investigate regions of ambiguity and, if desired, update the mapping process.

This approach combines the convenience and rigour of automated mapping, with the flexibility of human-in-the-loop steering of the desired outcomes. This section begins by outlining the workflow as shown in Figure 8.1. The following sections then demonstrates the procedure, first on a synthetic dataset, which allows for quantifiable comparison with ground truth, then on a diverse set of real experimental datasets. It is shown that this method can produce meaningful domain maps in different scenarios where different applications are required. The code required to run this method was made available in an open repository[151] with the intention that the flexibility of this approach will make it a valuable tool for the application of SEND to gain statistically relevant insights into composition-structure-property relationships in materials.

8.2 Workflow

The input SEND dataset is in the form of a 4-D data array (x,y | u,v): two dimensions corresponding to the x,y coordinates of the probe position in the real space plane; and two dimensions corresponding to u,v coordinate of the pixelated

detector recording diffraction data intensities in reciprocal space. This dataset is then pre-processed and used to train a VAE. The VAE is constructed in two parts, encoder and decoder, which are trained in tandem to produce the identity transformation, passing through an n-dimensional latent-space (where n $<<$ number of detector pixels). Thus the VAE learns to produce a compressed representation of the data. Once trained, the two halves can be used independently, to convert diffraction signal into low dimensional coordinates (encoder) or low dimensional coordinates into diffraction signal (decoder).

As standard, this workflow uses two dimensions for its low dimensional coordinates. The properties of both the encoder and decoder are leveraged by this workflow, but in the first case the encoder is used to convert the pre-processed patterns of the input data into a collection of points within a 2-D plane (green section to yellow section in Figure 8.1). As a result of the constraints of the tandem training of the components of the VAE, the 2D representation outputted from the encoder will group patterns of similar diffraction character proximally within this latent-space but a given level of closeness between two points does not guarantee similarity. This makes directly clustering these points based on 2D coordinate alone challenging.

This occurs as a result of the metrics which are minimised upon training. There are two distinct factors at play, the reconstruction loss and the KL-divergence. The reconstruction loss is attempting to minimise the difference between the input patterns and the output patterns. The KL-divergence is minimising the difference between the distributions being learned and the Normalised Gaussian (the distributions being learned are the ones from which the latent variable are being sampled to pass into the decoder).

This means the contribution of the KL-divergence is pulling the training processes

towards a situation where every input pattern is encoded to the same distribution, one centred on $(0, 0)$ with a standard deviations $(1, 1)$. Such a situation cannot achieve non-trivial reconstruction losses however, as there is no way to discriminate what output pattern should be produced if all encodings are taken from the same distribution.

Conversely with no KL-divergence term to contribute to the loss, the model can learn an essentially infinite number of trivial encodings. Even considered generally, the input space is a finite set, due to there being a fixed number of pixels in the input and each pixel having a finite level of precision (detectors collect a 16-bit integer number). Considering the specific case, where the model is only concerned with a specific dataset, the possible inputs are even more limited. With no KL divergence, the model can collapse the standard deviations of the sampling distributions towards 0 and the encoder can directly pass exact latent encodings to the decoder. In this case the only factor enforcing similarity between reconstructions and their latent encodings is the depth of the decoder network (a shallower network has fewer applications of non-linear transformations and so is more limited in how different it can make two inputs with very similar values).

Thus both factors are required for the model to learn non-trivial solutions. The balance between these two factors, essentially drives the model to find a solution where not only are latent encodings for similar inputs nearby (as these can have larger standard deviations to reduce the KL-divergence, without strongly effecting the reconstruction loss), but all encoding are as close to each other and $(0, 0)$ as possible.

The way the model optimises the encoding positions is by updating the parameters in the encoder, effectively altering the projection of the many-dimensional dataset down into the latent space. As the position of the encoding depends both on the

distribution in the high dimensional space and the projection function, by updating the encoder parameters and changing the projection function, the fixed-shaped distribution in the high-dimensional space can form different shaped regions in the latent space. Depending on the shape of the distribution in the projection, the overall loss can be optimised in terms of the competing influence of the reconstruction loss and KL-divergence.

This means that, for our datasets, while similar patterns will be encoded near to each other, there is not a 1:1 relationship between encoded distribution and similarity. This can leads to ambiguity in identifying the boundaries between regions from point locations alone. Further, as the encoding is a function of optimising the projection of unknown but fixed shape, high dimensional distributions, with respect to the competing factors of reconstruction loss and KL-divergence, the shapes of the distributions in the latent space can prove tricky for standard 2D clustering approaches (such as Gaussian Mixture Models) to handle.

These properties may rightly lead some to question the use the VAE at all? The simple answer to this, is that while there may be a better approach yet to be found, this workflow here is able to produce meaningful domain maps despite the challenges imposed by the VAE, and work better than other approaches attempted. It seems that the primary advantages offered by including a VAE are that they are inherently robust to random noise (as this cannot be encoded non-trivially) and can be optimised to align it's encoding of domain similarity with the goals of the experiment.

In order to leverage the advantages of the VAE, while also effectively and efficiently cluster patterns into separate domains, the workflow uses a four step process, which is expanded upon later: (i) estimate where the boundaries of domain regions could be in the latent-space; (ii) estimate where a representative pattern of

each domain is located; (iii) use the Structural Similarity Index (SSI)[152][153] of representative patterns and decoded boundary patterns to define the boundary of each domain's region in the latent-space; and (iv) use the location of encoded data relative to these boundaries to cluster our dataset. Using this workflow, our sample data can quickly be assigned to clusters and these labels can be used to produce domain maps of the sample. We now explore each of these steps in more detail.

8.2.1 The Variational Autoencoder

A Variational Autoencoder is used to embed the patterns in a latent space. The AutoEncoder was first developed in the late 1980s[116, 117, 118] and was later extended to the VAE by replacing each single latent-space coordinate with a pair of values which define the statistical distribution from which that coordinate could be sampled [119]. The function of the VAE within our workflow is to act as a flexible approach for identifying the most important sources of variation within the diffraction dataset and embedding these patterns conveniently into a low dimensional representation.

The VAE was built using Tensorflow [78] and the model architecture is summarised in Figure 8.2.

Upon training, the VAE has no inclination as to the physical phenomena underlying the observed data nor an appreciation of what aspects of the dataset one might be interested in. As the VAE trains it seeks to minimise the loss function as discussed above. Throughout the training process, due to the gradient descent optimisation, the performance is biased towards the separation of the largest sources of variance within the training dataset first and then refining the model for smaller and smaller sources of variance. Both these biases define how the model will sep-

arate out the input dataset and essentially defines the model's understanding of similarity.

In order to use the VAE most effectively, it is therefore necessary to manipulate these biases to align the VAE's interpretation of similarity with the definition of similarity imposed by the focus of the particular experiment. This is most effectively achieved by modifying the training set to contain the representative proportions of the relevant sources of variance and truncating training once all the desired variance has been identified. This truncation is achieved by saving iterations along the training process and evaluating the performance at different stages. This evaluation can be done by visually comparing the patterns inputted and outputted by the model and also by comparing decoded points from the different regions in the latent-space to input the patterns that have been encoded into those regions. If the bias is controlled effectively, the VAE can be used in vastly disparate application cases to give a primary identification of which patterns fall within which domain.

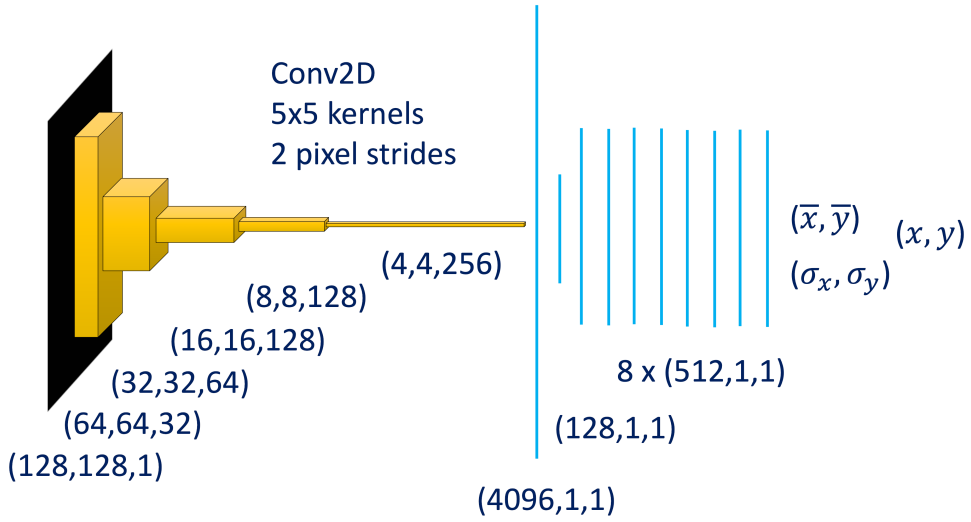


Figure 8.2: Architecture of Encoder half of VAE, the decoder is the same model structure in reverse.

8.2.2 Latent Clustering

Clustering of the patterns into single domains is achieved using the 2D encodings in tandem with the decoder and the Structural Similarity Index (SSI)[152][153]. The aim is to find and define the regions within the 2D latent-space, that correspond to each domain. To achieve this, the workflow seeks to identify the locations of the all boundaries between these domain regions, and also identify a typical pattern for each domain (yellow and orange sections of Figure 8.1, respectively). Then, by passing points along the boundaries through the decoder to generate diffraction patterns, the workflow uses the SSI image similarity metric between these decoded patterns and the domain typical patterns, to identify which boundaries points correspond to each domain.

8.2.3 Boundary Estimation

Boundary estimation is the first requirement for the clustering approach, which aims to identify candidate points that could define the edge of each domain. It is assumed, due to the bias of the VAE, that the encoded data will be grouped such that similar patterns are nearby and so each domain will occupy a continuous, definable region, within the latent-space. The challenge, is that the same features of the VAE which allows for each domain to occupy a continuous, smoothly varying region within the latent space, also leads to the boundaries between these regions becoming more smoothly varying. This makes it difficult to identify the locations of the boundaries between domains, simply by evaluating changes in the decoded patterns as one moves through latent space. This is because there tend not to be discrete boundaries between the patterns of different domains. While these smoother inter-domain transition regions can occur within the latent space, the patterns associated with these regions are often not of physical relevance and

are not observed within the actual data-set being processed. As a result of this, the locations of the encoded points of the actual data-set patterns can help identify where relevant domain boundaries occur, as there should be a higher density of encoded points in the regions of the latent space associated with the domains of interest and lower density in the inter-domain boundaries.

The final challenge is that the quantifiable "higher density of points" within domains can vary from domain to domain, across the latent-space, depending on the size of the region and the number of patterns of that domain present in the sample data-set. As such the ultimate approach attempts to leverage both changes in patterns across the latent space and point density within the latent space to identify the boundaries of the domain regions. A density-based evaluation is used to identify possible locations of domain boundaries using the gradient of the density of encoded points. It is assumed that areas where the density of points changes quickly (large gradient) will likely be these inter-domain regions, but this will also identify some erroneous areas as a result of the varied point density discussed earlier (the later centroid treatment is designed to minimise these issues caused by these erroneous points). Specifically, the density-based evaluation is achieved by using a Kernel Density Estimation (KDE). The KDE is performed using built-in functions in sklearn[76]. The KDE is used to approximate the density function for the entire latent-space. The first derivative of this function is then computed and gradient values within a set range are used to define a uniform probability distribution function. The potential boundary points are then sampled randomly from this probability distribution function. An example result of this is shown in Figure 8.3

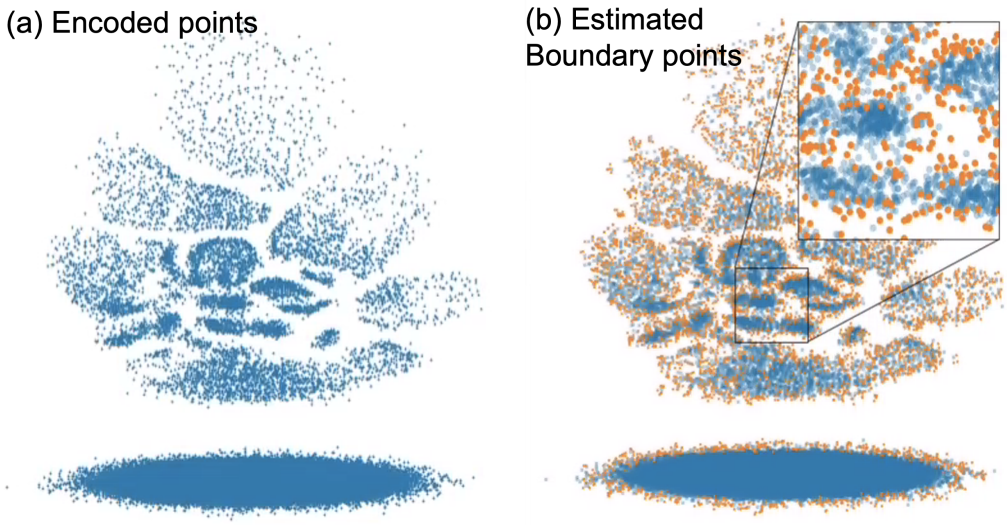


Figure 8.3: (a) 96,000 diffraction patterns encoded into 2D space. (b) the same encoded patterns in blue, with 5,000 sampled boundary point candidates shown in orange. A sub-region is highlighted in the inset.

8.2.4 Centroid estimation

Centroid estimation was used to apply the sampled candidate boundary points in the definition of bounding areas for each domain region. The approach used is to identify a set of representative patterns (referred to as centroids), one for each domain, that can be compared to the decoded boundary patterns and used to assess which region the boundary point most closely resembles. The comparison metric that is used in this workflow is the SSI, and while other metrics are not implemented in the code, it would be straightforward to add and use any such metrics. The requirement of the comparison metric is that the decoded pattern of a general point sampled from within a given domain's boundary will give a stronger response when compared to that domain's centroid pattern than any of the other centroids. The SSI was chosen as this performance closely aligned with the observations upon inspection by a domain expert. Due to the continuous na-

ture of the latent-space, almost any point within the region should fulfil the role of a representative pattern. Thus the centroids can be estimated by a few approaches: using the derivatives of KDE to find stationary points in the density function which correspond to centres of boundary regions; being manually selected; or by using non-negative matrix factorisation (NMF) [154]. NMF is the most automated approach and is used as standard in this workflow. An overview of the mathematical formalisation of NMF is given in Chapter 6.

In order to find a set of patterns to be representative of the domains within the latent-space by NMF, the workflow aims to use a dataset of diffraction patterns that contains all the pattern variation which has been encoded within the latent-space. Then the NMF is applied to extract the different sources of variations in the patterns that are found within this dataset. Assuming that the patterns within each domain exhibit some characteristic of their pattern that distinguishes them from the other domains, then these characteristics are sources of variation within the dataset. As such, these characteristics will be identified by the NMF. As each characteristic will be shared by the patterns within this domain but not in the other domains it should fulfill the role of the centroid, as defined earlier.

The dataset used could just be the original data, but this will have performance issues due to the uneven sampling of domains. The original data will be larger than is required, due to oversampling points in the more populated domains, and so will take longer to perform the NMF. Further, as the less populated domains will be significantly undersampled, the source of pattern variation corresponding to these might prove harder to extract from the inherent noise within the more populated domains.

To generate a better performing dataset for centroid estimation, the workflow, evenly samples points across the populated regions of the latent-space, using the

already calculated density approximation to identify the populated regions. These points can be fed through the decoder to generate a collection of diffraction images representative of the pattern variance within the original dataset. This re-balanced dataset reduces the issues identified for the raw dataset. Firstly, it is a much smaller dataset while still containing all the variance the VAE has identified within training; this allows for speeds up in calculation of the NMF. It is also a significantly de-noised dataset, as it only contains the variances that have been identified as relevant during the VAE training. Removing the sources of variation not corresponding to changes in domain, will improve the applicability of these NMF components for use as centroids. Finally, the distribution of patterns within the dataset is less dependent on the occurrence frequency within the sample and will be proportional to the size of region in the latent-space. This increases the visibility of minor phases and in turn results in better the segmentations of less populated domains.

An NMF is performed on the re-balanced dataset. The workflow then compare these components to the patterns from the rebalanced dataset and replace the NMF component with the reconstructed pattern that most closely matches it, as measured by SSI. This is because the NMF component will contain only the distinguishing characteristic of that pattern, but ideally the centroid would be the entire diffraction pattern that is representative of that domain, not just part of it. This set of patterns will then correspond to the centroids of our domains. The total number of these centroids is therefore dependent on the number of components requested of the NMF. The fewer the components requested of the NMF, the larger the sources of variance in the dataset has to be in order to be classified as a component. To reduce the dependence of the clustering on this parameter, the workflow allows for comparisons and evaluation of domain maps using different level number of centroids. Additionally, once the centroid patterns have been identified by

NMF, they can be compared by SSI and patterns centroids which are sufficiently similar (by a user-defined threshold) can be merged (by averaging the point locations in the latent-space). By doing this, the number of domains is slightly more flexible and is driven by a maximum level of similarity between domains, rather than being directly determined by a chosen parameter.

8.2.5 Assigning data utilising convex hulls

Once the centroids and potential boundary points have been identified, the SSI can be used to assign each boundary point to the best matching centroid. Then, for each centroid, the classified boundary points can be used to define an outer limit of that domain region using a convex hull.

The Convex Hull is used as it allows for faster evaluation of points falling within the hull. A convex hull is applicable as despite not producing a perfect bounding hull, the inaccuracy should only produce false positive inclusions, which should then lead to multiple group inclusion conflicts which can be straightforwardly resolved. With the latent-space now defined by these hulls, encoded patterns can fall into one of three categories: not within any hull; within exactly one hull; within more than one hull. The vast majority of the points tend to fall within exactly one hull. The points shared between hulls are classified by direct SSI of the pattern with the conflicting hull centroids and the points not within any hulls are classified by direct comparison with all hull centroids.

8.2.6 Iterating

Once the points have all been classified into domains, an additional means of centroid approximation becomes trivial, taking the centre of mass (CoM) of the encoded points within each domain. This can offer a useful approach for refining

a centroid location and might improve the outcome of a successive iteration of the workflow. This can be beneficial if the region is approximately centrosymmetric and the CoM approximation moves the centroid more centrally within the region. As standard the workflow outputs the CoM of the identified region as well as the centroid used.

8.3 Example Datasets

In this next section, three case studies are presented to show the diverse use-cases for the workflow. These case studies are presented in order of increasing complexity: a dataset populated with simulated diffraction patterns for known, distinct orientations of a single crystal structure; an experimental SEND dataset from an Al-alloy thin foil with precipitates; and an experimental dataset of a P2 cathode material.

8.3.1 Simulated dataset

The simulated dataset is populated using 12 distinct zone axis orientations of a Layered Metal Oxide (LMO) structure and is shown in Figure 8.4(c). The regions 1–5 and 7–11 are each populated with the simulated pattern of a single zone axis orientation. The bar at the bottom containing labels 12, 13 and 14 is populated with the simulated pattern of a single zone axis, but the diffraction signal decreases in intensity from right to left, this is designed to emulate real world conditions where signal intensity might be reduced due to factors such as thickness changes or beam damage. Region 7 contains patterns for an amorphous region. The final region, 0, is populated with an unscattered electron probe to emulate the vacuum around particles of interest. The goal of the domain mapping for this dataset was to identify the areas of different zone axis orientation.

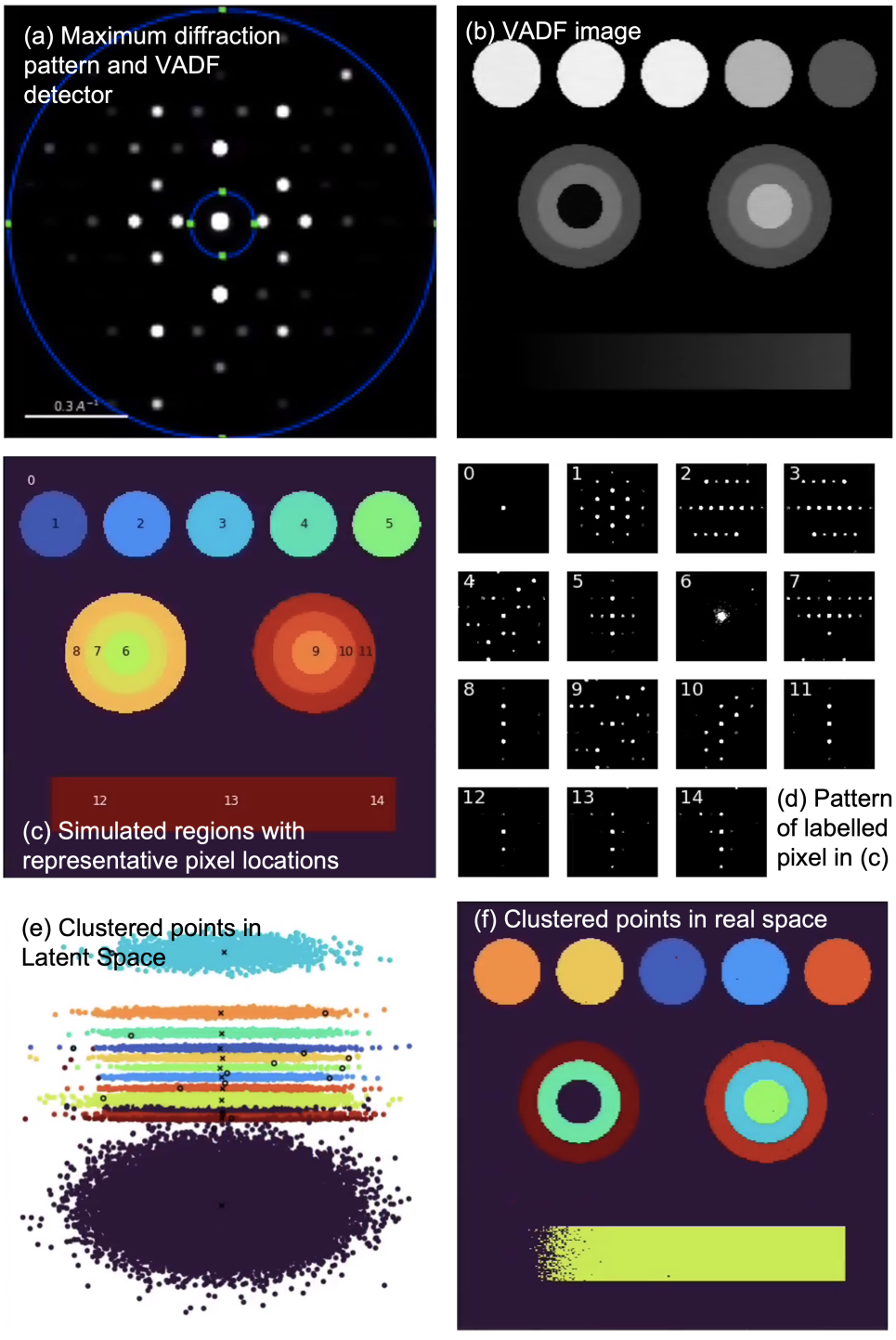

(a) Maximum diffraction pattern and VADF detector
0.3 A⁻¹
(b) VADF image
0
1
2
3
4
5
8 7 6
9 10 11
12
13
14
(c) Simulated regions with representative pixel locations
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
(d) Pattern of labelled pixel in (c)
(e) Clustered points in Latent Space
(f) Clustered points in real space

Figure 8.4

Figure 8.4: (a) the maximum diffraction pattern of the entire dataset with a virtual annular detector. (b) the Virtual Annular Dark Field (VADF) image for the simulated Layered Metal Oxide (LMO) sample dataset. (c) The region map used to populate the simulated dataset, each region contains patterns of the same crystal structure (or amorphous structure) and orientation but with added noise (and variation in thickness for the bar at the bottom). Pixel locations have been indicated with numbers to aid the understanding of the patterns which populate the dataset (d) the diffraction pattern corresponding to the pixel locations highlighted in (c). (e) the encoded positions of the patterns clustered by colour. The open black circles show the initial centroids identified by NMF and refined by SSI. The black crosses show the new centroids by averaging the locations of points within clusters. (f) the clustering on the real space pixel locations, a pixel location has been highlighted in each of the identified domains.

The dataset was generated with a 296×295 real space and 128×128 pixel array in each diffraction pattern. This pattern size is ideal for training the VAE and no pre-processing was required. The model was trained, from scratch, on all the shuffled input patterns with a random sample of 1000 patterns used to monitor performance. Training was truncated at the point where the model began identifying distinctions that were deemed no longer relevant to the aim of the segmentation. Once the best iteration of the model is selected the dataset is processed by the workflow and the results are shown in Figure 8.4(e) and (f).

Comparing the output to the ground truth the consistency of the clustering is above 98% for all regions excluding the intensity varying bar, which identified 86% of the total patterns. The drop in performance for the bar is a result of patterns with very low diffraction signal that closely resemble the vacuum. This resemblance grows increasingly along the bar such that at the far left the diffraction signal is far more akin to vacuum than the actual structure present, so this could be considered a valid interpretation of the domains.

8.3.2 Precipitates in Al alloy thin foil

Precipitation-hardening is an important mechanism for increasing the strength of commercial aluminium alloys. Gaining statistical insight into the population and crystallographic nature of the precipitates in these alloys is experimentally challenging and thus SEND is a suitable technique to tackle this problem[155]. The experimental dataset presented here is an example of such SEND data. This dataset is comprised of 511×511 probe positions in real space with 256×256 pixel array in diffraction patterns. The goal of the segmentation is to identify the diffraction character of the precipitates, which form a minor component of a dataset dominated by the diffraction of the Al substrate.

To optimise the training for this functionality the data is pre-processsed to crop to the central 128×128 of the diffraction pattern, where the majority of the desired information can be observed. The training data is then re-sampled using the Shannon entropy[156] of the processed patterns as an indicator of precipitate character. The Shannon entropy is a measure of information content in an image and is given by $-\sum_k p_k ln(p_k)$, where p_k is frequency/probability of pixels of value k. Shannon entropy is used as the precipitate patterns are expected to have a higher entropy due to the additional diffraction information and is based on a similar application of Shannon entropy by Slouf et. al.[157] The VAE was trained and the resulting latent-space clustered into two regions, which correspond to the precipitates and the background Al matrix.

Other methods for masking the precipitate region will yield similar results more straightforwardly, but this workflow was used to illustrate how VAE bias can be directed using iteration and how this might be applicable in less trivial cases where classical masking techniques aren't applicable. The mask was used to generate a new 4D dataset with the variance of the background Al substrate replaced with a

single mean pattern. This new dataset was then used to retrain the VAE and the model used to cluster the navigation region using 16 components in the NMF as indicated in Figure 8.5.

From the clustering in Figure 8.5(d), it can be observed that the boundaries of the classes do not precisely follow the gaps between the more dense point encodings (the yellow-green class, corresponding to Cluster 9 in the Figure, can be seen to span across various point densities). This is a similar case to that of the thickness effects for the simulated data, as again there is deviation from where one might anticipate boundaries, but again, this is the result of regions with insignificantly low diffraction intensity merging. Inspection of the resulting patterns in Figure 8.5(e) and (f), corroborates this observation. Separating out the low diffraction information regions from the strongly reflecting centres is an entirely valid domain distinction in this use case. This further highlights the utility of the proposed clustering approach, as solely interpreting the proximity of the latent coordinates is not entirely reliable for capturing the naunce of the dataset. This is because across the edge of the precipitate there are two factors of similarity being considered: how similar the precipitate reflections are; and how similar the matrix reflections are. The precipitate reflections are the factor that is being investigated here, however, in the matrix reflections there will be some fluctuations in intensity that will be correlated to position due to local factors such as matrix thickness.

The discrepancy seen between the latent encodings and the clustering are due to added weighting being given to the matrix reflections which are essentially unchanged moving from the exterior edge of the precipitate to the empty matrix right beside it, but are changing in some systematic way between the regions of the different precipitates. This added weighting manifests as the patterns corresponding to empty matrix besides a precipitate being encoded close to the edge of the given precipitate domain, and somewhat separated from the larger domain of

empty matrix patterns collected from elsewhere in the sample. Further improvement of the alignment between the latent encoding and the clustering could likely be achieved by masking out the matrix reflections in all patterns and retraining the VAE on this masked dataset.

From Figure 8.5, the utility of this workflow can be seen. The workflow has taken a sample of 261,121 diffraction patterns and condensed all the information relevant to the experiment into a domain map and 14 patterns. Additionally, in almost all the examples the cluster-averaged pattern for each domain contains better resolved and more usable diffraction information than in the individual patterns.

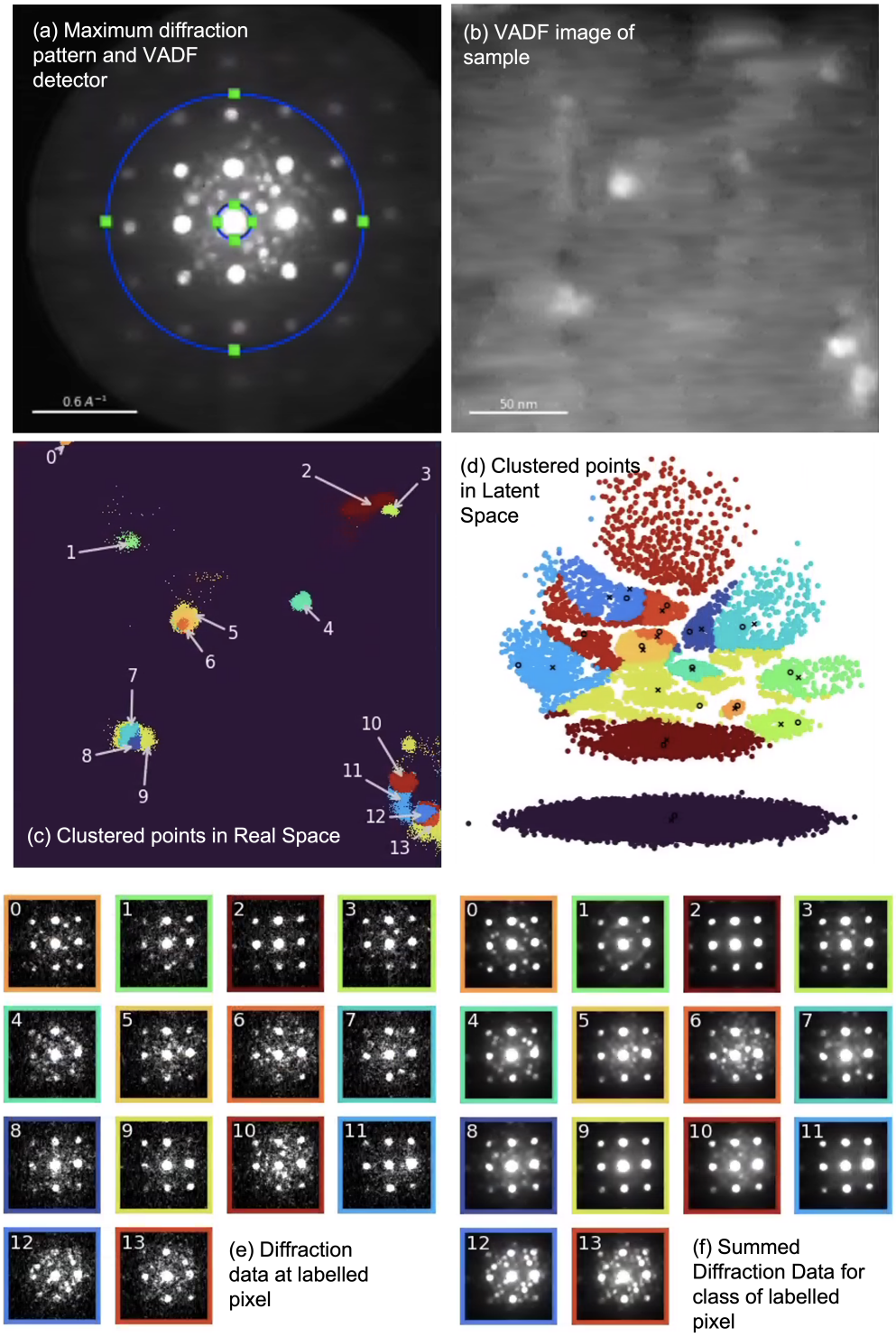

(a) Maximum diffraction pattern and VADF detector
0.6 Å⁻¹
(b) VADF image of sample
50 nm
0
1
2
3
4
5
6
7
8
9
10
11
12
13
(c) Clustered points in Real Space
(d) Clustered points in Latent Space
(e) Diffraction data at labelled pixel
(f) Summed Diffraction Data for class of labelled pixel

Figure 8.5

Figure 8.5: (a) the maximum diffraction pattern of the entire dataset with a virtual annular detector. (b) the Virtual Annular Dark Field (VADF) image for the precipitates in Al alloy thin foil dataset. (c) the clustering on the real space pixel locations, a pixel location has been highlighted in each of the identified domains. (d) the encoded positions of the patterns clustered by colour. The open black circles show the initial centroids identified by NMF and refined by SSI. The black crosses show the new centroids by averaging the locations of points within clusters. (e) The single diffraction pattern corresponding to the pixel highlighted in (c), the pattern border colour corresponds to the colour of the class in the segmentation. (f) The mean diffraction pattern corresponding to the class of the pixel highlighted in (c), the pattern border colour corresponds to the colour of the class in the segmentation.

8.3.3 Phase Distribution in P2 Cathode Material

The final case study is a standard $Na_{\frac{2}{3}}Mn_{\frac{2}{3}}Ni_{\frac{1}{3}}O_2$ P2 structure. The dataset is comprised of 256×256 probe positions in real space with 512×512 pixel array in diffraction plane. The goal of the segmentation is to identify different phases present within the sample, as well as variation in orientation or composition within these phases. For pre-processing, the diffraction information is centred and cropped to 256×256 and then down-sampled by a factor of 2 to get 128×128 input patterns. As before, the irrelevant vacuum and carbon support are masked out to remove unnecessary variance from the VAE training. This is achieved by the more straightforward approach than in the previous example: by masking the central beam and selecting patterns with low sum-total intensities. The remaining patterns are again sampled using the Shannon Entropy. The new training data is used to train the VAE and the workflow implemented to produce the domain map in Figure 8.6.

It can be seen from the latent-space and the subsequent domain map that this is a

more structurally rich sample than the previously studied examples but the workflow provides a robust segmentation, identifying minor phases in classes of fewer than 70 pixels alongside major classes with as many as 10,000 pixels. The segmentation breaks the sample down into 33 largely distinct regions. While some of these regions correspond to multiple classes containing similar diffraction character, this is a conscious decision for the dataset to search for minor factors that might arise due to changes in composition such as super-lattice reflections, and if desired we could tune this by altering the SSI threshold.

It is key to the use of the workflow that there is a human-in-the-loop, to ensure that obtained domains are meaningful statistically (as discussed later in Section 8.4) and then to extract the physical basis for each of the identified domains. Further work is ongoing to help aid the human-in-the-loop in interpreting the patterns, incorporating the orientation matching algorithms used in ACOM[143] to identify changes in crystal structure and orientation. A detailed discussion of its application to this data-set is, however, beyond the scope of this chapter.

A cursory review of the domains observed in Figure 8.6 shows that the domains result from to large differences in crystal orientation (an example seen in the different Bragg peaks present in patterns 3 and 7), minor misorientations (seen by changes in peak intensity in patterns 0 and 11) and changes in superstructure presence (seen by changes in observation of extra peaks in 6 and 8).

Comparing patterns in Figure 8.5 to Figure 8.6 the difference in challenge between the Al alloy with precipitates sample and the layered metal oxide sample can be seen. The individual patterns of the LMO dataset show better signal-to-noise than the precipitate dataset. As a result, there is less of a stark contrast in the domain patterns compared with the individuals, with regards to cluster averaging. However, the quality of the domain mapping can still be qualified by the domain

patterns presenting the same sets of reflections that are present in the individual. By this assessment, the clustering is successful in creating clusters where the deviation between the mean patterns and the individual patterns are small. There is also evidence of the level of variation within a phase being extracted, with patterns of the same phase with small intensity shifts (such as the row of reflections in 4 and 5) being separated out.

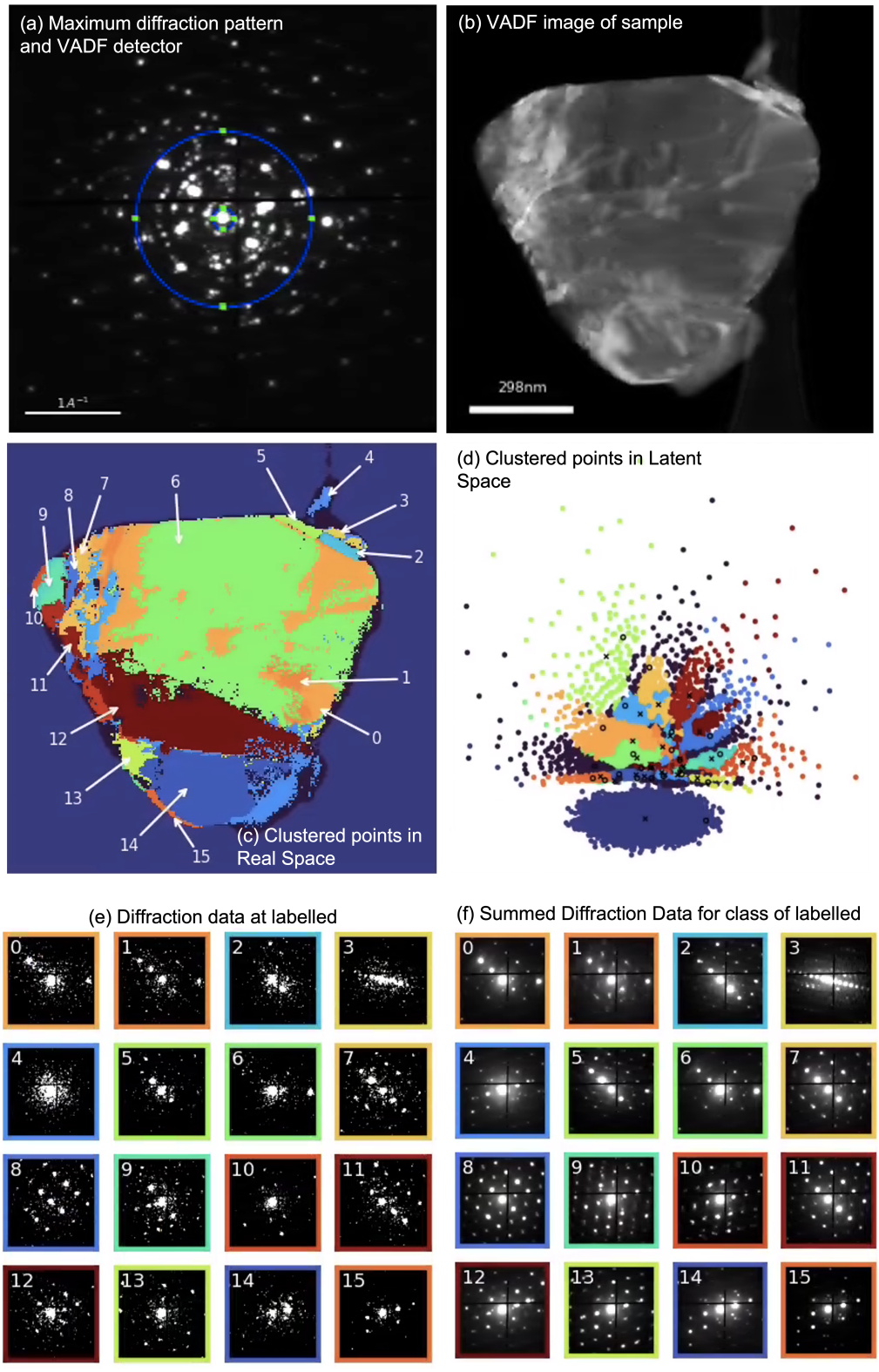

(a) Maximum diffraction pattern and VADF detector
1A⁻¹
(b) VADF image of sample
298nm
(c) Clustered points in Real Space
(d) Clustered points in Latent Space
(e) Diffraction data at labelled
(f) Summed Diffraction Data for class of labelled

Figure 8.6

Figure 8.6: (a) the maximum diffraction pattern of the entire dataset with a virtual annular detector.(b) the Virtual Annular Dark Field (VADF) image for the Layered Metal Oxide dataset. (c) the clustering on the real space pixel locations, a pixel location has been highlighted in each of the identified domains. (d) the encoded positions of the patterns clustered by colour. The open black circles show the initial centroids identified by NMF and refined by SSI. The black crosses show the new centroids by averaging the locations of points within clusters. (e) The single diffraction pattern corresponding to the pixel highlighted in (c), the pattern border colour corresponds to the colour of the class in the segmentation. (f) The mean diffraction pattern corresponding to the class of the pixel highlighted in (c), the pattern border colour corresponds to the colour of the class in the segmentation.

8.4 Evaluating Confidence in Domain Identification

As discussed in previous sections the exact domain map outputted is primarily dependent on the VAE model and the cluster centroids. Altering the training dataset and the length of training will alter the prominence of the different sources of pattern variation in the latent-space and this broadly defines the variation within the diffraction data that the VAE is able to differentiate. How finely/coarsely these variations are separated into different domains will then be determined by the number of cluster centroids and their locations within the latent-space. As such, altering any of the above factors can lead to a change in the final domain map. In order to increase the confidence that the best domain map is being used to evaluate the data (one with the most experimentally applicable separation of domains) a statistical approach was applied.

The motivation of a statistical approach is that while the ideal evaluation would be to go through each individual pattern, within a domain, and check that it has been classified correctly, this is untenable for most domains due to to sheer number of

patterns. Instead, using a statistical approach, the pixels that are most likely to change class can be identified and evaluation can be focussed on the classification of these patterns. Pixels likely to change class are determined by comparing maps generated with different parameterisations, and finding which pixels are consistently grouped together.

The consistency is formalised by comparing all pairs of maps (where one is assumed to be the Ground Truth Map and the other a Comparison) and evaluating for each pixel and its label: what proportion of the total pixels sharing this label in the Ground Truth Map also share their labels in the Comparison Map. These proportions for each pair of Ground Truth and Comparison Map are then averaged to give the overall consistency map. For a given pair of Ground Truth and Comparison Map, the proportion of the total pixels sharing one arbitrary label in the Ground Truth Map and another arbitrary label in the Comparison Map can be obtained from the Proportional Label Overlap Matrices (PLOM). The procedure for calculating a Statistical Consistency Map is summarised below and these steps are illustrated within Figure 8.8:

1. Calculate all Proportional Label Overlap Matrices (PLOM) for every Ground-Truth/ Comparison Map pair
2. For each pixel in real space:
   (a) Find the pixel label in the Ground-Truth Map and the label in the Comparison Map
   (b) for each Ground-Truth/ Comparison Map pair lookup the proportional overlaps for these two labels from the relevant PLOM
3. Average the proportional overlaps of the pixels from one Ground-Truth/ Comparison Map pair over all Ground-Truth/ Comparison Map pairs

Calculating the Proportional Label Overlap Matrix for Ground-Truth Map A and Comparison Map B is visualised in Figure 8.7 and is done using Equations 8.1 and 8.2: where ${M^{AB}}_{ij}$ is the element ij of the PLOM for GTDM Map A and comparison Map B; A^i is the Binary mask of label i in Map A; and A_{xy} is the class label of the pixel at coordinates (x, y) in Map A.

$${M^{AB}}_{ij} = \frac{A^i.B^j}{\sum A^i} \tag{8.1}$$

$$A^i = \begin{cases} 1, & \text{where } A_{xy} = i, \\ 0, & \text{where } A_{xy} \neq i. \end{cases} \tag{8.2}$$

This consistency map provides a visual summary for how invariant the different regions of the sample are to a change in the parameters of the workflow, an example for the precipitates dataset discussed above is shown in Figure 8.9.

Pixels with values close to 1.0 consistently appear in clusters with the same sets of pixels, whereas pixels with low values tend to change which sets of pixels they are grouped with. Lower values in the consistency map correspond to regions which are sensitive to hyper-parameter choice. This allows the consistency map to be used to help assess if a selected domain map provides an appropriate segmentation for the desired experiment. For a selected domain map, each individual domain can be applied as a mask to the consistency map. Assessment of the values of the consistency map within one domain can be used to identify which regions (if any) within this domain are likely to change with a change of clustering parameters. The average patterns associated with each of these identified regions can then be calculated and a visual assessment can be made if the differences between these patterns should, in fact, constitute separation into discrete domains. This can be

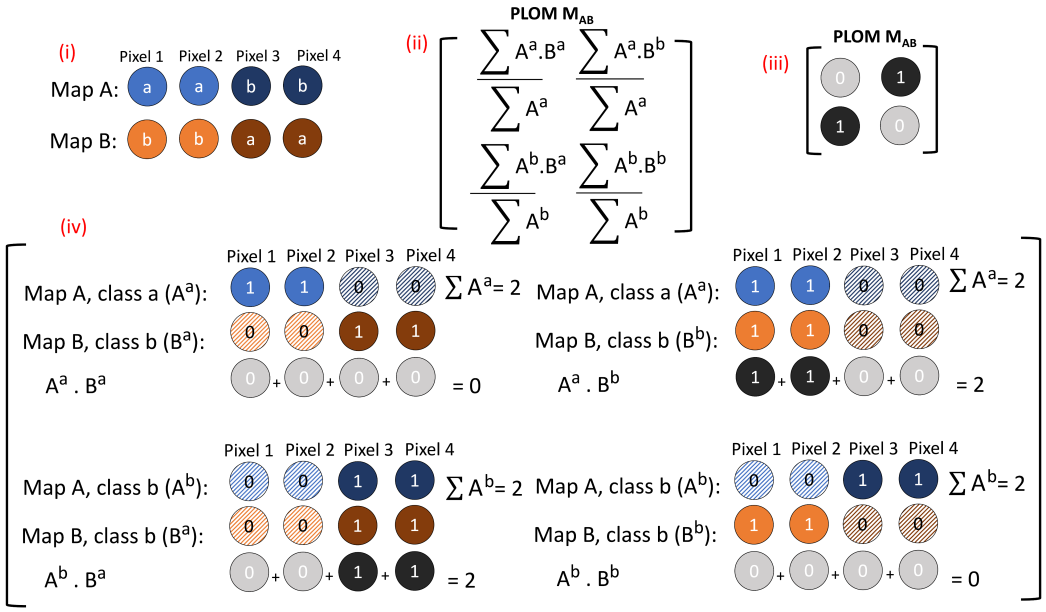


Figure 8.7: An example of the procedure for generating the Proportional Label Overlap Matrix for two 4x1 pixel maps, Map A and Map B (denoted by the red (i), to avoid confusion). Each map contains two arbitrarily assigned labels, a and b. The matrix marked (ii) has each element populated by the specific calculation required to evaluate that element. The values resultant from these calculations are given in (iii). The intermediate steps for evaluating (iii) are given in (iv) with each element showing: the binary masks for the respective labels of A and B; the total number of the specified A label; and the summation that yields the dot product of the binary masks.

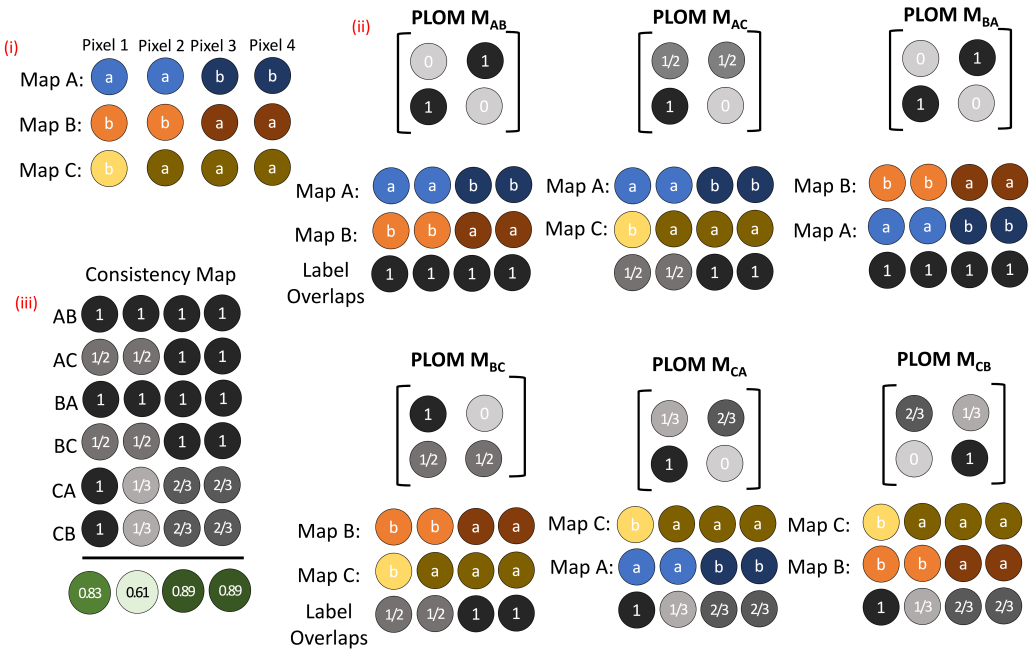


Figure 8.8: An example of the procedure for calculating the Statistical Consistency Map using 3 maps: Map A; Map B; and Map C (denoted by (i)). All the permutations of Ground Truth Map (the top row of each triplet of Map X, Map Y, Label Overlaps) and Comparison Map (the middle row of each triplet of Map X, Map Y, Label Overlaps) are given by (ii) along with the respective PLOM. For each pixel, the proportional label overlap is evaluated between the label of that pixel in the Ground Truth Map and the label of that pixel in the Comparison Map (this can be done by simply referencing the appropriate element of the PLOM). The Label Overlaps for each of the Ground Truth Map-Comparison Map pairs are then averaged in (iii) to give the final Statistical Consistency Map.

repeated for all the domains in a given map in order to increase the confidence in this map as a suitable representation of the data and that the best choice has been made for the number of centroids used in the segmentation.

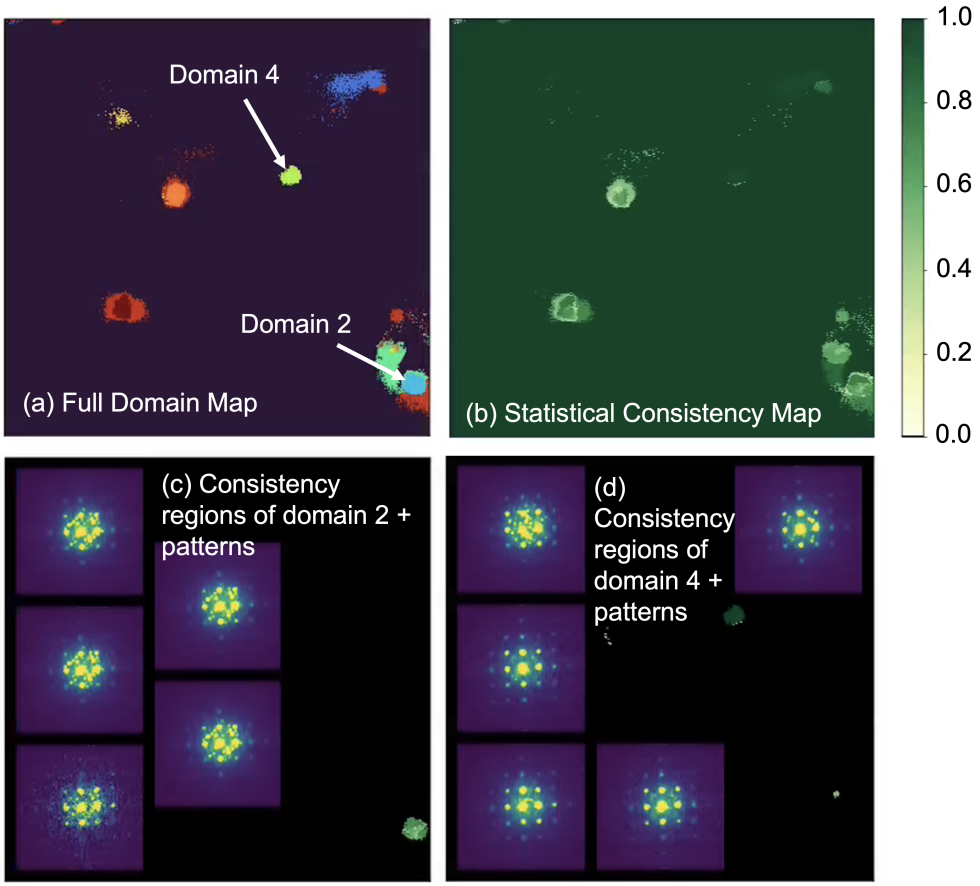


Figure 8.9: (a) domain mapping of Precipitates on Al alloy dataset using 10 components for centroid estimation. (b) Statistical Consistency Map for the domain mapping of Precipitates on Al alloy dataset, comparing 8 domain maps generated using 10,12,14,16,18,20,22 and 24 principal components for the centroid estimation. (c) and (d) Inspection of the diffraction character of identified regions of consistency for different labels (2 and 4) within the domain map.

Figure 8.9 shows an example of applying this protocol to two classes within a domain map. The top left image shows a domain map which was produced using 10 centroids identified using NMF. The top right shows the Statistical Consistency

Map generated by comparing the map using 10 centroids, to a series of maps using 12,14,16,18,20,22 and 24 centroids, respectively. The bottom left and bottom right images, show the regions of the Statistical Consistency Map which correspond to domains 2 and 4 in the 10 centroid map. These are the domain-masked consistency maps. Within each of these two domain-masked consistency maps, it can be seen that there are different consistency regions: regions with higher consistency values (darker green); and regions with lower consistency values (lighter green). Overlaid on to these domain masked consistency maps are the mean diffraction patterns averaged over each of these consistency regions. For domain 2, five consistency regions are identified, and while the bottom-left-most of these patterns is more noisy than the others, all patterns show good agreement in location and relative intensity of reflections, so this domain could be considered satisfactory. Conversely, domain 4 (bottom right), while also showing five regions, displays mean patterns with variations in both peak position and intensity, suggesting that the map in Figure 8.9 (a), is not performing a satisfactory segmentation, by classifying all these pixels as within the same domain.

This procedure can be repeated for each clusters within a given domain map - looking at the consistency regions within the cluster, and deciding whether they are suitably self-similar. If, upon inspection of all the clusters within a map, it is determined that the all the consistency regions are sufficiently self-similar (akin to the evaluation of the example of domain 2), then this should improve the confidence in the domain map as suitable. If however, upon inspection, many of the classes within the map produce consistency region patterns which are not very self-similar (more like the evaluation of domain 4), then this should reduce the confidence in that domain map. Such a situation might suggest investigating the use of different parameters for generating the mapping, or retraining the model with a different distribution of training - ideally a training distribution which more

prominently samples these inconsistent regions, encouraging the new model to better distinguish between their different features.

8.5 Comparison with Other Techniques

In this section, the outcome of the workflow presented in this paper is compared to some other more established approaches for processing 4D-STEM datasets. The techniques in question are a direct NMF of the sample data (without use of a VAE or the hull clustering technique) and a Feature Extraction based approach[149] utilising Py4DSTEM[158] for much of the data processing. For the sake of simplicity, the techniques are compared using the same single dataset, the precipitate on thin Al foil, but the technique is applied to both the raw dataset and also the precipitate masked dataset to see how this pre-processing step affects performance.

8.5.1 Direct Non-Negative Matrix Factorisation

Direct NMF is used on the dataset to extract 16 principal components (the same number used in the workflow example above) and the patterns are compared against these components using a root-mean-squared (RMS) comparison and a SSI comparison.

In Figure 8.10, the advantage of using the VAE to learn the sources of variance can be seen. As the VAE can be directed to learn to identify the precipitate peaks first, and can be truncated before learning the more subtle variations in the substrate matrix (as described in 8.2), the rebalanced dataset generated will only show variance corresponding to the reflections of the precipitate. Thus, when performing NMF, the set of components produced will also be dominated by these precipitate peaks (Figure 8.10, right). This is not the case when performing NMF directly on the raw dataset, even if only using patterns from within the precipitate regions

to perform the NMF, the resultant components are dominated by variation in the intensity of the substrate reflections (Figure 8.10, left).

In order to use the components to then generate clusters, a typical approach would be to compare the sample patterns to the principal components using RMS. Using SSI can provide a superior comparison metric but takes longer, and is still ultimately limited by the weak component detection and in this case produces an even poorer domain map than RMS (see Figure 8.11. Neither of the maps in Figure 8.11 are able to differentiate between the different precipitates in a satisfactory manner. This can be seen explicitly by looking at the patterns in Figure 8.5 (e) and (f). For example, the single patterns of pixels 3 and 12 are clearly from different orientations and the mean patterns in (f) corroborate this. Both methods in Figure 8.11 would have these pixels in the same domain (and would in-fact combine almost all of the domains identified in Figure 8.5 into one).

Inspecting the components generated during the centroid estimation, the benefit of the VAE workflow approach becomes apparent, as while the different components identified are indeed driven by the different precipitate patterns, the artifacts of the NMF limit the direct applicability of comparison metrics. Figure 8.12 shows the centroid estimation of each of the components identified by the workflow. The use of these centroid patterns in the comparison metrics as opposed to the components identified by direct NMF in Figure 8.10 accounts for the majority of the performance improvements.

8.5.2 Feature Extraction

Feature Extraction is an alternative contemporary approach outlined in Bruefach et. al.[149] This approach performs three methods of feature extraction on the dataset to represent the diffraction information and then for each, an iterative NMF

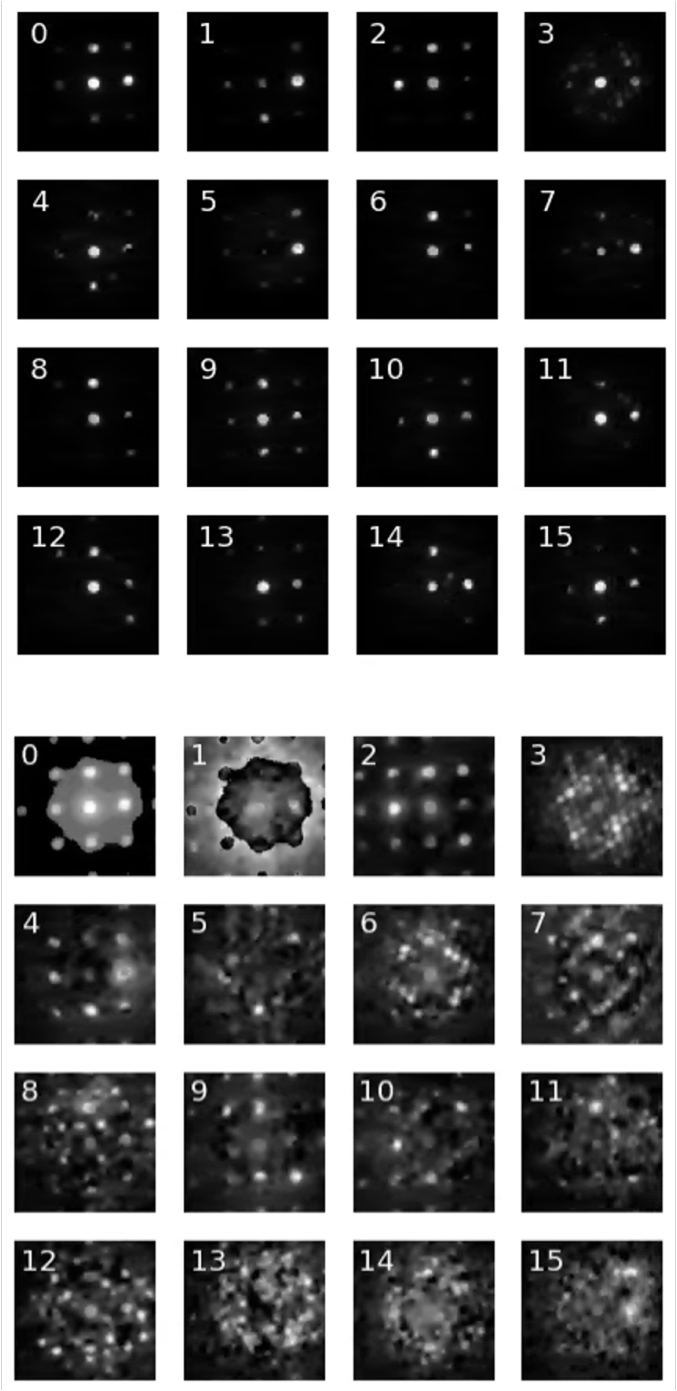


Figure 8.10: The 16 principal components extracted: from the masked Precipitates on Al alloy dataset (top); and from the redistributed dataset generated from decoding the latent-space of the masked Precipitates on Al alloy dataset (bottom).

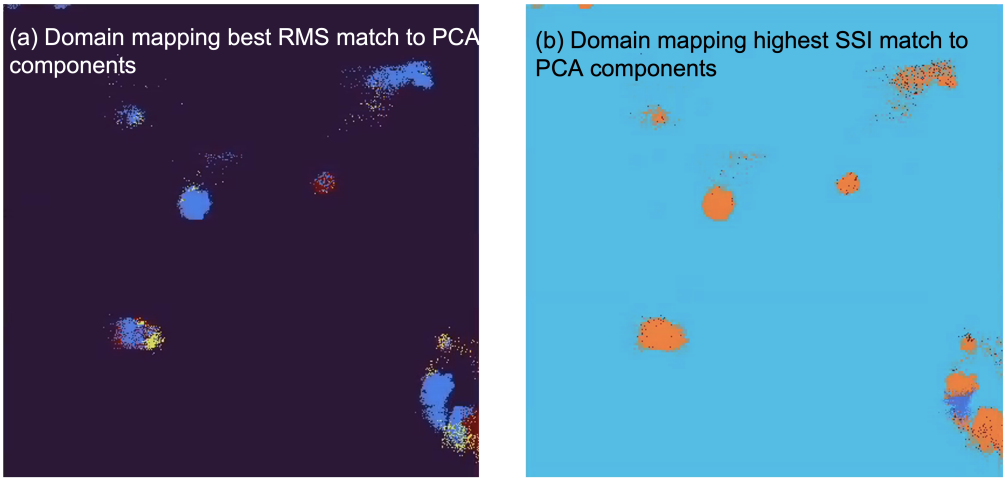


Figure 8.11: The domain maps produced by direct NMF using a RMS comparison metric (left) and an SSI comparison metric (right).

decomposition[159] to produce domain maps. The feature extraction methods are Radial Variance, Angular Averaging and Bragg Disk Detection. More information of the specifics are available in the manuscript. As before, the workflow is applied to both the raw thin foil Al alloy and the precipitate masked thin foil Al alloy datasets.

For the radial variance feature extraction, a polar elliptical transformation is applied to convert the Cartesian coordinates of the detector grid into polar coordinates. The variance for each row of pixels (corresponding to a single radius from the centre or a radial ring in the original pattern) is then calculated.

The angular averaging requires finding the local minima of the radial integral of the maximum diffraction pattern. This is used to approximate the location of the radial rings of reflections. The largest and smallest rings are used to define a reflection range and every 5 rows are averaged to return the angular average.

The Bragg disk detection is performed using built in methods of py4DSTEM[158]. It involves convolving a probe kernel with each pattern and detecting the max-

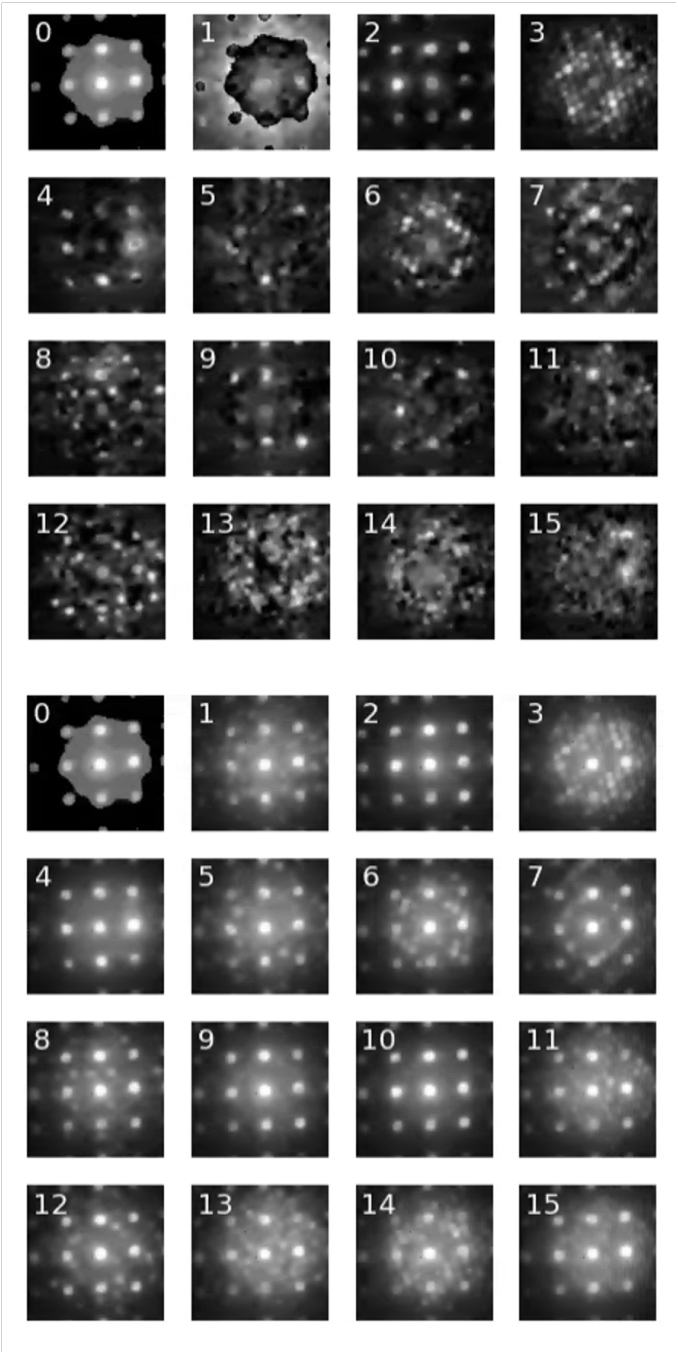


Figure 8.12: The 16 principal components extracted from the redistributed dataset from the masked Precipitates on Al alloy latent-space (top) and the centroid pattern used to represent each component (bottom).

ima of the cross correlation to identify diffraction peaks. These peaks are then rasterised into a grid with 3×3 bins to reduce the dimensions and this grid is vectorised to produce the features list.

As before, the workflow is applied to both the raw thin foil Al alloy and the precipitate masked thin foil Al alloy datasets.

The results of the feature extraction workflow can be seen in Figure 8.13. The unmasked dataset shows that the substrate variation dominates the domain maps for all three feature approaches. For the masked dataset the Bragg Peak method performs the best, identifying some regions within precipitates, but the segmentations are noisy and the precipitates are over-segemented. These shortcomings are likely due to the low precipitate peak intensity present in a single patterns resulting in these features being poorly expressed when principal components are identified using NMF. It is plausible that through optimal parameter selection the results of this method could be improved however, with the size of this dataset ($511 \times 511 \times 256 \times 256$) and the default processing pipeline, the iteration time for parameter tuning is prohibitive. An approximate comparison of the compute times, on the same hardware, for the Bragg Peak Feature Extraction workflow and our workflow are given in Table 8.1. The run times were calculated running the code on a cluster on the Diamond Light Source jupyterhub[160] equipped with 16 CPU cores, 128 GB RAM, Nvidia Tesla V100 GPU, Cuda 10.1.

The iteration time is the time required to complete the workflow if a change is made to the given process. As completing the workflow is often required to evaluate the quality of the parameter change, this metric can be used to indicate how quickly parameter iterations can occur at each stage of the workflow. From the iteration times in Table 8.1, it can be seen that parameter tuning, even for the last step of the Feature Extraction workflow, will take in the order of hours, whereas

Bragg Peak Feature Extraction workflow			VAE Workflow		
Process	Process Duration	Iteration Time	Process	Process Duration	Iteration Time
Bragg Disk Detection	12hr 25min 9s	19hr 39min 31s	Model Training	35min 33s	43min 16s
Disk Intensity Calculation	1hr 26min 54s	7hr 14min 22s	Pattern Pre-processing	1min 46s	7min 43s
Universal Thresholding	1hr 26min 35s	5hr 47min 28s	Pattern Encoding	32s	5min 57s
Iterative NMF	4hr 20min 53s	4hr 20min 53s	Boundary Point Estimation	1min 14s	5min 25s
			Centroid Estimation	1min 35s	4min 11s
			Clustering	2min 36s	2min 36s

Table 8.1: Comparison of approximate durations of processes in Bragg Peak Feature Extraction workflow with processes in the workflow present in this paper, on the same hardware. The iteration time given is the approximate time to complete the domain mapping if a parameter adjustment is desired for that process. The iteration time of the first process gives the approximate duration of the entire workflow.

our workflow can be iterated from complete model retraining in under an hour with the clustering behaviour tuned in closer to 5 minutes per iteration.

The memory usage for the two workflows was also calculated by running the same operations on the Diamond Light Source Hamilton Cluster equipped with Nvidia P100 GPU and 128 GB RAM. The Bragg Peak Feature extraction used a maximum of 36 GB of virtual memory (vmem) whereas the workflow in this paper used a maximum of 116 GB of vmem.

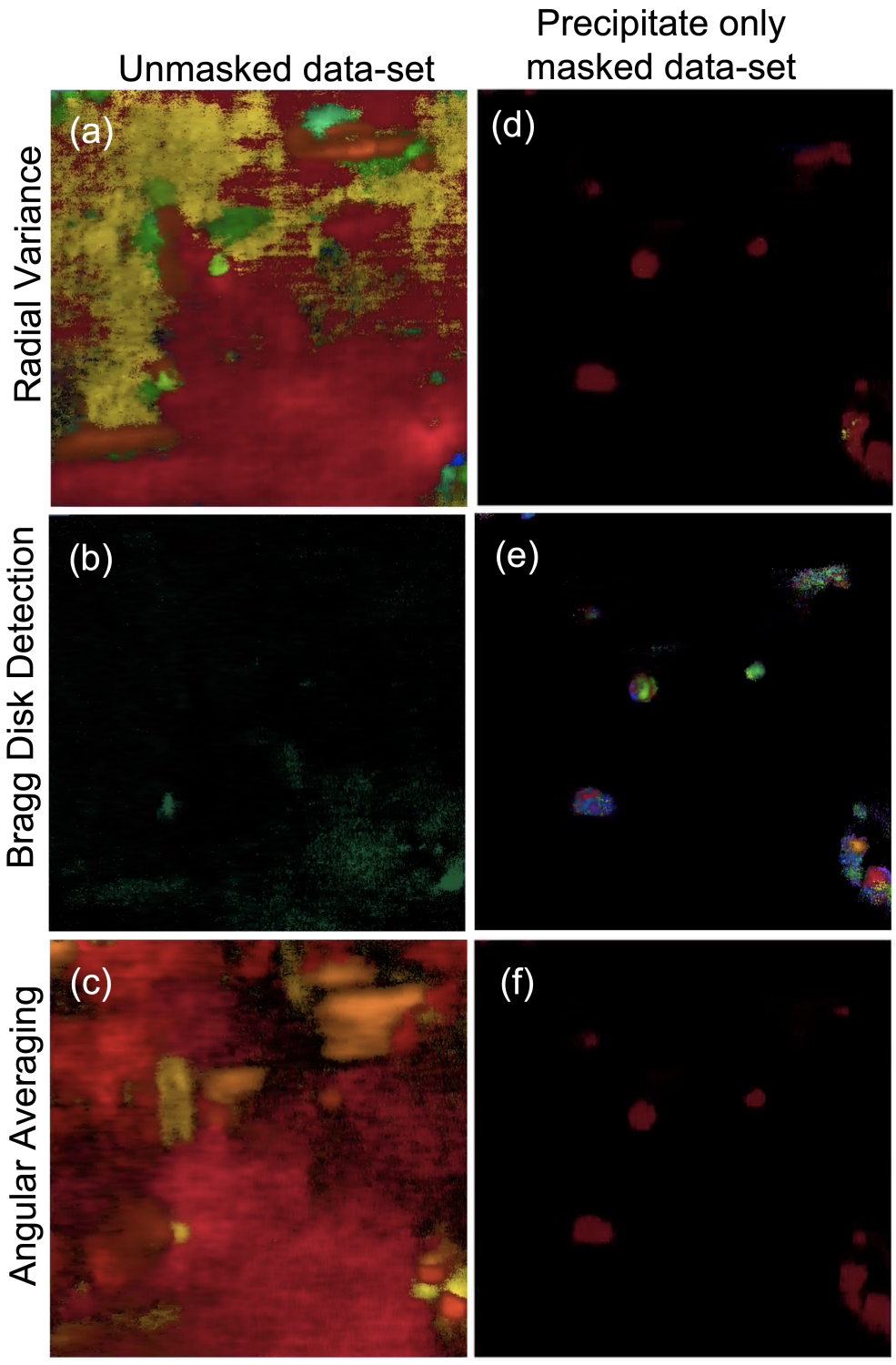

Unmasked data-set
Precipitate only masked data-set
Radial Variance
Bragg Disk Detection
Angular Averaging
(a)
(b)
(c)
(d)
(e)
(f)

Figure 8.13

Figure 8.13: The left column is the unmasked full thin foil Al alloy dataset clustered using (a) Radial Variance, (b) Bragg Disk Detection and (c) Angular Averaging. The right column is the precipitate only masked thin foil Al alloy dataset clustered using (d) Radial Variance, (e) Bragg Disk Detection and (f) Angular Averaging.

8.6 Workflow Automation

The discussions throughout the previous sections have intended to highlight the flexibility of the workflow in different use cases, but also emphasise that a versatile approach, requires directing towards producing domains optimised for the investigations of each experiment. This then promotes questions of how much intervention and resources are required for analysing each new dataset. The drive towards more autonomous workflows is an aim to promote higher throughput analysis and improve the statistical significance of the observations that can be drawn out of these domain maps.

The first area for automation is at the VAE training. While curating the training data is required to optimally bias the encoder towards learning relevant variation, this processing pipeline should not require alteration from sample to sample within the same experiment. This means that once a pre-processing routine has been identified which satisfactorily biases the encoder, this routine can then be applied automatically to other datasets with the same desired domain separation. For example, for the Layered Metal Oxide dataset, the pre-processing requires removal of the carbon support and vacuum patterns, and then balancing the dataset to increase the occurrence of high entropy patterns. With this identified, to produce the domain map of other LMO samples, the same pre-processing workflow

could be applied automatically to each dataset to produce the training set.

With the training of the VAE requiring the bulk of the total processing time, automating this process provides a significant improvement in the rate at which domain maps can be produced (see Iteration Time in Table 8.1). The testing of different parameterisations in the clustering step can also be automated, to produce a set of domain maps which can be used for producing a Statistical Consistency Map. The final step of selecting whether an outputted domain map is suitable, however, currently requires human assessment.

An additional automation step can be inserted here: replacing the PCA-based centroid identification (which requires setting a desired number of clusters) with an automated ML-based centroid identification routine. This is a trade-off rather than an outright improvement, as it requires giving up some of the fine grain control over the granularity of the clustering in exchange for less user-input. Such a ML based approach has been implemented and will be outline below.

8.6.1 ML Centroid Detection

From Figure 8.3 and as discussed throughout Section 8.2.1, it can be seen that the bias of the VAE encoder leads to separation into distinct regions within the latent space. The approach for identifying these regions autonomously is non-trivial however, primarily due to the density variations in region populations. The centroid approach detailed above works well for separating out these regions and is especially useful when the boundaries between regions are not obvious divides. The PCA centroid detection can be used in a hierarchical manner to interpret region boundaries without requiring obscure parameter tuning for individual datasets. The limitation of this approach is that some user intervention is required even when the number of regions is more apparent from inspection of the latent

space.

To account for these situations, an additional ML-based centroid detection approach was developed which can be used in place of the "Estimate Centroids" section of the workflow (Figure 8.1).

This ML based approach leverages the Kernel Density Approximation that has already been generated to estimate the boundary points as this can provide fixed size input data that can be fed in a small neural network.

This neural network takes in a 2-d array of the Kernel Density approximation of the latent space encoded data, feeds it through a small Convolutional network, and outputs the same 2-d array but with softmax activations with values of one at predicted centroid locations and zeros elsewhere.

For this network to function, this requires labelled data for training. As the purpose of this ML-based approach is for situations where the experimental data is encoded into a latent space with well defined regions of varying density, the model can be trained by simulating such datasets.

The simulated data for training were generated by randomly produces a random number of centroid locations. From these a voronoi tessellation is produced around the centroid locations to create well defined boundaries. For each voronoi region a random point density is then generated and each region is uniformly populated with sample points based on this density. The centroids are then recalculated to be the mean of these uniform points. The uniform points are then randomly adjusted in the direction of the new centroid location to produce boundaries between the regions and a general fall off in point density towards the edge of the region. The final step in the simulation is to randomly remove some of the regions to remove the expectation of the regions to fully tile the latent space.

From this simulation data, the input and target data for network training can be generated using the KDA and the calculated centroid locations.

The application of these steps are shown in Figure 8.14 and this ML-based centroid detection is also used later for the SEND processing shown in Figure 9.2.

8.7 Workflow Limitations and Resource Requirements

The resource requirements of this process are high, due to the training of the VAE requiring GPU acceleration to execute in a workable time, and this is certainly a limitation of the workflow. To mitigate this training time, if there is access to sufficient memory, it is possible to aggregate all the pre-processed training datasets together and train one model on all of the data; this model should have a larger latent dimension to handle the added variance and avoid overly compressing the data. As a result of the architecture of the model, transfer learning can then be used to help improve the training time. As the model is made up of 2D Convolution (Conv2D) layers, followed by densely connected (Dense) layers, it is expected that the Conv2D layers learn the low level concepts and then these are passed to the Dense layers which learn to apply these concepts to the specific sample case. Therefore we can apply transfer learning to take the weights of the outer Conv2D layers of this bulk-trained model and apply them to the standard VAE architecture, which is then trained on the specific training dataset in reduced time.

Alongside limitations imposed by resource requirements we would also like to highlight other potential limitations of the workflow and areas where future work will be directed. The first limitation to discuss is the quality of the diffraction data, while the workflow will function with any 4D-STEM data-set, there is a limit to what can meaningfully be extracted with noisy/ low signal data. Additionally to this, the sampling rate of the scan data across the sample will also play a role in

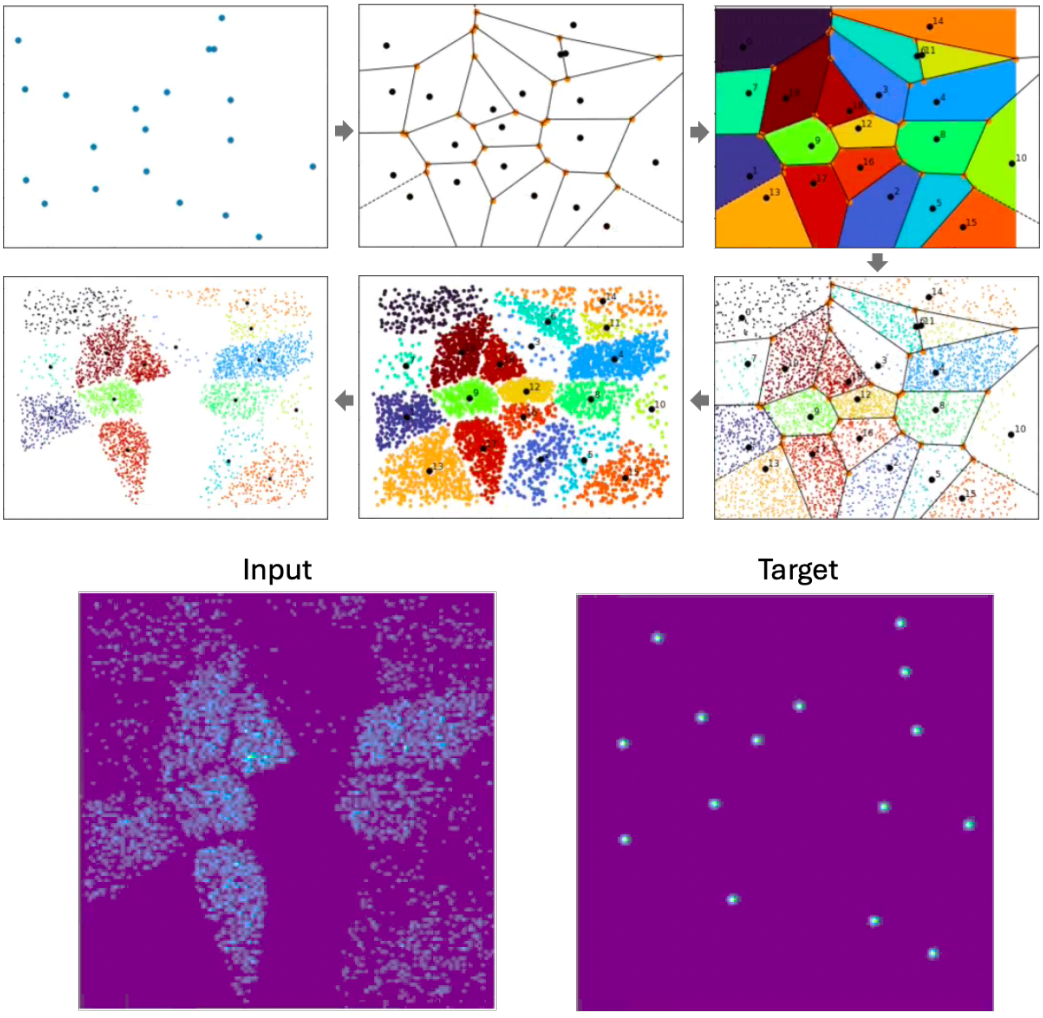


Figure 8.14: Production of the input and target data for the ML-based centroid detection.

the quality of the segmentation.

When the probe size is large compared to the size of the domains of interest, such that a significant proportion of the recorded patterns show overlapping diffraction information (this might be anticipated in a poly-crystalline sample where grains with similar orientations or twins are present across a very small area) then the workflow may struggle to satisfactorily identify the unique domains. We do not, however, observe that the presence of probe scans at locations with overlapping grains is an inherent limitation. Provided there are scan locations where just the individual grain are present, the workflow should be able to identify the separate domains. For such situations with overlapping grains we might anticipate a response similar to the thickness variation in the simulated data-set, where the exact edge becomes less well defined or, depending on the size of the overlap in the sample and the parameterisation of the segmentation, the overlapping patterns might be identified as a separate unique domain. Neither of these scenarios should cause issue in further structural analysis, as the subsequent zone-axis indexing step being developed utilising ACOM[143] is capable of to dealing with overlapping patterns. In more difficult cases, where most patterns in the sample show some form of grain overlap/twinning, we would also still expect the workflow to help cluster regions where the nature of this overlap/twinning is similar, if such regions exist. In the most extreme case, however, where there are highly varied systems with features of interest that are small relative to the probe size, this workflow would probably not be very applicable.

Additionally, the workflow is limited by the requirement of some guidance from a domain expert. The workflow relies on having a human-in-the-loop to help direct which similarities within the diffraction data are relevant to the particular experiment, and that a pre-processing pipeline can be setup and applied in order to enforce this bias on the VAE.

To summarise, the limitations of this technique are a result of it being reliant on the information present within the distribution of patterns available and it requiring the input of a domain expert to aid in biasing the workflow. If the features of interest cannot be extracted from within the distribution of patterns then the workflow will not provide a satisfactory segmentation. Difficulty extracting these features will most likely be a result of: features being unobservable in the data-set due to poor signal to noise; features being poorly resolved spatially; a prohibitively large number of features being present; or that the pre-processing workflow has not effectively biased the VAE to extract them.

8.8 Conclusion

Within this chapter an automated workflow for analysing SEND datasets has been developed. The approach presented while largely automated, is flexible enough to allow the adaptation of inductive biases in the model, to account for different experimental targets. The power of this methodology has been shown first on a simulated dataset where ground truth is available for evaluation, then by extracting useful domain maps from a heterogeneous set of experiments where the data and the desired outcomes are significantly different.

The workflow was able to extract sensible domain maps with minimal human intervention (aside from deciding the aim of the experiment) and within a reasonable time frame. The workflow was then compared to other common approaches for SEND data analysis and showed that this approach is more versatile and more efficient, ultimately yielding better more meaningful domain maps. Finally, this chapter demonstrated production of confidence maps associated with the SEND analysis, that would allow the experimenter to identify and investigate regions where domain classifications are ambiguous.

With this workflow now available as a tool for structural investigation, it can be applied to SEND datasets of cathode materials to aid in the structural studies of the following two chapters.

The breadth of the applicability of the workflow beyond the following chapters however, is yet to be fully established. An option for testing the limits of the process would be to investigate how it is effected by using different convergence semi angles. As discussed elsewhere, small convergence semi-angles where used for investigating the samples in this work as they display large c lattice parameters so have correspondingly small $\mathbf{c}^*$ vectors. In order to resolve the positions of these $\mathbf{c}^*$ vectors, small semi-angles are necessitated, however, in many systems this will not be the case. Larger semi-angles will allow for a smaller incident probe on the sample and so would allow for investigating domain sizes with further improved spacial resolution. In the extreme, where the semi-angle is large enough that bragg discs begin to overlap, these overlapping discs will contain phase information which is difficult for humans to interpret, but application of a dense neural network like a VAE might extract meaningful insight.

The limit of this approach could also be tested, as noted before, it is anticipated that this approach would struggle with poly-crystalline materials, where the physical domains are small with respect to the probe size. It would be interesting to study increasingly complex system of this type in order to see where the exact limit of the workflow is. It seems likely that what ultimately limits the approach is the interplay of the different factors to the loss metric in the VAE and how many different unique domains can be distinctly projected into the latent space while satisfying those constraints. With optimised tuning of the weighting of between the reconstruction loss and the KL-divergence, it may be possible to mitigate the difficulties of many-domain datasets to an extent, but it is probable that a limit will be reached eventually.

Further applications to crystal systems similar to the ones chosen could be potentially be achieved with the addition of precession to the probe. This would allow for better agreement between the obtained patterns with the kinematic model, which could lead to improvements in the quality of domain separation and in the subsequent identification of orientation. Specifically, for the P2/O3 mixed phase systems, the use of this workflow, along with precession, could allow statistically significant insight into the size and distributions of the relevant stacking regimes. Precession would be advantageous here as very similar bragg reflections locations can be achieved with specific orientations of both the P2 and O3 structural model, so having meaningful intensities would be a key step in distinguishing between the two phases.

A final avenue for application is to implement this as a step in a larger scale autonomous process. As SEND data collection routines become more and more automated, a full screening pipeline could be developed for material systems where finding suitably large regions of uncommon orientations would be advantageous (such as identifying regions of the P2 material with the **c** axis in plane for HR-STEM). The microscope could be set to find and collect SEND datasets of all the particles on the grid and then these could each be passed through the workflow to identify the domains and produce cluster-averaged patterns. All these domains could then be autonomously orientation matched and the domain size quantified and the regions which fit the desired size and orientation criteria could then be identified for further study.

9 | Investigation of Intra-layer Ion Orderings in Sodium Ion Cathode Materials

In the following two chapters develops an improved structural model of both the intra-layer and inter-layer ordering of the sodium and transition metal ions across a stoichiometric series of formally 'P2-type' systems. The series is N33, N30 and N25, so named for the ratio of Ni in the transition metal layer (for example N25 is $Ni_{0.25}Mn_{0.75}$). The full target compositions are given in 7.2. The aim of this investigation will be to produce a single model that is more consistent with all the collected structural data than both the standard P2 model but also the 'Large-ZigZag model. This will be done by independently using an empirical simulated annealing approach and a symmetry driven approach to derive plausible sodium ordering schemes that are consistent with observed electron and x-ray diffraction patterns. To model the transition metal ordering, the same empirical approach will be used, as well as probabilistic stacking models to compare the results with experimental neutron diffraction data.

9.1 Evidence of Sodium-Ion Superstructure in Experimental Data

As discussed in Chapter 5, the currently accepted model of sodium ordering in the N33 system is the Large Zig Zag (LZZ) model. Throughout our studies of this material system, we have observed numerous indications of extended sodium ion ordering occuring within these P2-type compositions. Despite this, the experimental findings we have collected appear to be inconsistent with the basic model of the LZZ suggesting that this description, is at the very least, incomplete.

Specifically, the evidence of extended sodium ordering is given by superstructure reflections, which have been observed in SEND, HR-STEM and PXRD. These are shown in Figure 9.1.

The hexagonal 'snowflake' intensity pattern observed and shown in Figure 9.1 was obtained from the application of the workflow outlined in the previous chapter (Chapter 8) to the SEND dataset of an N33 sample. A summary of the output domain mapping, along with the patterns associated with some of the major regions is shown in Figure 9.2. From this figure, it can be seen that there are a few large regions displaying this sodium ordering structure clearly, as well as some more regions where there appear to be two overlapping grains, rotated slightly in-plane about the **c** axis, but both of which show the same snowflake pattern.

These experimental diffraction patterns can be directly compared to simulations for the LZZ structure and evaluated for consistency. Noting that exact reflection intensities will depend upon the specific registration of susbequent LZZ layers, as well as dynamical effects of multiple reflection. A typical pattern is shown in Figure 9.3.

From the comparison of the LZZ simulation, it can be seen that the 'snowflake' intensity pattern observed in the SEND and the HRSTEM data does not seem to be replicated by the LZZ, however, many positions of bragg reflections themselves do align well. This suggests that the sodium ions order with the same periodicity as the LZZ, within a $2\sqrt{3}$ supercell of the standard $P6_3/mmc$ P2 unit cell. Additionally to the electron diffraction evidence, it is also observed that the LZZ model can not be used to provide a good fit for the PXRD super lattice reflections, as will be discussed in Section 10.3.1 and shown in Figure 10.13.

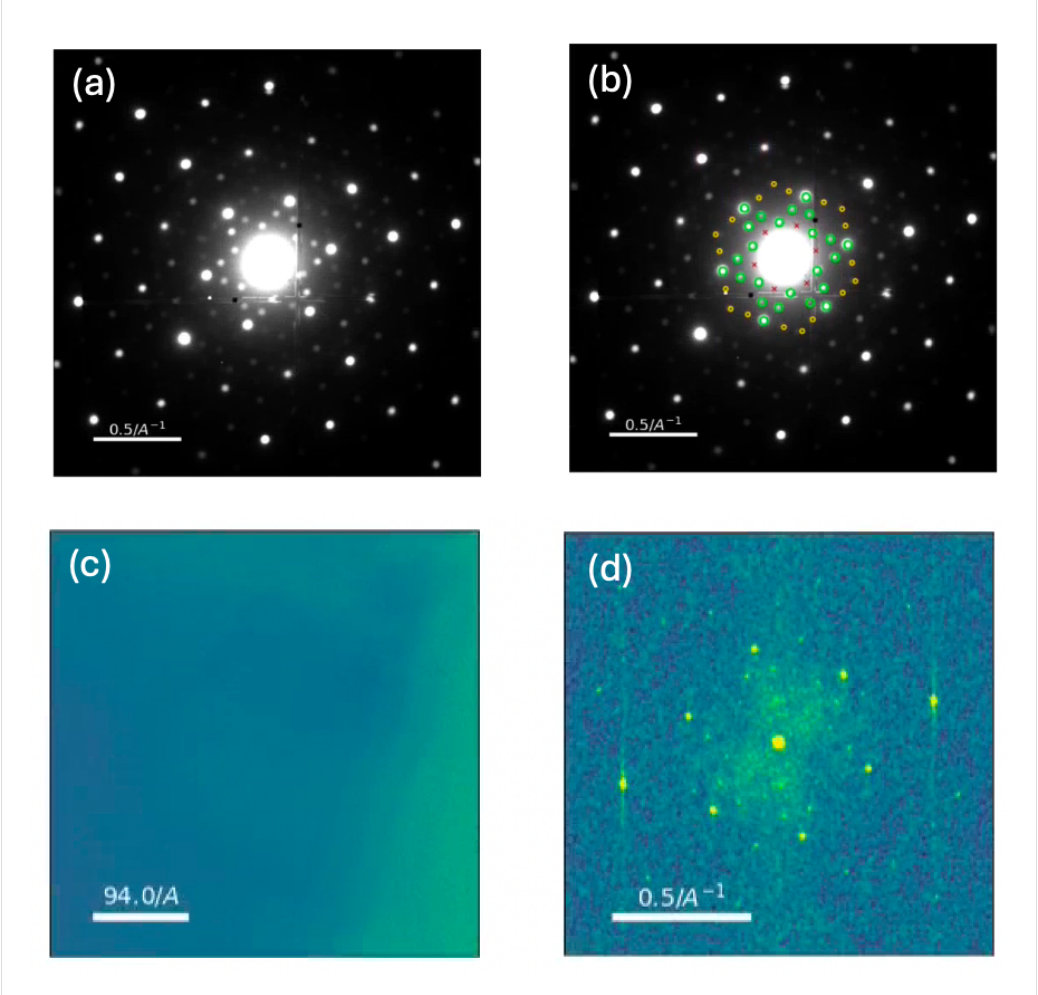


Figure 9.1: Superstructure evidence. (a) Experimental SEND data with extra superstructure reflections. (b) Experimental SEND data with the primary superstructure reflections shown in green, the secondary ones in yellow and the characteristic absences shown in red. (d) The Fourier transform of HR-STEM data (c).

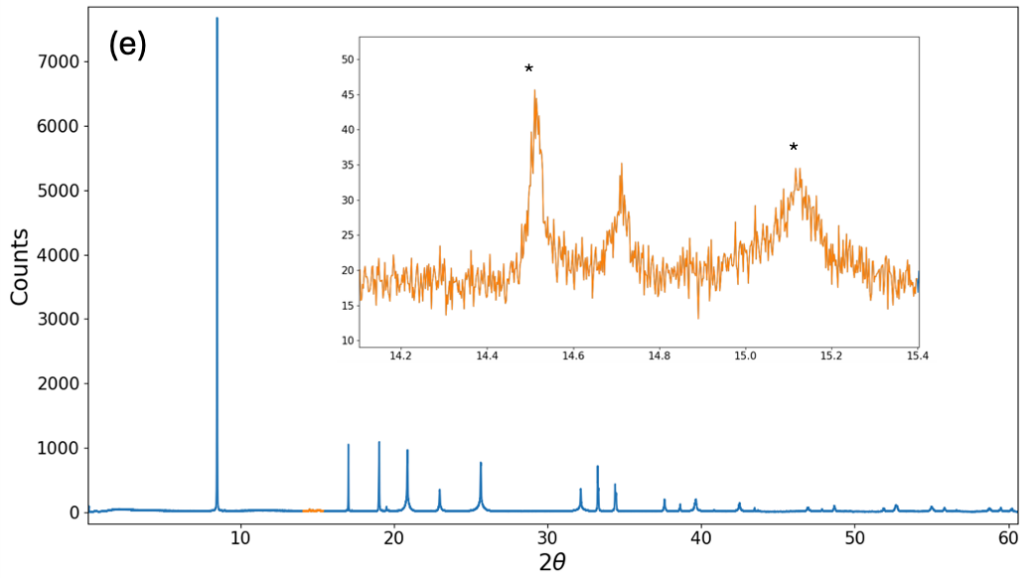


Figure 9.1: Superstructure evidence. (e) PXRD data with superstructure reflections shown by * on the insert

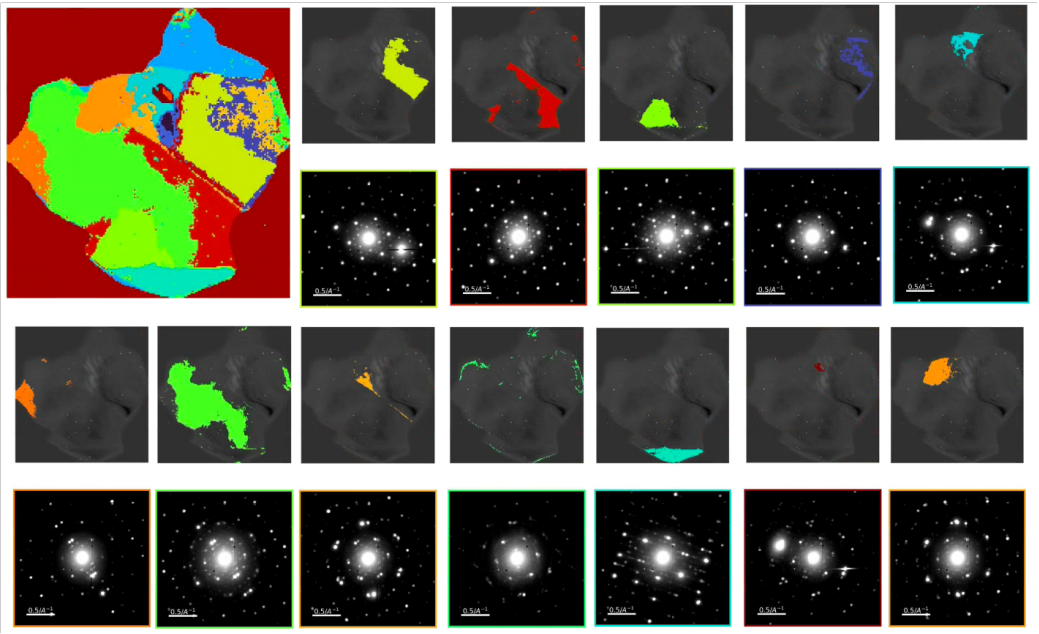


Figure 9.2: SEND automated domain mapping of a N33 sample, extracting regions of 'snowflake' sodium ion cathode superstructure.

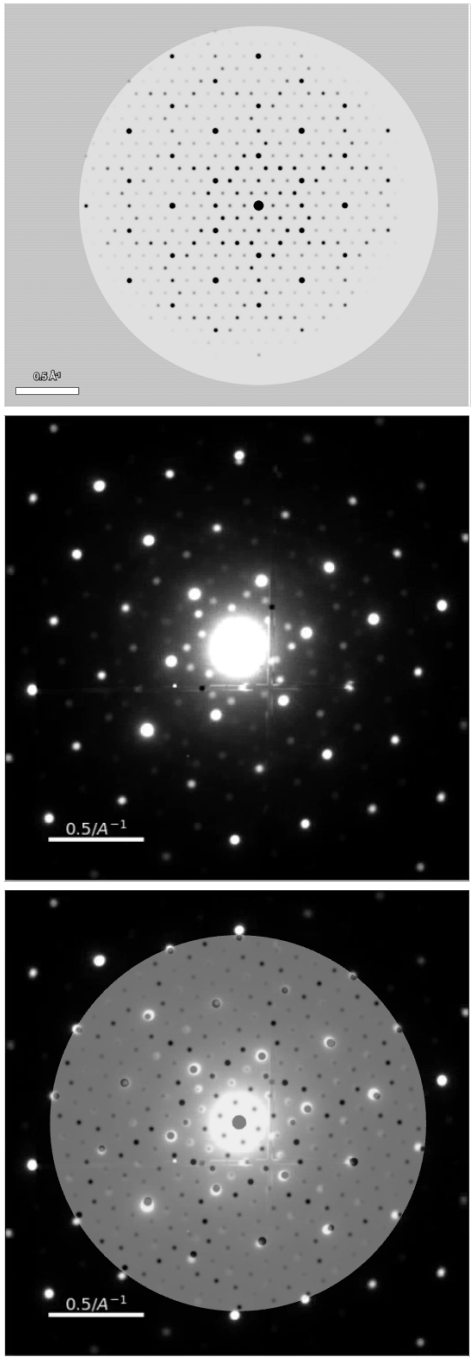


Figure 9.3: (left) A TEM simulation of the 2D electron diffraction pattern for the LZZ structure. (centre) The SEND summed diffraction pattern showing superstructure reflections. (right) An overlay of the LZZ on the experimental reflections.

9.2 Development and application of structural model using an empirical potential

There are a few standard approaches that one might employ to try and decipher the exact nature of the sodium ordering, such as the use of Virtual Aperture Dark Fields (VADFs) or the application of inverse Fourier transform to the subsets of the observed reflections. Perhaps unsurprisingly, there was little progress to be made via these approach, as if such methods were sufficient, then we would likely already have comprehensively characterised ionic orderings. We will, however, save the discussion of why they were not successfully applicable for the end (Section 10.4). This is done for narrative clarity, as it allows us to use the insight from the models we develop throughout the following chapters to aid in understanding why the more traditional approaches proved challenging.

Failing to gain insight by standard methods, the first approach to find a more appropriate sodium ordering scheme was to use simulated annealing in tandem with an empirical potential to optimise the positions of the sodium ions and minimise the energy of the empirical potential. It was chosen to use an empirical potential rather than a more rigorous DFT-type approach to optimise iteration times and gain a more insightful understanding of the interplay between competing energetic factors that would be lost using an 'out of the box' pseudo-potential.

The empirical potential used was based on Coloumbic interactions and evaluated both scalar magnitudes of repulsive coloumbic forces between cations in the structure and the resultant vector from the same repulsive forces. This empirical potential approach solely intends to be used to guide exploration and as an indication of possible low energy structures. Ultimately the determination of the quality of the super-structure models output by this approach could only be evaluated by

external metrics, such as consistency with experimental data, rather than any internal measures of relative energy according to the potential itself. With this being said, the applicability of a coloumbic-based empirical potential as a guide appears appropriate, due to the energetics of the closed shell sodium ions likely being dominated by such coloumbic interactions.

9.2.1 Development of the model and the potential

The code for the empirical model was written in Python and, other than a single, initial, nearest neighbour evaluation taken from pymatgen[161], the code was essentially written from scratch using Numerical Python[162]. The reason for this was performance. While Python itself is notorious for slow performance, the almost exclusive use of Numerical Python, 'numpy', allowed for implementing most of the code using, just in time, numba decorators[163] to convert key functions straight to machine code. This, alongside other vectorising and algorithmic improvements discussed later, allowed the calculations to move from tens of iterations per second to tens of thousand of iterations per second. These speed ups were integral to performing a sufficient number of iteration steps required for an effective optimisation. To further improve performance, maximal care was taken to ensure that only the components of the empirical energy function that directly change as a result of a simulated annealing swap were recalculated, so as few additional calculations were required. This involved representing much of the data in terms of abstract neighbouring charge arrays. Additionally to the empirical potential, the code also implements a measure of Pauling's electrostatic valence principle (EVP)[164]. Initially, this was calculated and used directly as an energy measure, however, this did not yield productive results, due to many degenerate solutions to the EVP being obtainable. Instead, it was used as an additional qualitative metric, with sodium ordering patterns which closely follow the EVP being

considered as more plausible than those which displayed high deviations.

Despite a focus being placed on performance optimisation, this was approached within the context of modularity, with the code designed to be applicable to all structures within this class of systems. The code allowed for fully customisable starting configurations within an arbitrary sized unit cell.

A summary of the approach is given in the following section.

9.2.2 Summary of the model and the potential

The first step is to load a base structure. This structure should be a layered structure, like those described above (Section 4.1), and should contain the locations of all the possible sites that ions will be able to occupy.

Next the unit cell is expanded to create whatever size supercell is desired. In the N33 case, where the base structure used was the $\sqrt{3}$ larger P2 cell, so the final supercell size is $2\sqrt{3} \times 2\sqrt{3}$.

After the supercell is created, the structure is split into its constituent layers in the z-direction. These layers are then populated to the desired specifications, either with a random population of the desired number of each ion, a random population sampled from the proportion of each ion, or with a predefined starting configuration.

The final setup step required is to set the parameters for the empirical potential and the annealing steps. These parameters and their effect on the potential will be noted as they are discussed throughout the rest of the summary.

To begin the simulated annealing, the empirical energy of the starting configuration must first be calculated. This is done by taking each of the cations in the structure and finding the relative position and effective charge of their cation

neighbours within a defined radius. The effective charge is calculated using two additional parameters defining shielding coefficients, to account for the interaction between transition metal ions and sodium ions and differences in whether these ions are interacting through the face or edge of the surrounding oxygen polyhedron. Once this is done, the electrostatic repulsive force is calculated for each cation-cation interaction, given by the formula in Equation 9.1

$$F = k\frac{q_1 q_2}{r^2} \tag{9.1}$$

where k would be given by $4\pi\epsilon_0$ but for the sake of relative energies can be set to 1 and ignored as a common factor.

The total electrostatic force experienced by each cation is then calculated as well as the magnitude of the resultant force vector. These values are then combined using another parameterised weighting, between 0 and 1, and then summed for all cation sites. This total force is then used to evaluate the empirical energy.

The distinction between the two computed components (total force and net force vector) for each site is fairly subtle, with the latter being a more physically relevant measure of the magnitude of the vector sum of all cation-cation repulsions. The inclusion of the former is an attempt to incorporate a distinction between sites with strong repulsive forces that equally cancel, and weak repulsive forces that equally cancel, in most cases the weighting can be set to 1 and this distinction can be ignored.

Once the energy has been calculated for the initial structure it is then time to optimise the structure using simulated annealing. As discussed in the Computational Methods chapter (Section 6.7), this requires a change be made in the structure, the change of internal energy between the states be calculated, and this change

being accepted if it either: results in lower energy, or is probabilistically accepted based on the current simulation temperature. As further noted, one of the most important aspects to successful optimisation using simulated annealing is to set the cooling profile, to determine the simulation temperature at each time step. The code developed here allows for different cooling profiles to be set and allows for viewing the relative probabilities of a given energy difference being accepted at each time step. For optimal function of a simulated annealing algorithm, it is important to set the cooling profile such that local minima can be escaped towards the beginning of the simulation but that a minimum is reached by the end of the simulation.

The final component of the optimisation to define is how the technique determines new candidate updates to the structure and how many of these to attempt. Setting the number of changes is straightforward, and is simply limited by execution time, but can be set in conjunction with the cooling scheme to ensure a good balance of exploration and brevity occurs in the simulation; the selection of which changes to make more scope for alteration.

The most basic option is to take a Monte-Carlo approach: defining which ions are allowed to interchange with which other ions in their immediate vicinity and selecting one of these swaps each time at random. Additionally, the code also implements gliding, which allows an entire layer to systematically move ions to adjacent positions. This glide functionality is proved especially useful for optimising the relative layer stackings with significantly fewer annealing steps.

Once all the annealing steps have been taken, the final structure is returned, along with the calculated deviation from the EVP.

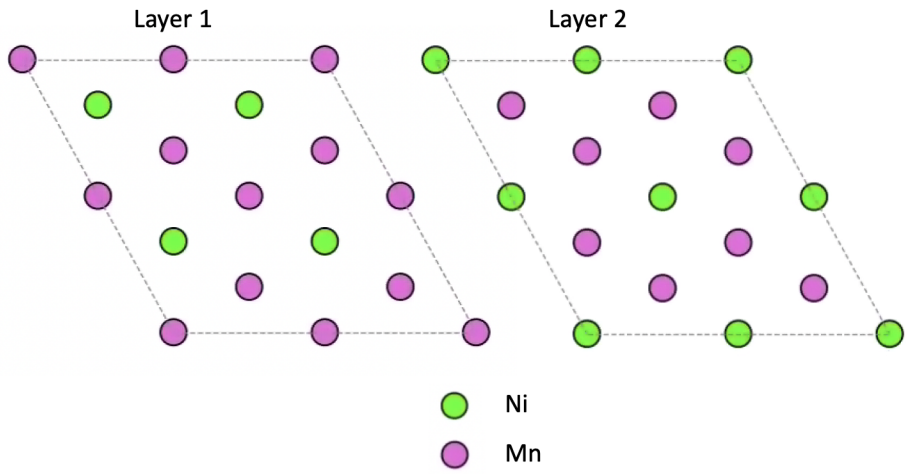


Figure 9.4: Transition Metal ordering of the two layers of the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell with N33 ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$) composition. The corresponding sodium ion orderings are discussed in Section 9.2.3.2

9.2.3 Resulting insights of intra-layer ion orderings using the empirical model

9.2.3.1 Transition Metal Ordering

To begin with, we will investigate the transition metal ion ordering. Initially, using the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell (and recalling that a $1 \times 1 \times 1$ P2 cell has ABBA stacking and therefore two transition metal layers and two sodium layers), the N33 composition and a default parameters for the empirical potential gave an annealing result with the transition metals order into the honeycomb pattern as shown in Figure 9.4 and reported in the literature[99][85].

This honeycomb pattern can be seen to persist over larger cells with the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cell shown in Figure 9.5.

Changing the transition metal composition to N30 and N25, the empirical model

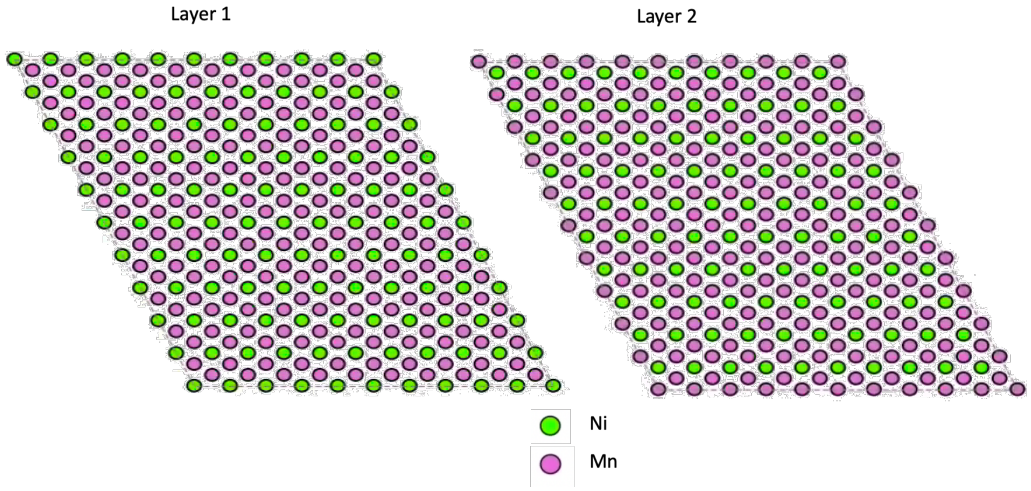


Figure 9.5: Transition Metal ordering of the two layers of the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cell with N33 ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$) composition. The corresponding sodium ion orderings are discussed in Section 9.2.3.2

predicts the honeycomb ordering to be maintained throughout the bulk of the layer, as shown in Figure 9.6. The simulations shows that as the Mn concentration increases slightly from N33 to N30, this manifests as increased amounts of Ni sites swapped to Mn in the underlying honeycomb matrix. This trend continues as the Mn concentration increases in the N25 composition. The empirical simulation predicts the 'defects' in the honeycomb to aggregate within the layers and somewhat between layers (seen more so in the N25), almost phase segregating into MnO_2 regions and Mn_2NiO_2 regions within the transition metal layers. Despite the segregation displayed in the transition metal layers, the model suggests that the Na ions do not similarly segregate, instead displaying a fairly uniform distribution across the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cell (shown in Figure 9.7) which is largely uncorrelated with the regions of high Mn concentration in the transition metal layers.

This can be seen more clearly in Figure 9.8, where a Gaussian blurring function has been applied to the point cationic charges to highlight the distribution of pos-

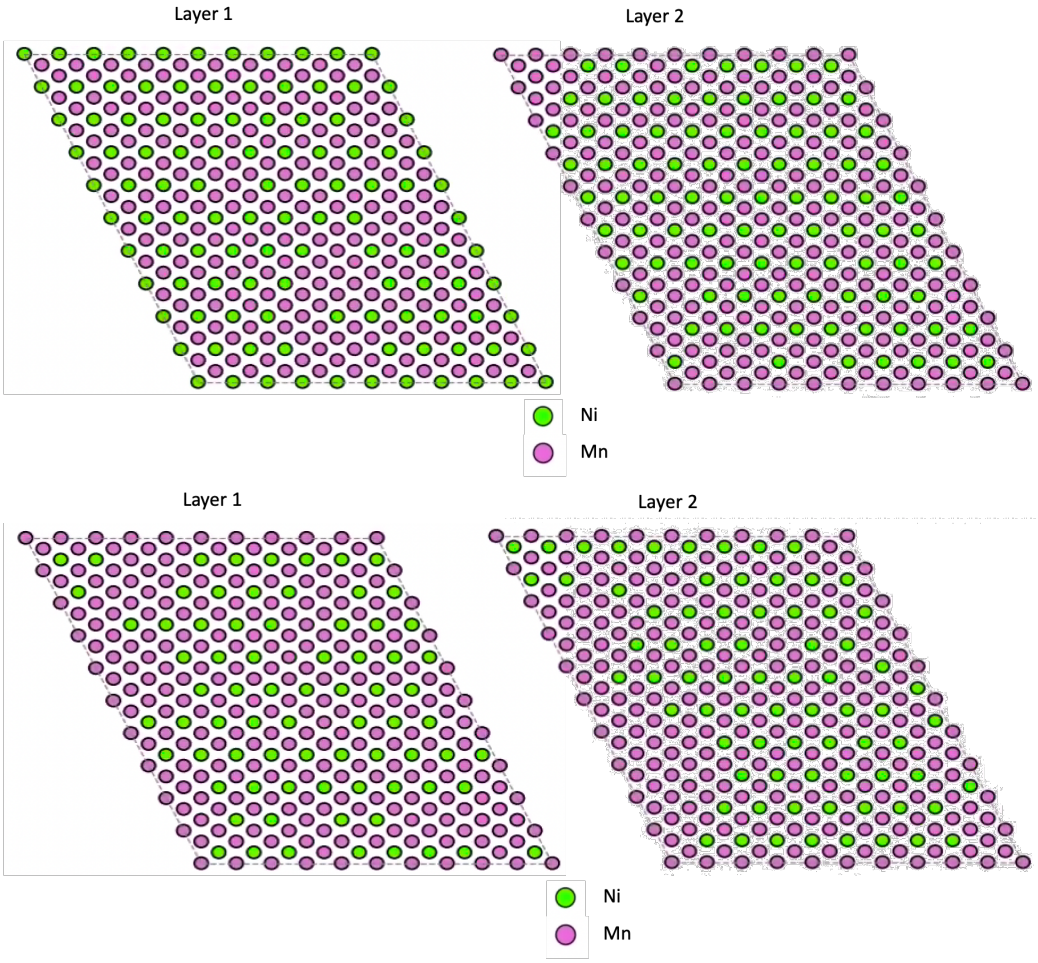


Figure 9.6: Transition Metal ordering of the two layers of the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cell with top: N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$), and bottom: N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$)compositions

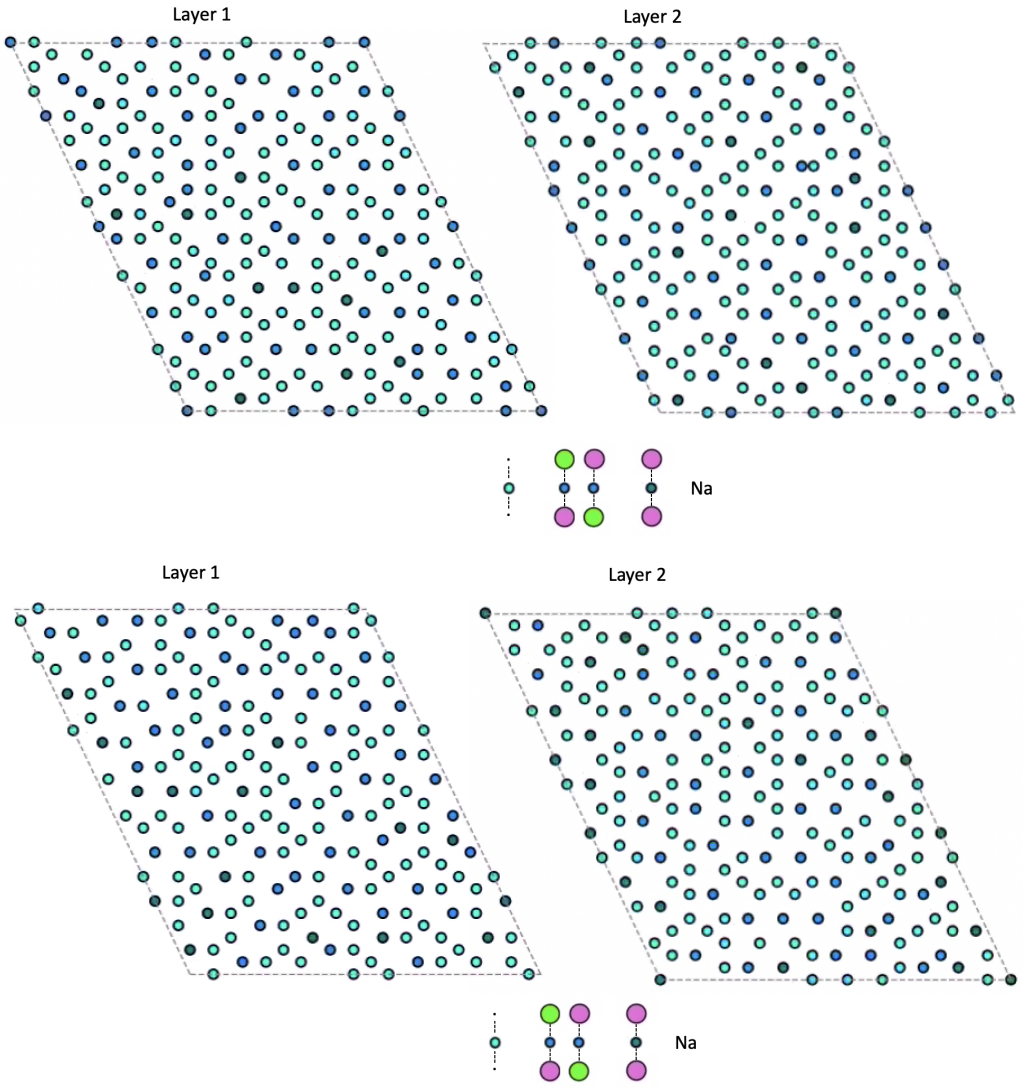


Figure 9.7: Sodium occupancies of the two layers of the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cell with top: N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$), and bottom: N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$) compositions. The different colours are explained in the key and correspond to whether the sites immediately above and below contain Ni, Mn or vacancies.

itive charge across the layers. This representation makes it clear to see where the deviations in the honeycomb ordering occur and how there is more correlation between these positions in the N25 than the N30. This representation also shows that the charge distribution in the sodium layer is somewhat noisy for both the N25 and N30 and that the regions of low charge in the sodium layer don't appear to strongly correlate with the regions of high charge in the transition metal layer, suggesting that minimising the intra-layer repulsion between Na ions is the more dominating factor. We can use these density plots to attempt to quantify the disorder of these models. By simply calculating the standard deviation of the charge densities (which are in turn a convolution of the point charge locations with a Gaussian filter), we get a crude measure of the variations in the local environments present within the model (as the disorder leads to areas of higher or lower charge density which in turn increases the standard deviation). By varying the size of the Gaussian filter, the range of the over which the disorder is quantified can be altered. The ultimate limitation of this will be the fact that it assumes a single underlying distribution, so it will give a disproportionately high measure for when two ordered domains phase segregate (such as the N25 transition metal layers).

Figure 9.9 shows the standard deviations and the charge density histograms of the respective ionic orderings. From this the quality of the standard deviation as a measure of disorder can be assessed. For the sodium layers it appears to be a good measure of the disorder, with an approximately normal distribution of the charge density across the layers. By this measure, it can be seen that the degree of disorder is largely unchanged from N30 to N25, but a slight broadening of the distributions indicates that the N25 is slightly more disordered. The change in standard deviation of the metal layers are more challenging to interpret directly as it can be seen from the histograms that from the N30 to N25 model the primary change is the growth of a second, essentially ordered, Mn segregated region.

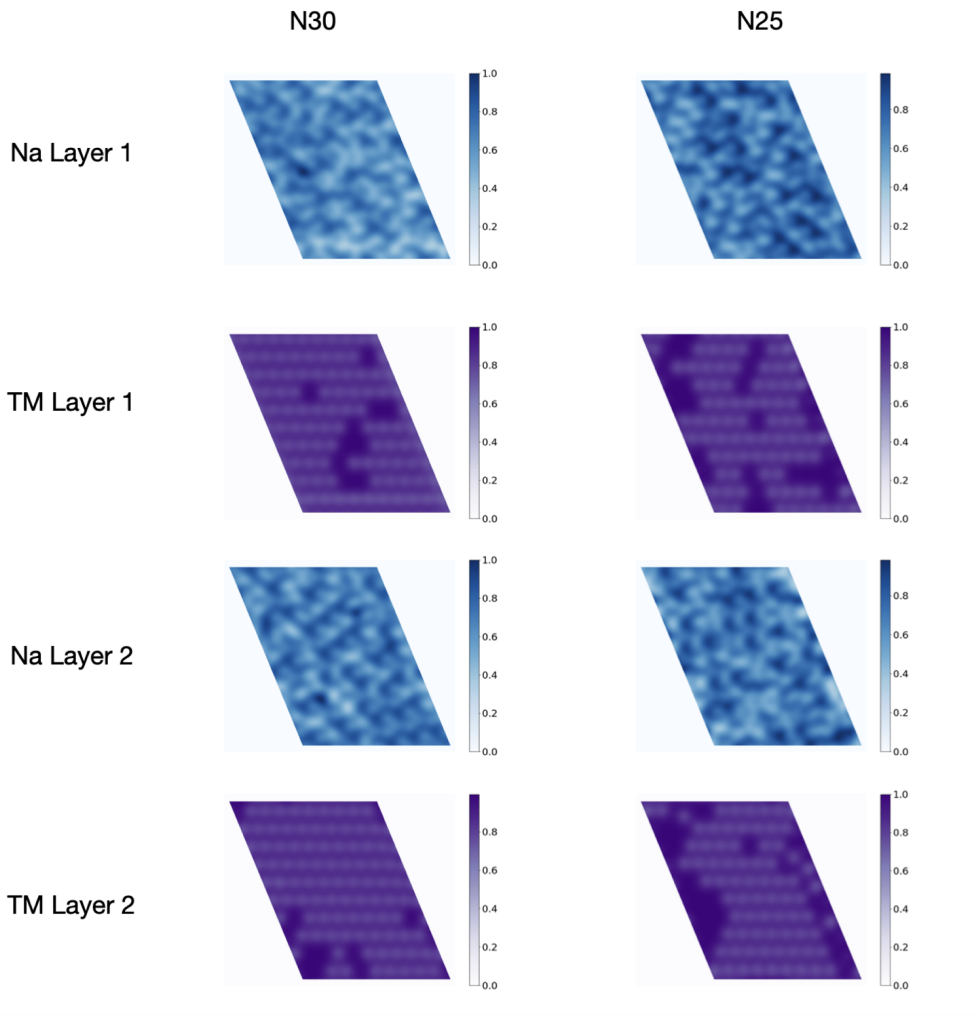


Figure 9.8: Approximate, normalised, charge distribution across the four metal layers of the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cells with N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$) and N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$) compositions

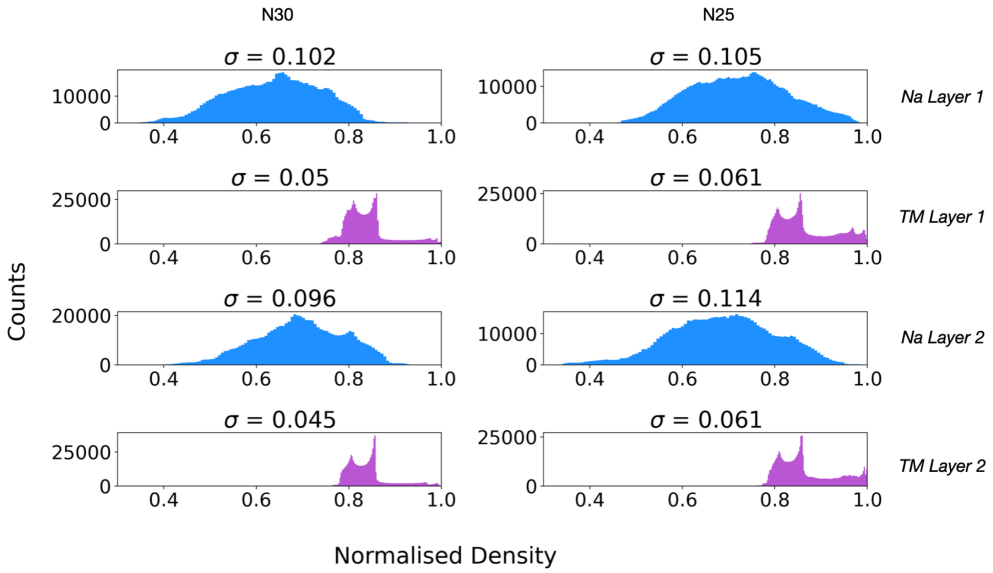


Figure 9.9: Histogram of approximate, normalised, charge distribution across the four metal layers of the $10\sqrt{3} \times 10\sqrt{3} \times 1$ super cells with N30 ($Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$) and N25 ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$) compositions

As noted in the initial development summary, however, these results are somewhat dependent on the parameterisation of the empirical potential. As such, these results should not be taken at face value as correct. They do, however, provide an initial indication of ordering which the subsequent neutron diffraction studies (Section 10.1.3.2) can build on.

9.2.3.2 Corresponding Sodium Ordering

From both the PXRD and SEND analysis of the structures, it is observed that there are various sodium ordering patterns observed, especially for the N33 composition. Of these orderings, the 2D patterns that appears to be qualitatively more prevalent is the 'snowflake' pattern shown in Figure 9.1. In addition to this, it is also common to observe patterns with lower dimensional ordering and other ordering patterns, especially for the N30 and N25 compositions. For the initial application of this approach, we will determine a potential sodium ordering scheme

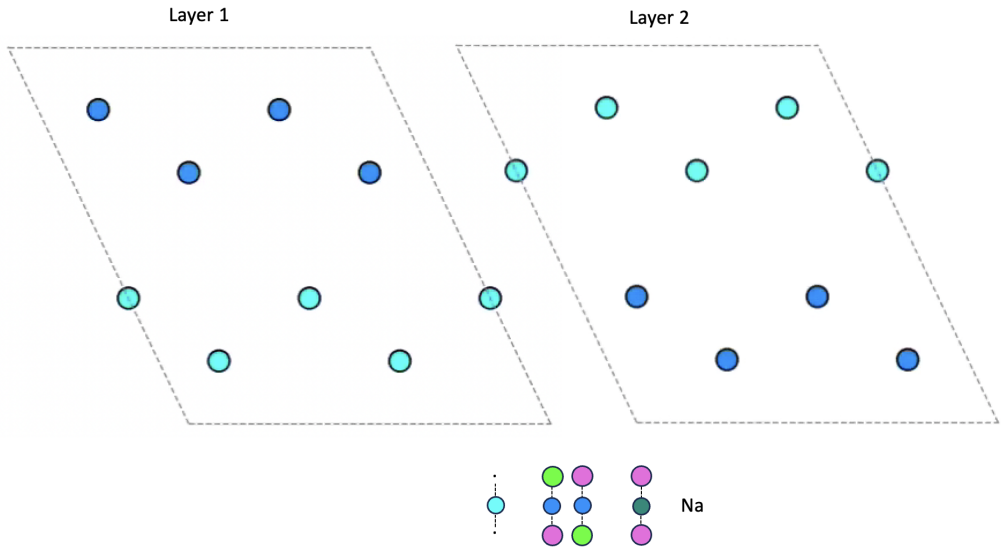


Figure 9.10: Sodium ordering of the two layers of the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell with N33 ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$) composition

which is consistent with the experimentally observed 'snowflake'.

To do this, we return to the $2\sqrt{3} \times 2\sqrt{3} \times 1$ supercell and the N33 composition. If we look at the corresponding sodium layers to the transition metal layers in Figure 9.4 we observe the ordering shown in Figure 9.10. The ordering observed is two zigzag structures, with one zigzag with the sodium located on the formally 'low energy' site (with no transition metals directly above or below). The other zigzag is then on the 'higher energy' site, in all cases this site displays neighbouring with one Mn ion and one Ni either above or below. We will refer to this as the small zig-zag structure or SZZ. This is different to the ratio and distribution of sodium sites seen in the LZZ (both have been shown again in Figure 9.11 for clarity), with the LZZ characterised by a hexagonal arrangements of the 'high energy' sodium sites and a three fold, triangular arrangement of 'low energy' sodium sites in a $1 : 2$ ratio. The LZZ sees larger separation between the high energy sites but closer proximity of the low energy sites than the corresponding SZZ ordering.

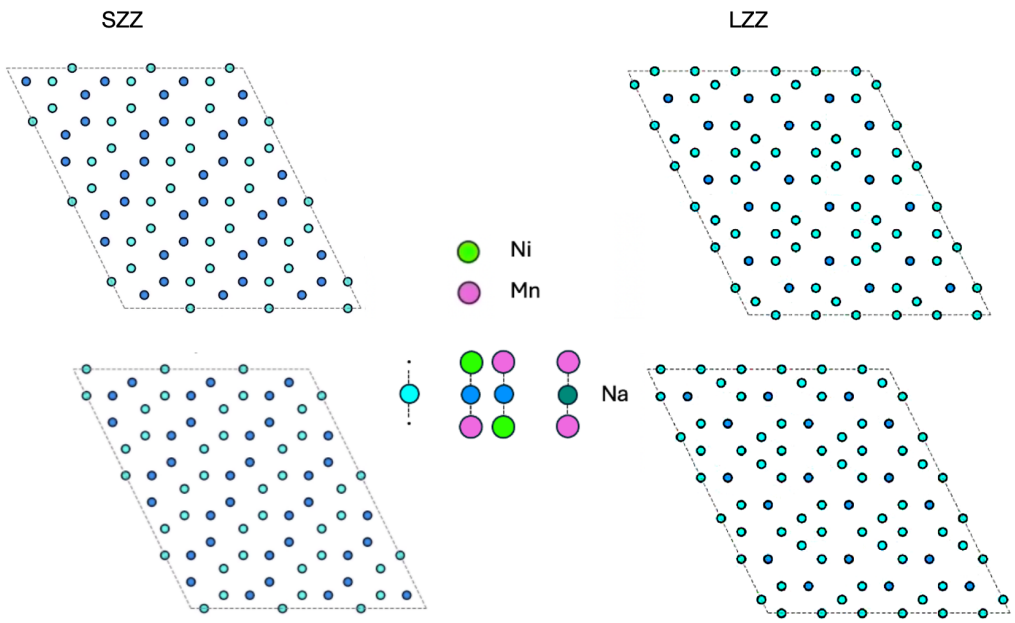


Figure 9.11: Sodium ordering of the SZZ ordering scheme compared to the LZZ ordering scheme.

Rerunning this simulation multiple times, we see that many degenerate variations of these structures can be obtained. The three primary classes of orderings can be defined in the direction of the zigzags, running horizontally (as seen in Figure 9.10), forward biases or backward biases (as seen in Figure 9.12). Additionally, the location of the zigzags can translate within the unit cell. The 'higher energy zigzag' seem to display a preference for the sodium ions occupying 'Mn – Na – Ni' sites. As the location of these sites is determined by the relative position of the honeycomb transition metal orderings above and below the layer, this suggests a strong link between the transition metal ordering and the sodium ordering.

Simulating the 2D electron diffraction pattern for this structure for just two layers of the forward biased direction gives the pattern shown on the left of Figure 9.13. Assuming an approximately equal population of all configurations can occur within the multiple sodium layers of a single structure, an averaged pattern of

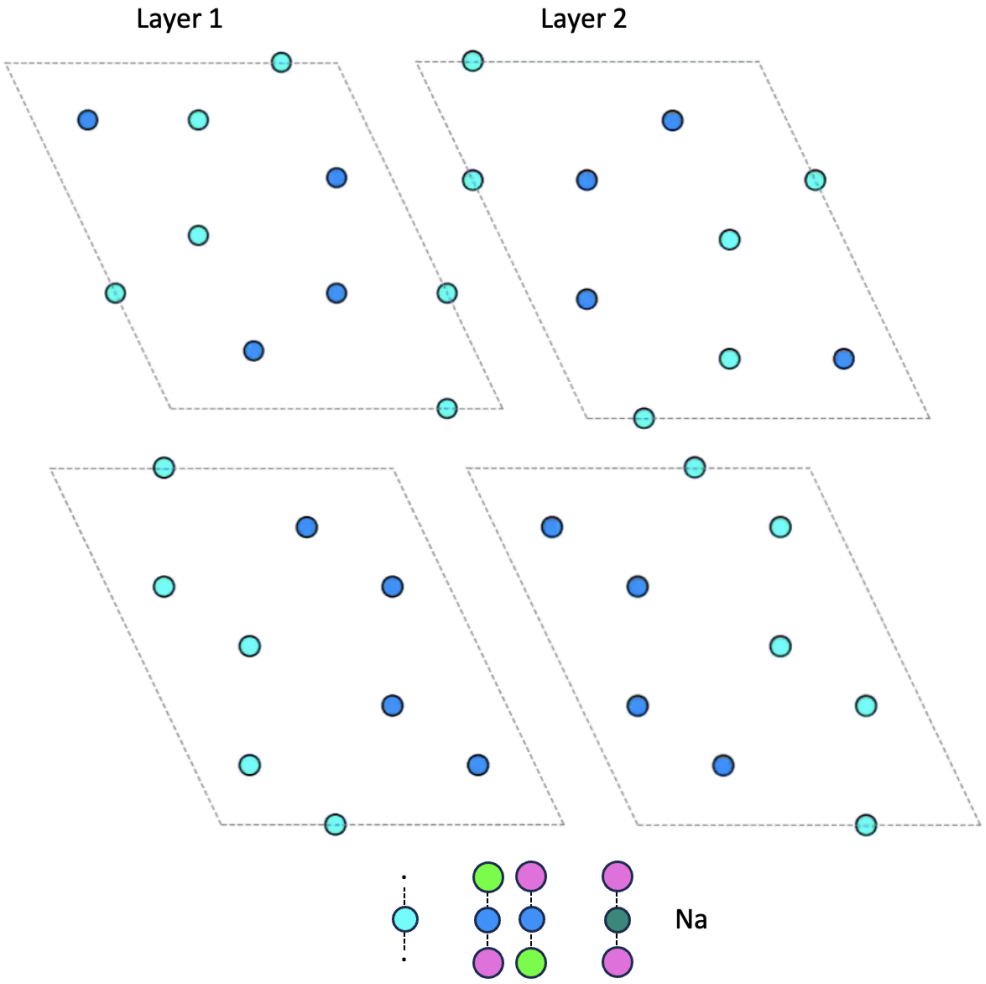


Figure 9.12: Other potential sodium ordering of the two layers of the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell with N33 composition, showing a forward biased zigzag (top) or a backwards biased zigzag (bottom)

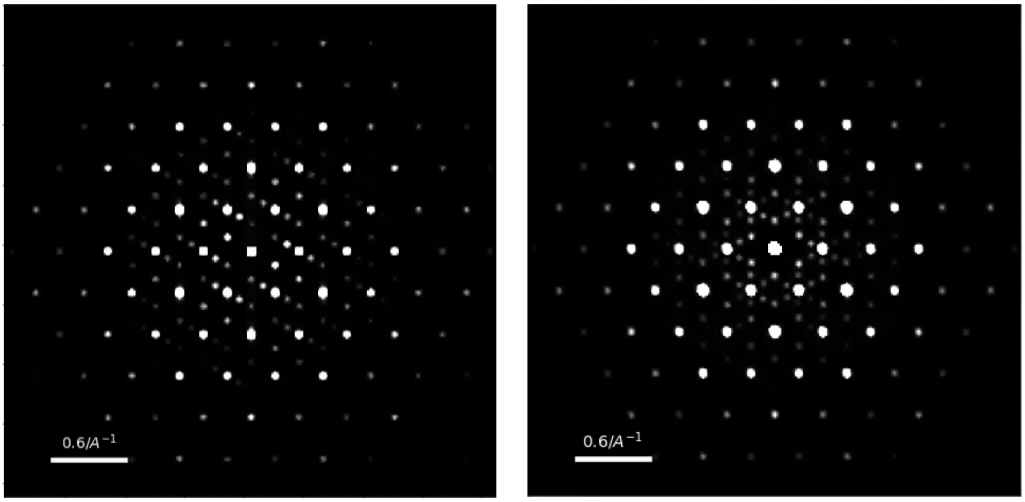


Figure 9.13: (Left) A TEM simulation of the 2D electron diffraction pattern for the forward biased structure. (Right) An TEM simulation of the 2D electron diffraction pattern averaged over the three ordering directions.

the three directions can be obtained (right of Figure 9.13). It can be seen that this closely resembles the 'snowflake' pattern observed.

9.2.4 Discussion of sodium orderings produced by the empirical potential

The main point of discussion for these results is the reliability of them, in reference to the empirical potential and the relative energetics of the systems. This is the primary motivation behind having a secondary approach which can corroborate the zigzag pattern as a plausible ordering scheme. It is worth reiterating that by altering the empirical potential, specifically the relative coloumbic screening values of the oxygen layer faces compared to edges, it is possible to output different ordering schemes. An additional caveat is that if insufficient annealing steps are provided relative to the number of atoms, alternative structures are also found.

Despite this, of all the ordering obtained by different parameterisations, the SZZ

is the only one for which a simulated TEM pattern could be produced to match the experimental data. This provides some confidence that the values for the empirical potential are suitable. Another confidence factor is the EVP adherence of these orderings. As noted earlier, the code integrates evaluation of the EVP to a given configuration, and the zigzag orderings paired with the honeycomb ordering of the transition metal layers all perfectly obey the EVP.

9.3 Development and application of structural model using a symmetry-driven approach

The second approach used to study the sodium ion ordering was an exhaustive approach generating all the symmetry unique ordering schemes for a given supercell.

9.3.1 Development

The code for this was again written in python but more readily leveraged existing packages, specifically: pymatgen[161], ASE[165], and pyprismatic[166]. The use of existing libraries allowed for faster development as many of the routines required were already available but did result in somewhat of a performance bottleneck as often the routines are written in Python/numpy which is slower than when machine-coded functions can be leveraged.

The approach taken to tackling the sodium ordering from a symmetry-driven methodology was to generate all the combination of possible sodium site occupancies within the unit cell and then remove the structures which are symmetry equivalent. While this is a slower approach, requiring a lot of computational work to identify if structures are symmetry equivalent, it is the more straight-forward

approach to implement and could be executed in a reasonable amount of time so optimising for performance was not critical. Such a method also theoretically ensures consideration of all orderings, not just the ones obtainable by tuning the empirical potential outlined above, however, some restrictions were placed on the system to prevent all possible orderings being included.

9.3.2 Summary of the implementation of the symmetry-driven approach

The first step of the approach is to generate the layered structure with all the possible sodium sites, much in the way described at the start of Section 9.2.2. To keep the number of combinations as manageable as possible it was decided to only investigate a single layer of sodium ions. This immediately poses an issue, however, due to the changing directionality of the metal oxide layers in P2, the smallest unit cell necessitates two sodium layers present. It was decided to accommodate this requirement by ensuring the second sodium layer occupancies were derived from the first, related by a 2_1 screw axis parallel to the **c** axis. This limited the number of combinations but also imposes a plausible but not necessarily correct constraint upon the possible systems. Despite the constraint, it does ensure that all configurations in the single layer are identified and investigated, which in itself would provide a useful initial insight.

Once the base structure has been generated, the next step is to then generate all the configurations. This is done by indexing the sodium sites in the first layer, then generating all the combinations with n desired indices (where n is the population of the sodium layer). The second layer is then populated using a precomputed look-up of the screw relationship between indices in the first layer and the second layer. This now gives all the possible sodium orderings and a list of the sites they

occupy.

Additionally to the constraint of the relationship between the two sodium layers, the transition metal layers also affect the symmetries of the resulting orderings. From the empirical model as well as the literature, it is assumed that the transition metal layers will display the honeycomb ordering, and that these will be stacked with no 'AA' type stacking[167][100]. This results in each sodium configuration generated being paired with the 'AB', 'AC', 'BA', 'BC', 'CA' and 'CB' transition metal layer stackings. The reason for this is the code assumes a 2_1 screw axis relation between sodium layers located in the centre of the cell, but this is not necessarily the position of the screw with relation to the transition metal layers. It is safe to assume that this approach will produce duplicate structures, therefore the next step is to remove these duplicates.

Using the list of sites, the structures can be created within the pymatgen ecosystem. Once the pymategen structures have been created, they must be compared to determine which of the structures are symmetrically identical. This could be done using a 'StructureMatcher' in built module to pymatgen. This function, however, is prohibitively slow to compare all structures for larger supercells, so two pruning strategies were implemented to reduce the total number of comparisons.

The first pruning approach is to simply identify the space groups of each structure, which can be done quickly using the pymatgen 'SymmetryAnalyzer'. This then allows the total collection of the structures to be broken down into smaller groups which share the same space group and only compared to other structures with the same space group, dramatically reducing comparisons.

The second pruning approach is to quantify structures in terms of their mean n nearest neighbour distances.

This can be done quickly using the a built in method to the pymatgen Structure

class. This pruning step requires a balance of time and accuracy, as it is faster to calculate the distance to fewer neighbours, however, more structures will have the same average distance when taken over fewer neighbours. In the extreme where infinite nearest neighbours are calculated, only symmetry identical structures would have the same mean nearest neighbour distances. Eight neighbours seems to be a good compromise and allows breaking the space group groupings down into further subgroups which all share the same average distance for each ions 8 nearest neighbours.

With the final subgroups created, the 'StructureMatcher' can be used to iterate over all the structures in a subgroup and remove duplicate, symmetry, equivalent structures. After this step, a unique set of possible sodium orderings for a single sodium layer are returned.

These possible structures are then compared against experimental results, to help indicate which are most likely to be the observed structure. This is done comparing simulated SEND patterns to the experimental SEND pattern and from Figure 9.1. The XRD could also be used as a comparison metric, but due to the nature of powder diffraction, knowledge of the periodicity in the c-direction would play a major role, so this evaluation left for a separate the stacking investigation.

9.3.2.1 Experimental Comparison - SEND

The experimental comparison for the SEND data was done using pyprismatic, a multi-slice electron diffraction simulation package. The approach for this was as follows. For each structure, an atomic model was generated using the atomic simulation environment, ASE. This model is then passed to pyprismatic along with the beam conditions of the experimental SEND data. The simulated SEND data is then compared to the experimental pattern and a similarity metric is used

to identify the best matches.

9.4 Resulting insights of intra-layer ion orderings using the symmetry-driven approach

9.4.1 $2\sqrt{3} \times 2\sqrt{3} \times 1$ Simulation

The simulation of the $2\sqrt{3} \times 2\sqrt{3} \times 1$ supercell was done using the summary outlined above, with the structure generation step initially producing 30975 symmetry unique sodium ordered structures. For each of these structures the six transition metal orderings were then computed resulting in 185850 structures. Each of these structures were then sorted into space groups and compared for symmetry equivalence. This resulted in the distribution of structures given in Table 9.1.

Table 9.1: Number of structures generated within each space group for the $2\sqrt{3} \times 2\sqrt{3} \times 1$ super cell

Space Group Symbol	Number of Structures
$P1$	121804*
$P2_1$	30457
$P2$	14
$C2$	470
$C222$	470
$P2_12_12_1$	6
$P3$	29
$P6_3$	29
$P6_322$	5
$P222_1$	14
$P312$	5

* Not refined for duplicate structures

From Table 9.1, it can be seen that the majority of structures produced are low symmetry structures, which are unlikely to be low energy ordering solutions. As-

suming this to be the case, the $P1$ and $P2_1$ structures were omitted from the rest of the analysis. The $P2$ structures were retained despite also being a low symmetry space group due to the more manageable group size. The remaining 1042 structures were investigated using SEND simulation, XRD simulation and energy quantification.

9.4.1.1 SEND Comparison

The 1042 structures $2\sqrt{3}\times2\sqrt{3}\times1$ were all simulated in pyprismatic to generate a SEND pattern. The resulting patterns were compared to the observed 'snowflake' pattern using two metrics. First, the simulated patterns were evaluated by summing the intensity of the characteristic absences of the 'snowflake' pattern. The patterns with intensity below a certain threshold were then evaluated based on the intensity of characteristic snowflake reflections. This was done on both the as-simulated pattern, and the rotationally averaged patterns. The rotationally averaged patterns were generated under the same assumption as in Section 9.2: as some of the patterns display ordering in one of three degenerate directions, an averaged pattern of these three direction may also be of merit to investigate, effectively including a small subset of multi-layered orderings in this study of single layer orderings.

Both the as-simulated and rotationally averaged patterns returned the same ordering of fit-quality for the best few structures. The four structures which produced the best fit were all from the $P2_12_12_1$ and owing to their essentially identical superstructure Bragg reflection locations. The four structures are shown in Figure 9.14 and 9.15. Of these, the fourth structure was the SZZ ordering as found by the empirical model. This is shown in Figure 9.16, along with the 3-fold rotationally averaged pattern.

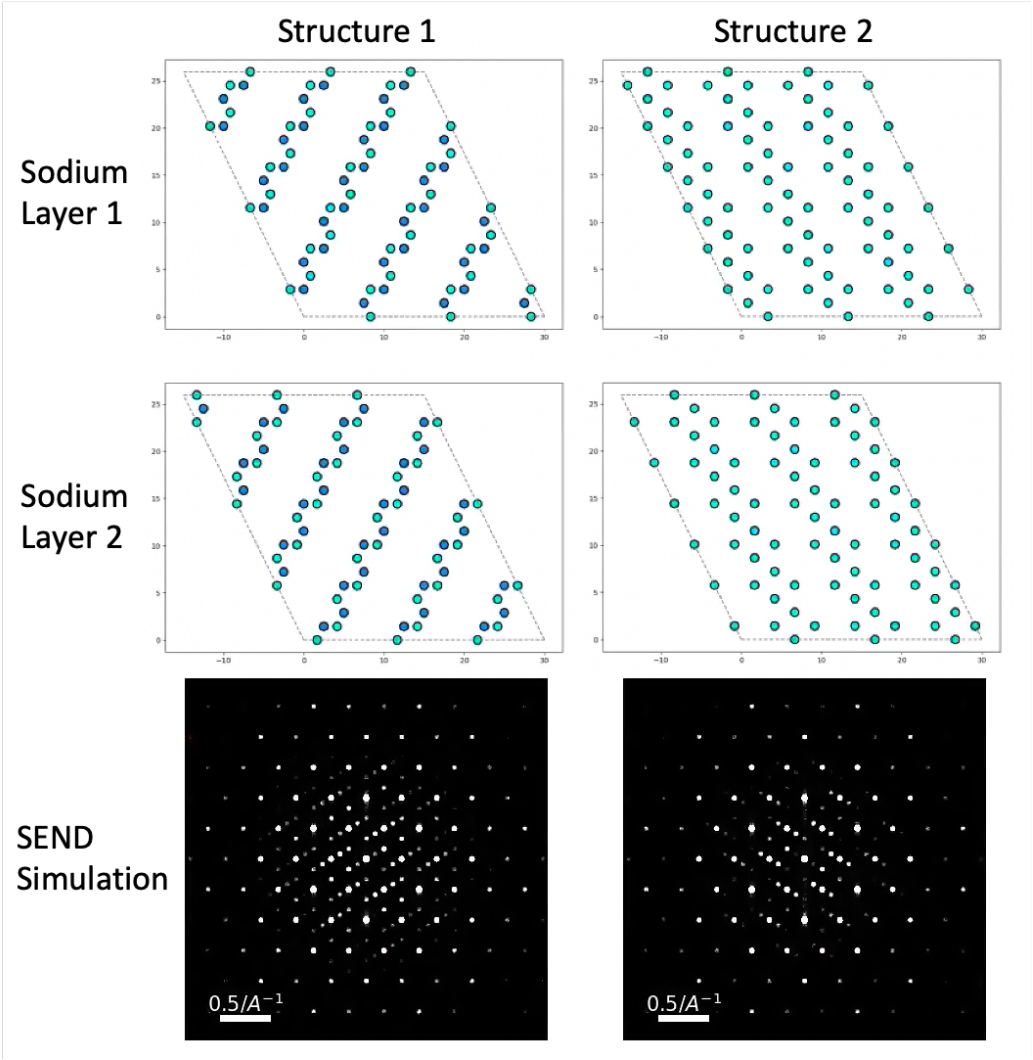


Figure 9.14: Summary of the four best fitting structures to experimental SEND pattern. Each structure is represented by a column containing images of the two sodium layers and the SEND simulation.

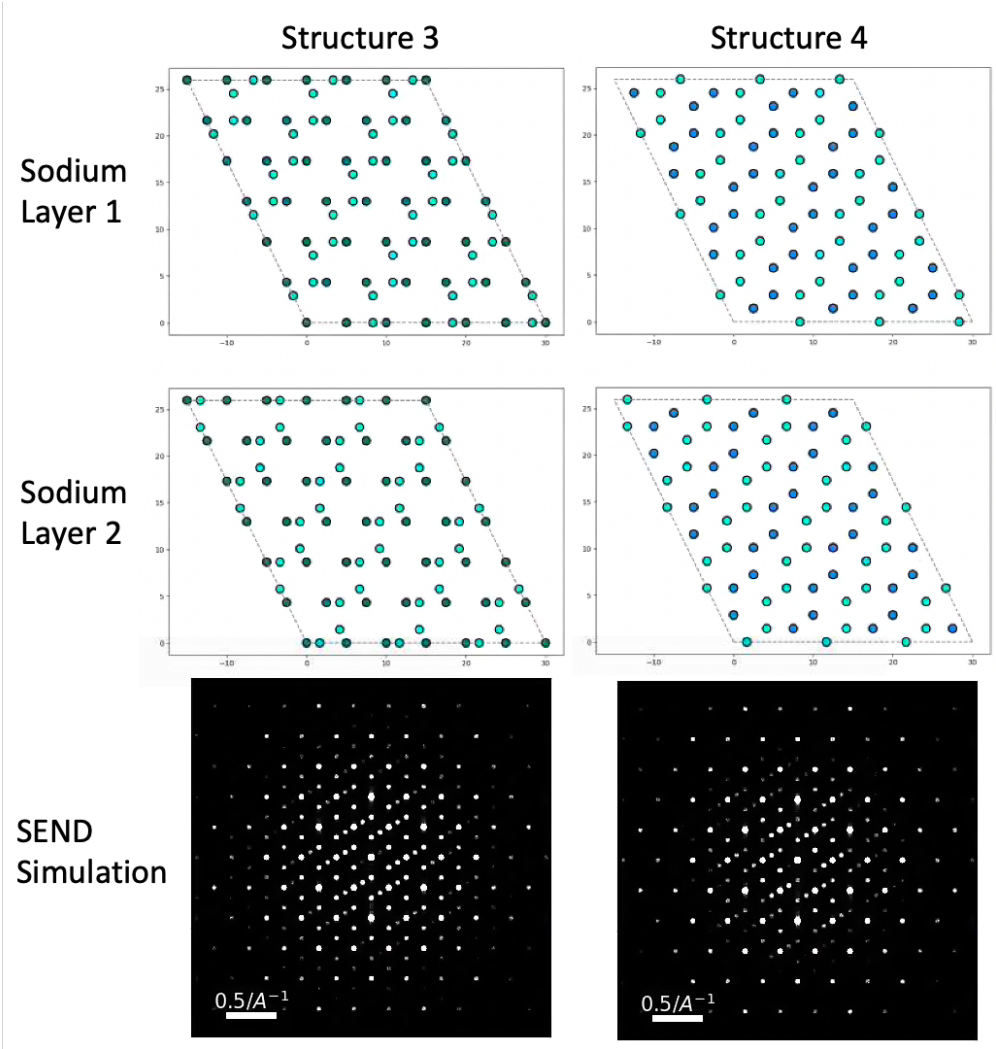


Figure 9.15: Summary of the four best fitting structures to experimental SEND pattern. Each structure is represented by a column containing images of the two sodium layers and the SEND simulation.

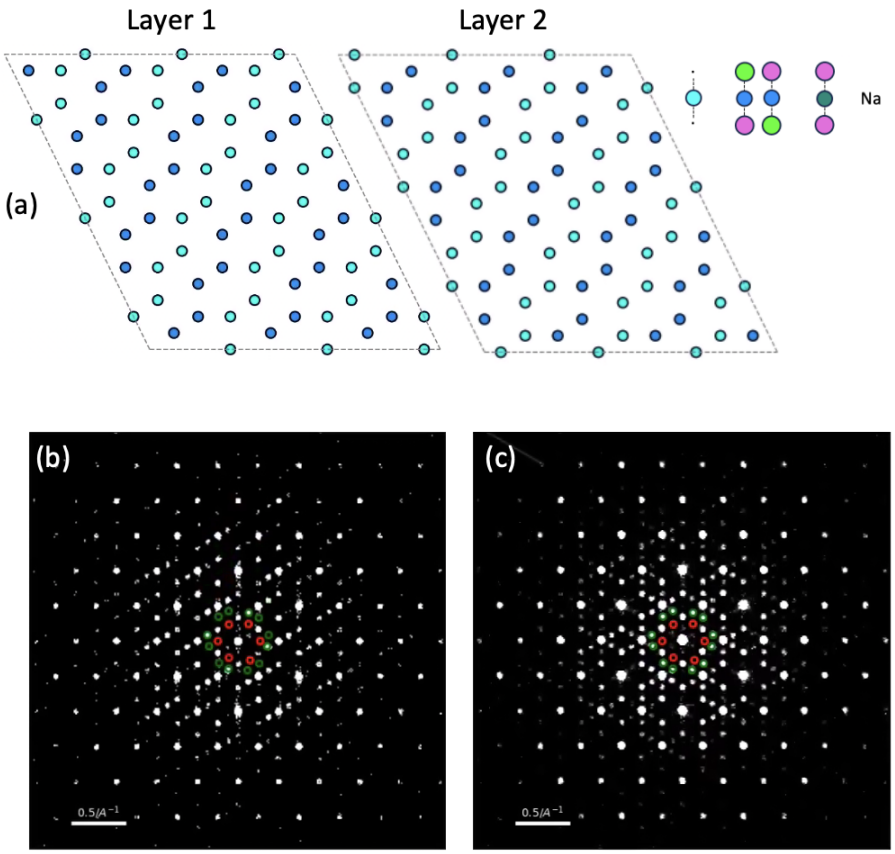


Figure 9.16: Summary of the zigzag structure, fitted to experimental SEND pattern. (a) Sodium ordering in the two layers of the structure. (b) Simulated summed SEND pattern for the structure with the characteristic reflections shown in green and the characteristic absences shown in red. (b) Simulated summed, rotationally averaged, SEND pattern for the structure with the characteristic reflections shown in green and the characteristic absences shown in red

It is worth noting that the 3-fold rotationally averaged pattern of all four of the identified structures would produce very similar patterns to that seen in Figure 9.16, varying solely by the reflection intensities of given peaks, but all showing the same characteristic extinction (indicated by the red peaks). As the intensities of the experimental patterns are likely to depend on dynamical effects, this extinction comparison is the extent of the discrimination between likely ordering schemes that can be confidently determined from the SEND data alone. It can, however, be seen visually in Figures 9.14 and 9.15 that this fourth structure has the most evenly distributed sodium ions within the layers. This is also shown in Figure 9.17 by the mean Na – Na separation calculated for different number of nearest neighbours, where Structure 4 has the highest or equal highest mean distance for up to 8 nearest neighbours. This more spaced out structure so it is likely to correspond to it being energetically favourable over the other structures due to coloumbic considerations.

This evaluation of energetic was further supported by implementing a Machine Learning Force Field based analysis of structural energies using 'm3gnet'[168], with the predicted energies from the model shown in Figure 9.18. This predicts Structure 4 to be ranked 21st lowest energy of the 1042 structures, within $0.32\,\mathrm{eV}$ of the lowest energy prediction.

9.4.2 Conclusion of combined results of two approaches for investigating the intra-layer ion orderings

The result of this empirical and symmetry-driven approaches to investigating the sodium ordering within a layer seems to suggest that the SZZ should be considered as a potential super-structure ordering present within these systems. As indicated by the predicted structural energy, in real systems at room temperature, there will

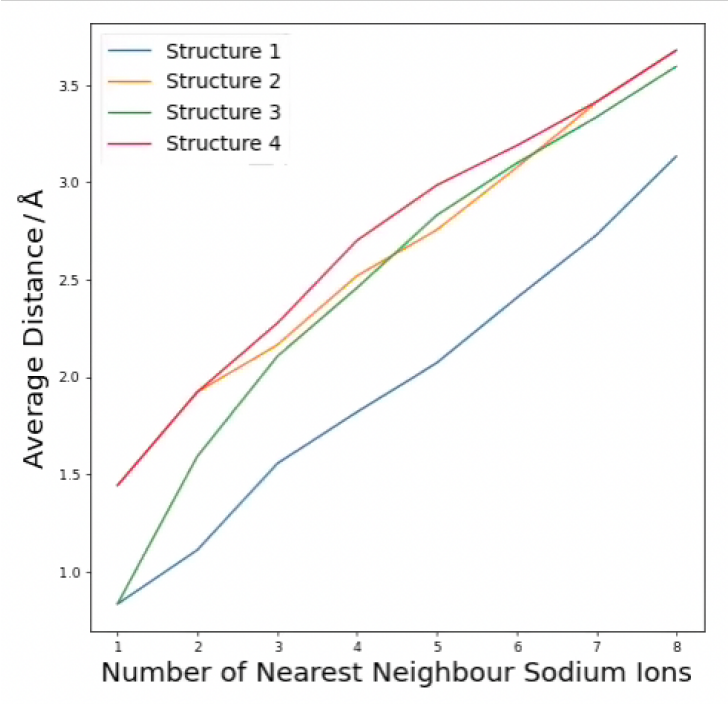


Figure 9.17: Mean distance between sodium ions and their n nearest neighbours shown for the four best fitting structures produced by the symmetry driven approach.

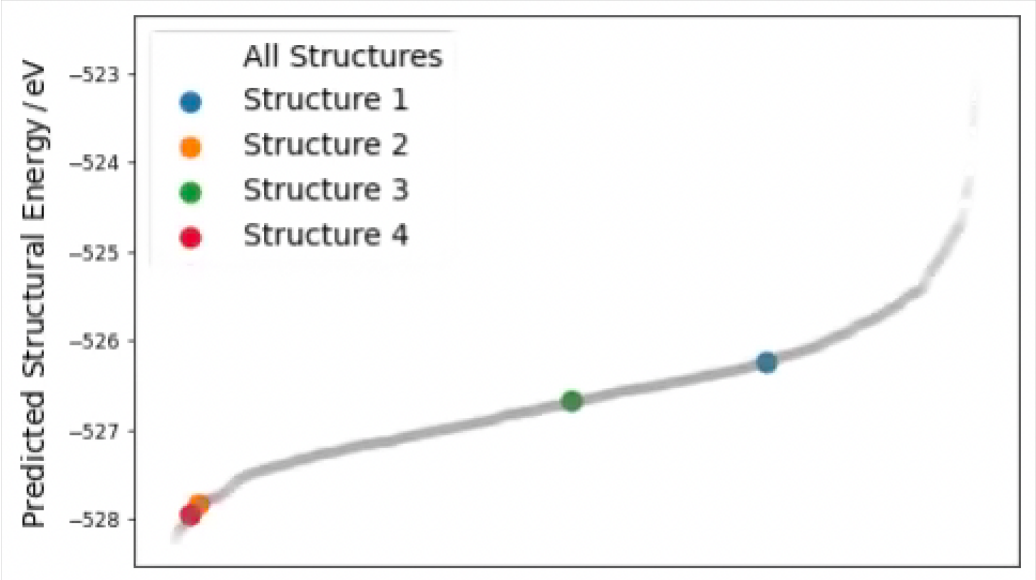


Figure 9.18: Structural Energy predictions of all the $2\sqrt{3} \times 2\sqrt{3} \times 1$ supercell structures, with the four structures best fitting the SEND data highlighted.

only be a small thermodynamic difference between any of the given structures, despite this, the presence of direct experimental evidence of this SZZ ordering at room temperature, in contrast to the absence of equivalent evidence for the LZZ, should merit a reconsideration of the prevalence with which the LZZ structure is ascribed.

10 | Investigation of Inter-layer Ion Orderings in Sodium Ion Cathode Materials

Having considered the ordering of ions primarily within individual layers, the logical extension is to ask how this ion ordering are dependent on the other layers, and so investigate inter-layer ordering. To do this, we can define models building up layers of the intra-layer ordering obtained from the results of the previous techniques. We then model the system with the position of the ions in of one layer being probabilistically dependent on the position in the neighbouring layers.

To investigate the inter-layer ion ordering in the samples, two modelling approaches were ultimately used, investigating three orderings across the structure: the oxygen ion ordering; the transition metal ion ordering; and the sodium ion ordering. Both approaches were to produce structural models by the probabilistic addition of new layers to existing layers, with the only difference being the required complexity. The first approach used was to produce structural models purely in TOPAS Academic, and this approach allowed for investigation of the oxygen ion ordering and the transition metal ion orderings. The second approach, for studying the inter-layer sodium ion ordering, required more flexibility in the layer populations and inter-dependencies than the TOPAS software natively accommodates, resulting in a different approach where the structural model was produced in python and then given to TOPAS to perform Rietveld analysis of the experimental data.

10.1 Oxygen and Transition Metal Ordering

The first approach, using TOPAS Academic, aimed to investigate how the oxygen ions and the transition metal ions order in the three cathode material stoichiometries: N33 , N30 and N25 ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$, $Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$, and $Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$).

10.1.1 Oxygen Ion Ordering

As discussed in Chapter 4, these layered structures can be considered in terms of layers of oxide ions, with alternating intermediate layers of sodium ions and transition metal ions. The positions of the oxide ions within each layer, and the periodicity of these positions between layers, is the primary factor which determines the classification of the structure type.

Recalling that the stoichiometries in question are targeting the P2 structure type, one would expect the oxygen ions to 'order' with the ABBA repeating scheme as derived in Section 4.2. It was also noted within that chapter, however, that other orderings are possible, and that the overall structure formally denoted as P2 may contain some degree of deviation from the strict ABBA scheme. Another ordering scheme outlined was the P3, following an ABBCCA scheme. It can be seen that starting from a stack of oxygen ions ABB, the next oxygen layer will determine whether the local stacking regime is P2-type or P3-type by whether the next layer is A or C.

As such, the structure is modelled within this approach by adding new oxygen layers based upon the current local stacking regime and two set probabilities P_a and P_b.

The construction works as follows: If the previous three layers are ABB, and the

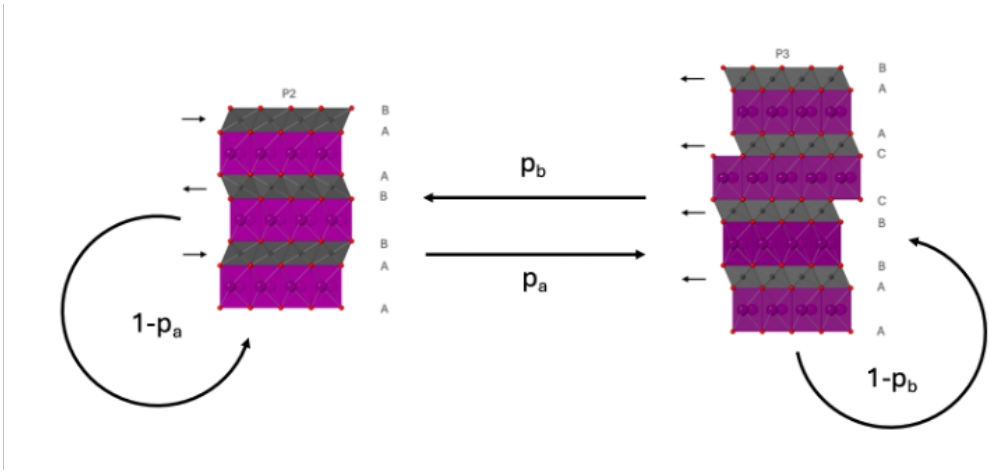


Figure 10.1: Oxygen layer orderings for both the P2 and P3-type stacking regimes and the transition probabilities of transferring between the two types of regime.

structure is currently within in a P2-type regime the next oxygen layer added will continue that scheme and so will be an A. This is unless a stacking fault occurs, determined by the probability P_a, in which case the other oxygen layer position, C, is added instead, which additionally transforms the current stacking into the P3-type regime.

From here, with the previous three layers, BBC, the next oxygen layer must be C regardless of whether the current regime is P2 or P3-type in order to maintain the prismatic sites for the sodium interstitials. The subsequent layer, however, has another choice. It is built upon the layers BCC and is in the P3-type regime. It will therefore continue to add the next oxygen layer to continue the P3-type regime (to form BCCA) unless a fault occurs with probability P_b (forming BCCB). This is shown schematically in Figure 10.1. By tracking the current regime and using these transition probabilities it is possible to generate models with different compositions of P2 and P3-type local stackings.

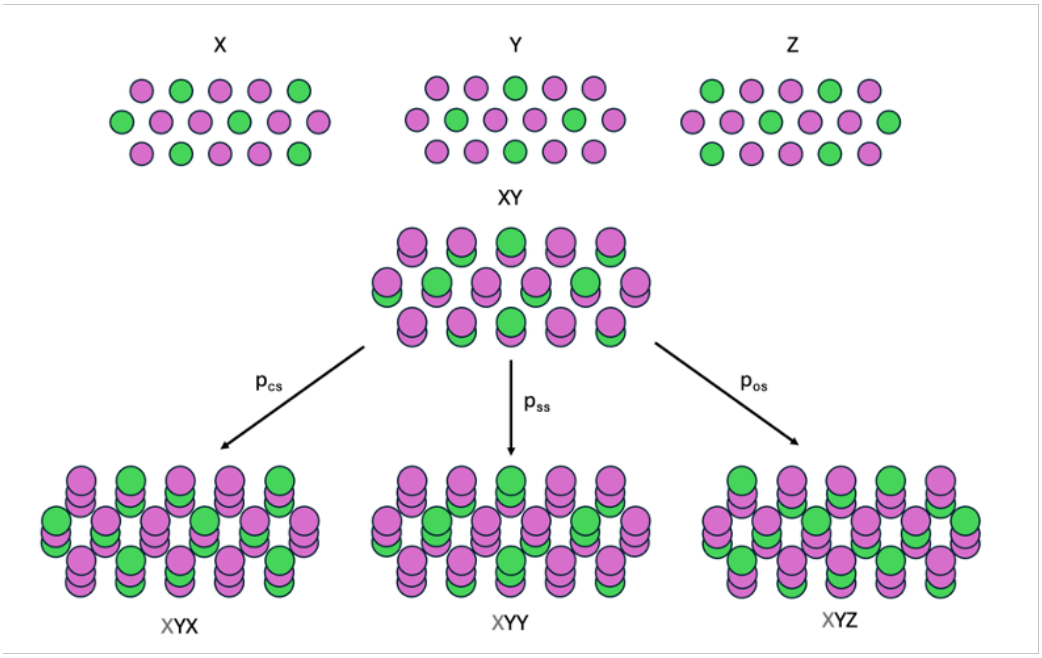


Figure 10.2: Three positional configurations of the honeycomb transition metal ion ordering of Ni (green) and Mn (purple). The transition probabilities for stacking multiple layers on top of each other are also shown.

10.1.2 Metal Faulting

Analogous to the oxygen ion positions, the transition metal ions also form layers that occupy the positions A, B and C. If the oxygen ions above and below the transition metal are in positions A and B, then the transition metals will be in position C. Additionally, when considering the honeycomb ordering of the N33 composition, the layer can also occupy three unique positional configurations, here labelled: X; Y; and Z, and shown in Figure 10.2.

Taking the P2-type stacking regime ABBA, the first transition metal layer has oxide layers A and B and so the transition metals occupy the C positions. Similarly the second transition metal layer is surrounded by oxide layers B and A so also has transition metals in the C positions. In such a stacking scheme the transition metal ions will all occupy the same atomic columns and so, if the honeycomb

ordering is present, the relative configurations of these individual layers as X, Y and Z will result in different structures. Due to this, the TOPAS model developed also incorporates the probabilities P_{cs}, P_{os}, and P_{ss} to control the relationship between the Ni positions in neighbouring layers as shown in Figure 10.2. P_{cs} is the probability of 'consistent stacking' and continues with the same alternating regime it currently adopts (in the example of the Figure, this is XY). P_{os} is the 'other stacking, where it swaps from the alternating stacking it is currently adopting, to the other alternating stacking (which in this case would be YZ). The final P_{ss} is the 'same stacking' where the Ni ions are stacking within the same columns. General wisdom is that this stacking does not occur and so $P_{ss} = 0$, however the option is included within the model for completeness. In the situation where $P_{ss} \neq 0$ the following transitions are determined by the probabilities P_{ss} to continue with another 'same stacking' or $(1 - P_{ss})/2$ to adopt either of the alternating stackings.

Taking instead a P3-type stacking ABBC, it can be determined that the transition metals will be in positions C and A respectively. As a result, the configurations of any ordering of transition metals within layer C will be independent of the orderings in the other layer A as they do not share atomic columns. As a result the model selects a random configuration for each honeycomb ordering when in the P3 regime. A more subtle implication of this is that the presence of a single P3-type fault layer can disrupt an established XYXY... transition metal ordering between P2-type layers, even if $P_{cs} = 1$.

For the N30 and N25 compositions, and as shown using the empirical model, a perfect periodic ordering of transition metals does not occur. Instead, the empirical potential suggests that the honeycomb ordering is maintained where it can be and this is accommodated by a fraction of the Ni sites being occupied by Mn to compensate. The transition metal ordering for the different Ni and Mn stoichiometries is handled within this model by two values S and D. The first value

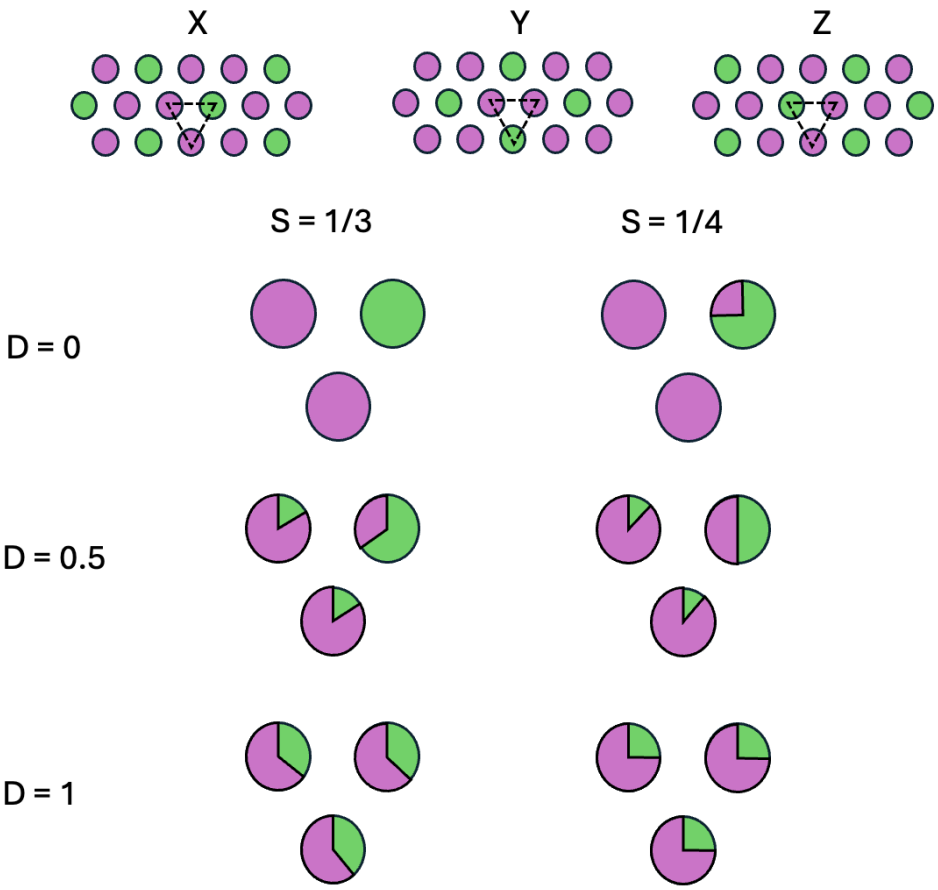


Figure 10.3: Schematic showing how the intra-layer transition metal ordering is treated for different stoichiometries of Ni (green) and Mn (purple).

S determines the overall proportion of Ni present in the transition metal layer and the second value D determines how ordered these Ni are into an approximately honeycomb arrangement. Some example values of S and D are shown in Figure 10.3.

The use of these parameters allows for the N30 and N25 stoichiometries to express consistency of inter-layer ordering if D is low, as the Ni ions are largely localised in the same periodic columns. The alternative situation, where this long range

ordering is not present, can also be modelled by setting D to be closer to 1.

10.1.3 Ionic orderings fit by TOPAS stacking model

10.1.3.1 Powder X-ray Diffraction Stacking Model

As explained in Chapter 5 the x-ray scattering factors of Ni and Mn are very similar, providing little contrast between the two in PXRD, and resulting in little information being decipherable about the relative inter-layer transition metal orderings. The oxygen ion ordering can however, be investigated, due to the effect that the changes have upon local environments and periodicity along the stack.

In the extreme case, where all the oxygen ion stacking is in the P2-type regime, the symmetry of the structure is $P6_3/mmc$. Alternatively, in the other extreme, where all oxygen ion stacking is in the P3-type regime, the symmetry of the structure is $R-3m$. These different structural symmetries give rise to different intensities at different Bragg reflections. If two of these pure P2 and pure P3 phases were present within the same sample the diffraction profiles would show peaks corresponding to a linear combination of both these base structures. In the intermediate case where both stacking regimes occur within a single structure it is less straight forward to qualitatively predict the effect of the probabilities P_a and P_b on the final pattern. It would be expected that the loss of periodicity along the stacking direction would result in a degree of broadening to a subset of peaks. From the experimental pattern it can be observed that characteristic broadening of the (10l) peaks are present that have been attributed to single layer P3 stacking faults by Pfeiffer et al.[102].

To investigate the nature of the oxide layer stackings in the N33, N30 and N25 samples, synchotron PXRD diffraction patterns were collected, a structural model was generated and refined for a grid of P_a and P_b values and the resulting Good-

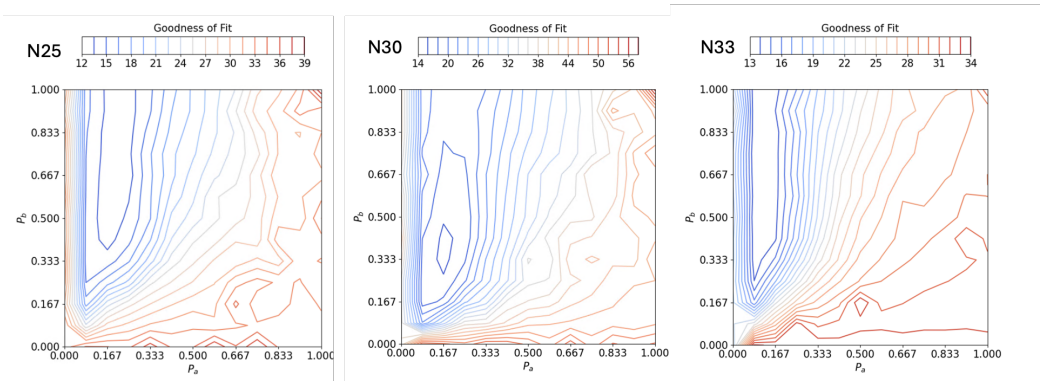


Figure 10.4: Contours of the Goodness of Fit surface across the different parameterisations, P_a and P_b, of a stacking model.

ness of Fit metrics were used to create contour plots for the parameter space. These contour plots are shown in Figure 10.4.

These contour plots can be used to guide further search and indicate the ideal parameterisation to model the experimental structure. As the models are statistical, this surface minima approach is more robust to variation in fit metrics that result from different iterations of the stack generations. Additionally these parameter space samples can be used to visualise the effect of each stacking parameter on the quality of the experimental fit as shown by the modelling of two features in Figures 10.5, 10.6, and 10.7. The ranges chosen to be displayed in these plots are somewhat arbitrary, with the $19° - 20°$ range chosen as it shows significant variation across the parameter space. The $25° - 26°$ range, however, was chosen as it shows the sensitivity of the fitting to the parameter change for one of the 'volcano peaks' (specifically the <104> reflections). It is also worth making explicitly clear that the refinements are performed on the full diffraction pattern giving the GoF contours seen in Figure 10.4 and the regions in Figures 10.5, 10.6, and 10.7 are taken from these refinements. From the contours and the refinements, it can be seen that the fit quality is more sensitive to changes in P_a than P_b, characterised by the broad minima in the vertical dimension. The broader the minima in P_b,

the more it suggests that the P2-type stacking is dominant for that sample. This is because a change in the transition probabilities from P3 to P2-type stacking (P_b) will have a small effect on the overall model, if the P3-type stacking is rare.

The best parameterisations identified from the contour plots are given in Table 10.2 and the final fits are given in Figure 10.8, with the corresponding fit metrics given in Table 10.1. A comparison to a completely ordered, standard model is given in the Appendix (Figure A.2 and Table A.1).

Table 10.1: N25, N30, and N33 PXRD fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF)

Sample	Target Stoichiometry	Fit Quality Metrics		
		R_{exp}	R_{wp}	GoF
N25	$Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$	0.446	4.176	9.369
N30	$Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$	0.420	5.582	9.173
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	0.568	6.173	10.865

Table 10.2: N25, N30, and N33 PXRD fit parameterisations

Sample	Target Stoichiometry	Parameter	
		P_a	P_b
N25	$Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$	0.20	0.63
N30	$Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$	0.18	0.36
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	0.11	0.88

It can be seen from the contour plots in Figure 10.4 that the goodness of fit is rather insensitive to the P_b parameter and a factor in this is that the length of the faulting P3 stack has relatively minor changes in the XRD profile as it disrupts the periodicity of the P2 phase similarly whether it is a single layer fault or a multi-layer fault. Before interpreting the implication of these parameters P_a and P_b, it is worth investigating the neutron diffraction where greater differences between P2 stacking and P3 stacking are likely to be seen as a result of the stronger

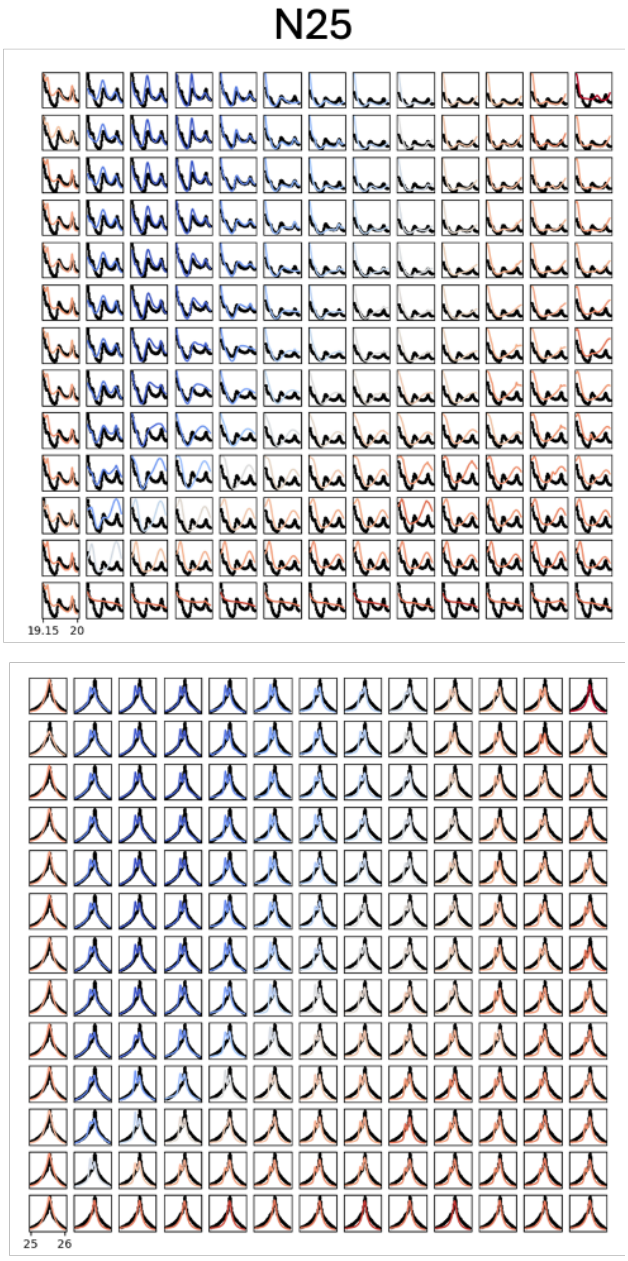


Figure 10.5: The fits (coloured) to the experimental data (black) of the N25 stoichiometry ($Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$), provided by different parameterisations, P_a and P_b, of the stacking model for two characteristic 2θ regions (indicated on the bottom left pattern). The P_a values increase along the horizontal of the grid and the P_b value increase along the vertical, with the span of the grid matching the axes of the plots in Figure 10.4.

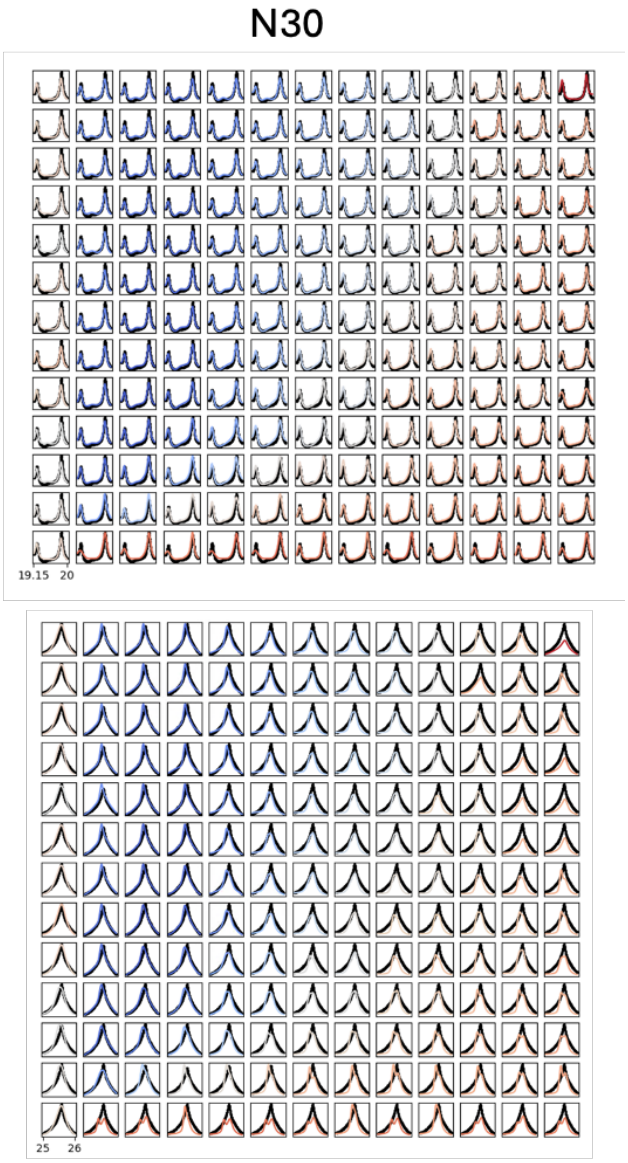


Figure 10.6: The fits (coloured) to the experimental data (black) of the N30 stoichiometry $Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$, provided by different parameterisations, P_a and P_b, of the stacking model for two characteristic 2θ regions (indicated on the bottom left pattern)



Figure 10.7: The fits (coloured) to the experimental data (black) of the N33 stoichiometry ($Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$), provided by different parameterisations, P_a and P_b, of the stacking model for two characteristic 2θ regions (indicated on the bottom left pattern)

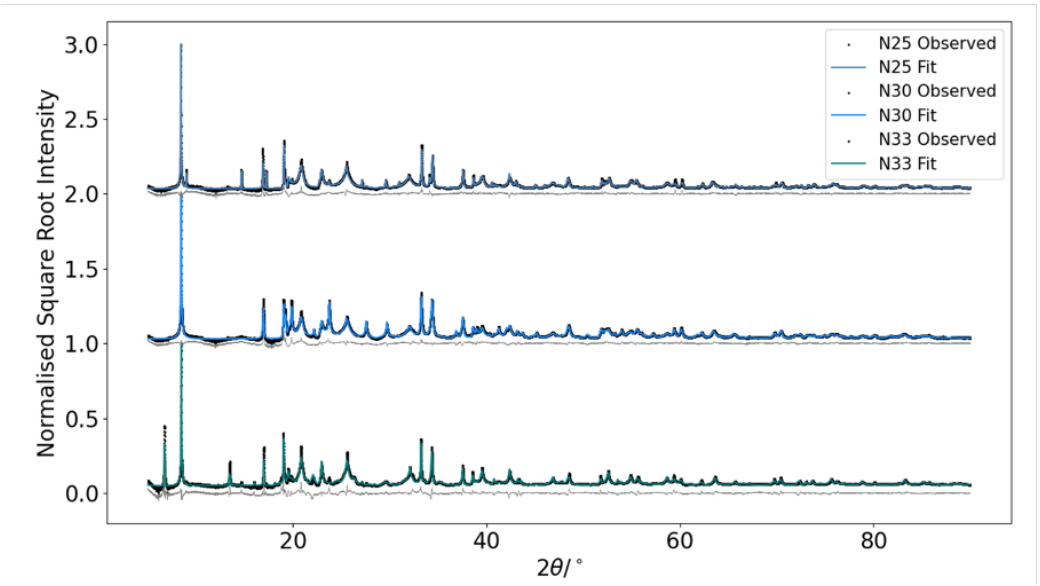


Figure 10.8: The experimental PXRD data for the three compositionally varied samples (N25, N30 and N33) and their corresponding Rietveld refinement fits. The difference between the data and the fits are shown below each sample in grey.

contrast in Ni and Mn scattering and the corresponding change in transition metal environments from P2 to P3.

10.1.3.2 Powder Neutron Diffraction Stacking Model

Essentially the same approach was taken for fitting the PND data as the PXRD data, with the difference that the neutron scattering factor requires the treatment of the transition metal ion ordering as well.

The same three stoichiometric samples, N33, N30 and N25 were investigated again, with the honeycomb ordering and S set to the target stoichiometry and $D = 0$. Due to the larger parameter space, an exhaustive grid sampling was not feasible in a reasonable amount of time, but the P_a and P_b values from the XRD fits were used to guide the initial search.

For the N30 and N25 samples, two independent stacking models were required

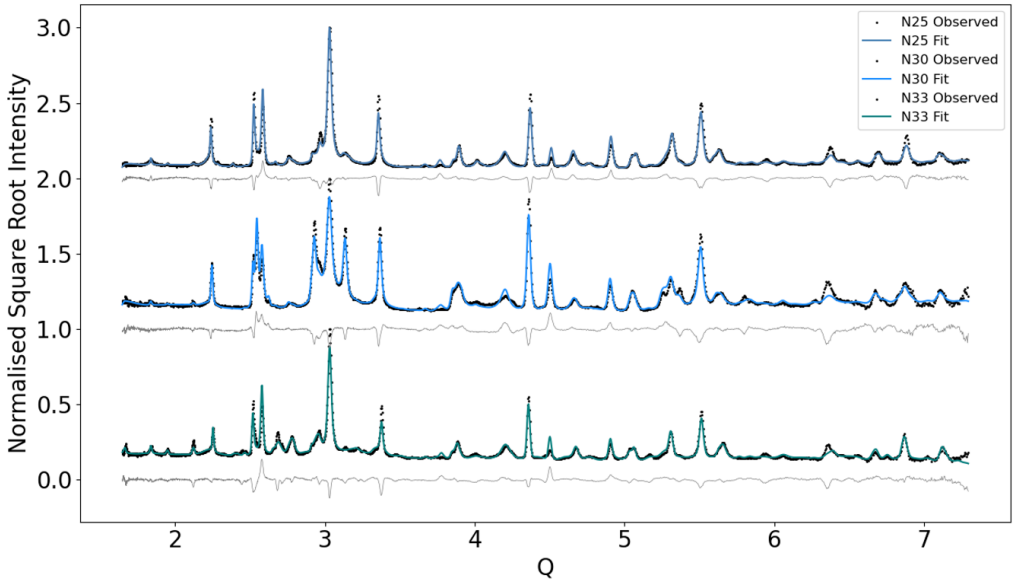


Figure 10.9: The experimental PND data for the three compositionally varied samples (N25, N30 and N33) and their corresponding Rietveld refinement fits. The difference between the data and the fits are shown below each sample in grey.

to most accurately fit the data, with the second model using a different P_a and P_b parameterisation which favoured the P3 stacking regime much more heavily.

The best fits, their metrics and the corresponding parameters are shown in Figure 10.9, Table 10.3, and Table 10.4 respectively. Again, a comparison to a completely ordered, standard model is given in the Appendix (Figure A.3 and Table A.2).

Table 10.3: N25, N30, and N33 PND fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF)

Sample	Target Stoichiometry	Fit Quality Metrics		
		R_{exp}	R_{wp}	GoF
N25	$Na_{0.62}Ni_{0.25}Mn_{0.75}O_2$	1.30	9.76	7.50
N30	$Na_{0.656}Ni_{0.3}Mn_{0.7}O_2$	1.34	8.93	6.66
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	1.12	7.44	6.63

The difficulty of dealing with a larger parameter space is evident here, as it is hard

Table 10.4: N25, N30, and N33 PND fit parameterisations as well as the parameterisation for the additional P3 phase present in the N25 and N30 samples

Sample	Target Stoichiometry	Parameter				
		P_a	P_b	P_{cs}	P_{os}	P_{ss}
N25	$Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$	0.15	0.90	1.00	0.00	0.00
N30	$Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$	0.10	0.25	0.78	0.22	0.00
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	0.15	0.95	0.73	0.27	0.00
N25-P3	$Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$	0.50	0.04	1.00	0.00	0.00
N30-P3	$Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$	0.50	0.04	0.78	0.22	0.00

to be confident of the specific values and tuning values cannot be done independently due to the correlation between effects. Regardless, reasonable quality fits were obtained and the differences in the parametrisations of these models, in conjunction with the XRD models can provide some insight into the different sample structures.

10.1.4 Discussion of results of PXRD and PND stacking models

From the stacking models obtained both from the PXRD data and the PND data, the complexity of these systems are clear. Even the N33 sample, which is regarded as the standard P2 system shows significant nuance in both the oxide stacking and transition metal ordering.

Using the PND transition matrices, it is possible to visualise the oxide layer stackings for the three samples by generating example 300 layer stacks, shown in Figure 10.10.

From Figure 10.10 it is evident why such a stacking approach is required to capture the complexities of these systems. The models show that within the P2 favouring parameterisations, the oxide layers of N25 and the N33 systems stack much more similarly than those of the N30, with the N30 system showing larger uninter-

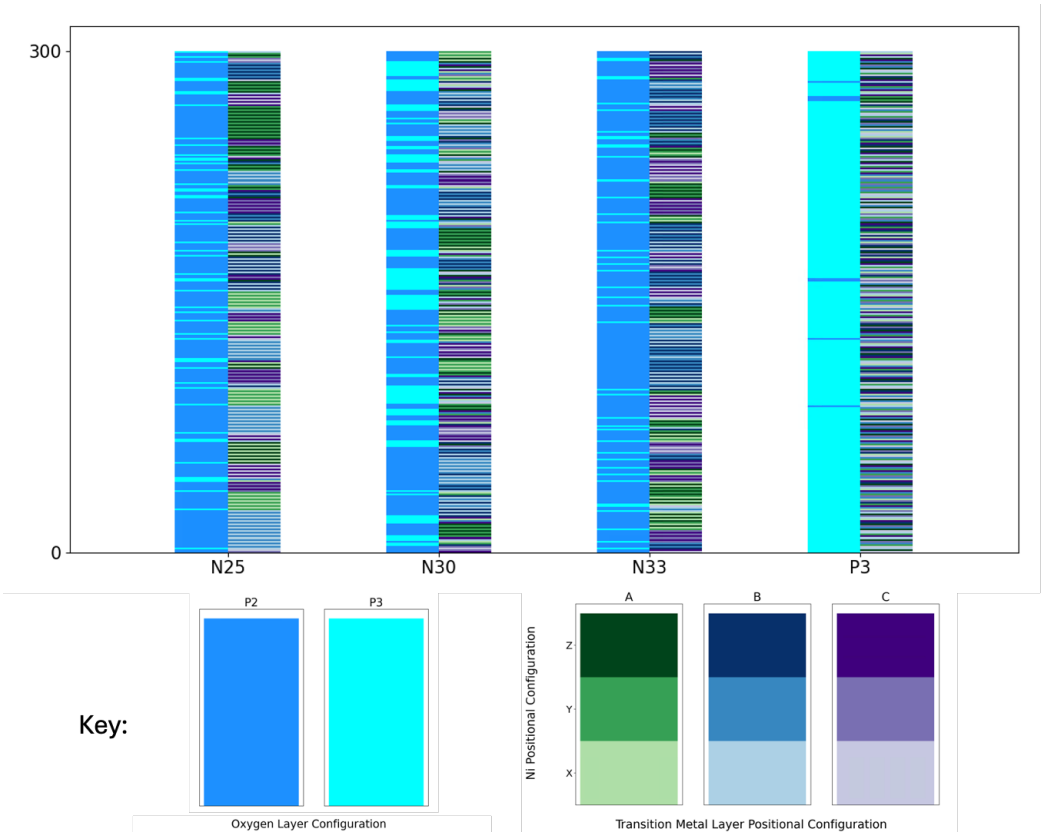


Figure 10.10: Example stackings generated from the transition matrix parameters fit to the PND data for the three stoichiometric samples (N25, N30, N33) and the extra P3 phase. The left hand side of each stack shows the ordering of the oxide layers, and whether they conform to a P2 or P3-type stacking regime. The right hand side of each stack shows the positional configuration of the Ni columns within the honeycomb ordering scheme. The different options for the positional configurations are indicated by the key at the bottom.

rupted blocks of P3-type oxide layers. This observation is in line with a qualitative assessment of the PXRD and PND data, where diffraction profiles of the N25 and N33 samples are also more similar to each other than the N30 sample. Despite this, there are subtleties between the N25 and N30 models, seen in both the oxide layering but also the transition metal layering. The structural models suggest that the P3-type oxide stacking within the N25 sample persists slightly longer than the equivalent faulting in the N33 (owing to the smaller P_b for N25). Additionally, it is worth restating that both the N25 and N30 samples required the additional P3 dominated stacking while the N33 sample did not.

Looking at the transition metal ordering, further nuance between the N25 and the N33 can be seen, with the N25 showing more consistent inter-layer ordering in general, but specifically within the uninterrupted P2-type oxide sections. This is an unexpected result, as the N33 has the $2:1$ Mn:Ni ratio required for perfect honeycomb ordering, whereas the N25 must have interruptions to the ordering within the layer. One possible justification for this is that as the honeycomb ordering cannot be present ubiquitously across the layers of the N25 sample, the magnitude of the effect it has on the PND data is likely to be smaller resulting in a weaker response for tuning the transition metal ordering parameters, more uncertainty in the value and an erroneously more ordered inter-layer system. Alternatively, the parameter fit could be correct, and the fact that the Ni:Mn ratio in the N25 sample prevents layers from forming large spanning honeycomb orderings means that the smaller regions which do exhibit the ordering, will have a smaller energetic barrier to the rearrangement required for the seemingly preferential XYXY-type ordering.

From the N25 stacking in Figure 10.10 the effect that even single layers of P3 faulting has on the transition metal ordering can be seen. As the P3 faulting layer disrupts the transition metal positional configuration from all being localised to

the same A, B, or C sites, it leads to a relatively short range order of consistent XYXY-type ordering, despite the P_{cs} parameter being set to 1.

The additional P3 stacking shown also highlights the massive impact that the seemingly minor change from a P2-type to a P3-type oxide layering scheme has on the inter-layer transition metal orderings. This difference in inter-layer orderings suggests that the changes in the local transition metal ion environments between P2 and P3-type structures will be much more meaningful than might be expected upon first inspection. As these materials are intended to be used in dynamic redox systems, understanding the local environment of the ions is imperative to fully understanding the electronics of these structures.

Finally, this modelling approach in general highlights the limitation of referring to materials as 'P2 structures' or 'P3 structures' as such labels are likely to be significant oversimplifications in most practical systems. Additionally, with the prevalence of single layer faults, referring to these as P3-type faults is somewhat contradictory to the 'P3' referring to sodium ions in prismatic sites with *three* repeating MO_2 units. Ideally the labelling system would incorporate more of the information available in Figure 10.10 to make such a name more expressive of the nuances of the underlying structure. A possible option for this could be an extension of the $A^{+/-}$ and $P^{+/-}$ nomenclature proposed by Brittain et. al.[83]. The extension would include a subscript for the proportion population of each type of layer so rather than referring to N25, N30, and N33 all as 'P2 structures', they could be $P^{-}_{86}P^{+}_{14}$, $P^{-}_{73}P^{+}_{27}$, and $P^{-}_{87}P^{+}_{13}$ structures (these could be shorthanded to $P^{-/+}_{86}$, $P^{-/+}_{73}$, and $P^{-/+}_{87}$). This scheme would allow scope for naming materials with the O2 and O3-type stackings, as well as larger combinations of the four. For example if a phase exhibited all four of the stacking environments it could be denoted as $P^{-}_{a}P^{+}_{b}A^{-}_{c}A^{+}_{d}$ where a, b, c and d are the fractional populations of the different types of layer.

This proposed naming scheme would still be limited from expressing the consistency of these stacking regimes as fully as the transition probabilities P_a and P_b but with the advantage of scaling more straightforwardly to more complicated systems.

10.2 Sodium Ion Ordering

This model, unlike the previously discussed TOPAS-based stacking model, was implemented wholly in python. The transition metal ordering used was again the honeycomb ordering described in the Structure chapter and derived in the empirical model. The sodium ordering used was the SZZ arrangement derived by both the empirical and symmetry driven model. The relative position of the Ni ions between layers have three positional options, A, B and C. The SZZ has three directional options and each of these then have three positional options. The stacking model defines three 'faulting' probabilities, metal layer faulting, sodium position faulting and sodium screw-type faulting. The structural model is built up generating the next layer based on the previous layers and these probabilities. The effect of these faults upon the stacking are discussed below.

10.2.0.1 Sodium Faulting

As mentioned, within the SZZ scheme, the sodium layers can have three directions and each can have three locations. As the three directions are linked by 60° rotations, when there are two subsequent sodium layers, the pair of directions are linked as part of a screw-type relationship. When generating a new stack, if screw-type is unfaulted, the next stack will just follow the same screw type relationship as the previous layers. The screw types were denoted as 6_1, 6_5, and 2_1, referencing only the periodicity of the sodium layer directions. The directions were theme-

selves denoted as h, f, and b after the ASCII symbols hypen (-), forwardslash (/) and backslash (\) and their relation to the directions of the sodium ion zigzags.

For example, if the previous two layers are: hh; ff; or bb, this is the 2_1 type screw axis stacking that was presupposed during the symmetry driven approach. It is 2_1 due to the inversion of site locations in sequential sodium layers of P2 so despite having the same h/f/b direction, they don't repeat positions until after two layers.

If instead the two layers are: hf; fb; or bh this is the 6_1 screw, where the unit cell contains 6 sodium layers. The alternative is: hb; bf; fh, which corresponds to the 6_5 screw axis over the 6 sodium layers. It is worth highlighting that this is only for the independent sodium layers, when these are taken in the context of the whole structure, they do not necessarily translate to the presence of these type of screw axes.

If the stacking continues with an unfaulted screw-type, the next direction of the established screw type is added. For example: hh continues to hhh, fb continues to fbh, and bf continues to bfh. If, instead, the stacking is faulted in the screw type, a random direction is chosen and this establishes a new screw-type regime. For example if bf faults to give bff, the final ff has swapped the regime from 6_5 to 2_1, if the next layer is stacked unfaulted, it would now continue with another f.

In a similar manner to the metal ordering, these sodium zigzags can be located on three dissimilar sites. By preference it is assumed they should stack upon each other in an 'AA' type stacking. If instead they stack in a faulted manner a random location is chosen and the 'AA' type stacking continues, centred on this new location. For example, if the stacking is AA and a location fault occurs the next layer might go AAB, if the next layer after is unfaulted it will continue AABB.

10.2.0.2 Implementation

These faulting probabilities were used to generate test structures which were then refined and fit to the diffraction data using TOPAS. The R_{wp} of the different refinements were then used to evaluate the quality of fit and the probability values iterated. This was done using a Bayesian optimisation to try and minimise the R_{wp} as a function of the probabilities.

10.3 Results of intra-layer and inter-layer sodium ion ordering model

10.3.1 Powder X-ray Diffraction

As noted earlier little information about the transition metal ordering can be corroborated using the PXRD data due to the similarity in x-ray scattering factors between Ni and Mn. What can be studied, however, is the sodium ordering, which manifests as small peaks in the low two theta region of the pattern. Interestingly, the visibility of these superstructure peaks can vary meaningfully from batch to batch. Two samples, both of the same N33 stoichiometry with similar synthesis conditions displayed different profiles in the region where super structure would be anticipated, as shown in Figure 10.11.

From Figure 10.11 it can be seen that the sodium ordering on the bulk scale is inconsistent from sample to sample. Both patterns do, however, display a raised hump between 10° and 14° and one of the patterns (DK) additionally displays two resolved peaks that can be attributed to sodium ordering. Despite the lack of resolved peaks for the TW sample, this is the exact sample that both the SEND and the HR-STEM data was collected from. This means that the explanation for the

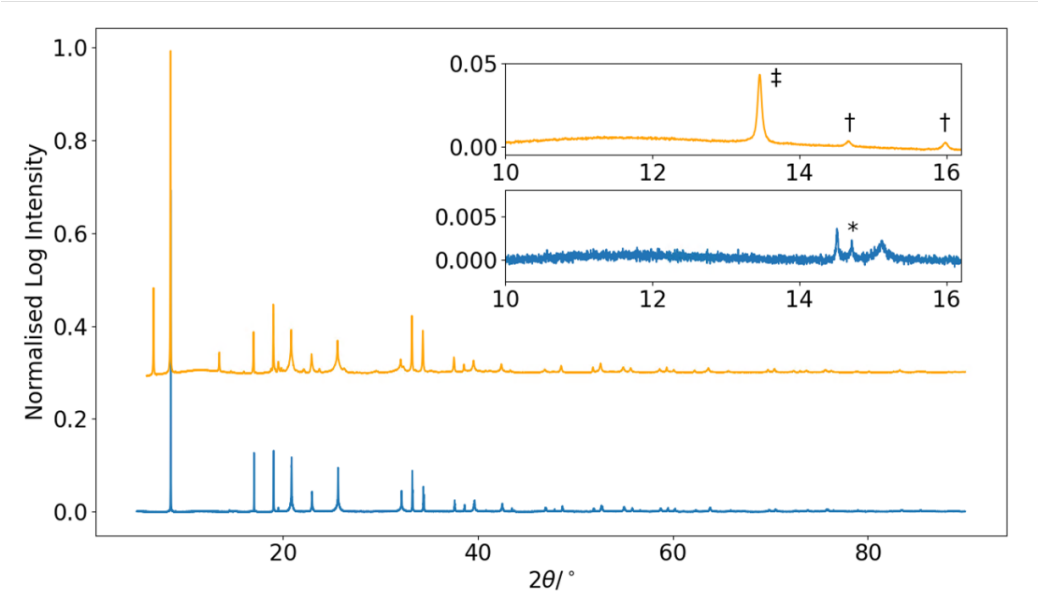


Figure 10.11: Powder X-ray Diffraction Patterns for two samples, DK (blue) and TW (orange), both with N33 composition. The inset axes show the 2θ region between $10°$ and $16°$ where superstructure peaks would be anticipated. Impurity peaks are annotated for $NaWO_4$(*), Na_2CO_3(†), and hydrated P2 (‡).

lack of super-structure peaks is not that there is simply no ordering in this particular sample. In both cases the super-structure peaks are small, perhaps suggesting that ordering is not ubiquitous and that disordered sodium layers also exist however the Z dependence of scattering factors of x-rays will also contribute to these low intensities. One possible explanation for the lack of observed superstructure peaks in the TW PXRD, is that when they are observed in the electron diffraction, the ordering is inconsistent over the length scale of unit cells, displaying high disorder both within layers and between layers, resulting in slight variation in peak positions and intensity. This hypothesis would be consistent with the DK sample as well. While this sample displays peaks, these cannot be straightforwardly fit by ordered models of either the SZZ or LZZ, due to both fits displaying additional peaks that are not observed in the data. These ordered models display the sodium ordering described for single layers (Figure 9.11) and assume these orderings are

entirely consistent within layers and are consistently related by a 2_1 screw axis between layers. The observation that these ordered models give fits with extra, unseen reflections is consistent with the hypothesis that both inter-layer and intra-layer disorder is present within the materials.

Modelling this disorder required an expanded approach from the more straightforward stacking model that was used for the neutron data. This model incorporated the neutron model for the metal layers but modelled the sodium layers separately. This is because the sodium model was constructed as a supercell in the ab plane as well as stacking along c. Thus the model was able to incorporate disorder both within layers and between them. The sodium model was produced as a 3D array, where the ordering in one cell was determined by both neighbouring cells in the ab plane and the sodium layer above and below. The 3D array was populated according to a probability matrix of how likely these neighbouring interactions should be. The base sodium orderings used were the SZZ in the three directions (h, f, b) and three locations (A, B, C) as defined earlier. The sodium model was defined by 5 parameters; two parameters estimating the probability that neighbouring cells are the same SZZ direction (one for neighbours within a layer, one for neighbours between layers); two parameters for estimating the probability that neighbouring cells are the same SZZ location (one for neighbours within a layer, one for neighbours between layers) and a final parameter which weighs the contribution of the inter-layer probabilities against the intra-layer probabilities

10.3.1.1 DK Model

The DK pattern was modelled using the following parameters: 0.6 for the intra-plane directions; 0.99 for the inter-layer directions; 0.75 for the intra-layers locations; 0.8 for the inter-layer location and a weighting of 2. The fit to the data is shown in Figure 10.12

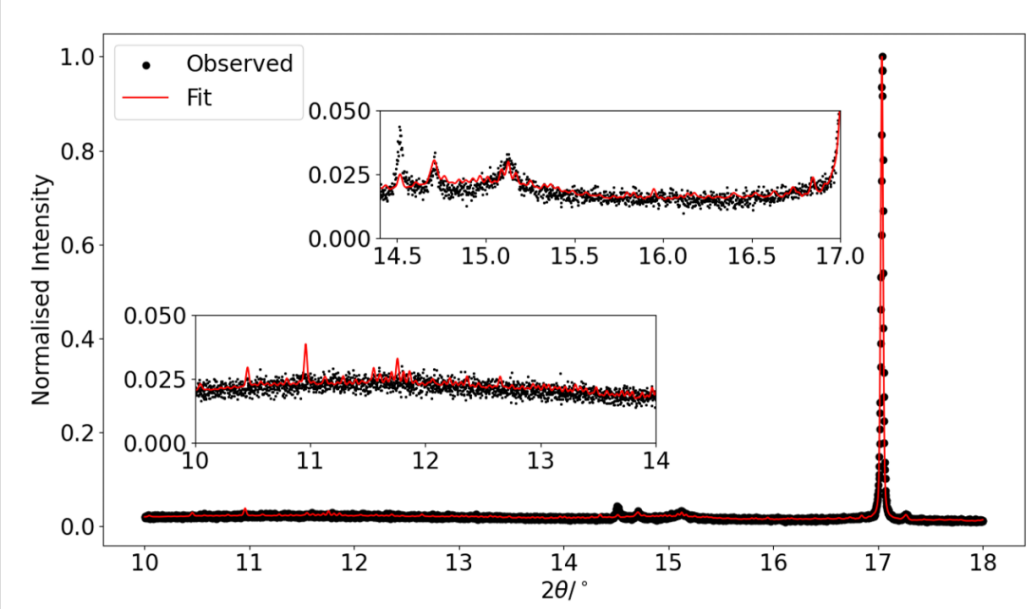


Figure 10.12: Fit to the DK-N33 composition PXRD data using the disordered sodium super-structure model.

Figure 10.12 shows a relatively good agreement between the fit and the experimental data. The model accounts for the broader peak shape at $15.2°$ as well as qualitatively accounting for both the hump between $10°$ and $14°$ and the hump leading into the large peak at $17°$. Some of the deviation present can be attributed to the finite size of the model, using a $5 \times 5 \times 30$ supercell of the base $2\sqrt{3}$ sodium ordered SZZ cell. This finite cell results in some jagged, discrete features occurring in place of seemingly continuous features and this is likely the explanation for the deviation of the peak at $15.2°$ and the more general features of both humps. There are also some peaks that are prominently above the background and the under representation of the peak at $14.5°$. It is possible that these too could improve with a larger model, as the probabilities are represented more accurately, but these could also be a limitation of what is being accounted for by the model, and an indication of there still being more to understand about these structures. Figure 10.13 shows the comparison between this probabilistic model and the fit achieved

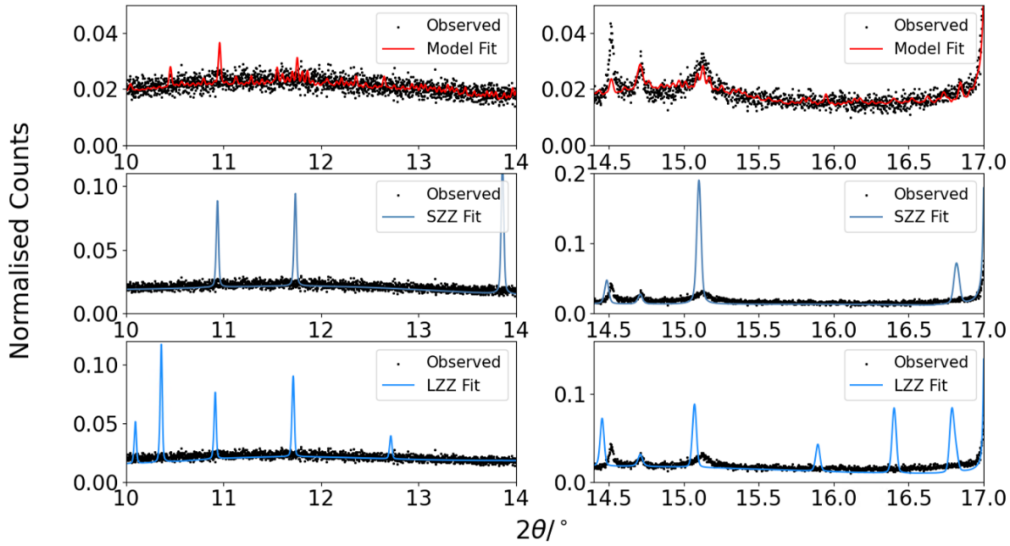


Figure 10.13: Comparison of Probabilistic model fit to the standard SZZ and LZZ models

with just a standard SZZ or LZZ model.

It can be seen that the probabilistic stack model provides an improved fitting of the data than either of the standard models, supporting the general hypothesis that intra-layer and inter-layer disorder is present in these structural systems.

10.3.1.2 TW Model

Despite the contrast in superstructure peaks visible in Figure 10.11 the same model provides a relatively consistent fit to the data, as shown in the first row of Figure 10.14. This suggests that the primary factor motivating the difference in superstructure between the DK and TW sample is not the underlying ordering but the factors which are affecting peak shape, such as the crystallite size and strain.

Additionally to this there is an improvement to fit quality when using a more disordered model as shown in row 2 of Figure 10.14. This more disordered structure was modelled using the following parameters: 0.6667 for the intra-plane direc-

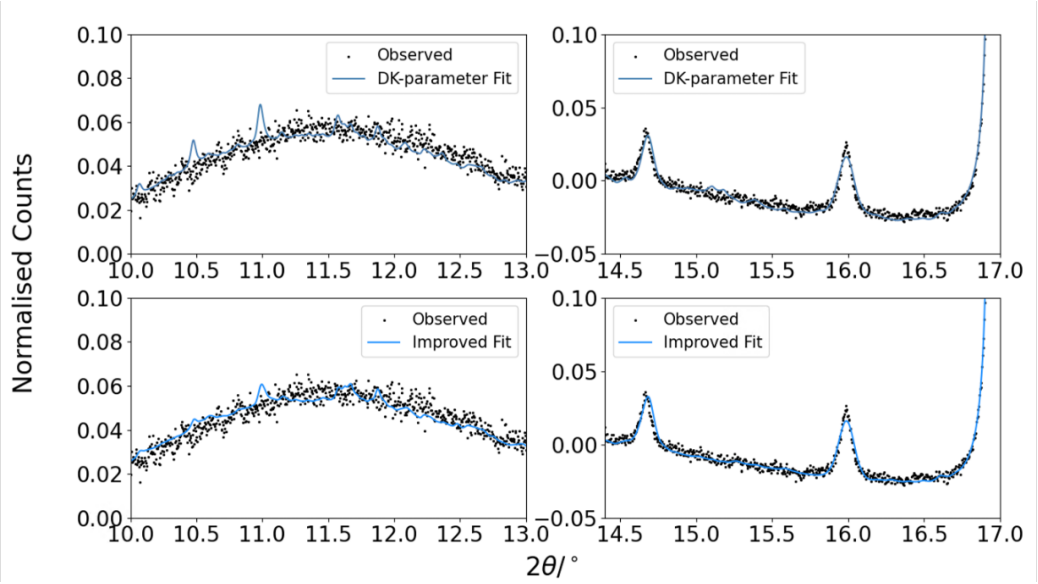


Figure 10.14: XRD fits of probabilistic models with the same parameterisation as for DK shown in the first row, and a moderately more disordered model shown in the second row.

tions; 0.6667 for the inter-layer directions; 0.6667 for the intra-layers locations; 0.6667 for the inter-layer location and a weighting of 2.

This modelling of the sodium ordering was then used to try and corroborate the findings of SEND and HR-STEM

10.3.2 XRD-derived model for Microscopy simulation

The final check of the quality of the probabilistic-derived model of the structural ordering was to try and use the bulk model to replicate the results observed on the electron microscope. This was already achieved using the perfect ordering model derived by the Empirical and Symmetry Driven approaches to fit the SEND 'snowflake' pattern in Figure 9.13, however, the insight from the XRD modelling suggests that there is much more disorder present in the sodium super-structure than this ideal model accounts for.

This hypothesis is supported by the HR-STEM data, which requires a broad sampling by means of a large field of view before the super-structure peaks are observable in the Fourier Transform. This is shown in Figure 10.15

The result of this is that it is difficult to accurately simulate the HR-STEM data using the derived atomic model as the required field of view is prohibitively large. Some approximations can be simulated on account of the reduced noise in the simulation. This allows for better resolution of peaks for a given field of view. The largest simulation achievable with the memory available on 4 GPUs on the Diamond HPC IRIS cluster allowed for a field of view of around 40 Å, comparable with the smallest field of view shown in the bottom row of 10.15. The HR-STEM was simulated for three sodium ordering models, where the first model uses the same structural model as the PXRD fit for the DK sample. The other two simulations use parameterisations with more local order, with both displaying high consistency of the ordering within layers, but one displaying high consistency between layers and one not. The three simulations are shown in Figure 10.16.

Comparing the simulations in 10.16 to the different field of view regions in 10.15, it can be seen that all the simulations display an intensity of superstructure reflection comparable with that displayed in the smallest field of view of the experimental data. It can further be seen that the most comparable model to the largest field of view is the model which is well ordered within the layers but weakly ordered between the layers. This is consistent with the theory that the sodium positions are much more strongly correlated within layers than they are between layers.

Without performing a simulation of the models over a comparable field of view to the experimental data, however, it is difficult to evaluate how consistent the structural model used to fit the DK sample for the PXRD is over length scales beyond $100\,\mathrm{nm}$. What has been shown, however, is that the SZZ sodium ordering

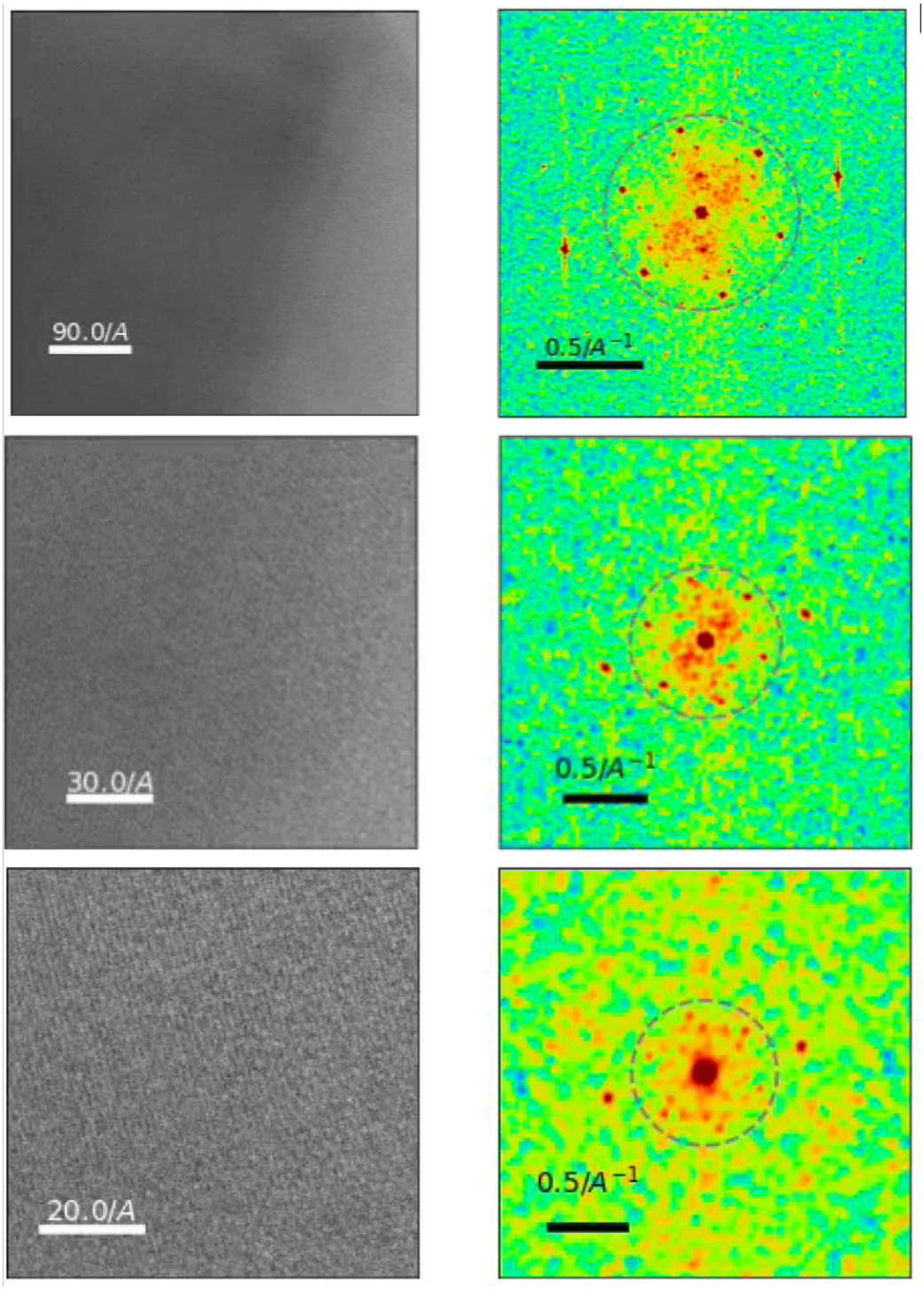


Figure 10.15: Regions of HR-STEM Annular dark field data (left) and the corresponding Fourier Transforms (right) shown for different size regions (top to bottom).

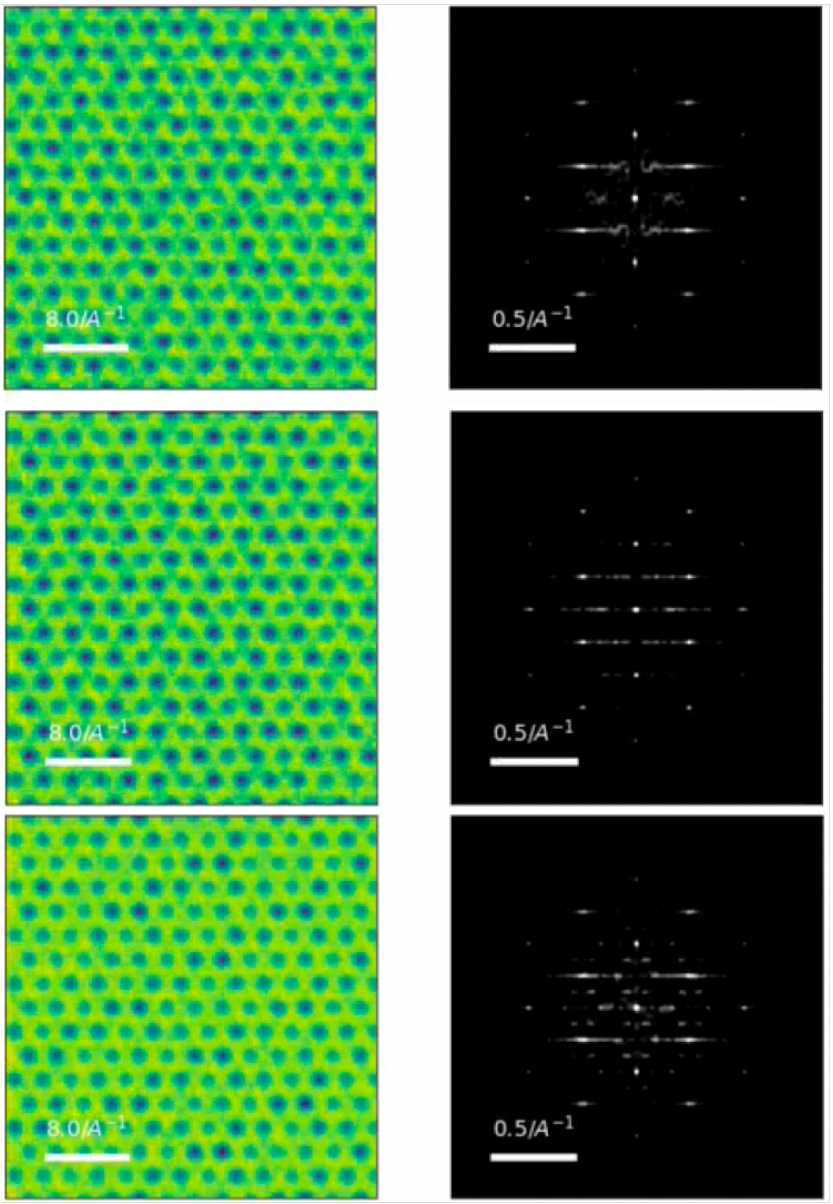


Figure 10.16: Regions of HR-STEM Annular Dark field data (left) and the corresponding Fourier Transforms (right) shown for: The model used to fit the DK sample PXRD (top); a model with 90% consistency between the ordering of neighbouring cells both within layers and between layers (middle); a model with 90% consistency between the ordering of neighbouring cells within layers but only 33% consistency between layers (bottom).

pattern can be incorporated into structural models that do not contradict any of the experimental data observed over the three different length scales of HR-STEM, SEND and PXRD.

Another consideration is the selection bias present in the more local techniques such as HR-STEM and SEND. As the proportion of the total population of particles that can be sampled by these techniques is much smaller than the corresponding proportion for PXRD, it is possible that the bulk of the material might be more accurately represented by a disordered ordering model, but the smaller particles (sampled in SEND and HR-STEM), may be better represented by a more ordered model. Additionally in investigating these superstructure peaks, there is a bias to seek out more ordered regions within these particles.

10.4 Understanding why traditional approaches structure solution approaches struggle

The models developed by complementary application of these scattering techniques, highlight the surprising level of disorder present within these ion 'orderings'. They further explain why the application of VADF and inverse Fourier transform were unsuccessful in gaining significant insight into the nature of the ordering. The modelling suggests that the 3-fold diffraction pattern observed in the SEND dataset, is infact the result of the averaging of different orientation of an underlying 2-fold ordering motif, varying both in plane within a single probe scan location and between the stacking of these layers along the view direction. As a result of this it is unsurprising that little insight could be gained from either a VADF or inverse FFT approach. As the intensity of individual superstructure reflections are a consequence of this complex disorder, averaged within individual probes

locations, and additional dynamical interactions, using subsets of superstructure reflections did not give interpretable changes in the VADF images. Likewise, using the inverse Fourier transform which gives essentially a 2-dimensional pair distribution function (showing the present inter-atomic positions) also fails to give much insight, as it is giving an averaged representation of all the inter-atomic positions present across the layers, and between the layers. This average, essentially results in returning an image with some fractional population of all the possible sodium-sodium inter-atomic separations. In general, the short comings of the application of these techniques come from the crucial information which has been lost as a result of measuring only a plane of the intensity of the complex electron wave function. A wave function which is itself a function of the interaction of many electron wave functions with the different structural arrangements along the beam path within the illuminated probe.

10.5 Conclusions

In summary, within this chapter we have attempted to model the ion ordering in P2 type structures with an appropriately high level of rigour, where traditional approaches have been unsuccessful. In doing so we have identified nuanced differences in the oxygen and transition metal orderings and outlined a structural modelling approach to fit, visualise and understand the effect of these on the structure in question. If this approach could be extended to more general cases and to incorporate the full possibilities of oxygen layer orderings (not just P2 and P3-type stackings) it could be a useful general purpose approach to understanding a broad range of these material systems. We have additionally suggested an updated naming convention to more appropriately represent the underlying structures of these materials

As well as the transition metal and oxide ordering, we have identified a sodium ordering regime that while occurring prominently within our sample appears to be undocumented within the literature. By evaluating the model against experimental data over different length scales we demonstrate the ability to develop structural models that are more detailed and representative of the underlying systems than could be produced by one technique alone. Going forward these models, and similarly produced models for other systems, can and should be used in conjunction with electrochemical analysis to guide the development of improved cathode materials.

11 | Investigating Mn Formal Valence

In the previous chapter we initiated an investigated into the transition metal ion ordering of three compositions of cathode material: N25, N30 and N33 (referring to % Ni in the transition metal layer, again full stoichiometeries in Table 7.2). Using the empirical potential we generated plausible ion orderings for within the metal layers. As the ratio of Ni to Mn changed we hypothesised deviation from ideal honeycomb ordering. Assuming Ni will maintain the Ni^{2+} valence state this deviation must either be accompanied by a reduction in Na^{+} occupancy or some reduction of Mn^{4+} to Mn^{3+}, to maintain charge neutrality. For the functionality of these samples as cathode material, maximising sodium occupancy is preferable, such that the storage capacity of the battery is maximised. To this end, the Na occupancy within the target sample stoichiometry was kept high, with the idea that charge neutrality would be maintained through a fractional presence of Mn^{3+} ions.

It has been seen in the literature that fine-tuning of transition metal ratios is a key strategy for optimising cathode performance[169][170][171][172]. This fundamental study of Ni:Mn ratio aims to outline procedure for investigation more complex systems closer to the state-of-the-art in terms of performance and demonstrate an approach to probe Mn formal valence across individual layers of similar layered materials. Understanding this in more complex systems may help identify in-homogeneity within Mn valence states and environments, which could in turn aid in elucidating any detrimental structural or electrochemical consequences hampering battery performance.

To conduct this investigation of Mn formal valence across structural layers we first revisit the powder diffraction analysis from the previous chapter to identify

evidence of Mn^{3+} incorporation. From here, we outline the involved procedure of using various electron diffraction based approaches for identifying and sampling suitable regions where the layer stacking is present within the plane. For these regions concurrent spectroscopic EELS data will then be collected and processed to give insight into a layer by layer formal valence.

11.1 Evidence of Mn^{3+}

Using the Rietveld analyses of the PND and PXRD data from the previous chapter, two insights were extracted and compared in search of evidence of Mn^{3+} ions. The first point of investigation was using the site occupancies parameters to estimate the stoichiometric composition of each sample and calculate a subsequent formal Mn valence. The second was investigating the lattice parameters of the P2 phase in each of the samples.

11.1.1 Compositional Analysis

To calculate a formal Mn valence, the compositions of the three cation species Na, Mn and Ni were required. It was assumed that the oxygen occupancy was constant and that Na and Ni would appear solely as Na^{+} and Ni^{2+}. The transition metal site occupancies were estimated using the PND data and the S and D parameters for altering the honeycomb ordering as described in Section 10.1.2. The PND data was chosen as the improved contrast between Ni and Mn scattering for neutrons should allow for more accurate compositional estimation than the PXRD equivalent would. As the S and D parameters could not be refined natively within the refinement software a grid search was performed to indicate the optimal configuration of these two parameters. The resulting contour plots are shown in Figure 11.1.

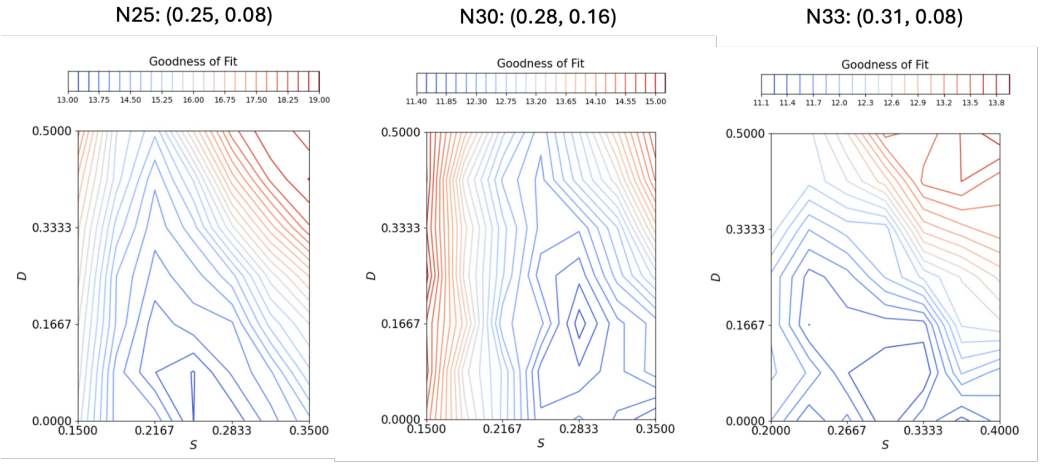


Figure 11.1: Grid search for optimal values of S and D parameters for the PND refinement of the three stoichiometric samples N25, N30 and N33. The co-ordinate of the approximate minima are shown in brackets above.

From Figure 11.1 relatively good agreement is found between the target stoichiometry and calculated stoichiometry, with N30 and N33 both showing slightly lower Ni occupancy than planned. Additionally a similar value for the disorder parameter D is observed for all three samples, with the higher value for N30 appearing consistent with the other observed trends in the analysis of these samples (the N25 and N33 have, in general, displayed more similar results than the N30 intermediate). Across all three samples, however, the refinement appears less sensitive to the D parameter than the S parameter. Additionally, the D parameter does not effect the overall ion ratios so is not imperative to identify accurately for estimating the Mn formal valence. The consistently low D parameter estimation does also fit with the result of the empirical potential investigation which predicted regions of Ni absence compensating for otherwise neatly ordered honeycomb regions.

The Na site occupancy, unlike the transition metal occupancies, was estimated from the PXRD but with greater uncertainty, due to the relatively weak scattering

of Na for both x-rays and neutrons. A more rigorous approach to calculating sodium occupancy would have been ideal but must fall to recommended future work. The occupancy predictions of the cationic sites and the resultant calculated Mn formal valence is summarised in Table 11.1

Table 11.1: N25, N30, and N33 calculated cation stoichiometry and subsequent Mn formal valence

Sample	Target Stoichiometry	Calculated Stoiciometries			Mn Valence
		Ni	Mn (Mn^{3+} + Mn^{4+})	Na	
N25	$Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$	0.25	0.75 (0.20 + 0.55)	0.70	3.73
N30	$Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$	0.28	0.72 (0.14 + 0.58)	0.70	3.81
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	0.31	0.69 (0.04 + 0.65)	0.66	3.94

The compositional analysis results provide supporting evidence of Mn^{3+} presence within the N25 and N30 samples. The calculations of predicted Mn formal valence all output smaller values than the expected value from the target compositions. This is primarily due to the higher than anticipated sodium populations. The presence of these higher sodium contents may be due to inaccuracies in the estimation methodology, but are also likely to be a consequence of the presence of the P3-type oxide layer stackings observed within the samples. The presence of these types of stackings give rise to sodium layers where there is no distinction between site energies, resulting in the ability for such layers to accommodate greater amounts of Na.

11.1.2 Lattice Parameters

The refined lattice parameters from these powder diffraction analyses could also be investigated for evidence of Mn^{3+}. Due to the high resolution of the synchotron diffraction patterns, the lattice parameters could be calculated to a high precision. The trade off is that these high precision measurements can only be indirectly

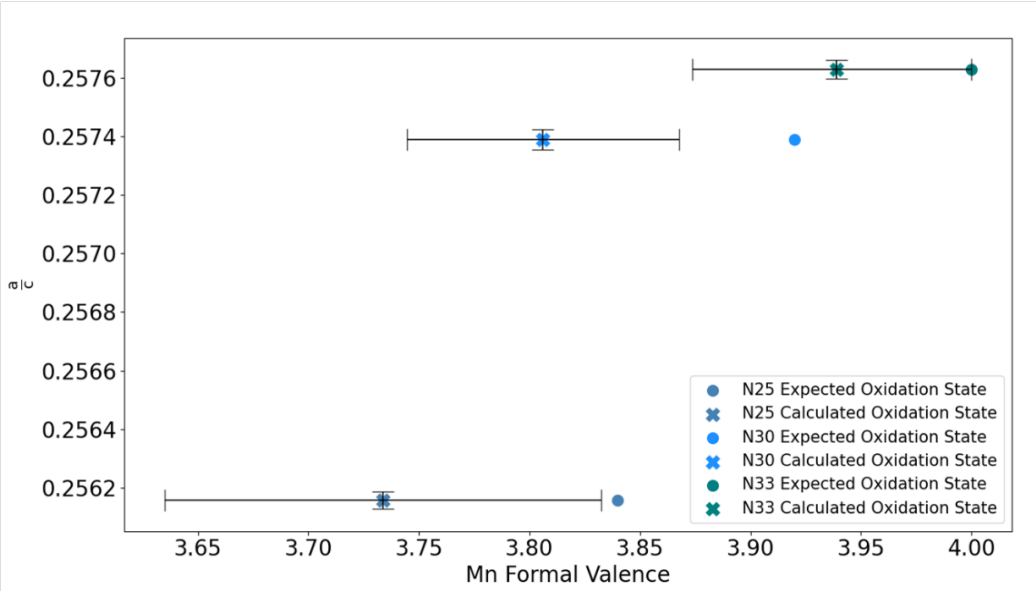


Figure 11.2: Comparison between the ratio of calculated lattice parameters and the trend in calculated Mn formal valence (circles) for the three compositionally varied samples (N25, N30 and N33). The large uncertainties from the XRD compositional refinements are shown in the horizontal error bars and the small uncertainities in the lattice parameters are shown as vertical error bars. The expected oxidation state, calculated using each samples target composition are shown as crosses

correlated to the Mn formal valence, as only the trend in the lattice parameters can be measured and other factors will additionally contribute to these changes.

The comparison of the compositional analysis and the lattice parameter analysis is shown in Figure 11.2.

The ratio of the lattice parameters is used rather than direct measures as it allows multiplicative errors in measurements to cancel. The ratio calculated show a trend correlated to the targeted and calculated Mn formal valence, however, this cannot be ruled out as just the effect of the changing sodium occupancies.

This is because as the sodium occupancy decreases there is less electrostatic attraction between the sodium layer and the neighbouring oxide layers. Additionally

there is an increase in repulsion between these oxide layers as sodium content decreases. Both these effects result in an increase in the distance between adjacent MnO_2 layers resulting in an increase in the c lattice parameter. As the sodium occupancy will have little effect on the a lattice parameter, solely a change in sodium occupancy would be expected yield a similar correlation to that observed.

This trend in the a/c lattice parameter ratio does also support the hypothesis of more Mn^{3+} in the N25 compared to the N30, compared to the N33.

The justification of this is that $Mn^{3/4+}$ have a d^4 and d^3 electronic configurations respectively. Due to the high effective nuclear charge experienced by the weakly self-shielding d-orbitals, this results in radially contracted orbitals with high electron pairing energy. As such, these Mn ions are prone to adopting high spin electron configurations. The energy of the individual orbitals occupied by these unpaired electrons is determined by the symmetry permitted mixing of the 5d orbitals with, primarily, the p orbitals of the surrounding 6 oxygen ions.

As the P2 structure is of the $P6_3/mmc$ space group and the Mn sites within this structure are of $\bar{3}m$ point symmetry, the oxygen ions are constrained to occupy three equivalent, trigonally coordinated sites above and below the Mn sites. The separation between the three upper and three lower sites is not, however, constrained by the symmetry. This distance can change and will, in turn, change the degree of mixing of the d-orbitals with the oxygen p orbitals which will alter the corresponding energy of each of the resulting hybrid orbitals.

Varying the separation between the layers, there is a distance such that the inter-atomic distance between oxygen ions, both in the same layer is the same as the inter-atomic distance between oxygen ions, both in different layers. At this point, locally to the Mn centre, the ligand environment strongly resembles octahedral symmetry. This local, psuedo-octahedral, environment is then moved away from,

as either the inter-layer separation between the oxygen ions above and below contracts or expands relative to the intra-layer separation. By refining the oxygen locations in these structures, as well as comparison with literature fits, it is accepted that the inter-layer separation is contracted for the N33 P2 structure (with full Mn^{4+} occupancy).

If we assume that the ligand environments move between a generically contracted oxygen ion inter-layer spacing to the pseudo-octahedral inter-layer spacing, we can further assume that the Mn orbital energies will tend to resemble those in an octahedral field more, as the contraction is reduced, and less as the contraction is increased. This is shown in Figure 11.3 using the orbital energy splittings given by Trimarchi et. al.[173]

Using Figure 11.3 we can explain why we would expect a higher proportion of Mn^{3+} to be associated with a smaller a/c ratio, as the Mn^{4+} ions have two $e_g^{'}$ electrons stabilised upon compression and only one a_{1g} electron destabilised, whereas the Mn^{3+} ions have an additional $e_g^{''}$ electron which will be destabilised upon compression. The result of this is that the Mn^{3+} ions will favour a more expanded inter-layer spacing relative to the Mn^{4+} ions and so a higher proportion of Mn^{3+} ions is expected to result in a smaller a/c ratio.

For completeness it is worth mentioning that the single electron in the degenerate $e_g^{''}$ orbital for the d^4 Mn^{3+} ion is likely to induce a Jahn-Teller distortion. It is probable this distortion would be similar to the tetragonal distortion described by Trimarchi for the later d-block transition metals as this would remove the degeneracy of the three oxygen positions coordinated above and below the transition metal through an uneven expansion/contraction. The presence and exact nature of this Jahn-Teller distortion is, however, beyond the scope of this thesis and we shall assume that the former distortion is the dominant lattice parameter factor.

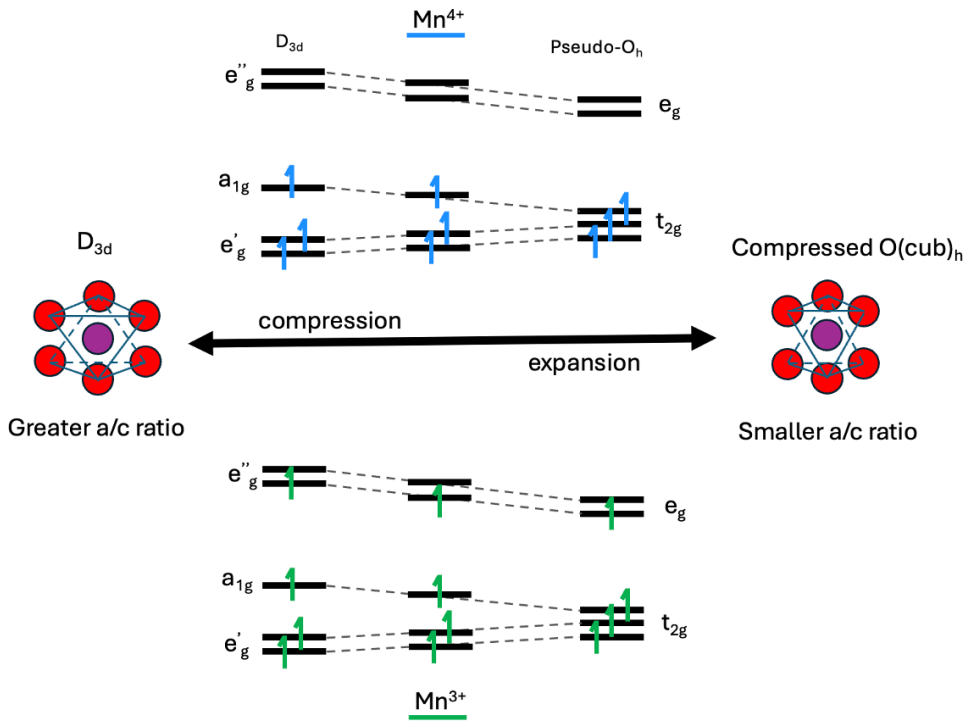


Figure 11.3: Expected relationship between the d-orbital energies in D_{3d} symmetry and the compression/expansion of the inter-layer oxygen-oxygen atomic distances. The electron configuration of Mn^{4+} and Mn^{3+} ions are shown in blue and green respectively.

The inability to confidently separate whether the trend in a/c ratio is simply due to sodium content, or whether changes in Mn formal valence are also present further highlights the need for using more direct approaches than bulk PXRD, such as HR-STEM with EELS, to probe Mn oxidation states. Not only will this offer spectroscopic support of the evidence of Mn^{3+} ions but also, crucially, allow spatially correlating this evidence within the samples.

11.2 Achieving atomic resolution imaging with the c axis in-plane

Collecting direct images of the layering within these structures is a non-trivial challenge. The major difficulties come as a result of inherent experimental requirements to obtain atomic resolution imaging, namely, thin sample along the beam direction. Along with this, the sample would ideally display a consistent structural motif along the beam direction, to maximise the coherent, constructive interference and provide optimal resolution of the atomic columns. The absence of well resolved atomic columns with all other requirements satisfied, may therefore provide evidence of structural disorder along the beam direction. Finally, the sample would also ideally display the desired orientation over a fairly large in-plane area ($\geq 1\,\mu\mathrm{m}$). This is not an imaging requirement, but a practical experimental requirement, as larger areas are much easier to determine the orientation of, and subsequently align to zone-axis than smaller areas through use of Kikuchi Patterns.

Upon preparation and preliminary study of the electron microscopy samples of the various P2 compositions, the majority of particles on the sample holder did not satisfy the requirements to image the **c** axis stacking. Instead the majority

of thin particles at neutral stage tilt appeared to be orientated with the **ab**-plane in-plane and the **c**-axis parallel to the beam direction. It was hypothesised that this is due to the inter-layer bonds being weaker than the intra-layer bonds, leading to preferential shearing perpendicular to the stacking direction upon sample grinding.

11.2.1 Screening particles to find suitable candidates

The procedure to identify suitable specimens for atomic resolution imaging was two-fold: firstly, the samples were loaded onto the Research Complex at Harwell microscope, where an initial screening could be done; then the identified particles were investigated by SEND to identify the structural domains using the Automated Workflow from the previous chapter (Chapter 8).

The screening resulted in identification of candidate particles which might be suitable for HR-STEM, and details of this screening are given more thoroughly in the Methods chapter (Section 7.4). Additionally to identifying well orientated particles, this primary screening allowed for compositional analysis by EDX to confirm that the particle was of the target structure and not an impurity.

A particle was identified as a candidate for HR-STEM if upon screening it was thin enough to give good quality TEM diffraction and the Selected Area Electron Diffraction (SAED) pattern displayed reflections associated with the c-axis. Due to the time challenges in acquiring many quality SAED patterns, this initial screening gave very little indication as to the size of the suitably orientated region. As the region needs to be suitably large to acquire a Kikuchi pattern, this necessitated the secondary screening using SEND.

Once all the candidate particles had been identified and their locations recorded (as again described in Section 7.4) they were loaded onto the e02 microscope in pencil

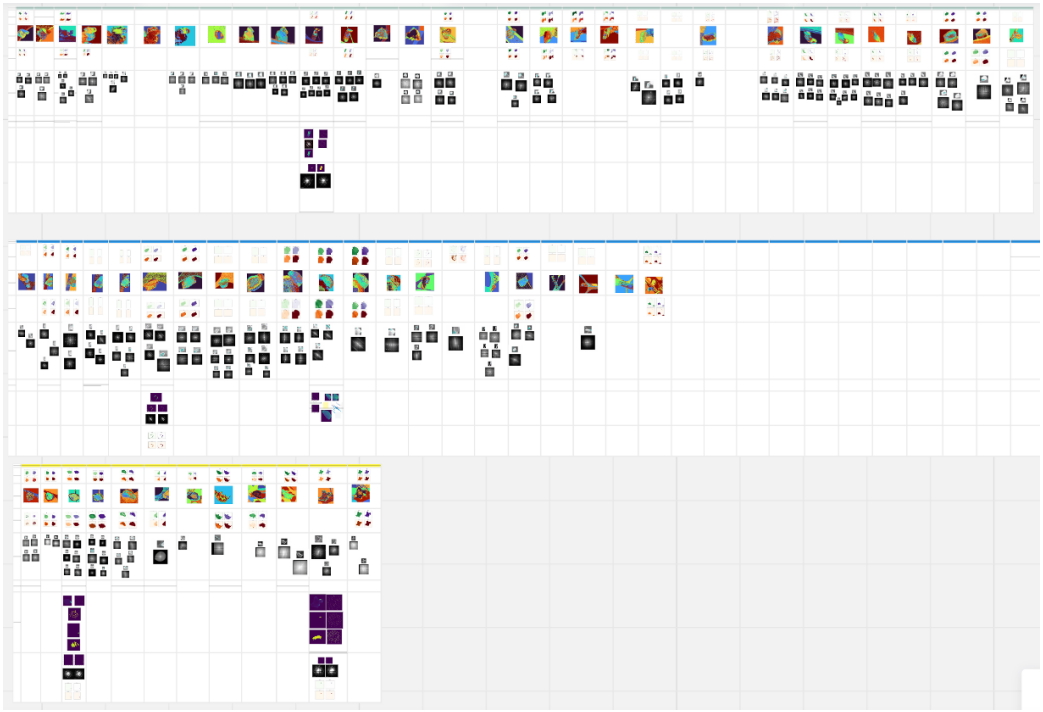


Figure 11.4: Complete reference workbook, each block of rows is a different composition and sample grid and each column is a different SEND dataset collected from one of the pre-screened samples. The sample location is recorded in the column header and the grid reference images are stored elsewhere.

beam mode and SEND datasets were collected. These SEND datasets were then autonomously processed to provide domain maps, identifying both the location of the desired orientation within the particle, but also the size of the region displaying the desired orientation. These were documented in a reference workbook. In total more than 60 SEND datasets were collected and documented in order to identify suitable regions for HR-STEM. The reference workbook is shown in Figure 11.4 as an indication of the scale of the second round of screening. An example of the key information documented is shown roughly in Figure 11.5.

From Figure 11.4 it is possible to see that not all the datasets contain the same amount of information, and some gaps are present. This is partly due to some additional analysis also being recorded for some of the more unusual domains

(shown by extra rows of images beneath the common ones) but also some data absences are present where either the domain mapping was inconclusive or the EDX signal was too weak to acquire compositional analysis (recalling that dual EDX with SEND provides sub-optimal beam conditions for x-ray detection).

After the SEND screening there were several promising candidate regions, thin enough to collect atomic resolution images of the layer stacking, and large enough to give allow aligning of the zone axis.

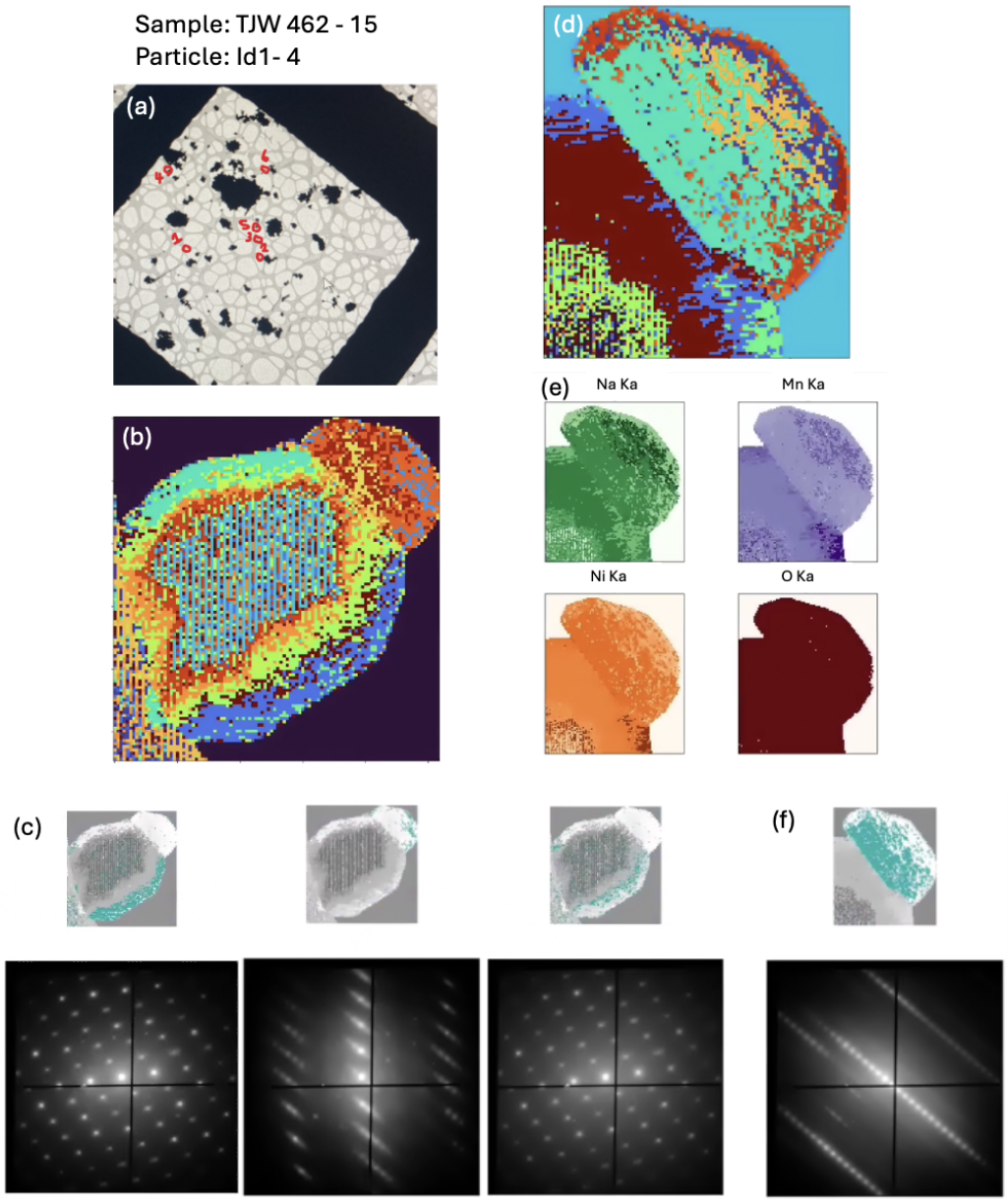


Figure 11.5

Figure 11.5: The information collected for each of the pre-screened candidate particles during the SEND screening. Sample TJW 462 refers to the synthesis batch sample (in this case it is one of the composition N25), the 15 refers to position 15 in the TEM sample grid holder. (a) The low magnification image of grid bar Id1, with the annonated screened particles, with the subject of this being particle number 4. (b) The coarse domain map for the whole of particle 4, produced using the automated SEND workflow. (c) Three regions and their corresponding summed diffraction patterns, the second of which shows high frequency reflections associated with the c-axis. (d) A higher magnification and slightly realigned, domain mapped SEND dataset of the region in the primary sample displaying c-axis reflections. (e) A rough compositional analysis using concurrently collected EDX signal to check that all the expected elements are present. (f) A summed pattern for one of the larger domains displaying high quality c axis reflections

11.2.2 HR-STEM Processing

Once the samples were screened, HR-STEM data was collected with concurrent Electron Energy Loss Spectroscopy (EELS). This allows for both visual imaging of the structure but allows for spatially resolution of the corresponding spectroscopic investigation. As the dwell time for each scan location was limited to reduce electron dose and minimise beam damage, the signal to noise ratio of individual EELS spectra were fairly low signal to noise ratio, as shown in Figure 11.6.

To generate usable data for further analysis, the signal for a given layer of the structure was summed. To do this, the individual layers had to be identified. From the summed spectral image in Figure 11.6 it can be seen that there are bright bands and dark bands. These bright bands primarily correspond to increases in intensity in the Mn and Ni K edges, so these are these bright bands are the transition metal layers. Conversely, the dark bands must be the Na layers. The dark field image (shown in Figure 11.7) shows a complementary intensity pattern but with higher contrast and reduced noise, so that was used to identify the layers rather than the

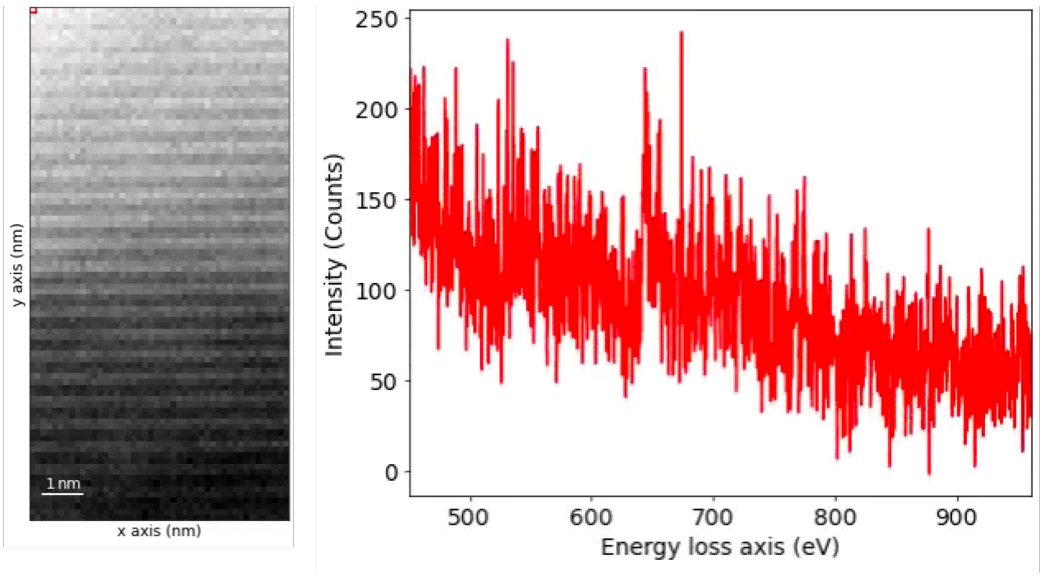


Figure 11.6: The summed EELS spectral intensities for each scan location in a HR-STEM experiment (left) with the individual spectra corresponding to the first pixel (right).

EELS data.

Even without true atomic resolution imaging, such as that achieved, it can be seen from Figure 11.7 that the layers are not perfectly orientated parallel with the pixel grid, and while the layers are approximately horizontal, the brightest pixel in each layer is not necessarily at the same y coordinate across a given layer. This can be a result of slight misorientation of the detectors in the experimental collection but could also be a sign of deviations from perfectly parallel layers within the atomic structure. Regardless of the origin, fitting the location of the layers to the data, rather than assuming horizontal layers, is required for improving the quality of the summed data.

Fitting the layers was achieved by splitting the dark field image into multiple narrower columns, and summing the intensity. This summed intensity signal was then fit with a n-order polynomial to deconvolve the general decrease in intensity

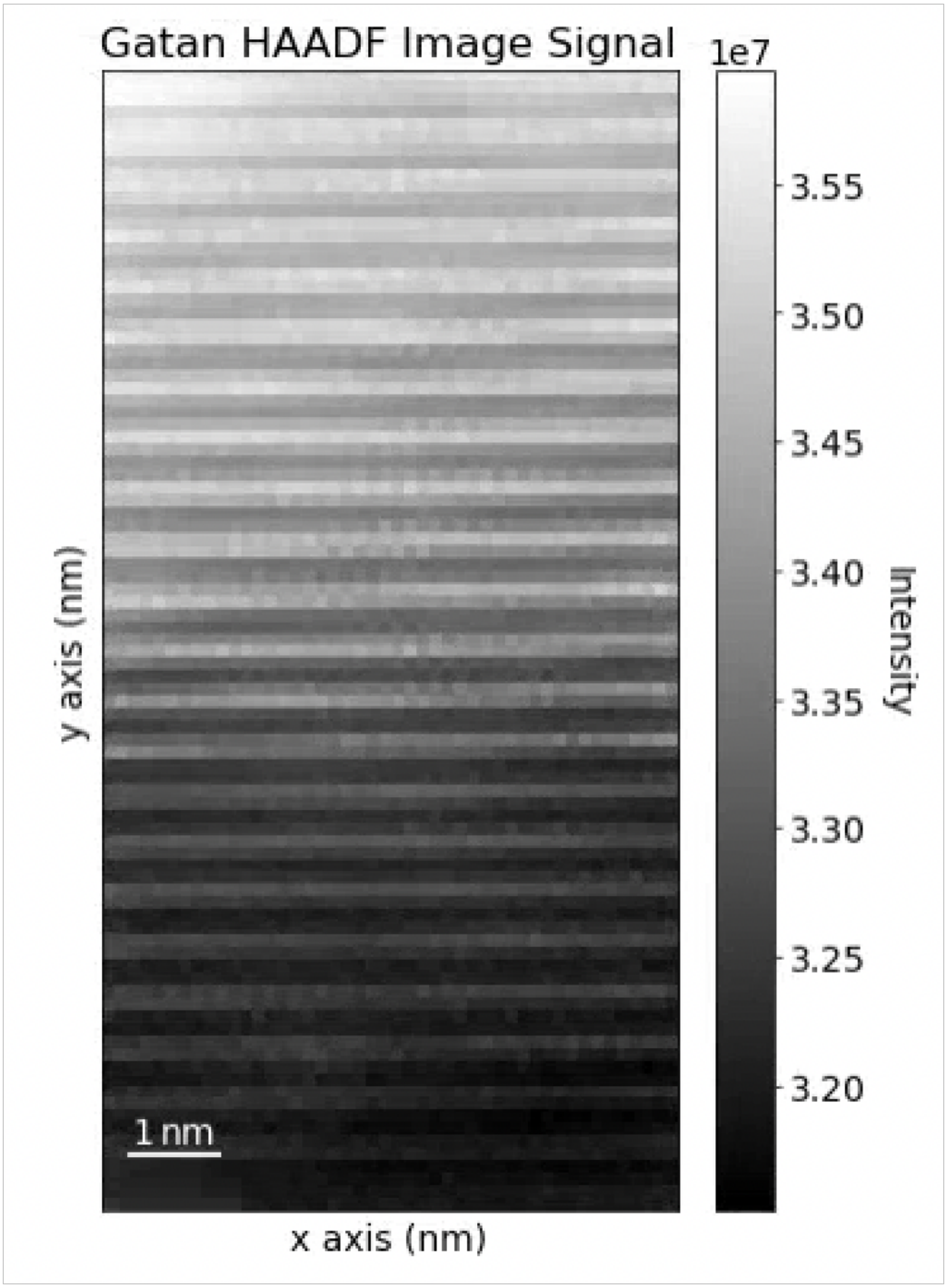


Figure 11.7: The Dark Field image for the same scan as the EELS spectra shown in Figure 11.6

observed from top to bottom. With this polynomial background removed, the maxima in the summed intensity was interpolated and smoothed with a Gaussian kernel, and the maxima used to identify sub-pixel approximations of the centre of each layer. This procedure is shown in Figure 11.8.

Once this has been done for all the columns a fit is calculated interpolating between the identified locations to get a layer centre at each horizontal location. The summed pattern for the layer was then calculated by taking a weighted sum of the two pixel spectra that the calculated height falls between. This is shown in Figure 11.9.

11.2.3 Mn Oxidation State Estimation

These spectra were then used to estimate the average Mn oxidation state for each layer. There is no singularly recognised method of calculating Mn oxidation state from EELS spectra, the main techniques used are: White Line intensities, Energy Offsets and Linear Combinations. Additionally to these, an exciting recent development is the publication of MnEdgeNet, a neural network based approach.

11.2.3.1 White Line Intensities

This is perhaps the most common approach for estimating oxidation state. The method takes advantage of the observed correlation between the relative intensities of the two fine structure peaks associated with the Mn L-edge and the Mn oxidation state. For this work we use the method and calibration curve outlined by Zhang et. al.[174], where the background is initially fit by a Power Law function and the contribution to the continuum is fit using a two step function centred at each peak. Additionally the low loss spectra which is collected concurrently is used to adjust the zero-loss peak and deconvolve the multiple scattering events.

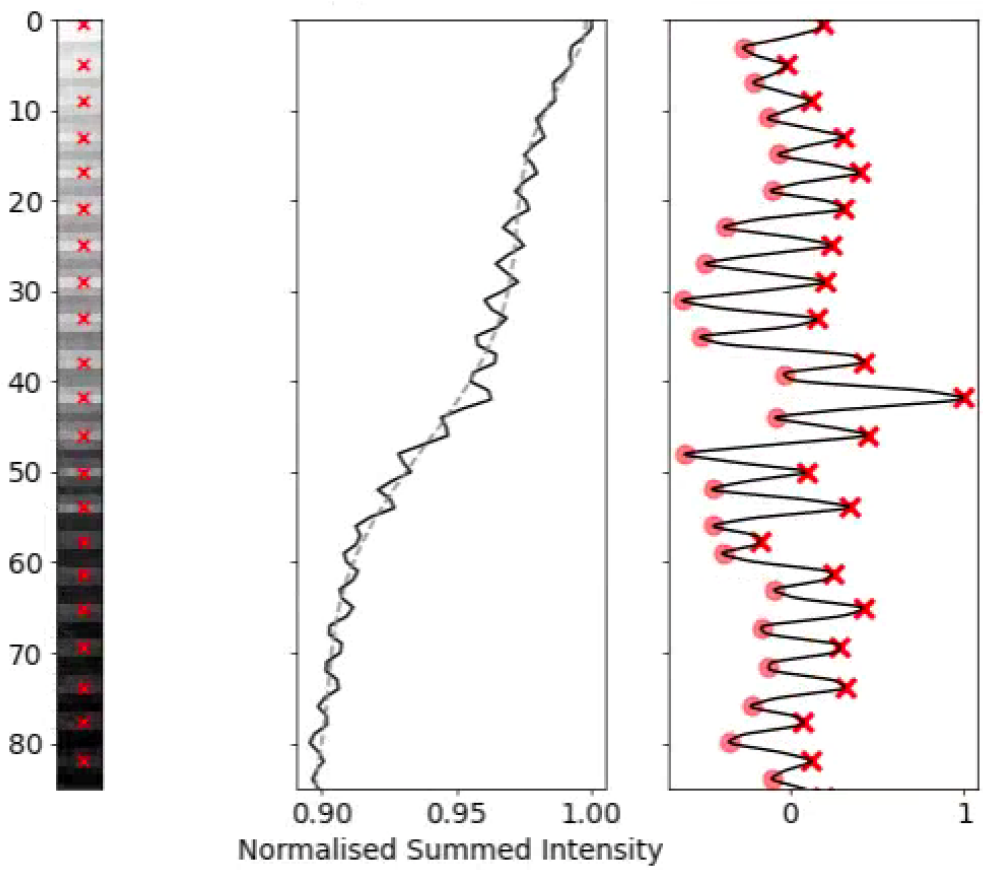


Figure 11.8: (Left) Narrow column of the HR-STEM data and fit the intensity maxima. (Centre) Summed horizontal intensity of the image column with the polynomial background fit shown in grey. (Right) Interpolated, smoothed and background subtracted image data processed to identify maxima (cross) and minima (circles).

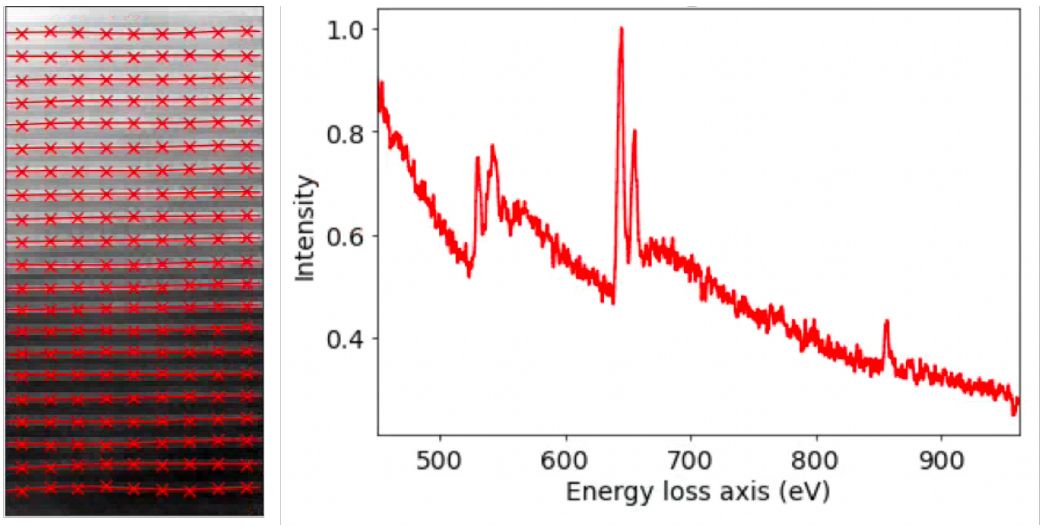


Figure 11.9: (Left) Approximate metal layer locations interpolated across the layer (red line). (Right) Example weighted summed EELS spectra for the first layer.

The intensities are then integrated in a $4\,\mathrm{eV}$ window centred at the peak maxima. The resulting ratio of the L3 to L2 intensity is then used in conjunction with the Quadratic calibration function, Equation 11.1, to calculate the prediction of Mn oxidation state.

$$y = -0.734x^2 + 4.638x - 5.014 \tag{11.1}$$

An example of the calibration and intensity quantification is shown in Figure 11.10.

11.2.3.2 Energy Offsets

As well as the ratio of the L3 and L2 peaks, another common methodology involves comparison of the edge onset for the Mn L3 peak and the O K edge. These two onsets are chosen due to their similar dependencies. It is also possible to compare the L3 and L2 onsets, but this can be unreliable due to the Coster-Kronig

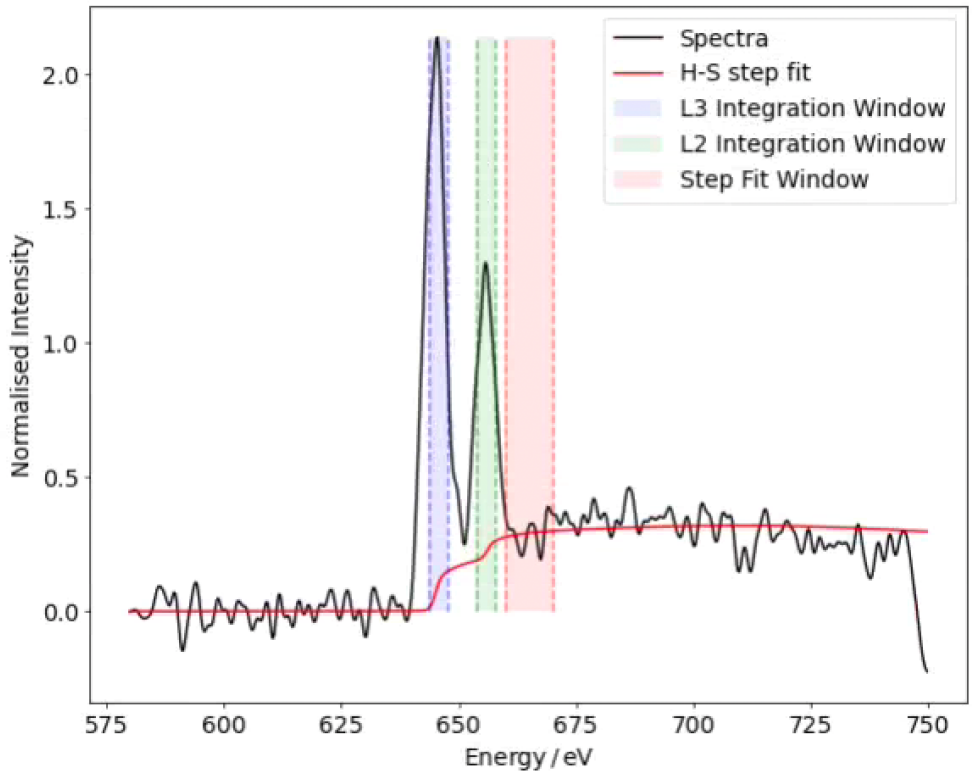


Figure 11.10: Example layer spectra and the regions used for calculating white line intensities. The red line shows the double-step function fit using Hartree-Slater GOS (generlised oscillator strength). This is fit with a 2:1 ratio of step height at L3:L2 to match the step fit window in red. The white line intensity is the integrated area between the spectra and the double-step function, within the L3 window (blue) and the L2 window (green).

Auger decay effect causing L2 broadening and significantly varying the L2 onset as a result. Again the calibration curve used is referenced from Zhang, this time with the linear relationship shown in Equation 11.2.

$$y = 2.65x + 103.45 \tag{11.2}$$

This energy onset approach is the most straighforward of the method to apply as it doesn't require any of the background or two step calibration, but as a result it also doesn't account for any fine structure changes at all.

11.2.3.3 Linear Combinations

This is perhaps the most reliable approach provided high quality reference data is available. The spectra are processed as with the white line intensity approach. These spectra are then fit as a linear combination of other reference spectra with known Mn oxidation states. Provided such spectra exist and the oxidation state in the reference spectra is reliably known, this has been shown to provide good estimation of Mn oxidation state in novel spectra. Unfortunately, we did not collect concurrent reference spectra due to time limitations and as of submission, the quality of reference data in publicly available EELS databases were inadequate to achieve a reliable results from this approach.

11.2.3.4 MnEdgeNet

The final approach utilised was a neural network approach outlined by Ji et. al.[175]. The code for this network is publicly distributed and the authors claim high predictive accuracy without need for data calibration. The standard library is packaged with 8 predictive models, with the following names: model1 (uniform scaling), model2 (trained without noise), model3 (correct scaling), model4

(fixed mean and std, correct scaling), model5 (Used mean and std of each spectrum), model6 (Increased Mn metal appearance), model7 (no metal), and model8 (with metal). As neither the code nor the paper make any reference to the differences of the models it is unknown which model was used to produce the results documented in the paper. This is increasingly problematic due to the range of prediction provided by the different models. Ultimately MnEdgeNet was used in two ways to inform the Mn oxidation states of the layers: either Model 5 was used on its own as it qualitatively provided the most similar estimations to the other approaches, or all the models were used to provide a distribution of predictions and the mean value used for any further calculations. Which of the two approaches is used for any given result should be made clear in the corresponding discussion and figures.

11.2.4 Results

11.2.4.1 Dark Field Images

In total, six usable datasets were collected with simultaneous EELS detection, four of these were of N25 composition and two were of N30 composition. The dark field images of the six datasets are shown in Figure 11.11.

For the remainder of this discussion these six samples shall be referred to as Sample N25-1, N25-2, N25-3, N25-4, N30-1 and N30-2 for (a)-(f) respectively in Figure 11.11. From these dark field images various observations can be made. Firstly the resolution of the atomic columns within the layers varies from sample to sample but qualitatively the N30 composition images appear to resolve individual columns better. This might indicate that the atomic coordinates within a layer are more consistent, which would be consistent with a more ordered layer, as simulated in the previous chapter. Alternatively, this may just be the result of

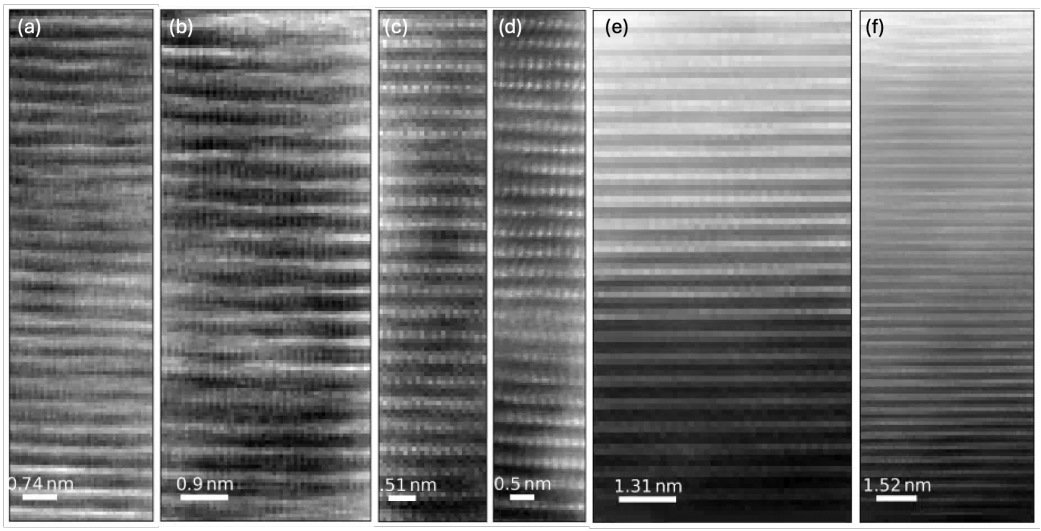


Figure 11.11: Dark Field Images of six high resolution images which were collected with simultaneous EELS. Images (a-d) are of N25 composition, (e and f) are of N30 composition.

the specific samples used and may not be representative of the compositions generally. Repeated experiments following the same procedure could help determine if this is a consistent observation across the two sample compositions.

Further, images (c) and (d) provide more evidence for the incorporation of P3 type stacking faults within these P2 structures, as there are some sequential layers present that seem to be biased in opposite directions (corresponding to the P2 type stacking) and some where they appear biased in the same direction (which would correspond to P3 type stacking).

11.2.4.2 Oxidation State Estimation

Using the summed spectra for each layer, as described earlier, the Mn oxidation states were approximated using two calibration curves. The curve for each of the datasets is shown in Figure 11.12.

From Figure 11.12, it can be seen that the white line intensity curve provides

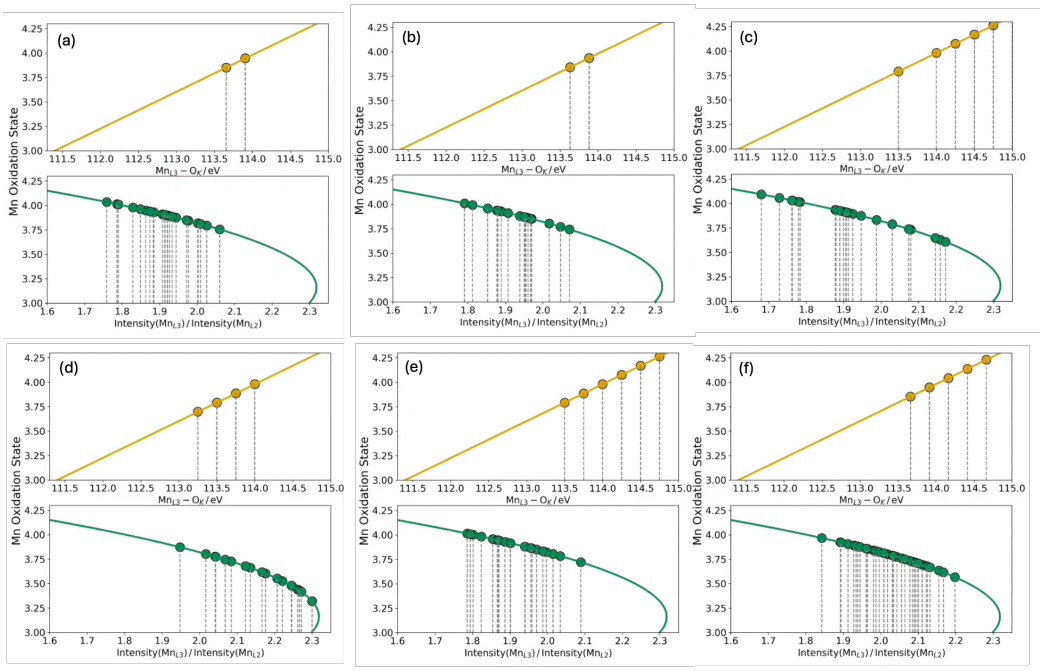


Figure 11.12: Calibration Curves and the corresponding location of each layer in the samples (a-f) from Figure 11.11

very plausible results, ranging between oxidation states of 3 and 4 with a sensible distribution across the layers. The comparison between the Mn L3 edge onset and the O K edge onset is more prone to over estimation, frequently predicting values above 4. Additionally, this measure is more dependent on the discrete energy steps sampled by the experimental data. This could be improved by estimating the peak maxima of interpolated data, but this would not significantly address the over estimation so it has been left as calculated.

The predictions in Figure 11.12 were used in conjunction with the neural network based predictor, MnEdgeNet to produce the final layer based predictions. Figures 11.13-11.18 show the results of combining these techniques, where in each case, the first plot uses only one of the MnEdgeNet models to calculate the oxidation state estimate and the second plot uses the average prediction of all the models for the estimation.

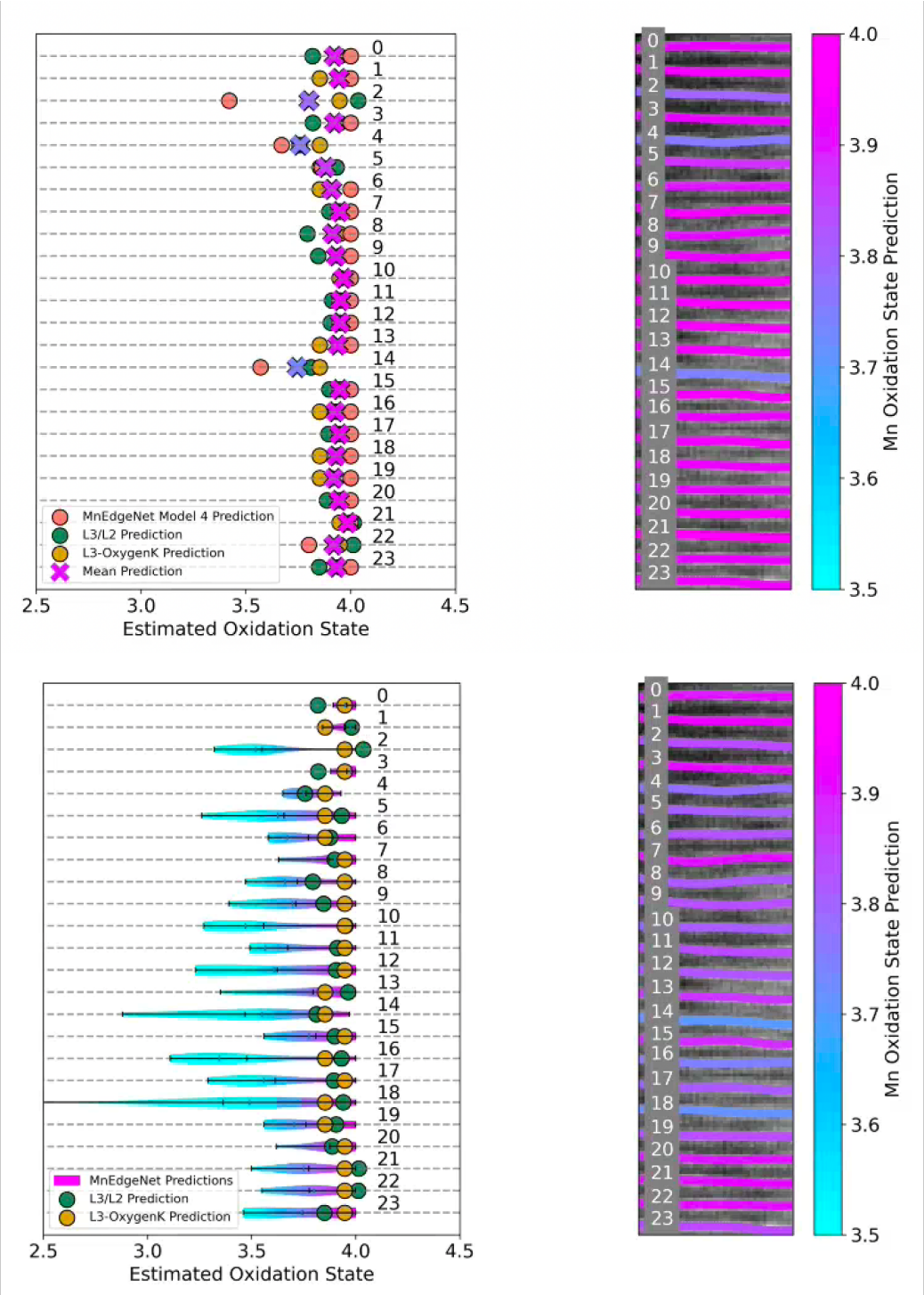


Figure 11.13: Oxidation state estimations per layer for Sample N25-1 ((a) in Figure 11.11) using the white line intensities, edge energy onsets and (Top) MnEdgeNet Model 4, (Bottom) all MnEdgeNet Models

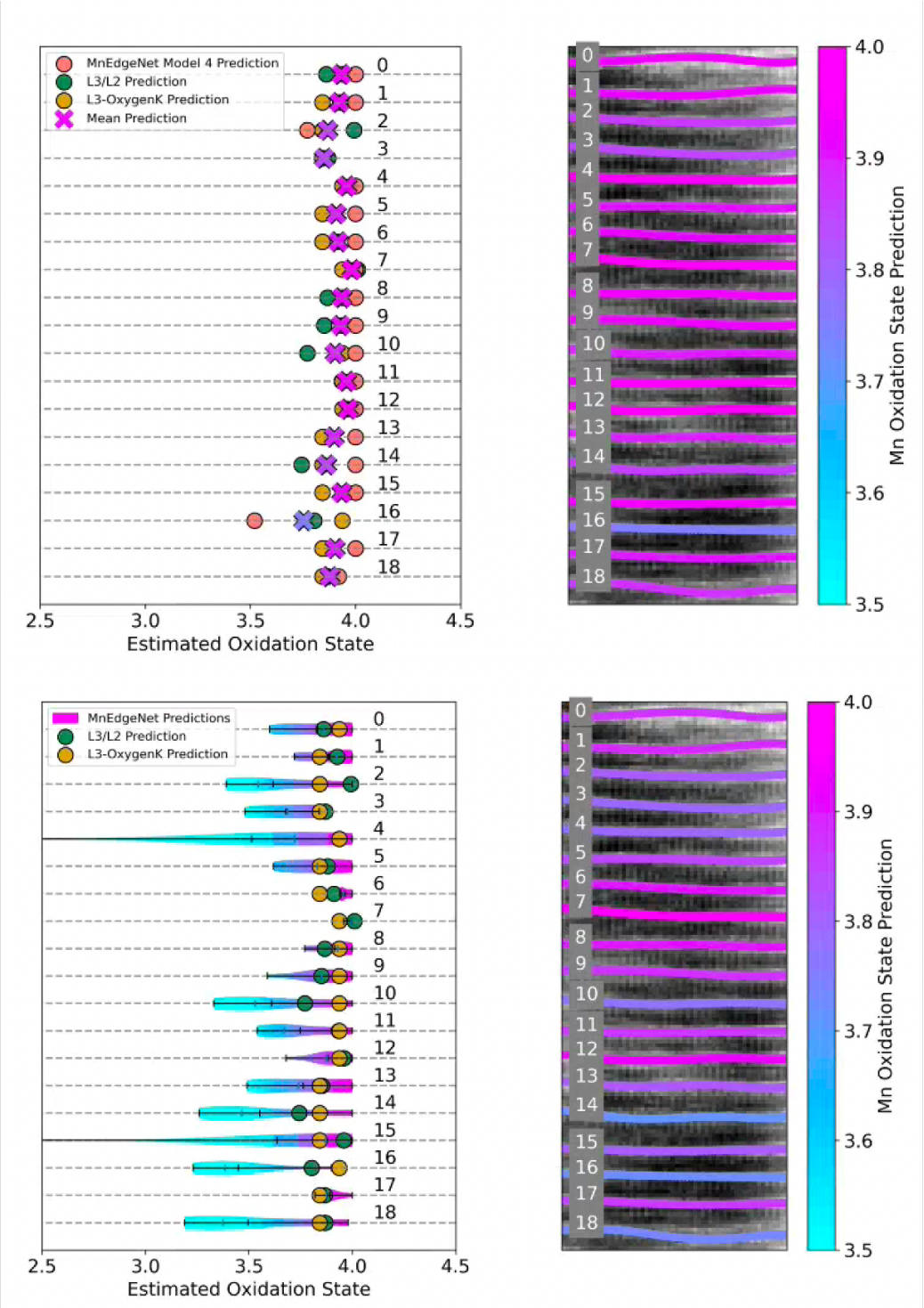


Figure 11.14: Oxidation state estimations per layer for Sample N25-2 ((b) in Figure 11.11) using the white line intensities, edge energy onsets and (Top) MnEdgeNet Model 4, (Bottom) all MnEdgeNet Models

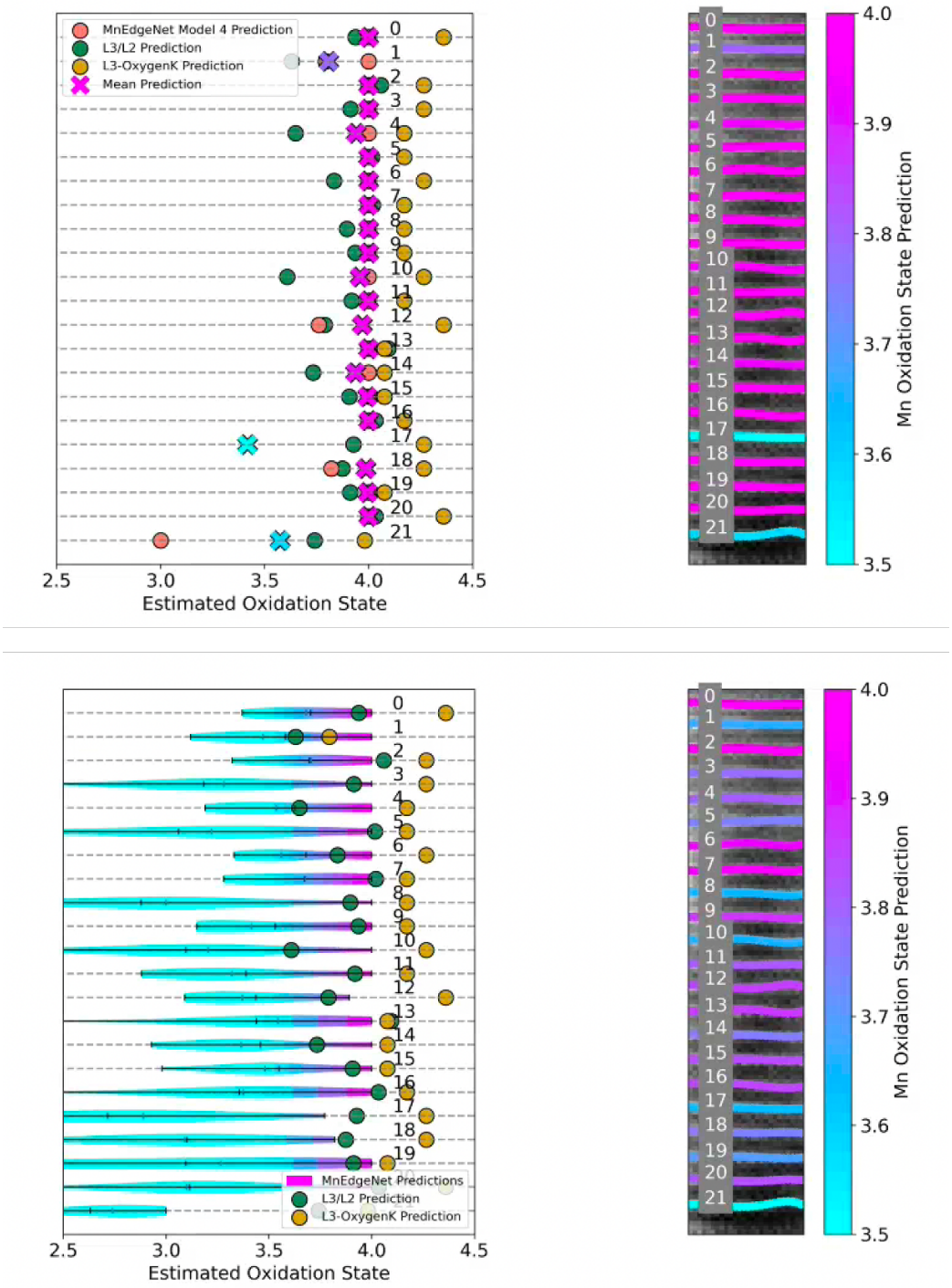


Figure 11.15: Oxidation state estimations per layer for Sample N25-3 ((c) in Figure 11.11) using the white line intensities, edge energy onsets and (Top) MnEdgeNet Model 4, (Bottom) all MnEdgeNet Models

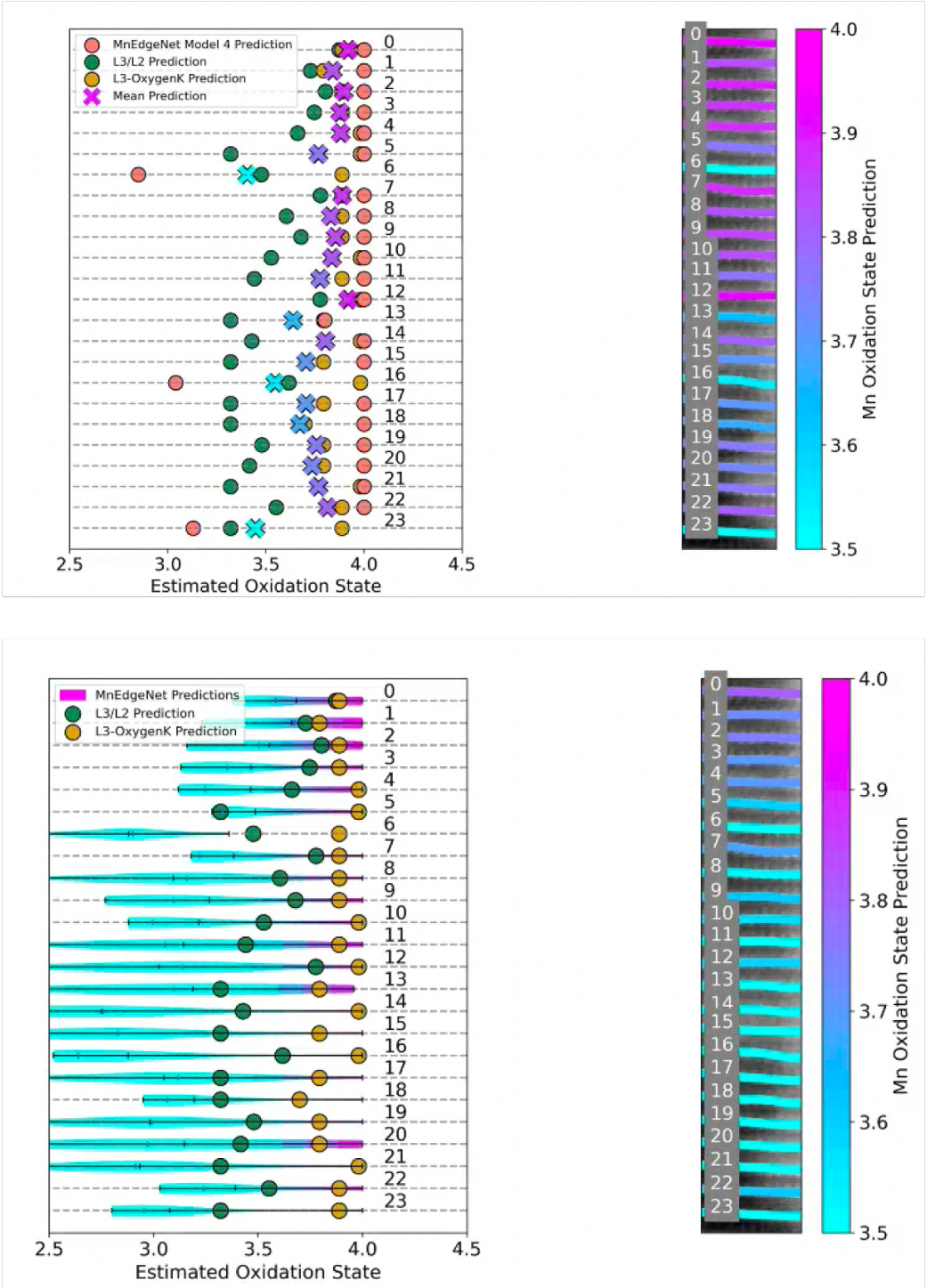


Figure 11.16: Oxidation state estimations per layer for Sample N25-4 ((d) in Figure 11.11) using the white line intensities, edge energy onsets and (Top) MnEdgeNet Model 4, (Bottom) all MnEdgeNet Models

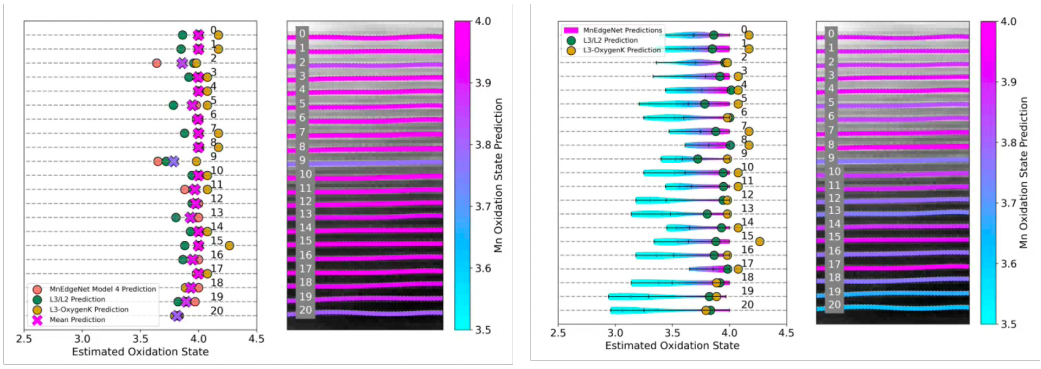


Figure 11.17: Oxidation state estimations per layer for Sample N30-1 ((e) in Figure 11.11) using the white line intensities, edge energy onsets and (Left) MnEdgeNet Model 4, (Right) all MnEdgeNet Models

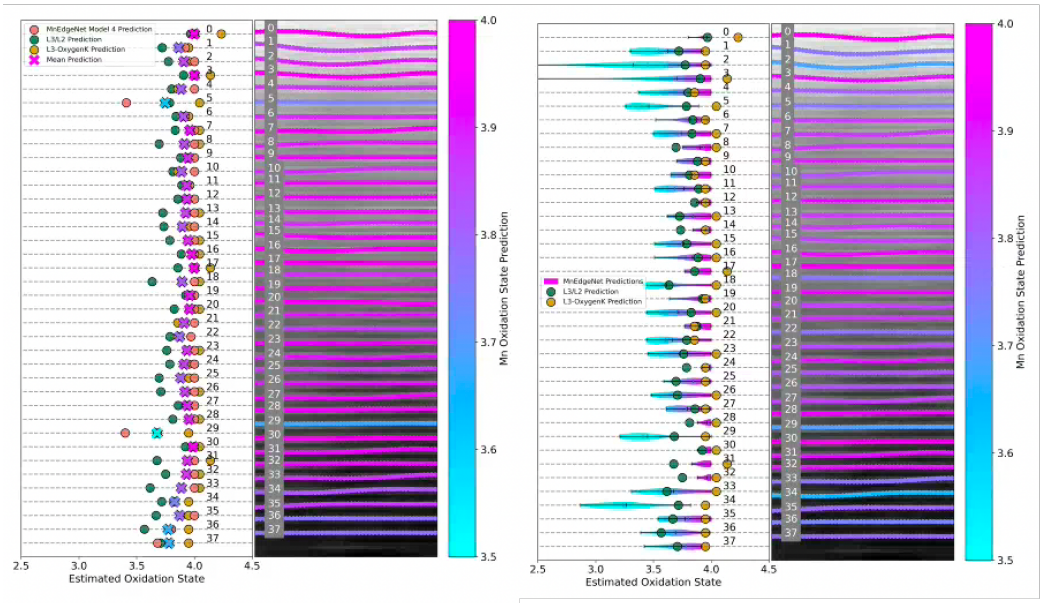


Figure 11.18: Oxidation state estimations per layer for Sample N30-2 ((f) in Figure 11.11) using the white line intensities, edge energy onsets and (Left) MnEdgeNet Model 4, (Right) all MnEdgeNet Models

These Figures show the disparity between the MnEdgeNet models, primarily through the broad distributions of predictions that can be seen in the second plot of each Figure. The result of this is that when all models are averaged, the oxidation state estimate is lower and the variance from layer to layer tends to increase.

The interpretation of the EELS results is challenging without more confidence in the estimations being provided. The least consistent predictions across the different methods is for samples N25-3 and N25-4 (Figures 11.15 and 11.16). This is due to the relatively small field of view horizontally, so fewer spectra are included in the final summed spectra for the layer. This results in lower signal to noise ratio and this can be seen to lead to more disparity in the estimation metrics, particularly across the different MnEdgeNet models. Another observation of these samples is they display the best resolution of atomic columns within each of the layers. This means that in the layer sum, there is sampling between the atomic columns which may correspond to reduced Mn signal and add further challenges to estimating the average oxidation state.

For the remaining samples, N25-1, N25-2, N30-1, and N30-2 (Figures 11.13, 11.14, 11.17 and 11.18) using including the single MnEdgeNet model in the final estimation shows fairly consistent oxidation state values. The distribution of layer oxidation states are shown in Figure 11.19.

The layer distributions are mostly consistent across the N30s and N25s despite the different compositions. The two N25 samples display more self-consistency in their distributions than the two N30s, although this is likely just a reflection of the relatively small sampling. The N25 sample also shows the lowest estimated oxidation state on average, which would be fit with the hypothesis of increased Mn^{3+} incorporation due to formally having the highest population of Mn.

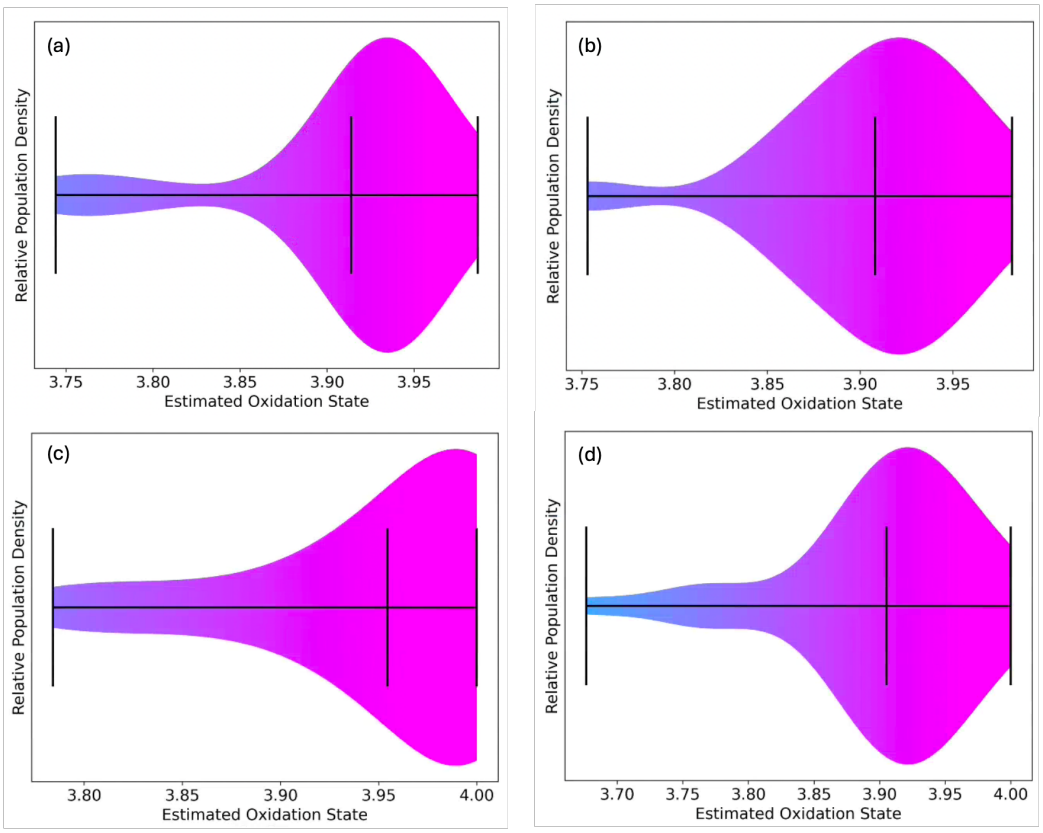


Figure 11.19: Distribution of Estimated Oxidation State for each layer in (a) Sample N25-1, (b) N25-2, (c) N30-1, and (d) N30-2.

11.2.5 Conclusions

Here we have shown how it is possible to investigate the structural stacking of these layered structures and how corresponding spectroscopy can be leveraged to provide some insight into the Mn environment and oxidation states within these layers. The investigation has highlighted the need for more statistical significance and sampling of more regions with more layers to draw robust conclusions about structure and composition. Our leveraging of pre-screening and SEND workflows to identify suitable candidate regions provide the framework required for this scale up in data collection. Additionally, this identifies the variation of using multiple methodologies to study Mn oxidation states and highlights concerns of reporting oxidation states using only a single approach.

12 | Conclusions and Future Work

The work in this thesis outlines how multi-modal approaches can be used to better understand the complex material systems of sodium ion cathode materials.

Within the thesis, a novel and widely applicable workflow for processing and analysing SEND datasets has been developed, investigated, and applied. The workflow created offers a methodology for studying local atomic structure, with a statistical appreciation for the range over which these structures are observed. Additionally, the same workflow has been demonstrated to provide a valuable component in the screening of samples for acquiring high quality atomic resolution STEM images. The open source and versatile nature of the workflow, along with the integration into the ePSIC computational architecture makes it applicable for many use cases beyond the specific materials investigated within this thesis.

Additionally to the development of this analysis workflow, tangible structural insight has been gained into the ionic orderings of the P2-type materials. Furthermore, the framework and methodologies have been created for extending these studies to analogous or even more complex systems. As part of these studies, a seemingly unreported sodium ion ordering scheme has been identified, which fits with the collected experimental data better than the currently reported model. This new model has been reported cautiously, with the understanding that even this model is likely incomplete, but is reported with multiple modalities of corroborating evidence.

As well as insight into the ionic orderings, this thesis has also outlined an extensive pipeline for screening, collecting and analysing spatially correlated valence state electronic studies of Mn species within layered systems such as these. This work has highlighted the difficulties of acquiring consistent measures of Mn va-

lence using widely reported results, encouraging caution when reporting values calculated by only a single approach.

12.1 Future Work

The work in this thesis can be continued by extending the application of the methodologies developed within or the scope of the material systems studied. One of the key motivations for the depths to which the structural modelling has been approached is to elucidate the best strategies for improving material application. Looking at the development of improved battery materials as an optimisation problem, in the almost infinite parameter space of possible structures, it becomes clear that the application of some form of random search strategy (analogous to varying a single element proportion or step in the synthesis protocol) is not an efficient approach.

Instead, the approach should be more akin to a Bayesian Optimisation Approach, where sample points in the parameter space are used to update an approximation of the underlying surface. That is to say, the better the understanding the effect that each parameter has upon the structures, the better ones approximation of the surface being optimised is, as a function of the parameter space. With a sufficiently good approximation of the surface, it is then straightforward to identify where optimal solutions will be.

To achieve this understanding of the relationship between material performance and parameters, it is imperative that the detail of the structural modelling is as accurate as possible. This is because it is the interplay of these subtle structural factors that ultimately drives the material performance and so it is these structural factors that need to be understood to achieve this accurate anticipation for novel materials.

The approaches to achieving detailed structural models developed within this thesis should offer a viable means of understanding the structural factors at play. If these modelling approaches were applied to a wide range of state-of-the-art cathode materials, the underlying structural factors driving performance differences might become more evident and this in turn would allow for directing the search for even better candidate materials.

Further applications of the VAE workflow to new systems should be straightforward as has already been discussed, due to the general nature of the design, but also the application of the ideas and approaches to novel SEND processing workflows could further enhance the efficacy of SEND as a key structural study technique. The workflow in its current state still has plenty of scope for refinement and improvement, not least as a result of the progress that has been made in machine learning research over the last couple of years.

A logical next step, as touched upon in Section 8.8, would be the incorporation of this in a broader autonomous SEND pipeline, where the microscope is run to automatically identify and collect samples and the workflow is used to identify individual domains within each sample. Aggregation of all the datasets for all particles on a TEM could create something analogous to a powder diffraction pattern, but a powder diffraction patterns that can be probed and studied alongside every 2D diffraction pattern it is composed off. Such an approach would be extremely valuable for more directly bridging the gap between nano-structure and bulk structure, and is only really computationally feasible when individaul datasets can be compressed down from many thousands of patterns into tens of domain averaged patterns.

One interesting route for investigation might be experimenting with changing out the convolutional filters in the VAE for something like the attention blocks re-

cently popularised for their use in large language models. The idea behind the attention blocks is that the same data can take on a different meaning depending on the context it is found within. This is used in language models to encode and distinguish between homographs (such as a ship's bow compared to an archer's bow). The information implied by a given bragg reflection is similarly dependent on the context of the other reflections observed. This is currently captured by successively downscaling the image and applying convolutional filters, however such a filter based approach is strongly influenced by the orientation of the pattern on the detector pixel grid. An attention based approach may prove robust to such changes and capture a more appropriate understanding of the data.

In order for such attention mechanisms to be successfully trained, large datasets are required, so such an approach would likely require training on vast simulated patterns rather than on the individual per-dataset approach. This in itself might offer a more attractive solution in the long term, with a pre-trained general purpose model offering potential for much faster iteration over datasets with reduced demands on computational resources. A potential limitation of an attention-based approach might be difficulty with amorphous regions (which are currently handled well) where there less correlation between data points.

Remaining closer to the workflow in its current state, modification of the loss function might offer an alternative manner for dealing with the current challenges of clustering the latent space. If the loss function could be altered to incentivize learning latent representations which are more receptive to traditional clustering approaches, then that might offer a more streamlined processing pipeline. The challenge with this approach is establishing a way to do this, that doesn't make the optimisation too much more challenging, as adding complexity into the loss function will likely lead to a less smooth cost surface.

The ionic ordering studies presented within were also designed and reported in a way that should allow for the logical extension to novel systems, with the empirical model especially developed with modularity of the underlying system in mind. Finding other analogous materials, with experimentally reported superstructures and applying these approaches, might yield similar insights to those gained for the P2 structure. Similarly, taking other target stoichiometries and other standard cathode phases and applying the HR-STEM pipeline would offer a reliable way of directly studying the stacking of the layers, while also allowing scope for studying spatially resolved oxidation state variation.

To gain further confidence in the 'disordered sodium ordering' scheme identified within, multislice-ptychography would make an excellent candidate for verifying the model derived. Beyond the corroboration of multiple techniques, as used here, a volumetric approach to the investigations would be the ideal for structures such as these cathodes, where the structural complexity and disorder extends along all directions. The current state of multislice-ptychography is probably not quite suitable, offering a depth resolution of around $5\,\mathrm{nm}$, but if that can be taken down to $0.5\,\mathrm{nm}$ it would be theoretically possible to collect direct experimental evidence that could be compared to the proposed structural model.

Beyond theses avenues for further explorations, there are likely countless other ways to expand upon the ideas and approaches developed here. With the applications of new techniques like SEND, and the ever evolving landscape of Machine Learning research, our scientific ability to characterise and optimise materials should only keep improving. The hope is that these improvements can be applied to the development and deployment of efficient, economical, energy storage devices and to swiftly facilitate the global move towards becoming a sustainable species.

Bibliography

[1] S. Arrhenius, "On the influence of carbonic acid in the air upon the temperature of the ground (1896)," in *The Future of Nature* (L. Robin, S. Sörlin, and P. Warde, eds.), pp. 303–315, Yale University Press, Dec. 2017.

[2] R. Monroe, "The Keeling Curve." https://keelingcurve.ucsd.edu, Jan. 2024.

[3] Intergovernmental Panel On Climate Change, *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 1 ed., July 2023.

[4] W. Cheng, L. Dan, X. Deng, J. Feng, Y. Wang, J. Peng, J. Tian, W. Qi, Z. Liu, X. Zheng, D. Zhou, S. Jiang, H. Zhao, and X. Wang, "Global monthly gridded atmospheric carbon dioxide concentrations under the historical and future scenarios," *Sci. Data*, vol. 9, pp. 83–96, Mar. 2022.

[5] Statista, "Dataset: Global CO_2 emissions by year 1940-2022." https://www.statista.com/statistics/276629/global-co2-emissions/, Dec. 2022.

[6] NASA, "Dataset: GISS Surface Temperature Analysis (v4): Analysis Graphs and Plots." https://data.giss.nasa.gov/gistemp/graphs_v4/, Dec. 2024.

[7] M. R. Allen and W. J. Ingram, "Constraints on future changes in climate and the hydrologic cycle," *Nature*, vol. 419, pp. 224–232, Sept. 2002.

[8] K. L. Findell, P. W. Keys, R. J. van der Ent, B. R. Lintner, A. Berg, and J. P. Krasting, "Rising Temperatures Increase Importance of Oceanic Evapora-

tion as a Source for Continental Precipitation," *J. Clim.*, vol. 32, pp. 7713–7726, Nov. 2019.

[9] K. L. Rasmussen, A. F. Prein, R. M. Rasmussen, K. Ikeda, and C. Liu, "Changes in the convective population and thermodynamic environments in convection-permitting regional climate simulations over the United States," *Clim. Dyn.*, vol. 55, pp. 383–408, July 2020.

[10] Y. Qing, S. Wang, Z.-L. Yang, and P. Gentine, "Soil moisture-atmosphere feedbacks have triggered the shifts from drought to pluvial conditions since 1980," *Commun. Earth Environ.*, vol. 4, pp. 1–10, July 2023.

[11] T. Kjellstrom, "Climate change, direct heat exposure, health and well-being in low and middle-income countries," *Glob. Health Action*, vol. 2, pp. 1–3, Mar. 2009.

[12] J. Liu, B. M. Varghese, A. Hansen, Y. Zhang, T. Driscoll, G. Morgan, K. Dear, M. Gourley, A. Capon, and P. Bi, "Heat exposure and cardiovascular health outcomes: A systematic review and meta-analysis," *The Lancet Planetary Health*, vol. 6, pp. 484–495, June 2022.

[13] M. B. Rice, W. Li, E. H. Wilker, D. R. Gold, J. Schwartz, A. Zanobetti, P. Koutrakis, I. Kloog, G. R. Washko, G. T. O'Connor, and M. A. Mittleman, "Association of outdoor temperature with lung function in a temperate climate," *Eur. Respir. J.*, vol. 53, pp. 612–622, Jan. 2019.

[14] CDC, "Government information website: Managing diabetes in the heat," *Centers for Disease Control and Prevention*, July 2022.

[15] S. Dasgupta, N. Van Maanen, S. N. Gosling, F. Piontek, C. Otto, and C.-F. Schleussner, "Effects of climate change on combined labour productiv-

ity and supply: An empirical, multi-model study," *The Lancet Planetary Health*, vol. 5, pp. 455–465, July 2021.

[16] J. Verschuur, E. E. Koks, and J. W. Hall, "Systemic risks from climate-related disruptions at ports," *Nat. Clim. Chang.*, vol. 13, pp. 804–806, Aug. 2023.

[17] A. F. S. Ribeiro, A. Russo, C. M. Gouveia, P. Páscoa, and J. Zscheischler, "Risk of crop failure due to compound dry and hot extremes estimated with nested copulas," *Biogeosciences*, vol. 17, pp. 4815–4830, Oct. 2020.

[18] R. Rojas, L. Feyen, and P. Watkiss, "Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation," *Glob. Environ. Change*, vol. 23, pp. 1737–1751, Dec. 2013.

[19] S. L. Gariano and F. Guzzetti, "Landslides in a changing climate," *Earth-Sci. Rev.*, vol. 162, pp. 227–252, Nov. 2016.

[20] A. M. Stewart-Ibarra, "Climate change and infectious diseases: Research and policy actions needed to address an inequitable health crisis," *One Earth*, vol. 5, pp. 333–335, Apr. 2022.

[21] C. Müller, W. A. Ouédraogo, M. Schwarz, S. Barteit, and R. Sauerborn, "The effects of climate change-induced flooding on harvest failure in Burkina Faso: Case study," *Front. Public Health*, vol. 11, pp. 1–11, Aug. 2023.

[22] A. Arnall, "A climate of control: Flooding, displacement and planned resettlement in the Lower Zambezi River valley, Mozambique," *Geog. J.*, vol. 180, pp. 141–150, Mar. 2014.

[23] O. Landeg, G. Whitman, K. Walker-Springett, C. Butler, A. Bone, and S. Kovats, "Coastal flooding and frontline health care services: Challenges

for flood risk resilience in the English health care system," *J. Health Serv. Res. Policy*, vol. 24, pp. 219–228, Oct. 2019.

[24] A. Pal, T. W. Tsusaka, T. P. L. Nguyen, and M. M. Ahmad, "Assessment of vulnerability and resilience of school education to climate-induced hazards: A review," *Dev. Stud. Res.*, vol. 10, pp. 1–16, Dec. 2023.

[25] P. Cianconi, S. Betrò, and L. Janiri, "The Impact of Climate Change on Mental Health: A Systematic Descriptive Review," *Front. Psychiatry*, vol. 11, pp. 1–16, Mar. 2020.

[26] M. Santini, S. Noce, M. Antonelli, and L. Caporaso, "Complex drought patterns robustly explain global yield loss for major crops," *Sci Rep*, vol. 12, pp. 5792–5809, Apr. 2022.

[27] R. Lal, "Climate Change and Soil Degradation Mitigation by Sustainable Management of Soils and Other Natural Resources," *Agric. Res.*, vol. 1, pp. 199–212, Sept. 2012.

[28] K. L. Ebi, J. Vanos, J. W. Baldwin, J. E. Bell, D. M. Hondula, N. A. Errett, K. Hayes, C. E. Reid, S. Saha, J. Spector, and P. Berry, "Extreme Weather and Climate Change: Population Health and Health System Implications," *Annu. Rev. Public Health*, vol. 42, pp. 293–315, Apr. 2021.

[29] L. C. Stringer, A. Mirzabaev, T. A. Benjaminsen, R. M. B. Harris, M. Jafari, T. K. Lissner, N. Stevens, and C. Tirado-von der Pahlen, "Climate change impacts on water security in global drylands," *One Earth*, vol. 4, pp. 851–864, June 2021.

[30] Y. Wang, Y. Xie, W. Dong, Y. Ming, J. Wang, and L. Shen, "Adverse effects of increasing drought on air quality via natural processes," *Atmos. Chem. Phys.*, vol. 17, pp. 12827–12843, Oct. 2017.

[31] K. Teather, N. Hogan, K. Critchley, M. Gibson, S. Craig, and J. Hill, "Examining the links between air quality, climate change and respiratory health in Qatar," *Avicenna*, vol. 2013, pp. 9–20, Sept. 2013.

[32] A. I. Filkov, T. Ngo, S. Matthews, S. Telfer, and T. D. Penman, "Impact of Australia's catastrophic 2019/20 bushfire season on communities and environment. Retrospective analysis and current trends," *J. Saf. Sci. Resil.*, vol. 1, pp. 44–56, Sept. 2020.

[33] P. T. Brown, H. Hanley, A. Mahesh, C. Reed, S. J. Strenfel, S. J. Davis, A. K. Kochanski, and C. B. Clements, "Climate warming increases extreme daily wildfire growth risk in California," *Nature*, vol. 621, pp. 760–766, Sept. 2023.

[34] F. N. Kamta, J. Schilling, and J. Scheffran, "Insecurity, Resource Scarcity, and Migration to Camps of Internally Displaced Persons in Northeast Nigeria," *Sustainability*, vol. 12, pp. 6830–6845, Jan. 2020.

[35] M. T. Klare, "Climate Change, Water Scarcity, and the Potential for Interstate Conflict in South Asia," *J. Strateg. Secur.*, vol. 13, no. 4, pp. 109–122, 2020.

[36] P. M. Regan and H. Kim, "Water scarcity, climate adaptation, and armed conflict: Insights from Africa," *Reg. Environ. Change*, vol. 20, pp. 129–149, Nov. 2020.

[37] UNFCCC, "International treaty: The Paris Agreement." https://unfccc.int/documents/184656, Nov. 2015.

[38] U. K. Government, "Government bulletin: UK becomes first major economy to pass net zero emissions law."

https://www.gov.uk/government/news/uk-becomes-first-major-economy-to-pass-net-zero-emissions-law, June 2019.

[39] U. K. Government, "Government bulletin: Plans unveiled to decarbonise UK power system by 2035." https://www.gov.uk/government/news/plans-unveiled-to-decarbonise-uk-power-system-by-2035, Oct. 2021.

[40] U. K. Government, "Government bulletin: UK enshrines new target in law to slash emissions by 78% by 2035." https://www.gov.uk/government/news/uk-enshrines-new-target-in-law-to-slash-emissions-by-78-by-2035, Apr. 2021.

[41] U. K. Government, "Uk legislation: The net zero transformation: Delivery, regulation and the consumer." https://committees.parliament.uk/publications/9141/documents/159550/default/, Mar. 2022.

[42] S. Wheatley, B. K. Sovacool, and D. Sornette, "Reassessing the safety of nuclear power," *Energy Res. Soc. Sci.*, vol. 15, pp. 96–100, May 2016.

[43] Y. Wu, Z. Chen, Z. Wang, S. Chen, D. Ge, C. Chen, J. Jia, Y. Li, M. Jin, T. Zhou, F. Wang, and L. Hu, "Nuclear safety in the unexpected second nuclear era," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, pp. 17673–17682, Sept. 2019.

[44] P. Lappi and J. Lintunen, "From cradle to grave? On optimal nuclear waste disposal," *Energy Econ.*, vol. 103, pp. 105556–105567, Nov. 2021.

[45] G. G. Goswami, U. Rahman, and M. Chowdhury, "Estimating the economic cost of setting up a nuclear power plant at Rooppur in Bangladesh," *Environ. Sci. Pollut. Res. Int.*, vol. 29, pp. 35073–35095, Jan. 2022.

[46] L. M. Miller and D. W. Keith, "Climatic Impacts of Wind Power," *Joule*, vol. 2, pp. 2618–2632, Dec. 2018.

[47] A. Jäger-Waldau, I. Kougias, N. Taylor, and C. Thiel, "How photovoltaics can contribute to GHG emission reductions of 55% in the EU by 2030," *Renew. Sustain. Energy Rev.*, vol. 126, pp. 109836–109844, July 2020.

[48] Y. Zhou, J. Liu, W. Ge, C. He, J. Ma, and S. Tao, "Toward carbon neutrality: Projecting a desert-based photovoltaic power network circumnavigating the globe," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 2, pp. 1–10, Apr. 2023.

[49] M. Barooni, T. Ashuri, D. Velioglu Sogut, S. Wood, and S. Ghaderpour Taleghani, "Floating Offshore Wind Turbines: Current Status and Future Prospects," *Energies*, vol. 16, pp. 1–28, Jan. 2023.

[50] M. J. Burke and J. C. Stephens, "Political power and renewable energy futures: A critical review," *Energy Res. Soc. Sci.*, vol. 35, pp. 78–93, Jan. 2018.

[51] LondonDatastore, "Dataset: Photovoltaic (PV) Solar Panel Energy Generation data." https://data.london.gov.uk/dataset/photovoltaic–pv–solar-panel-energy-generation-data, Dec. 2013.

[52] A. Sánchez, Q. Zhang, M. Martín, and P. Vega, "Towards a new renewable power system using energy storage: An economic and social analysis," *Energy Convers. and Manag.*, vol. 252, pp. 115056–115069), Jan. 2022.

[53] W.-P. Schill, "Electricity Storage and the Renewable Energy Transition," *Joule*, vol. 4, pp. 2059–2064, Oct. 2020.

[54] M. Y. Worku, "Recent Advances in Energy Storage Systems for Renewable Source Grid Integration: A Comprehensive Review," *Sustainability*, vol. 14, pp. 5985–6003), Jan. 2022.

[55] M. S. Ziegler, J. M. Mueller, G. D. Pereira, J. Song, M. Ferrara, Y.-M. Chiang, and J. E. Trancik, "Storage Requirements and Costs of Shaping Renewable Energy Toward Grid Decarbonization," *Joule*, vol. 3, pp. 2134–2153, Sept. 2019.

[56] J. Mitali, S. Dhinakaran, and A. A. Mohamad, "Energy storage systems: A review," *Energy Storage and Saving*, vol. 1, pp. 166–216, Sept. 2022.

[57] J. Twitchell, K. DeSomber, and D. Bhatnagar, "Defining long duration energy storage," *J. Energy Storage*, vol. 60, pp. 105787–105794), Apr. 2023.

[58] A. A. Kebede, T. Kalogiannis, J. Van Mierlo, and M. Berecibar, "A comprehensive review of stationary energy storage devices for large scale renewable energy sources grid integration," *Renew. Sustain. Energy Rev.*, vol. 159, pp. 112213–112232), May 2022.

[59] M. R. Chakraborty, S. Dawn, P. K. Saha, J. B. Basu, and T. S. Ustun, "A Comparative Review on Energy Storage Systems and Their Application in Deregulated Systems," *Batteries*, vol. 8, pp. 124–152, Sept. 2022.

[60] P. Albertus, J. S. Manser, and S. Litzelman, "Long-Duration Electricity Storage Applications, Economics, and Technologies," *Joule*, vol. 4, pp. 21–32, Jan. 2020.

[61] S. G. Nnabuife, E. Oko, B. Kuang, A. Bello, A. P. Onwualu, S. Oyagha, and J. Whidborne, "The prospects of hydrogen in achieving net zero emissions by 2050: A critical review," *Sustainable Chemistry for Climate Action*, vol. 2, pp. 100024–100041), Jan. 2023.

[62] M. Yang, R. Hunger, S. Berrettoni, B. Sprecher, and B. Wang, "A review of hydrogen storage and transport technologies," *Clean Energy*, vol. 7, pp. 190–216, Feb. 2023.

[63] S. Giddey, S. P. S. Badwal, C. Munnings, and M. Dolan, "Ammonia as a Renewable Energy Transportation Media," *ACS Sustainable Chem. Eng.*, vol. 5, pp. 10231–10239, Nov. 2017.

[64] K. Machaj, J. Kupecki, Z. Malecha, A. W. Morawski, M. Skrzypkiewicz, M. Stanclik, and M. Chorowski, "Ammonia as a potential marine fuel: A review," *Energy Strategy Rev.*, vol. 44, pp. 100926–100945, Nov. 2022.

[65] A. E. Yüzbaşıoğlu, C. Avşar, and A. O. Gezerman, "The current situation in the use of ammonia as a sustainable energy source and its industrial potential," *Curr. Res. Green Sustain. Chem.*, vol. 5, pp. 100307–100312, Jan. 2022.

[66] Statista, "Dataset: Energy storage global cumulative deployment 2031." https://www.statista.com/statistics/728652/projected-energy-storage-market-capacity-deployment-globally/, Dec. 2023.

[67] K. Mizushima, P. C. Jones, P. J. Wiseman, and J. B. Goodenough, "Li_xCoO_2 ($0 < x < 1$): A new cathode material for batteries of high energy density," *Mater. Res. Bull.*, vol. 15, pp. 783–789, June 1980.

[68] Statista, "Dataset: Lithium global reserves top countries 2023." https://www.statista.com/statistics/268790/countries-with-the-largest-lithium-reserves-worldwide/, Dec. 2023.

[69] W. P. Review, "Dataset: Cobalt Reserves by Country 2024." https://worldpopulationreview.com/country-rankings/cobalt-reserves-by-country, May 2024.

[70] B. K. Sovacool, "When subterranean slavery supports sustainability transitions? power, patriarchy, and child labor in artisanal Congolese cobalt mining," *Extr. Ind. Soc.*, vol. 8, pp. 271–293, Mar. 2021.

[71] P. Li, N. Hu, J. Wang, S. Wang, and W. Deng, "Recent Progress and Perspective: Na Ion Batteries Used at Low Temperatures," *Nanomaterials (Basel)*, vol. 12, pp. 3529–3558, Oct. 2022.

[72] K. Wang, P. Yan, and M. Sui, "Phase transition induced cracking plaguing layered cathode for sodium-ion battery," *Nano Energy*, vol. 54, pp. 148–155, Dec. 2018.

[73] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*. New York: Springer, 2 ed., July 2009.

[74] M. M. Ripoll and F. H. Cano, *Crystallography in a Nutshell*. Jan. 2003.

[75] G. Sanderson, "Video lectures: Neural Networks." https://www.3blue1brown.com/topics/neural-networks, Oct. 2017.

[76] M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[77] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and t. scikit-image contributors, "Scikit-image: Image processing in Python," *PeerJ*, vol. 2, pp. 453–481, June 2014.

[78] M. Abadi, A. Agarwal, P. Barham, and E. Brevdo, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," Mar. 2005.

[79] A. Bridger, W. I. F. David, T. J. Wood, M. Danaie, and K. T. Butler, "Versatile domain mapping of scanning electron nanobeam diffraction datasets utilising variational autoencoders," *NPJ Comput. Mater.*, vol. 9, pp. 1–18, Jan. 2023.

[80] C. Delmas, C. Fouassier, and P. Hagenmuller, "Structural classification and properties of the layered oxides," *Physica B+C*, vol. 99, pp. 81–85, Jan. 1980.

[81] Z. Lu and J. R. Dahn, "In Situ X-Ray Diffraction Study of P2-$Na_{2/3}Mn_{2/3}Ni_{1/3}O_2$," *J. Electrochem. Soc.*, vol. 148, no. 11, pp. 1225–1229, 2001.

[82] C. Delmas, J.-J. Braconnier, C. Fouassier, and P. Hagenmuller, "Electrochemical intercalation of sodium in Na_xCoO_2 bronzes," *Solid State Ion.*, vol. 3–4, pp. 165–169, Aug. 1981.

[83] J. M. Brittain, *Structural and Electrochemical Analyses of a P2|O3 Mixed-Phase Sodium-Ion Positive Electrode Active Material*. PhD thesis, University of Oxford, Nov. 2021.

[84] P.-F. Wang, Y. You, Y.-X. Yin, and Y.-G. Guo, "Layered Oxide Cathodes for Sodium-Ion Batteries: Phase Transition, Air Stability, and Performance," *Adv. Energy Mater.*, vol. 8, no. 8, pp. 1701912–1701935, 2018.

[85] A. Gutierrez, W. M. Dose, O. Borkiewicz, F. Guo, M. Avdeev, S. Kim, T. T. Fister, Y. Ren, J. Bareño, and C. S. Johnson, "On Disrupting the Na+-Ion/Vacancy Ordering in P2-Type Sodium-Manganese-Nickel Oxide Cathodes for Na-Ion Batteries," *J. Phys. Chem. C*, vol. 122, pp. 23251–23260, Sept. 2018.

[86] C. Delmas, "Sodium and Sodium-Ion Batteries: 50 Years of Research," *Adv. Energy Mater.*, vol. 8, pp. 1703137–1703148, Feb. 2018.

[87] S. Guo, Y. Sun, J. Yi, K. Zhu, P. Liu, Y. Zhu, G.-z. Zhu, M. Chen, M. Ishida, and H. Zhou, "Understanding sodium-ion diffusion in layered P2 and P3 oxides via experiments and first-principles calculations: A bridge

between crystal structure and electrochemical performance," *NPG Asia Mater.*, vol. 8, pp. 266–275, Apr. 2016.

[88] S. P. Ong, V. L. Chevrier, G. Hautier, A. Jain, C. Moore, S. Kim, X. Ma, and G. Ceder, "Voltage, stability and diffusion barrier differences between sodium-ion and lithium-ion intercalation materials," *Energy Environ. Sci.*, vol. 4, pp. 3680–3688, June 2011.

[89] E. Lee, J. Lu, Y. Ren, X. Luo, X. Zhang, J. Wen, D. Miller, A. DeWahl, S. Hackney, B. Key, D. Kim, M. D. Slater, and C. S. Johnson, "Layered P2/O3 Intergrowth Cathode: Toward High Power Na-Ion Batteries," *Adv. Energy Mater.*, vol. 4, pp. 1400458–1400466, July 2014.

[90] Q. Bai, L. Yang, H. Chen, and Y. Mo, "Computational Studies of Electrode Materials in Sodium-Ion Batteries," *Adv. Energy Mater.*, vol. 8, pp. 1702998–1703027, Mar. 2018.

[91] M. H. Han, E. Gonzalo, G. Singh, and T. Rojo, "A comprehensive review of sodium layered oxides: Powerful cathodes for Na-ion batteries," *Energy Environ. Sci.*, vol. 8, pp. 81–102, Dec. 2014.

[92] P. F. Wang, H. R. Yao, X. Y. Liu, Y. X. Yin, J. N. Zhang, Y. Wen, X. Yu, L. Gu, and Y. G. Guo, "Na+/vacancy disordering promises high-rate Na-ion batteries," *Sci. Adv.*, vol. 4, pp. 1–10, Mar. 2018.

[93] Y. S. Meng, Y. Hinuma, and G. Ceder, "An investigation of the sodium patterning in Na_xCoO_2 ($0.5 \leq x \leq 1$) by density functional theory methods," *J. Chem. Phys.*, vol. 128, pp. 1–8, Mar. 2008.

[94] X. Cao and M. Dolg, "Pseudopotentials and modelpotentials," *Comp. Mol. Sci.*, vol. 1, pp. 200–210, Feb. 2011.

[95] Y. Shi, P. Jiang, S. Wang, W. Chen, B. Wei, X. Lu, G. Qian, W. H. Kan, H. Chen, W. Yin, Y. Sun, and X. Lu, "Slight compositional variation-induced structural disorder-to-order transition enables fast Na+ storage in layered transition metal oxides," *Nat. Commun.*, vol. 13, pp. 7888–7900, Dec. 2022.

[96] H.-R. Yao, W.-J. Lv, X.-G. Yuan, Y.-J. Guo, L. Zheng, X.-A. Yang, J. Li, Y. Huang, Z. Huang, P.-F. Wang, and Y.-G. Guo, "New insights to build Na+/vacancy disordering for high-performance P2-type layered oxide cathodes," *Nano Energy*, vol. 97, pp. 107207–107216, June 2022.

[97] J. W. Somerville, A. Sobkowiak, N. Tapia-Ruiz, J. Billaud, J. G. Lozano, R. A. House, L. C. Gallington, T. Ericsson, L. Häggström, M. R. Roberts, U. Maitra, and P. G. Bruce, "Nature of the "Z"-phase in layered Na-ion battery cathodes," *Energy Environ. Sci.*, vol. 12, pp. 2223–2232, May 2019.

[98] E. N. Bassey, I. D. Seymour, J. D. Bocarsly, D. A. Keen, G. Pintacuda, and C. P. Grey, "Superstructure and Correlated Na+ Hopping in a Layered Mg-Substituted Sodium Manganate Battery Cathode are Driven by Local Electroneutrality," *Chem. Mater.*, vol. 35, pp. 10564–10583, Dec. 2023.

[99] Z. Lu, R. A. Donaberger, and J. R. Dahn, "Superlattice Ordering of Mn, Ni, and Co in Layered Alkali Transition Metal Oxides with P2, P3, and O3 Structures," *Chem. Mater.*, vol. 12, pp. 3583–3590, Dec. 2000.

[100] E. Gabriel, Z. Wang, V. V. Singh, K. Graff, J. Liu, C. Koroni, D. Hou, D. Schwartz, C. Li, J. Liu, X. Guo, N. C. Osti, S. P. Ong, and H. Xiong, "Influence of Interlayer Cation Ordering on Na Transport in P2-type $Na_{0.67-x}Li_yNi_{0.33-z}Mn_{0.67+z}O_2$ for Sodium-Ion Batteries," *J. Am. Chem. Soc.*, pp. 15108–15118, May 2024.

[101] N. Yabuuchi, M. Kajiyama, J. Iwatate, H. Nishikawa, S. Hitomi, R. Okuyama, R. Usui, Y. Yamada, and S. Komaba, "P2-type $Na_xFe_{1/2}Mn_{1/2}O_2$ made from earth-abundant elements for rechargeable Na batteries," *Nat. Mater.*, vol. 11, pp. 512–517, June 2012.

[102] L. F. Pfeiffer, N. Jobst, C. Gauckler, M. Lindén, M. Marinaro, S. Passerini, M. Wohlfahrt-Mehrens, and P. Axmann, "Layered P2-$Na_xMn_{3/4}Ni_{1/4}O_2$ cathode materials for sodium-ion batteries: Synthesis, Electrochemistry and Influence of Ambient Storage," *Front. Energy Res.*, vol. 10, pp. 1–17, May 2022.

[103] J. M. Paulsen and J. R. Dahn, "Studies of the layered manganese bronzes, $Na_{2/3}Mn_{1-x}MxO_2$ with M=Co, Ni, Li, and $Li_{2/3}Mn_{1-x}M_xO_2$ prepared by ion-exchange," *Solid State Ion.*, vol. 126, pp. 3–24, Nov. 1999.

[104] Y.-T. Tu, "Ray optics simulation software." 10.5281/zenodo.10547599, Jan. 2024.

[105] W. S. Mcculloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, pp. 115–133, Dec. 1943.

[106] N. M. Nawi, W. H. Atomi, and M. Z. Rehman, "The Effect of Data Preprocessing on Optimized Training of Artificial Neural Networks," *Procedia Tech.*, vol. 11, pp. 32–39, Jan. 2013.

[107] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Adv. Neural Inf. Process. Syst.*, vol. 25, Curran Associates, Inc., May 2012.

[108] T. Szandała, "Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks," in *Bio-Inspired Neurocomputing*

(A. K. Bhoi, P. K. Mallick, C.-M. Liu, and V. E. Balas, eds.), pp. 203–224, Singapore: Springer, Jan. 2021.

[109] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," *IEEE Trans. Comput. Imaging*, vol. 3, pp. 47–57, Mar. 2017.

[110] S. H. Haji and A. M. Abdulazeez, "Comparison of optimization techniques based on gradient descent algorithm: a review," *PalArch's J. Archaeol. Egypt/ Egyptol.*, vol. 18, pp. 2715–2743, Feb. 2021.

[111] T. Yu and H. Zhu, "Hyper-Parameter Optimization: A Review of Algorithms and Applications," *arXiv*, pp. 1–56, Mar. 2020.

[112] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A.-L. Boulesteix, D. Deng, and M. Lindauer, "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," *WIREs Data. Mining. Knowl. Discov.*, vol. 13, pp. 1484–1527, Nov. 2023.

[113] R. Chauhan, K. K. Ghanshala, and R. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pp. 278–282, Dec. 2018.

[114] M. A. Hossain and M. Sajib, "Classification of Image using Convolutional Neural Network (CNN)," *Glob. J. Comp. Sci. Tech.*, vol. 19, pp. 13–18, May 2019.

[115] B. Chidester, T. Zhou, M. N. Do, and J. Ma, "Rotation equivariant and invariant neural networks for microscopy image analysis," *Bioinformatics*, vol. 35, pp. 530–537, July 2019.

[116] P. Gallinari, Y. Lecun, S. Thiria, and F. Fogelman Soulie, "Memoires associatives distribuees: Une comparaison (Distributed associative memories: A comparison)," *Proceedings of COGNITIVA 87, Paris, La Villette, May 1987*, 1987.

[117] Y. Lecun, *PhD Thesis: Modeles Connexionnistes de l'apprentissage (Connectionist Learning Models)*. Universite P. et M. Curie (Paris 6), June 1987.

[118] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *MIT Press*, vol. 1, pp. 318–362, Aug. 1986.

[119] P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel, "The Helmholtz machine.," *Neural Comput.*, vol. 7, pp. 889–904, Sept. 1995.

[120] H. M. Rietveld, "Line profiles of neutron powder-diffraction peaks for structure refinement," *Acta Cryst.*, vol. 22, pp. 151–152, Jan. 1967.

[121] B. H. Toby, "Rietveld refinement," *J. Appl. Crystallogr.*, vol. H, pp. 465–472, Apr. 2019.

[122] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, vol. 2, pp. 559–572, Nov. 1901.

[123] M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori, and T. Tanji, "Sparse modeling of EELS and EDX spectral imaging data by nonnegative matrix factorization," *Ultramicroscopy*, vol. 170, pp. 43–59, Nov. 2016.

[124] B. R. Jany, A. Janas, and F. Krok, "Retrieving the Quantitative Chemical Information at Nanoscale from Scanning Electron Microscope Energy Dis-

persive X-ray Measurements by Machine Learning," *Nano Lett.*, vol. 17, pp. 6520–6525, Nov. 2017.

[125] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *J. R. Stat. Soc. Ser. C*, vol. 28, pp. 100–108, Jan. 1979.

[126] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference And Applications To Clustering*. J. Am. Stat. Assoc., Jan. 1988.

[127] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, (Portland, Oregon), pp. 226–231, AAAI Press, Aug. 1996.

[128] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-Based Clustering Based on Hierarchical Density Estimates," in *Advances in Knowledge Discovery and Data Mining* (J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, eds.), (Berlin, Heidelberg), pp. 160–172, Springer, Apr. 2013.

[129] R. E. Burkard and F. Rendl, "A thermodynamically motivated simulation procedure for combinatorial optimization problems," *Eur. J. Oper. Res.*, vol. 17, pp. 169–174, Aug. 1984.

[130] S. P. Thompson, J. E. Parker, J. Potter, T. P. Hill, A. Birt, T. M. Cobb, F. Yuan, and C. C. Tang, "Beamline I11 at Diamond: A new instrument for high resolution powder diffraction," *Rev. Sci. Instrum.*, vol. 80, pp. 075107–075107, July 2009.

[131] C. Ophus, P. Ercius, M. Sarahan, C. Czarnik, and J. Ciston, "Recording and using 4D-STEM datasets in materials science," *Microsc. Microanal.*, vol. 20, pp. 62–63, Aug. 2014.

[132] A. A. Bunaciu, E. gabriela Udriştioiu, and H. Y. Aboul-Enein, "X-Ray Diffraction: Instrumentation and Applications," *Crit. Rev. Anal. Chem.*, vol. 45, pp. 289–299, May 2015.

[133] A. K. Cheetham and J. C. Taylor, "Profile analysis of powder neutron diffraction data: Its scope, limitations, and applications in solid state chemistry," *J. Solid State Chem.*, vol. 21, pp. 253–275, Jan. 1977.

[134] E. M. James and N. D. Browning, "Practical aspects of atomic resolution imaging and analysis in STEM," *Ultramicroscopy*, vol. 78, pp. 125–139, Dec. 1998.

[135] I. MacLaren, T. A. MacGregor, C. S. Allen, and A. I. Kirkland, "Detectors-The ongoing revolution in scanning transmission electron microscopy and why this important to material characterization," *APL Mater.*, vol. 8, pp. 1109101–1109114, Nov. 2020.

[136] C. Ophus, "Four-Dimensional Scanning Transmission Electron Microscopy (4D-STEM): From Scanning Nanodiffraction to Ptychography and Beyond," *Microsc. Microanal.*, pp. 563–582, Mar. 2019.

[137] G. Brunetti, D. Robert, P. Bayle-Guillemaud, J. L. Rouvière, E. F. Rauch, J. F. Martin, J. F. Colin, F. Bertin, and C. Cayron, "Confirmation of the domino-cascade model by $LiFePO_4$/$FePO_4$ precession electron diffraction," *Chem. Mater.*, vol. 23, pp. 4515–4524, Sept. 2011.

[138] A. Kobler, A. Kashiwar, H. Hahn, and C. Kübel, "Combination of in situ straining and ACOM TEM: A novel method for analysis of plastic deformation of nanocrystalline metals," *Ultramicroscopy*, vol. 128, pp. 68–81, Jan. 2013.

[139] M. Gallagher-Jones, C. Ophus, K. C. Bustillo, D. R. Boyer, O. Panova, C. Glynn, C. T. Zee, J. Ciston, K. C. Mancia, A. M. Minor, and J. A. Rodriguez, "Nanoscale mosaicity revealed in peptide microcrystals by scanning electron nanodiffraction," *Commun. Biol.*, vol. 2, pp. 1–8, Jan. 2019.

[140] A. Nalin Mehta, N. Gauquelin, M. Nord, A. Orekhov, H. Bender, D. Cerbu, J. Verbeeck, and W. Vandervorst, "Unravelling stacking order in epitaxial bilayer MX2using 4D-STEM with unsupervised learning," *Nanotechnology*, vol. 31, pp. 1–8, Aug. 2020.

[141] G. C. Ortiz, M. Zhu, L. Dou, and J. Hwang, "4D-STEM Quantification of Nanoscale Ordered Domains in Organic Semiconducting Polymers," *Microsc. Microanal.*, vol. 26, pp. 1740–1742, July 2020.

[142] S. V. Kalinin, M. P. Oxley, M. Valleti, J. Zhang, R. P. Hermann, H. Zheng, W. Zhang, G. Eres, R. K. Vasudevan, and M. Ziatdinov, "Deep Bayesian local crystallography," *NPJ Comput. Mater.*, vol. 7, pp. 181–193, Nov. 2021.

[143] C. Ophus, S. E. Zeltmann, A. Bruefach, A. Rakowski, B. H. Savitzky, A. M. Minor, and M. C. Scott, "Automated Crystal Orientation Mapping in py4DSTEM using Sparse Correlation Matching," *Microsc. Microanal.*, vol. 28, pp. 390–403, Nov. 2022.

[144] N. Cautaerts, P. Crout, H. W. Ånes, E. Prestat, J. Jeong, G. Dehm, and C. H. Liebscher, "Free, flexible and fast: Orientation mapping using the multi-core and GPU-accelerated template matching capabilities in the Python-based open source 4D-STEM analysis toolbox Pyxem," *Ultramicroscopy*, vol. 237, pp. 113517–113527, Mar. 2022.

[145] T. Meyer, B. Kressdorf, V. Roddatis, J. Hoffmann, C. Jooss, and M. Seibt, "Phase Transitions in a Perovskite Thin Film Studied by Environmental In

Situ Heating Nano-Beam Electron Diffraction," *Small Methods*, vol. 5, July 2021.

[146] H. Liu, W. Deng, X. Gao, J. Chen, S. Yin, L. Yang, G. Zou, H. Hou, and X. Ji, "Manganese-based layered oxide cathodes for sodium ion batteries," *Nano Select*, vol. 1, pp. 200–225, June 2020.

[147] A. Valery, E. F. Rauch, L. ClÉment, and F. Lorut, "Retrieving overlapping crystals information from TEM nano-beam electron diffraction patterns," *J. Microsc.*, vol. 268, pp. 208–218, June 2017.

[148] R. Vincent and P. A. Midgley, "Double conical beam-rocking system for measurement of integrated electron diffraction intensities," *Ultramicroscopy*, vol. 53, pp. 271–282, Oct. 1994.

[149] A. Bruefach, C. Ophus, and M. C. Scott, "Analysis of Interpretable Data Representations for 4D-STEM Using Unsupervised Learning," *Microsc. Microanal.*, vol. 28, pp. 1998–2008, Dec. 2022.

[150] C. Shi, M. C. Cao, S. M. Rehn, S.-H. Bae, J. Kim, M. R. Jones, D. A. Muller, and Y. Han, "Uncovering Material Deformations via Machine Learning Combined with Four-Dimensional Scanning Transmission Electron Microscopy," *NPJ Comput. Mater.*, vol. 8, p. 114, May 2021.

[151] A. Bridger, "Open-source software: Stemsegment." https://github.com/andy-bridger/stemsegment, Oct. 2022.

[152] Z. Wang and A. C. Bovik, "Mean Squared Error : Love It or Leave It ?," *IEEE Signal Process. Mag.*, pp. 98–117, Jan. 2009.

[153] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, Apr. 2004.

[154] T. D. Hien, D. V. Tuan, P. V. At, and L. H. Son, "Novel algorithm for non-negative matrix factorization," *New Math. Nat. Comput.*, vol. 11, pp. 121–133, July 2015.

[155] J. K. Sunde, *The Effect of Elevated Temperatures on Precipitation in Aluminium Alloys*. PhD thesis, Norwegian University of Science and Technology, Sept. 2020.

[156] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. tech. j.*, vol. 27, pp. 623–656, Oct. 1948.

[157] M. Slouf, R. Skoupy, E. Pavlova, and V. Krzyzanek, "High resolution powder electron diffraction in scanning electron microscopy," *Materials*, vol. 14, pp. 7550–7572, Dec. 2021.

[158] B. H. Savitzky, S. E. Zeltmann, L. A. Hughes, H. G. Brown, S. Zhao, P. M. Pelz, T. C. Pekin, E. S. Barnard, J. Donohue, L. Rangel Dacosta, E. Kennedy, Y. Xie, M. T. Janish, M. M. Schneider, P. Herring, C. Gopal, A. Anapolsky, R. Dhall, K. C. Bustillo, P. Ercius, M. C. Scott, J. Ciston, A. M. Minor, and C. Ophus, "Py4DSTEM: A Software Package for Four-Dimensional Scanning Transmission Electron Microscopy Data Analysis," *Microsc. Microanal.*, vol. 27, pp. 712–743, May 2021.

[159] F. I. Allen, T. C. Pekin, A. Persaud, S. J. Rozeveld, G. F. Meyers, J. Ciston, C. Ophus, and A. M. Minor, "Fast Grain Mapping with Sub-Nanometer Resolution Using 4D-STEM with Grain Classification by Principal Component Analysis and Non-Negative Matrix Factorization," *Microsc. Microanal.*, vol. 27, pp. 794–803, May 2021.

[160] "Computing Resource: Diamond Light Source Jupyterhub." https://jupyterhub.diamond.ac.uk/, Oct. 2020.

[161] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, "Python Materials Genomics (pymatgen): A Robust, Open-Source Python Library for Materials Analysis.," *Comp. Mater. Sci.*, vol. 68, pp. 314–319, July 2013.

[162] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020.

[163] S. K. Lam, A. Pitrou, and S. Seibert, "Numba: A LLVM-based Python JIT compiler," in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, (Austin Texas), pp. 1–6, ACM, Nov. 2015.

[164] L. Pauling, "The principles determining the structure of complex ionic crystals," *J. Am. Chem. Soc.*, vol. 51, pp. 1010–1026, Apr. 1929.

[165] A. Hjorth Larsen, J. JØrgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. SchiØtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment - A Python library for working with atoms," *J. Phys.: Condens. Matter*, vol. 29, pp. 273002–273033, June 2017.

[166] L. Rangel DaCosta, H. G. Brown, P. M. Pelz, A. Rakowski, N. Barber, P. O'Donovan, P. McBean, L. Jones, J. Ciston, M. C. Scott, and C. Ophus,

"Prismatic 2.0 – Simulation software for scanning and high resolution transmission electron microscopy (STEM and HRTEM)," *Micron*, vol. 151, p. 103141, Sept. 2021.

[167] J. M. Paulsen, R. A. Donaberger, and J. R. Dahn, "Layered T2-, O6-, O2-, and P2-type $A_{2/3}M'^{2+}_{1/3}M^{4+}_{2/3}O_2$ Bronzes, A = Li, Na; M' = Ni, Mg; M = Mn, Ti," *Chem. Mater.*, vol. 12, pp. 2257–2267, Aug. 2000.

[168] C. Chen and S. P. Ong, "A universal graph deep learning interatomic potential for the periodic table," *Nat. Comput. Sci.*, vol. 2, pp. 718–728, Nov. 2022.

[169] L. Yang, X. Li, J. Liu, S. Xiong, X. Ma, P. Liu, J. Bai, W. Xu, Y. Tang, Y.-Y. Hu, M. Liu, and H. Chen, "Lithium-Doping Stabilized High-Performance P2–$Na_{0.66}Li_{0.18}Fe_{0.12}Mn_{0.7}O_2$ cathode for Sodium Ion Batteries," *J. Am. Chem. Soc.*, vol. 141, pp. 6680–6689, Apr. 2019.

[170] R. J. Clément, J. Billaud, A. R. Armstrong, G. Singh, T. Rojo, P. G. Bruce, and C. P. Grey, "Structurally stable Mg-doped P2-$Na_{2/3}Mn_{1-y}Mg_yO_2$ sodium-ion battery cathodes with high rate performance: Insights from electrochemical, NMR and diffraction studies," *Energy Environ. Sci.*, vol. 9, pp. 3240–3251, Aug. 2016.

[171] N. Bucher, S. Hartung, J. B. Franklin, A. M. Wise, L. Y. Lim, H.-Y. Chen, J. N. Weker, M. F. Toney, and M. Srinivasan, "P2–$Na_xCo_yMn_{1-y}O_2$ (y = 0, 0.1) as Cathode Materials in Sodium-Ion Batteries—Effects of Doping and Morphology To Enhance Cycling Stability," *Chem. Mater.*, vol. 28, pp. 2041–2051, Apr. 2016.

[172] Y. J. Park, J. U. Choi, J. H. Jo, C.-H. Jo, J. Kim, and S.-T. Myung, "A New Strategy to Build a High-Performance P2/-Type Cathode Material through

Titanium Doping for Sodium-Ion Batteries," *Adv. Funct. Mater.*, vol. 29, pp. 1901912–1901922, May 2019.

[173] G. Trimarchi, Z. Wang, and A. Zunger, "Polymorphous band structure model of gapping in the antiferromagnetic and paramagnetic phases of the Mott insulators MnO, FeO, CoO, and NiO," *Phys. Rev. B*, vol. 97, pp. 1–35, Jan. 2018.

[174] S. Zhang, K. J. T. Livi, A.-C. Gaillot, A. T. Stone, and D. R. Veblen, "Determination of manganese valence states in (Mn3+, Mn4+) minerals by electron energy-loss spectroscopy," *Am. Min.*, vol. 95, pp. 1741–1746, Nov. 2010.

[175] Z. Ji, M. Hu, and H. L. Xin, "MnEdgeNet for accurate decomposition of mixed oxidation states for Mn XAS and EELS L2,3 edges without reference and calibration," *Sci. Rep.*, vol. 13, pp. 1–14, Aug. 2023.

A | Appendix

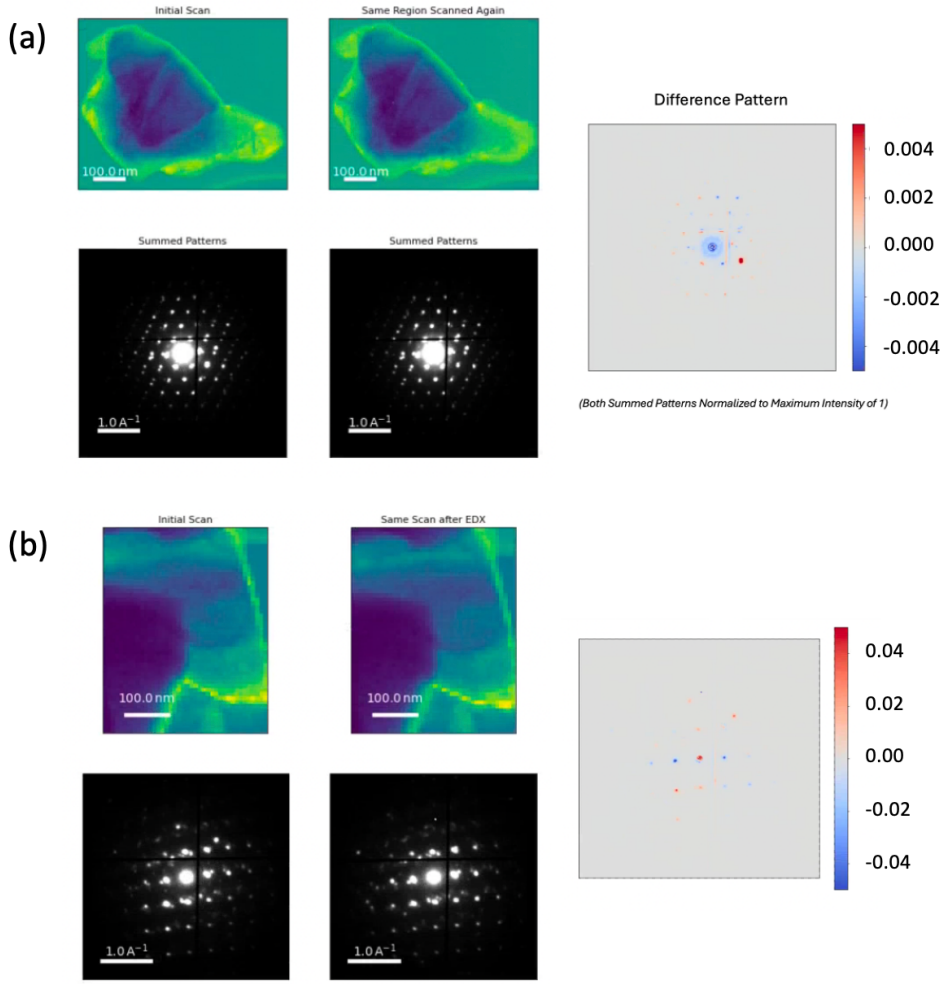


Figure A.1: Electron Dose Stability of Samples. (a) Shows the same P2-type sample subject to two consecutive collections of SEND data with the summed 2D diffraction pattern for each dataset. The difference between the two summed patterns is also shown, with diffraction patterns intensity changing at the order of 0.5% of the central beam intensity. (b) Shows the same sample subject to two consecutive collections of SEND data with an intermediate EDX data collection. The summed 2D diffraction pattern for each dataset is shown below. The difference between the two summed patterns is also shown, with diffraction patterns intensity changing at the order of 5% of the central beam intensity.

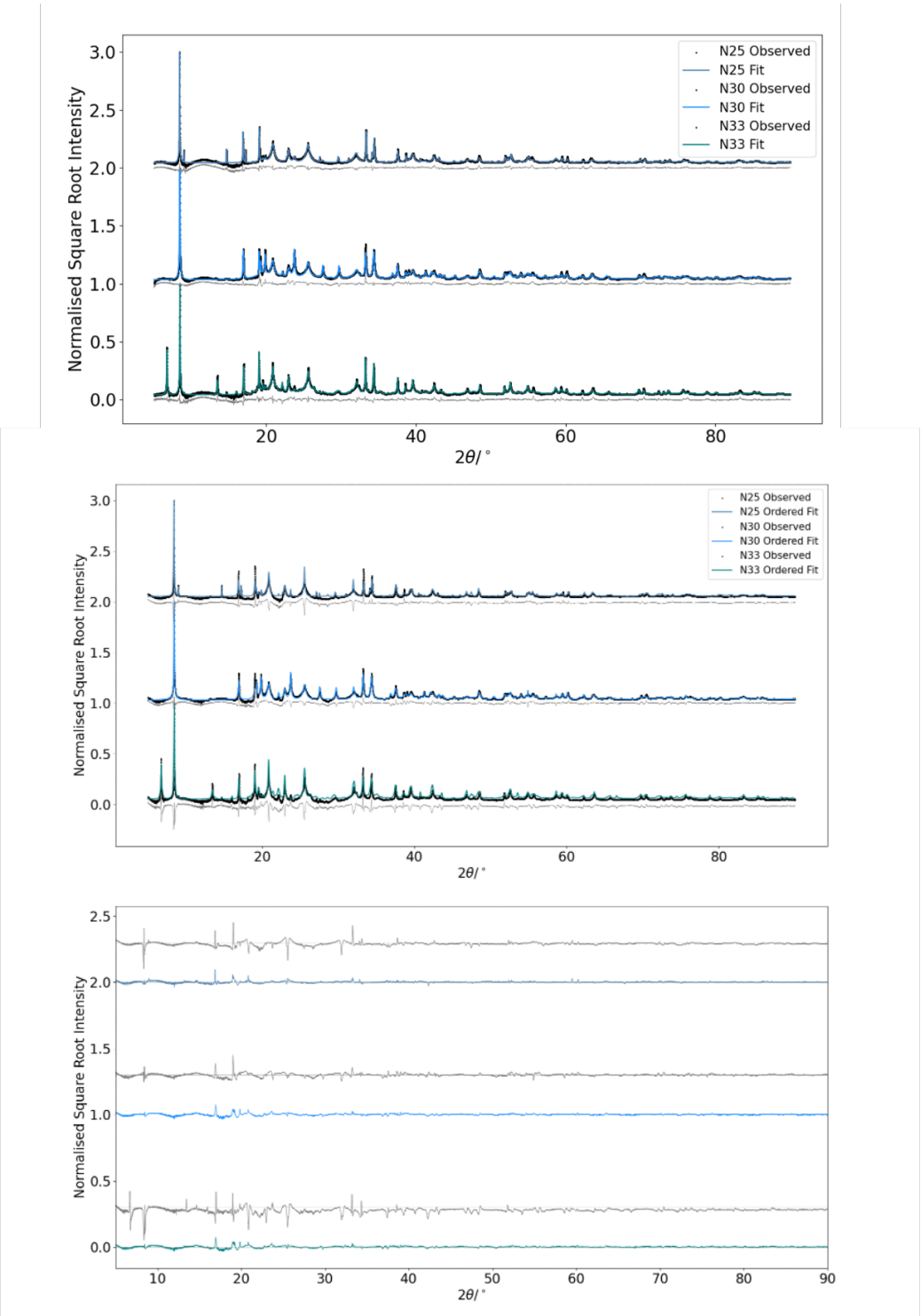

N25 Observed
N25 Fit
N30 Observed
N30 Fit
N33 Observed
N33 Fit
Normalised Square Root Intensity
2θ/°
N25 Observed
N25 Ordered Fit
N30 Observed
N30 Ordered Fit
N33 Observed
N33 Ordered Fit
Normalised Square Root Intensity
2θ/°
Normalised Square Root Intensity
2θ/°

Figure A.2

Figure A.2: Top: Reprint of Figure 10.8, for ease of comparison (The experimental PXRD data for the three compositionally varied samples (N25, N30 and N33) and their corresponding Rietveld refinement fits. The difference between the data and the fits are shown below each sample in grey.). Middle: The experimental PXRD data for the three compositionally varied samples (N25, N30 and N33) and their corresponding Rietveld refinement fits using a standard, ordered model. The difference between the data and the fits are shown below each sample in grey. Bottom: Comparison of the fit differences for the two models, with each disordered model shown in colour (N33 sat with 0 intensity at 0, N30 with 0 intensity at 1, and N25 with 0 intensity at 2) and the ordered models shown in grey above them (N33 sat with 0 intensity at 0.3, N30 with 0 intensity at 1.3, and N25 with 0 intensity at 2.3). The differences for the metrics are shown in Table A.1.

Table A.1: N25, N30, and N33 PXRD fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF), shown for the developed 'disordered' models and the standard, 'ordered' models.

Sample	Target Stoichiometry	Fit Quality Metrics					
		R^{d}_{exp}	R^{o}_{exp}	R^{d}_{wp}	R^{o}_{wp}	GoF^{d}	GoF^{o}
N25	$Na_{0.620}Ni_{0.25}Mn_{0.75}O_2$	0.446	0.446	4.176	11.223	9.369	25.152
N30	$Na_{0.656}Ni_{0.30}Mn_{0.70}O_2$	0.420	0.420	5.582	10.206	9.173	24.304
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	0.568	0.568	6.173	12.021	10.865	21.159

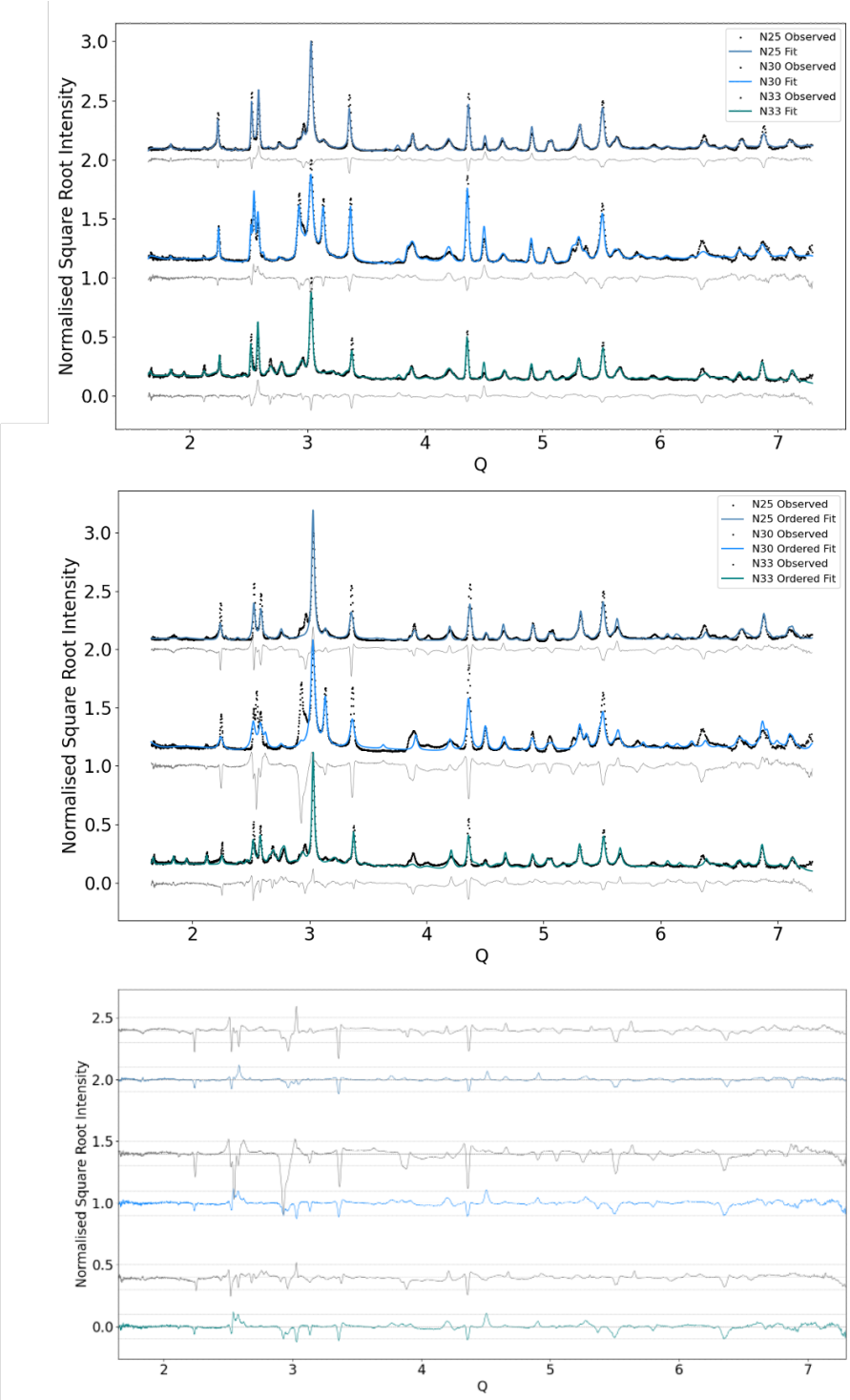

N25 Observed
N25 Fit
N30 Observed
N30 Fit
N33 Observed
N33 Fit
Normalised Square Root Intensity
Q
N25 Observed
N25 Ordered Fit
N30 Observed
N30 Ordered Fit
N33 Observed
N33 Ordered Fit
Normalised Square Root Intensity
Q
Normalised Square Root Intensity
Q

Figure A.3

Figure A.3: Top: Reprint of Figure 10.8, for ease of comparison (The experimental PND data for the three compositionally varied samples (N25, N30 and N33) and their corresponding Rietveld refinement fits. The difference between the data and the fits are shown below each sample in grey.). Middle: The experimental PND data for the three compositionally varied samples (N25, N30 and N33) and their corresponding Rietveld refinement fits using a standard, ordered model. The difference between the data and the fits are shown below each sample in grey. Bottom: Comparison of the fit differences for the two models, with each disordered model shown in colour (N33 sat with 0 intensity at 0, N30 with 0 intensity at 1, and N25 with 0 intensity at 2) and the ordered models shown in grey above them (N33 sat with 0 intensity at 0.4, N30 with 0 intensity at 1.4, and N25 with 0 intensity at 2.4). The differences for the metrics are shown in Table A.1.

Table A.2: N25, N30, and N33 PND fit quality using the metrics of: Expected R factor (R_{exp}); Weighted Profile R factor (R_{wp}); and Goodness of Fit (GoF), shown for the developed 'disordered' models and the standard, 'ordered' models.

Sample	Target Stoichiometry	Fit Quality Metrics					
		R^{d}_{exp}	R^{o}_{exp}	R^{d}_{wp}	R^{o}_{wp}	GoF^{d}	GoF^{o}
N25	$Na_{0.62}Ni_{0.25}Mn_{0.75}O_2$	1.30	1.31	9.76	17.46	7.50	13.38
N30	$Na_{0.656}Ni_{0.3}Mn_{0.7}O_2$	1.34	1.35	8.93	20.83	6.66	15.47
N33	$Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$	1.12	1.13	7.44	11.16	6.63	9.86