Quantitative electron microscopy for microstructural characterisation

Thomas Patrick McAuliffe

Department of Materials
Imperial College London

Submitted in part fulfillment of the requirements for the degree of Doctor of Philosophy at Imperial College London, 2021

Abstract

Development of materials for high-performance applications requires accurate and useful analysis tools. In parallel with advances in electron microscopy hardware, we require analysis approaches to better *understand* microstructural behaviour. Such improvements in characterisation capability permit informed alloy design.

New approaches to the characterisation of metallic materials are presented, primarily using signals collected from electron microscopy experiments. Electron backscatter diffraction is regularly used to investigate crystallography in the scanning electron microscope, and combined with energy-dispersive X-ray spectroscopy to simultaneously investigate chemistry. New algorithms and analysis pipelines are developed to permit accurate and routine microstructural evaluation, leveraging a variety of machine learning approaches.

This thesis investigates the structure and behaviour of Co/Ni-base superalloys, derived from V208C. Use of the presently developed techniques permits informed development of a new generation of advanced gas turbine engine materials.

Preamble

1. Introduction & Thesis Summary

The mission of this thesis is development and application of new data-driven electron microscopy characterisation techniques to solve challenging materials science problems. Specifically, one desires accurate knowledge of metal alloy precipitate phases' crystallography and chemistry, and quick, statistically representative assessments of microstructure. Machine learning driven approaches to correlative analysis are well suited to this purpose. Armed with accurate, precise, and fast characterisation of the microstructure, candidate development alloys can be rapidly screened. This in turn permits efficient development of a new generation of Co/Ni-base superalloys for high temperature (gas turbine jet engine) applications, which possess properties notoriously sensitive to microstructural constituents' chemistry and structure. Physical understanding of the alloys' microstructure provided by the techniques presently developed informs alloy design, with an aim to improving high temperature strength, oxidation resistance, thermal processability over current generation superalloys.

The dual objectives of this work are thus:

1. To develop approaches to rapidly and accurately assess precipitates and microstructure.
2. To leverage these approaches to design superalloys with superior thermomechanical capability.

Supplementary to developing electron microscopy characterisation and superalloys, statistical electron backscatter diffraction indexing has been advanced, and lattice strain localisation is characterised in TWIP steel nanotwins. This work is presented in the appendices.

A summary of this thesis' chapters follows:

- A **Literature Review** discusses superalloy physical metallurgy, deformation mechanisms, and the associated requirements for gas turbine jet engines. The state-of-the-art in characterisation-driven alloy development and electron microscopy (with a focus on EBSD) is presented and reviewed. Developments in machine learning, with a focus on computer vision and materials science applications, are set out, and opportunities explored.
- **Chapter 1** presents a new method for correlative scanning electron microscopy, using unsupervised machine learning (principal component analysis, PCA) to extract significant, latent features in a multimodal dataset. The effect of varying the variance weighting of the variable types is investigated. Scan points in the area of interest can subsequently be assigned to a representative, 'characteristic' signals (typically one per grain). Refined template matching (RTM) of these latent signals to dynamical simulations yields structural and orientation solutions to the characteristic features, with correlated, quantified chemistry.
- **Chapter 2** uses the PCA method presented in **Chapter 1** to investigate C, B, Zr and Ti additions to the V208 series of Co/Ni-base superalloys. Chemical and structural fingerprints for grain boundary phases are identified, and the presence of the M_2B, M_6C, ZrC and B2 phases mapped across the composition space. This is compared to predictions from thermodynamic 'calculation of phase diagrams' (CALPHAD) modelling. The influence of these alloying additions on grain boundary serration is also investigated.
- **Chapter 3** addresses segmentation of matrix and superlattice in a Co/Ni-base superalloy, using three machine learning approaches: PCA, non-negative matrix factorisation (NMF) and an autoencoder neural network. Scan

points are separable in the latent space, and subtle features are further characterised with an inverse gnomonic projection onto the diffraction sphere, where cross-sectional band profiles can be evaluated. This permits quantification and imaging of specific band features, which are used to develop 'spherical-angular dark field' microscopy.

- **Chapter 4** develops the Co/Ni-base superalloys with optimal grain boundary character (thoroughly evaluated in **Chapter 2**) for improved strength, oxidation resistance and density, while maintaining the low solvus advantage of Co-base superalloys, *via* additions of Nb and Ti. The effect of these elements on physical properties is investigated.
- A **Discussion** follows and **Conclusions** are drawn from the development of new methods, experimental results, and impact on the state-of-the-art in characterisation and materials science. Future research avenues building upon the findings of this thesis are suggested.

The appendices include supplementary work on characterisation with electron diffraction:

- **Appendix 1** validates a pattern centre (PC) and orientation refinement routine used in **Chapter 3**.
- **Appendix 2** discusses a novel normalisation method for cross-correlation in the RTM approach, used to evaluate acceptability of candidate structure/orientation solutions.
- **Appendix 3** investigates the lattice stress and strain state surrounding a TWIP steel nanotwin, using 4D-STEM. Using the diffraction patterns to map spatial resolution of elastic strain, and comparison to an analytical model of the stress field, provides insight into twin thickening, to inform future alloy design.

2. Publications

- A version of **Chapter 1** has been published in *Ultramicroscopy*:
  T. P. McAuliffe, A. Foden, C. Bilsland, D. Daskalaki-Mountanou, D. Dye, and T. B. Britton, "Advancing characterisation with statistics from correlative electron diffraction and X-ray spectroscopy, in the scanning electron microscope," *Ultramicroscopy*, no. c, p. 112944, Jan. 2020.

- A version of **Chapter 2** has been accepted for publication in *Metallurgical and Materials Transactions A*. A pre-print is available at:
  T. P. McAuliffe, I. Bantounas, L. R. Reynolds, A. Foden, M. C. Hardy, T. B. Britton, and D. Dye, "Quantitative precipitate classification and grain boundary property control in Co/Ni-base superalloys," September 2020. ArXiv ID: 2009.00948.

- Aspects of preliminary work for **Chapter 2** were presented at *Microscopy & Microanalysis 2019*, and published in the conference proceedings:
  T. McAuliffe, L. Reynolds, I. Bantounas, T. Britton, and D. Dye, "The Use of Scanning Electron Beam-based Phase Classification as a Crucial Tool in Alloy Development for Gas Turbine Engine Applications," *Microsc. Microanal.*, vol. 25, no. S2, pp. 2402–2403, 2019.

- A version of **Chapter 3** has been published in *Ultramicroscopy*:
  T. P. McAuliffe, D. Dye, and T. Ben Britton, "Spherical-angular dark field imaging and sensitive microstructural phase clustering with unsupervised machine learning," *Ultramicroscopy,* vol 219, p. 113132, Dec. 2020.

- The alloys developed in **Chapter 4** have been protected with a patent. At the time of writing, the application has been submitted to the Greek patent office under number GR20200100502, and UK patent office under GB2015106.4.

- A pre-print of **Appendix 3** is available at:
  T. P. McAuliffe, A. K. Ackerman, B. H. Savitzky, T. W. J. Kwok, M. Danaie, C. Ophus, D. Dye, "4D-STEM elastic stress state characterisation of a TWIP steel nanotwin," April 2020. ArXiv ID: 2004.03982.

3. Open Source Software

Most of the anlalysis methods presented in this thesis are generally available on the author's *Github* page: www.github.com/tmcaul. Additionally, specific packages have been developed and are available open source:

The PCA correlative EBSD/EDS analysis pipeline has been incorporated into *AstroEBSD*, written in *MATLAB*, and available open source at www.github.com/benjaminbritton/AstroEBSD. Example datasets are listed there, hosted on *Zenodo*.

The spherical-angular dark field analysis approach is available at www.github.com/tmcaul/SphericalAngularDF, written in *MATLAB*.

A data handling, background correction, and machine learning preparation package for EBSD data, *ebspy*, is available at www.github.com/tmcaul/ebspy, written in *Python 3*. This also contains example scripts for unsupervised learning of latent features in EBSD datasets.

4. Acknowledgements

First and foremost I am especially grateful to my supervisors Prof David Dye and Dr Ben Britton. It has been an honour and pleasure to work with both of them, on both a professional and personal level. They have been very patient, incredibly supportive towards my goals, and given me more than ample freedom to explore my own ideas as well as their own. I honestly couldn't have asked for more in either a technical or pastoral regard.

I would also like to especially thank Dr Vassili Vorontsov, Dr Ioannis Bantounas, Dr Thibaut Dessolier, Alex Foden, and Dr Lucy Reynolds for their support and friendship on this journey.

I am indebted to my friends (both those I've made along the way at Imperial, and who have stuck around from previous lives). Couldn't and wouldn't have done it without you.

Last and not least, my family, who are an *infinite* well of encouragement and support.

5. Declaration of Originality

I, Thomas Patrick McAuliffe, declare that all of the research included in this thesis is my own work and all the sources I have used or quoted from are appropriately referenced within this thesis.

6. Copyright Declaration

7. Glossary

A non-exhaustive list of acronyms and abbreviations used throughout this work is provided here. In each chapter they are additionally defined at their first point of use.

4D-STEM	Four-dimensional scanning transmission electron microscopy
AOI	Area of interest
CALPHAD	Calculation of phase diagrams
CPU	Central processing unit
EBSD	Electron backscatter diffraction
EDS	Energy-dispersive X-ray spectroscopy
FCC	Face centred cubic
GPU	Graphical processing unit
ML	Machine learning
NMF	Non-negative matrix factorisation
PC	Pattern centre
PCA	Principal component analysis
SEM	Scanning electron microscope/microscopy
TEM	Transmission electron microscope/microscopy
XCF	Cross-correlation function

Mathematical terms are defined as presented through the document. Notation is reused where possible, but in some cases the same symbols are used to represent different quantities. Lower-case, non-bold letters are used to represent scalar quantities; lower-case, bold letters are used to represent vectors; upper-case, bold letters are used to represent matrices and tensors. If a specific component of a vector or tensor is being referenced, the symbol will include a subscript.

8. Table of Contents

Appendix 3: 4D-STEM elastic stress state characterisation of a TWIP steel nanotwin

Appendix 4: Supplementary data

10. List of Figures

Literature Review

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix 1

Appendix 2

Appendix 3

Appendix 4

11. List of Tables

Litearture Review

Chapter 2

Chapter 3

Chapter 4

Appendix 1

Appendix 2

Appendix 4

Literature Review

1. Introduction

Micro- and nano-scale analysis of microstructure can broadly be split into (linked) discussions of spatial distributions of crystallography and chemistry. Progress towards rapid, accurate and statistically robust characterisation of microstructures has been made in recent years with developments in both experimental techniques and data processing. There has been interest in 'correlative' microscopy, where multiple techniques are employed to access independent information channels sampled from the same area of interest (AOI) [**1**]–[**5**]. Successful use of correlative microscopy yields superior characterisation capability (as limitations of individual techniques may be mitigated) and provides better confidence in phase assignment if independent classifications are mutually inclusive. This thesis explores new opportunities for microstructural characterisation.

Accurate phase identification requires knowledge of precipitate crystal structure as well as chemistry. Here, phase classification is described in terms of assigning a crystal with specific structure and chemistry to a common label or class. In many cases, classification can be performed with only chemistry or only structure if the domain of the problem is constrained (e.g. there is knowledge of the thermodynamics and kinetics for microstructural formation). This is different to phase identification, where the structure and chemistry of the phase is unknown. To perform accurate classification, one must sample both the chemistry and structure.

Data science exists at the intersection of statistics and computer science. As a discipline it is rapidly evolving, driven by the desire to fully utilise the richness of information contained within the ever-increasing modality and quantity of data we collect. Machine learning (ML) is very broad terminology for statistical techniques that in some way 'learn' from data. Such techniques are usually divided into 'supervised' and 'unsupervised' approaches. In supervised learning a relationship between independent and dependent variables is established, then the trained model is used for evaluation of unseen independent variables to predict a response. Generally, this is performed in a regression (real-valued numeric response) or classification (categorical response) setting. Supervised learning covers a wide range of techniques, including straightforward linear regression, support vector machines, and many neural networks, with plenty of variants on each of these. Unsupervised learning does not involve a specific dependent variable, but rather aims to explore the underlying structure of data. In some ways this is a more challenging task, because there is no 'answer' with which one can validate a model. Such techniques often attempt to establish a new representation of the data, perhaps encoding a more interpretable, relevant, or interesting feature space. A major component of this thesis is the application of unsupervised learning approaches to microscopy data for superior microstructural insight in order to aid alloy design.

In the subsequent literature review unique challenges presented in gas turbine alloy development are discussed. The state-of-the-art in characterisation-driven alloy development is established, and advances in electron microscopy (with a focus on EBSD) are presented. Developments in ML, with scope generally limited to focus on computer vision and application in materials science, are discussed.

2. Unique requirements for gas turbine engine superalloys

Jet engine thermal efficiency is strongly dependent on the gas stream's 'turbine entry temperature' (TET). Over a 70 year period, materials scientists have driven up the TET by over 300℃ [**1**]–[**6**]. The demand for better efficiency of natural and financial resources is greater than ever, and there is a constant motivation to increase operating temperature further. Ni-base superalloys have been the *de facto* choice for high temperature applications since the development of the very first jet engines. They maintain their strength up to very high temperatures, often up to 90% of their melting temeprature [**6**]. Superalloys also exhibit excellent creep resistance, and can be alloyed to imbue good oxidation resistance. Their outstanding high temperature mechanical properties are the result of the interaction between the face-centred cubic (FCC) γ and primitive $L1_2$ γ′ phases, a phenomenon discussed extensively in the literature [**6**], [**7**].

In 2006 Sato *et al* [**8**] discovered that the Co-Al-W ternary includes an $L1_2$ phase. It was shown that this phase field is continuous with Ni_3Al γ′, permitting the inception of novel Co-base superalloys [**9**]. Due to a slightly higher melting temperature and competitive mechanical properties, Co-base polycrystalline alloys have since emerged as competitors to Ni [**1**], [**10**]–[**15**]. Additions of Ni have been found to improve high temperature strength, by stabilising γ′ and raising the solvus [**16**], [**17**]. These additions are necessary for microstructural stability (and inhibition of decomposition into DO_{19} and other products) upon addition of further alloying elements, such as Cr [**18**], [**19**]. However, one of the potential advantages of the Co system is a low solvus (and wide processing temperature window). This is reduced with addition of Ni.

The mechanisms controlling high temperature thermomechanical capability are an essential consideration for optimisation of alloy microstructure and performance. Chemistry and microstructure play a critical role in determining which mechanisms operate, acting for example through grain boundary precipitation and stacking fault energy (SFE).

2.1 High temperature creep strength

In the context of $L1_2$ strengthened superalloys, creep has historically only been a significant issue for (high pressure) turbine blades, which experience temperatures of up to 1200˚C: roughly 90% of the absolute melting point [**6**]. Turbine blades are investment cast as single crystals in order to eliminate grain boundaries: features prone to sliding and cavitation. As the TET is raised, creep is of increasing concern for polycrystalline disc alloys, as thermally activated (diffusion-driven) creep mechanisms initiate [**20**]. Behaviour has been divided into three types: primary (strain hardening due to dislocation propagation and generation), secondary (achievement of a steady state), and tertiary (strain softening to failure). This is schematically shown in **Figure LR-1**.

Superalloys exhibit three principal creep regimes, named after the types of creep extensively exhibited: 'primary' (low temperature, high stress), 'tertiary' (intermediate temperature, low stress) and 'rafting' (high temperature, very low stress). These have been extensively reviewed [**6**], [**21**]. Typical behaviours for these regimes are shown in **Figure LR-2**. The creep mechanisms that operate are very sensitive to stress, temperature, composition and orientation. Tertiary creep is accepted to occur by the movement of a/2<110> dislocations through the γ-channels, for the most part leaving the γ′ untouched. The mechanisms for γ′ shearing during primary creep are more numerous and complex. Directional coarsening in the rafting regime generally occurs only at the highest temperatures, in conditions only subjected to single crystal blade alloys. Rafting will not be significantly discussed further. While many historical accounts are focussed on Ni-base superalloys, the discussion is applicable to Co-base alloys due to the fundamental and extensive similarities in microstructure.

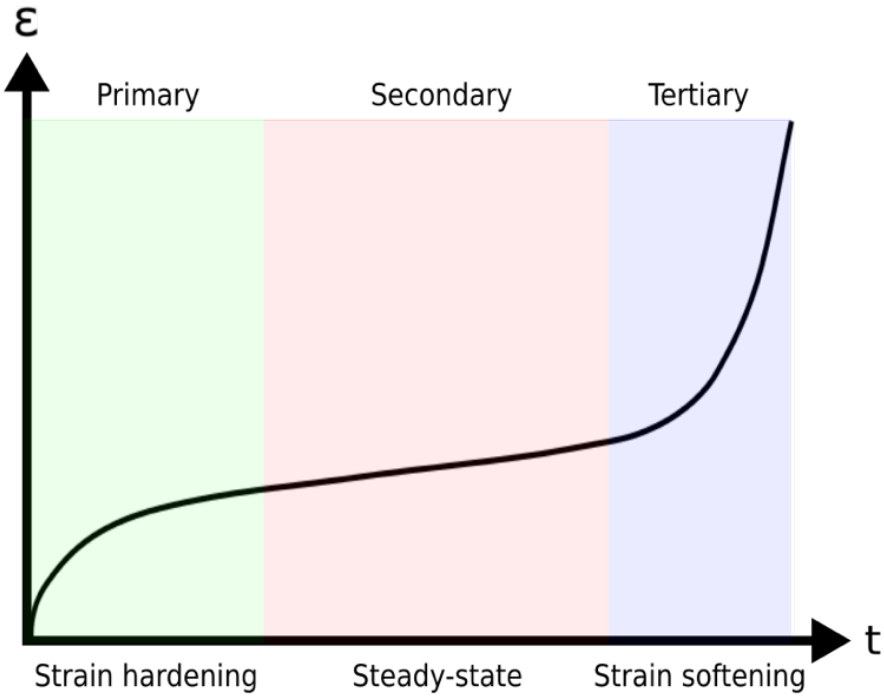

Figure LR-1: Archetypal creep regimes in a polycrystalline superalloy, after Reed [**6**].

A short period of approximately zero strain rate commonly precedes primary creep. The incubation period is accepted to occur due to the generation of ɣ-channel dislocations [**6**]. Dislocations grow out with the same Burgers' vector from 'grown-in' networks in the ɣ-channels, working against a solid solution resistance [**21**]. In a negatively misfitting alloy (where the ɣ′ lattice parameter is less than that of the ɣ) , a/2<110> dislocations glide preferentially through the horizontal channels due to superposition of misfit stresses. Co-base superalloys generally have a positive misfit, so it is expected and observed that the dislocations glide in the vertical channels [**1**], [**15**]. Change of ɣ-channel requires cross-slip of the leading screw segment. It is therefore not surprising that incubation time increases with decreasing temperature and stress [**21**]. Greater thermal activation and resolved shear stress will drive cross-slip limited dislocation propagation.

2.1.1 Mechanisms of primary creep

After incubation, if the dislocations are able to enter the precipitates primary creep begins. The a/2<110> reactions that take place at the precipitate interfaces determine the subsequent deformation behaviour. The principle mechanisms observed are stacking fault (SF) shear, twinning, and antiphase boundary (APB) shear. Unsurprisingly they are intimately related.

Most commonly during primary creep (at lower temperature and higher stress), the ɣ′ is sheared via the progression of net a<112> Burgers vector dislocation ribbons through the microstructure [**22**], [**23**], [**25**]–[**28**]. These are observed to consist of alternating intrinsic and extrinsic stacking faults. Kear *et al* [**24**] proposed reaction of two different a/2<110> dislocations for the ribbon's formation. For example, a/2[011] and a/2[$\bar{1}$01] forming a/3[$\bar{1}$12] and a/6[$\bar{1}$12]. There is experimental evidence for this dissociation [**26**].

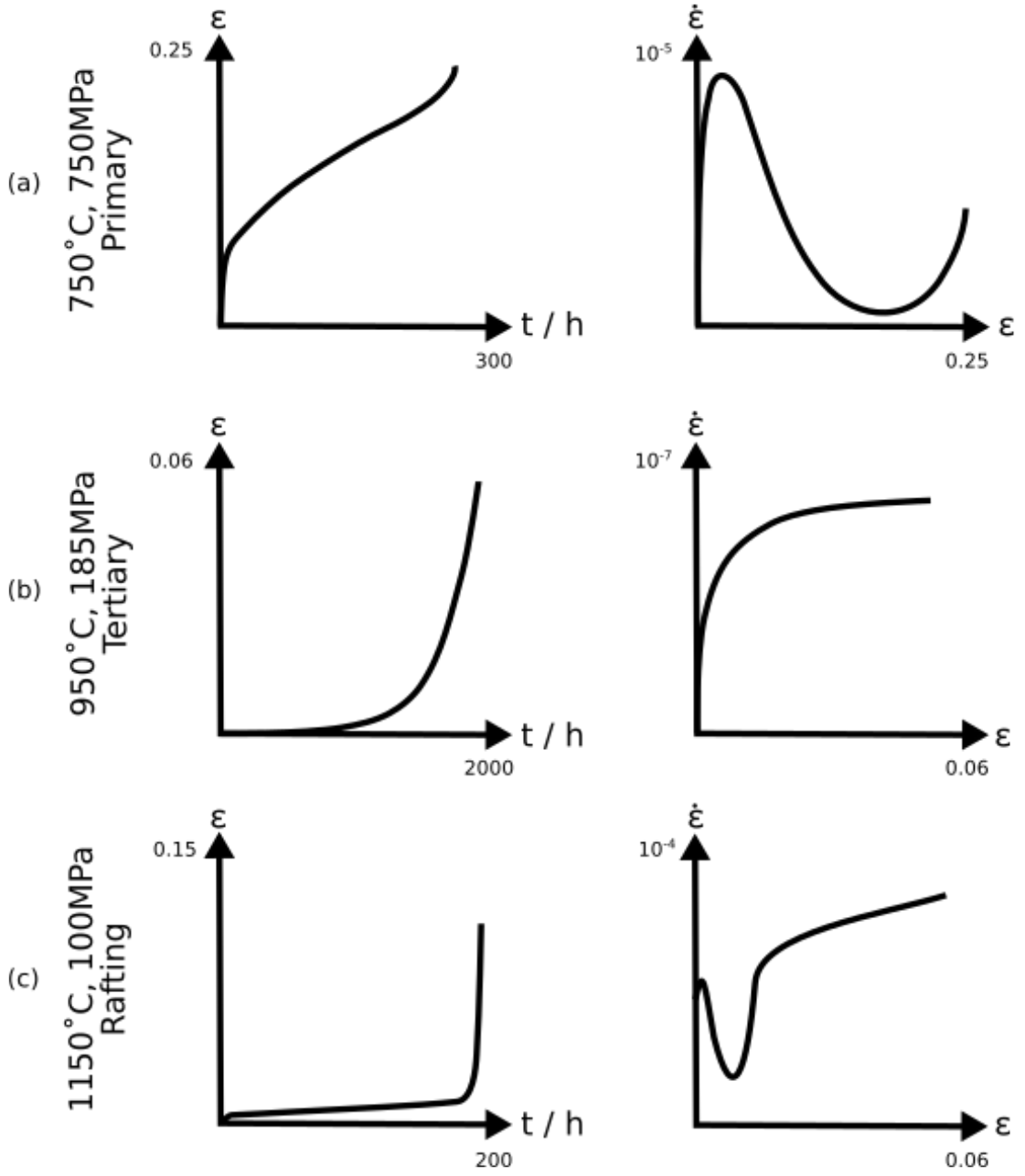

Figure LR-2: General creep behaviours, after Reed [**6**].

Movement of the a/3[$\bar{1}$12] (dislocation 1, d1) into the ɣ is glissile, though a SISF is left in its wake. Further reactions lead to the first a/6[$\bar{1}$12] (d2), a further a/6[$\bar{1}$12] (d3) and a second a/3[$\bar{1}$12] (d4) entering the ɣ. A SISF separates (d1) and (d2), an APB (d2) and (d3), and a SESF (d3) and (d4) [**23**], [**28**]. The collective 'ribbon' shears the ɣ′ in a glissile fashion, though as pointed out by Leverant & Kear [**29**] and acknowledged by Rae & Reed [**28**], a rationalisation of the stacking sequence requires atomic adjustments subsequent to the passage of (d1). The balance between SESFs and SISFs that are produced depends on the loading direction, and therefore the resolved shear stress on stacking fault shear slip systems that favour the production of one or the other [**26**].

Alternative mechanisms for SF shear have been proposed, particularly with regard to production of a single SISF rather than a full ribbon. For example, several authors have suggested that the a/3<112> dislocation forms via the dissociation

of a single a/2<$\bar{1}$01> into a/3<211> and a/6<1$\bar{2}$1> [**27**]. The passage of the a/3<211> leaves behind a SISF, with the a/6<1$\bar{1}$2> Shockley partial pinned at the interface, as its passage would create a high energy APB. However, the stacking fault ribbon idea coined by Rae *et al* [**30**], utilising the Kear mechanism, has been shown to be dominant [**26**], [**27**].

In this picture, the single ɣ SISF is observed (rather than the full ribbon) due to only the a/3[$\bar{1}$12] having entered the precipitate at this point [**28**]. The details of the mechanism by which the first a/6[$\bar{1}$12], d2, is able to enter the ɣ by reaction with d3 and d4 yet remain to be clarified.

In addition to nucleation of a<112> ribbons, Rae *et al* [**28**] note that extensive primary creep requires a suitable degree of work hardening such that the ribbons have sufficient time to propagate before the ɣ-channels saturate. This raises an interesting question of why the ribbons propagate viscously in ɣ′ if their net Burgers vector is a superlattice vector. It has been suggested that achievement of the steady state is due to a balance between SF shear strain and work hardening resulting from the filling of the γ-channels with a/2<110>{111} dislocations [**28**]. Thus, inhibition of a/2<110>{111} dislocations in the matrix leads to extended primary creep.

Primary creep also progresses under certain conditions by the propagation of deformation twins. Unocic *et al* [**32**] and Kovarik *et al* [**33**] have extensively reviewed the variety of proposed mechanisms for twin propagation. Like stacking faults, twins have been observed both isolated to the ɣ′ [**9**], [**34**], and extended across the matrix [**34**]–[**40**].

Early mechanisms postulate that micro and deformation twins build up via passage of subsequent SISFs on consecutive planes [**23**]. A build-up of Shockley partials at the interface creates a high stress network, which recovers into continuous microtwins [**41**]. Dissociation of a single a/2<110> dislocation was suggested, occurring despite being energetically unfavourable due to the presence of the ɣ/ɣ′ interface. The newly formed a/3<121> shears the particle, while the a/6<112> is pinned at the interface. The twins grow by successive a/3<112> dislocations passing on adjacent {111} planes [**41**]. Chen & Knowles [**26**] suggest a similar mechanism, except that two different a/3<112> dislocations (each formed from the combination of two different superpartial dislocations) pass on subsequent {111} planes [**26**]. It is effectively the direct creation of a SESF and thickening of the altered (twinned) stacking sequence. The opposite sense of shear leads to an equivalent process with a SISF. Notably, microtwins were only observed in orientations where SESFs are also observed.

Segregation of alloying elements to stacking fault interfaces (including twin boundaries) has been observed in both Ni-base and Co-base superalloys [**35**], [**36**], [**38**], [**42**], [**43**]. An example is shown in **Figure LR-3**. Twin boundaries locally create an HCP structure, so attract low SFE elements. Ni and Al, which raise the SFE, are locally depleted, while Co and Cr, which lower it, are enriched [**35**], [**36**], [**43**]. Freund et al [**38**] propose that in the precipitates these ɣ stabilising elements create a region of greater disorder. This in turn reduces stacking fault energies, such that both phases may be more easily cut by passing dislocations. Propagation of the fault therefore requires diffusion of the segregating elements, which act as Cottrell atmospheres around the twinning dislocations. Thus their motion is inhibited [**38**]. This introduces further time dependency to microtwinning, in addition to the atomic reordering required by Kovarik et al [**41**]. Notably for the Freund mechanism in Co-base superalloys, atomic adjustments are necessary but not as critical as in the Kolbe mechanism, due to locally reduced complex fault energy in the disordered region [**38**].

The mechanism that is observed to operate will be the lowest energy deformation mode, determined by a balance of the variety of influencing factors already discussed. Rationalising dislocations' behaviours into general schemes is a complex task. There is a consensus in the literature that creep resistance is obtained by finer secondary ɣ′ size, narrow

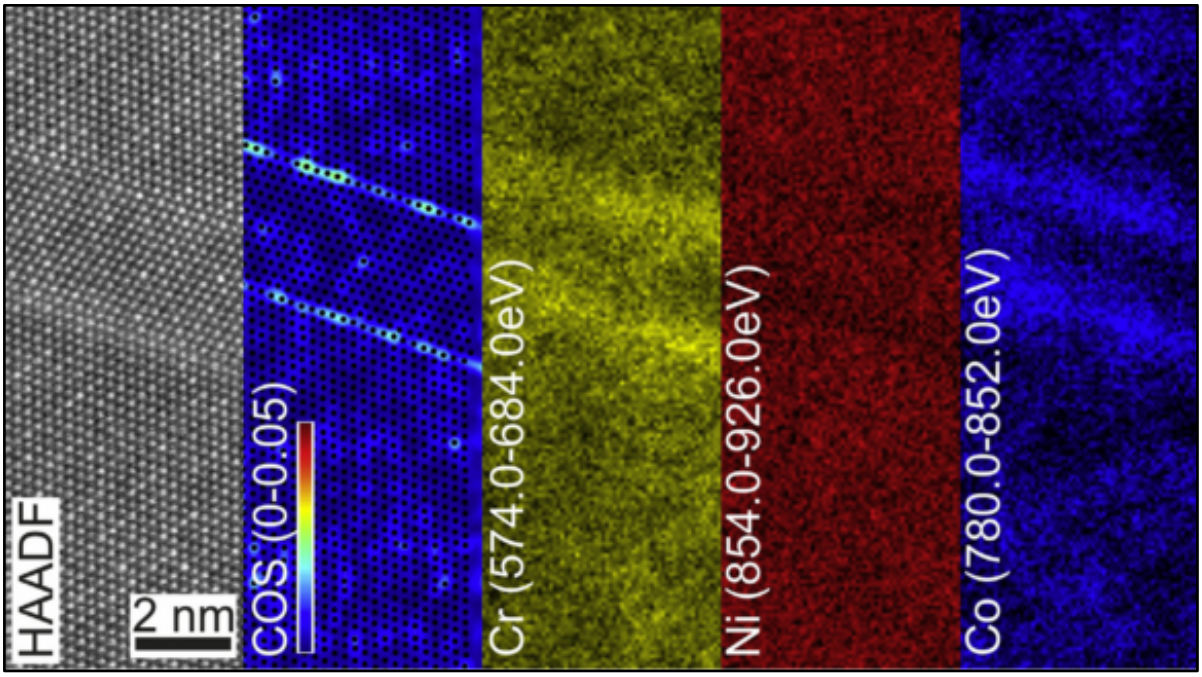

Figure LR-3: Stacking fault segregation in a series of Co-base superalloys [**38**].

ɣ-channel width and smaller volume fraction of tertiary ɣ′ [**32**]. All of these factors promote primary creep by restricting the movement of a/2<110> dislocations in the ɣ-channels. It seems likely that SF shear and microtwinning modes will be favoured when dislocation reactions and dissociations are easier. Kovarik *et al* [**41**] set forth important influencing factors for dissociation of a/2<110> dislocations into Shockley partials (which may then form microtwins) at a secondary ɣ′ interface. Firstly, as discussed, a low matrix SFE promotes decorrelation. Secondly, below the Orowan bowing stress for a/2<110> dislocations, dissociation is favoured for smaller secondary ɣ′ spacing. Finally, if the resolved shear stress is greater for the trailing partial, decorrelation is favoured. Above the Orowan bowing stress, matrix dislocations remain undissociated regardless of SFE.

It has been experimentally verified that isolated fault formation is observed for coarse precipitates, while microtwinning is favoured for finer ɣ′ [**32**], [**41**]. If the dispersion is fine, too high a stress will lead to Orowan looping of a/2<110> dislocations around precipitates, with little microtwinning [**41**]. This general behaviour has been confirmed in René 88DT by Viswanathan *et al* [**36**], who reported microtwinning at 50℃ and 838MPa, but Orowan looping around ɣ′ precipitates at 650℃ and 976 MPa. APB shearing is typically observed for high creep stress and low temperatures, closer to yield conditions [**32**]. Its occurrence at high stress is likely due to the significant activation penalty the APB formation presents. At low temperatures many other mechanisms may not operate, due to the thermal activation requirement of many required dissociations. APB shearing occurs by the passage of matrix dislocations in the precipitates, so no dissociations or reactions are required.

Questions remain regarding the process by which a steady state is achieved. Does it occur due to balance between dislocation generation and recovery, or are primary creep mechanisms inherently viscous? In the latter case, deformation would be limited by a rate determining mechanism, most likely diffusion of some kind. During stacking fault shear, it has been suggested that the activation of a second shear system leads to work hardening [**44**]. In microtwinning, based upon experimental observations it has been suggested that the rate limiting process is atomic reordering within the ɣ′ precipitates [**33**], [**41**]. Kovarik *et al* [**41**] suggests that microtwinning shares a rate limiting mechanism with SF shear,

namely thermally activated reordering required for Shockley partial propagation through ɣ′. SISF formation upon passage of a/3[112] is suggested in practice to not be as straightforward as in theory. Rather, the diffuse nature of a real dislocation core requires atomic reordering in order to correct the crystal into a perfect SISF [**41**]. In the microtwinning regime the kinetics are shown from first principles to be comparable to self-diffusion of Ni in simplified binary, ordered ɣ′. Unfortunately a similar comparison would not be helpful for SF shear in order to determine the rate limiting mechanism, as the most likely competitor is dislocation recovery, also controlled by Ni self-diffusion. As previously discussed, segregation of certain elements to stacking faults (including twin boundaries) will introduce further time dependency due to a Cottrell atmosphere effect, where the segregated elements must diffuse and catch up to the boundary in order to lower its energy [**35**], [**38**]. However, it has been shown that the kinetics of the steady state are associated with an activation energy comparable to that of Ni [**41**]. This excludes this mechanism from being rate controlling, because several elements besides Ni segregate [**35**], [**36**], [**43**]. Nonetheless, this effect would benefit the alloy's strength and creep resistance.

The creep strength of Co-base superalloys is comparable to early Ni-base [**1**], [**10**], [**12**], [**15**]–[**17**], [**35**], [**38**], [**45**], [**46**]. Their intermediate temperature (primary) creep resistance is particularly good. Neumeier *et al* [**10**] demonstrate a creep resistance at 750℃ superior to that of Udimet 720Li and Waspaloy. It is suggested that this is due to either a high content of W limiting diffusion, as also suggested by Knop *et al* [**1**], [**12**], less primary ɣ′ (and therefore a higher volume fraction of intragranular precipitate), or a difference in creep mechanism. Considerable variety in deformation behaviour has already been observed: Titus *et al* [**35**] report stacking fault shear at 900℃ and 275 - 345 MPa, Eggeler *et al* [**43**] describe an APB shearing mechanism being dominant at the same temperature and stress, while Freund *et al* [**38**] observe microtwins and SF shear at 750℃ and 530 MPa. Optimisation of alloy composition for creep strength clearly requires further work.

2.1.2 Mechanisms of tertiary creep

If the applied stress is insufficient to induce the required reactions, a/2<110> dislocations remain in the ɣ. Tertiary creep will progress if the temperature is high enough or the precipitates small enough for them to cross-slip into a new channel [**28**], [**32**]. In their seminal work on CMSX-3, Pollock & Argon [**21**] reported extensive creep in the ɣ-channels via a/2<110> dislocations. Rae *et al* [**28**] ascribe an insufficient stress as explanation of Pollock & Argon's observed lack of ɣ′ penetration (and therefore any true primary creep). Prior to failure, $L1_2$ strengthened superalloys exhibit a tertiary creep regime associated with strain softening, as depicted in **Figure LR-1**. As discussed, strain is accumulated from dislocations confined to ɣ-channels. This has been repeatedly experimentally verified [**21**], [**44**], [**47**]. Failure after extensive creep is often due to cavitation, thought to be induced by the combined effect of stress and grain boundary sliding (GBS). Edington *et al* [**48**] set forth GBS as a necessary but not sufficient requirement for cavitation, as cavities geometrically cannot form from vacancy clustering alone. SEM studies have shown that cavity nucleation is intimately related to grain boundary behaviours [**49**], [**50**]. As such, design of creep resistant polycrystalline alloys has been focussed on the strengthening of grain boundaries. Tolerable creep strains in engine components are generally in the region of less than 5%. This is the reason for a historical focus on plastic strain mitigation rather than cavitation characterisation.

If a threshold stress for dislocation entry into ɣ′ precipitates is not reached, tertiary creep operates via a/2<110>{111} dislocation propagation in the superalloy ɣ-channels [**6**], [**21**]. Dislocation density rises with increasing creep strain, and cross-slip is required for precipitate circumvention. Factors such as a higher SFE increase the ease of cross-slip [**2**]. The

creep rate has been shown to be proportional to an increasing dislocation density as strain is accumulated, leading to exponential strain softening behaviour [**44**], [**51**]. Coakley *et al* [**47**] have investigated lattice strain evolution during tertiary and primary creep regimes in CMSX-4, and found that matrix deformation releases the misfit stress, resulting in elastic compression of ɣ in the loading direction. Results were indicative of a load transfer from the ɣ to the ɣ′ in the transverse direction, corresponding to a significant amount of ɣ-channel creep, in accordance with Pollock & Argon [**21**].

GBS is a deformation mechanism exhibited in a wide variety of polycrystalline materials. Experiments have shown that it can contribute to upwards of 70% of macroscopic strain during superplasticity [**52**], corresponding to tertiary creep in superalloys. In general, GBS may be divided into two behavioural regimes: those of Lifshitz [**53**], and Rachinger [**54**]. In the former, the grains' shape change accommodates diffusional creep, and there is no net increase in their number lying along the tensile axis. In the latter, dislocations are required to accommodate the strain, such that there is no net grain shape change, but an increase in number along the tensile axis. Very little of the GBS literature focusses on superalloy systems. Due to the mechanism's importance in superplasticity, the vast majority of previous work is focussed around such (often binary) alloys.

GBS accommodated by diffusional creep (Lifshitz sliding) is well established and reviewed [**52**], [**53**], [**55**], [**56**]. In their seminal work, Ashby & Verrall [**57**] present a thorough analysis of GBS in superplasticity accommodated by diffusional creep. The mechanism differs from straightforward Nabarro-Herring or Coble creep (occurring up to ten times faster than these at high strains) in that grains switch neighbours without significant elongation, schematically shown in **Figure LR-4**. The process is limited either by diffusion (bulk or grain boundary), or the capacity for a grain boundary to act as a source or sink for point defects. Ashby & Verrall's diffusional accommodation is presented in **Figure LR-5**. The work done on a unit group of four hexagonal grains by four constituent processes is used to consider a quasi-analytical model. These are: the diffusive current required to accommodate shape change; the chemical potential activation penalty for the 'interface reaction', where boundaries source/sink defects; shear displacements of grains sliding past one another; fluctuations in boundary area over the course of the mechanism.

Langdon *et al* [**58**] have proposed a unified Rachinger sliding model for superplasticity and creep. In this scheme GBS occurs by the propagation of dislocations, with their motion principally diffusion controlled. A variety of strain accumulation mechanisms exist. These include glide and climb in adjacent grains after Gifkins [**59**], groups of grains sliding with stress accumulations driving climb after Ball & Hutchinson [**60**], and production of dislocations at protrusions after Mukherjee *et al* [**61**]. Differences in mechanism empirically manifest as variety in the numerical prefactor and choice of diffusivity constant. The transition between regimes is due to variation in the details of the rate limiting step [**58**]. In fine grained superplastic samples it is the climb of dislocations across grain boundaries. In coarse grains undergoing creep it is annihilation and interaction at sub-grain boundaries. This is schematically shown in **Figure LR-6**.

GBS may act as a nucleation mechanism for cavities, which are commonly the source of creep failure [**62**], [**63**]. Mitigation of sliding and boundary strengthening is employed to maximise tertiary creep life. A variety of cavitation nucleation and growth mechanisms have been suggested. These include an agglomeration of spherical cavities at triple junctions, or by void growth at transverse boundaries [**63**]. GBS may lead to accelerated cavitation by raising the local stress state at transverse boundaries. Additionally, cavities may nucleate and grow due to dislocation pile up at grain boundaries or hard particles, or because of vacancy diffusion under a stress induced chemical potential gradient.

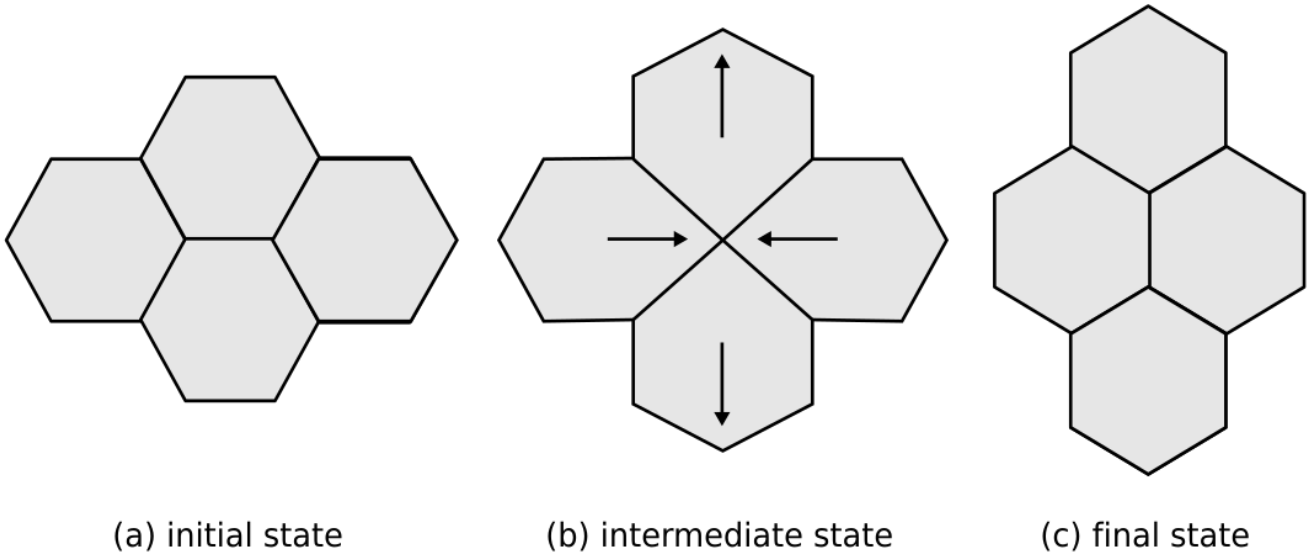

Figure LR-4: Lifshitz grain boundary sliding, redrawn from [**57**].

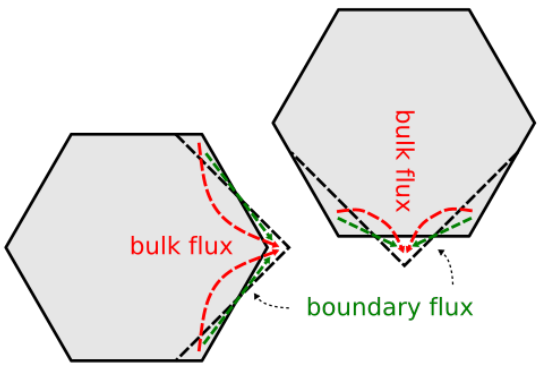

Figure LR-5: Ashby mechanism for Lifshitz sliding, redrawn from [**57**].

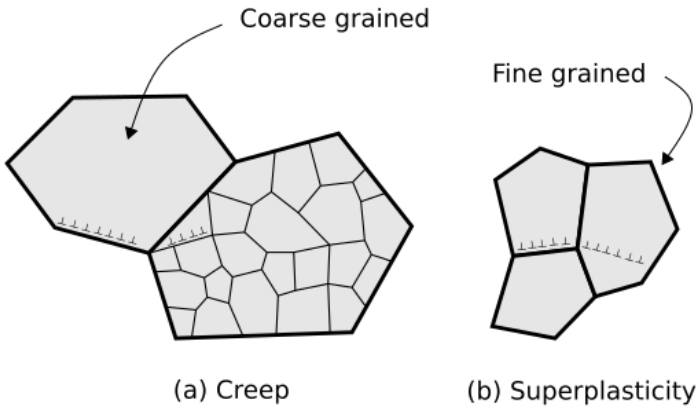

Figure LR-6: Langdon mechanism for Rachinger sliding, redrawn from [**58**]

2.2 Grain boundary engineering

B, C and Zr are commonly alloyed to improve grain boundary cohesion [**1**], [**10**], [**12**]. Such elements are used due to their tendency to segregate to the boundaries, forming monolayers and precipitates that increase the work of cohesion. This inhibits grain growth and GBS, acting as a useful tool in microstructure development and extending both creep and fatigue life. It has been demonstrated that maximisation of benefit from B additions requires elemental segregation. Precipitation of borides occurs after additions greater than the solubility limit of B; upon their formation no further gains in strength or life are achieved [**64**], [**65**].

2.2.1 Intergranular precipitation

Of particular concern is precipitation of minor phases either during heat treatment or in-service. Engineering a desirable grain boundary character is essential to the optimisation of superalloy microstructure and performance. Toward this endeavour we require precise knowledge of chemistry, distribution, and structure of intra/intergranular precipitates, particularly carbides and borides. These phases form through all stages of alloy processing, from initial casting through to final ageing treatments. In the literature, refractory-rich grain boundary precipitates may be referred to as a 'carbide' without any effort to differentiate between $M_{23}C_6$, M_6C or boride structures. Due to preferential elemental segregation of high Z-number elements, these all exhibit high backscatter scanning electron microscope (SEM) contrast. There is significant evidence that these precipitates' exact character has a significant effect on mechanical and environmental stability, especially in high temperature applications, despite their small volume fraction [**64**], [**66**]–[**68**]. A secondary effect of grain boundary precipitate interactions is the ability to facilitate grain boundary serration mechanisms, also an essential consideration in modern superalloy grain boundary engineering [**69**], [**70**].

A wide variety of carbide and boride structures are known to precipitate in superalloys, including but not limited to MC, $M_{23}C_6$, M_6C, M_2C, M_2B, M_3B_2 and M_5B_3, where M is a dominant metallic enrichment, generally a refractory element. A selection of prior studies within superalloys and steels with similar refractory contents are presented in **Table LR-1**. Current thinking is that a significant mismatch between solute atomic radii to the average of the pseudo-FCC (ɣ / ɣ′) matrix (smaller for C and B, larger for refractories) leads to a driving force for these elements to migrate to grain boundaries. Here the strain dipoles accommodating lattice parameter mismatch from interstitial or substitutional solutes can be relieved. The migrated elements tend to combine, forming intermetallic and ceramic compounds such as topologically close packed (TCP) phases, carbides and borides.

MC carbides generally precipitate at high temperatures (for example over 1100˚C in Udimet-520 [**71**]), during casting ('primary' MC carbides) or homogenisation treatments ('secondary'). These are stable at high temperature, and are difficult to dissolve once formed without risking incipient melting. MC carbides may exhibit significant variation in composition due to forming over a wide temperature range during solidification [**72**]. Their relative stability also means they are a common decomposition product of other carbides, such as M_2C [**72**]–[**75**], with the associated refractory rejection possibly also leading to TCP phase precipitation [**76**]. In superalloys their presence is generally welcomed, with certain systems utilising them during supersolvus forging, providing hot ductility, or for the prevention of hot zone cracking [**77**].

Table LR-1: Carbides and borides commonly observed in superalloys, with the corresponding refractory affinities measured with a variety of characterisation techniques.

Type	System	'M' Enrichment	Material	Technique	Ref
MC	Superalloy	Ti	René 88DT	XRD	[**90**]
		Nb	Inconel 718	TEM-EDS	[**91**]
		Ti, Mo	Udimet 520	EDS	[**71**]
		Ta	STAL-15CC	APT	[**65**]
		Ti, Ta, Nb, Mo	ME3	STEM-EDS	[**92**]
		Zr, Ta	Co/Ni superalloy	EBSD+EDS	[**93**]
	Steel	Mo, V	AISI M2 HSS variants	APFIM	[**94**]
		Mo, V, W	AISI M2 HSS variants	TEM-EDS	[**73**]
		Nb, Ta, V	Ferritic/Martensitic Steel	STEM-EDS	[**95**]
		V, Cr, Mo	AISI M2 HSS variants	EDS	[**74**]
$M_{23}C_6$	Superalloy	Cr, Mo	RR1000	STEM-EDS	[**96**]
		Cr, Mo	Udimet 720	TEM-EDS	[**97**]
		Cr, Mo	Udimet 520	EDS	[**71**]
		Cr	Inconel 738	STEM-EDS	[**98**]
		Cr, Mo	STAL-15CC	APT	[**65**]
		Cr, Mo, W	ME3	TEM-EDS	[**92**]
	Steel	Cr, Co	Ferritic/Martensitic Steel	STEM-EDS	[**95**]
		Cr	CrMnFeCoNi HEA	STEM-EDS	[**99**]
M_6C	Superalloy	W, Mo	K465 superalloy	STEM-EDS	[**100**]
	Steel	W, Mo	AISI M2 HSS variants	STEM-EDS	[**72**]
		Mo, W	AISI M2 HSS variants	STEM-EDS	[**73**]
		Fe, Co	Ferritic/Martensitic Steel	STEM-EDS	[**95**]
M_2C		Mo, Cr	Hastelloy N	EDS	[**78**]
		Mo, Cr	AISI M2 HSS variants	APFIM	[**94**]
M_3C	Steel	Co	Ferritic/Martensitic Steel	STEM-EDS	[**95**]
M_5B_3	Superalloy	Cr, Mo, W	René 88DT	XRD	[**90**]
		Cr, Mo, W	STAL-15CC	APT	[**64**]
		Cr, Mo	In 738	STEM-EDS	[**101**]
		Cr, Mo	STAL-15CC	APT	[**65**]
		Cr, Mo, W	ME3	TEM-EDS	[**92**]
M_3B_2	Superalloy	Cr, Mo, W	René 88DT	XRD	[**90**]
		Nb, Mo, Cr	Inconel 718	EDS	[**102**]
M_2B	Steel	Cr, Fe	Austenitic stainless steel	STEM-EDS/EELS	[**103**]
		Cr, Fe	AISI 1045 steel	TEM-EDS	[**104**]
		Cr, Fe	18/20 Ni/Cr stainless steel	XRD	[**105**]

$M_{23}C_6$ and M_6C phases often form on grain boundaries at intermediate temperatures (700 – 1000˚C), coinciding with optimum ɣ′ ageing regimes. They may also form from the decomposition of other carbides such as M_2C or MC, and on occasion precipitate intragranularly [**78**], [**79**]. As presented in **Table LR-1**, M_6C tends to exhibit a greater W affinity than $M_{23}C_6$, so is observed in Co-base superalloys that require high atomic fractions of W [**79**]–[**81**].

2.2.2 Boundary serration

GBS can further be inhibited by serration of grain boundaries. Development of such morphologies increases resistance to creep crack growth for mechanical reasons, as well as due to an increase in length of diffusion paths [**69**], [**82**], [**83**]. Serrations develop due to strain induced movement of primary ɣ′ (or other pinning phases) during cooling. Carter *et al* [**84**] have characterised the effect of grain boundary serration on strain localisation and accumulation in René 104. The study showed that strain tended to accumulate at triple junctions and low symmetry boundaries in a standard microstructure, with multiple slip systems active. Sliding boundaries tended to have a single slip system active. There was less correlation at serrated boundaries, which were shown to be more resistant to sliding. Furthermore, regions of strain localisation around serrated boundaries tend to be more diffuse, extending further into the grain interiors.

Heat treatments designed to mechanically lock grain boundaries together have been used since first being reported for superalloys in 1976 by Miyagawa *et al* [**85**] and Larson [**86**], though the morphology was reported in austenitic stainless steels 10 years prior [**87**]. They universally involve slow cooling through an intergranular phase solvus temperature. There is significant evidence that engineering such microstructures improves creep ductility and life, where intergranular cracking and cavitation modes are exhibited [**70**], [**85**]–[**87**]. Serrations directly affect the character of strain evolution in high temperature deformation regimes. *In-situ* digital image correlation (DIC) studies on René-104 have shown that serrations reduce strain concentration around microstructural features such as annealing twin (Σ3) boundaries, and plasticity is distributed more evenly across the microstructure [**88**], [**89**]. This is as originally suggested by Larson [**86**] and later validated by Carter *et al* [**88**] who showed that strain concentration fields at serrated grains boundaries are smaller in magnitude and extend further into grain interiors than for non-serrated. This is likely due to a reduction in the accommodation of strain through GBS, which reduces the onset of tertiary creep.

The processes by which grain boundaries serrate during heat treatment are still the subject of some debate. The dominant mechanism varies between alloy systems as a function of ɣ′ and intergranular phase solvus temperatures. Larson [**86**] showed that an air cool through the ɣ′ solvus develops a serrated microstructure in Inconel 792, with the mechanism attributed to ɣ′ intergranular nucleation followed by subsequent boundary migration. Highly serrated boundaries are observed to only exhibit large, globular carbides. This is attributed to a preference for this morphology to precipitate at serration nodes during intermediate temperature ageing. Lower temperature ageing prior to air cooling produced smoother grain boundaries with film-like $M_{23}C_6$ carbides. The inference drawn at the time was that serrated boundaries promoted globular $M_{23}C_6$. An alternative, non-mutually exclusive interpretation is that the film-like $M_{23}C_6$ carbides, if aged in to the microstructure at intermediate temperature (even if above the ɣ′ solvus), prevent serration. Miyagawa *et al* [**85**] explain their similar observations (after comparable heat treatments to Larson) through nucleation and resulting preferential growth of $M_{23}C_6$ precipitates during cooling. In this alternative scheme, carbide growth into the grain bulk along a preferred crystallographic direction provides a lower interfacial energy plane for the grain boundary to migrate to.

The ɣ′-driven mechanism proposed by Koul & Gessinger [**69**] involves preferential intergranular precipitation due to superlattice misfit relief. They argue that the strain energy difference between the boundary and matrix facing sides of

the ɣ′ particle provides a driving force for its migration in the direction of the boundary normal, until this energy is matched by the boundary line tension. For this mechanism to operate effectively the grain boundary carbide ($M_{23}C_6$ in their study) solvus temperature should to be lower than that of ɣ′. The key requirement is that grain boundary segments between ɣ′ particles must be mobile during the extended nucleation period. Consequently ɣ′ and carbide-driven mechanisms cannot operate simultaneously: if a precipitate is already pinning the boundaries, slow precipitation of a lower temperature phase will not successfully serrate the boundaries (regardless of whether this is ɣ′ or carbide). This effect is presented in Nimonic 105, where $M_{23}C_6$ precipitation precedes ɣ′ and does not itself serrate the boundaries. The pinning precipitates then prevent ɣ′ from subsequently doing so.

3. Developments in electron microscopy

Here simple diffraction physics and the formation of both 'spot' and Kikuchi diffraction patterns are outlined. Advances in orientation microscopy and phase-ID are presented, with a focus on the scanning electron microscope (SEM) and electron backscatter diffraction (EBSD). Additionally, advances in chemical analysis using energy-dispersive X-ray spectroscopy (EDS) and the capabilities of scanning transmission electron microscopy (STEM) are discussed.

3.1 Electron diffraction: spot and Kikuchi pattern formation

High magnification and diffraction imaging modes of the transmission electron microscope are presented in **Figure LR-7**. Conventional analysis usually involves employing the electron beam in one of three operational modes: bright or dark field (BF, DF) imaging (differentiated by the position of the objective aperture), selected area electron diffraction (SAED), or scanning transmission electron microscopy (STEM). In STEM, a converged probe is employed, schematically shown in **Figure LR-8**.

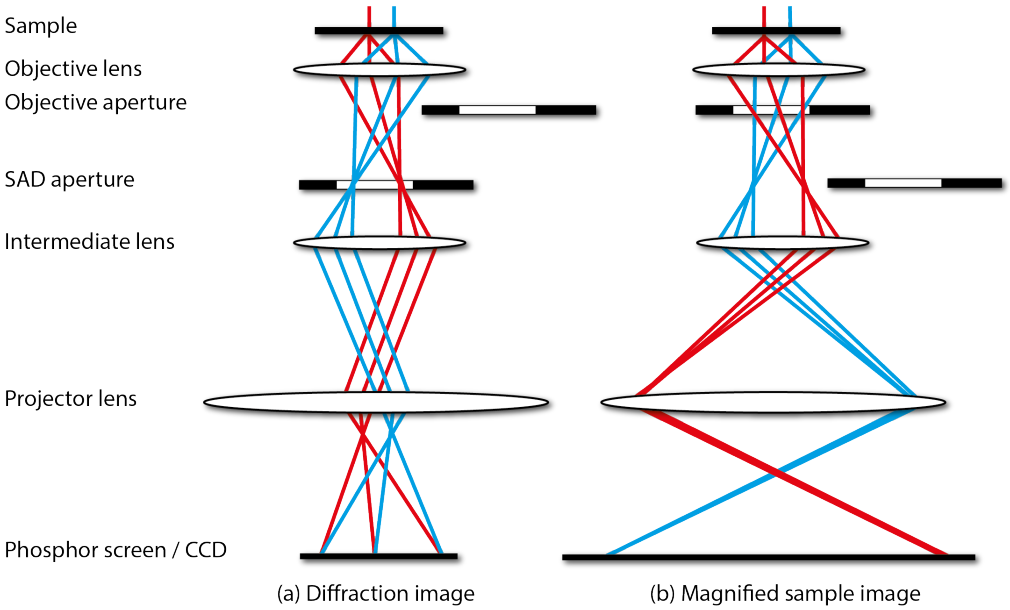

Figure LR-7: Diffraction (a) and sample (b) imaging from a parallel beam, redrawn from [**107**]

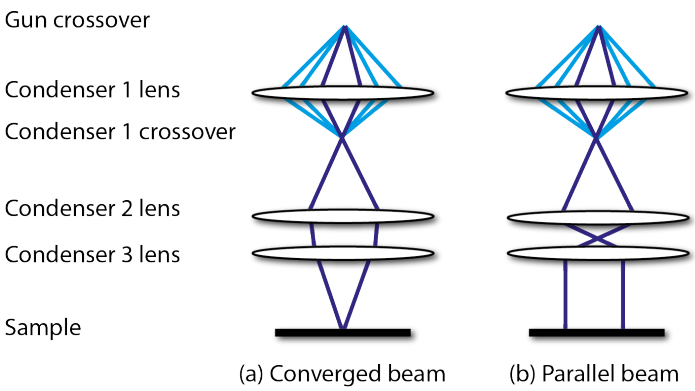

Figure LR-8: Converged (a) and parallel (b) probe formation, redrawn from [**107**]

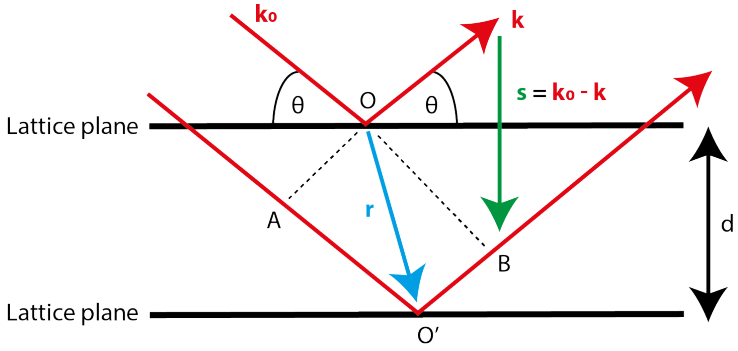

Figure LR-9: Schematic of Bragg's law, redrawn from [**106**].

3.1.1 Bragg's law and (near) elastic diffraction

Illumination of a periodic crystal with long-range order and regularly spaced atomic planes will lead to constructive interference and far-field (Fraunhofer) diffraction for the incoming / outgoing angle θ, in scheme with Bragg's law. This is presented in **Figure LR-9**. A derivation after Phillips [**106**] is presented.

For an incoming plane wave with wavevector $\mathbf{k_0}$ diffracting first off electron density at O, then again at O' separated by real space vector $\mathbf{r}$. We define the magnitude of the wavevectors $\mathbf{k}$ and $\mathbf{k_0}$ to be $1 / \lambda$ .The difference in path length between these two outgoing vectors is given by the distance $AO' + O'B$. This path difference is given as:

$$\begin{aligned} AO' + O'B &= \boldsymbol{r} \cdot \frac{\boldsymbol{k_0}}{|\boldsymbol{k_0}|} + \boldsymbol{r} \cdot \frac{\boldsymbol{k}}{|\boldsymbol{k}|} \\ &= \lambda\,(\boldsymbol{r} \cdot \boldsymbol{k_0}) + \lambda\,(\boldsymbol{r} \cdot \boldsymbol{k}) \\ &= \lambda \boldsymbol{r} \cdot (\boldsymbol{k} - \boldsymbol{k_0}) \\ &= \lambda \boldsymbol{r} \cdot \boldsymbol{s} \end{aligned} \qquad \textit{LR-1}$$

There is therefore a phase difference between radiation diffracted at O and O' of magnitude $2\pi\, \boldsymbol{r} \cdot \boldsymbol{s}$, with difference in magnitude given by:

$$\frac{\psi}{\psi_0} \propto e^{-2\pi i\,(\boldsymbol{r} \cdot \boldsymbol{s})} \qquad \textit{LR-2}$$

Where the exit wavefunction is given by ψ , and the incident wavefunction as ψ_0. Constructive interference occurs when $(\boldsymbol{r} \cdot \boldsymbol{s})$ is an integer: the complex exponent is a multiple of 2π. It can be shown that the length of $\mathbf{s}$ is given by:

$$\begin{aligned} |\boldsymbol{s}| &= 2\,|\boldsymbol{k}|\,\sin(\theta) \\ &= \frac{2\,\sin(\theta)}{\lambda} \end{aligned} \qquad \textit{LR-3}$$

Such that upon constructive interference:

$$\begin{aligned} (\boldsymbol{r} \cdot \boldsymbol{s}) &= n \\ (\boldsymbol{r} \cdot \boldsymbol{s}) &= d\,|\boldsymbol{s}| \\ n\lambda &= 2\,d\,\sin\,(\theta) \end{aligned} \qquad \textit{LR-4}$$

For integers n, which is Bragg's law [**106**], [**107**].

In a TEM or SEM, where sufficient incoming angles are sampled, this results in coherent, elastic scattering at angles θ to all possible diffracting Miller planes {hkl}, with spacings d_{hkl}. The intensity of an elastically scattered beam from plane (hkl) is proportional to the magnitude of the corresponding 'structure factor' $\boldsymbol{F}_{hkl}$, which accounts for the possibility of destructive interference due to symmetries of the crystallographic unit cell.

The structure factor is given by the Fourier transform of the electron density function $f(\boldsymbol{r})$ over the crystal [**106**]:

$$F(\boldsymbol{s}) = \int f(\boldsymbol{r})\, e^{-2\pi i\,(\boldsymbol{r}\cdot\boldsymbol{s})}\, d\boldsymbol{r} \qquad \textit{LR-5}$$

Which upon substitution of **s** for reciprocal lattice vector (h,k,l), and **r** for unit cell normalised position vector (x, y, z), and evaluated at all atomic locations a gives the complex F_{hkl}:

$$F_{hkl} = \sum_{a=1}^{N} f_a\, e^{2\pi i\,(hx_a + ky_a + lz_a)} \qquad \textit{LR-6}$$

With x_a, y_a, z_a corresponding to the atomic sites positions within the unit cell of the a^{th} atom, and f_a the 'atomic structure factor', a property that describes the propensity for a given element to scatter a kind of radiation (X-ray, electron, *etc*). Finally:

$$\text{Intensity} \propto |F_{hkl}|^2 \qquad \textit{LR-7}$$

The electron density function is dependent on incident angle and beam energy amongst other things, but is considered as a constant for the purposes of determining systematic diffracting beam absences due to intra-cell destructive interference (especially as the structure factor only provides proportionality to total intensity rather than an absolute determination) [**108**]. Subsequently in this thesis, diffraction as described will be referred to as 'Bragg' or 'elastic' interchangeably.

Notably, the symmetry of the face-centred cubic (FCC) unit cell is such that Miller planes with mixed odd and even indices (such as {101}, {112}) do not constructively interfere and have $\boldsymbol{F}_{hkl} = 0$ for such conditions. The primitive system does not have such a restriction. This is extremely useful for discriminating for example $L1_2$ (primitive) and FCC structures in dark-field TEM, as the objective aperture on the diffraction plane can be used to isolate and re-interfere intensity from diffraction spots forbidden in the FCC structure [**6**], [**30**], [**107**].

SAED is the traditional method for structure determination in electron microscopy. Spatial resolution of this approach is limited by the size of the selected area aperture employed, unless overlapping diffraction patterns can be successfully deconvolved. In practice, after (usually challenging) sample preparation a nearby zone axis is located and a two-dimensional projection of the sampled region's reciprocal lattice is measured. These can be compared to (evenly sampled in orientation space) libraries of kinematically or dynamically simulated spot patterns for candidate crystal structures, but there may be cases of pseudosymmetry and strong pattern similarity, especially between phases with near identical structure. This is confounded by the fact that upon rotation to a zone axis one only ever samples two coplanar

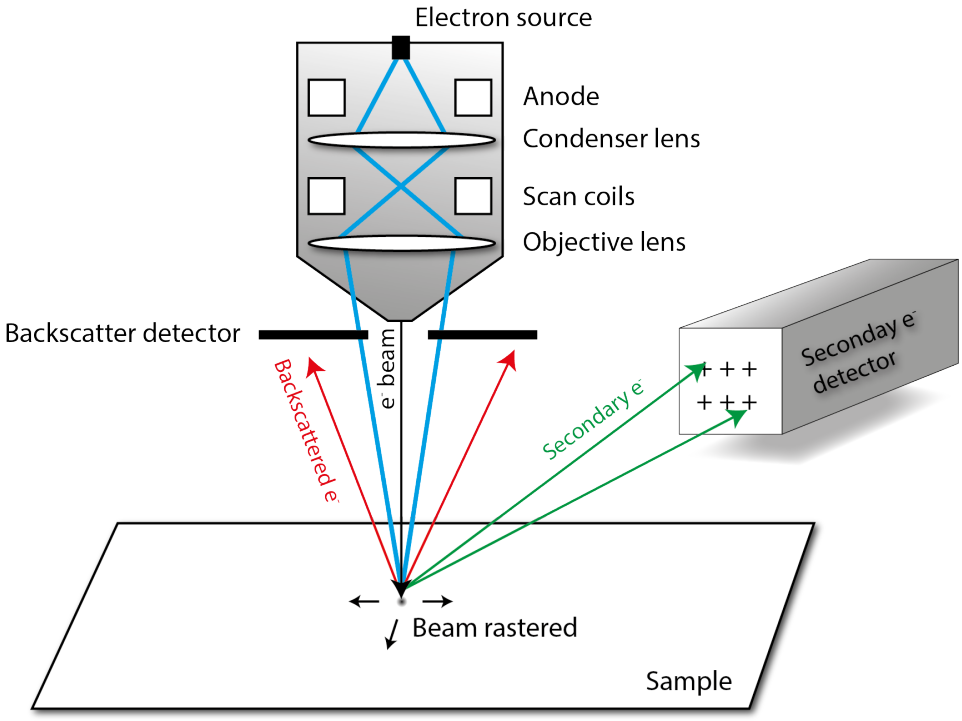

Figure LR-10: Scanning electron microscope imaging schematic.

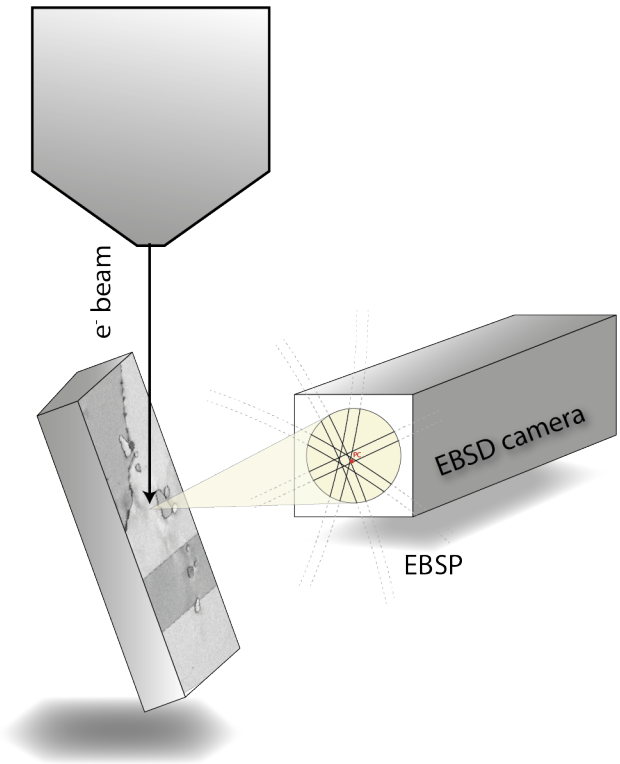

Figure LR-11: Typical setup for an EBSD measurement.

reciprocal lattice dimensions. This leads to a 180˚ ambiguity in crystal orientation: three non-coplanar vectors are required for a full description [**109**]. However, spot diffraction patterns are extremely difficult to interpret and index if the electron beam is not aligned to a high symmetry zone, unless the full 3D reciprocal lattice is measured with a tomographic approach [**110**]. Furthermore, dynamical effects can lead to activation of disallowed reflections, which can confuse structural analyses.

3.1.2 Kikuchi diffraction and EBSD

The SEM is maybe a more convenient instrument than the TEM, permitting analysis of bulk samples across areas of interest (AOIs) up to mm-scale, with spatial electron beam position resolution of 100s of μm down to a few nm. A schematic is presented in **Figure LR-10**. Upon electron incidence in a material (typically the first few 1 - 10 μm of depth), inelastic backscattering occurs [**107**], [**111**]. Electrons are scattered in all directions, generating a full, new set of uncorrelated incident angles to the lattice planes. Subsequent Bragg diffraction of the inelastically scattered beams generates so-called 'Kossel cones' of electron intensity. Such cones form for every diffraction condition. Intersection of these with a planar detector (forming a gnomonic projection of the full backscatter-diffraction sphere) allows the capture of 'Kikuchi' or 'electron backscatter' patterns (EBSPs); they present as a series of crossing bands of raised intensity [**112**]. A typical experimental setup is presented in **Figure LR-11**. When a sample is 15-20˚ from the optic axis of the microscope patterns are reasonably intense [**113**]. EBSPs are formed from a relatively large solid angle of diffracted electrons, so more than three geometrical conditions (in this case Kikuchi bands corresponding to lattice planes) are normally sampled, as required to determine a unique orientation solution. Conventionally a Hough (Radon) transform and set of interplanar angle lookup tables are employed to index crystallographic planes to determine phase and orientation [**112**], [**114**]. Collection of EBSPs across a wide AOI by electron beam rastering in an SEM is known as electron backscatter diffraction (EBSD), usually performed for the purposes of orientation mapping. Kikuchi patterns also appear in transmission diffraction, though the smaller volume of material sampled generally leads to weaker patterns. Transmission Kikuchi diffraction (TKD) is seeing increased use in the SEM due to the much finer probe size that is achievable, while maintaining the relatively large solid angle of incident electron intensity provided by the SEM (as supposed to the TEM) [**109**], [**115**], [**116**].

3.2 Advances in EBSD post-processing

Indexing of EBSPs (assignment of orientation and crystal structure to the scan location) is generally performed *online* (at the time of measurement) with a fast interplanar angle lookup table (LUT) approach. Such methods are briefly reviewed, then advances in slower *offline* approaches such as template matching are discussed.

3.2.1 Kikuchi band analysis for orientation and structure

With good knowledge of detector, sample, and scanning geometry a full set of Euler angles can be obtained for each scan point. These are used for the generation of pole figures, inverse pole figures (IPFs), and IPF maps. Britton *et al* [**117**] highlight the importance of correct geometry calibration for the calculation of accurate orientation solutions. Upon measurement, patterns' Hough (discrete Radon) transforms are calculated, as suggested by Krieger-Lassen *et al* [**118**]:

$$\rho = x\cos\omega + y\sin\omega$$
$$-R < \rho < R \qquad \text{LR-8}$$
$$0 < \omega < \pi$$

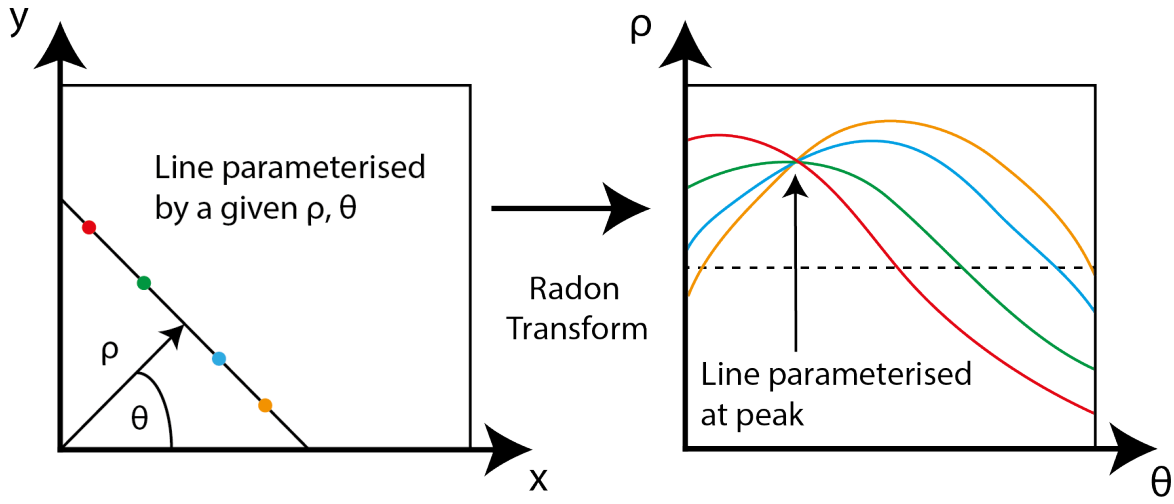

Figure LR-12: Operation of the Radon transform.

Where ω is the angle of any given line with respect to the x-axis, ρ the distance to the image centre, and $2R$ the diameter of the pattern. "Hough-space" is parameterised by the basis (ω, ρ). This is visualised in **Figure LR-12**. In practice, it is discretised into steps $\Delta\omega$ of for exampe 1˚, and steps $\Delta\rho$ of $\sqrt{2}$ px [**109**]. Maxima/minima in the Hough space correspond to straight lines of raised/lowered intensity in the EBSPs, which can be easily computed and the Kikuchi bands identified by convolution with an appropriate (for example Sobel) kernel. Given a band centre identified by (ω, ρ) in the Hough transform, with appropriate treatment of the pattern centre a plane normal can be identified:

$$T_x = \frac{S_x}{S_y}\left(\frac{\frac{S_x+1}{2}-1}{S_x-1}\right) - \frac{PC_x}{PC_z} \qquad \textit{LR-9}$$

$$T_y = \left(\frac{\frac{S_y+1}{2}-1}{S_y-1}\right) - \frac{1-PC_y}{PC_z} \qquad \textit{LR-10}$$

$$T_s = \frac{1}{PC_z\,(S_x-1)} \qquad \textit{LR-11}$$

$$\hat{\boldsymbol{n}} = \left[\,0, -T_x + T_y\tan(\omega) - \frac{T_s\,\rho}{\cos(\omega)}, \tan(\omega)\,\right] \times \left[\,\tan(\omega), 1, 0\,\right] \qquad \textit{LR-12}$$

For screen size S_x-by-S_y in px and pattern centre co-ordinates ($PC_{x,y,z}$) in fractions of EBSP width, height and height respectively[1], and the cross-product $\hat{\boldsymbol{n}}$ gives the plane normal [**119**]. The PC is a critical parameter, corresponding to the position on the detector with shortest Euclidean distance to the scan point. The vector $\hat{\boldsymbol{n}}$ is inverse-gnomonically projected back into the microscope frame of reference to derive the orientation of the considered plane. If sufficient

[1] As a convention, in this thesis PC_x is measured from the left edge of the pattern, parallel to the sample tilt axis (normalised by pattern width). PC_y is measured from the top edge (normalised by pattern height), and PC_z is the detector distance (normalised by pattern height).

planes' orientations are well identified for a given EBSP, calculation of interplanar angles and comparison to possible structures allows phase and orientation determination [**109**]. Straightforwardly:

$$\cos(\gamma) = |\hat{\boldsymbol{n}}_1 \cdot \hat{\boldsymbol{n}}_2| \qquad LR\text{-}13$$

For crystallographic planes with normals $\hat{\boldsymbol{n}}_1$ and $\hat{\boldsymbol{n}}_2$, and γ the interplanar angle. The procedure of Wright & Adams [**120**] compares identified plane normals' interplanar angles γ_n to the corresponding set a given crystal symmetry. They did not directly Hough transform measured patterns, instead convolving a kernel with the Hough transform of pixels clustered according to the direction of their intensity gradient. After interplanar angle identification, the set is compared to geometrically expected pairings for candidate structures. For example in the FCC structure 25.94˚ separates (200) and (311) ; 29.50˚ separates (111) and (131), *etc* [**120**]. All possible triplets of measured interplanar angles for each candidate phase are identified, and (geometrically) possible solutions identified for each band, agreeing with the lookup tables to within a specified precision. The most likely {hkl} and phase assignment for each band is accordingly identified, by voting across all triplets. At this point the pattern is indexed, and crystal orientation and phase are determined. This led to the first presentations of (now ubiquitous) automatically indexed orientation images [**111**], [**121**]. A confidence index (CI) in the assignment is often employed, equal to the difference in votes between the two most popular candidates divided by total number of votes [**122**].

Further developments in LUT indexing of EBSPs include treating the problem as an analogue of identifying stars in the night sky [**119**]. Generation of an interplanar angle LUT for each given triplet of $\hat{\boldsymbol{n}}_i$ and candidate structure is performed, and a 'characteristic triangle' with sides $\boldsymbol{v}_i = \hat{\boldsymbol{n}}_j - \hat{\boldsymbol{n}}_k$ inferred. The ratios of the side lengths, $|\boldsymbol{v}_i| \,/\, |\boldsymbol{v}_j|$ are stored and compared to those calculated for triplets of experimentally identified bands to obtain a solution. In such a scheme, the pattern centre accuracy can be determined *via* the mean angular error between bands {hkl} and their locations in the best matching solution.

There is rarely a unique orientation solution, especially in metallurgy where highly symmetric crystals are typically of interest. A cubic orientation for example has 24 equivalent orientations. Dingley & Wright [**123**] further developed phase identification in EBSD *via* analysis of zone axis symmetries. Potential symmetry axes (PSAs) are identified by measuring orientations with respect to all possible pairs of orientation solution to an identified triclinic unit cell. The PSA type is identified: one of 60˚, 90˚, 120˚, 180˚ rotations. All possible triplets of PSA are identified, and the corresponding crystal structure type voted on (*eg*. tetragonal structures only exhibit 90˚ and 180˚ PSA types) and a candidate unit cell determined.

The richness of information in EBSPs is not limited to band locations and symmetry. If the pattern centre is accurately known, lattice parameter ratios and angles between basis vectors can be determined [**114**]. The angular width of the Kikuchi bands is approximately twice the Bragg angle for any given plane, but the gnomonic projection destroys inter-band proportionalities in the detector frame. Dingley & Wright suggested an accuracy limit of approximately 5% for band-width determined lattice parameter observations [**123**]. By correcting gnomonic distortions and fitting zone axis locations and manually measured band widths, Li & Han [**124**] surpassed this limit, determining the Bravais structure and unit cell parameters to within 4%, intentionally without any *a priori* knowledge of the material they were characterising. As pointed out by Nolze & Winkelmann [**114**], band width measurements are extremely sensitive to accuracy in knowledge of the pattern centre, due to the inverse gnomonic transform that must be employed. Sensitive analysis of high quality and correctly treated EBSPs permits mapping of for example *c*/*a* ratio in steel martensites at

very high spatial resolution, making no assumptions about the stress state and relying solely on the properties of the projective transformation that is the gnomonic projection [**125**].

3.2.2 Improving pattern fidelity

When signal-to-noise is poor, data reduction or pattern averaging approaches can be immensely useful to obtain more confident and accurate orientation and phase solutions. Kikuchi bands may be weak due to very short exposure times, extensive plastic strain (such that lattice rotation gradients within the sampling volume disrupt elastic diffraction), or a material/phase-specific tendency to weakly diffract (due to a low atomic scattering factor). This is the case in many Zr and Ti alloys, for example [**126**], [**127**]. Microstructural AOIs that can be scanned with EBSD are typically limited in size/resolution by the time taken to expose so many patterns. Wright *et al* [**128**] introduce two new methodologies that leverage locality of measurements to improve confidence in indexing when patterns are noisy. The 'extra solutions clean-up' approach examines the CI of an orientation solution and its runners up (from the triplet method [**120**]), and if below a given threshold its neighbours are examined. If a less popular orientation in the voting scheme matches its neighbours to within a tolerance, this is selected over the (low confidence) most popular triplet solution. Additionally neighbour pattern averaging and re-indexing (NPAR) is presented, in which any one measured pattern is kernel averaged with its neighbours (with equal weighting). The remarkable results of this approach on improving indexing with simulated pattern noise are presented in **Figure LR-13**, with noise levels 0, 0.5, 0.8, 1. The noise levels correspond to the standard deviation of a normalised and mean centred Gaussian probability density function. The orientation map with NPAR is fully recovered, even at extremely high artificial noise levels. As the authors point out, errors in indexing are particularly prevalent at grain boundaries when using this naïve averaging due to the enforced pattern overlap and blurring, especially at higher step sizes [**128**]. Spatial resolution drops, preventing observation of fine features, even when including a similarity-based weighting into the averaging kernel.

An improved version of NPAR, non-local pattern averaging and re-indexing (NLPAR), has recently been presented by Brewick *et al* [**129**]. For a test EBSP, the corresponding NLPAR pattern is a weighted average of all patterns in a given window (which could be the entire dataset), and weighted by a function of the L2 norm of the separating vector in feature space. The approach includes normalisation for the effect of expected noise in the dataset, which is calculated with the variation of individual pixels. An NLPAR-corrected pattern $\widehat{\boldsymbol{p}}_i$ is given by:

$$\widehat{\boldsymbol{p}}_i = \sum_{j=1}^{W} w_{ij} \boldsymbol{p}_j \qquad \textit{LR-14}$$

For measured (flattened) pattern vectors $\boldsymbol{p}_j$ in a search window W. The weighting coefficient w_{ij} is given as:

$$w_{ij} = \frac{1}{Z_{ij}} \exp\left(-\frac{d_{ij}^*}{\lambda^2}\right) \qquad \textit{LR-15}$$

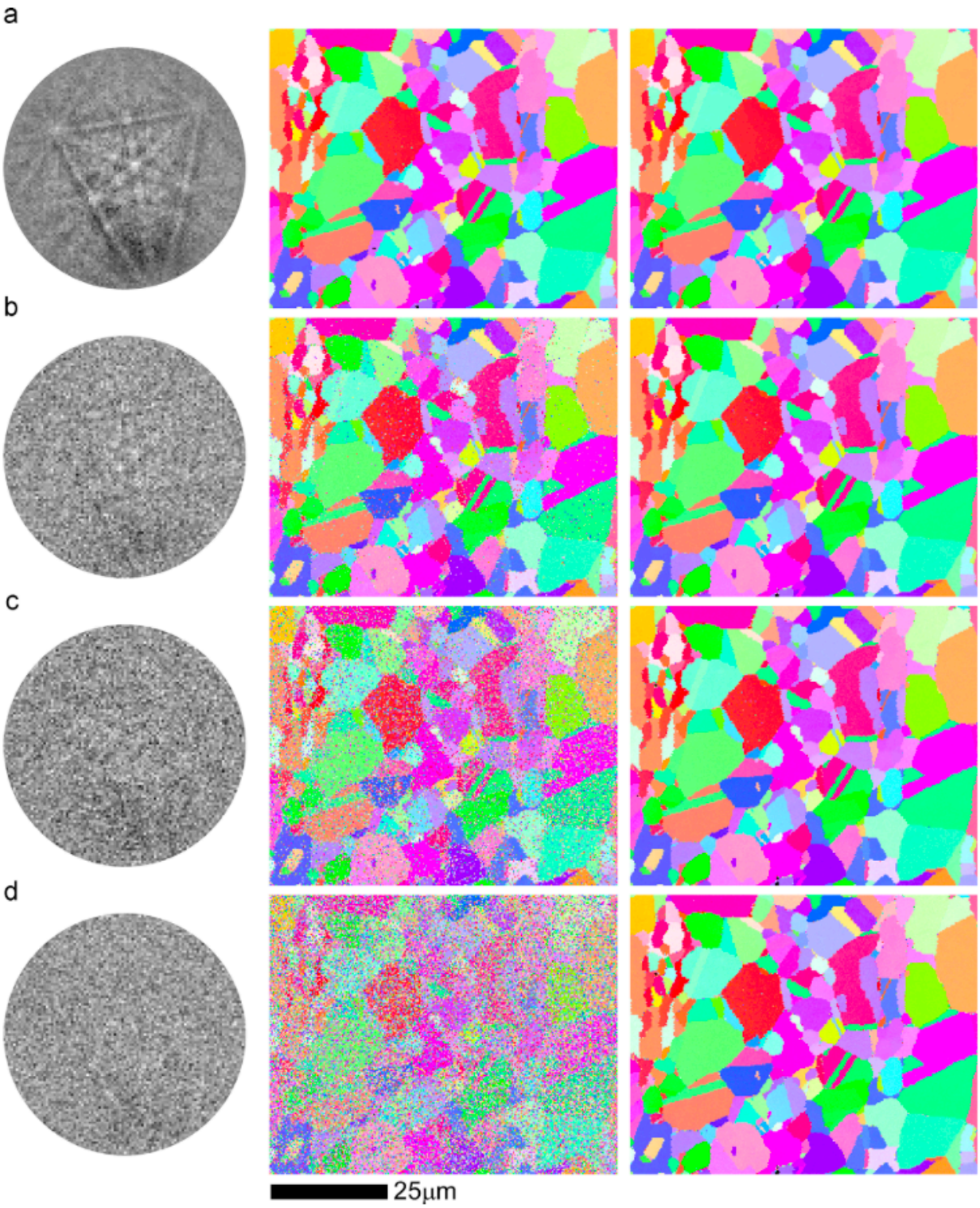

Figure LR-13: The NPAR approach, with noise levels at (a) 0, (b) 0.5, (c) 0.8, (d) 1. Reproduced from [**128**].

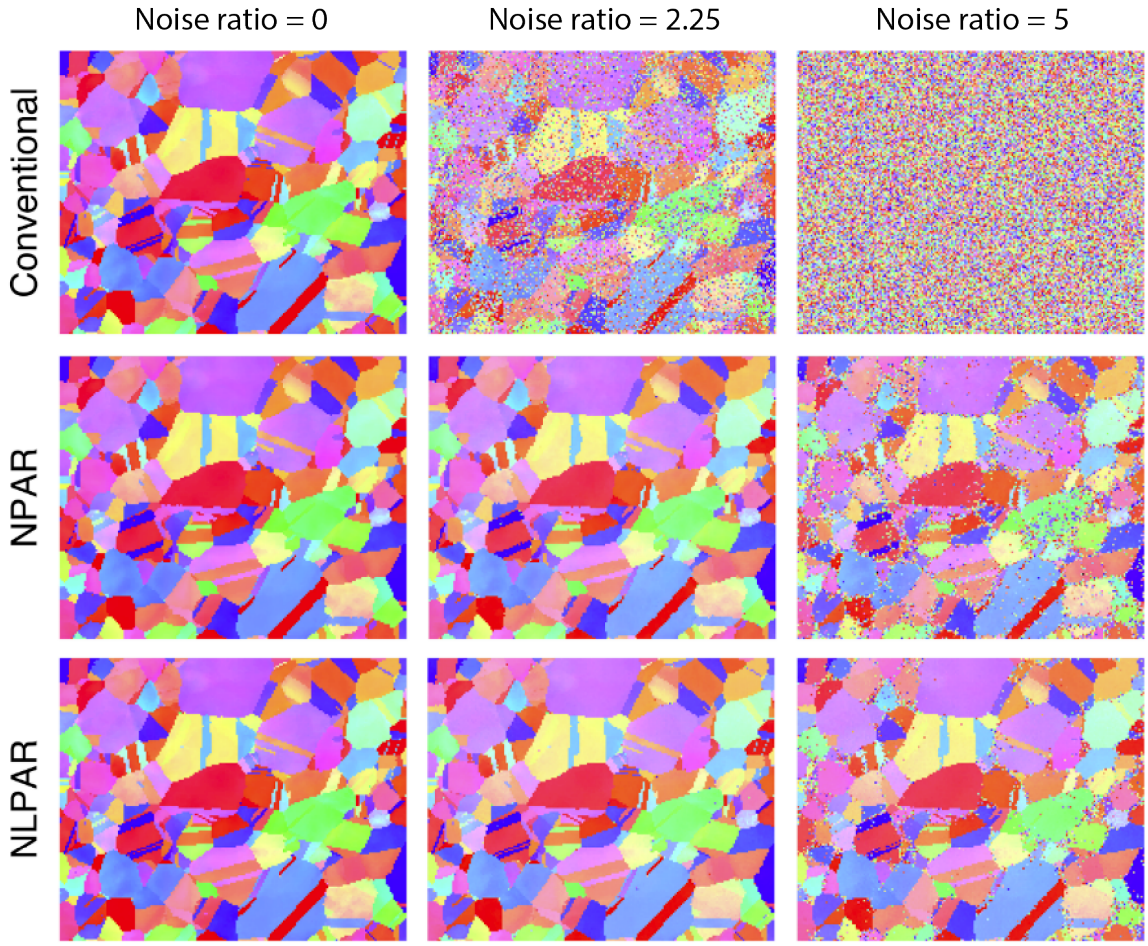

Figure LR-14: The NLPAR approach, operating with noise levels of 0, 2.25, and 5. Reproduced from [**129**].

$$Z_{ij} = \sum_{j=1}^{W} \exp\left(-\frac{d_{ij}^{*}}{\lambda^2}\right) \qquad \textit{LR-16}$$

Where Z_i is employed for normalisation, and tuning parameter λ used. The expectation and variance of the L2 norm of the feature difference vector are straightforwardly:

$$E\left(\left\|\boldsymbol{p}_i - \boldsymbol{p}_j\right\|_2^2\right) = \left\|\boldsymbol{p}_i - \boldsymbol{p}_j\right\|_2^2 + n\left(\sigma_i^2 + \sigma_j^2\right) \qquad \textit{LR-17}$$

$$\mathrm{Var}\left(\left\|\boldsymbol{p}_i - \boldsymbol{p}_j\right\|_2^2\right) = 4\left(\sigma_i^2 + \sigma_j^2\right)\left\|\boldsymbol{p}_i - \boldsymbol{p}_j\right\|_2^2 + 2n\left(\sigma_i^2 + \sigma_j^2\right) \qquad \textit{LR-18}$$

For n the length of pattern vectors $\boldsymbol{p}_i$, and σ_i the standard deviation of the (Gaussian, mean-centred) noise in the i^{th} pattern. Assuming dominance of the second term in the variance, the normalised distance metric d^* is then given by:

$$d_{ij}^{*} = \frac{\left\|\boldsymbol{p}_i - \boldsymbol{p}_j\right\|_2^2 - n\left(\sigma_i^2 + \sigma_j^2\right)}{\sqrt{2n}\left(\sigma_i^2 + \sigma_j^2\right)} \qquad \textit{LR-19}$$

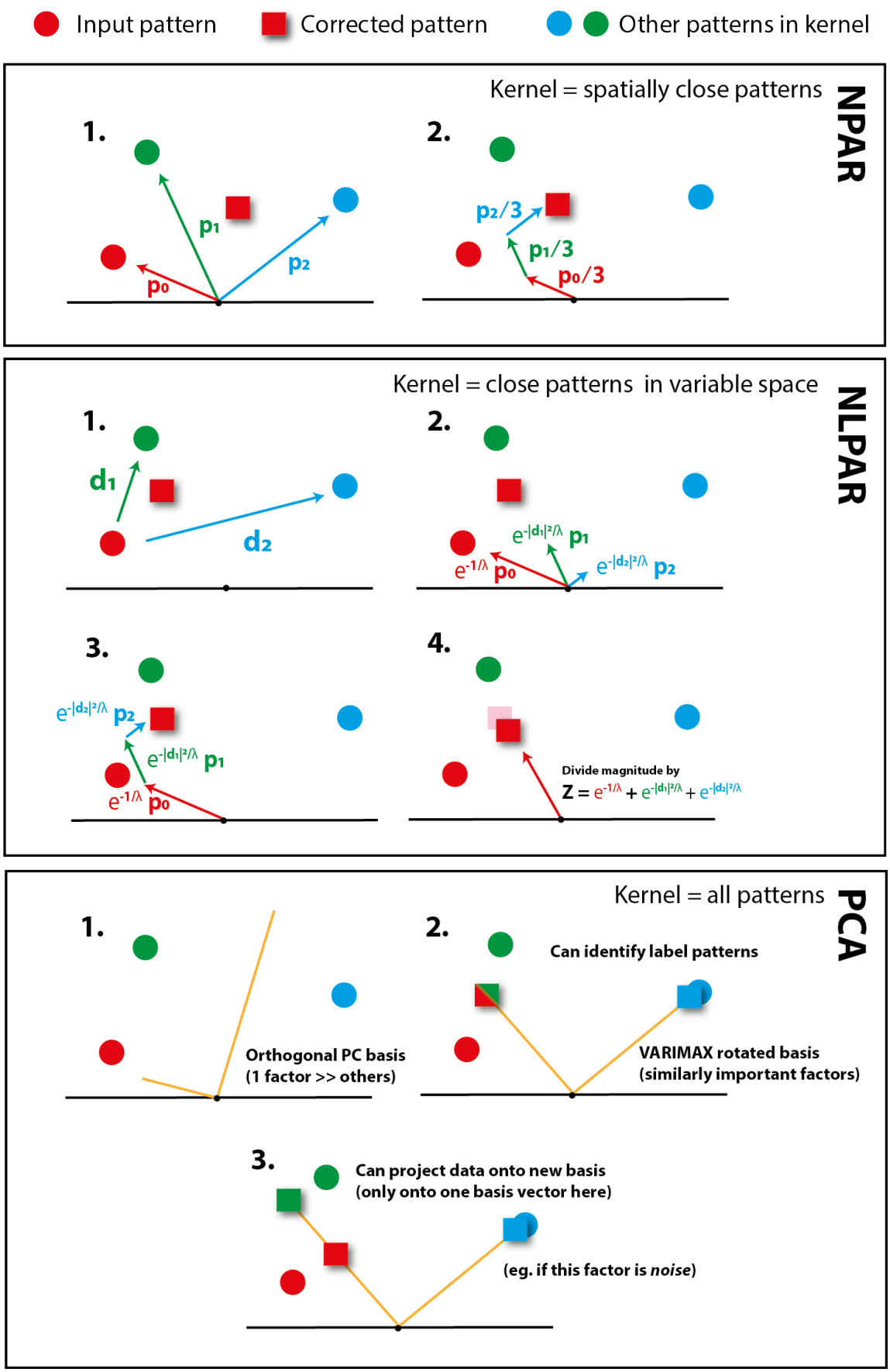

Figure LR-15: A geometrical comparison between the operation of NPAR, NLPAR, and PCA.

The standard deviations in the noise are calculated pattern-specifically by assuming that neighbouring patterns should have zero feature space Euclidean separation and equal noise Gaussian variance. Accordingly, the σ_i^2 are selected to be equal to *2n* times the minimum L2 Euclidean separation of the i^{th} pattern's nearest neighbours.

The comparison presented by Brewick *et al* [**129**] between NLPAR, NPAR and conventional pattern treatment (all Hough/triplet indexed) is shown in **Figure LR-14**. Noise is added to the EBSPs with standard deviation proportional to the overall standard deviation of the patterns themselves, with scalar multiple equal to the noise ratio. Both NPAR and NLPAR present a vast improvement over conventional indexing at high noise levels.

Brewer *et al* [**130**] and Wilkinson *et al* [**131**] have developed an approach to EBSP clustering based on principal component analysis (PCA), which is developed further in this thesis. A detailed exploration of PCA is presented in **Section LR.4.2.1**, and the approach makes up a substantive part of this thesis. Briefly, a new orthogonal basis is sought for the data, onto which the original measurements are projected. The new basis necessarily explains the greatest amount of dataset variance for a given data matrix rank. A rotation of the set of principal component EBSD patterns that maximises the variance between each member of the set effectively reduces a full EBSD dataset down to a single representative, or 'characteristic' EBSD pattern for each commonly labelled domain. If the number of components is well selected, prior to VARIMAX rotation, then these can correspond to a single pattern per grain, and for oversampled or deformed grains the domains may also correlate with sub-grains [**131**]. This approach (in its basic form) completely discards spatial information, unlike NLPAR which retains a small amount in the form of limiting the search window. PCA fully reduces the dataset to representative patterns with minimised L2 Euclidean feature difference vectors to the new basis.

NLPAR does not calculate such a basis or reduce the dataset, but in the limit of considering the full map as the search window a similar solution could be obtained by a PCA reconstruction (taking each scan point as the sum of the PCA projections and basis vectors). The difference between the approaches would then lie in the fact that the output pattern in the NLPAR scheme is an exponential sum of all patterns in the dataset, with weighting equal to the dot product of difference vectors with themselves, while the PCA reconstruction sees a linear sum of the first *k* de-noised representative patterns, with weighting equal to the dot product of the test pattern and the representative pattern. The effect in both cases is to reduce the noise: in the NLPAR case this is by decaying the random variations by averaging a negative exponential, and in the PCA case it is by throwing away higher order latent signals that contribute minimally to total dataset variance (arguably the definition of noise). A 2D (ie. n=2) graphical comparison between NPAR, NLPAR, and PCA is presented in **Figure LR-15**. In this illustrative schematic the noise standard deviations in NLPAR are assumed to be one.

3.2.3 Simulation & template matching

Due to advancements in feasibility of data transfer and storage, full sets of EBSPs can now be captured and stored at high resolution for *offline* analysis. This permits more computationally heavy indexing approaches [**129**]. Combined with data reduction approaches such as PCA, very swift and accurate microstructural phase analysis becomes achievable. 'Template matching' or 'dictionary indexing' approaches compare experimental patterns to libraries of (dynamically) simulated candidates at known orientation, for a given, calibrated microscope setting.

Accurate template matching relies on well simulated candidate EBSPs. Kinematical simulations are able to provide expected band presence / absence, but in order to accurately model *intensities* dynamical simulations must be performed

[**122**]. Dynamical diffraction modelling is challenging enough for elastic, TEM spot patterns, and the prospect of including prior inelastic scattering is a daunting challenge [**107**], [**132**].

Signal in a simulated pattern is treated as proportional to the square of the intensity of the superposed Bloch wave field for a given exit point on the phosphor screen [**133**]. A point within a simulated crystal can be treated as a virtual incoherent source, with the backscattering event randomising the phase difference between the point source and the incident Bloch wave. This assumption effectively separates the inelastic scattering process from the subsequent elastic diffraction, such that no coherence to the incident electron beam is left [**132**]. The problem is not sensitive to assumptions on the prior distribution of inelastically scattered electron density (before elastic diffraction), as of greater interest than absolute intensities are the relative differences between bands and features on a diffuse inelastic background.

The reciprocity principle states equivalence between starting a plane Bloch wave at the detector and tracing it back to the scattering unit cell, and the reverse. The former is employed in dynamical simulations, as it permits the use of a single plane wave to infer intensities at the scattering atoms. After Callahan & de Graef [**134**], intensity I for a wavevector $\boldsymbol{k}$ is given:

$$I(\boldsymbol{k}, E, z_0(E)) = \sum_{i}^{N} \frac{1}{z_0} \int_0^{z_0(E)} \lambda(E, z)\, |\Psi(\boldsymbol{r}_i)|^2 \, dz \qquad \textit{LR-20}$$

Where Ψ is the electron Bloch wavefunction at atom position $\boldsymbol{r}_i$ in the unit cell, z is sample thickness up to z_0, E is electron energy, N is the number of atoms in the unit cell. The modification presented by Callahan & de Graef [**134**] is to include the function λ, accounting for variation in electron yield with energy and depth. Additionally, the Bloch wavefunction is corrected for thermal excitations using the Debye-Waller factor, and weighted proportionally to the atomic number to account for propensity for inelastic scattering. Typically, one specifies a set of lattice planes to contribute to the solution, for example all reflectors within 50˚ of the [001] zone axis with minimum spacing 0.35 Å. The strongest beams are usually directly accounted for, with a larger number included according to the Bethe perturbation, which transforms the strong reflectors' calculated periodic potential (which is an important parameter of the wavefunction) by an excitation error due to the presence of the weaker reflectors [**135**].

Once a suitable library of accurate EBSPs has been generated, it can be sampled for candidate orientations (for example evenly in the SO(3) rotation group) and gnomonically projected into the detector frame. Thus a set of templates for varied crystal unit cell orientation is obtained, and the most similar template to the measured pattern provides the orientation of the unit cell, to within sampling precision.

For a template library of vectorised patterns, The normalised dot product (NDP) comparison metric presented by Chen *et al* [**136**] is given:

$$C = \frac{1}{d} \frac{\boldsymbol{m}^T \boldsymbol{e}}{\boldsymbol{m}^T \boldsymbol{m} \times \boldsymbol{e}^T \boldsymbol{e}} \qquad \textit{LR-21}$$

For (vectorised) template pattern $\boldsymbol{m}$ and experimental pattern $\boldsymbol{e}$, and d the length of the pattern vectors (the number of pixels). The NDP is used to determine an experimental pattern's k nearest neighbour dictionary patterns in orientation space. Given these, the average orientation of the experimental pattern is inferred using a maximum likelihood

optimisation procedure. The nearest k orientations form the set $\{\boldsymbol{\theta}_j\}^k$, for $\boldsymbol{\theta}_j$ each a vector of Euler angles. The ML estimate of $\boldsymbol{\theta}$, the experimental orientation is given by the estimate $\widehat{\boldsymbol{\theta}}$:

$$\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} \prod_{j=1}^{k} f(\boldsymbol{\theta_j}\,;\,\boldsymbol{\theta}) \qquad \textit{LR-22}$$

Such that the $\boldsymbol{\theta_j}$ are independent and identically distributed with an underlying marginal probability density $f(\boldsymbol{\theta_j}\,;\,\boldsymbol{\theta})$ that operates on the orientation sphere**.** This formulation identifies the $\boldsymbol{\theta}$ for which the product of probabilities of obtaining $\boldsymbol{\theta_j}$ with a given $\boldsymbol{\theta}$ is maximised**.** The probability density f is taken as the Von Mises-Fisher density, which evaluates onto a unit sphere in quaternion space with a given mean location and spread.

As discussed by Foden *et al* [**137**], the NDP metric is translationally sensitive, such that semantically similar patterns translated by a few pixels may seem to have total dissimilarity, when there is in fact a great degree of shared information. The NDP is a special case of the normalised cross correlation function (XCF), specifically its value at zero lag in the horizontal and vertical direction. Fourier cross-correlation sees many applications requiring registration of images and interpretation of translations in signals [**138**]. In the formalism of Foden et al [**137**]:

$$C_{ab} = \sum_{x,y} M_{x+a,y+b}\, E^*_{x,y} \qquad \textit{LR-23}$$

$$C_{ab} = \mathrm{F}^{-1}[\,\mathrm{F}[M_{x+a,y+b}] \circ \mathrm{F}[E_{x,y}]^*\,]$$

The XCF determines how well reference and test EBSPs align at lag i and j in the horizontal and vertical directions respectively, and F denotes the 2D discrete Fourier transform (DFT), calculated for example with the Cooley-Tukey fast Fourier transform (FFT) algorithm [**139**].

$$\mathrm{F}[\,M(\mathrm{x},\mathrm{y})\,] = m(u,v) \qquad \textit{LR-24}$$

$$m(u,v) = \sum_{y=1}^{Y}\sum_{x=1}^{X} \frac{M(x,y)}{\sqrt{XY}} \exp[-2\pi i\,\left(\frac{ux}{X}+\frac{uy}{Y}\right)] \qquad \textit{LR-25}$$

For image dimensions (X,Y). The inverse is given:

$$M(\mathrm{x},\mathrm{y}) = \mathrm{F}^{-1}[\,m(u,v)\,] \qquad \textit{LR-26}$$

$$M(x,y) = \sum_{v}^{Y}\sum_{u}^{X} \frac{m(u,v)}{\sqrt{XY}} \exp[2\pi i\,\left(\frac{ux}{X}+\frac{uy}{Y}\right)] \qquad \textit{LR-27}$$

Such that:

$$C_{ab} = \sum_{v}^{Y}\sum_{u}^{X} m(u,v)\; e^*(u,v) \exp[2\pi i\,\left(\frac{ua}{X}+\frac{ub}{Y}\right)] \qquad \textit{LR-28}$$

The maximum value of the XCF determines absolutely how well images *can* align with translational freedom (which could be attributed to an incorrectly simulated pattern centre, or misorientation). Using the FFT and an appropriate sub-pixel registration algorithm, resolved image shifts can very quickly be determined [**138**]. Similar approaches have been successfully used to infer the elastic strain state within a crystal by tracking EBSP sub-regions of interest [**127**], [**140**], [**141**]. By permitting translational freedom in the comparison metric far fewer templates need to be sampled to obtain the same orientation resolution. This provides massive computational speedup, as the FFT leverages computational complexity of $O(\, n \log n \,)$ [**139**]. In order to achieve subpixel registration (i.e. efficiently upsample and interpolate in the frequency domain), Foden *et al* [**137**] employ the cross-correlation registration algorithm developed by Guizar-Sicairos et al [**142**]. This avoids the computationally expensive procedure of embedding $m(u,v)\, e^{*}(u,v)$ into an upsampled grid of zeros.

In an upsampled DFT, the need for zero-embedding is alleviated if the peak is expected to be close to autocorrelation. Guizar-Sicairos *et al* [**142**] implement this as the product of three matrices, with dimensions $(1.5\alpha,\, Y)$, $(Y,\, X)$, and $(X,\, 1.5\alpha)$, after translating the registration to be close to an initial estimate in peak location from a non-upsampled cross-correlation. Subsequently, the XCF is computed for an upsampled square array of dimension 1.5α [**142**]. This presents computational complexity $O(MN\sqrt{\alpha})$, much faster than the zero-embedding and inverse transform approach, which requires complexity $O(XY \log(\alpha X) + \alpha \log(\alpha Y))$, for $Y \leq X$.

Foden *et al* [**137**] efficiently compare a set of templates to an experimental pattern. The translational insensitivity of the XCF permits the sampling frequency of the SO(3) rotation group to be fairly large, for example 8°, much higher than required by the NDP approach. This permits a swift pass through the library. SO(3) is split into $S^2 \times S^1$, with S^2 the set spanning the surface of a sphere (the polar angle split being equidistantly and azimuthally sampled) and S^1 (rotation around points on the sphere). The best matching template has its orientation refined.

Small rotations (misorientations between template and measured EBSP) about the sample *x* and *y* axes correspond to translations in the EBSP, which can efficiently be detected in the XCF, provided the PC is well known. The corresponding misorientation (about *x, y* axes) is given simply as shift / detector distance. The rotation about *z* is inferred upon translation to log polar space, parameterised by $\log(r)$ (with r the distance from the image centre) and η the anticlockwise angle from the horizontal axis. Rotations about *z* then correspond to translations in η. A full misorientation between template and measured EBSP can thus be inferred by iteration of these calculations while optimising the pattern centre, and the absolute orientation of the measured pattern determined. This is presented in **Figure LR-16**.

Both the NDP and FFT approaches to template matching yield substantive improvement in orientation precision than the standard Hough-based method [**136**], [**137**], [**143**], [**144**]. The FFT approach yields accuracy of < 0.2° for 128-by-128 px EBSPs [**137**]. The NDP approach yields accuracy of approximately 0.2° for 120-by-120 px EBSPs [**143**]. These are vast improvements over the conventional Hough transform approach, which is typically on the order of 0.7° for 120-by-120 px EBSPs [**143**].

3.2.4 Spherical EBSD

A major caveat in analysis of the gnomonic projection of the diffraction sphere onto an inclined planar screen leads to *hyperbolic* band edges rather than straight lines [**109**], [**119**], [**145**]. A more natural basis in which to conduct EBSP analysis is the inverse gnomonic projection of the measured patterns back onto the diffraction sphere. The subject has

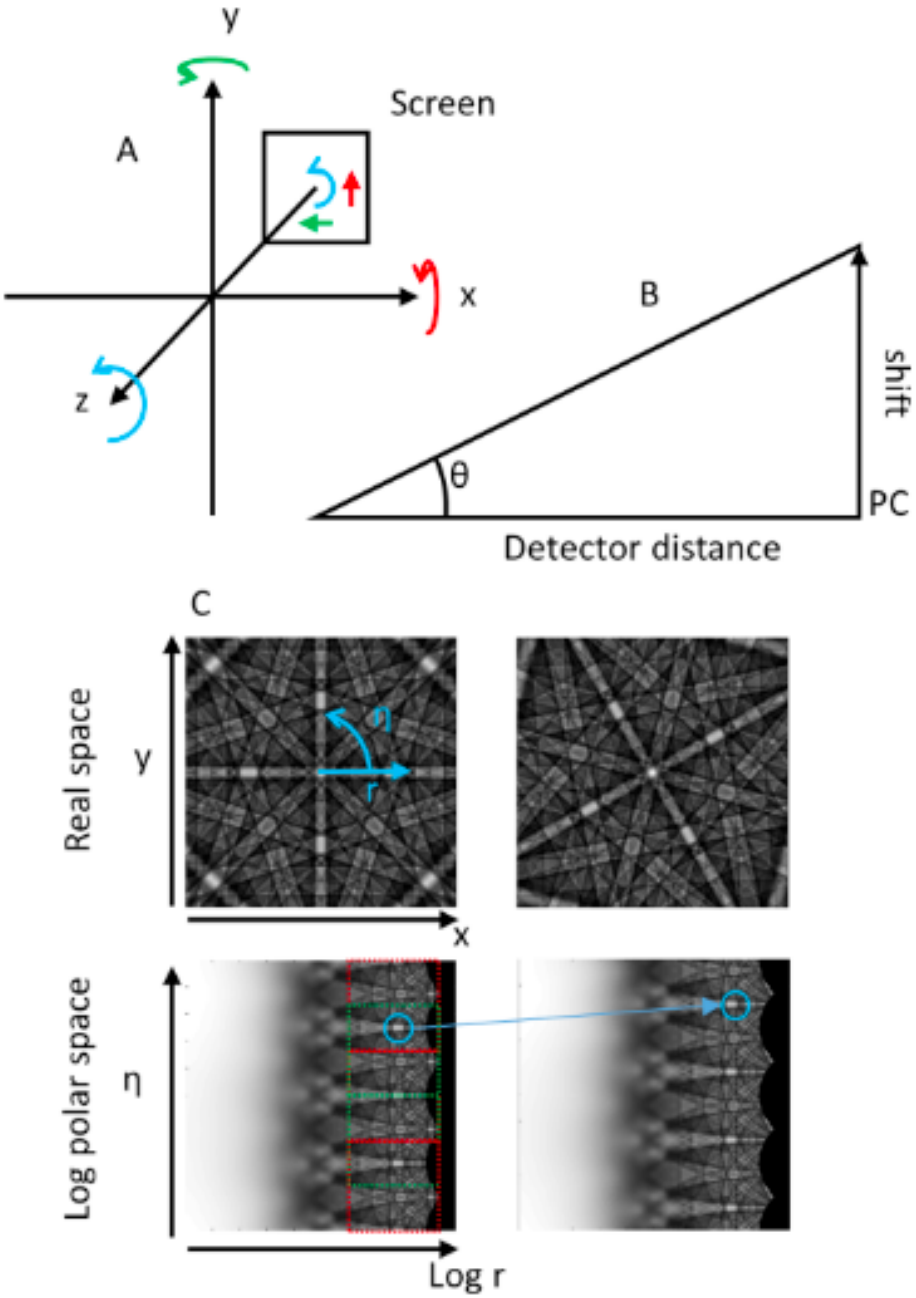

Figure LR-16: Log-polar refinement after and reproduced from Foden et al [**137**].

been reviewed and opportunities for the field discussed by Day [**146**]. An illustration of the problem is presented in **Figure LR-17**. Of particular note is the overlap of this problem with the field of robotics; a spherical Hough transform has been developed in order to assist robots navigating through straight lines in an office space. Kikuchi bands 'on-the-sphere' are inverted to discs centred at the plane normal, with diameter equal to the band width. Analysis 'on-the-sphere' permits spherical cross-correlation and accurate analysis of Kikuchi band cross-sectional profiles, as the hyperbolic divergence is removed. These problems have been presented and solved by Hielscher *et al* [**145**]. This inverse gnomonic projection is achieved by calculating a function f with respect to diffraction directions $\boldsymbol{\xi}$. This follows the expansion:

$$f(\boldsymbol{\xi}) = \sum_{a=0}^{N} \sum_{b=-\beta}^{\beta} \hat{f}(a,b)\, Y_a^b(\boldsymbol{\xi}) \qquad \textit{LR-29}$$

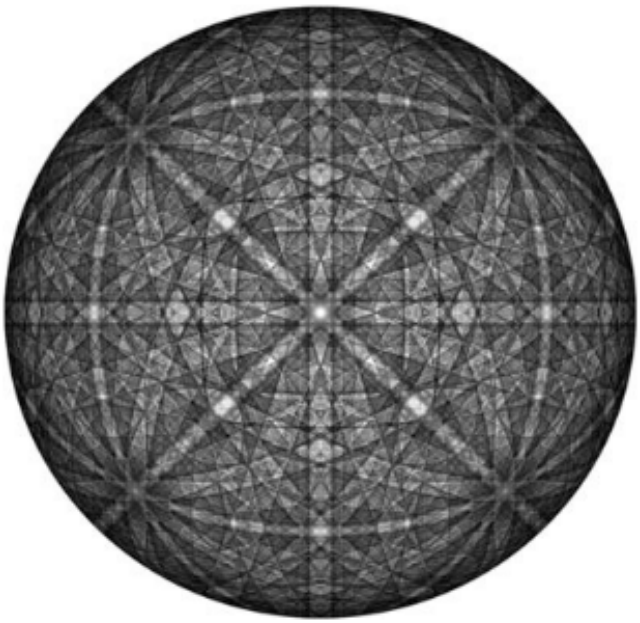

Figure LR-17: A spherical visualisation of the diffraction sphere, reproduced from Day [**146**].

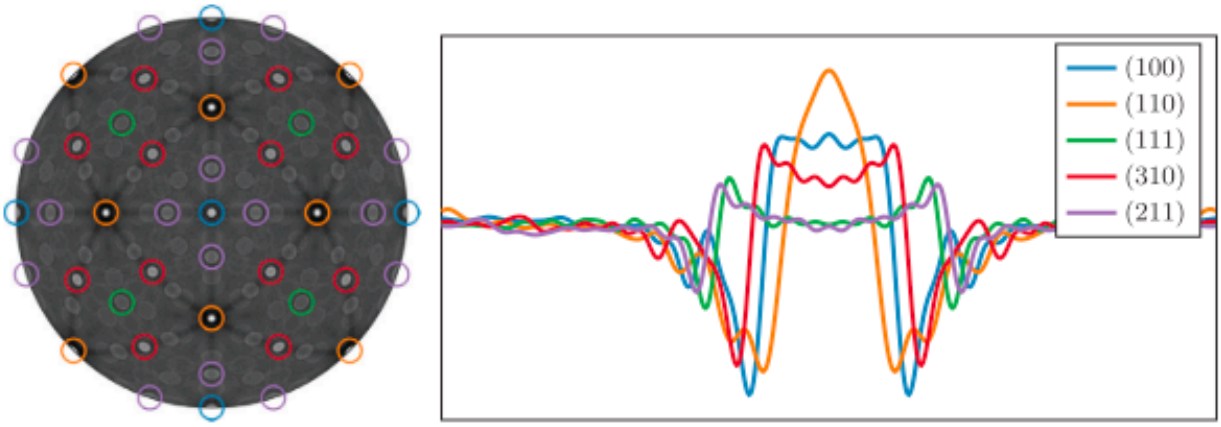

Figure LR-18: (a) the 2D projection of the spherical Radon transform, and (b) integrated band profiles of a diffraction sphere, reproduced from [**145**].

Where Y_a^b are the spherical harmonic functions, and $\hat{f}(a,b)$ are the Fourier coefficients of f. N corresponds to the degree of harmonic employed for the expansion. The spherical harmonics replace the basis of the clasical Fourier expansion. Hielscher *et al* [**145**] discuss several approaches for calculating the Fourier coefficients from diffraction intensities, including estimating using quadrature (exploiting the orthogonality of the spherical harmonic basis), directly interpolating from the measured values, or approximation by fitting a hyperplane to minimise a square error loss function in an overdetermined system of equations. Once the inverse projection has been mapped, spherical Fourier and Radon transforms as well as cross-correlation can be achieved.

If the pattern centre and corresponding orientation are well known, crystallographically expected band traces can be mapped onto a spherical inverse-gnomonically projected EBSP. This permits integration along all the small circles

following the surface of the sphere outwards from the band centre, to calculate an average band cross-sectional profile. Such an analysis on a dynamically simulated pattern is presented in **Figure LR-18**.

The shapes of these profiles are characteristic to the projecting bands. Leveraging this for new crystallographic imaging is explored further in **Chapter 3**.

3.3 4D-STEM analysis

Four dimensional scanning transmission electron microscopy (4D-STEM) is a relatively new technique in which an electron spot diffraction pattern is acquired at every point of a scan grid in a TEM. In this regard it is similar to electron backscatter diffraction (EBSD), now a routine method for microscale structural analysis [**109**], [**117**], but a much finer 'pencil-beam' probe permits sub-nm spatial resolution. The trade-off is that a zone-axis generally must be identified and aligned with the transmitted beam, inherently limiting knowledge of the reciprocal lattice to two coplanar vectors. A comprehensive review of 4D-STEM and its applications in strain mapping, imaging, and ptychography is available in ref [**147**]. In this work we employ the py4DSTEM open source software package, developed by Savitzky *et al* [**148**]. Lattice strain measurement with this approach is becoming fairly routine. It has been used to investigate resistivity in semiconductors [**149**], [**150**], and more recently begun to be applied to polycrystalline materials [**151**]. Pekin *et al* [**152**], [**153**] have measured the strain field around austenitic (FCC) stainless steel features. They observed a ~4% variation in strain across their area of interest, which included dislocations and an annealing twin boundary.

3.3.1 Bragg disc location

As discussed by Pekin *et al* [**153**], an extension to cross-correlation (extensively discussed in **Section LR.3.2.3**) is phase correlation, in which only the frequencies of the Fourier components (and not their magnitudes) are allowed to influence the measure:

$$\mathrm{F}[C_{ab}] = \frac{\mathrm{F}[M_{x+a,y+b}] \circ \mathrm{F}\left[E_{x,y}\right]^{*}}{\left|\mathrm{F}[M_{x+a,y+b}] \circ \mathrm{F}\left[E_{x,y}\right]^{*}\right|^{r}} \qquad \textit{LR-30}$$

In which r is the power to which the Fourier coefficients are weighted. Cross-correlation corresponds to $r = 0$, and phase-correlation when $r = 1$. Such an approach can be useful when edges and other high frequency components are of interest, in their case when aligning a circular probe with TEM discs, for example. This is not of particular concern in EBSD analysis, in which the alignment of centres of mass of images is of primary concern (the low frequency terms are more important). In 4D-STEM strain analysis, for example, a hybrid cross-correlation with $0 < r < 1$ is often employed [**154**].

3.3.2 Measuring strain

After Bragg discs have been identified for each diffraction pattern, a relative in-plane lattice strain map can be produced by comparing their locations in reciprocal space to a reference lattice. This reference lattice is usually taken to be a partial average, or reference portion of the area of interest [**148**]. In the scheme of Béché *et al* [**155**]:

$$\mathbf{G} = \begin{bmatrix} g_{1,i} & g_{2,i} \\ g_{1,j} & g_{2,j} \end{bmatrix} \qquad \textit{LR-31}$$

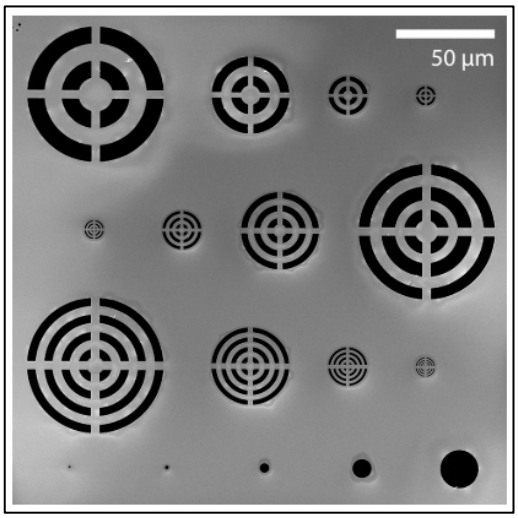

Figure LR-19: Novel aperture designs for better registration of Bragg discs, after and reproduced from [**157**].

Where the vector component $g_{1,i}$ for example refers to the i (row) component of the first reference vector in the diffraction pattern, tracking the position of one (*hkl*) spot. The components $g_{2,i}$ and $g_{2,j}$ track a second location[2]. Given an 'unstrained' set of basis vectors:

$$\mathbf{G_0} = \begin{bmatrix} \overline{g_{1,i}} & \overline{g_{2,i}} \\ \overline{g_{1,j}} & \overline{g_{2,j}} \end{bmatrix} \qquad \textit{LR-32}$$

The distortion, **D**, strain **ϵ**, and rotation **Ω** can be obtained, using an infinitesimal decomposition:

$$\mathbf{D} = (\mathbf{G_0^T})^{-1}\mathbf{G}^T - \mathbf{I} \qquad \textit{LR-33}$$

$$\boldsymbol{\epsilon} = \frac{1}{2}(\mathbf{D} + \mathbf{D}^T) \qquad \textit{LR-34}$$

$$\boldsymbol{\omega} = \frac{1}{2}(\mathbf{D} - \mathbf{D}^T) \qquad \textit{LR-35}$$

A more detailed derivation, and procedures for correcting elliptical distrotions, has been presented in Appendix F of the recent work of Savitzky *et al* [**148**], and included in the open source *py4DSTEM* software package. This approach to strain mapping is very sensitive to spot location, and improvements in precision thereof directly translate to higher resolution strain mapping. Béché *et al* [**155**] perform an analysis of SiGe thin films, and determine a strain precision of 6×10^{-4} at 2.3 nm step size, though they acknowledge in reality that in practice accuracy is limited by stress relaxation of samples. An improvement in spatial resolution was presented by Ozdol *et al* [**156**], measuring at 10×10^{-4} and 1 nm step size, using modern direct electron detection and fast scan times.

Recently, Zeltmann *et al* [**157**] have proposed an improved approach for identifying Bragg discs to high precision, by including a mask in the objective aperture to produce a specific shape (such as a bullseye) with sharp edges that track

[2] The positions tracked by these vectors may be linear combinations of spots, in order to maintain the required orthogonality in crystal systems or diffraction patterns without four-fold (90˚) symmetry.

well in hybrid cross-correlation. Unfortunately the probe size diverges with these modifications, resulting in a loss of scanning spatial resolution. The apertures employed are presented in **Figure LR-19**. Of particular note is that an improvement in strain mapping precision of thick samples (in which there is considerable dynamical electron scattering and uneven disc illumination) of up to 29 times can be achieved through use of these apertures. This vastly offsets the trade-off in spatial resolution.

4. Machine learning for materials science

Especially in the last decade, the remarkable successes of machine learning have inevitably led to its application in a wide range of fields, and materials science is of course no exception. In fact, a variety of classic metallurgical problems perhaps lend themselves very well to ML approaches: alloy design [**158**], microstructural defect identification [**159**], and property prediction [**160**], [**161**] are natural applications of supervised models. Application of such approaches to explore the *very* wide range of compositional possibilities in multi-principle element alloys is perhaps another exciting domain [**162**].

In this chapter, the nomenclature 'model' refers to a computationally implemented approximation to a function, rather than a set of physical rules. On a related note, 'parameters' are said to be optimised during learning, while 'hyperparameters' are fixed constants that define the model's structure and properties, for example the number of layers in a neural network or a regularisaiton coefficient.

Understanding the structure of data with unsupervised learning methods is becoming an important consideration with developments in data capture and storage, especially in electron microscopy and X-ray / neutron diffraction experiments [**130**], [**131**]. In this thesis such approaches are developed further, and unsupervised ML is used to better understand electron microscopy measurements of microstructure. In this section algorithm development and training, neural network structure, and relevant unsupervised methods are reviewed. Subsequently, applications of such approaches in materials science, with a focus on metallurgy and electron microscopy, shall be discussed.

4.1 Learning from data

In order to train a model to accurately perform a specific task, within a given framework, the model must evaluate its 'success' on training examples. This is quantified through use of an objective (or 'loss') function, in a simple case this can be the squared difference between the predicted and true value [**163**]. If a problem is convex there is a single global minimum in the loss function. Equivalently, the second derivative of the function must be positive everywhere. In the multivariate case this is a statement of the Hessian matrix being positive semidefinite. In convex optimisation, the presence of a global minimum means that eventual convergence is guaranteed [**164**]. Machine learning approaches usually attempt to tackle non-convex problems, and as such cannot rely on loss function convergence to determine when the problem has been solved. Instead, the model is fed 'training' examples, and parameters adjusted algorithmically to iteratively improve performance. Subsequently the model is evaluated on unseen 'validation' data, the performance on which gives an indication on how well the trained model can generalise [**163**], [**165**], [**166**]. The phenomenon of 'overfitting' refers to when a model is fed too much training data, such that it learns the labels from specifically those examples. The training error rate and loss function are low, but the model is completely useless when asked to evaluate unseen data. For example, there are sufficient degrees of freedom to exactly fit a 5^{th} order polynomial to three data points, but such a model is very unlikely to accurately represent the actual relationship. Such a case is easily be identified when the model is evaluated on validation data, as the error rate and loss will be extremely high.

This is the paradigm in which supervised machine learning models are usually developed. By dividing known, labelled data into training and validation sets the model can be both fitted and accurately evaluated. In the case of deep learning with neural networks, the model is usually continuously trained through many 'epochs' (rounds of seeing the training data), with parameters adjusted to reduce the loss function at each step, until the validation error doesn't reduce any

further. If the validation error starts to rise the model is overfitting, and the 'learning rate' at which parameters are updated should be decreased [**165**], [**167**].

Most machine learning models are trained using some form of gradient descent, first presented by Cauchy [**168**]. In this scheme a function's minimum is simply identified by continuously following the gradient. The gradient is zero at a local (or global) minimum. In the multivariable case this takes the following form [**164**] :

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} - \alpha\, \mathbf{g}_{n-1} \qquad \textit{LR-36}$$

$$\mathbf{g}_n = \nabla_{\boldsymbol{\theta}_n} \mathrm{L}(\boldsymbol{\theta}_n) \qquad \textit{LR-37}$$

For a loss function L, parameter vector $\boldsymbol{\theta}$, iteration n, and learning rate α. In this scheme the learning rate determines how quickly the function approaches the minimum. If set too low, the optimisation will be very slow. If set too high, the algorithm can quickly diverge, depending on the shape of the loss surface. With knowledge of the gradient, this process is quick and simple to numerically evaluate. Computationally more complicated approaches include a line search, in which the learning rate r itself is optimised, and Newton's method, which considers the loss function as a Taylor series and leverages the Hessian / second derivative to achieve convergence in quadratic time [**169**], [**170**]. In most ML applications a modified version of the vanilla gradient descent is employed, as it is very expensive to calculate second derivatives and there are a lot of parameters to optimise. Furthermore, a less-than-optimal step size can actually lead to improvement in generalisation (and prevention of overfitting) [**171**].

Usually this procedure is slightly modified, as it is expensive to calculate derivatives and adjust parameters after evaluating loss at each and every training example. In the extreme, 'batch' gradient descent evaluates the expected value of the gradient over all the training data. This single average gradient is used to update all the model parameters. However, this is very inefficient and can easily lead to overfitting as the model sees every training example in every iteration. To introduce more variance into the parameters, a 'minibatch' is randomly sampled from the training data, and an expected gradient calculated. This is known as stochastic gradient descent (SGD) [**171**]–[**173**]. 'True' SGD involves randomly sampling a minibatch of size 1, with replacement, at each training step [**170**]. In practice, the dataset is randomly split into large minibatches, without replacement, at the onset of training. These are then reused every epoch [**164**]. As such, SGD in common parlance is something of a misnomer, and the process is sometimes referred to as 'minibatch gradient descent'. Choice of batch size becomes a very important training parameter, and is usually limited by a balance between data type (images are very large, for example) and GPU memory [**167**]. An additional practical caveat lies in 'learning rate annealing' – the enforced decay of the learning rate over time, which can in practice help improve accuracy while descending into a minimum of the loss surface [**171**].

A common modification (and vital to current levels of neural network performance) to learning algorithms is the inclusion of so-called 'momentum'. In this scheme, the (expected) gradient from the previous epoch is allowed to influence the current parameter update direction [**165**]:

$$\begin{aligned} \boldsymbol{\theta}_n &= \boldsymbol{\theta}_{n-1} + \boldsymbol{m}_{n-1} \\ \boldsymbol{m}_n &= \mu\, \boldsymbol{v}_{n-1} - \alpha\, \mathbf{g}_{n-1} \end{aligned} \qquad \textit{LR-38}$$

In which μ is an update parameter. This has benefits when the current parameters are migrating across ridges in the loss surface, for example. A rationalisation lies in that directions in which the parameter set *should* be moving accrue over iterations, and oscillations either side of the ridge tend to cancel out [**174**]. Of course, the same scheme is beneficial to

training when the loss surface is fairly flat, and progress can quickly be made despite low gradient magnitude. This was expanded by Nesterov [**174**] to incorporate momentum into the parameter values themselves, and an updated gradient expectation calculated from the average position of these in parameter space. This is known as 'Nesterov's Accelerated Gradient' (NAG). Duchi *et al* [**175**] have presented *AdaGrad* (adaptive gradient algorithm), which emphasises the importance of rarely seen input features. In their scheme, the optimiser is required to 'take notice' when an infrequent feature is parsed. Its update rule takes the form:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} - \alpha \left(\epsilon \mathbf{I} + diag(\boldsymbol{G}) \right)^{-1/2} \mathbf{g}_{n-1} \qquad \textit{LR-39}$$

$$\mathbf{G} = \sum_{\tau=1}^{n} \mathbf{g}_t \, \mathbf{g}_t^{\mathrm{T}} \qquad \textit{LR-40}$$

In which the matrix **G** is the sum of self-outer products of gradients from all previous time steps. The ϵ and identity are included for numerical stability. As presented here, the term preceding the gradient in **Equation LR-39** is a slight abuse of notation, but the inverse square root of a diagonal matrix can be computed elementwise. This scheme was developed upon by Kingma & Ba [**176**] with their 'adaptive moment estimation' algorithm, *ADAM*, which has seen huge popularity. This approach employs a statistics-driven rationalisation, attempting to combine the good behaviour for sparse gradients enjoyed by *AdaGrad* with an update parameter de-biasing, and suitability for loss functions that don't necessarily see any minimum:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} - \alpha \frac{\widehat{\boldsymbol{m}}_n}{\sqrt{\widehat{\boldsymbol{v}}_n} + \epsilon} \qquad \textit{LR-41}$$

Where $\widehat{\boldsymbol{m}}_n$ and $\widehat{\boldsymbol{v}}_n$ are de-biased exponentially weighted moving averages of the mean of the gradient, and the uncentered variance of the gradient (first and second moments). Again the denominator of the second term is an abuse of notation, and refers to an element-wise division (with ϵ included for numerical stability). The de-biasing is performed as follows:

$$\widehat{\boldsymbol{m}}_n = \frac{\boldsymbol{m}_n}{1 - \beta_1^n} \qquad \textit{LR-41}$$

$$\boldsymbol{m}_n = \beta_1 \, \boldsymbol{m}_{n-1} + (1 - \beta_1) \, \mathbf{g}_{n-1} \qquad \textit{LR-42}$$

$$\widehat{\boldsymbol{v}}_n = \frac{\boldsymbol{v}_n}{1 - \beta_2^n} \qquad \textit{LR-43}$$

$$\boldsymbol{v}_n = \beta_2 \, \boldsymbol{v}_{n-1} + (1 - \beta_2) \left(\mathbf{g}_{n-1} \odot \mathbf{g}_{n-1} \right) \qquad \textit{LR-44}$$

Where β_1 and β_2 are exponential weighting parameters, and $\odot$ is the pointwise (Hadamard) product. Kingma & Ba [**176**] set out that the de-biasing is necessary in order to correct for the fact that the variance and mean exponential averages start off as vectors of zeros. The advantages of this algorithm include that the step sizes are approximately bounded in magnitude by the learning rate, α, the updates are independent of the scaling of the gradient, and a form of learning rate annealing occurs automatically. Furthermore, it was shown that *AdaGrad* is in fact a special case of *ADAM* for β_1 equal to zero [**176**].

Of course, many, *many*, learning algorithms have been developed in the literature. In terms of applied ML, it seems to be the case that *ADAM*, with suitable choices of hyperparameters, leads to good results in a wide range of domains [**177**]–[**181**]. Optimisation routines of the kinds discussed in this section are by no means only applicable to neural networks or even supervised settings, and see application in problems such as linear least squares [**163**], logistic regression [**165**], [**176**], and non-negative matrix factorisation [**182**].

4.1.1 Neural networks for supervised learning

Probably the most ubiquitous (and important) machine learning approach in the 21st century is the artificial neural network. Very generally inspired by the operation of the human brain, though the analogy is perhaps somewhat contrived, they have been able to achieve unprecedented classification and regression accuracy in a wide range of domains [**183**], [**184**]. Schematically, vanilla versions are simply a sequence of matrix/tensor multiplications, additions, and kernel operations (such as convolution and averaging across certain dimensions), combined with application of pointwise 'activation functions'. There is a considerable amount of obfuscating language surrounding these relatively straightforward concepts. They have been shown in fact to be a universal approximator: with a non-linear activation function a sufficiently 'deep' network (many layers of operations) can replicate the output of any function, which could be numerical, like predicting future rainfall levels, or very abstract, such as determining whether an image is a cat or a dog [**185**]. These of course are not necessarily convex functions. Despite possessing the ability to represent the form of any function that returns real values, there is no guarantee a neural network can be trained well enough to accurately replicate it [**164**]. The surge in popularity began in 2006, when Hinton et al [**186**] presented a breakthrough in ease of network training, difficulty in which had previously plagued the community.

Neural networks usually appear in a supervised setting, where there is a large amount of training data which must be used to predict a response for test data. After a good choice of initial parameters, the network is fed examples and the response predicted. This is used to calculate a value of a loss function, quantifying the difference between predicted output and true output. Differentiation of the loss function with respect to the many thousands (probably millions) of parameters in the model, inferred through a process called backpropagation, allows each parameter to be updated to be closer to an optimum value, such that the loss function is minimised. This is further discussed in **Section LR.4.1.2**. Knowledge of the parameter gradients allows the model to be optimised through (stochastic) gradient descent , often with modifications [**183**], [**187**], or with an alternative optimiser such as *ADAM* [**176**]. Development of object-oriented and easy to implement (Python) libraries for deep learning, especially *TensorFlow* [**178**], [**179**] and *PyTorch* [**177**] have drastically improved accessibility to the field.

In the following sections a brief introduction to the basic structure and operation of neural networks is presented, followed by common modifications for optimum performance in computer vision settings (which is of particular interest for analysis of EBSPs).

4.1.2 Network structure

A simple 'feedforward' neural network (alternatively 'multilayer perceptron') is a series (set of layers) of matrix multiplications and additions (an affine transformation of the input), followed by application of a pointwise, differentiable, activation function to each output. The outputs of each layer are known as a vector of 'neurons'. For an input vector $x^{(n)}$, with superscript denoting the layer number, the vector (neurons) fed through to the subsequent $(n+1)^{\text{th}}$ layer is given by:

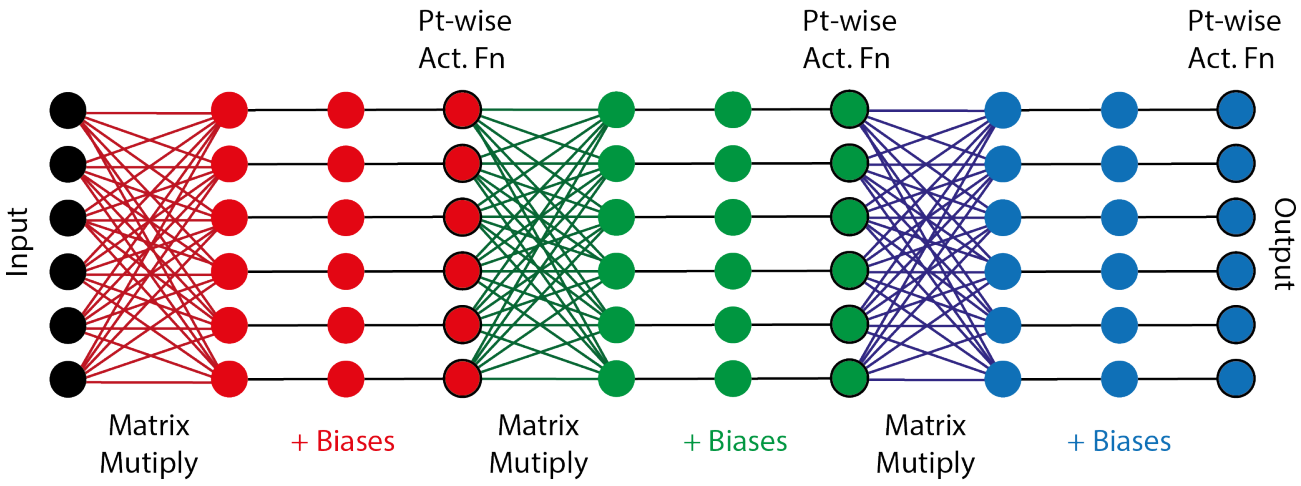

Figure LR-20: A simple neural network architecture. Circles are 'neurons', holding 'activations'. Connections represent operations performed on the network (such adding biases, participating in a matrix multiplication, etc).

$$x^{(n+1)} = f(\mathbf{W}^{(n)}\,\boldsymbol{x}^{(n)} + \mathbf{B}^{(n)}) \qquad \textit{LR-45}$$

Where the matrices $W^{(n)}$ and $B^{(n)}$ are known as the weights and biases of the n^{th} layer , and f is the activation function (applied to each element of its vector elements), whose presence introduces non-linearity. The most common activation function is the rectified linear unit (ReLU) [**188**], defined elementwise simply as:

$$f_{ReLu}(x_i) = \max(x_i, 0) \qquad \textit{LR-46}$$

If the argument is negative the output of the function (known as the 'activation') is zero. This has been shown to be empirically an extremely effective choice, and is thought to be somewhat responsible for the surge in popularity and accuracy of deep learning in the past few years [**183**]. Historically, functions such as tanh and logistic sigmoid were employed [**164**].

Dimensions of the weights and biases are chosen to engineer the size of the activations, such that the final layer has dimensions of the required number of outputs. In a classification setting this may be the number of possible classes, and the outputs interpreted as probabilities after applying an alternative activation function to the final layer, the softmax function:

$$g_{softmax}(x_i) = \frac{e^{x_i}}{\sum_j^k e^{x_j}} \qquad \textit{LR-47}$$

For an output vector x with length k. Applying this function pointwise to a vector x produces a vector of outputs whose elements all lie in the range (0,1) and sum to one. It is fairly common to apply a different activation function, g, to the final layer than to the middle 'hidden' layers, as the range of one's problem may not align with that of f, which is $(0,\infty)$ in the case of the ReLU.

In summary: the affine transformation / non-linear function process is repeated over many layers, multiplying by weights and adding biases, which need not have constant dimensionality across the layers. In the simple case of three layers, for example, the predicted response point estimate $\hat{y}$ is given:

$$\hat{y} = g(\mathbf{W}^{(3)}\, f(\mathbf{W}^{(2)}\, f(\mathbf{W}^{(1)}\,\boldsymbol{x} + \mathbf{B}^{(1)}) + \mathbf{B}^{(2)}) + \mathbf{B}^{(3)}) \qquad \textit{LR-48}$$

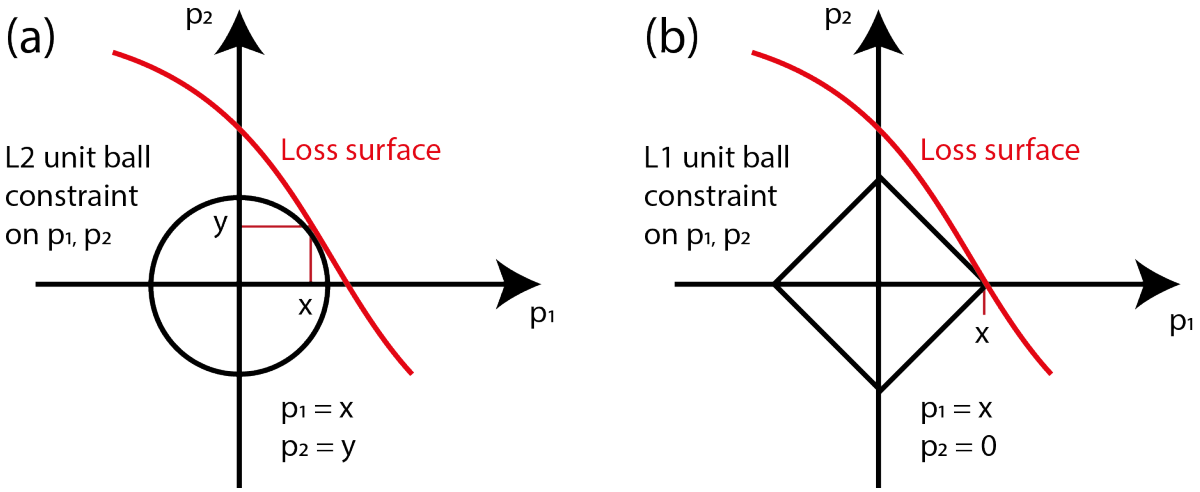

Figure LR-21: L2 and L1 regularisation. For a given set of parameters, constrained by their norm, the shape of the L1 unit ball means such regularisation tends to gives sparse solutions [**163**].

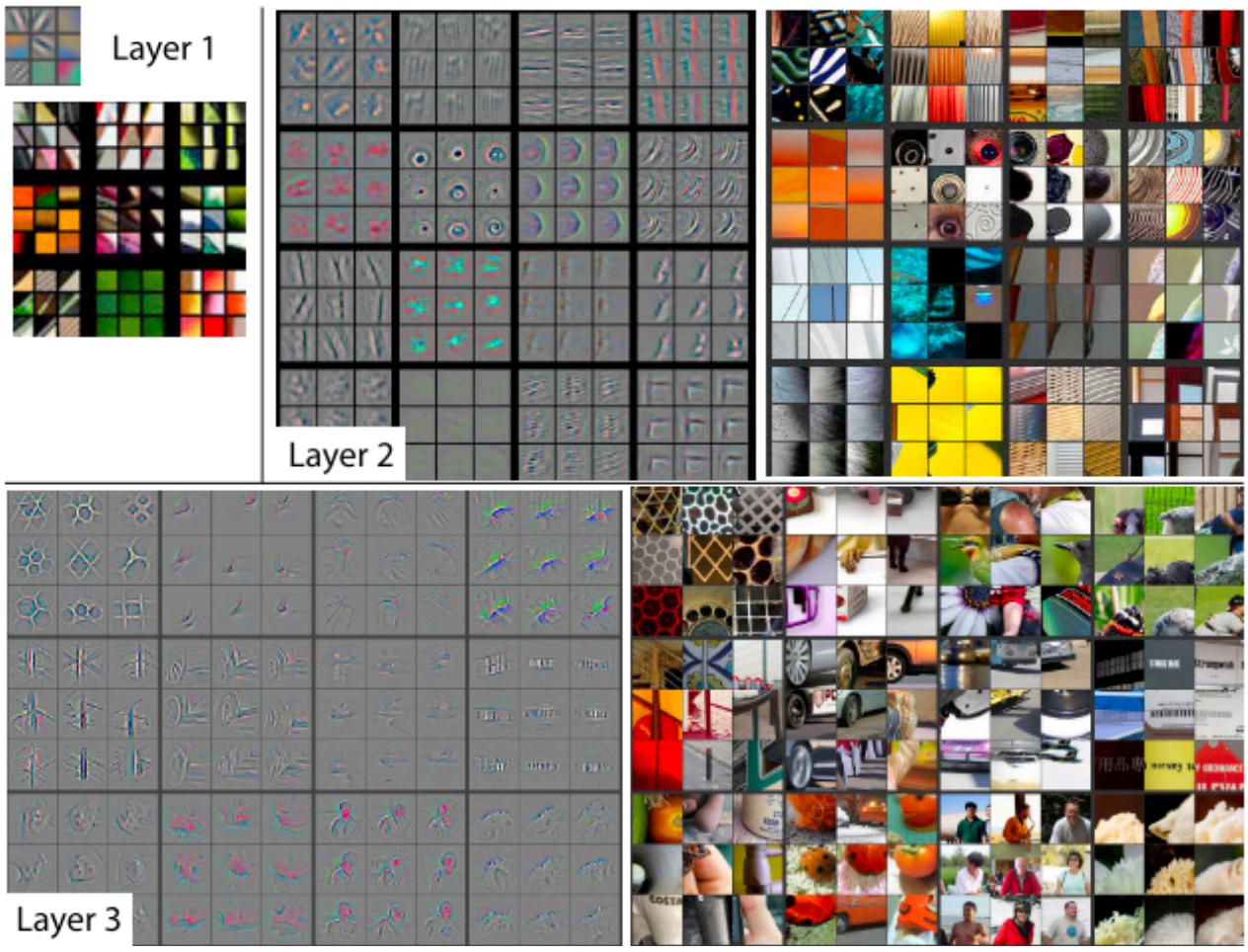

Figure LR-22: A visualisation of the features identified in sequential layers of an Imagenet trained model for individual input images. Reproduced from [**193**].

The dimensionality of predicted response $\hat{y}$ (and the true y) in a classification setting will be equal to the number of possible classes (usually a binary true/false specification for the true y) [**164**]. In a regression setting a number, vector, or tensor is predicted. The overall operation of this simple network is schematically presented in **Figure LR-20**. Neural networks are ubiquitously trained by updating their weights and biases through backpropagation. Many descriptions are

available; Goodfellow *et al* [**164**] provide a detailed discussion and review. Briefly, the chain rule is employed to calculate the (tensor) derivatives of a loss function with respect to the neural network's parameter matrices. Accordingly, each weight and bias matrix can be updated (as a function of the loss function's gradient with respect to that matrix, depending on the optimisation routine employed).

4.1.3 Improving accuracy

Common approaches to improve generalisability involve the concept of 'regularisation'. This is usually achieved by imposing additional constraints to model architecture or optimisation routine. The most basic kind, and probably the most widely used across a variety of ML approaches, is L2-regularisaiton, also known as weight decay [**163**], [**166**], [**189**]. In this scheme, the sum of the squares of the model weights (a scalar) are simply added to the loss function [**163**]. This has the effect of penalising large parameter values, which otherwise could lead to sharp gradients in a fitted function. It is closely related to L1-regularisation, in which the sum of the absolute values of the weights are added to the loss function [**165**]. In the case of linear least-squares, an L2-regularised model is known as 'ridge regression', and an L1-regularised model is called the 'lasso' [**190**]. Functionally, L1-regularisation tends to lead to a sparse set of parameters (it is said to perform 'feature selection'), a concept often rationalised by considering the proximity of the loss surface to an expanding unit ball constrained by the L1 (a diamond at the origin) or L2 (a circle) norm. This is schematically shown in **Figure LR-21**. An alternative explanation lies in that the posterior distribution of parameters, $p(\boldsymbol{\theta}|data)$, tends towards a zero-centred Gaussian in the case of L2 regularisation, and a zero-centred double exponential distribution (with a large probability mass at zero) in the L1 case [**165**]. An L2-regularised neural network loss function usually takes the form:

$$L_{reg}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) + \frac{1}{2}\lambda\left(\sum_{c=1}^{n} \mathbf{W}^{(c)\mathrm{T}}\,\mathbf{W}^{(c)}\right) \qquad \textit{LR-49}$$

Where $\mathbf{W}^{(c)}$ are the weights of layer c, for n total layers, and λ is an adjustable hyperparameter. Here $\boldsymbol{\theta}$ refers to all of the model parameters, including weights, biases, learning rate, momentum parameters, *etc*. Krogh & Hertz [**189**] emphasise the value of L2 regularisation in terms of its reduction in model complexity, and suppression of noise in the labels during training.

Dropout is a very widely employed method for regularisation, specific to neural networks, first presented by Srivastava *et al* [**187**]. This approach prevents a model from overfitting by simply turning off a random selection of neurons during each training epoch. It can be regarded as an approximation to averaging all possible sets of acceptable parameters, with each weighted by its posterior probability given the training data, $p(\boldsymbol{\theta}|data)$. This approach has proved very successful, is implemented in the common deep learning frameworks [**177**]–[**179**], and is consistently seen in state-of-the-art models for a wide range of tasks [**167**], [**183**].

Convolutional neural networks (CNNs) are particularly well suited to, and have seen fantastic success on image analysis problems. The introduction of a convolutional layer, in which a group of pixels in the input are summed within kernels, spaced with regular 'stride', introduces locality to the information contained within an array of otherwise independent pixel array values [**183**]. Despite the name, the order in which elements are multiplied and summed (kernels are scanned top left to bottom right) means that these layers actually perform cross-correlation, but the common nomenclature will be retained here. Such operations can easily be represented as matrix multiplications, so the network can retain the form of **Equation LR-48** [**191**]. Images are often represented as a 3D RGB multi-dimensional array, so the 'convolutions'

are commonly extended into three dimensions (and summing performed in a 3D kernel). These layers are usually combined with some sort of pooling operation, the combination of which reduces the dimensionality of the network, and has been shown to compress information down to latent features [**164**]. Combinations of 'convolution' and pooling were historically developed in similar models to represent the human visual system [**192**]. Implementations of this approach tend to identify very high level features in the early layers, (parallel lines, criss-crossing patterns, circles), and combine these to highlight the presence of traits (perhaps eyes, whiskers, legs) and finally combine these into an abstract classification ('cat-ness', 'dog-ness', 'strawberry-ness') [**193**]. This is represented in **Figure LR-22**. Application of this approach has reached the point of accurate 'semantic segmentation', in which images can be quantitatively divided into regions of multiple labels. Shelhamer *et al* [**194**] are able to segment horse, rider, and cars in a single image, for example. Generally, the field is moving towards the concept of 'transfer learning', in which models are extensively pre-trained on a related problem before 'fine-tuning' to suit specific objectives [**164**], [**167**], [**195**], [**196**].

4.2 Unsupervised learning

When there is no training or labelled data from which to construct a classifier or regression model, but instead we wish to learn about the *structure* of the data, unsupervised learning is of great importance. Schematically, we may wish to identify the latent features of a dataset, and how much each observation correlates with each of these features. If the relationship is linear, this can be understood as a matrix decomposition of a observation-by-variable matrix ('data matrix'). This is the basis for principal component analysis (PCA) and non-negative matrix factorisation (NMF), which are of particular concern in this thesis. PCA was formulated by Pearson [**197**] as a method of resolving the discrepancy in 'line of best fit' when regressing y on x as supposed to x on y. If the relationship is non-linear then a neural network based approach may be employed. This was first presented by Hinton & Salakhutdinov [**198**] as the 'autoencoder' neural network.

4.2.1 PCA

In order to extract the directions of maximal variance within the dataset, PCA extracts the sequence of linear combinations of feature-space variables that sequentially maximise variance, while remaining (linearly) uncorrelated with one another [**199**]. That is to say, for a data matrix $\mathbf{X}$, the first principal component (PC1) describes the linear combination of its columns (a vector of features) that has maximal variance. The second component is the linear combination of columns of ($\mathbf{X}$-PC1) that has maximal variance, while being orthogonal to PC1. PC3 must be orthogonal to PC1 and PC2, *etc*. In order to describe the full dimensionality of the dataset there must in the limit exist as many components as the rank of the data matrix. The principal components are conventionally ranked by contribution to overall dataset variance, so one usually takes the first few. A geometric interpretation is presented in **Figure LR-23**.

Calculation of principal components is equivalent to the singular value decomposition (SVD) of the (n-by-m) data matrix, $\mathbf{X}$. This matrix has n rows of variables (EBSP pixels) and m columns of observations (scan points). Principal components are stored in the columns of the orthogonal matrix $\mathbf{U}$ (the left singular vectors), and 'scores', descriptors of the extent to which a data point is represented by that vector, are given by $\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}}$ (with $\mathbf{V}^{\mathrm{T}}$ orthonormal and $\mathbf{\Sigma}$ diagonal):

$$\mathbf{X} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}} \qquad \textit{LR-50}$$

This SVD is equivalent to the eigen-decomposition of the sample covariance matrix, $\mathbf{S} = \mathbf{X}\,\mathbf{X}^{\mathrm{T}} / (n - 1)$, leading to eigenvectors in the columns of $\mathbf{U}$ and eigenvalues $\mathbf{\Sigma}^2$. As such, the principal components are an orthonormal set of

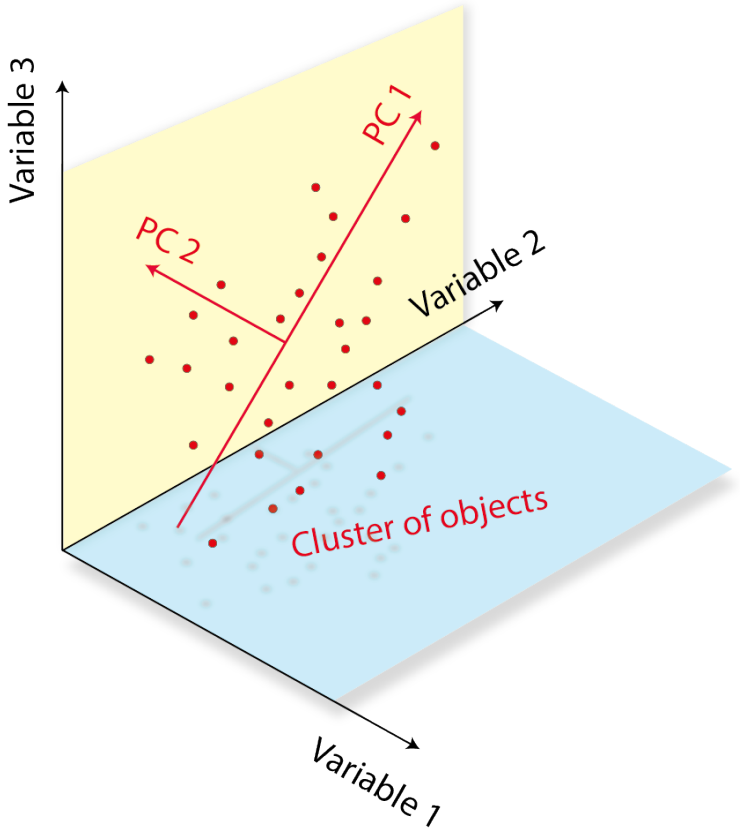

Figure LR-23: A geometrical representation of PCA in 3D, with 2 PCs highlighted.

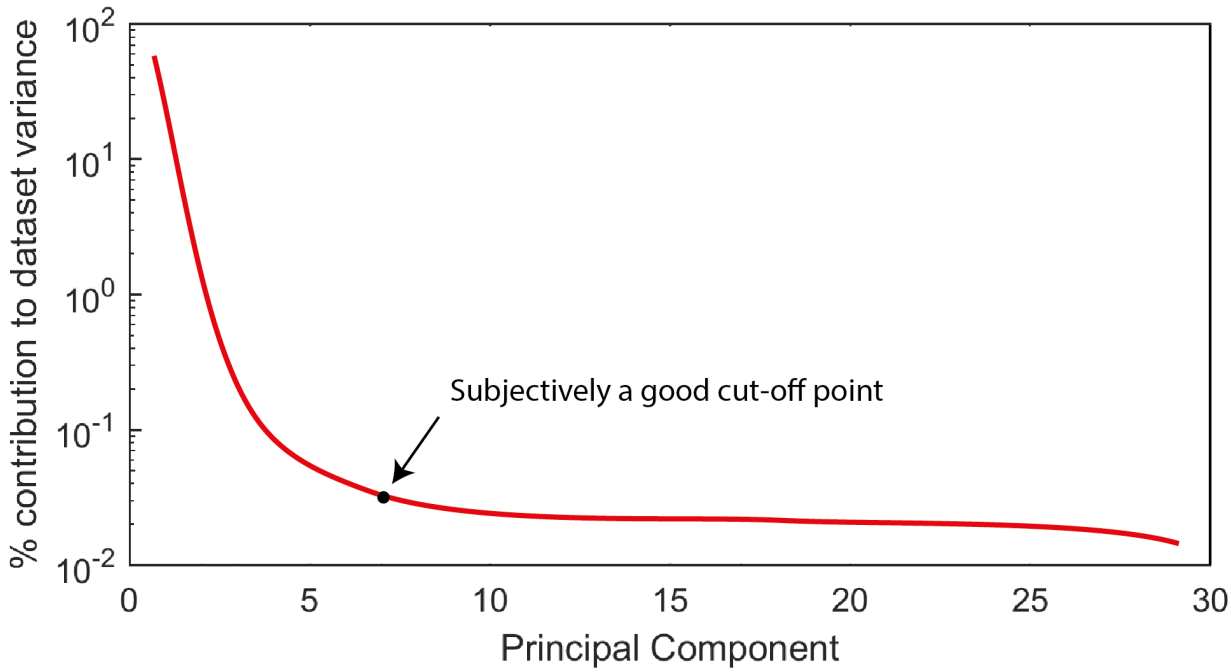

Figure LR-24: Identification of a subjective cut-off point for component retention using a Scree plot.

vectors in variable space, with each minimising square distance between itself and the data in the orthogonal directions. When ordered by eigenvalue (or singular value), the principal components correspond to the directions that contribute the most variance (elements of the diagonal matrix Σ^2) to the dataset, and the data matrix can efficiently approximated by a reduced number of principal component vectors (*k)*:

$$\mathbf{X} \approx \mathbf{X}_k = \sum_{i=1}^{k} \boldsymbol{u}_i \sigma_i \boldsymbol{v}_i^T \qquad \textit{LR-51}$$

This reduction allows us to retain the *k* most significant features, which in an EBSD decomposition, for example, will correspond to *k* representative Kikuchi patterns, for example one per grain, sub-grain, or precipitate [**131**]. This is extremely useful when the full rank of the dataset is equal to the number of scan points (and there could be hundreds of thousands of scan points). A reduction of this type is a natural fit for problems where *a priori* we know that the AOI will only exhibit a few archetypes of Kikuchi pattern and we are not interested in intra-class variation (e.g. corresponding to crystal disorientation).

The number of components to retain is often a subjective choice. The contributions to dataset variance of each of the PCs are known from the singular values, given the orthonormality of both $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$:

$$\begin{aligned} \text{var}(\boldsymbol{u}_i) = \boldsymbol{u}_i^T \mathbf{S}\ \boldsymbol{u}_i &= \frac{1}{n-1}\left(\boldsymbol{u}_i^T \mathbf{X}\right)\left(\mathbf{X}^T \boldsymbol{u}_i\right) \\ &= \frac{1}{n-1}\left(\sigma_i \boldsymbol{v}_i^T\right)\left(\sigma_i \boldsymbol{v}_i\right) \\ &= \frac{\sigma_i^2}{n-1} \end{aligned} \qquad \textit{LR-52}$$

Sometimes these variances are scaled to be proportions of the overall dataset variance (which can be found from the trace of $\mathbf{\Sigma}$). Plotting var($\boldsymbol{u}_i$) as a function of *i* yields a 'Scree' plot, which is often used by a human interpreter to determine a cut-off point, at which the important information has been extracted and further components are just contributing noise [**200**]. This procedure is presented in **Figure LR-24**. The set of remaining components may then be interpreted as-is or after some orthogonal rotation, such as VARIMAX [**131**]. Alternative approaches to component selection have been developed, including calculating the 'Scree test optimal co-ordinate', and 'Scree test acceleration factor' [**200**]. In this thesis it was found that simply providing a cutoff for the amount of variance the final component must contribute worked well, and is discussed further in **Chapter 1**.

The difference between PCA and straightforward least squares linear regression (eg. regressing *y* on *x*) can be understood by considering what is being minimised. Considering a 2D example, in the former case, the square *perpendicular* distances, in fact along the second principal component are minimised in PCA. In a conventional linear regression, the differences between data and model *only in the response variable(s)-direction* are minimised.

A recent review by Jolliffe & Cadima [**199**] sets out historical developments in application of PCA to a wide range of problems. Of particular modern note is the development of so-called 'Robust' PCA, in which a dataset is decomposed into a low-rank component (as conventional PCA does) as well as a sparse component. The sparse matrix is designed to account for secondary disturbances (for example identifying a moving person in the foreground of CCTV footage) [**201**].

Inferring principal components can be performed with the same approaches as calculating singular vectors and eigenvectors. Efficiently computing these is unsurprisingly the subject of significant research. By default, *MATLAB*, *numpy*, and *SciKit-Learn* software packages return *all* the principal components of a data matrix, in which case the problem is identical to getting the full set of singular vectors. This never involves directly computing the covariance matrix itself, which alone requires O(nm^2 (for m<n) ; (mn^2 (for m>n)) operations [**202**]. Even with knowledge of the covariance matrix, computing the solution to the eigenvector equations $|\mathbf{A} - \lambda\mathbf{I}| = 0$ is completely intractable, and is not how eigenvectors are computed in practicality.

MATLAB's implementations of *eig*() and *svd*() are proprietary, there is no transparency to the algorithms used to perform these decompositions. In the case of open-source *numpy*, it is evident that the *svd*() function is a wrapper around the corresponding LAPACK (Fortran Linear Algebra Package) routine [**203**]. This is not the case in *MATLAB*, which instead has LAPACK implementations specified as different functions, such as *svd_lapack*(). LAPACK includes several algorithms to compute the SVD of an array (and computing an eigendecomposition is the SVD of a square matrix). The most widely used, *xGESVD*, implements the approach first presented by Golub & Kahan [**204**]. In this scheme, a series of Household transformations are successively applied, reducing the matrix to a bidiagonal representation. These are followed by a QR factorisation[3], reducing the matrix to a diagonal array of singular values [**205**]. This has computational and stability advantages when using floating point arithmetic. A modern 'divide-and-conquer' approach, *xGESDD* was implemented in version 3.0 of LAPACK, and is significantly faster for large arrays [**203**], [**206**].

If fewer components are requested than the full set, alternative algorithms discussed by Nash & Shlien [**207**] may be more efficient. If a posterior distribution of component scores is desired, the 'expectation-maximisation' algorithm developed by Roweis [**202**] can be employed to quickly gather the most significant vectors. This has complexity O(*nmk*), and generalises to account for a prior distribution of input data, which can account for arbitrary noise distributions.

4.2.2 NMF

When a system must (for example physically) be the sum of positive or zero elements, NMF is helpful and can lead to easily interpretable latent features. First popularised by Lee & Seung [**208**], this procedure identifies a dataset linear decomposition with all elements strictly positive:

$$\mathbf{X} \approx \mathbf{A}\,\mathbf{S} \qquad \textit{LR-53}$$

Where *A* and *S* are (*n*-by-*k*) and (*k*-by-*m*) matrices, corresponding to components and scores respectively. Note that *X* can never fully be represented by *A* and *S* as the rank of their product is at most *k*. This is always an approximate factorisation and inherently a less accurate one than that provided by the SVD (for the same *k*) by the Eckart-Young-Mirsky theorem [**209**]. A simple rationalisation is that the boundary condition imposed by element non-negativity is such that degrees of freedom that would otherwise allow a closer solution are lost. This boundary condition can be extremely helpful as a necessary condition that the noise-filtered solution must fulfil. Unlike PCA/SVD, there is not exact analytical solution to an NMF factorisation (even though in reality as discussed in the previous section PCA is not calculated through the covariance matrix), so parameters have to be learned through an optimisation [**182**].

[3] The QR factorisation decomposes a rectangular matrix into orthogonal (**Q**) and upper triangular (**R**) matrices. This in turn can be computed for example through the Gram-Schmidt approach..

In gradient descent approaches, A and S are randomly initialised, and a loss function L involving the Frobenius norm is used to alternately minimise A and S:

$$L = \frac{1}{2} \|\mathbf{X} - \mathbf{A}\,\mathbf{S}\|_F \qquad \textit{LR-54}$$

This is the simplest and probably the most instructive description of the procedure. Other algorithms exist, and are well reviewed by Berry *et al* [**182**]. The solution is dependent on the number of components one asks the algorithm to return. Components in NMF are not orthogonal, and a unique solution is not guaranteed due to the non-convexity of L (there will be local minima) in both **A** and **S** [**182**]. There is not an obvious geometrical analogue. Additional regularisation terms can be added to the loss function in order to encourage smoothness, *etc* [**210**].

As discussed by Lee & Seung [**208**], NMF tends to learn a 'parts-based' representation, as supposed to the 'holistic' representations learned by PCA. This is evident in **Figure LR-25**. This is what leads to more intuitive human interpretability of the factorisation. Intrinsic sparsity in the factorisation (many latent scores tend to zero, as they would otherwise be negative) leads to input features easily being modelled by a subset of the latent features. This is converse to PCA, in which generally all the latents contribute to each solution unless some form of regularisation is employed. The sparsely activated latents thus can more intuitively represent a 'part' [**208**].

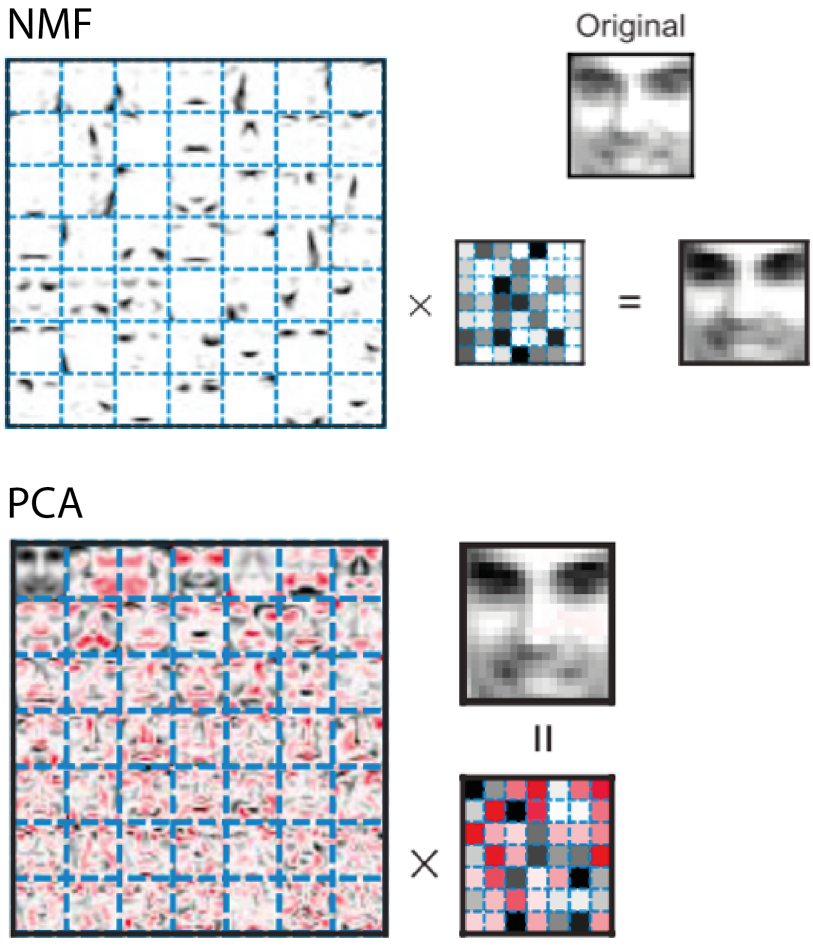

Figure LR-25: An illustrative example of the 'parts-based' basis that NMF tends to learn, reproduced from [**208**].

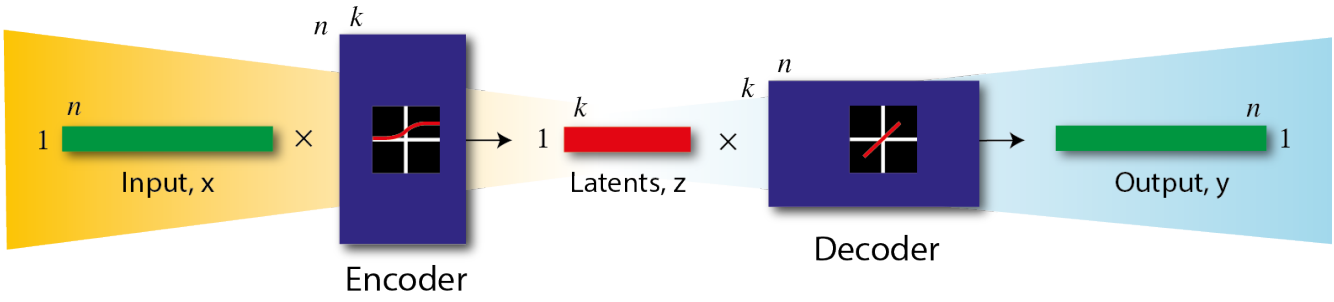

Figure LR-26: Schematic of a shallow autoencoder for learning latent features of an input vector.

4.2.3 Autoencoders

As previously discussed, neural networks have the advantage of being able to approximate the form of broadly any function, regardless of convexity and complication. Upon training, layer tends to encode a latent or 'hidden' feature [**193**]. Hinton & Salakhutdinov [**198**] present an autoencoder neural network, which takes advantage of this to intentionally learn the latent feature space of a dataset for subsequent interpretation. This is achieved by combining an 'encoder', which combines conventional fully-connected or convolutional layers to reduce the dimensionality of the input tensors, with a 'decoder', which upsamples the latent feature space to provide an output of the same dimensionality as the input. The autoencoder is trained to learn the identity function for all of the measurements (which are tensors – usually vectors or matrices), minimising a loss function such as the mean squared error. The structure is schematically shown in **Figure LR-26** for a shallow network with logistic encoder activation function.

As pointed out by Valpola [**211**] in 2015, supervised learning with neural networks had seen much more success in recent years than developments in neural unsupervised learning. This was especially in the context of the latter informing pre-training for supervised learning, in a context helpfully known as 'semi-supervised learning'. They suggest this is because (despite much more obvious success metrics being present in the supervised case) supervised procedures aim to filter out all irrelevant (latent) features for a given task, while the objective of an unsupervised approach is to accurately represent as much of the input information as possible. Higgins *et al* [**212**] suggest that leveraging autoencoders to accurately pre-train neural networks to represent the 'objectness' of features in datasets, which can subsequently be used in deeper parts of a network, is critical to developing artificial intelligence operating with human-like reasoning. This is in effect a process of learning a 'disentangled' posterior distribution of generative factors.

Using a neural network in this way for unsupervised learning permits exact engineering of the probability distributions surrounding latent activations. PCA, for example, returns orthogonal latent feature vectors, which are *linearly* independent [**213**]. This is potentially a considerable downside to one of the approach's most useful features: there are no guarantees that feature vectors are not correlated in some higher degree. In contrast, a highly customisable neural network architecture permits visibility and loss function penalties in a very controllable manner [**198**]. Hidden layer (latent) activations can be sampled, and the loss function can penalise deviation from the desired distributions. This is the basis of modern 'variational' autoencoders (VAEs) [**212**]. Including Kullback-Leibler (KL) divergence, applied to the hidden layer activations, encourages the network to learn a sparse representation, whereby there is a loss penalty if a training set produces too inhomogeneous a probability distribution of hidden neuron activations. Goodfellow *et al*

[**164**] provide a good description of the derivation of KL divergence between activation distributions with respect to information theory.

The KL-divergence is as:

$$K = \sum_{i=1}^{k} \rho_0[\ln \rho_0 - \ln \bar{\rho}_i] + (1-\rho_0)[\ln(1-\rho_0) - \ln(1-\bar{\rho}_i)] \quad \textit{LR-55}$$

Where ρ_0 is the desired sparsity proportion and $\bar{\rho}_i$ is the average likelihood of observing activating the i^{th} hidden neuron:

$$\bar{\rho}_i = p(z_i \mid x) = \frac{1}{m}\sum_{s=1}^{m} z_{i,s} \quad \textit{LR-56}$$

The full loss function for the variational autoencoder becomes:

$$L = \sum_{i=1}^{h}\sum_{j=1}^{h}(x_{i,j} - y_{ij})^2 + \lambda W + \omega K \quad \textit{LR-57}$$

Where x is the function input, y the output, W is a L2-regularisation term and K is the Kullback-Leibler divergence. These are modulated by hyperparameters λ and ω. VAEs have seen successful application in a range of standard validation datasets, such as MNIST number recognition and the 'Frey face' facial recognition challenge [**214**].

4.3 ML for predicting material properties

Occurrences of ML in materials science tend to fall into one of two camps. Firstly, directly predicting material properties is an obvious, extremely challenging, and financially lucrative challenge. Such approaches often treat composition, temperature, and occasionally processing parameters as independent variables, and properties as dependent. ML techniques may seem attractive for modelling complicated processes, as they do not require a physical interpretation for their form. Much work treats ML techniques as impenetrable black boxes, and consideration of *what* the algorithm has learned can be somewhat lacking. This is confounded by one of the major drawbacks of neural networks (which are by far the most popular choice of model) – the sheer number of parameters makes the networks extremely difficult to interpret. This is likely more attractive for industrial materials scientists and engineers, perhaps wishing to make accurate predictions about behaviour without concern for underlying mechanisms. Secondly, ML techniques can be used (usually but not always in an unsupervised setting) to improve understanding of physical phenomena and aid interpretation of experimental measurements in the context of current literature or theory. The work in this thesis primarily falls into the second camp; ML technologies can be used to better understand the relationships between crystal structure, chemistry, and microstructure. This is further discussed in the subsequent section.

Leveraging the techniques discussed in the previous sections to better understand, predict, and develop material properties is an area of ongoing research. Algorithmic materials design is complicated: there are many factors at play, with differing degrees of uncertainty in both their value and validity of measurement. Many variables are strongly covarying (grain size, oxidation resistance, and hardness for example) and processing/temperature exposure plays a critical role. There are perhaps far more trade-offs in different aspects of performance. Fatigue life, creep resistance, strength, are all opposing functions of grain size, itself a dependent variable, for example. Moreover, there are vast

regions of parameter space that are completely unexplored [**215**]. The key limitation for the field is the significant cost in obtaining a single data point for a material property. Obtaining an expected value and variance for this property, in completely independent instances of a composition / processing pipeline, is prohibitively expensive. As discussed by Kuehmann & Olson [**216**], probabilistic models that can accurately characterise variability, in the presence of limited data, must be the way forward. Accordingly, linear dependence in material properties has been explored by Toda-Caraballo *et al* [**217**] using PCA, who show that 80% of dataset variance in ten important material properties (heat capacity, density, melting point, yield strength, heat of fusion, thermal expansion, thermal conductivity, electrical resistivity, Young's modulus, yield strength, fracture toughness, ductility) was explainable by three PCs. This may not be particularly surprising, as many of these properties are clearly related and driven by the same physical mechanisms. This somewhat lends credence to the use of neural networks in property prediction, as circumventing the obvious correlations between properties and learning the important non-linear relationships between latent factors is their key strength [**164**].

As discussed by Ramprasad *et al* [**218**] in their recent review of ML approaches to materials problems, quantifying and evaluating models with respect to the domain of prior data is critical to development of the field. To this end, Joo *et al* [**219**] present a neural network trained with a genetic algorithm, in which the best performing of individual seed sets of parameters are 'selected' to source the next generation. The probability of selecting an individual is proportional to its relative suitability compared to its siblings This work was performed before modern advances in optimisation with stochastic gradient descent, ReLUs, and deep learning became mainstream. Even so, the model was able to learn a well-rationalised latent space for the influence of C, Mn, Si, *etc* on the elongation and strength of transformation-induced plasticity (TRIP) steels.

A related approach to maintaining statistical validity has been presented by Conduit *et al* [**158**]. A neural network was used to refine a nickel-base superalloy composition with optimal strength, fatigue life, density, cost and oxidation resistance. Furthermore, processing parameters were included as alloy properties – an absolutely essential consideration often overlooked. The authors' approach combined three streams of models: a network to predict each property separately was first trained, followed by a Bayesian bootstrap model (after Papadopolous *et al* [**220**]) to predict the probability that a proposed composition satisfies all design criteria, and finally a model to search composition space for the best solution (most likely to satisfy all design criteria). The advantage of such an approach is that when the model explores composition space far from any experimental measurements, it is reflected in the overall pipeline maximisation metric. The chosen alloy experimentally performs better than the baseline RR1000, and simultaneously satisfies all the design criteria, which no competitor alloy was able to do. This scheme is sophisticated and well thought through, in contrast to pipelines that aim to explicitly predict the composition that will best maximise properties without any thought to uncertainty or covariance [**160**], [**221**].

In recent years there has been a push for open availability to materials property data to facilitate development and training of the types of models discussed in this section. Of particular note are AirFlow [**222**], The Materials Project [**223**], and NOMAD [**224**], databases for (usually density functional theory) simulated physical properties such as elastic modulus and band structure. Towards a similar aim, the international crystal structure database (ICSD) hosts an extremely extensive collection of experimentally measured structure types, unit cell parameters *etc* for organic and inorganic crystals [**225**]. The Crystallography Open Database (COD) hosts similar information, with the ability to query standard 'crystallographic information files' (CIFs) using structured query language [**226**]. Repositories such as these

are absolutely essential to ongoing research in data-driven materials science, and their continued open-access is extremely welcome.

Sophisticated interpolation and model fitting is obviously not new to materials science. Calculation of phase diagrams (CALPHAD) approaches have been employing Bayesian statistical methods for assessing phase stability (which in modern parlance would certainly be referred to as ML) since the early 2000s [**227**], [**228**]. Ramprasad *et al* [**218**] point out that relationships such as Hall-Petch and the Hume-Rothery rules, absolutely fundamental to materials engineering, are effectively classes of the same problem. This analogy summarises probably the best case for future ML applications to materials design. Extracted relationships are empirical, but can function extremely well. They can then accurately

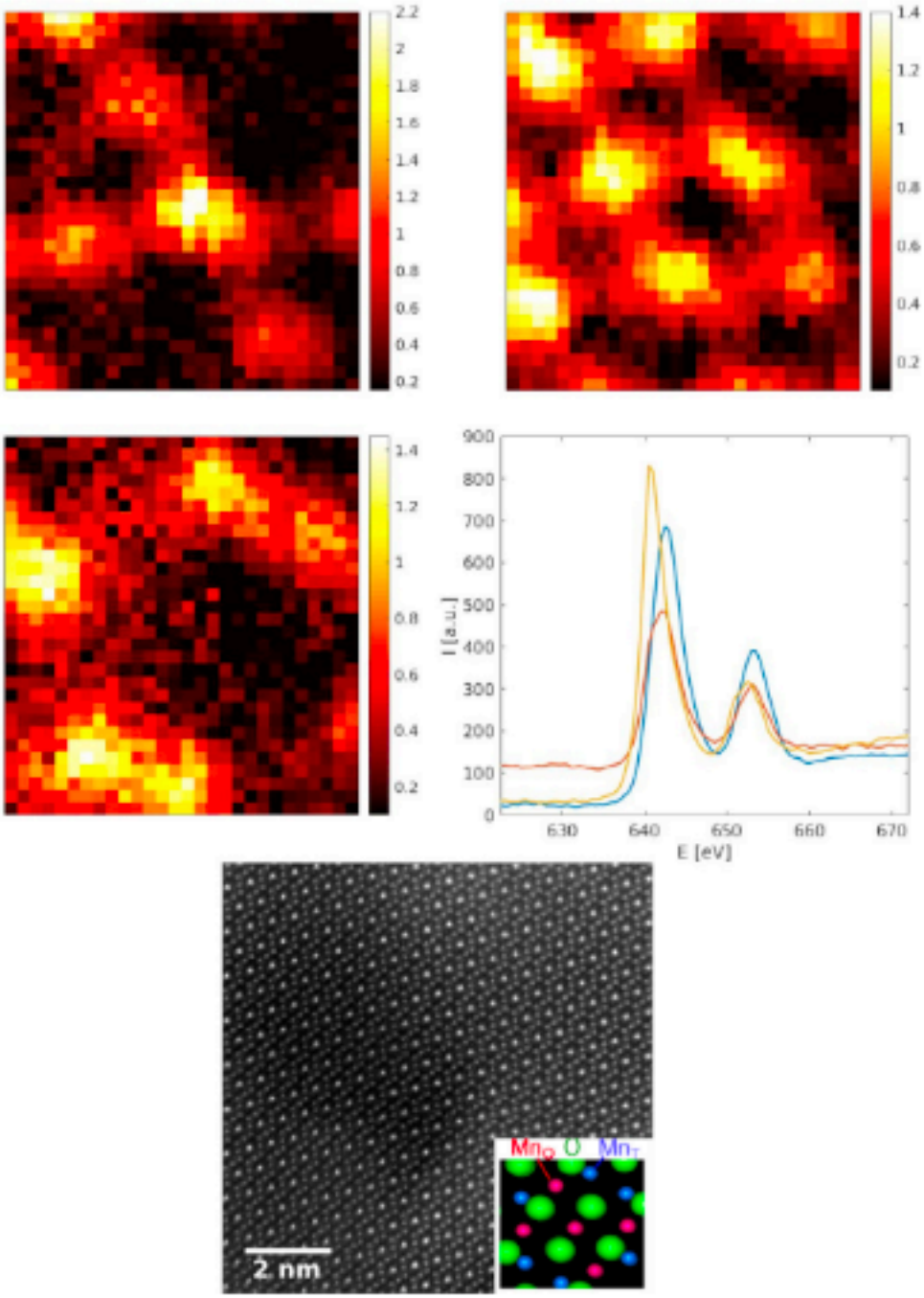

Figure LR-27: Latent factors (bottom right) and scores (top left three maps) for the NMF spectral decomposition, and atomically-resolved structure showing supposed chemistry. Reproduced from [**236**].

and helpfully inform research directions into underlying mechanisms, to subsequently generate even better models and improve material performance.

4.4 ML for improving analysis

Using ML tools to enhance understanding of physical processes by improving materials characterisation pipelines is a focus of this thesis. This is generally performed using unsupervised learning, but applying well trained supervised models to new data can also yield novel insight.

4.4.1 Hyperspectral decomposition

PCA and NMF have historically seen extensive use in analyses of chemical spectra (for example X-ray [**229**], Raman [**230**], mass [**231**]). This generally involves background removal, or separation of signal components. Watanabe *et al* [**229**] show that this can greatly improve precision of chemical quantification in STEM-EDS datasets. If the signals from independent contributors are additive (and separation is desirable), such as in X-ray spectra, NMF may be particularly suitable. There seems to be an excess of terminology in the literature referring to essentially the scheme of decomposition: 'linear mixture models', 'multivariate curve resolution', and 'hyperspectral unmixing' describe very similar concepts.

PCA has been used by Parish *et al* [**232**] to develop 'quantitative spectral imaging'. In this work the authors de-noise STEM-EDS datasets by retaining only the first six principal components. These reconstructed spectra, in the form of a lower rank approximation to the original data matrix (discussed previously in **Section LR.4.2.1**, **Equation LR-51**), are least-squares fitted with reference spectra for accurate quantification. This approach is developed to account for Poisson noise and develop confidence intervals in a multiphase PLZT thin film [**233**]. Scores are calculated for each PC, and phases segmented on their histograms. It was observed that the phase separation computed from PC1 gives generally the same segmentation as directly on the HAADF intensity histogram. Separating on higher order components appeared to give physically sensible results. Averaging the raw spectra associated with each segmentation permits straightforward treatment of Poisson noise, such as previously performed by Kotula & Keenan [**234**]. The uncertainty (standard deviation) in cation composition ratio (C_a:C_b) can be computed:

$$\frac{C_a}{C_b} = k_{ab}\frac{I_a}{I_b}\sqrt{\left(\frac{I_a}{I_a}\right)^2 + \left(\frac{I_b}{I_b}\right)^2 + \left(\frac{\Delta k_{ab}}{k_{ab}}\right)^2} \qquad \textit{LR-58}$$

Where I_x is the (averaged) intensity of the peak for cation a, and k_{ab} the Cliff-Lorimer sensitivity factor. Δk_{ab} is the uncertainty (standard deviation) in sensitivity factor, which can be obtained from a t-statistic analysis [**233**].

Li *et al* [**235**] present a novel method for NMF spectral signal separation. Their primary concern is in selecting the correct number of non-negative factors, and develop a pipeline called Robust Collaborative NMF (R-CONMF). The motivation for this derives from the fact that unlike in PCA, NMF solutions are a function of the number of components selected. In the scheme of **Equation LR-39**, they adjust the conventional NMF loss function to include additional regularisation:

$$L = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\,\mathbf{S}\|_F^2 + \alpha\|\mathbf{S}\|_{2,1} + \frac{\beta}{2}\|\mathbf{A} - \mathbf{P}\|_F^2 \qquad \textit{LR-59}$$

$$\|\mathbf{S}\|_{2,1} = \sum_{i=1}^{k} \|\mathbf{S}_{i,:}\|_2 \qquad \textit{LR-60}$$

Where $\mathbf{S}_{i,:}$ denotes the ith row of score matrix **S**, k is the number of retained components, and α, β are hyperparameters. **P** is a matrix of constraint spectra, and the third term of L 'pushes' the solution for **A**, the component spectra**,** towards the constraints. For example, **P** can be a repeated array of the average spectrum in the dataset, or specified as the limits of the convex hull of the data space. The approach is applied to a standard Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) Cuprite mineral dataset, and shows good performance and accurate signal separation.

Spiegelberg *et al* [**236**] point out that in the context of identifying additive source spectra, of primary concern should be the generation of a unique solution for the dataset, independent of initialisation parameters. Here, the NMF solution is made unique (which isn't generally the case, as discussed in **Section LR.4.2.2**) by resampling the data points such that a 'pure pixel assumption' holds. This corresponds to a sparse, orthogonal solution for the NMF expansion (see **Equation LR-39**): the score matrix **A** has elements=0 for all but one of its columns. This translates to an orthogonality condition for the non-negative factor spectra, **S** (which is usually not the case for raw, experimental data). Resampling the NMF-derived spectra is achieved by computing the PCA of the original data matrix, and picking random, positive, linear combinations of the basis vectors. The sample is only retained if the reconstructed spectrum is fully non-negative. This interesting, well though through approach yields excellent separation of spectral signals, and as presented in **Figure LR-27**, is able to spatially resolve chemical differences in atomic resolution EELS.

4.4.2 Quantitative analysis of electron diffraction patterns

In addition to the analysis of 1D chemical spectra, unsupervised learning techniques have been successfully applied to electron diffraction (conventionally 2D image) problems. Applying PCA for factorisation of EBSD datasets was first presented by Brewer *et al* [**130**]. This is achieved by edge filtering (Sobel convolution), possibly Hough transforming the filtered images, and then reshaping the resulting raw or transformed patterns into 1D vectors of intensity. This is schematically presented in **Figure LR-28**.

A VARIMAX rotation is applied to the PCs in order to 'maximise their mutual simplicity', in other words more equally distribute the variance. This has been previously discussed in **Section LR.3.2.2**, and shown schematically in **Figure LR-15**. This approach recovers characteristic patterns well, and was validated with simulated mixing of measured patterns as well as a small grid (10-by-10, 20-by-20) of aluminium EBSPs. The authors effectively compare the confidence in assignment to a given class by plotting a composite map of the weighted assignment number, and compare decompositions of raw patterns, Hough transformed patterns, and Sobel filtered then Hough transformed patterns. It was found that the Sobel filtered and Hough transformed decomposition was the most robust. This is effectively separating the dataset along the pure kinematic diffraction information: no band contrast or background contributes to variance after the Sobel filtering. It seems likely that the strong, diffuse background present in their raw EBSPs, and its variation across the field of view with pattern centre, contributed significantly to variance. Hough transforming improved band contrast in the inverse Radon transformed component EBSPs. As the Hough transform is itself a linear parameterisation of the dataset, and therefore will not directly affect variance, the improvement in separation compared to decomposing the raw patterns is likely due to the transform extracting information on top of the diffuse background. It is effectively performing a background correction. This work was performed in 2008, and the authors discuss

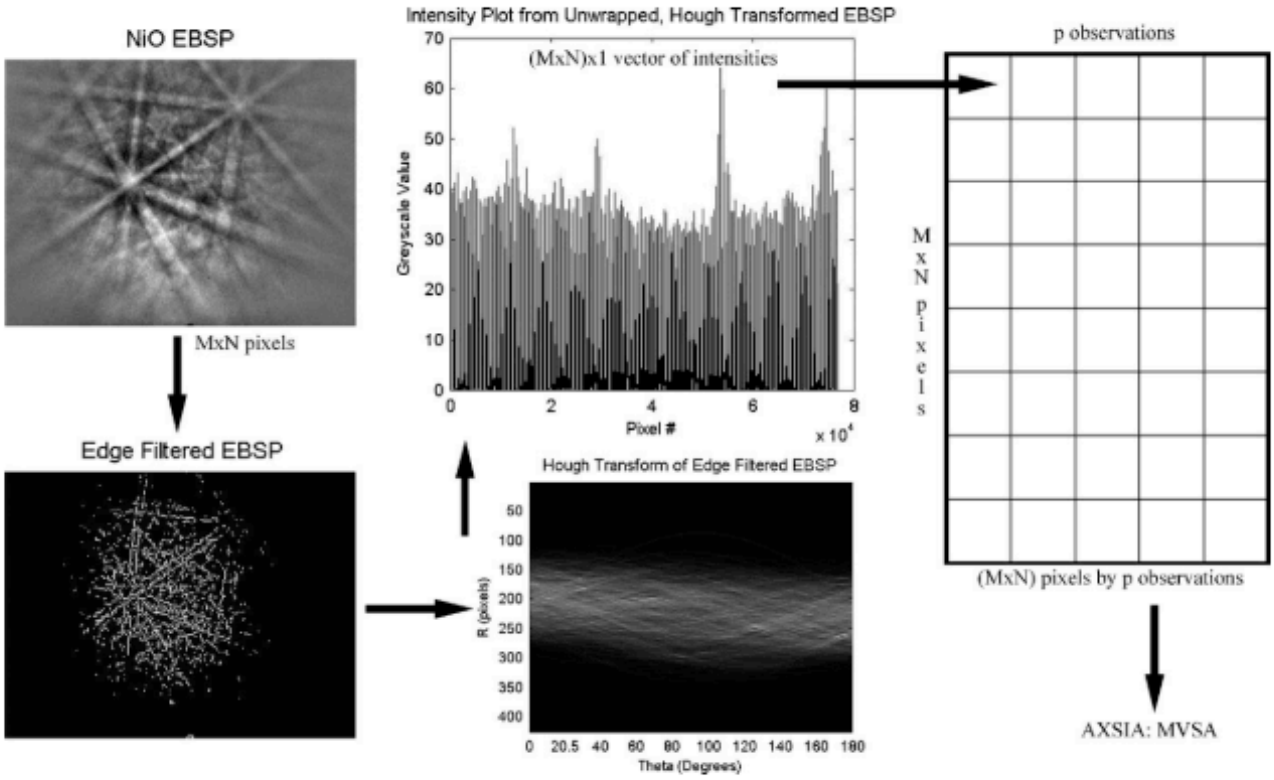

Figure LR-28: A schematic of the approach adopted in, and reproduced from, [**130**].

difficulties in taking their analysis further due to memory limitations. Despite this, the approach was shown to be extremely useful in separating dissimilar patterns for crystallographic phase identification.

Ten years later (with significantly more memory and computing power available), Wilkinson *et al* [**131**] reviewed the PCA/VARIMAX approach and presented a comparison to a *k*-means clustering ML routine. Their analysis is performed directly on the measured patterns, after dividing by a Gaussian kernel filter. This does an excellent job at removing the diffuse background that likely caused issues in the analysis of Brewer *et al* [**130**]. After applying the VARIMAX transformation, the scores and components represent a reasonable segmentation of the microstructure, as presented in **Figure LR-29**. Grains are segmented by assigning the index of the highest scoring rotated component.

The approach occasionally yields multiple components for the same grain, in which case there is sufficient misorientation between intragranular regions to generate substantively different gnomonic projections of EBSPs. The PCA approach is contrasted with *k*-means clustering, in which *k* centroids in variable space are seeded, and iteratively refined to reduce the Mahalanobis distance between cluster centres and data points. The cluster centres are used for characteristic patterns. The authors subsequently index the diffraction patterns, using a normalised dot product approach for the correlation quality. It was found that both clustering approaches yielded great improvement in the signal-to-noise of the characteristic patterns as compared to the raw data. This permits quantitative analysis of relatively weak diffraction patterns, such as those obtained in TKD. Both *k*-means and PCA were found to be useful. However, the VARIMAX rotation applied to a subset of the PCs is much faster in iterating the choice of number of characteristic signals than re-running the full *k*-means analysis. Additionally, choosing *k* in the context of PCA has a direct statistical interpretation as the contribution to total dataset variance is present in the singular values. This makes automatic selection of *k* far easier and more systematic in the context of PCA than *k*-means, as is discussed further in this thesis.

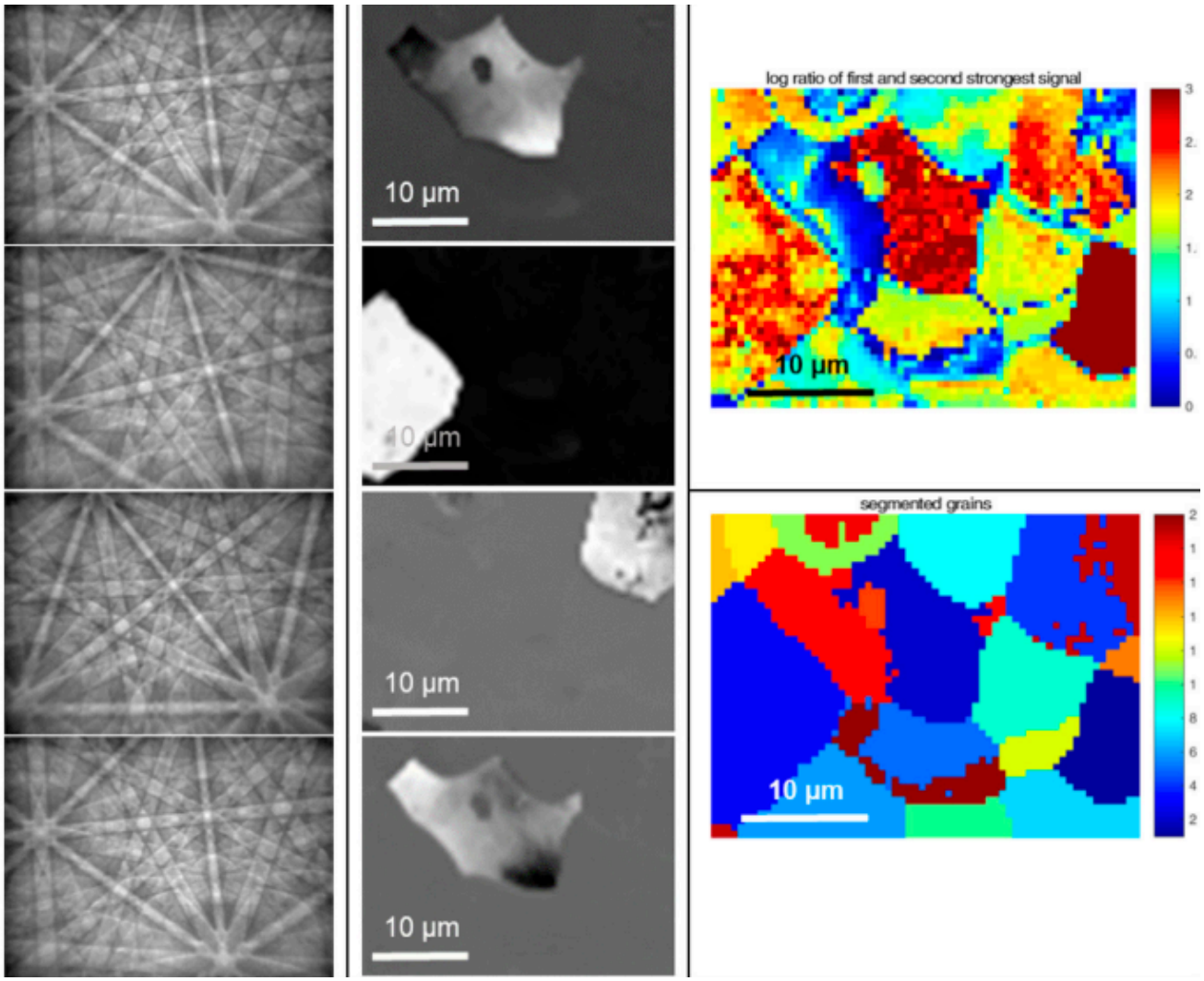

Figure LR-29: Grain segmentations, scores, and latent feature EBSPs after PCA and VARIMAX segmentation, reproduced from [**131**].

Generally, the challenge with a PCA approach is that the components returned represent the statistical dominance of each characteristic signal within the data set, and they are not intrinsically physical. A VARIMAX rotation results in an easier to interpret label, where each label can be uniquely applied to each point within the map. This works for an EBSD pattern, as the variance between two Kikuchi-based diffraction patterns for different phases is relatively small. For a TEM spot-based diffraction pattern, rotating the data according to a variance model may not be reasonable, because the spot patterns for different phases may have a stronger variation in variance (e.g. due to a different number of reflectors that create spots within the pattern). Accordingly, Eggeman *et al* [**110**] use NMF to separate superposed diffraction patterns in a precession electron diffraction (PED) tilt series for tomography of a Ni-base superalloy. This is a particularly challenging problem due to the coherence of the precipitates with the matrix. A related approach was employed by Einsle *et al* [**237**] to analyse spinodal decomposition in a meteorite microstructure. Signals were first denoised using PCA, then non-negative factors calculated. Subsequently, characteristic diffraction patterns were calculated from cluster centres from a fuzzy c-means algorithm. This is a variation on *k*-means in which members are permitted to influence and be assigned to more than one class. The subsequent cluster centre patterns saw better distinction between subtle features than the NMF factors themselves.

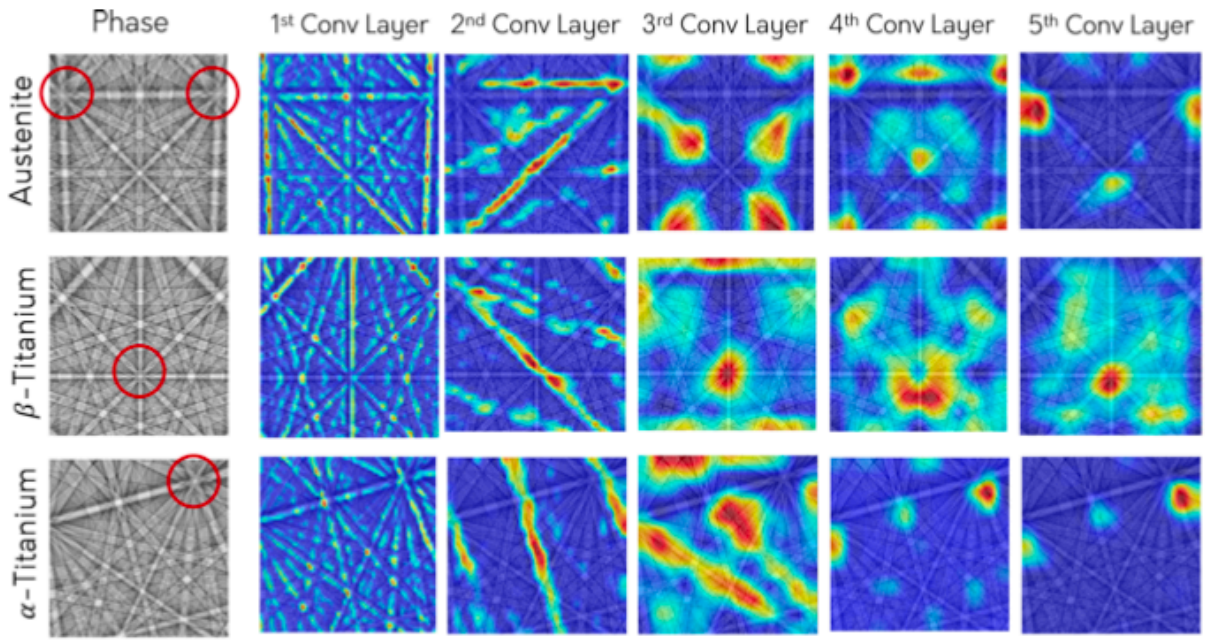

Figure LR-30: Feature visualisations for three simulated candidate structures, reproduced from [**241**].

4.4.3 Neural networks in EBSD

Alongside increased usability of neural network implementations, a trend is emerging in using trained CNNs for indexing and interpretation of diffraction patterns, especially in EBSD. The first implementation of a CNN for EBSD analysis was presented by Liu *et al* [**238**]. In this scheme, a CNN was trained to perform an orientation regression, separately predicting three Euler angles for a given input pattern. A significant contribution of their paper is to propose an intuitive loss function with the correct periodicity:

$$L = \arccos(\cos(\,|y_i - \hat{y_i}|\,)) \qquad \textit{LR-61}$$

In which y_i and $\hat{y_i}$ are one of the three Euler angles describing crystal orientation (the authors' model predicts each separately). The employed model is composed of four convolutional layers (some with max pooling) and two fully connected layers. ReLU activation functions, dropout regularisation, and momentum are used. The model gives good accuracy, with validation mean angular error reaching (2.5˚, 1.8˚, 4.8˚) for (ϕ_1, Φ, ϕ_2) after 300 epochs of training.

As has previously been discussed in **Section LR.4.1.1**, transfer learning is seeing ever-increasing popularity in the data science community, especially given the scale and depth of modern models. Shen *et al* [**239**] realise this concept, training an orientation regression model first on simulated EBSPs, then transfer learning to an experimental paradigm. The first stage of their approach uses a conventional CNN architecture to predict four quaternion components simultaneously, and accordingly is able to use a simple mean absolute error loss function. The quaternion vector should of course have unit norm, and this is not specifically expressed in the model (though the authors acknowledge it could be included as an additional boundary condition). Instead, it is left up to the model to train this condition, and the quaternion norm is used as a kind of sensibility metric. After pre-training on simulated data, the model is transferred to an experimental context. The first approach uses a 'domain transform', in which experimental patterns are transformed into 'simulated' patterns using a U-net structure [**240**]. This includes skip connections between the innermost and outermost layers of the network. It is an architecture known to mitigate the so-called 'vanishing gradient' problem, in which loss function

gradient components, especially for early layers, can become very small due to the depth of the network. The outputs of the U-net are then fed into the orientation regression model. The second approach, transfer learning, partially retrains the original orientation regression CNN with new, experimental EBSPs. It was found that the transfer learning approach performed slightly better. The trained model is very robust to experimental noise, and very noisy (short exposure) patterns can quickly (< 1 ms per EBSP) be characterised. This has the potential to vastly speed up on-line orientation analysis and indexing.

Focussing on a structure classification task, Foden *et al* [**241**] have adapted the *AlexNet* CNN [**183**] using transfer learning for classification of simulated EBSPs to the correct phase (including FCC, BCC, diamond, and HCP structures). Visualising layer activations (similarly to **Figure LR-22**) shows that the network has learned to generally identify Kikuchi bands, then 'major' bands, and finally zone axes. This is sequentially the same process as human indexing of an EBSP, and is presented in **Figure LR-30**.

Kaufman *et al* [**242**] present an ambitious CNN model trained to absolutely the classify space group of input EBSPs. Using *Resnet50* [**243**] and *XCeption* [**244**] architectures yielded similar results, with accuracies of 93.5% and 91.2% respectively. Visualising the important input features, similar to the model of Foden *et al* [**241**], showed that the network learned to assign high importance to zone axes. Interestingly, the network had difficulty in assigning structure when only one high symmetry zone axis was present in the EBSP. This makes sense, as traditional indexing of a diffraction pattern typically requires three identified zone axes to find a unique solution. Even considering the dynamical band features that the network has likely learned to classify with, patterns with just one zone axis are a reasonable stumbling block. Arguably, if the model had learned to accurately identify such patterns it could be a case of overfitting. Applying this network to a wide range of materials, such as quartz, a meteorite, $Ni_{80.0}B_{13.6}Si_{5.4}Fe_{0.2}$ blended with 40wt% WC, 430 stainless steel, and an Fe-Al intermetallic laminate all showed excellent results [**245**]. Using problem specific fine tuning, the CNN is able to give good performance on even the most challenging of tasks, such as differentiating ferrite from martensite in the 430 stainless steel sample.

The ultimate advantage of CNNs in EBSD seems to be the speed at which a well trained model can classify structure or regress orientation. Given modern advances in data transfer and storage, this is likely to only be of considerable advantage in industry, where a neural network could function as an enhanced 'black box' over the Hough transform method. Despite rationalisations in layer-wise activation, neural networks do not seem to be particularly useful for improving understanding of physical diffraction, metallurgical, or geological processes. Perhaps a sensible implementation would be for a neural network to index and locate bands, including phase ID, but then a more traditional method (such as triplet indexing discussed in **Section LR.3.2.1**) used for orientation determination. In terms of accurate, interpretable phase ID and orientation analysis, the template matching approaches discussed in **Section 3.2.3** remain the most suitable.

Chapter 1 - Advancing characterisation with statistics from correlative electron diffraction and X-ray spectroscopy

1. Chapter Summary

The routine and unique determination of minor phases in microstructures is critical to materials science. In metallurgy alone, applications include alloy and process development and the understanding of degradation in service. Here a correlative method is developed, exploring superalloy microstructures which are examined in the scanning electron microscope (SEM) using simultaneous energy dispersive X-ray spectroscopy (EDS) and electron backscatter diffraction (EBSD). This is performed at an appropriate length scale for characterisation of carbide phases' shape, size, location, and distribution. EDS and EBSD data are generated using two different physical processes, but each provide a signature of the material interacting with the incoming electron beam. Recent advances in post-processing, driven by 'big data' approaches, include use of principal component analysis (PCA). Components are subsequently characterised to assign labels to a mapped region. To provide physically meaningful signals, the principal components may be rotated to control the distribution of variance. In this work, this method is developed further through a weighted PCA approach. The EDS and EBSD signals are used concurrently, thereby labelling each region using both EDS (chemistry) and EBSD (mainly crystal structure) information. This provides a new method of amplifying signal-to-noise for very small phases in mapped regions, especially where the EDS or EBSD signal is not uniquely sufficient for classification.

2. Introduction

Confidently assessing microstructure is of significant concern in materials science and engineering, as well as in the earth and planetary sciences. In the present work, a new approach is developed using an example in Co/Ni-base superalloys. In these alloys, there are carbide precipitates that are known to strongly influence fatigue and tertiary creep performance, as discussed in **Section LR.2.2** [**32**], [**65**], [**68**], [**70**], [**246**]. The precipitates are thought to increase boundary cohesivity and to mitigate sliding. However, their high temperature oxidation reduces grain boundary strength and permits easier intergranular crack propagation. Some precipitate phases are thought to exhibit better oxidation properties than others, conferring superior enviro-mechanical stability across deformation regimes [**64**]. To assist in understanding these phases, we can use EBSD and EDS analysis for characterisation. With EDS alone, it can be difficult to distinguish two phases of similar chemistry but different structure, for example $M_{23}C_6$ and M_6C carbides. Similarly, using EBSD alone it can be difficult to distinguish two phases of similar structure but different chemistry, for example the pseudo-FCC matrix and MC carbide. Applying correlative EBSD and EDS offers a solution to this problem.

Very briefly, the 2D EBSD pattern captured using conventional EBSD is created from near surface (< 20 nm) scattering and diffraction events [**247**]. The raw signal within the EBSD pattern is semi-quantitative, due to the many transfer processes and image processing stages required to generate useful patterns for analysis. These patterns can be indexed to reveal the orientation and structure of the crystal, provided the signal-to-noise ratio is high enough and a series of input phases are used as input 'classifiers' which the indexing algorithm is tested against. Further details are discussed in **Section LR.3.1**. EBSD analysis is challenging if two phases have similar crystal structures (e.g. only slight changes in lattice parameter or subtle differences between symmetrically-related structures, particularly if an orientation relationship is present) or the signal strength is poor (e.g. a small phase). Signal classification can be improved through the use of template matching [**126**], [**131**], [**137**], [**143**], [**144**] against simulated patterns using dynamical diffraction theory [**132**], [**134**], [**247**], [**248**]. This is discussed in **Section LR.3.1.2**.

To amplify signal-to-noise for poor quality patterns, principal component analysis (PCA) can be used. PCA is a data processing approach used to reduce and project measured variables of a group of objects onto an orthogonal set of basis vectors: the principal components. The PCA method is described in detail in **Section LR.4.2**. This approach is useful, as it can amplify signal-to-noise in data, especially where the data at each point is oversampled and noisy (e.g. a diffraction pattern or a highly resolved energy dispersive X-ray spectrum) and these measurements can be used as the PCA 'variables'. The dimensions of these variables are the diffraction vectors (the intensities of each pixel in a vectorised EBSD pattern) and the counts for each EDS spectrum energy bin. When a scan is performed on a sample in a scanning electron microscope one 'object' (a full set of measurements) is collected per measurement 'point'.

For the present work, we can apply PCA, but we note that the principal components of diffraction data may be difficult to interpret. This is because PCA extracts and ranks principal component vectors by the strength of each component signal, and many signals will contribute to each point in our map. From the physics of our problem, we know (broadly) that the variance of the signal between one phase and another should be similar, and that we would like (typically) only one signal type to label each point in the map. This motivates us to develop the work of Wilkinson *et al* [**131**] and Brewer *et al* [**130**], who have shown that the rotation of the principal component EBSD patterns which maximises the variance between each member of the set reduces a full EBSD dataset down to representative, or 'characteristic' EBSD patterns. For a well selected choice in number of components to retain, these can correspond to a single pattern per grain. For oversampled or deformed grains the domains may also correlate with sub-grains. In this work, we refer to the VARIMAX rotated components as 'rotated characteristic components' (RCCs). Each RCC contains a characteristic electron backscatter diffraction pattern (RC-EBSP).

In this work, the Wilkinson *et al* [**131**] method is used as the starting point, and now the challenge of including EDS spectra is addressed. Each EDS spectrum contains chemical information related to the interaction volume associated with the generation and escape of X-rays, which are counted by a detector. The number of X-rays generated for each energy is a function of the electron transitions. Characteristic X-rays are generated from the primary beam promoting a core electron, and that core electron subsequently 'falling down' to a lower energy level to generate the radiation. These peaks are superposed on the Bremsstrahlung. The spectra contain digitised signals of the number of counts per energy level, as detected (in this case) using a silicon drift detector (SDD). The signal also contains a broadening function related to the detector and instrument noise [**249**].

In the first instance the EDS spectra can be appended onto the end of the diffraction pattern vector. However, in practice the variance in the EDS and EBSD signals may be significantly different, and the number of channels in each can vary significantly. These properties are important for statistical analysis. Finally, the interaction volume of the electron beam (and the scattering to generate the X-ray signal) may be substantively different to the volume that generates the bands within the diffraction signal. For the same map point the information within each data type may represent different volumes of matter. To address these challenges, in the present work a weighted PCA approach is explored, prior to VARIMAX rotation, using appropriately prepared and standard deviation normalised EDS and EBSD data.

In this new approach, there are three critical aspects to select in the weighted PCA and VARIMAX rotation method: (1) background correction and data normalisation prior to statistical treatment; (2) the number of RCCs to retain and rotate, corresponding to under- or oversampling of the data, and the variation in the signal for each phase and orientation; (3) the weighting of the signals, dependent on the variance of the EBSD and EDS information, as well as the number of channels in each data set.

In addition to amplification of signal-to-noise, this statistical approach has computational advantage. Instead of characterising each signal independently, the weighted PCA with VARIMAX rotation can be used to select a reduced set of characteristic signals to quantify. There are a few options; each characteristic label can be directly quantified, but this has to be performed with care as the normalisation and statistical reduction may induce uncertainty (especially where discrete peaks are being quantified in data associated with EDS spectra). The labelled regions can also be used to re-generate amplified means, using averaging, to group together physical data that represents similar phases. Here these two approaches are explored.

Upon classification of scan points to an RC-EBSP and a RC-spectrum, computationally expensive analyses can be applied to a reduced dataset. A ~40,000 point map can be reduced to a few hundred RC-EBSPs and RC-spectra each with a superior signal-to-noise ratio than a single EBSD pattern or EDS spectrum. In this chapter the EBSD patterns are analysed using the refined template matching approach developed by Foden *et al* [**137**] with a selection of possible candidate structures. Similar methods have been developed by Ram *et al* [**143**], [**144**]. Adopting a template matching approach allows utilisation of the fine detail in RC-EBSPs (weakly reflecting bands, band widths, *etc*) that PCA is able to extract. A Hough/Radon-based method would not see significant benefit from this approach, as it is based on comparing angles between the most prominent Kikuchi bands to an interplanar angle lookup table for candidate phases. More structures may be template matched to the reduced set of RC-EBSPs than would be viable for matching to the full experimental dataset, permitting greater confidence in the phase assignment. Characteristic spectra can then be used to quantitatively probe the chemistry of the classifications, and statistically robust comparisons between structure and chemistry can be made. The average and RCC EDS spectra are analysed with commercial EDS analysis software.

3. Materials & Methods

3.1 Experimental

The alloy characterised in this work, V208C, is part of a development series of Co/Ni-base superalloys, engineered for high temperature gas turbine applications. Its composition is 36 Co – 34 Ni – 15Cr – 10.5 Al – 3 W – 1 Ta – 0.150 C – 0.200 B – 0.040 Zr (at.%). This alloy was arc melted, vacuum homogenised at 1250˚C for 48 h, then hot rolled at nominally 1250℃, with 10-15% reductions from 23 mm square cross-section down to 14 mm. Finally, a sample was solution heat treated at 1100˚C for 1 h, followed by ageing at 800˚C for 4 h. This is the same as HT1 in **Chapter 2**. V208C is intraganularly dual-phase, with approximately 55% $L1_2$ γ′ volume fraction in a face-centred cubic (FCC) γ matrix. These two phases have very similar chemistry at the SEM length scale. The EBSD patterns are also extremely similar and can both be indexed with the FCC phase. Refractory element precipitates, *a priori* believed to be carbides, decorate the grain boundaries. A variety of alloying elements are used: Al, Ta and W for γ′ stabilisation; Cr for oxidation resistance; Mo for solid solution strength; small additions of C, B, and Zr for grain boundary precipitation. Further details of alloy development have previously been provided elsewhere [**1**], [**12**], [**13**].

Samples were prepared with standard metallographic grinding procedures, finishing with a neutralised colloidal silica polish for 1 h. A *Zeiss Gemini Sigma300* FEGSEM equipped with *Bruker e⁻Flash^{HD}* EBSD detector and *XFlash 6160* EDS detector was used for this work. A dataset was captured with 20 kV accelerating voltage at ~ 10 nA and 21.5 mm working distance, with the sample tilted to 70° with respect to the sample being perpendicular to the incident beam. A step size of 100 nm was employed with a pixel time of 8.3 ms. 200-by-150 px EBSD patterns were collected at 16 bit

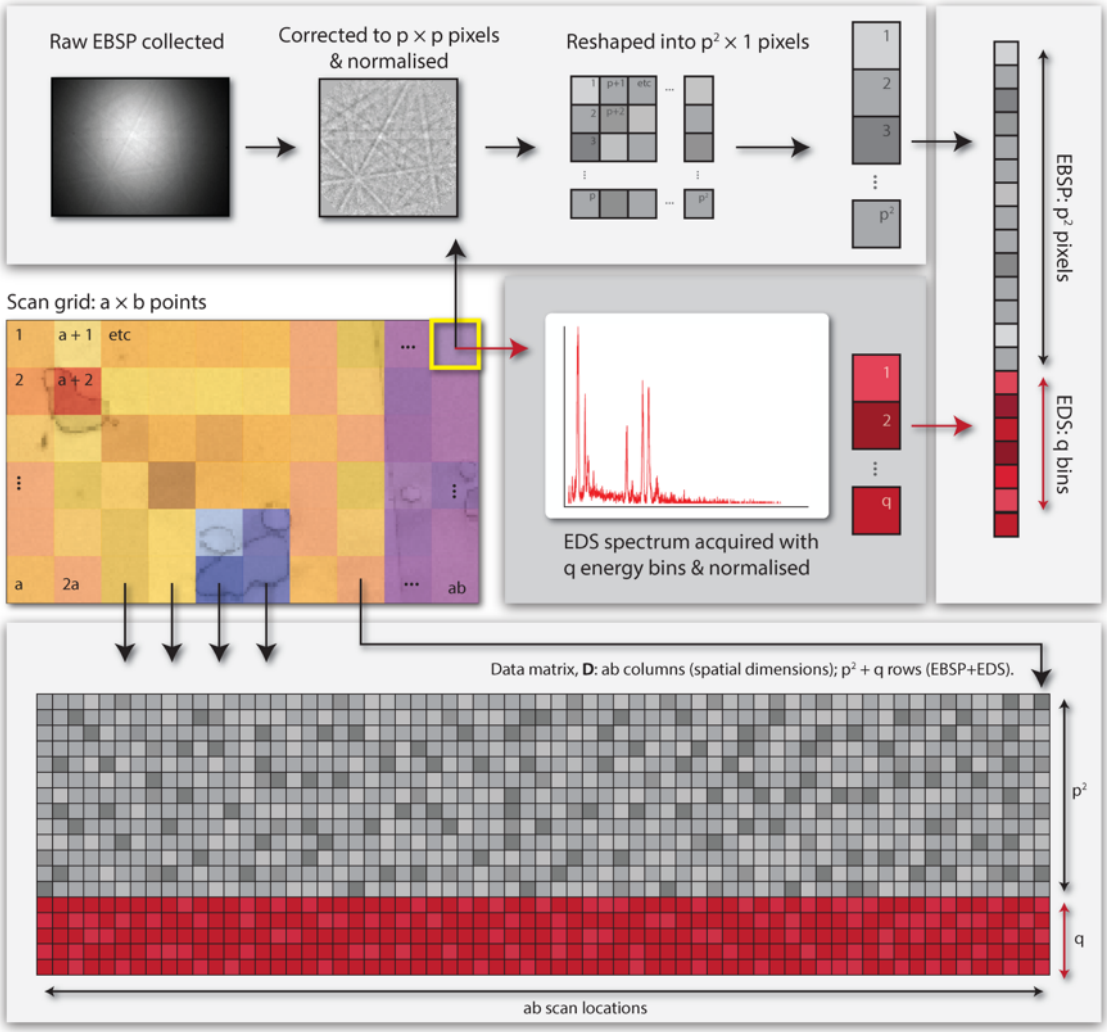

Figure 1-1: The work flow for construction of the data matrix, **D,** used for the weighted PCA method.

depth, and EDS spectra were captured with 2048 energy channels at 100 eV resolution. The captured data was extracted and stored in a HDF5 file for processing and analysis.

EBSD patterns for each point were processed in *MATLAB*, with background correction, radial cropping, and hot-pixel/split-chip fixes performed using the *AstroEBSD* package developed and presented by Britton *et al* [**119**]. EBSD patterns were originally captured with an aspect ratio of 4:3, but these were cropped to squares to simplify the refined template matching indexing (as discussed by Foden *et al* [**137**]) prior to creation of the data matrix. EDS spectra were processed in *MATLAB*. The only pre-processing performed on the spectra was background subtraction and standard deviation normalisation, which we discuss further.

3.2 Data treatment and PCA operation

At each map point the corrected EBSD patterns were vectorised and EDS spectra appended, presented in **Figure 1-1**. EBSD patterns were background corrected using the *AstroEBSD MATLAB* package, in which each pattern is divided by a 2D fitted gaussian. Patterns are then centered (mean set to zero and standard deviation set to one). EDS spectra were background corrected using *eSprit* 2.1 to remove the Bremsstrahlung, then divided by their standard deviation similarly to the EBSD patterns. In the Data Matrix, **D**, each column of data then contains the background corrected and variance normalised EBSD and EDS signals for each measurement point. Each row is the signal for a particular pixel in the binned EBSD pattern or a particular energy in the EDS signal. This data matrix follows the formulation of Wilkinson *et*

al [**131**]. The (square cropped) EBSD vector consists of p^2 pixels. The EDS signal consists of q bins. **D** therefore contains $p^2 + q$ rows. The data matrix contains measurements from a-by-b points now populating each row, and accordingly the matrix has ab columns. Each column is a PCA 'object': a full set of EBSD and EDS measurements ('variables').

The action of PCA and the geometric interpretation of variance (or standard deviation) weighting are presented in **Figure 1-2**. A specified number (n) of orthogonal principal components are calculated *via* singular value decomposition (SVD). These high-dimensionality vectors (in measured variable space) represent the directions in the dataset that explain the most variance. 'Scores' of the components are calculated for each object, corresponding to projections of the principal

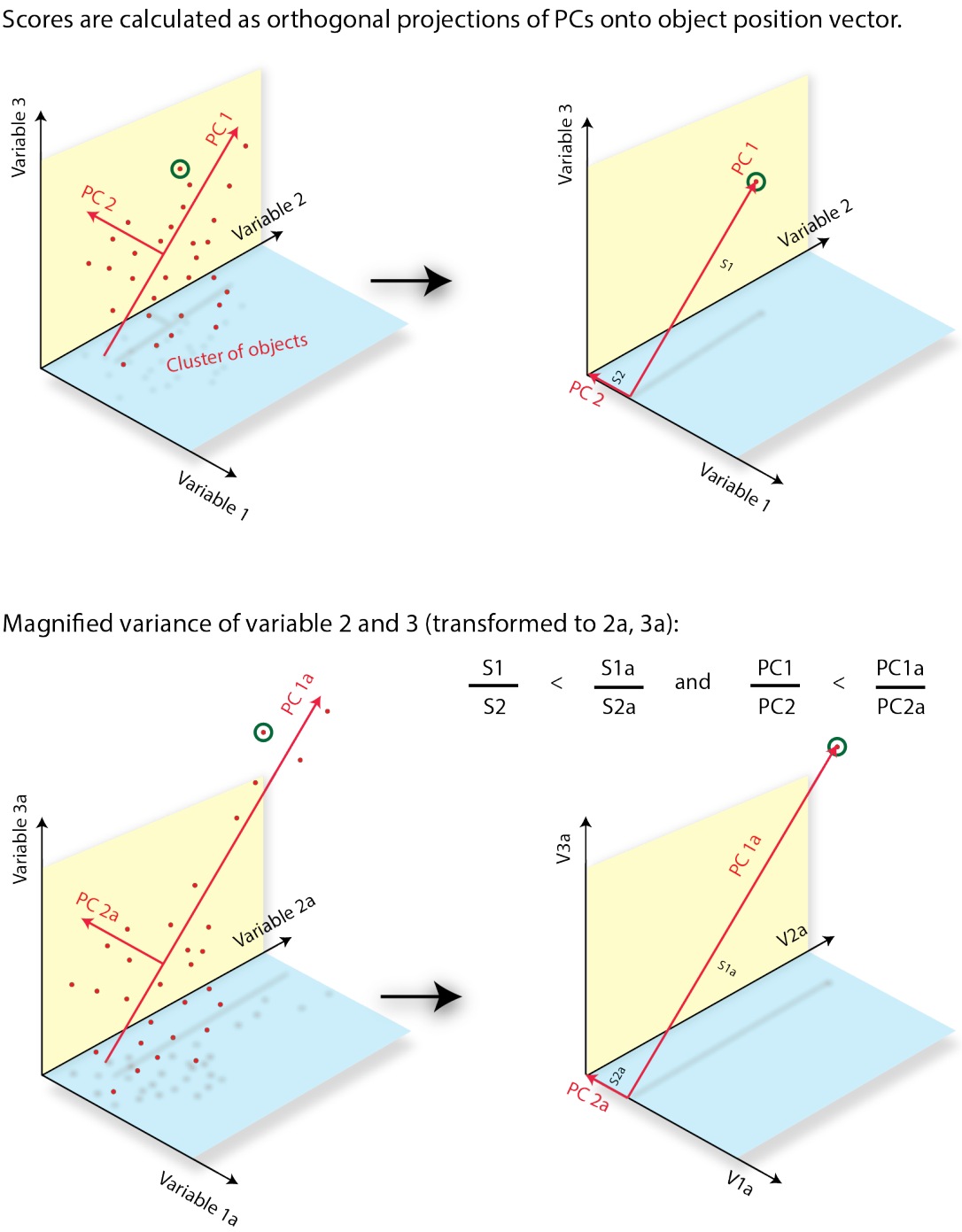

Figure 1-2: Action of PCA for a schematic dataset with many objects and three variables. (a) shows how the PCA reduces the data set to show strong variation along one principal axis, which may not be an axis of the initial data set; (b) shows how varying the weighting of two combined data sets, which present as information along different axes, can change the variance and therefore the separation of the data sets. Note that this is a simplified schematic for the purposes of visualisation, as the present datasets contain tens of thousands of variables.

components onto the position vector of an object in (high-dimensional) variable space. They represent how strongly a set of measurements is represented by a principal component. Each object (scan point) has a score for each principal component. In the case of **Figure 1-2b**, we scale up variables 2 and 3 to new variables 2a and 3a. This increases the variance in these directions. The dataset is extended in the direction of PC1a, and the object is proportionally loaded more by PC1a than it is by the unscaled PC1.

As previously described, each EBSD pattern and EDS spectrum is normalised with respect to its standard deviation before concatenation and insertion into the data matrix as a column. This is required in order to retain meaningful principal components, and is known to be an important aspect of data pre-treatment due to the variance seeking nature of the principal component fitting process [**213**]. Variables with high variance across their observations will dominate the principal components, with large corresponding scores for many objects. This is depicted in **Figure 1-2b**. To control the weighting of the principal components with respect to EBSD or EDS, it is essential to normalise the variance of the different variable types upon the data matrix's construction. Without this normalisation of variance for the EBSD pattern and EDS spectrum separately, the weighting parameter would act statistically non-uniformly on each column of $\mathbf{D} \in \Re^{p^2+q \times ab}$, leading to a confused scoring output[4].

The matrix of principal components, $\mathbf{C} \in \Re^{p^2+q \times n}$, and that of their scores at each point, $\mathbf{S} \in \Re^{ab \times n}$ are given by:

$$\mathrm{D} = \mathrm{C}\,\mathrm{S}^{\mathrm{T}} \qquad 1\text{-}1$$

The parameter n is the number of principal components to be fitted, and takes a maximal value of p^2+q. **C** and $\mathbf{S}^{\mathrm{T}}$ are calculated from the singular value decomposition of **D**, equivalent to the eigendecomposition of the covariance matrix $\mathbf{D}\mathbf{D}^{\mathrm{T}}/(r\text{-}1)$, where r is the data matrix's rank. The variance is contained (singular values of the SVD) in **S**, leaving the principal components themselves with unit length. After Wilkinson *et al* [**131**] a VARIMAX rotation, $\mathbf{R} \in \Re^{n \times n}$, is then employed, such that:

$$\mathrm{D} = \mathrm{C}\,\mathrm{R}\,\mathrm{R}^{\mathrm{T}}\,\mathrm{S}^{\mathrm{T}} \qquad 1\text{-}2$$

The matrices **C**, **S** and **R** are numerically calculable in *MATLAB* using the *Statistics & Machine Learning* toolbox. The VARIMAX rotated characteristic components (RCCs) are held in the rows of the matrix **CR**, with the corresponding scores for every point given by the rows of **SR**. RC-EBSPs and RC-spectra may then be re-constructed from the first p^2 and final q rows of **CR** respectively. An a-by-b assignment map can be constructed with the same spatial dimensions as the original scan grid. Each point is assigned a number, m, corresponding to the RCC with the highest score. For each scan point's corresponding row in **SR**, m is the number of the column that takes the greatest value. Each point is thus classified to one of n labels to construct an assignment map. Each label is associated with a characteristic EBSP (RC-EBSP) and spectrum (RC-spectrum).

As this algorithm has been constructed, the weighting term introduced acts to reduce the variance of the EBSD variables, magnifying the relative standard deviation observed for each EDS variable (across all points in the map, columns of the

[4] Note that we do not normalise each row of the data matrix with respect to the variance of the row, as we do not want to treat each channel and pixel number as totally independent measurements. Normalising the variance of each individual measurement across all scan points would reduce or completely eliminate the prominence of features such as EDS peaks and Kikuchi bands in the principal components.

data matrix) by a factor *w*. This has the effect of increasing the influence of the EDS variables on the principal components through the action of **Figure 1-2b**. In practice, this means that there is an observable transition from EBSD-dominated through to EDS-dominated behaviour as *w* is decreased, this is presented and discussed in **Sections 1.3.2** and **1.3.3**.

Upon construction of the data matrix **D**, the PCA algorithm employed has two parameters that require selection. These are: the number of components that we choose to retain for the VARIMAX rotation, *n*, and the EBSD standard deviation weighting, *w*. Selection of these will be discussed subsequently. To make the processing tractable on a reasonable computer, there is a requirement to divide the full region of interest into smaller tiles so that there is sufficient memory available for the SVD algorithm (we need to hold in memory a rank *r* square matrix). 3-by-3 tiling was employed for the datasets presented in this chapter. The RAM requirement for processing each tile of the dataset considered in this work is 58 gigabytes. Principal components and RCCs are calculated for each tile fully independently, with slight consequence discussed in **Section 1.3.3**.

3.3 Analysis of PCA output

Reshaping the first p^2 rows in all *n* columns of **CR** into *p*-by-*p* images reconstructs *n* RC-EBSPs. The final *q* rows for all *n* columns correspond to RC-spectra. These are separately analysed and quantified. The reduced dataset of characteristic patterns and spectra has a superior signal-to-noise ratio to the experimental measurements. The data in the form recovered from the weighted PCA and VARIMAX rotation can be analysed, but there may issues for instance where two carbides have the same chemistry and phase (i.e. similar EDS spectra) but different crystal orientations (varying EBSD data). Therefore it is useful to analyse the labels in more detail.

The data are labelled first using EBSD pattern analysis. In this work the refined template matching (RTM) approach developed by Foden *et al* [**137**] is applied to assign phase and orientation to each of the point labels. This method involves cross-correlation of test EBSPs (in this case the RC-EBSPs) with a database of library patterns, sampled with the fundamental zone of SO(3) space for each crystal structure at a specified angular frequency. The master patterns were dynamically simulated using *Bruker DynamicS* [**132**], [**247**] and reprojected in *MATLAB* using the pattern centre calibrated from the Ni-rich matrix. Sampling of SO(3) was performed with a frequency of 7° and refinement was used to upsample the orientations. Templates were generated using five input candidate crystal structures were considered for RTM template matching: FCC Co, $M_{23}C_6$, M_2C, M_6C and MC selected from the literature [**2**], [**68**], [**250**], [**251**]. The RC-EBSPs in these datasets were indexed as FCC Co, M_6C or MC (selecting each phase based upon the highest ranked cross correlation value and scrutiny of pattern matching).

For EDS, the data is exported in a format that can be analysed directly in *Bruker eSPRIT* 2.1, and quantify RC-spectra using a 'ZAF' correction algorithm that accommodates the 70˚ sample tilt required for EBSD. We additionally compute and compare the average measured EDS spectra from all points assigned to a given RCC label.

4. Optimising weighting and VARIMAX rotation

Here the effect of two important input parameters, *n* and *w,* will each be discussed. These correspond to the number of principal components we retain and the weighting factor for EBSD information that leverages PCA in favour of either EBSD of EDS.

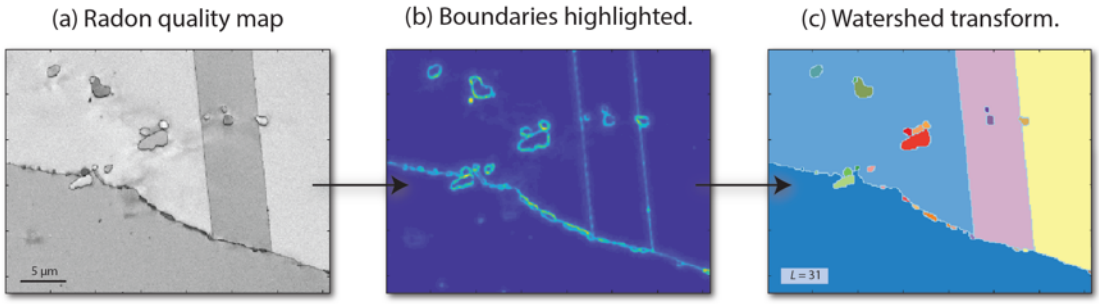

Figure 1-3: Filtering and watershed transform of Radon quality map to determine a value of L for VARIMAX rotation.

4.1 Retention of components

While the number of principal components worth retaining, n, can be selected by hand based upon a qualitative assessment of the appearance of our final micrographs, it is also useful to explore whether there are quantitative, or semi-quantitative assessment processes that can guide the selection. This is a problem that has been considered extensively [**200**], [**213**], [**230**], [**231**], [**252**].

Here, a VARIMAX rotation is performed on the first n principal components. A single label is expected to represent structural and chemical information contributed from one grain. If n is greater than the number of grains in the AOI, the additional labels will correspond to sub-grains (which may be advantageous but will reduce the signal to noise ratio of the RCCs). If n is too large, the calculated scores of nearby (and similar) points will be high for multiple labels and there will be noise in the assignment map. If n is too low, the grains will not be properly segmented and information will be lost as the dataset is over-reduced. This work presents two approaches for selecting n. The first involves counting the number of grains in the measured EBSD-based Radon quality map. The second imposes a limit on the contribution of the first n principal components to the total variance of the dataset.

4.1.1 Counting grains to select n

A reasonable value of n is an estimate of the number of grains in the EBSD-based Radon quality map, L. This was calculated with *Bruker eSPRIT* 2.1 and is an essentially 'free' microstructural image that is spatially consistent with the EBSD and EDS measurements. Several approaches have been reported for counting the number of grains in a microstructural image. These include the application of an H-concave transformation to channelling contrast forescatter electron images (with subsequent refinement) developed by Tong *et al* [**253**]. This grain counting step is employed to select the locations of a dramatically reduced number of EBSD patterns for an orientation map. The 24-bit information depth of the RGB colour image constructed from the forescatter diode intensities allows segmentation of scan points into labels of similar colour and contrast. A subsequent refinement step where each point is compared to the labels of its neighbours leads to a very accurate image reconstruction. Campbell *et al* [**254**] utilise a Watershed transform to identify and distinguish phase fractions and morphology in grayscale SEM images of Ti-6Al-4V. This algorithm treats an image as a topographic region of intensity basins. A labelled source is placed at each local minimum and allowed to flood the image. Image segmentations are delineated where floods from different sources meet [**254**]–[**256**]. When applied to the local gradient of a microstructural image (in order to highlight boundaries and leave grain interiors with low intensity), reasonably accurate intensity classification can be achieved. However, the algorithm has a tendency to over-segregate and assign too many labels.

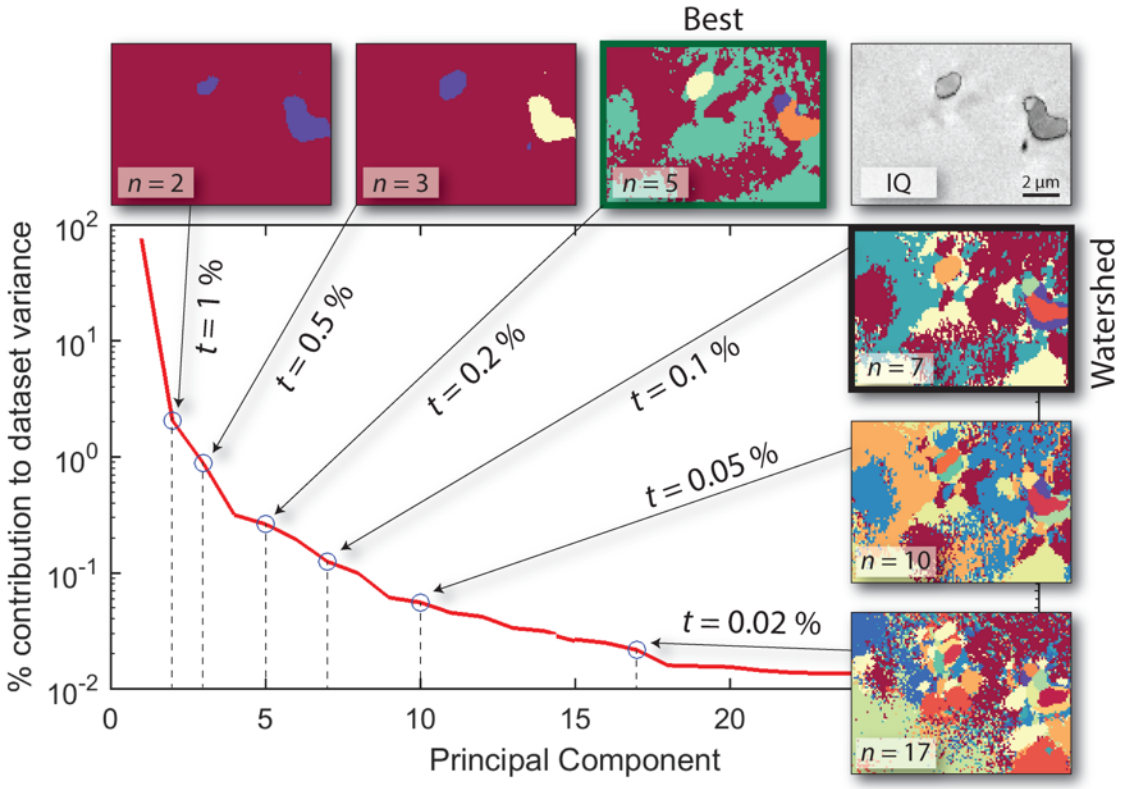

Figure 1-4: Selection of n based on applying a cut-off tolerance, t, for percentage contribution to total dataset variance of the n^{th} principal component. The third PC contributes < 1% of the total dataset variance. The 18^{th} PC contributes < 0.02%. The watershed algorithm discussed in **Section 1.3.1** selected seven components for this AOI.

For its speed and simplicity the Watershed algorithm was selected to quickly identify a value of L. The image processing steps are presented in **Figure 1-3a**. Starting from the EBSD-based Radon quality map (**Figure 1-3a**) a series of local averaging filters, ending with the image complement of a standard deviation filtered image and local minima flattening, are used to highlight boundaries (3b) A watershed transform then assigns labels to different regions (3c) following the topographical method of Meyer [**255**]. We then select n equal to L principal components, and then we perform the VARIMAX rotation.

This approach is fast and provides a reasonable estimation of n for an area of interest. As will be shown subsequently the watershed algorithm significantly oversamples the subsets, especially where coherent intergranular precipitates with similar orientation and chemistry are counted separately by the watershed algorithm. This leads to too great a value of n being selected and the signal-to-noise ratio of the characteristic patterns and spectra are not optimised.

Care should be taken, as the ability for the Watershed algorithm to determine the number of components will depend on the types of features presented. In this example, an annealed Co/Ni matrix is explored and so there is minimal contrast in that region, despite the desired focus being on the number of signals from the carbides. In noisier Radon quality maps it seems likely that the watershed algorithm would perform significantly worse. This could potentially be somewhat mitigated for a wider variety of datasets by further *ad hoc* filtering and image pre-processing.

4.1.2 Limiting variance contribution of the n^{th} principal component.

Conventional PCA relies on the relatively subjective identification of an inflection in a Scree plot (for an example, see **Figure 1-4**) to estimate the number of components that significantly and sufficiently describe the variance of the dataset [**200**], [**213**]. The Scree plot describes the explained variance contribution to the dataset as a function of principal component, derived from the corresponding eigenvalues of the covariance matrix (singular values of the data matrix).

In this plot the principal components are ordered by their contribution to the variance, with the first being the highest contributor (largest eigenvalue). The inflection point in a Scree plot may indicate where the principal components cease to add new substantive value to a description of the data, but often the inflection is unclear. Alternatively, the total explained variance for a certain number of retained principal components itself provides a metric for the utility of the retained components. This idea can aid in selection of n, the number of principal components to retain for VARIMAX rotation and generation of characteristic patterns (RC-EBSPs) and spectra (RC-spectra). The parameter n can be chosen as the integer for which explained variance of the $(n+1)^{th}$ principal component falls below a certain threshold, t, for example 0.2%. All n principal components then contribute a variance proportion greater than t to the dataset. Information contained in the remaining principal components is discarded (i.e. this is considered as noise). The short circuit caveat around this selection rule is that n must be greater than or equal to 2 for VARIMAX rotation, so in some cases (with high initial t) we are forced to select n with variance contributions greater than t. Then to choose n:

$$VarianceContribution\{\ (n+1)^{th}\ PC\ \} < t$$
$$If\ n < 2, set\ n = 2. \qquad \textit{1-3}$$

This is a relatively crude selection algorithm, but it is demonstrated that it works. More sophisticated methods discussed by Raîche *et al* [**200**] using non-graphical metrics such as the Scree test optimal co-ordinate or Scree test acceleration factor could be employed. These values of n correspond to analyses of the gradient of the explained variance (equivalently the eigenvalues of the covariance matrix) as a function of n. These criteria are not guaranteed to be entirely suitable for our datasets, and they systematically select the maximum statistically allowable number of components as described by Kaiser's rule, which stipulates that the eigenvalue of the n^{th} retained component must not be less than one [**257**]. As per **Equation LR-52**, the eigenvalues are a rescaling of the corresponding variance contributions.

Considering the first tile of the full AOI, **Figure 1-4** shows the Scree plot for a PCA with EBSD weighting parameter (discussed further in section 3.1), w, equal to one, and the resulting assignment maps after VARIMAX rotation for values of n selected with varying variance tolerance criteria, as well as the Watershed algorithm. The variance tolerance limit, t, is selected to vary between 1% (leaving two principal components) and 0.02% (leaving 17 principal components).

The quality of the label assignment to each point can be quantified by normalised cross-correlation of measured EBSD patterns and EDS spectra with a point's corresponding characteristic RC-EBSP and RC-spectrum. Maps of these correlation values are helpful to visualise how this varies across the AOI, and how well different precipitates and grains match to their corresponding RC-EBSPs and RC-spectra. Maps of these correlation qualities (normalised correlogram peak heights, χ_{EDS} and χ_{EBSD}) are presented in **Figure 1-5a**. In the presented case for $w = 1$ there is little variation in χ_{EDS} as dataset variance is dominated by EBSD information. A quadrature combination of EBSD and EDS correlogram peak heights, χ_{comb}, is also considered:

$$\chi_{comb} = \sqrt{\chi_{EBSD}^2 + \chi_{EDS}^2} \qquad \textit{1-4}$$

Four possible metrics for the classification quality of an analysis are suggested. These are each plotted as a function of the variance of the final, n^{th}, component in **Figure 1-5**:

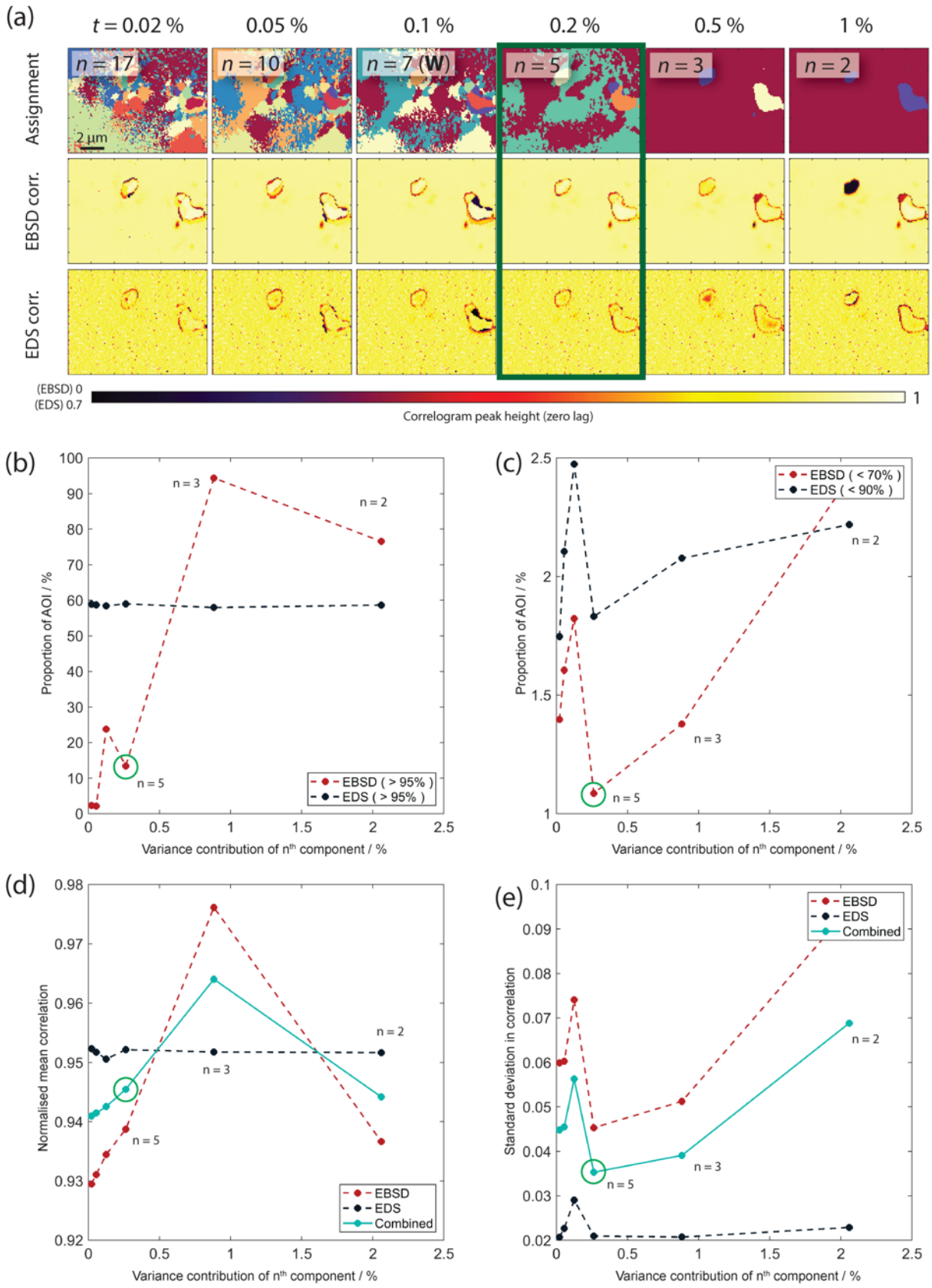

Figure 1-5: (a) Effect of varying variance rejection percentage on measured EBSD pattern and spectrum cross-correlation peak height with corresponding RC-EBSPs and RC-spectra. Colour is cross-correlation peak height. (b) and (c) show percentage of these EBSD and EDS cross-correlation peak heights greater or less than cut-off proportions of maximal correlation in that analysis. (d) and (e) show mean and standard deviation of the cross-correlation peak heights for EBSD, EDS and the quadrature combination of the two.

- Metric 1: The proportion of points that satisfy $\chi > 0.95\ \chi_{max}$ (for χ_{EBSD} and χ_{EDS}) - to be maximised, **Figure 1-5b**.
- Metric 2: The proportion of points that satisfy $\chi_{EBSD} < 0.7\ \chi_{EBSD,max}$ or $\chi_{EDS} < 0.9\ \chi_{EDS,max}$ - to be minimised, **Figure 1-5c**.
- Metric 3: Mean value of χ_{EBSD}, χ_{EDS} or χ_{comb} – to be maximised, **Figure 1-5d**.
- Metric 4: Standard deviation in χ_{EBSD}, χ_{EDS} or χ_{comb} – to be minimised, **Figure 1-5e**.

Selecting the 'best' value of n for an AOI can be made less subjective by choosing a variance tolerance limit that maximises or minimises one of these metrics. When the tolerance limit is relaxed (low t, high n, with n^{th} component contributing only a small amount of variance) the AOI is oversampled with principal components. This leads to a noisy assignment map (observed for the assignment map in **Figure 1-5a** with n = 17), as nearby (and similar) points have similar scores for several RCCs. As the tolerance limit is tightened (moving to the right in the graphs of **Figure 1-5b-e**), the assignment initially improves across most of the metrics. Percentage of correlogram peak heights close to the maximal values increases (**Figure 1-5b**), and percentage much less than the maximal decreases (**Figure 1-5c**). The mean increases (**Figure 1-5d**) and standard deviation falls (**Figure 1-5e**). This initial improvement is attributed to improving the signal to noise ratio of the RCCs, as superfluous principal components are rejected. As the tolerance limit is tightened further and n is reduced, the assignment moves past an optimal position. Beyond this point insufficient principal components are included in the VARIMAX rotation. This leads to insufficient (and inaccurate) RC-EBSPs and RC-spectra. Accordingly the cross-correlation peak heights for this dataset fall – the mean decreases and standard-deviation rises. It is also observed that the Watershed algorithm oversamples the dataset.

For the analysis presented in **Figure 1-5** a choice of t = 0.2% provides a good assignment. This was selected as the minimum in the standard deviation based metric 4. This t also performs best in metric 2 (**Figure 1-5c**) and second best in metric 3 (**Figure 1-5d**). Metric 4 appears to be a good choice for deciding the optimal value of t (and therefore selecting n in independent datasets). A choice of t = 0.5 % (n = 3) leads to superior metrics 1 and 3 than t = 0.2% (n = 5). However, the assignment map in **Figure 1-5a** shows better correlation (both EBSD and EDS) for t = 0.2% (n = 5). Evidently the matrix regions correlate slightly better for t = 0.5% than for t = 0.2%, raising the mean peak height despite an observed poorer correlation for the precipitate regions. An approach of minimising poor correlation, by either of metrics 2 and 4, is less sensitive to this effect, and t = 0.2% (n = 5) exhibits obvious minima. Furthermore, the percentage difference between measurements of mean χ_{comb} is far smaller than that for standard deviation of χ_{comb}, providing a more justifiable minimum.

It was found that metrics 1 and 2 are sensitive to the choice of proximity parameters (here 95% maximal for EBSD and EDS, 70% minimal for EBSD and 90% minimal for EDS). In contrast, it is observed that trends in standard deviation of χ_{comb} as a function of n^{th} component variance are stable between datasets, choice of w, and the specific values of n that the tolerance limits correspond to, shown in **Figure 1-6**. Based on the stability of the standard deviations in χ_{comb}, a variance tolerance t = 0.2% was selected for selection of n in subsequent analysis of the effect of the EBSD weighting parameter w.

4.2 Leveraging relative EDS variance

The dataset variance contribution of the EDS energy bins, the final q rows of **D**, is altered *via* a scaling factor in order to bias the PCA in favour of EBSD or EDS information. Without any weighting, the far greater number pixels in an EBSD pattern compared to energy bins in an EDS spectrum (65,536 for a 256-by-256 pixel EBSD pattern, and 2048 channels for our energy-binned spectra) dominate the variance of the dataset unless the former is dramatically scaled

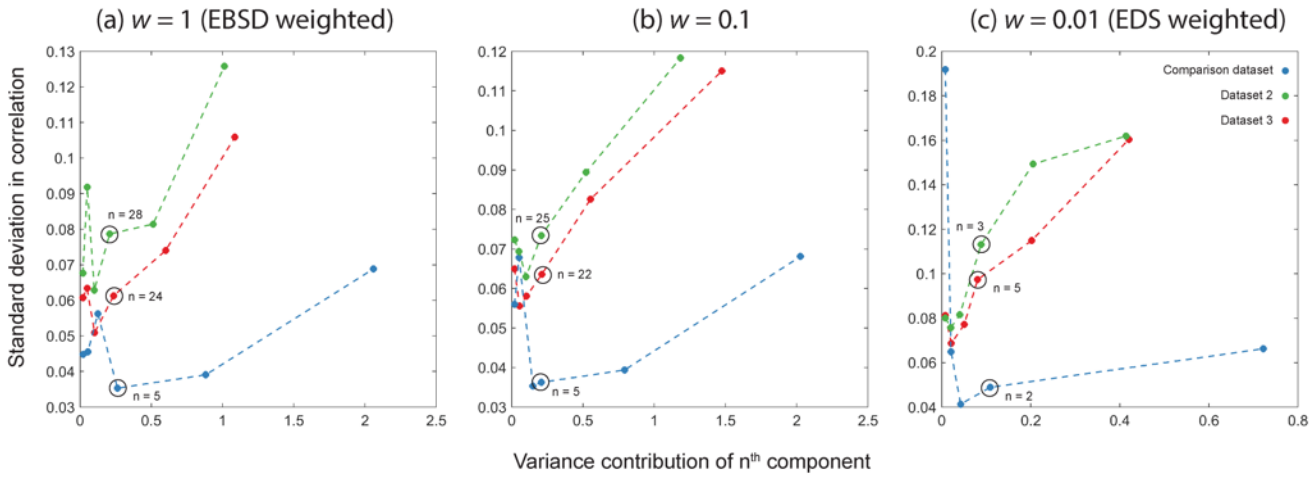

Figure 1-6: Standard deviation of normalised cross-correlation between measured EBSPs and spectra with their associated characteristic EBSPs and spectra, combined in quadrature. Shown for three alloys and three EBSD weighting parameters. As more principal components are retained, correlation initially improves as AOIs are better matched. After reaching the optimal level correlation degrades as principal components corresponding to noise are included in the VARIMAX rotation. The trend is stable (the minimum occurs in approximately the same place) across independent datasets and weighting parameters. n corresponding to $t = 0.2\%$ is circled in each case.

down. This is achieved by multiplying the EBSD and EDS rows of **D** by w and one respectively [**199**]. The variance normalisation step employed during the construction of **D** separately reduces the standard deviation of the input EBSPs and EDS spectra to one. The weighting multiplication therefore reduces the standard deviations of the input EBSPs to w, while that of the input spectra remains equal to one.

The tile AOI presented in **Figure 1-5** was processed with weighting parameter, w, varied between 0.001 and 1. A variance tolerance limit, t, was used to select the number of components to retain, n, varied between 0.02% and 1%. Maps of standard deviation and mean of χ_{comb} as a function of t and w are presented in **Figure 1-7**. A local minimum in standard deviation is identified. The corresponding assignment map is included (**Figure 1-7c**), along with maps of χ_{EBSD} (**Figure 1-7d**) and χ_{EDS} (**Figure 1-7e**).

The same trends in variance tolerance limit are observed as in **Figure 1-5** and **Figure 1-6**, which present χ_{comb} as a function of the variance contribution of the n^{th} component for a single w. As the tolerance limit is tightened and the n^{th} component has to contribute more variance (left to right in **7a**, **7b**), the standard deviation in χ_{comb} falls and the mean rises. As in **Figure 1-5**, a local minimum in standard deviation is observed as t is varied. Beyond this limit insufficient principal components are retained. As w is increased (tending towards EBSD weighting, upwards in **Figure 1-7a**, **Figure 1-7b**), standard deviation in χ_{comb} generally decreases. The mean value of χ_{comb} increases as w increases for loose variance tolerance limits (AOI oversampling), but at higher t (smaller n) there appears to be less of a correlation between mean χ_{comb} and w. Considering the standard deviation in χ_{comb} (metric 4 of **Section 4.1.2**) as a measure of assignment quality identifies a seemingly optimal combination of w and t. Associated label assignment, χ_{EBSD} and χ_{EDS} maps are presented (**Figure 1-7c-e**).

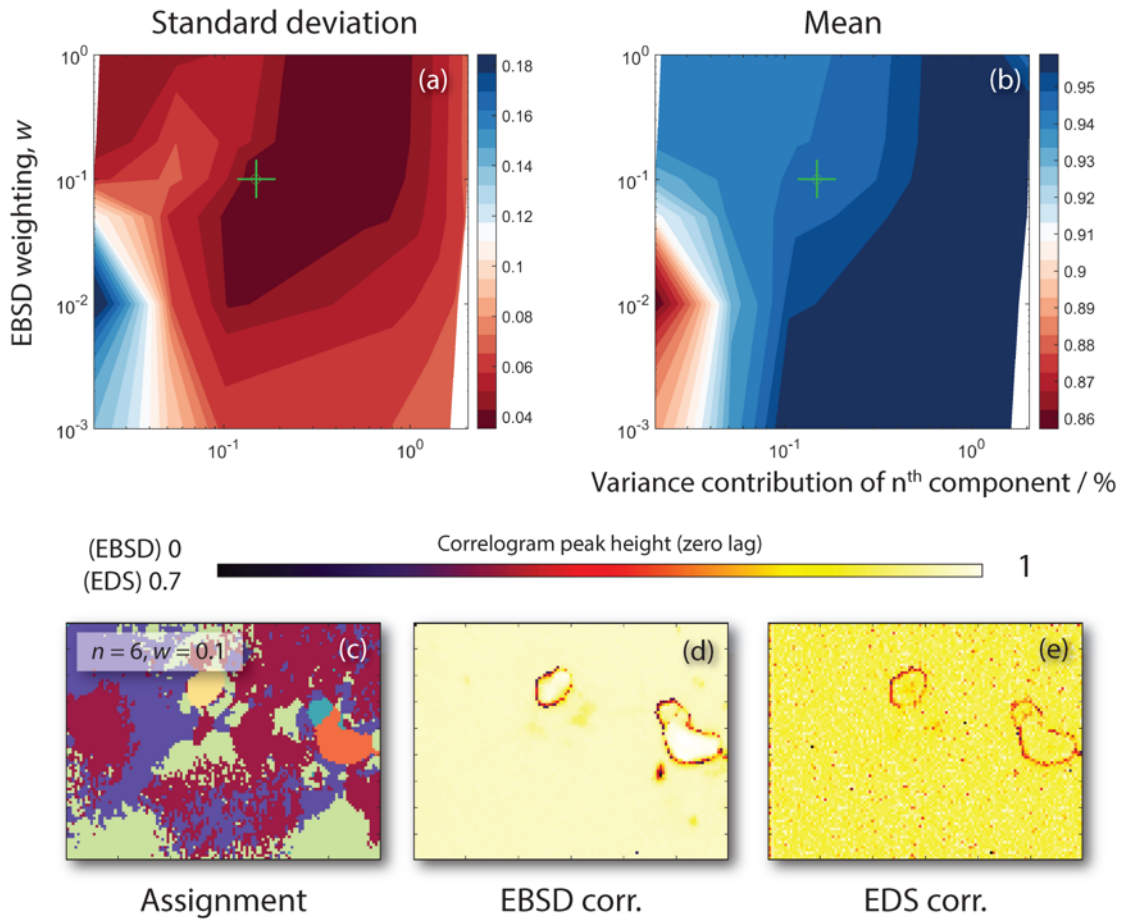

Figure 1-7: Standard deviation (a) and mean (b) of combined EBSD and EDS RCC/measurement cross-correlation peak heights, χ_{comb}, defined in **Section 1.4.1.2**. The 'best' assignment map is presented in (c), along with correlation peak height maps for EBSD, χ_{EBSD} (d) and EDS, χ_{EDS} (e). 36 data points are included in the t / w parameter space maps.

4.3 Full dataset processing and assignment artefacts

The PCA approach achieves the desired signal-to-noise improvement for poorly diffracting particles such as M_6C carbides. Example measured, label (RCC), and matched simulation patterns for the pseudo-FCC matrix and M_6C carbide are presented in **Figure 1-8**.

Considering the full AOI of nine tiles, **Figure 1-9** presents the dataset processed with $w = 1$ (EBSD weighting) and $w = 0.1$ (EDS weighting, identified as the local minimum in standard deviation of χ_{comb} in section 3.2). RC-EBSPs were indexed using the RTM procedure [**137**], and RC-spectra were quantified with *Bruker eSprit* 2.1.

When the EDS weighting is high (**Figure 1-9b,d,f,h**), label assignment is dominated by the magnified variance of the EDS spectrum energy bins. To qualify this, we explore the interaction volume using Monte Carlo simulations and the continuous slowing down approximation, this provides an indication of the interaction volume of the electron beam and X-ray generation. It is noted that the generation of the background signal for EBSD is likely to be smaller than predicted from the CSDA-approximation as the energy of the electrons that form the Kikuchi bands is constrained to be closer to the primary beam than the CSDA predicts [**258**]. The interaction volume of backscattered electrons in this system, simulated with CASINO at 70˚ sample tilt, is at most 100 nm. That for secondary (X-ray generating) electrons is

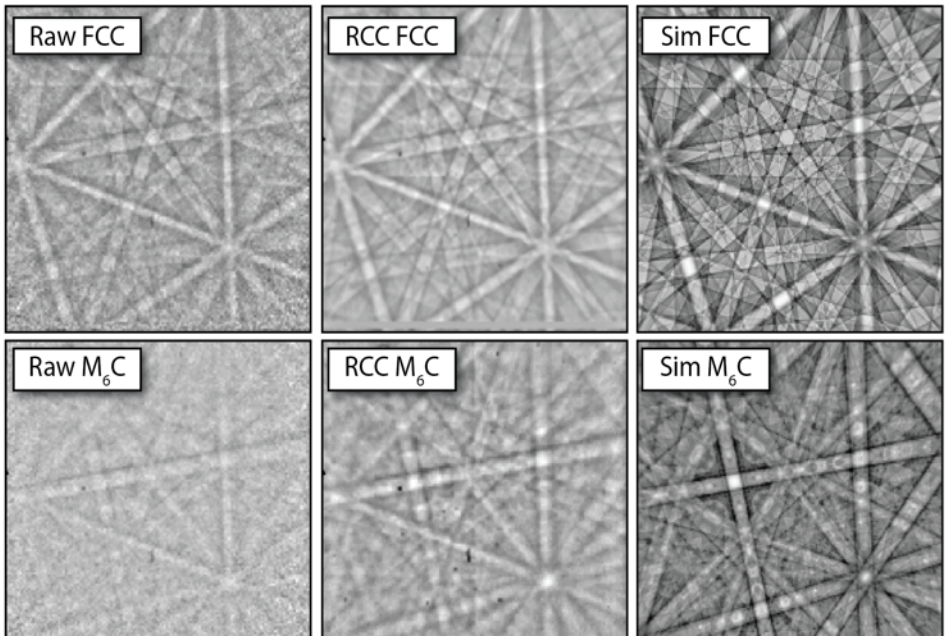

Figure 1-8: Comparisons of example raw, rotated characteristic component (RCC), and template matched dynamically-simulated EBSPs for the pseudo-FCC Ni/Co matrix and the M_6C carbide phase. This demonstrates how the method amplifies the quality of the minor phase substantially, which assists in unambiguous classification.

significantly larger. The Monte-Carlo simulation suggests that secondary electrons from the FCC matrix are generated up to perpendicular depths of 1 μm. Carbides exhibit even larger volumes, with Cr_6C and ZrC carbides interacting up to 1.4 μm and 1.7 μm respectively.

The consequence of large EDS interaction volumes is that at high magnification (where scan step is much less than the coarser technique resolution) two adjacent scan positions with measurably different crystal structure may exhibit very similar EDS spectra. The position vectors of these observation sets in variable space are very similar, despite differences in the measured EBSD pattern, due to the demagnification of EBSD pattern variance in this analysis. Effectively this leads to a loss of spatial resolution in label assignment, and as highlighted at position A, the possibility of missing fine precipitates from the classification. It can be seen in the EDS weighted assignment map **Figure 1-9d**, that an elongated characteristic region of C enrichment and different chemistry, starting at point A, follows the grain boundary. The greater number of matrix points in this region dominate the principal component-EBSP, and the RC-EBSP for this region indexes as FCC Co.

Another artefact we observe in the assignment is highlighted at position B. In the EDS-weighted PCA it is observed that the upper region of a precipitate (MC carbide) grain is assigned a different orientation to the remainder below. This artefact is the coincidence of two method limitations. The need to tile the dataset due to the significant memory requirement of the SVD algorithm means the upper and lower regions have to be assigned labels independently. This is not an issue for matrix regions, as there is a sufficient population (and therefore dataset variance contribution) to assign noise free and appropriate RC-EBSPs and RC-spectra. However, when the PCA is EDS-weighted there is insufficient EBSD variance (due to a small population of points in the upper half of the mis-assigned precipitate) to provide a label that contains sufficient orientation data for this region to be separated from the nearby second grain of chemically similar MC carbide. This results in the upper region of the split precipitate being assigned the correct chemistry and phase, but

an incorrect orientation. This could be compensated for by relaxing the variance tolerance limit imposed on this analysis or an alternative sampling strategy to retile for small segments towards the tile edge.

A third artefact is highlighted at position C. In the EBSD weighted **Figure 1-9f**, one MC carbide precipitate is identified and labelled. Two are assigned in EDS weighted **Figure 1-9e**. In this case the applied variance tolerance limit of 0.2% is insufficient to separate this region from the surrounding matrix. The region is highlighted in **Figure 1-10**, and score maps for the first six calculated principal components and the VARIMAX rotated RCCs are presented. The precipitate is mis-indexed in this analysis due to insufficient sampling of principal components. No PC strongly contributes to the precipitate grain (**Figure 1-10**, highlighted in 1-6), and little signal from this region is included in the VARIMAX rotation and calculation of the six RCCs. This results in the precipitate being labelled with an FCC Co matrix dominated RC-EBSP, and eventually indexed as such.

5. Discussion

Easy access to advanced statistical treatments enable us to treat microscopy data as a ‘big data’ problem and the community is likely to see increased use of these approaches. This work has illustrated that consideration of the data modality (e.g. physical processes to generate the signal, combined with the statistical variance of each data type) provides improved confidence in their use to reasonably, and usefully, segment large data sets.

The present work presents a limited size of region with only a few domains, but testing using a number of other (lower magnification) maps indicates that the variance tolerance limit is a good indicator of the number of domains - grains, sub-grains, and precipitates - within Ni-based maps containing many more of these features.

The combination of EDS and EBSD signals together using a weighted PCA approach, with subsequent label identification and characterisation, improves phase characterisation within the scanning electron microscope. To combine these modalities, one has to select an appropriate data processing pipeline to provide robust data mixing, with subsequent selection of an appropriate number of components for retention prior to VARIMAX rotation. This is required to inform correct identification of the labelled regions. This is achieved through selection of suitable values for the two independent parameters w and n (the latter varied through t, the variance tolerance limit). Here the presented approach is reviewed, then applications and potential utility of the technique discussed.

5.1 Parameter choice and data-type leverage

It has been shown that a PCA approach may be biased towards obtaining RCCs through identifying the strongest signals in either EBSD or EDS information. An EBSD-weighted PCA exhibits a finer effective spatial resolution in label assignment, due to the smaller interaction volume for electron backscattering than for X-ray generation. From this one obtains RC-EBSPs and RC-spectra identified from structurally contrasting points in an AOI. PCA can also be leveraged in favour of EDS spectrum dissimilarity and identify RC-EBSPs and RC-spectra accordingly. In this case a coarser assignment resolution is observed, but by slightly weighting towards their EDS signal the label assignment improves on several metrics.

Two approaches have been presented for selection of the number of principal components to retain for an AOI. Counting the number of grains in an EBSD quality map, for example with a Watershed algorithm, may be susceptible to

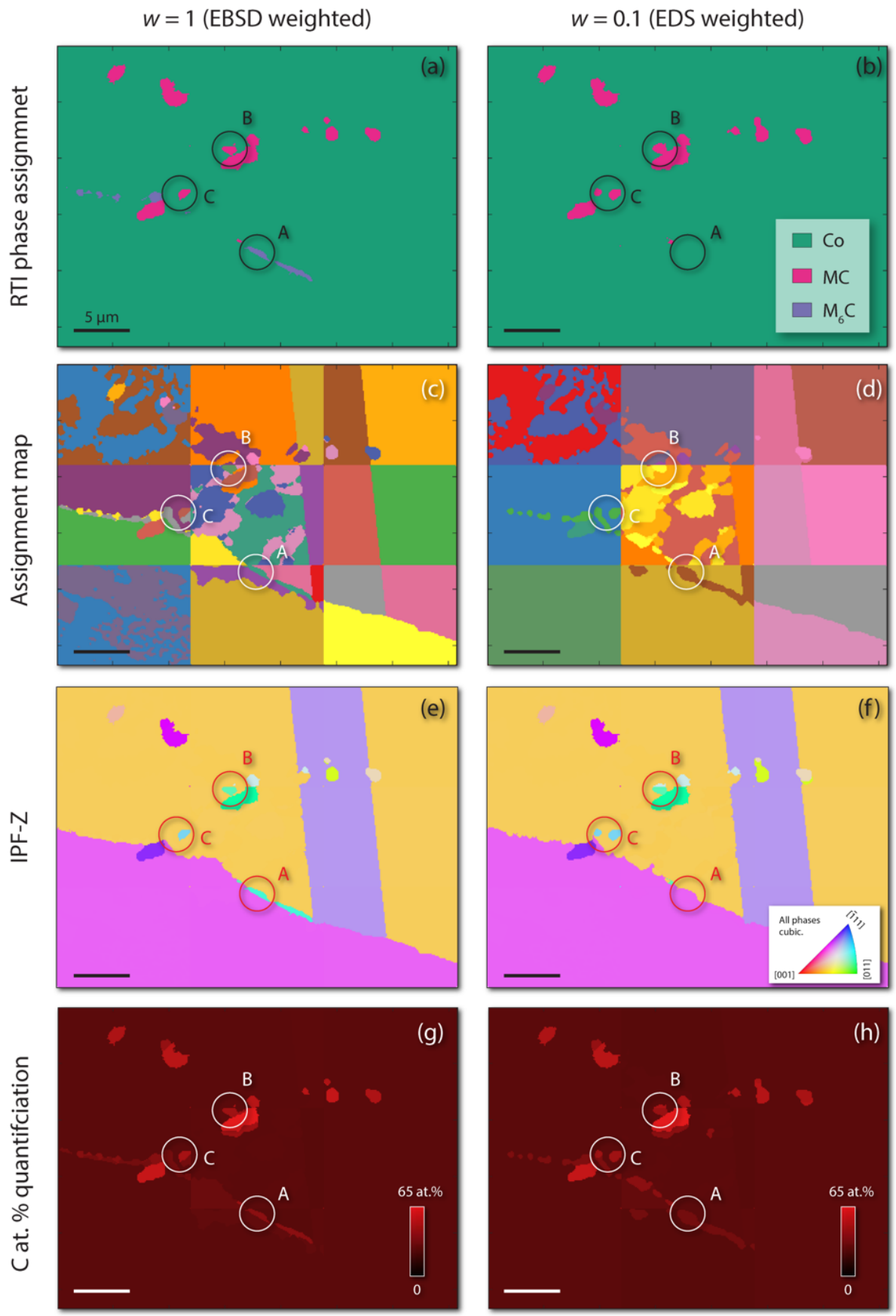

Figure 1-9: Comparison between phase assignment (a-b), label assignment, with arbitrary colouring (c-d), IPF-Z – out of plane (e-f), and C at.% from the RC-spectra quantified with Bruker eSprit 2.1 (g-h), after processing with w = 1 (EBSD weighted) and w = 0.1 (EDS weighted). Both analyses were performed with a variance tolerance limit t = 0.2 %.

oversampling. This can be due to the fact that coherent and chemically similar boundary precipitates will be counted separately but ideally should share a label. Slight oversampling is not a problem, but significant oversampling may make the components recovered after the VARIMAX rotation difficult to interpret. Furthermore it will reduce the ability of the method to amplify weak signals. A more systematic approach for selecting n is to consider the contributions of retained principal components to total dataset variance and impose a limit beyond which we discard residual information as noise. A small t (*eg*. 0.02%) corresponds to retaining many principal components as even those contributing relatively little signal are permitted to participate in the analysis. Increasing t restricts the number of principal components, as we impose a low pass filter on the proportional variance contribution to the dataset required. This improves the signal-to-noise ratio in the RC-EBSPs and RC-spectra. This leads to an increase in the mean and reduction in standard deviation of correlogram peak heights for cross correlation of measured and characteristic EBSD patterns and EDS spectra. When t gets too high, the mean and standard deviation of correlation peak heights falls as the number principal components required to segment an AOI is undersampled.

Optimal choice of parameters will depend on what dataset insight is required from an analysis. If one wishes to reduce a dataset to as few RCCs as possible then a fairly tight tolerance limit provides a mechanism for quantitatively limiting the significance requirement of features in an AOI. If precipitates/grains of interest are small, reducing the tolerance limit permits weaker dataset signals to be assigned their own component. This may lead to oversampling of the dataset, reducing signal to noise.

Weighting PCA in favour of EBSD yields a finer effective spatial resolution in assignment than EDS weighting. This is advantageous if spatial precision is required, and especially when analysing phase presence in an AOI. Leveraging towards EDS can improve RCC assignment by several metrics. In some cases assigning chemically similar but structurally different regions (due to overlap in EDS interaction volume) to the same label may be compensated for by relaxing the variance tolerance limit and encouraging dataset oversampling. It may be the case that an EDS-weighted PCA would prove useful in situations where crystal pseudosymmetry or other similarity in Kikuchi bands reduces EBSD pattern contrast between two regions.

5.2 Chemical analysis of labelled phases

Label chemistry can be quantified from RC-spectra independently of structure-ID from RC-EBSPs. Comparisons between chemistry and crystallography may then be made, with the benefit of a reduced signal to noise ratio of RC-spectra and RC-EBSPs than the raw measurements. RC-EBSPs and RC-spectra are simultaneously calculated and assigned to regions of an AOI. They are not independent, and reflect the most significant structural and chemical signals of points that they strongly load.

Figure 1-11 presents chemical maps (quantified RC-spectra) of a dataset for both directly quantified RC-spectra (**a**) and average spectra for a given RCC label (**b**). The same PCA parameters were employed as the results shown in **Figure 1-9**(**a**,**c**,**e**,**g**). They are almost identical. It is observed that all precipitates exhibit Ni and Al depletion. The intergranular M_6C carbides (**Figure 1-9**) show Mo enrichment, while the intragranular MC carbides exhibit Mo depletion.

Figure 1-12 shows elemental quantification of chemistry for each phase, along with standard deviations. This is performed for both directly quantified RC-spectra (**Figure 1-12a**) and quantified average spectra for a given label (**Figure 1-12b**). It is noted that the trend in refractory element segregation between the carbides. In this system the MC carbide is strongly Ta and Zr enriched. Mo, Cr and W segregate to the M_6C phase.

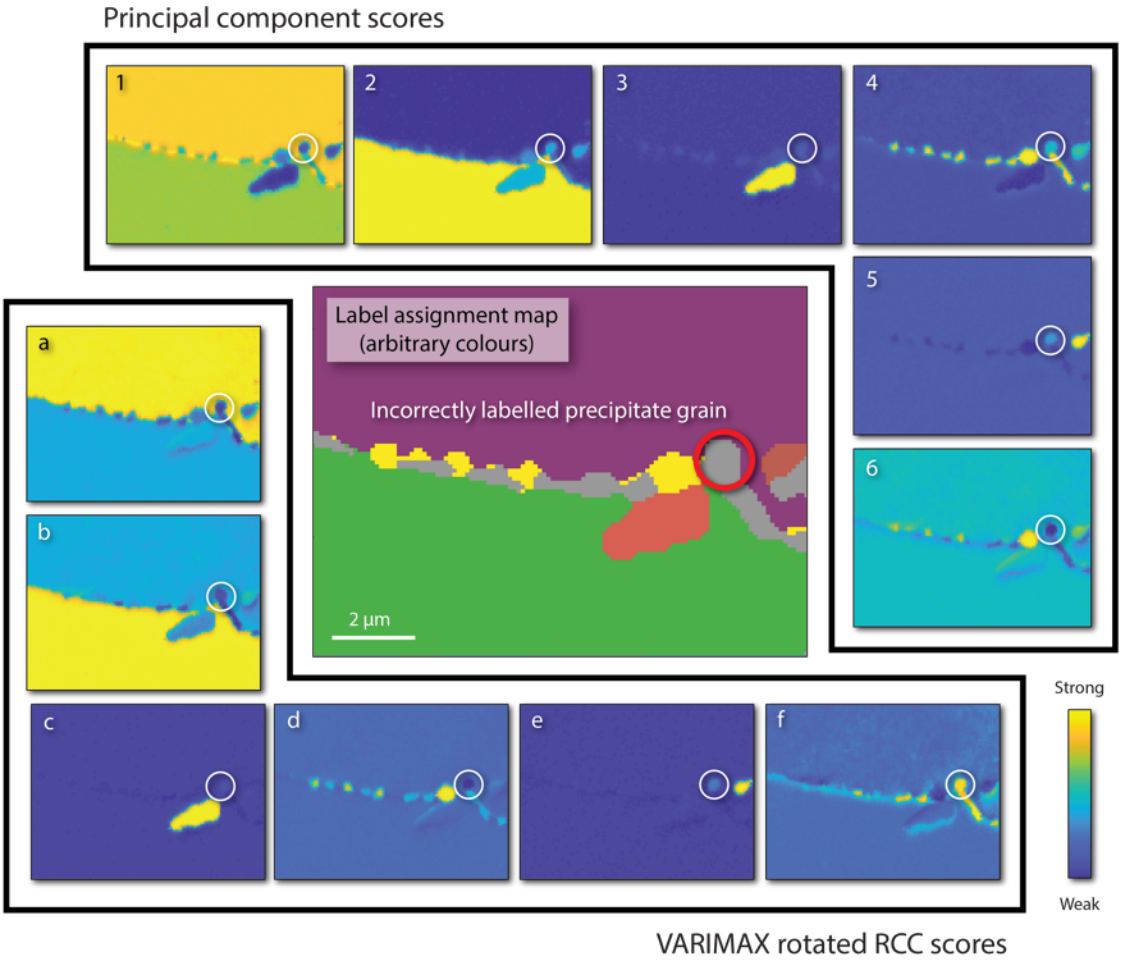

Figure 1-10: Second tile AOI (with arbitrary label colouring) of **Figure 1-9** with EBSD weighting, $w = 1$. The mis-labelled precipitate grain (point C in **Figure 1-9**) is highlighted. PC scores and VARIMAX rotated RCC scores are presented. Principal components are ordered by their contribution to dataset variance, but RCCs are calculated to contribute equal variance. Score colour map is unnormalised between maps to show contrast[1]. No principal component strongly contributes to the highlighted grain (1-6), and the VARIMAX rotated component with the highest score (f) includes significant signal from several other precipitates and matrix. This results in the corresponding RC-EBSP being dominated by FCC Co signal, and indexing accordingly (point C in **Figure 1-9**).

Similar observations of Ta and Zr enrichment in superalloy MC carbides has been reported by atom probe tomography studies [**64**], [**65**]. The technique presented in this work provides a means of confirming trends in precipitate composition between alloys, as well as where elements tend to segregate upon nominal enrichment of the bulk composition.

5.3 Qualitative comparison to state-of-the art post-processing approaches

The PCA approach is a method to amplify signal to noise, where any knowledge of the co-location of measurement points and sum similar signals is removed. In the limit, the EBSD pattern based neighbour pattern average method (NPAR) [**128**] improves signal to noise *via* a summation patterns within a local neighbourhood. This ignores extraction of the similarity of the signal obtained from each neighbouring pattern and could lead to adding of signals from two phases or grains which can affect interpretation of the average signal. The non-local pattern averaging reindexing (NLPAR) approach [**129**] provides delocalised smoothing, using a weighting function based upon the similarity of the pattern information in a moving window centred around a candidate point. Similar approaches have been adopted in the TEM community, particularly using the *Hyperspy* Python package, to obtain characteristic spot diffraction patterns for example with a cluster analysis approach [**237**]. So-called ‘cluster-centre’ diffraction patterns are calculated by grouping

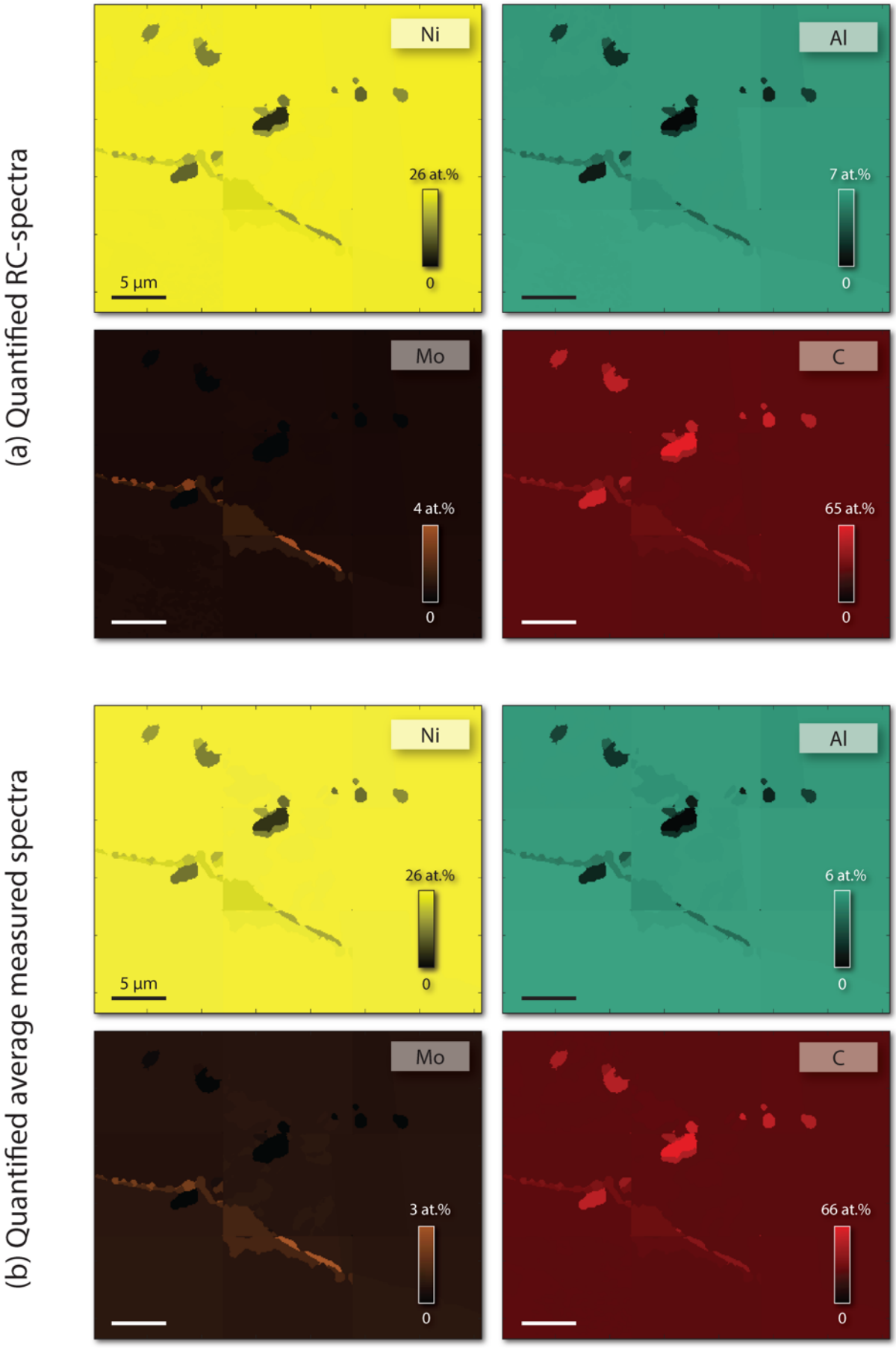

Figure 1-11: Chemical maps (at.%) of quantified RC-spectra. Only Ni, Al, Mo and C are shown for brevity. This analysis was performed with variance tolerance, *t*, of 0.2% and EBSD weighting, *w*, of 1. Maps are shown for directly quantified RC-spectra (a) and average spectra assigned to the same given label (b).

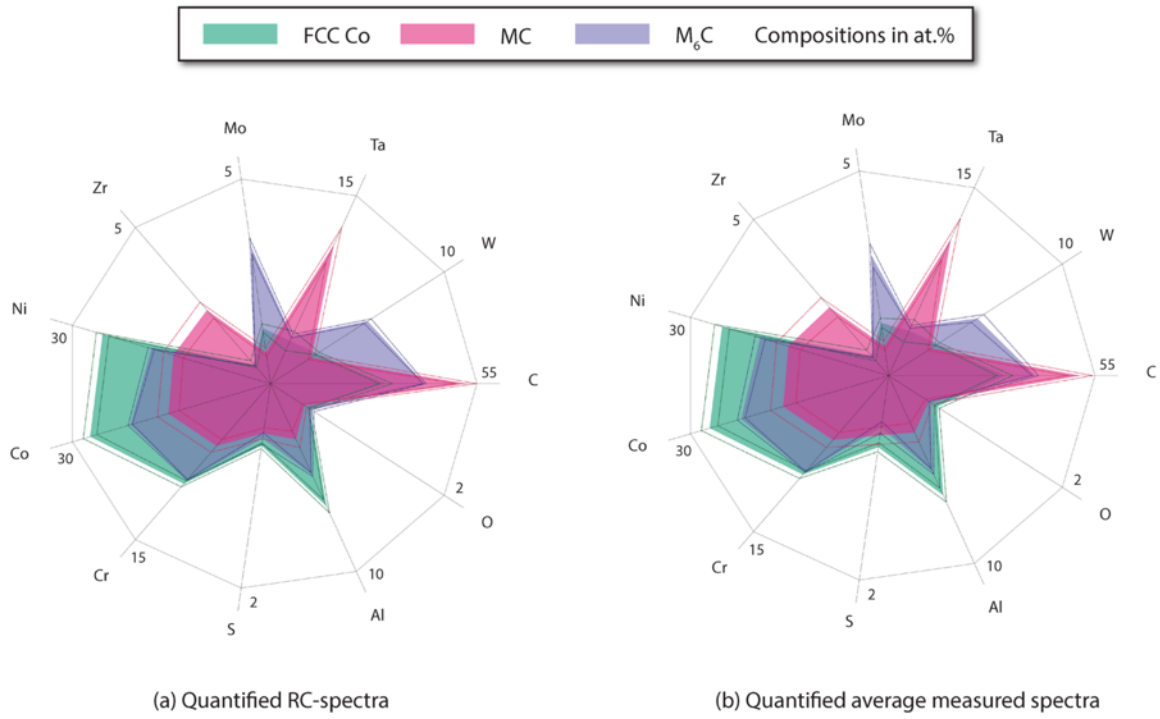

Figure 1-12: Average chemistry of the three identified phases in the dataset presented through this work. As with Figure 9a,c,e,g; Figure 10 and Figure 11, this analysis was performed with a variance tolerance, *t*, of 0.2% and an EBSD weighting, *w*, of 1. Filled regions plot average composition, dotted lines show +/- a standard deviation from the mean. This is shown for directly quantified RC-spectra (a) and average spectra assigned to the same given label (b).

and calculating the average of DPs transformed into variable space. Both cluster-centre analysis and PCA can be considered extreme cases of NLPAR, where the spatial location is discarded and instead only applied statistics are relied on to denoise, label, and index data. Einsle *et al* [**237**] note that raw PCA is not suitable for spot diffraction analysis due to strong similarity in many reflections observed across the area of interest.

Generally, the challenge with a PCA approach is that the components returned represent the statistical dominance of each characteristic signal within the data set, and they are not physical. A VARIMAX rotation for the combined EBSD pattern and EDS spectra results in an easier to interpret label, where each label can be uniquely applied to each point within the map. This works for an EBSD pattern, as the variance between two Kikuchi-based diffraction patterns for different phases is relatively small. For a TEM spot-based diffraction pattern, rotating the data according to a variance model may not be reasonable, because the spot patterns for different phases may have a stronger variation in variance (e.g. due to a different number of reflectors that create spots within the pattern). In practice, this may impact how the diffraction data is pre-processed before putting into the data matrix, as well as a selection of an appropriate weighting scheme when joining the diffraction-based structure data with the EDS-based chemical data.

Keenan and Kotula [**259**], [**260**] have explored scaling for multivariate statistical analysis for time of flight secondary ion mass spectrometry (TOF-SIMS) data. They focussed on how count-dependent pre-scaling impacts the distribution of variance and show that many scaling methods hinder multivariate statistical analysis. Their work highlights that variance scaling can be improved when the Poisson probability distribution of the raw data is accounted for, especially with regard to chemically significant minor features in the TOF-SIMS spectra. In contrast, the present work combines two data types with different noise and scaling methods, using the output signals in separate characterisation processes.

In considering the EDS signal, the heteroscedasticity of noise/error in the measurements is not just a function of the magnitude of the peak (which could be rectified by dividing by a Poisson correction factor for each energy bin, independent of the bin's energy) but also uncertainty in beam energy, sample elemental fluorescence and absorption, *etc*. Accounting for this could lead to formal normalisation of expected variance for a given energy bin, and would be a function of spectrum bin energy, identity of the chemical species, and intensity. The presented method works well for alloy EDS datasets as the X-ray spectra contain many interacting (and covarying) signals. In a dataset there may for example be a majority of scan points with intense Ni and weak Mo peaks, producing a principal component with that self-same pattern of intensity (intense Ni and weak Mo). An improved noise function would be useful where the EDS signal is less distinctly clustered. However, note that in the present workflow the final phase classification is performed on the EBSD signal, with the EDS data used for chemical quantification (interpreted as a function of structure). A similar but different noise model could be applied to each pixel within the EBSD pattern, dependent on the anisotropic spatial distribution of (near elastic) backscatter electrons that can be modulated in counts by Kikuchi diffraction. For example, this could be implemented through inclusion of a pixel and energy bin specific weighting function prior to operation of the 'macro' weighting term we have applied to mix the signals prior to applying the PCA that we have fairly extensively discussed.

In the absence of advanced noise models, this work employs a simpler method. This is focussed on (1) maximisation of the likelihood of successful segmentation of similar domains and the generation of appropriate characteristic signals; (2) the ability to register those domains against the EBSD signal for phase classification (which is augmented by the rotation of components to create uniform variance, as per Wilkinson *et al* [**131**]); (3) the subsequent EDS chemical signature analysis. This approach is simple and sufficiently successful, and has been made available open source *via AstroEBSD* [**119**].

Improvements to this methodology could be motivated by the discussed prior work of Keenan and Kotula [**259**], provided noise and variance models of the EBSD signal and EDS signal can be determined and validated. At present, these are limited and the origins of the signals are still somewhat disputed [**134**], [**143**], [**247**], [**258**]. While this method has a significant grasp of the EBSD signal it is limited in analysis of EDS information, and currently uses standard proprietary software such as *Bruker eSprit* 2.1. Further method development could involve implementing open source quantification (accounting for discussed effects), including better accounting for Poisson noise within our processing toolbox.

6. Conclusions

This chapter develops an analysis pipeline to provide robust correlative microscopy, mixing chemical information obtained using EDS and structural information obtained using EBSD. This enables us to observe small carbides and optimise signal-to-noise for the different phases present. PCA is an effective data processing technique for identifying regions of strong similarity in a dataset (microstructure), while remaining spatially unbiased (two adjacent points share the same propensity to be assigned the same principal component and RCC as two far-field points). Inclusion of both EBSPs and EDS spectra into the data matrix, D, provides a mechanism for obtaining simultaneous structural and chemical fingerprints of features in an AOI. It is possible to weight the identification of these characteristic EBSPs and spectra (RC-EBSPs and RC-spectra) in favour of similarity in crystallography or chemistry between points. The following observations are presented:

1. Selection of the number of principal components to retain for VARIMAX rotation and subsequent analysis can be made less subjective by counting grains in a Radon transform EBSP quality map, for example by a Watershed transform, or by selecting a tolerance limit (low pass filter) for the proportional explained variance of the retained principal components. Oversampling reduces signal to noise ratio in RCCs but reduces the risk of missing fine precipitates from the analysis.
2. An EBSD weighted PCA exhibits a finer effective spatial resolution in label assignment due to the smaller interaction volume of backscattered than secondary electrons, and therefore for EBSP than EDS-spectra generation.
3. Leveraging the PCA slightly towards EDS measurements can improve segmentation (lower standard deviation in cross-correlation peak height) of characteristic EBSPs and spectra.
4. Structural phase-ID of an AOI, for example by a refined template matching algorithm, can be enhanced *via* data reduction of a 40,000 point map to (in the case of the dataset presented in this work) 35 RC-EBSPs. This drives a significant processing speedup, and permits the trialling of many candidate structure libraries, improving confidence in assignment.
5. Quantification of RC-spectra, assigned a structure label by RC-EBSP refined template matching, permits measurements of average chemical segregation between phases.

7. Chapter Acknowledgements

This work was a collaborative effort, involving Alex Foden, Dafni Daskalaki-Mountanou, and Chris Bilsland in addition to TPM, TBB, and DD.

TPM, TBB and DD would like to acknowledge support from the Rolls-Royce plc - EPSRC Strategic Partnership in Structural Metallic Systems for Gas Turbines (EP/M005607/1), and the Centre for Doctoral Training in Advanced Characterisation of Materials (EP/L015277/1) at Imperial College London. AF acknowledges funding from EPSRC. DDM and TBB acknowledge funding from the Shell AIMS UTC. CB and TBB acknowledge funding from the EPSRC Centre for Doctoral Training in Nuclear Engineering (EP/L015900/1) and Rolls-Royce plc. TBB would like to acknowledge the Royal Academy of Engineering for funding his research fellowship. DD acknowledges funding from the Royal Society for his industry fellowship with Rolls-Royce plc. We thank Angus Wilkinson (Oxford) for sharing MATLAB code used in [**131**] and for useful discussions. We used the Harvey Flower Microscopy Suite at Imperial College London for collecting this data. We would like to thank Mark Hardy (Rolls-Royce), Ioannis Bantounas, and Lucy Reynolds (both ICL), as well as Paraskevas Kontis and Baptiste Gault (both MPIE) for ongoing and very helpful discussions.

TPM developed the majority of the *MATLAB* code in the work and drafted the initial manuscript of the associated publication. CB performed CASINO electron interaction simulations for the phases identified by the analysis. DD and TBB supervised the work.

8. Chapter Data Statement

Data is available at DOI: 10.5281/zenodo.3617455. The tools presented here have been integrated into the *MATLAB* based *AstroEBSD* analysis package.

Chapter 2 - Quantitative precipitate classification and grain boundary property control in Co/Ni base superalloys

1. Chapter Summary

A correlative approach is employed to simultaneously assess structure and chemistry of (carbide and boride) precipitates in a set of novel Co/Ni-base superalloys. Structure is derived from electron backscatter diffraction (EBSD) with pattern template matching, and chemistry obtained with energy dispersive X-ray spectroscopy (EDS). It is found that the principal carbide in these alloys is Mo and W rich with the M_6C structure. An M_2B boride, also exhibiting Mo and W segregation is observed at B levels above approximately 0.085 at.%. These phases are challenging to distinguish in an SEM with chemical information (EDS or backscatter Z-contrast) alone, without the structural information provided by EBSD. Only correlative chemical and structural fingerprinting is necessary and sufficient to fully define a phase. The identified phases are dissimilar to those predicted using ThermoCalc. An assessment of the grain boundary serratability in these alloys is performed, and it is observed that significant amplitude is only obtained in the absence of pinning intergranular precipitates.

2. Introduction

Engineering grain boundary character is essential to optimisation of superalloy microstructure and performance. Toward this endeavour we require precise knowledge of chemistry, distribution, and structure of intra/intergranular precipitates, particularly carbides and borides. These phases form through all stages of alloy processing, from initial casting through to final ageing treatments. In the literature, refractory-rich grain boundary precipitates may be referred to as a ‘carbide’ without any effort to differentiate between $M_{23}C_6$, M_6C or boride structures. Due to preferential elemental segregation of high Z-number elements, these all exhibit high backscatter SEM contrast. There is significant evidence that these precipitates’ exact character has a significant effect on mechanical and environmental stability, especially in high temperature applications, despite their small volume fraction [**64**], [**66**]–[**68**]. A secondary effect of grain boundary precipitate interactions is the ability to facilitate grain boundary serration mechanisms, also an essential consideration in modern superalloy grain boundary engineering [**69**], [**70**].

This work develops the V208 series first presented by Knop *et al* [**1**], [**12**], [**261**]. A set of Co/Ni-base superalloys based on V208C are presented, with Mo additions (for solid solution strength) and varied C, B, Zr, and Ti content for grain boundary chemistry adjustment. As-received coarse-grained RR1000 is also characterised, as well as cast and wrought V208C for comparison. Intergranular precipitates are quickly and accurately classified using a correlative electron backscatter diffraction (EBSD) / energy dispersive X-ray spectroscopy (EDS) approach [**93**]. The robustness of the classification is examined, chemical statistics as a function of precipitate structure are presented, and a comparison of observations to thermodynamic modelling is considered. The effect of intergranular phase morphology on grain boundary control is also briefly investigated. It is shown that in this alloy series boundaries may only be serrated in the absence of spatially dense precipitate boundary coverage, regardless of cooling rate from solution.

3. Background

In this study a new series of Co/Ni-base superalloys with varied C, B, Zr and Ti compositions are developed, with an aim to adjusting grain boundary chemistry. The formation of various intergranular phases are observed and discussed,

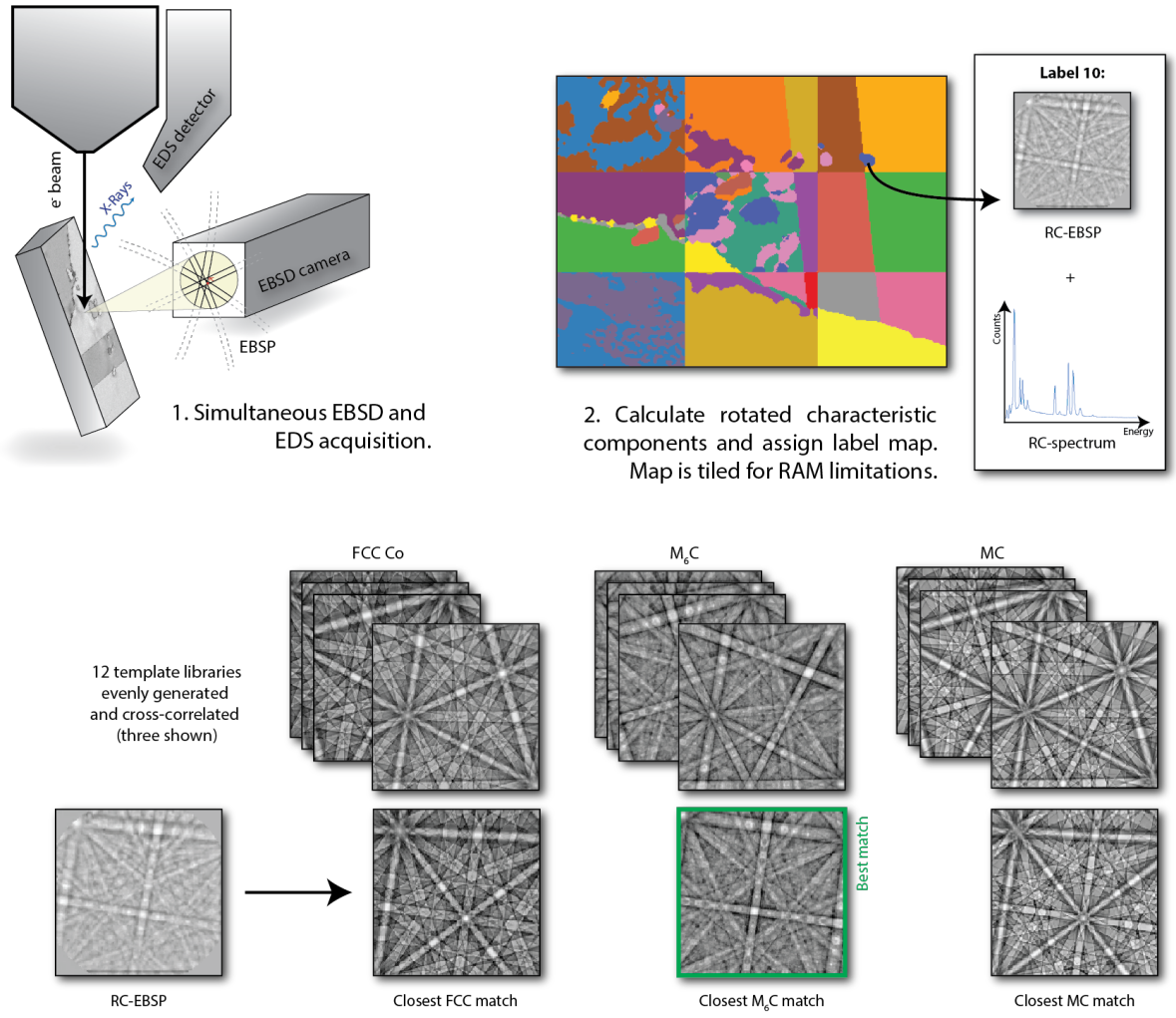

Figure 2-1: The EBSD/EDS PCA approach. Chemical and structural information is combined and corresponding relationships are extracted using statistical methods from the dataset.

and grain boundary morphologies investigated upon independently varied heat treatment. As-received coarse-grained RR1000 and cast V208C are also characterised and compared.

A correlative EBSD and EDS method is employed, together with statistical methods, to enable analysis of large area data sets. EBSD involves the measurement of electron backscatter diffraction patterns (EBSPs). These are formed from near elastic diffraction scattered electrons, and correspond to projections of the lattice planes [**114**], [**117**]. EBSPs are formed from a relatively large solid angle of diffracted electrons, so more than three geometrical conditions (in this case Kikuchi bands corresponding to lattice planes) are normally sampled, as required to determine a unique orientation solution. Conventionally a Hough (Radon) transform and set of interplanar angle lookup tables are employed to index crystallographic planes to determine phase and orientation.

An informatics-type approach is adopted, combining EBSD and EDS information into large data matrices from which we extract the strongest (correlated) structural and chemical signals using unsupervised machine learning (principal component analysis, PCA). These signals are the principal components of the data matrix. A VARIMAX rotation is performed on the principal components and corresponding (spatial) scores, as per Wilkinson *et al* [**131**] and including the EDS signal as per **Chapter 1** (McAuliffe *et al* [**93**]). This rotation acts to maximise and equalise the variance of the principal components. Positivity can then be enforced on the rotated components, mimicking statistics of experimentally measured EBSPs to obtain physically meaningful results. The procedure generates rotated characteristic components (RCCs), in turn containing rotated characteristic EBSPs (RC-EBSPs) and spectra (RC-spectra). These can be

Alloy	Co	Ni	Mo	Cr	Al	W	Ta	C	B	Zr	Ti
1	~36	~36	2	12	10	2.75	1.25	**0.300**	**0.085**	0.040	0
2	~36	~36	2	12	10	2.75	1.25	**0.296**	**0.043**	0.040	0
3	~36	~36	2	12	10	2.75	1.25	**0.224**	**0.041**	0.040	0
4	~36	~36	2	12	10	2.75	1.25	**0.100**	**0.042**	0.040	0
5	~36	~36	2	12	10	2.75	1.25	**0.100**	**0.020**	0.040	0
6	~36	~36	2	12	10	2.75	1.25	**0.180**	**0.085**	0.040	0
7	~36	~36	2	12	10	2.75	1.25	**0.250**	**0.125**	0.040	0
8	~36	~36	2	12	10	2.75	1.25	**0.250**	**0.200**	0.040	0
9	~36	~36	2	12	10	2.75	1.25	**0.300**	**0.110**	0.040	0
10	~36	~36	2	12	10	2.75	1.25	0.100	0.042	**0.020**	**0.1**
11	~36	~36	2	12	10	2.75	1.25	0.100	0.042	**0.020**	**0.2**
12	~36	~36	2	12	10	2.75	1.25	0.100	0.042	**0.020**	**0.3**
13	~36	~36	2	12	10	2.75	1.25	0.100	0.042	**0.040**	**0.3**
V208C [261]	~36	~34	0	15	10.5	3	1	0.150	0.200	0.040	0
L1	~34	~35	2	13	12	2.75	1.25	0.300	0.110	0.040	0
L2	~34	~35	2	13	12	2.75	1.25	0.300	0.440	0.040	0

Table 2-1: Nominal compositions (in at.%) for the alloys developed in this study, as well as V208C [**261**]. Alloys 1-9 alter C and B content. Alloys 10-13 adjust Zr and Ti composition.

independently indexed and quantified with higher confidence as they have enhanced signal to noise as compared to the raw data. The approach allows reduction of a (for example) 40,000 scan point map, each with 200-by-200 EBSP pixels and 2048 EDS energy bins down to as few RCCs as there are grains (or sub-grains if oversampling is permitted) in an area of interest (AOI). At intermediate magnification this value is usually 50 – 100 in these relatively coarse grained materials. This approach is presented and validated in further detail in **Chapter 1** and is schematically presented in **Figure 2-1**.

Combining structure assignment and chemical measurement permits classifying the phase of a labelled region (usually corresponding to a single precipitate grain). Statistics of chemistry and structure as a function of the other, for example, can then be gathered. In this work, structure is classified *via* the refined template matching (RTM) procedure developed by Foden *et al* [**137**] using dynamically simulated libraries of template EBSPs for each candidate phase, also depicted in **Figure 2-1**. The candidate phases may be selected from observations in previous studies, or *via* computational predictions (e.g. with *ThermoCalc*). The simulated library EBSPs are cross-correlated in Fourier space with the RC-EBSPs. A scan point is assigned to the structure and candidate orientation with the highest cross-correlation peak height. Subsequent iterative refinement determines the precise misorientation of the measured pattern to that of its best matching template. This approach allows us to distinguish phases of similar structure that share many Kikuchi band features, especially since many Radon transform based indexing approaches only consider up to a dozen interplanar angles, which for similarly symmetric structures may be shared. Distinguishing pseudo-FCC (γ / γ′) matrix from MC carbide with conventional EBSD, for example, is challenging.

4. Experimental methods

Sixteen approximately 415 g ingots were fabricated by vacuum arc melting. Compositions are presented in **Table 2-1** and **Figure 2-2**. Each was vacuum homogenised for 48 h at 1250˚C. Ingots were then hot rolled at nominally 1250˚C, with 10-15% reductions from 23 mm square cross-section down to 14 mm. Samples for heat treatment and microstructural characterisation were extracted from the rolled bar. Alloy samples were encapsulated in quartz tube backfilled with Ar for heat treatment; times and temperatures are presented in **Figure 2-3**. The solution stage aims to dissolve all the γ′ precipitated during hot rolling and uncontrolled cooling, in order to 'reset' the alloy and generate a controlled γ′ distribution upon ageing. Three different cooling rates were trialled for a subset of the alloys in order to investigate the effect of intergranular precipitation on boundary serratability, discussed in **Section 2.5.4**. As-received coarse grained RR1000 was also characterised. A standard metallographic polishing procedure was used to prepare samples for microscopy: 400 to 4000 SiC grit grinding, followed by a 1 h H_2O_2 neutralised colloidal silica polish.

SEM, EBSD and EDS were performed on a *Zeiss Gemini Sigma 300* FEGSEM, equipped with *Bruker e⁻Flash^{HD}* and *XFlash 6160* EBSD and EDS detectors respectively. *Bruker DynamicS* was used to dynamically simulate library EBSPs for each candidate structure [**132**], [**247**]. From within the fundamental zone for each phase, a SO(3) sampling frequency of 7˚ was employed for generation of an EBSP template library in the detector reference frame (pattern centre selected with *Bruker Esprit 2.1* from the well-indexed matrix regions and simulated patterns sampled as 200 by 200 px). RC-spectra were quantified with *Bruker Esprit 2.1* using a P/B ZAF correction algorithm accommodating the 70˚ sample tilt required for EBSD.

ThermoCalc was used to model the alloys' predicted phase composition. The TCNi-8 database was used.

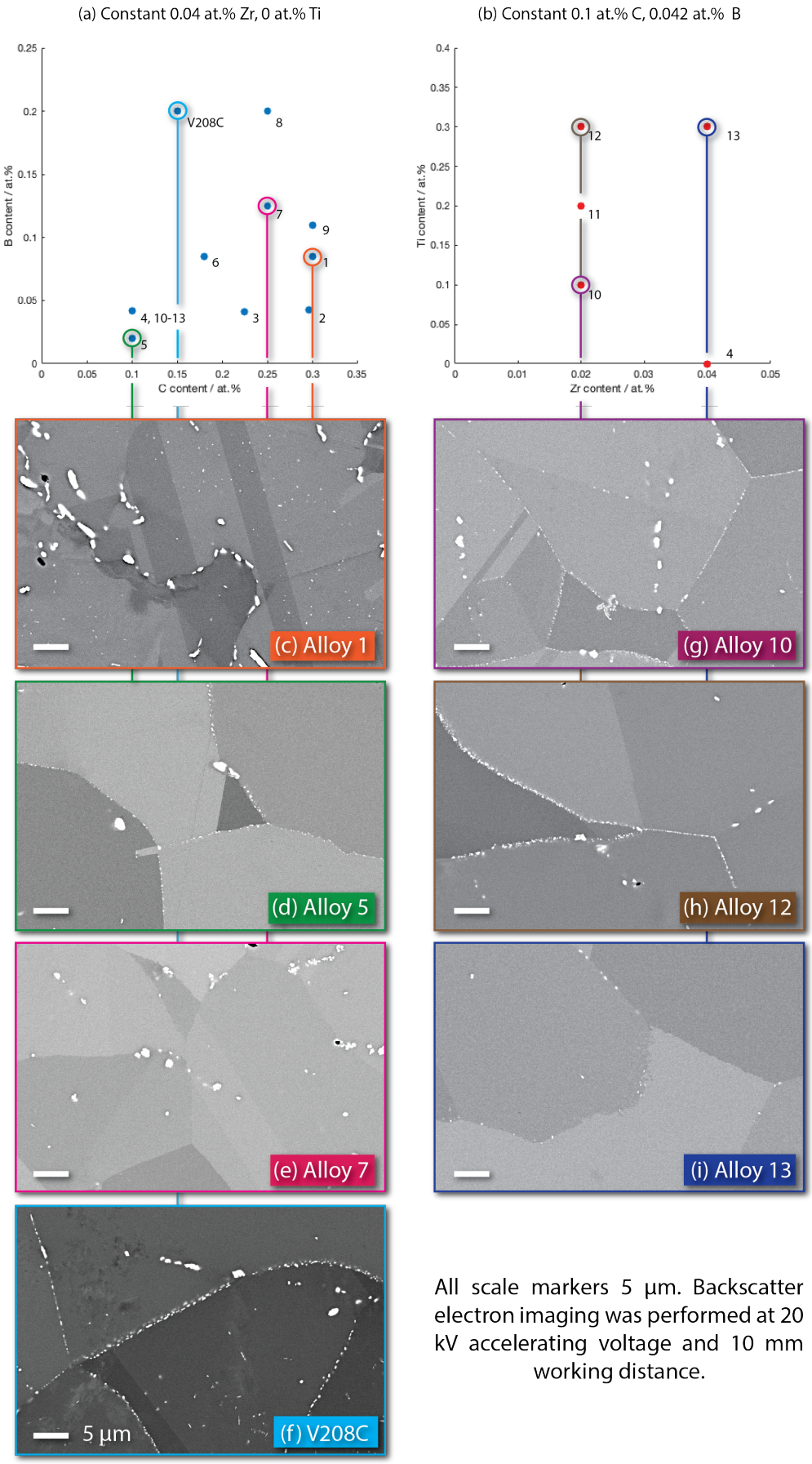

Figure 2-2: Nominal compositions (in at.%) for the alloys developed in this study. Alloys 1-9 and V208C (a) have constant 0.040 at.% Zr and 0 at.% Ti. Alloys 4, 10-13 (b) have constant 0.1 at.% C and 0.042 at.% B. (c)-(h) present representative BSE images of grain boundary morphology for a highlighted sub-selection of the alloys.

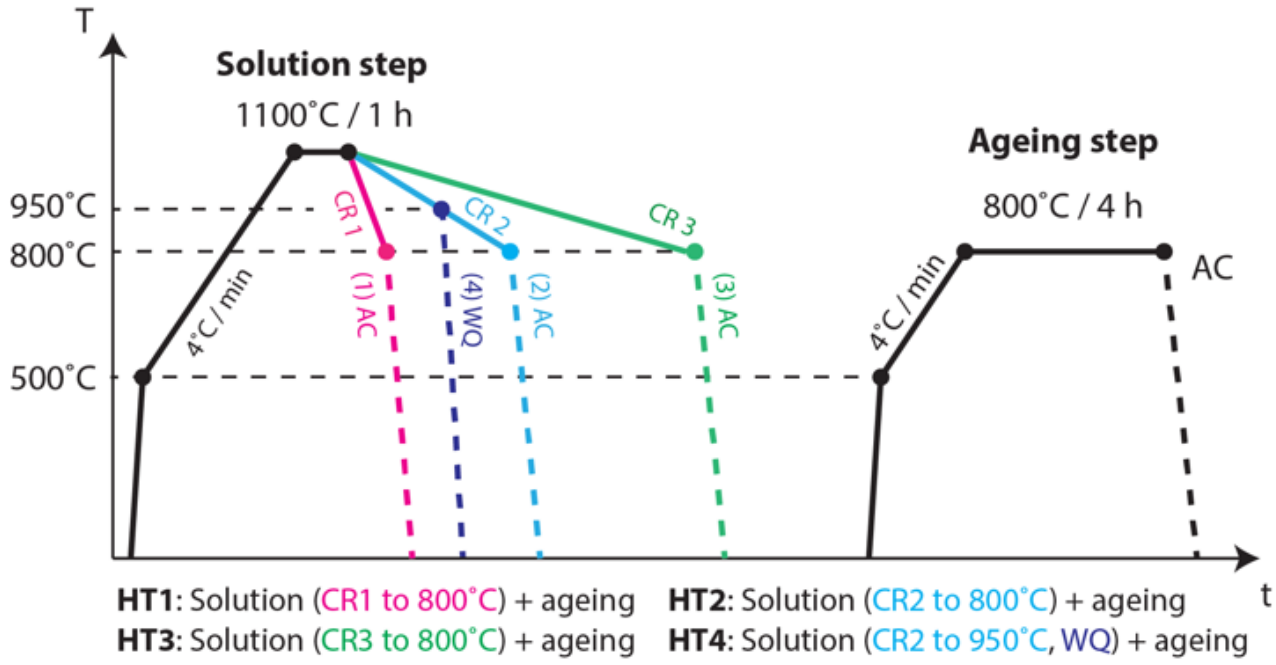

Figure 2-3: Post rolling heat treatments trialled in this study. The solution condition was 1100˚C for 1 h, followed by one of: CR1 (20˚C/min), CR2 (5˚C/min) or CR3 (1˚C/min). From 700˚C alloys were air cooled, or water quenched at 950˚C. All samples were aged at 800˚C for ɣ′ nucleation and growth.

5. Results

5.1 Grain boundary precipitation

Combinatorial PCA-EBSD/EDS was used to classify precipitates that nucleated on grain boundaries. EBSP template libraries for candidate structures were generated with *Bruker DynamicS*: FCC Co/Ni matrix; NiAl and Co_3W intermetallics; M_6C, $M_{23}C_6$ and MC carbides; MB, M_3B_2 and M_5B_3 borides; eta (Ni_3Ti), sigma, mu and P topologically close packed phases. The PCA approach then reduces large scans (often with over 40,000 points) down to representative patterns and spectra, with one corresponding to each grain (or sub-grain depending on the extent of oversampling). These are cross-correlated with template libraries for each candidate structure to identify crystallographic phase. The template with the largest correlogram peak height out of all structures' libraries is identified as the best match. Example RCC, phase assignment, and phase specific chemical distribution radar plots are presented for alloy 7, L2 and RR1000 in **Figure 2-4**. Alloy 7 exhibits precipitation of the MC and M_6C carbides and the M_2B boride. The MC carbide is mainly enriched in Ta and Zr, while the M_6C carbide is W and Mo rich. The M_2B boride is also Mo and W rich, making it difficult to distinguish from the carbide using conventional EDS chemical mapping. This intergranular morphology is typical of alloys 1 -13,. Alloy L2 additionally sees precipitation of the B2 NiAl intermetallic structure on the grain boundaries, which is (unsurprisingly) identified as being enriched in Ni and Al, and relatively depleted in other elements. An RR1000 dataset is also presented, seeing precipitation of the MC carbide and M_3B_2 boride, as thermodynamically simulated by Hardy *et al* [**2**]. The MC carbide sees Ta and Ti segregation, while the boride draws Ta and Mo.

The same analysis is conducted for all of the alloys investigated. The main alloy set (Alloys 1-13) shows varied precipitation of M_6C and M_2B across C, B, Zr and Ti composition space, with no observations of NiAl or any of the other eleven candidate phases. The majority of alloys exhibited MC carbides. Observations of M_6C carbide and M_2B boride precipitation are presented in **Figure 2-5**.

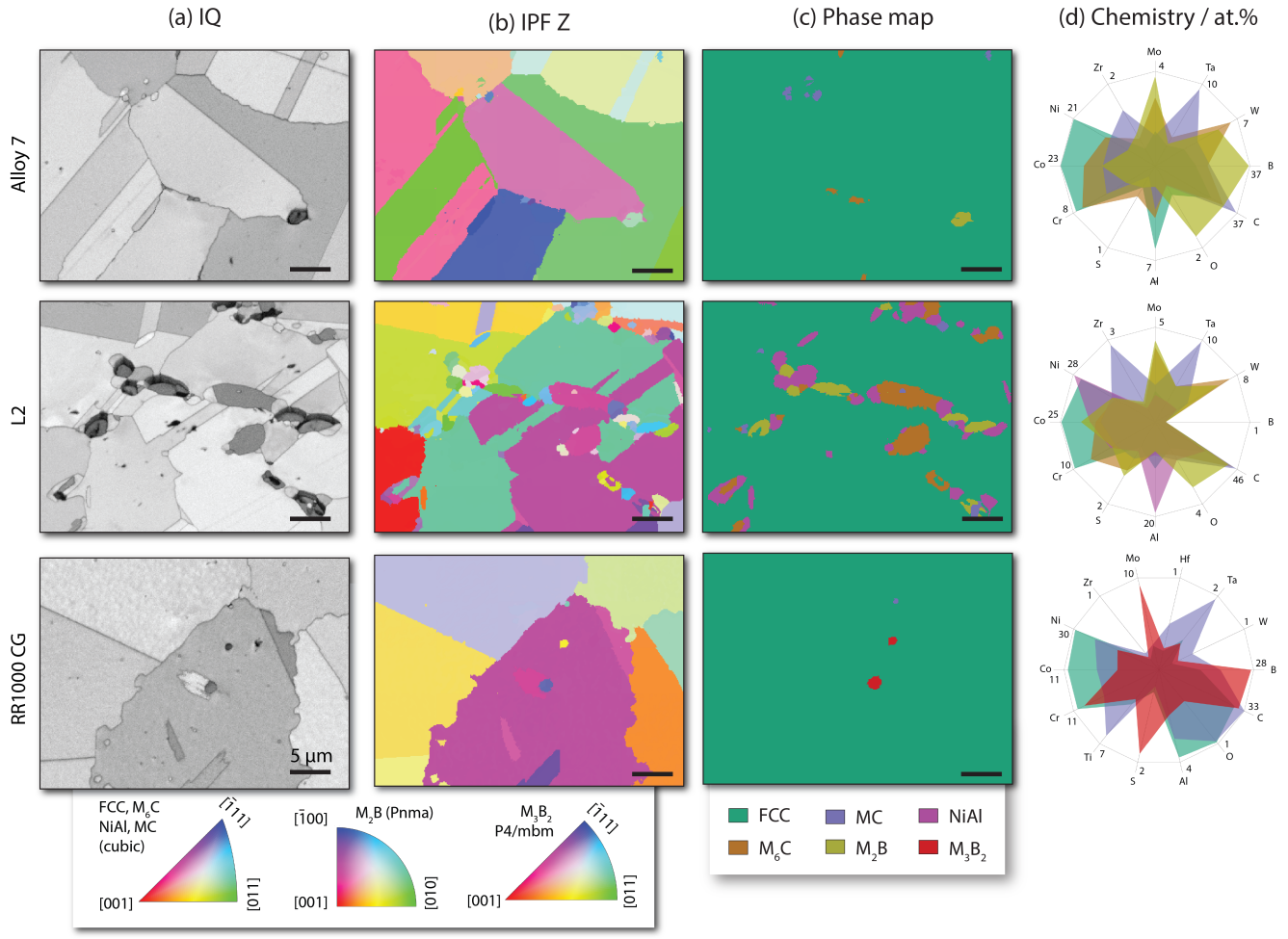

Figure 2-4: Pseudo-backscatter image (a), IPF-Z, Z out of plane (b), phase map (c) and chemical distribution (d) for three selected alloys.

Each RCC we extract from the dataset contains an RC-EBSP and RC-spectrum. Having classified each scan point to one of the candidate structures (using the EBSP), the average EDS spectrum for each scan point assigned to a given phase label, or independently the RC-spectrum itself, can then be quantified. As previously shown in **Chapter 1**, these give very similar results, as the VARIMAX rotation applied to make the principal component EBSPs physical has a similar effect on the EDS spectra [**93**]. Presently the average spectrum for each label is quantified using *Bruker Esprit 2.1*. Subsequently the at% chemistries of each phase is numerically averaged by alloy, and presented in **Figure 2-6** are trends in phase chemistry across the alloy series.

All observations of M_2B boride and M_6C carbide see enrichment in Mo and W. The MC carbide and 'matrix' are generally depleted in these elements. The boride is depleted in Cr and Al, while the M_6C carbide appears to have greater tolerance for these elements.

5.2 Confidence in phase assignment

The utility of the statistical chemistry-by-phase approach is dependent on accurate phase classification. The Fourier space cross-correlation peak height ('RTM peak height') provides a metric for this assignment quality. Higher values imply stronger similarity between test (RC-EBSP) and reference (library) patterns. The template matching process assigns phase and (unrefined) orientation based on the highest RTM peak height across the template libraries. **Figure 2-7** shows average peak height by assigned phase across the alloy series. It is observed that in almost every case the assigned phase is significantly higher than the second closest template structure, and outside of standard errors (error in

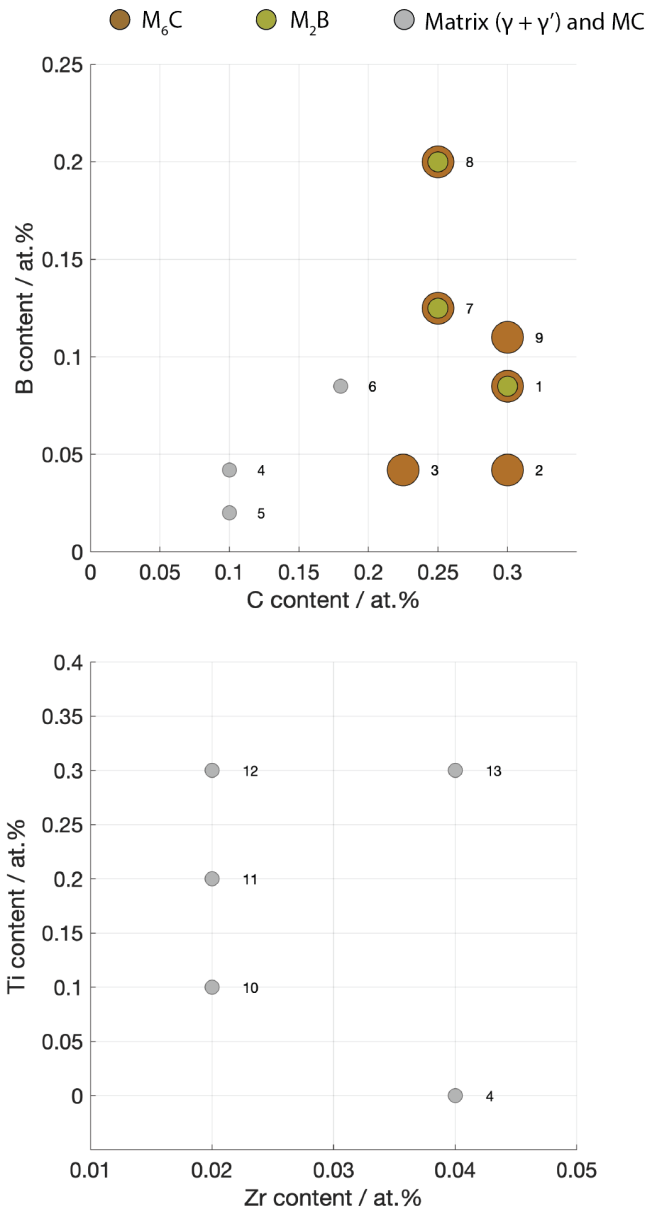

Figure 2-5: Phase diagram of precipitate observations across investigated composition space for alloys 1- 13.

the mean value of RTM peak height for all labels in an area of interest). Whenever a matrix label is assigned, the runner-up most prevalent phase is consistently the MC carbide. When an M_6C label is assigned, the runner-up template is always the $M_{23}C_6$ structure. To further characterise the accuracy of the phase assignment and demonstrate the viability of the RTM procedure, presented in **Figure 2-8** are examples of a matrix and M_6C assigned RC-EBSPs. Best, second-best and poorly matching template patterns, with corresponding RTM peak heights, demonstrate the separability of the assigned phase from the alternatives.

MC carbide templates are always runner-up to matrix assignments due to strong similarity in crystal structure Conventional Hough (Radon) transform EBSP indexing does not account for intra-pattern intensity variation, or presence of minor Kikuchi bands, especially in strongly diffracting crystals which often satisfy the common maximum number to consider for the interplanar angle lookup (often 12). The Fourier cross-correlation handles this well [**137**], and is able to distinguish the FCC label pattern in **Figure 2-8** from the very similar (in major band trace) MC carbide, with an approximately 32% difference in peak height. Distinguishing the $M_{23}C_6$ and M_6C carbides presents a similar case. There is similarity in the crystal structures leading to systematic relative proximity in peak height. Conventional

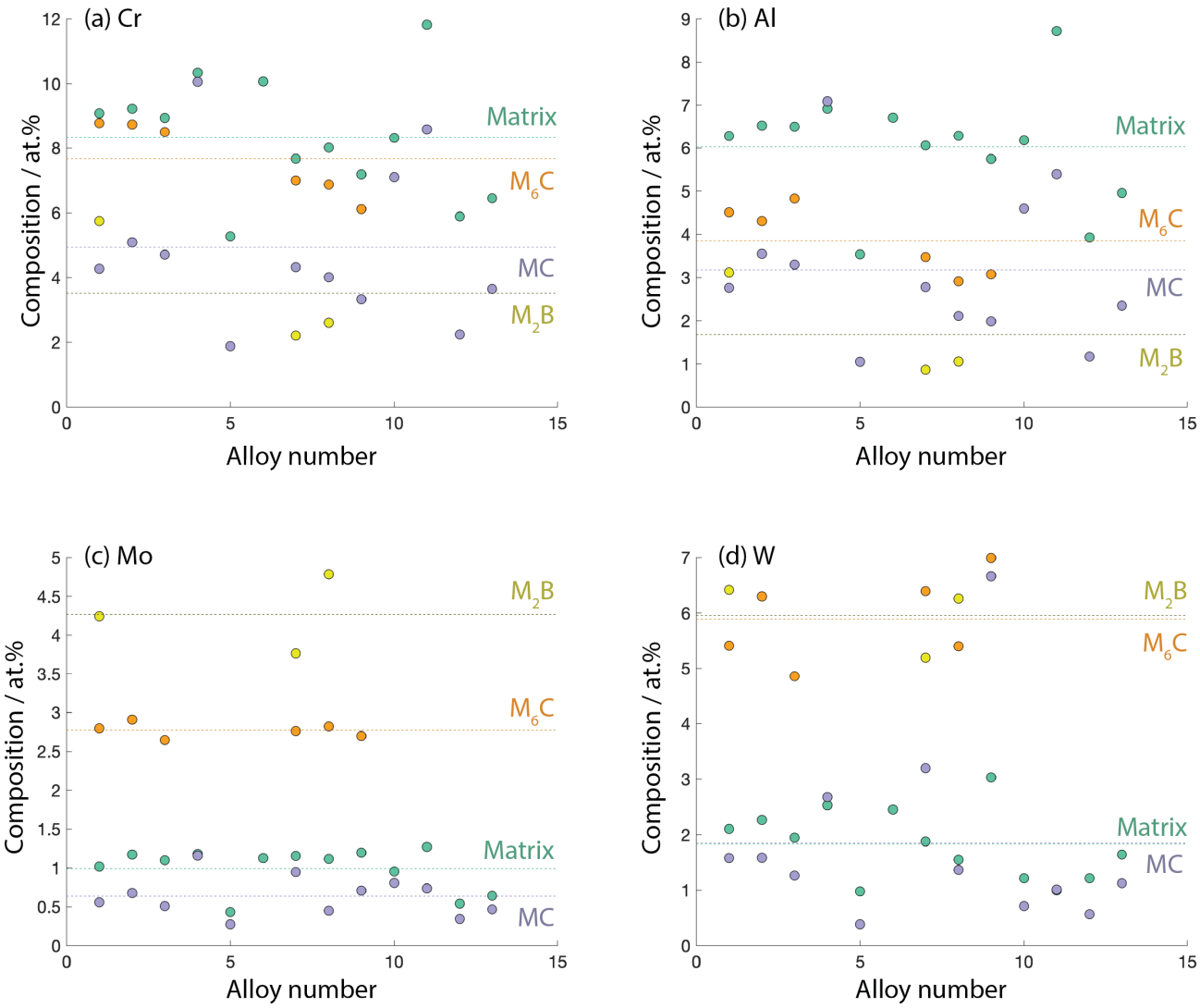

Figure 2-6: Average composition of the M_2B boride, M_6C carbide, ɣ + ɣ′ 'matrix' and the MC carbide across the main alloy series. Cr (c), Al (b), Mo (c) and W (d) are plotted here.

indexing of this area of interest does not robustly distinguish $M_{23}C_6$ from M_6C structure, and completely misses MC carbides. This issue together with the previously discussed similarity in EDS measured chemistry between M_6C carbide and M_2B boride demonstrates the advantages of a combined approach with independent and accurate quantification / indexing.

5.3 Comparison to CALPHAD thermodynamic modelling

Using ThermoCalc with the TCNI8 database, initially with no structures suspended, predicted phase fractions were calculated for each alloy in the main set. M_2B borides, $M_{23}C_6$ carbides, Ni_3Ti and TCP mu were predicted (along with an FCC matrix and various populations of L12 ɣ′, not plotted). Phase proportions as a function of temperature for alloy 7, whose characterisation is highlighted in **Figure 2-4**, are presented in **Figure 2-9**.

There is a discrepancy between phases predicted by ThermoCalc: M_2B, $M_{23}C_6$, mu and eta (Ni_3Ti), and those we observe (MC, M_6C, M_2B). The unobserved phases $M_{23}C_6$, mu and eta were included in the template matching procedure; their template libraries were dynamically simulated and cross-correlated with the label RC-EBSPs but were never matched.

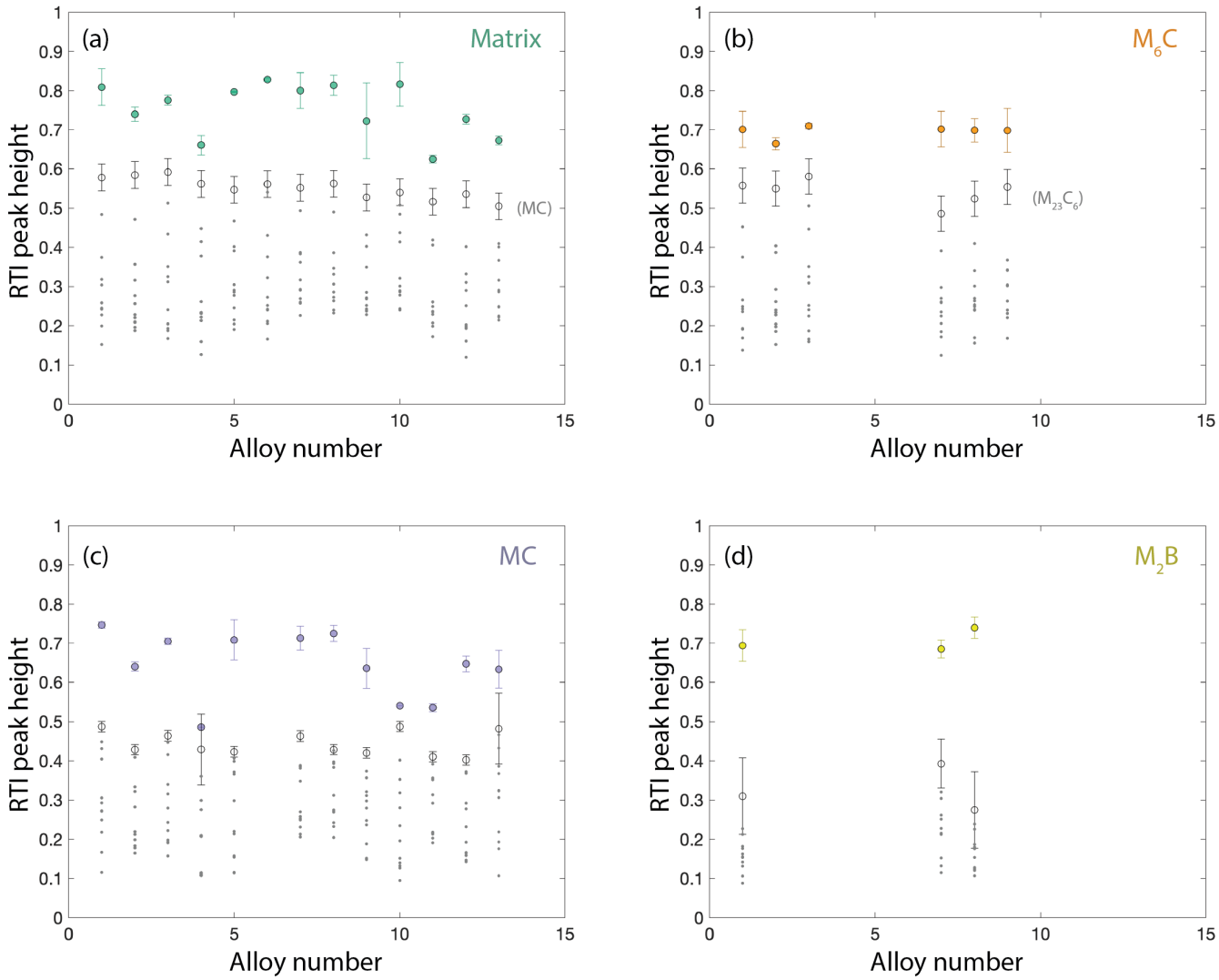

Figure 2-7: Average RTM peak heights (cross-correlation quality) for labels assigned to each of the identified phases, across the main alloy series, with standard-error errorbars. (a) presents the pseudo-FCC 'matrix', (b) the M_6C carbide, (c) the MC carbide, and (d) the M_2B boride. In each case the second best matching phase is also plotted with standard-error errorbars. This is consistently an MC template for matrix labels, and an $M_{23}C_6$ template for M_6C labels. There is no consistent runner-up for the MC carbide or M_2B boride.

As discussed in **Section 2.4.2** the RTM phase identification procedure is robust, so we have confidence precipitates have not been mis-identified (for example the $M_{23}C_6$ RTI peak height is consistently lower than that of M_6C for carbide assignment, and there are observable Kikuchi band discrepancies, see **Figure 2-7** and **Figure 2-8**).

In **Figure 2-10a** and **b** *ThermoCalc* predicted solvus temperatures are presented for the M_2B boride, this allows inference of how these precipitates are stabilised by variations in C, B, Zr and Ti across the alloy set. Also included are the solvuses for the unobserved $M_{23}C_6$ carbide, **Figure 2-10c** and **d**. From **Figures 10a** and **b** it is observed that the boride is destabilised (the solvus temperature is lowered) by additions of C, Zr and Ti. Additions of B raise its solvus temperature. A different trend is observed for $M_{23}C_6$. Additions of Ti reduce the carbide's solvus temperature, while Zr appears to raise it (it is not predicted to precipitate at all for alloys with low Zr content). Additions of B do not have a pronounced effect on the predicted stability of $M_{23}C_6$. Comparison to the phase diagram of precipitate observations, **Figure 2-5**, partially agrees with the *ThermoCalc* predictions. In the Zr and Ti enriched alloys the boride is not observed; it is only observed after about 0.085 at.% B addition.

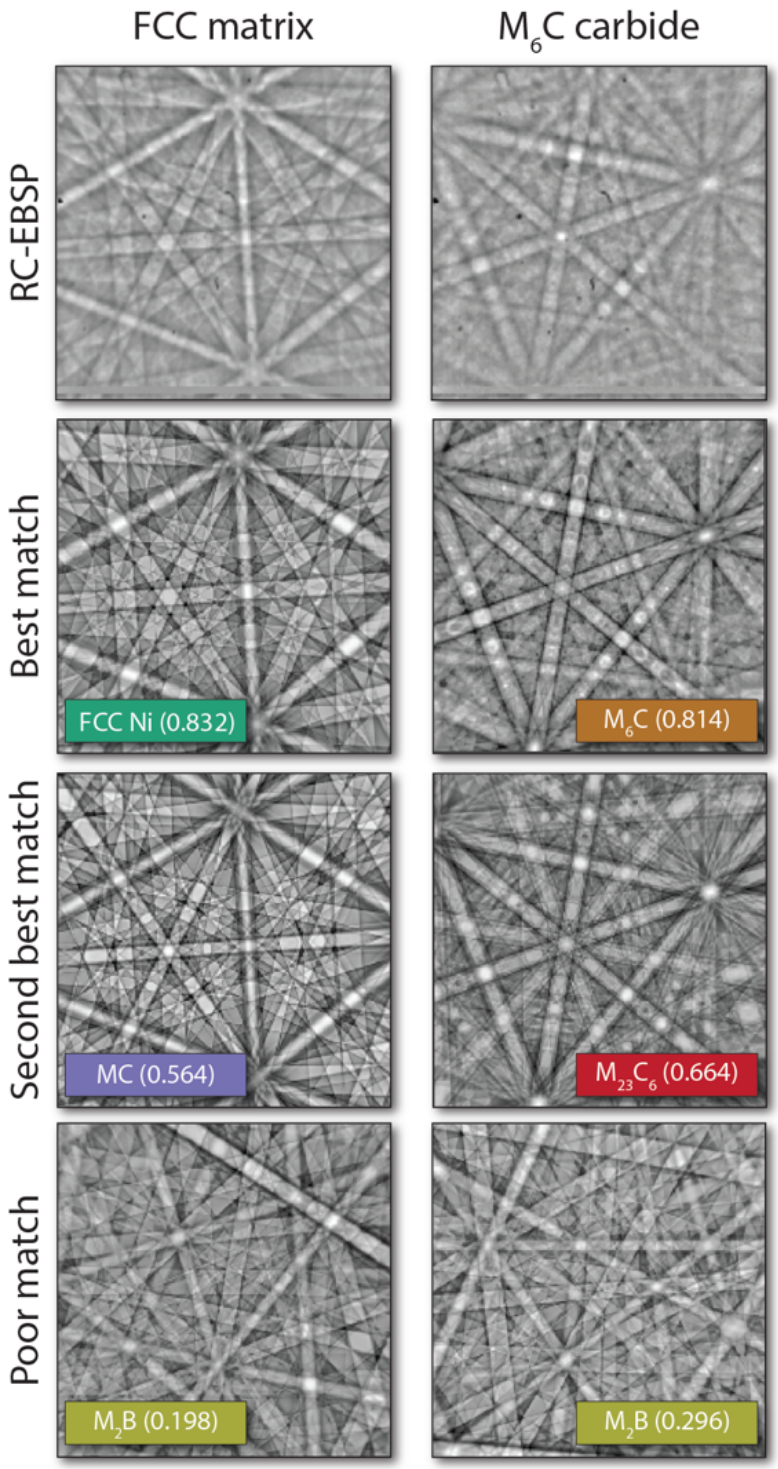

Figure 2-8: RC-EBSPs and the best, second-best and poorly matched simulated template EBSPs. For the FCC 'matrix', the best match (FCC) has an RTM peak height of 0.832. The second closest match, the MC carbide, has a peak height of 0.564 and the 'correct' assignment has a 32% advantage.

5.4 Grain boundary serration

Heat treating to produce serrated grain boundaries is known to improve superalloy mechanical performance, especially in deformation regimes where boundary sliding mechanisms are believed to be the principal contributor to strain accumulation [**84**], [**86**], [**88**], [**89**].

In addition to the standard HT1 employed before characterisation, three further heat treatments were applied to a selection of the alloys in order to promote grain boundary serration. HT2 involved a slow cool through the ɣ′ solvus of

5˚C / min followed by an air cool from 800˚C. HT3 employed an even slower cooling rate of 1˚C / min, followed by the same air cool from 800˚C. HT4 saw the same 5˚C / min cool from solution, but was followed by a water quench at 950˚C. All alloys were subsequently aged at 800˚C for 4 h to produce an optimal ɣ′ distribution. Examination of the heat treated microstructures showed that serrations were possible in this alloy series, albeit of a smaller amplitude than what is often observed, but are completely inhibited by grain boundary coverage of even the finest precipitates. In the scheme of Koul & Gessinger [**69**], the presence of fine boundary particles pins the boundaries, such that upon ɣ′ migration there is sufficient line tension to pull back the boundary and prevent extensive serration. In alloy 7, **Figure 2-11e** and **i**, precipitates are sparse (though with greater volume), and serration is readily achievable on many boundaries.

6. Discussion

Across an even sampling of C, B, Zr, and (low) Ti space, thirteen alloys were processed (as well as the additional L1, L2, V208C, and RR1000) and the distribution and chemistry of phases within their microstructures were characterised with a combinatorial EBSD/EDS approach. In this alloy series there is precipitation of intergranular M_6C and MC carbides, and M_2B borides. In alloys with a lower Cr:Al ratio there is extensive precipitation of the NiAl intermetallic. The observed structures are relatively exotic, with the majority of commercial superalloys exhibiting $M_{23}C_6$ carbides as

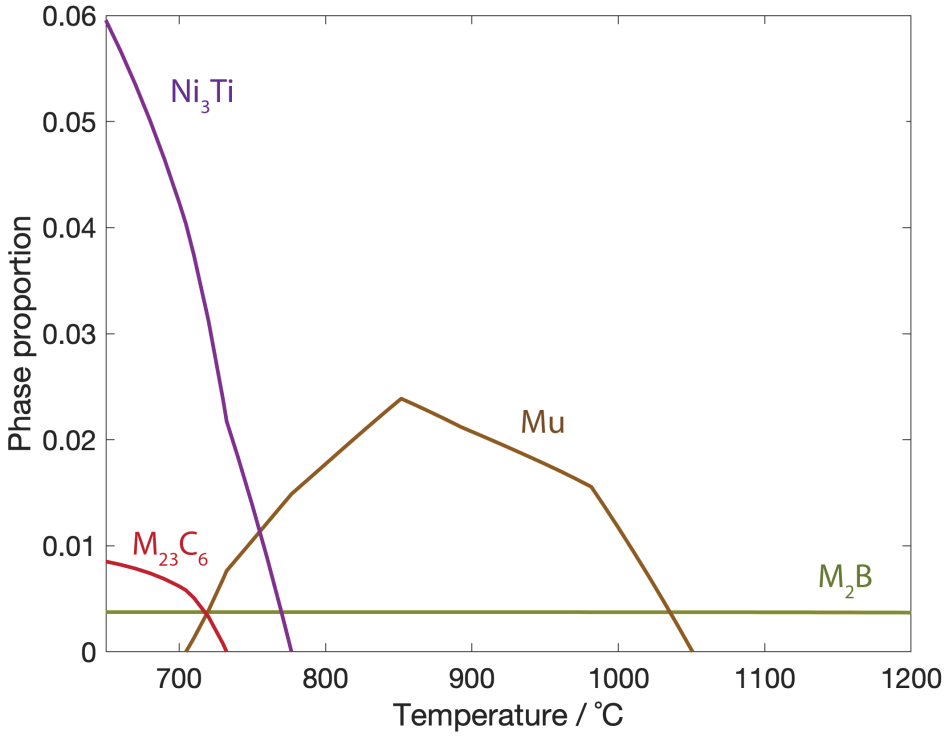

Figure 2-9: Phase proportions of thermodynamically predicted phases (matrix and ɣ′ distributions not plotted).

well as M_5B_3 and M_3B_2 borides. These structures were all included in the phase-ID stage of our method: evenly SO(3) sampled EBSPs were dynamically simulated for them and compared to the extracted representative patterns. The greater Mo and W content of our alloys might favour the M_6C carbide and M_2B boride as opposed to $M_{23}C_6$, M_5B_3, M_3B_2, etc. These structures consistently exhibit strong measured enrichment in these two elements. The consensus from the literature is that $M_{23}C_6$ generally has a greater affinity for Cr than M_6C [**65**], [**96**]. Significant enrichment of Cr in M_6C relative to the matrix has not been measured. Similarly, the M_5B_3 and M_3B_2 borides have been observed to exhibit greater affinity for Cr than what is seen for M_2B [**64**], [**90**], [**102**]. This likely has significant implications on oxide scale formation and possible ɣ′ depletion during high temperature (atmosphere exposed) deformation. Mo and W are not known to be beneficial to stable oxide formation, so precipitates rich in these elements may not be optimal for significant grain boundary coverage. Additionally, Mo and W are known to be slow diffusing elements, which may be an additional factor in the precipitates' ability to inhibit grain boundary serration that is observed.

Our experimental observations are compared to thermodynamic simulations, revealing a stark difference. *ThermoCalc* predicts precipitation of $M_{23}C_6$, M_2B, TCP mu and eta (Ni_3Ti). The M_2B boride is observed in this system, but $M_{23}C_6$, mu or eta were not once observed. These phase structures were included in the EBSP template matching. It may be that

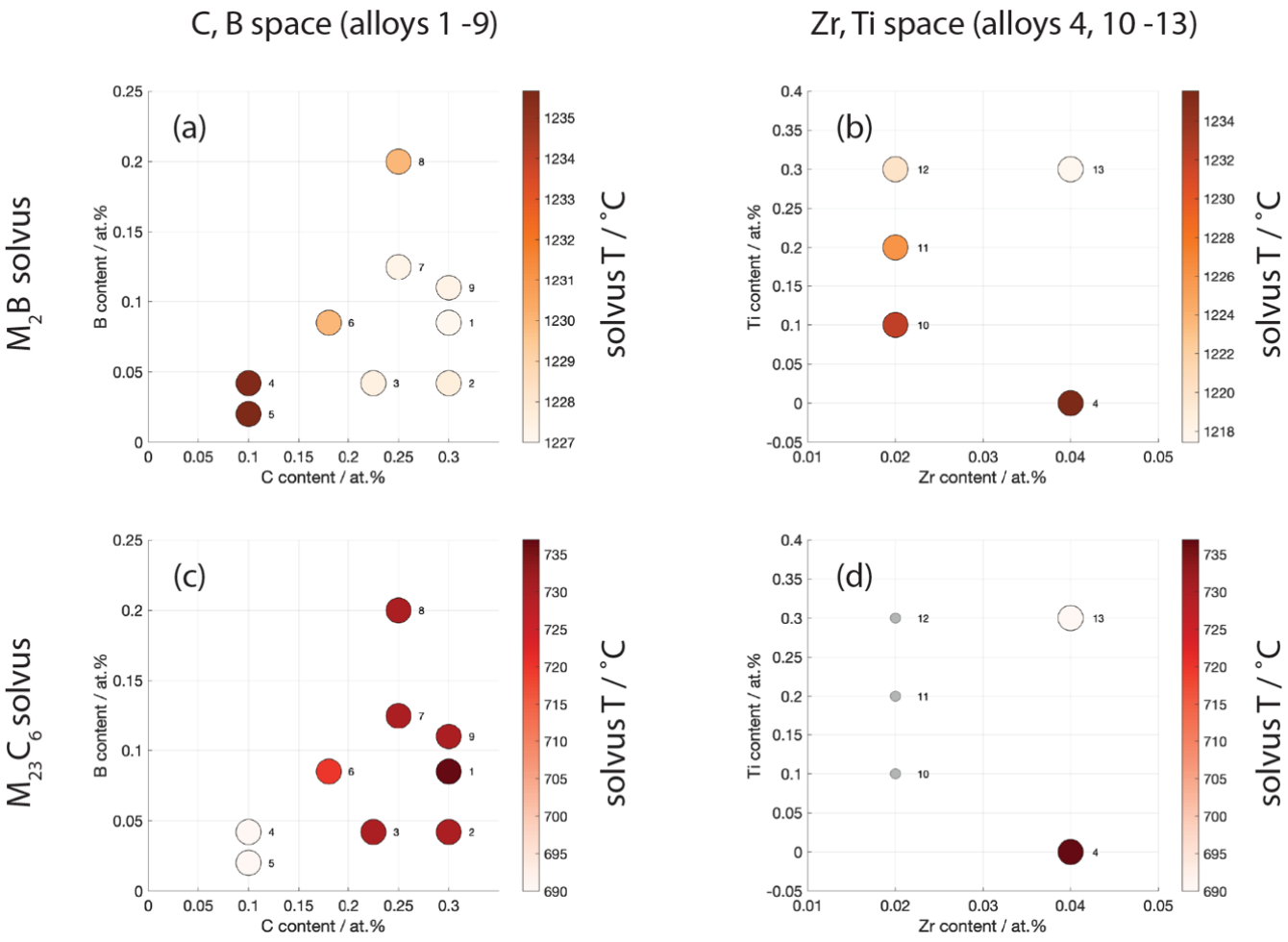

Figure 2-10: Modelled solvus temperatures for the (observed) M_2B boride and (not observed) $M_{23}C_6$ carbide.

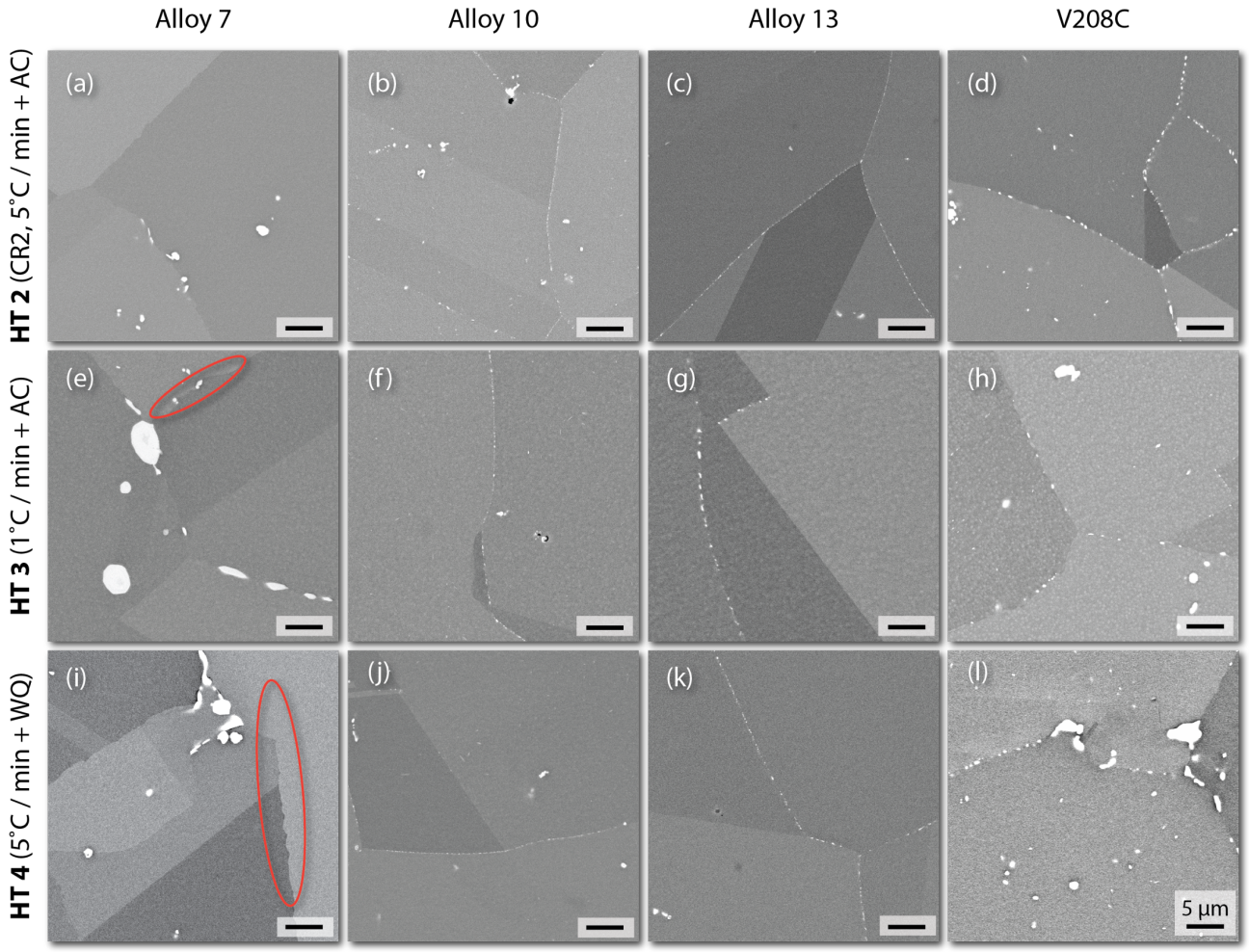

Figure 2-11: Grain boundary observations for four alloys and three heat treatments (varied cooling rate, see Fig. 3). Serrations are highlighted in orange circles. They are only observed when boundary coverage of precipitates is not dense. Backscatter imaging, 20 kV, 10 mm working distance.

there are insufficient accurate observations of these phases in the database we used (TCNi8) for *ThermoCalc* to successfully interpolate to our composition and accurately predict stability. A potentially useful conclusion we draw from the simulations is that M_2B precipitation appears to be inhibited by C, Zr and Ti, while C, B and Zr additions raise the solvus (stabilising) the $M_{23}C_6$ carbide.

This work demonstrates the utility of the PCA-EBSD/EDS method in extracting representative EBSPs and EDS spectra from combined and simultaneous datasets for the purposes of phase classification and observing trends in chemistry. Included within the spatial maps is information on the location, size and shape of each phase. In the present work these observations are not quantitatively analysed, but we note that this is likely important for the relative contributions to high temperature creep performance of these engineering alloys.

This study used a high SEM voltage and a coarse step size (100 nm) which limits our ability to classify ultra-fine precipitates that are occasionally observed. Adjustment to the sampling geometry, or use of alternative techniques such as TEM or transmission Kikuchi diffraction (TKD), may be required for greater spatial resolution. A combinatorial PCA approach, using VARIMAX, could be used for a TEM diffraction and EDS experiment, but care should especially be taken when considering a VARIMAX rotation for TEM spot diffraction classification as the experimental intra-pattern variance may not necessarily be equal across different structures. Furthermore, even at the relatively loose variance tolerance limit we have used in this work, if a unique signal is only coming from one or two scan points the PCA

approach may not see sufficient inter-pattern variance to extract a unique component. We have previously discussed this effect and other PCA artefacts in **Chapter 1**.

7. Conclusions

From the present study the following conclusions may be drawn:

1. Thirteen alloys across C, B, Zr and (low) Ti composition space have been characterised, as well as two alloys with lower Cr:Al ratio, V208C, and RR1000, to investigate intergranular precipitation. The main alloy series exhibits M_6C and (intragranular) MC carbide precipitation, as well as M_2B borides. A NiAl phase is observed in the reduced Cr:Al alloys, but only MC and M_3B_2 in RR1000.
2. Interrogation of the chemistries of the grain boundary precipitates in the main alloy reveal them all to be enriched in Mo and W. M_6C draws less Mo than M_2B, and appears to see a greater solubility for other elements. The NiAl phase is rich in Ni and Al.
3. Thermodynamic simulations inaccurately predict the precipitation we experimentally observe. $M_{23}C_6$, TCP mu and eta are all predicted but never observed. The modelling predicts the M_2B boride that is observed, and that it is destabilised by additions of C, Zr and Ti. This agrees with the lack of boride observation in Zr and Ti enriched alloys.
4. Grain boundary serrations were only achievable at slow cooling rates where grain boundaries did not see spatially dense coverage of M_6C or M_2B.

8. Chapter Acknowledgements

This work was a collaborative effort, involving Dr Ioannis Bantounas, Dr Lucy Reynolds, and Alex Foden, in addition to TPM, TBB, MH, and DD.

TPM processed and characterised the presented alloys and prepared the manuscript. Heat treatments were developed by IB, TPM, DD and MH. IB, LRR, TPM, MH, and DD discussed the project direction and selected compositions for investigation. TPM and TBB developed the PCA-EBSD/EDS approach. TBB and AF developed the EBSP refined template indexing procedure. DD and TBB supervised the work.

At Imperial College Thomas Kwok and Xin Xu are acknowledged for assistance in hot rolling, as well as Dafni Daskalaki-Mountanou, Chris Bilsland and Thibaut Dessolier for helpful conversations and participating in developing the PCA approach and characterisations of carbides. TPM, IB, TBB and DD would like to acknowledge support from the Rolls-Royce plc - EPSRC Strategic Partnership in Structural Metallic Systems for Gas Turbines (EP/M005607/1). TPM, TBB and DD acknowledge the Centre for Doctoral Training in Advanced Characterisation of Materials (EP/L015277/1) at Imperial College London. TBB would like to acknowledge the Royal Academy of Engineering for funding his research fellowship. DD acknowledges funding from the Royal Society for his industry fellowship with Rolls-Royce plc.

Chapter 3 - Spherical-angular dark field imaging and sensitive microstructural phase clustering with unsupervised machine learning

1. Chapter Summary

Electron backscatter diffraction is a widely used technique for nano- to micro-scale analysis of crystal structure and orientation. Backscatter patterns produced by an alloy solid solution matrix and its ordered superlattice exhibit only extremely subtle differences, due to the inelastic scattering that precedes coherent diffraction. This chapter shows that unsupervised machine learning (with PCA, NMF, and an autoencoder neural network) is well suited to fine feature extraction and superlattice/matrix classification. Remapping cluster average patterns onto the diffraction sphere lets us compare Kikuchi band profiles to dynamical simulations, confirm the superlattice stoichiometry, and facilitate virtual imaging with a spherical solid angle aperture.

2. Introduction

Electron backscatter diffraction (EBSD) is a common method for analysis of crystal structure and orientation in engineering materials. Typically, many thousands of electron backscatter patterns (EBSPs) are produced in a single scan of an area of interest (AOI), and there is strong motivation to develop understanding by taking advantage of the wealth of information contained within each pattern. Unsupervised machine learning allows exploration of the structure of EBSD datasets and idenficiation of latent features. In this chapter segmentations performed with principal component analysis (PCA), non-negative matrix factorisation (NMF), and an autoencoder neural network are compared. Performing post-segmentation analysis ‘on the sphere’ using a spherical harmonics approach allows comparison of Kikuchi band profiles for different latent patterns and class averages to dynamical simulations. It is found that the {100} and {111} crystallographic planes exhibit greater superlattice Kikuchi diffraction contrast than {110} and {131}. There is significant difference between simulated diffraction profiles of CoNi-Co_3(Al,W) and Ni-Ni_3Al ɣ - ɣ′ pairings, and it is confirmed that the V208C Co/Ni-base superalloy matches the former. These data driven techniques enable resolution and quantitative analysis of of ɣ - ɣ′ structures in the scanning electron microscope.

Superalloys enable modern engineering systems such as the high temperature gas turbine engine. To develop these alloys further, new tools are required to routinely analyse atomic scale ordering in microstructures. This presents a challenge at the micro-scale using EBSD, as existing analysis methods do not enable the separation of a parent crystal from its (ordered) superlattice. Often, more complicated and expensive TEM analysis using dark field imaging must be employed.

In a superalloy ɣ - ɣ′ system the ɣ matrix exhibits a face-centred cubic (FCC) structure, and ɣ′ superlattice precipitates with a primitive $L1_2$ structure form intragranularly below a solvus of usually 1000-1200˚C. The symmetry of the FCC matrix is such that elastic (Bragg) diffraction is not observed as coming from crystallographic planes with mixed-odd-and-even Miller indices. This is not the case for ɣ′, and so-called ‘superlattice spots’ are observed in transmission electron microscope (TEM) diffraction. These are widely used to differentiate matrix from precipitate in dark field imaging. Furthermore, dark field imaging can be used to image dislocations, as the Burgers vector of a dislocation locally transforms the crystal, systematically altering the Bragg condition, and creating strong contrast for known foil normal / Burgers vector combinations. These methods have revolutionised materials science, especially in the characterisation of industrially relevant alloy deformation mechanisms, and have lead to vast improvements in creep

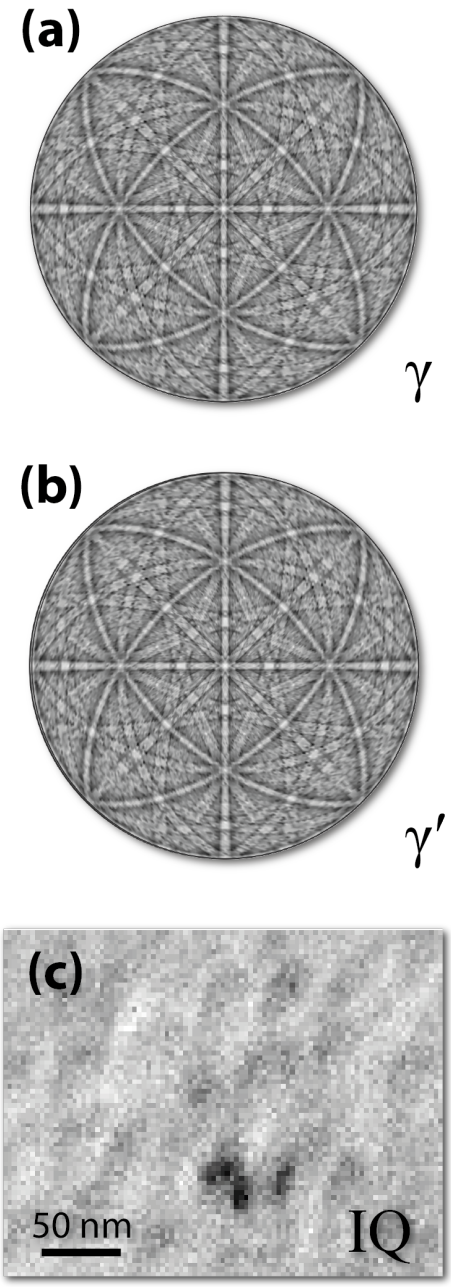

Figure 3-1: Similarity of EBSPs from diffracting ɣ and ɣ′. (a) and (b) show the (identical to the eye) upper halves of simulated diffraction spheres, and (c) presents a typical pattern quality map from an EBSD scan (Hough-indexed with *Bruker eSprit 2.1*).

and fatigue life of (for example) aerospace gas turbines [**30**], [**262**], [**263**]. However, transmission methods require thin foils, limiting the area that can be routinely studied. This motivates development of new approaches.

EBSD is performed in the SEM on the surface of well-polished 'bulk' materials. The method involves serial capture of 2D wide angle diffraction patterns, which contain rich structural information as they are produced from an incident electron beam scattering, diffracting, and escaping from the sample. A map can be formed when the beam is scanned in a controlled manner across the surface of the sample, and large areas at a range of step sizes can be interrogated easily.

In the superalloys, existing EBSD analysis methods index the ɣ′ and ɣ as the same (usually ɣ) phase, as unfortunately Kikuchi diffraction and the formation of EBSPs derives from extensive inelastic electron diffraction prior to elastic scattering and the formation of Kossel cones [**109**], [**114**], [**123**]. This means that the differences in electron scattering

behaviour of ɣ′ and ɣ is extremely subtle and difficult to detect [**107**], [**112**]. Diffraction patterns for the two phases, calculated using dynamical diffraction theory, are presented in **Figure 3-1**. Global EBSP simulations of ɣ and ɣ′ are performed to highlight their similarity, as well as a typical band image quality (IQ) image. The IQ map corresponds to average peak height of the Hough transformed EBSP, and is often used as a pseudo-backscatter mode in conventional EBSD-based imaging.

In this work, new, and now accessible, machine learning (ML) based-approaches are used to tackle this challenging segmentation problem in a Co/Ni-base superalloy.

PCA is a powerful tool in identifying (scan) points of self-similarity and can be used to reduce EBSD data in an observation-by-variable matrix ('data matrix') down to a handful of high quality patterns. These can then be efficiently indexed [**130**], [**131**]. Using PCA, **Chapters 1** and **2** have previously combined structural information from EBSD with chemical fingerprints from simultaneous energy-dispersive X-ray spectroscopy (EDS) to characterise the phase of a microstructural constituent and separate carbide types [**93**], [**264**].

NMF has seen extensive use in analysis of maps made up of (usually 1D) X-ray spectra, as the strictly positive nature of signal 'hit' counting provides an excellent boundary condition for decomposing a data matrix into its latent factors [**182**], [**236**], [**265**].

Autoencoder neural networks have not seen significant application in electron microscopy, and here the applicability is evaluated [**198**], [**211**], [**266**]. They come in many flavours: deep (many-layered) or shallow, linear or convolutional. All are based on the idea of training the identify function, with information channelled through a 'bottleneck' layer containing only a few nodes. This forces the network to learn the structure of the data, with bottleneck activations representing the 'firing' of latent features. A shallow (single hidden layer) autoencoder, employed in this work, can be thought of as performing a non-linear matrix decomposition. This is regularised with the L2 norm of the weights and a Kullback-Leibler divergence penalty to the hidden layer latents. Deep (convolutional) neural networks are starting to see deployment to supervised EBSD classification problems [**242**], but there has been little discussion of what crystallographic features neural networks are capable of learning. This is a secondary motivation for the present study.

These data science-based approaches have been performed through clustering of the captured diffraction patterns (which are 2D images captured in direct space). To provide evidence of their success, and reveal more about the microstructure of a Co/Ni-base superalloy, clustered patterns are analysed 'on the sphere'. This analysis is used to drive development of an EBSD-focussed angular resolved segmentation, and presented here is 'spherical-angular dark field imaging'.

3. Methods

3.1 Experimental

A sample of V208C Co/Ni-base superalloy, as developed by Knop *et al* [**1**], [**261**], was fabricated by vacuum arc melting and casting, then homogenised and hot rolled at 1200˚C. It was solution heat treated at 1100˚C, and aged at 800˚C for 4 h. The sample was prepared by standard metallographic grinding and polishing. EBSD was performed on a *Zeiss Sigma300* FEGSEM, equipped with Bruker e$^-$Flash^{HD} detector at 20 kV with an 120 µm aperture and high current mode, resulting in a probe current of ~10 nA. The sample was tilted to nominally 70˚. For the map analysed in this work, a step size of 30 nm was used in scanning of a 2.85 µm by 2.16 µm AOI, with patterns captured at a resolution of 600-by-800 pixels. EBSPs were simulated using *Bruker DynamicS* [**132**], [**247**], using a cut-off diffraction condition intensity

of 5.0% of the maximally scattering reflector, and minimum plane spacing of 0.1 Å. These simulations were performed for CoNi (FCC solid solution), $Co_3(Al,W)$ ($L1_2$ ordered), Ni (FCC), and Ni_3Al ($L1_2$ ordered), structures.

3.2 Data preparation

To eliminate systematic error from the analysis we normalise the EBSPs. Usually this entails subtracting the mean pixel value from each pattern and dividing by the standard deviation, which is the operation we perform for our PCA and autoencoder analysis. For NMF, as all data must be positive, we subtract the minimum value from each observation (EBSP) and divide by its standard deviation. Consequently there is an inhomogeneity in the mean value of NMF-normalised data, which is known to manifest in one of the calculated non-negative factors [**182**].

Without data adjustment the presented methods are intrinsically non-local. This means that the EBSPs could be unpacked and factorised in any order and there would be no difference to the calculated principal components (PCs), non-negative factors or autoencoder weights and biases. This has advantages, but when identifying very fine differences in latent variables it is useful to leverage localisation as additional input information to our problem. As discussed in **Chapter 1** this brings the PCA analysis closer to the NLPAR approach set forth by Brewick *et al* [**129**]. In the present work, a spatial weighting kernel is introduced to the pipeline, which clusters patterns from local spatial neighbourhoods. This weighting kernel is applied directly to the data, and the maps can be used with the non-local algorithms as previously introduced. For the autoencoder a convolutional layer could have been introduced to achieve a similar effect.

The spatial weighting kernel is described based upon the work of Guo *et al* [**267**], where for each EBSP P_{ij} in the i^{th} row and j^{th} column of the AOI, the locality-corrected pattern, P_{ij}^{local}, is introduced as:

$$\mathrm{P_{ij}^{local}(r)} = \sum_{k=1}^{n}\sum_{l=1}^{m} \eta\, \mathrm{P_{kl}} \qquad 3\text{-}1$$

With kernel value:

$$\eta = \left(1 - \frac{d^2}{r}\right)^4 \; for\; d < r, else\; 0 \qquad 3\text{-}2$$

Where:

$$d = \sqrt{(i-k)^2 + (j-l)^2} \qquad 3\text{-}3$$

This algorithm imposes a Gaussian-like kernel onto the AOI: patterns are averaged with weights decaying on their square separation within a kernel of consideration. Alternative kernel functions are of course possible, but it is found that this one is functional and useful. In this work a kernel size *r* is employed, equal to three steps within the map.

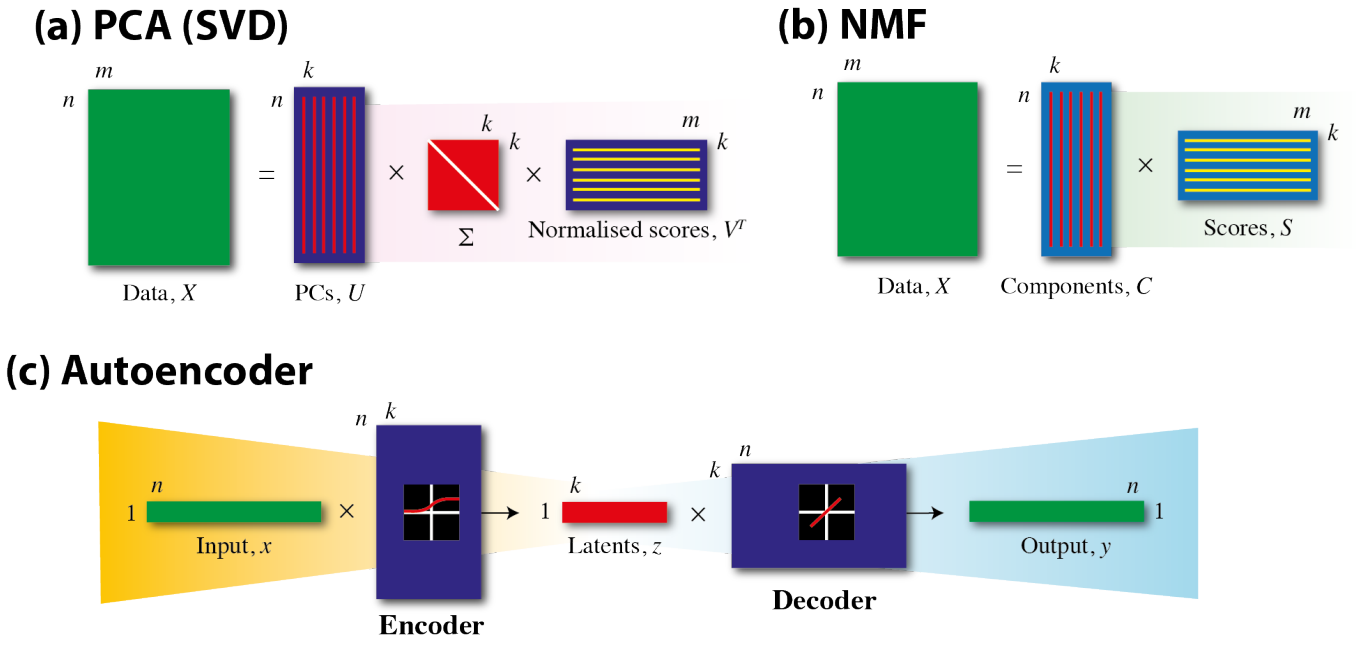

Figure 3-2: Visualisations of the matrix algebra involved in PCA (a), NMF (b), and the autoencoder (c).

3.3 Algorithms for learning latent factors

The methods employed in this chapter have previously been discussed in **Section LR.4.2**. Briefly, PCA acts to extract the dataset directions that maximise contribution to dataset variance in orthogonal (independently uncorrelated) directions. This is equivalent to the SVD of the data matrix, as presented in **Figure 3-2a**. NMF similarly finds a dataset decomposition, **Figure 3-2b**, but all the scores and loadings are greater than or equal to zero. As an approach to unsupervised ML, it is therefore particularly suited to signals that are intrinsically additive, such as counting electrons. The (fully-connected) autoencoder neural network **Figure 3-2c** employed has three layers: an input (of dimension n), a single hidden layer (of dimension k) and the output, with the same shape as the first. A logistic sigmoid activation function is applied for the sole hidden layer, with the network's outputs left unscaled after the second set of weight multiplication and bias addition. Because the output has the same shape as the input, the network's loss function can be calculated point-wise as the mean-squared-error, combined with some regularisation terms to prevent inflation of weights and encourage sparsity. L2-weight regularisation is used (a penalty to the magnitude of the weights and biases) along with a Kullback-Leibler (KL) divergence term (which penalises deviation of a set of activations from a chosen distribution). Labelled are m observations (scan points) and hidden activations z. In the autoencoder ρ_0 is selected as 0.05, with λ and ω as 0.15. The network is randomly initialised then trained with scaled conjugate gradient descent for 300 epochs (full cycles through every dataset EBSP). Scores were calculated as the values of z for each example. Latent patterns are taken as the columns of the encoder matrix.

The analysis presented in this work was performed on a 64-bit Windows 2019 Server PC, with an Intel ® Xeon ® Gold 6138 CPU and 256 GB of RAM. The *SciKit-Learn* Python 3.7 package was used for PCA and NMF decompositions, and the autoencoder developed in MATLAB 2019b with the *Statistics & Machine Learning* toolbox.

NMF and PCA EBSP dataset decomposition routines have been implemented in the *ebspy* Python package. The pattern centre refinement algorithm is available in *AstroEBSD*, written in MATLAB. These repositories are available open

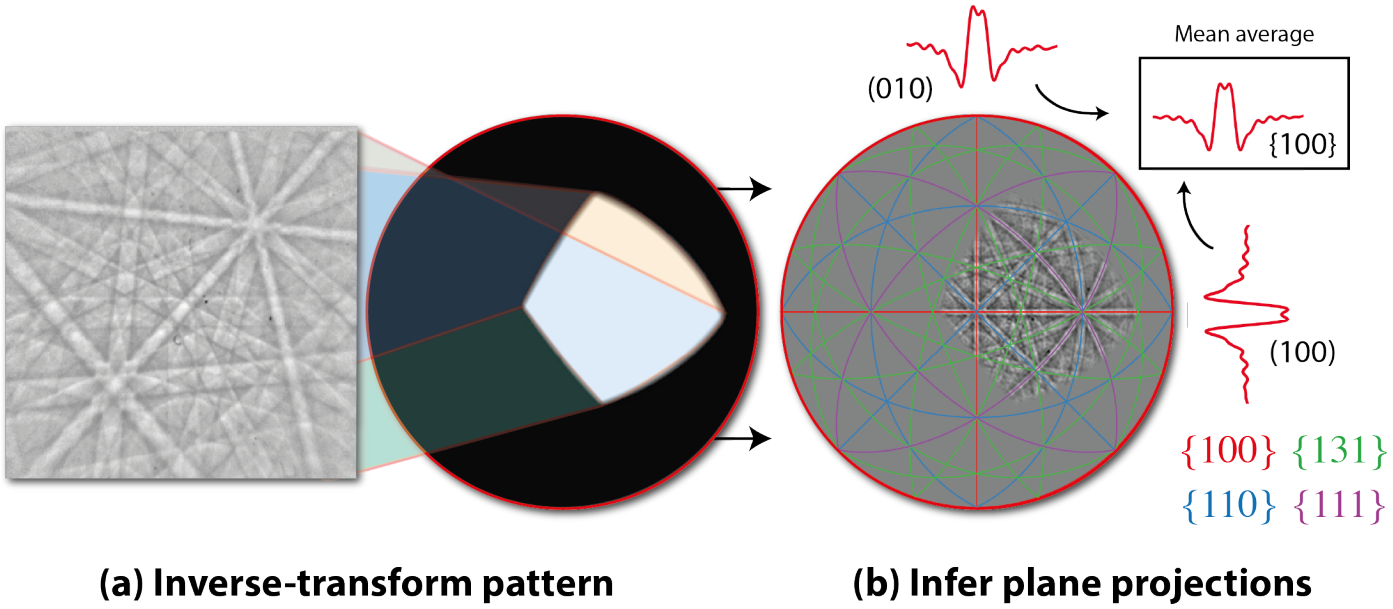

Figure 3-3: EBSPs are re-projected onto the diffraction sphere using precise knowledge of the pattern centre and crystal orientation.

access and can be found at github.com/tmcaul and github.com/benjaminbritton respectively. Band analysis workflows are presented at github.com/tmcaul/SphericalAngularDF, and will be incorporated into *AstroEBSD*. The data and analysis pipeline has been made available open access at DOI: 10.5281/zenodo.3837276.

3.4 Spherical EBSP analysis

Analysis of the profiles of Kikuchi bands in EBSPs is inherently better suited to a spherical co-ordinate system than the gnomonic projection. In the spherical projection, the band centre is a great circle which is the plane perpendicular to the plane diffracting normal. The band can be sampled by examining subsequent small circles, each perpendicular to the diffracting plane normal. Each band profile is calculated through integration 'on-the-sphere', where Kikuchi bands have parallel edges, as opposed to integration along hyperbolic lines in the gnomonic projection [**145**], [**146**]

Here the formulation of Hielscher *et al* [**145**] is adopted, and the measured EBSPs are re-projected onto a calibrated sphere. In order to do this, precise knowledge of the crystal orientation at the scan point is required, coupled with a precise measurement of the pattern centre. To simultaneously achieve this, a simple gradient ascent algorithm is implemented. This uses the peak height of the cross correlation function (XCF) of a candidate EBSP and an orientation refined, simulated template from the evenly SO(3) sampled library. Starting with an initial estimate of the pattern centre and crystal Euler angles (from *Bruker Esprit 2.1*), we simulate templates with increments in PCX, PCY and detector distance (DD), following conventions of Britton *et al* [**117**]. The gradient in XCF peak height with respect to PCX, PCY and DD is used to generate an updated centre, and template matching used to get an updated orientation. This procedure is implemented in *MATLAB* and iterated with decreasing step size in PCX, PCY and DD to generate a highly accurate pattern centre and orientation. This algorithm is discussed further in **Appendix 1**.

This geometry and crystal orientation is used to re-project candidate EBSPs onto the sphere for subsequent analysis. This is achieved by calculating a function f with respect to diffraction directions $\boldsymbol{\xi}$ that maps the inverse gnomonic projection. This follows the expansion:

$$f(\boldsymbol{\xi}) = \sum_{a=0}^{N} \sum_{b=-\beta}^{\beta} \hat{f}(a,b)\, Y_a^b(\boldsymbol{\xi}) \qquad 3\text{-}4$$

Where Y_a^b are the spherical harmonic functions, and $\hat{f}(a,b)$ are the Fourier coefficients of f. N corresponds to the degree of harmonic employed for the expansion. Several approaches for calculation of the Fourier coefficients have been discussed by Hielscher *et al* [**145**], and in this work the 'quadrature' method is employed with $N = 256$.

The diffraction pattern, now on the sphere, can be analysed. The band profiles can be extracted through projections of relevant crystallographic planes using the *MTEX* orientation analysis *MATLAB* package. This pipeline is presented in **Figure 3-3**.

With this calculated projection and the Kikuchi band profiles, the intensity can be integrated in spherical co-ordinates. This corresponds to summing the path-normalised sums of the small circles of the diffraction sphere, varied along the opening angle around a plane projection. Specifically, this is an integration with respect to all the possible rotations around the plane normal, R about angles θ [**145**]:

$$\phi(\boldsymbol{\xi}) = \int_0^{2\pi} \Psi(\, R(\theta)\, \boldsymbol{\xi}\,) d\theta \qquad 3\text{-}5$$

With $\phi(\boldsymbol{\xi})$ the resultant band profile and Ψ the spherically projected pattern. This analysis is performed for both global simulations (in which the full Ψ is calculated) and re-projected patterns in experimental co-ordinates, for which $\Psi = f(P)$, with P a single (square) pattern. In this work, if multiple {hkl} plane projections are present in the field of view the profiles are mean-averaged.

4. Results & Discussion

In a Co/Ni-base superalloy sample known to contain a high ɣ′ volume fraction, EBSPs are collected across an area of interest of 2.85-by-2.16 µm, employing a scanning step size of 30 nm.

4.1 Decomposition evaluation

The latent factors uncovered by unsupervised learning of our AOI are compared, coefficients (latent pseudo-EBSPs) are evaluated and corresponding scores are uncovered by PCA, NMF and the autoencoder neural network. All three approaches are able to extract subtle, physically significant features.

The first two principal components (PCs) (**Figure 3-4-1**,**2**; **b-1**,**2**) represent information specific to ɣ and ɣ′. PCs 1 and 2 provide reasonable distinction between precipitate and matrix, and we attribute contrast in the higher order components to sample topography (and subsequent impact on electron exit angle). This appears to be the case especially for PC 5, where there is a horizontal spatial mode. Such influence is not insignificant, as reflected in the magnitude of the scores

Factor	*1*	*2*	*3*	*4*	*5*
1	0.2036	0.0595	0.0392	0.0630	0.0064
2	0.0595	0.1930	0.0509	-0.0179	0.0301
3	0.0392	0.0509	0.1917	0.0205	0.0605
4	0.0630	-0.0179	0.0205	0.2010	0.0945
5	0.0064	0.0301	0.0605	0.0945	0.2043

Table 3-1: Covariances of autoencoder latent factor scores. Shown in **bold** are the intra-factor variances.

of the higher order (topography-driven) components being of a similar magnitude to the structure-derived differences for PCs 1 and 2.

The solution identified by NMF appears physical in origin, as compared to the statistical solution from PCA. These factors are not ordered by dataset contribution (as the PCs are). Human based analysis of the NMF solutions reveals high contrast segmention of γ and γ′ in Factor 3 (**Figure 3-4c-3**; **d-3**), with signal from the γ regions positively aligning to a vector from γ′ to γ clusters in variable space. Factor 1 is attributed to variation in background (as discussed our positivity condition necessitates a roaming EBSP mean), and the remaining factors to a combination of weak variations in topological mode as in PCA. The factor aligned with precipitate difference (Factor 3) concentrates intensity in the band interiors, which is where the band contrast lies as observed in **Figure 3-5**.

Finally, the decomposition identified by the autoencoder neural network appears as heavily binarised versions of those seen in the matrix decompositions. Human based analysis of the autoencoder solution can be used to attribute Latent 1 to our expected deviation between precipitate and matrix, with the others not being so easily interpretable. Binarisation in Latent 1 is desirable, as the model attempts to separate two very similar crystal structures. For further insight into the network output we present in **Table 1** the covariances of the latent factor scores. As requested the variance in each factor (the diagonal components) are approximately 0.2, with the off-diagonal components smaller. There remains a fairly substantial degree of covariance (for example Factors 4 and 5), even after significant KL-divergence regularisation. As most notably comparable with NMF (PCA does not present a single latent factor solely attributable to γ′) contrast between matrix and precipitate is mostly contained to the band interiors, observed in Latent 1(**Figure 3-4e-1**; **f-1**). The sense of this vector is opposite to that in the NMF, despite similar sense in γ / γ′ separation in scores. This is likely due to the difference in normalisation required by NMF and which side of the mean the vector is operating from. The latent signals are much noisier than those identified with the linear methods, despite a low ultimate error rate and very accurate reconstruction.

4.2 Superlattice segmentation

All three approaches are able to identify subtle differences between γ and γ′. Firstly, analysing the PCA-reduced dataset, the components are binarised at the 0.38 quantile point (subjectively identified) on the sum of the first three PCs (which we interpret to contain significant matrix / precipitate contrast). For the NMF and autoencoder separations, the binarisation is performed at the 0.38 quantile on Factor 3 and Latent 1 respectively. The segmentations are presented in

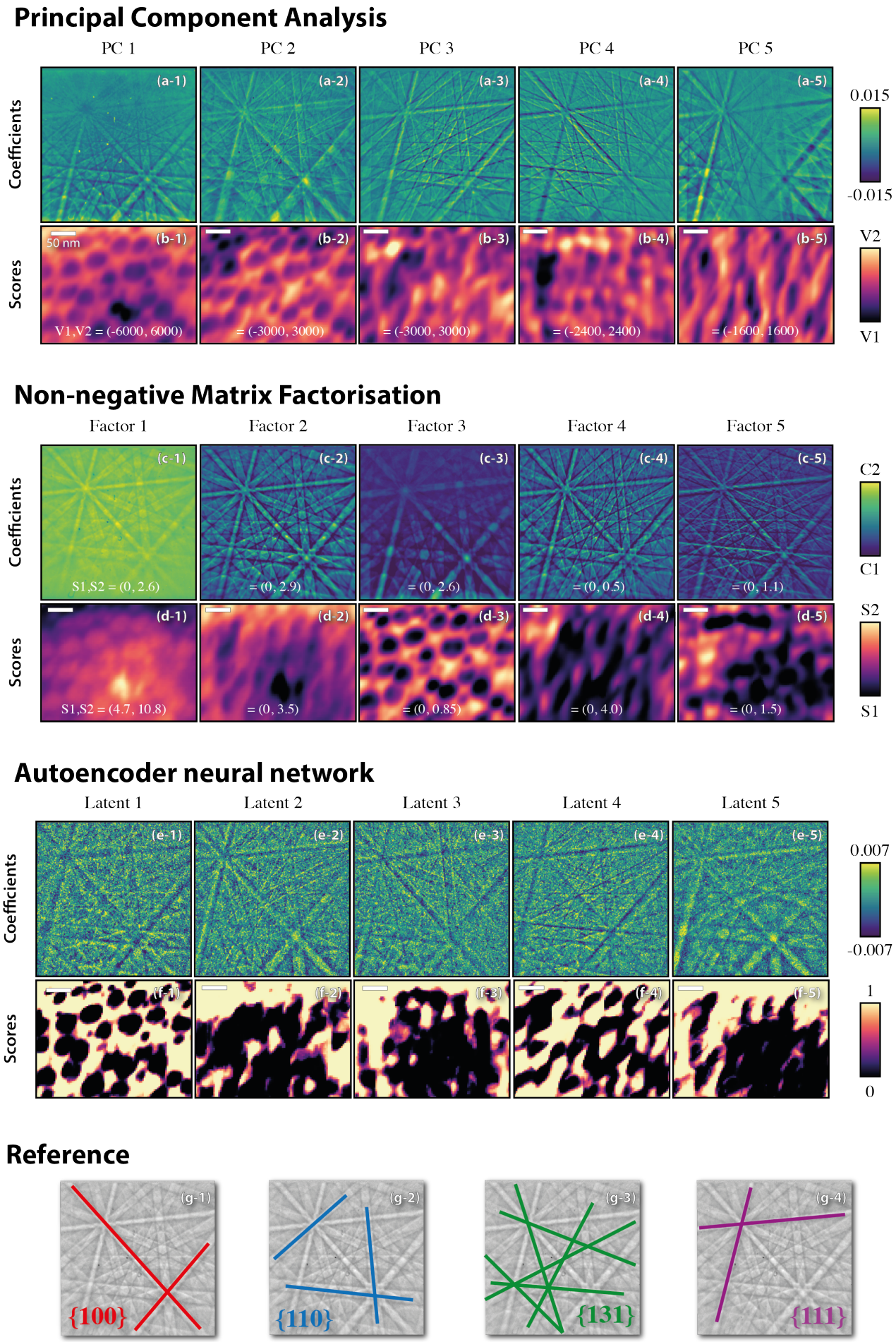

Figure 3-4: Latent factors identified by our three unsupervised machine learning algorithms. PCA - (a,b), NMF – (c,d), and the autoencoder (e,f). Kikuchi bands are labelled in (g) to show corresponding plane projections.

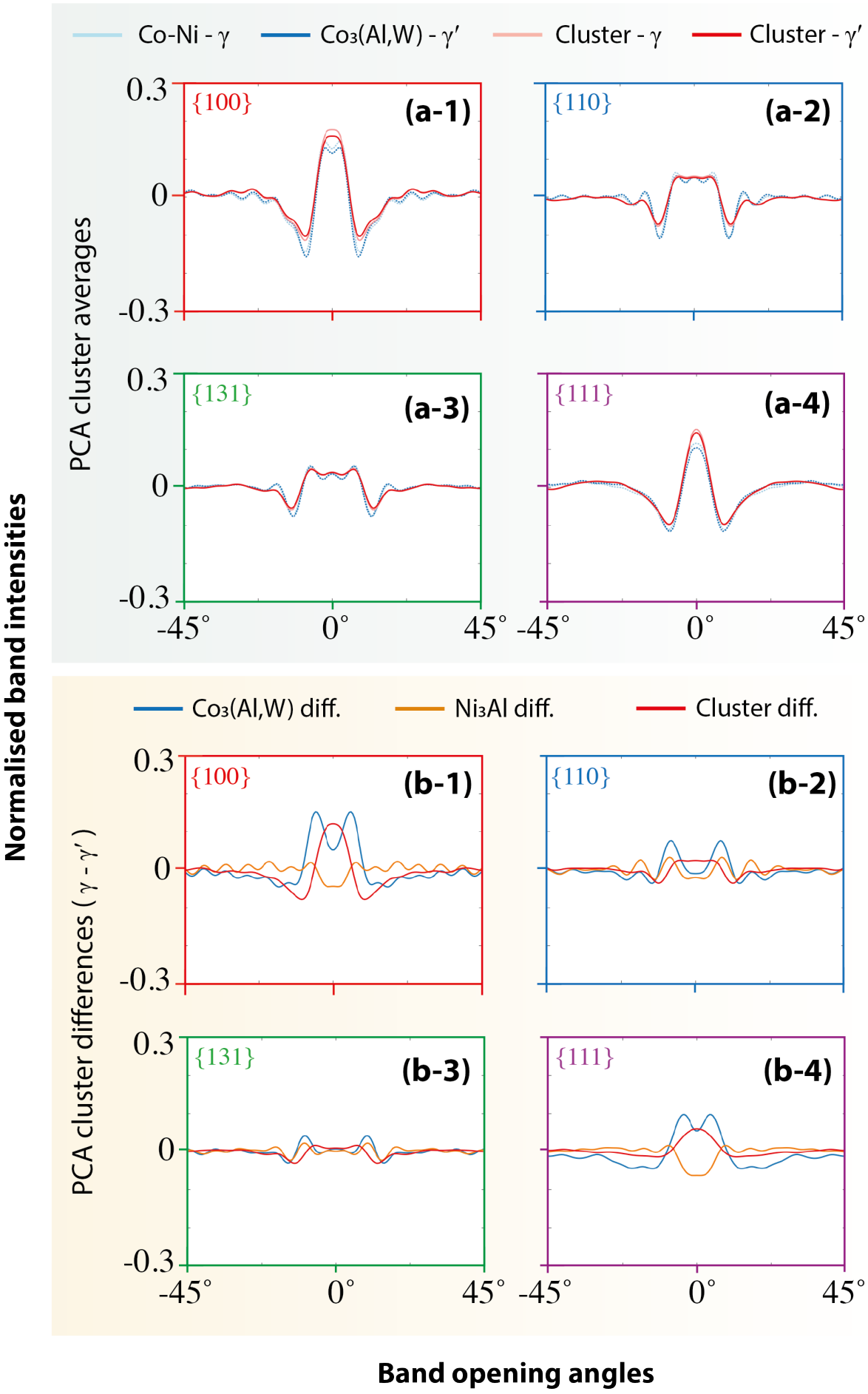

Figure 3-5: Comparison between Kikuchi band profiles obtained from (a) averages of the PCA-identified clusters, and (b) differences between profiles for simulated CoNi-$Co_3(Al,W)$ and Ni-Ni_3Al, and PCA-cluster ɣ - ɣ′ systems.

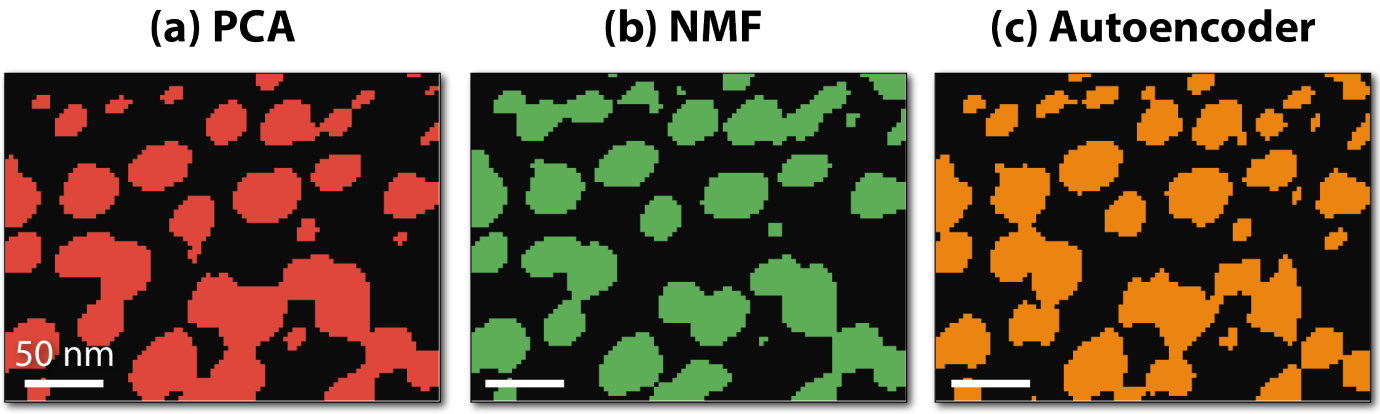

Figure 3-6: Clustering performed on the PCA (a), NMF (b), and autoencoder (c) decompositions, in order to identify ɣ (black) and ɣ′ (coloured).

Figure 3-6. All three lead to reasonable and spatially consistent classification. Subsequently, the average measured EBSPs from each of the classified regions are compared to dynamical simulations 'on-the-sphere' after Hielscher *et al* [**145**] in order to correct for hyperbolic divergence of as-measured (gnomonically projected) Kikuchi patterns. The spherically integrated profiles, Φ_{ij}^{hkl}, of crystal plane families {hkl} for each of the ɣ - ɣ′ class-average EBSPs are plotted in **Figure 3-9**. These are directly compared with dynamical simulations (for a replica of the crystal orientation and camera geometry). Due to the similarly good performance in classification, the cluster-average patterns are extremely similar for PCA, NMF and autoencoder approaches. This can clearly be seen in **Figure 3-7**.

Figure 3-5a shows the same pattern in ɣ and ɣ′ intensities for the clustered experimental EBSPs as for the dynamical simulations of the CoNi-Co_3(AlW) system: ɣ′ generally diffracts less at small opening angles than ɣ, as was qualitatively observed in the latent factor backscatter patterns (**Figure 3-4**). This is the case across the AOI for the {100}, {110}, {131}, and {111} band profiles. The differences between matrix and precipitate profile, presented in **Figure 3-5b**, shows the experimental (clustered) EBSPs retain consistently the same sign in profile difference as the Co pairing, and are opposite in sign to the Ni pairing. In order to confirm that this observation is not an artefact of our detector undersampling the diffraction sphere, global simulations are performed in addition to the inverse-gnomonically projected single templates analysed in **Figure 3-5**. Global simulations of the full diffraction sphere, presented in **Figure 3-8**, show the same fingerprint.

The difference in scattering behaviour between Ni and Co-base systems is attributed to chemical segregation between matrix and precipitate. In Co-base superalloys the ɣ′ precipitates require stabilisation with W to prevent formation of extraneous microstructural phases such as B2 CoAl and DO_{19}. This necessitates segregation of W to the precipitates, and will result in a greater tendency for inelastic electron scattering [**107**], [**135**], [**268**]. This in turn reduces the elastically diffracted signal for the ɣ′ relative to the ɣ. The lack of heavy element segregation to ɣ′ in the Ni / Ni_3Al system leads to a reversal in contrast, which is not observed the experimental patterns. Winkelmann & Vos [**133**] have previously established the importance of the atomic species and degree of localisation of the scattering inelastic source in Kikuchi band formation. These results, and attribution of the segregation of W to the ɣ′ system, agree with their conclusion that eventual measured intensity distributions are sensitive to scattering of the incoherent point source in the unit cell.

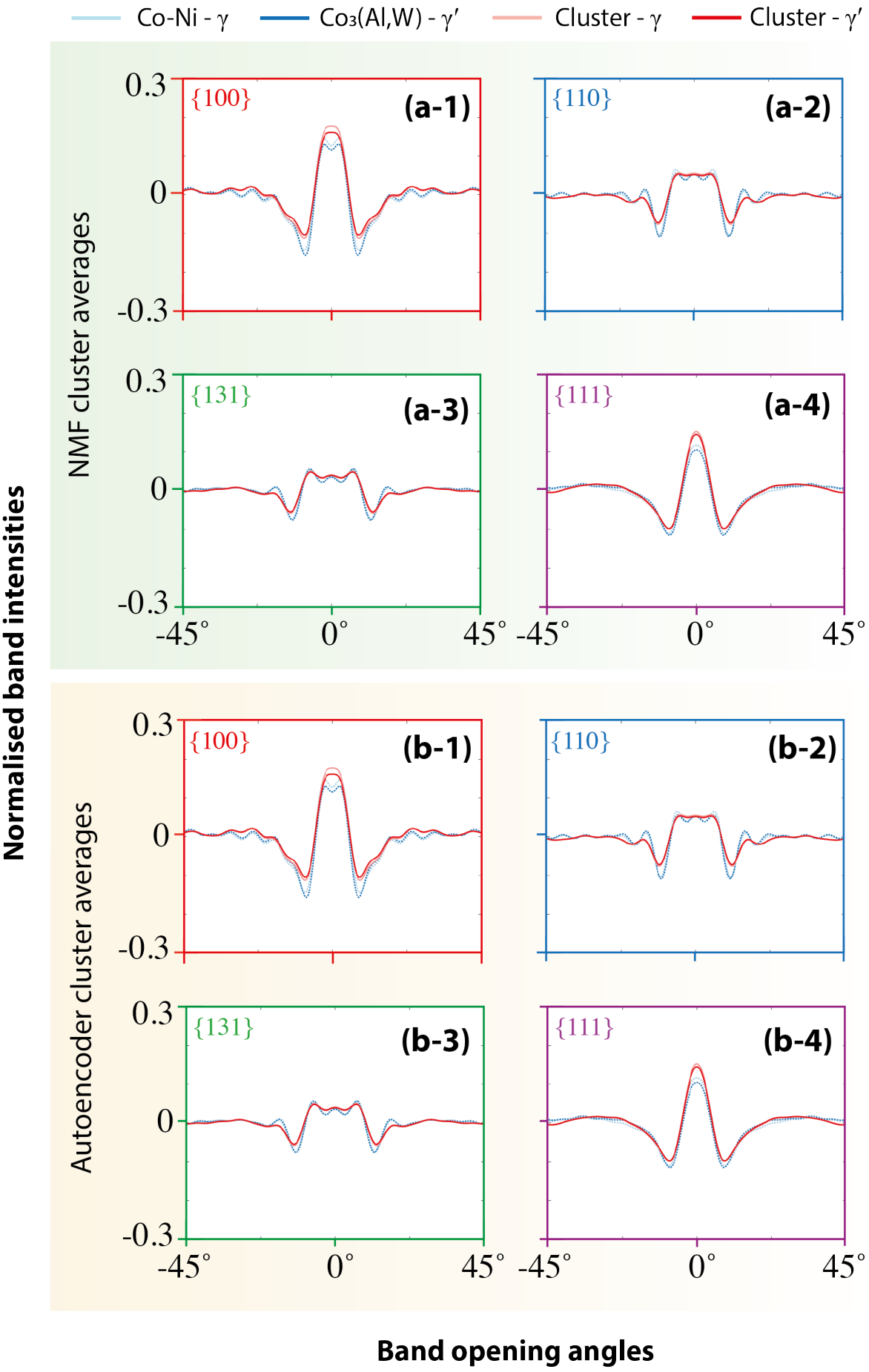

Figure 3-7: Comparison between Kikuchi band profiles obtained from (a) averages of the NMF-identified clusters, and (b) averages of the Autoencoder-identified clusters.

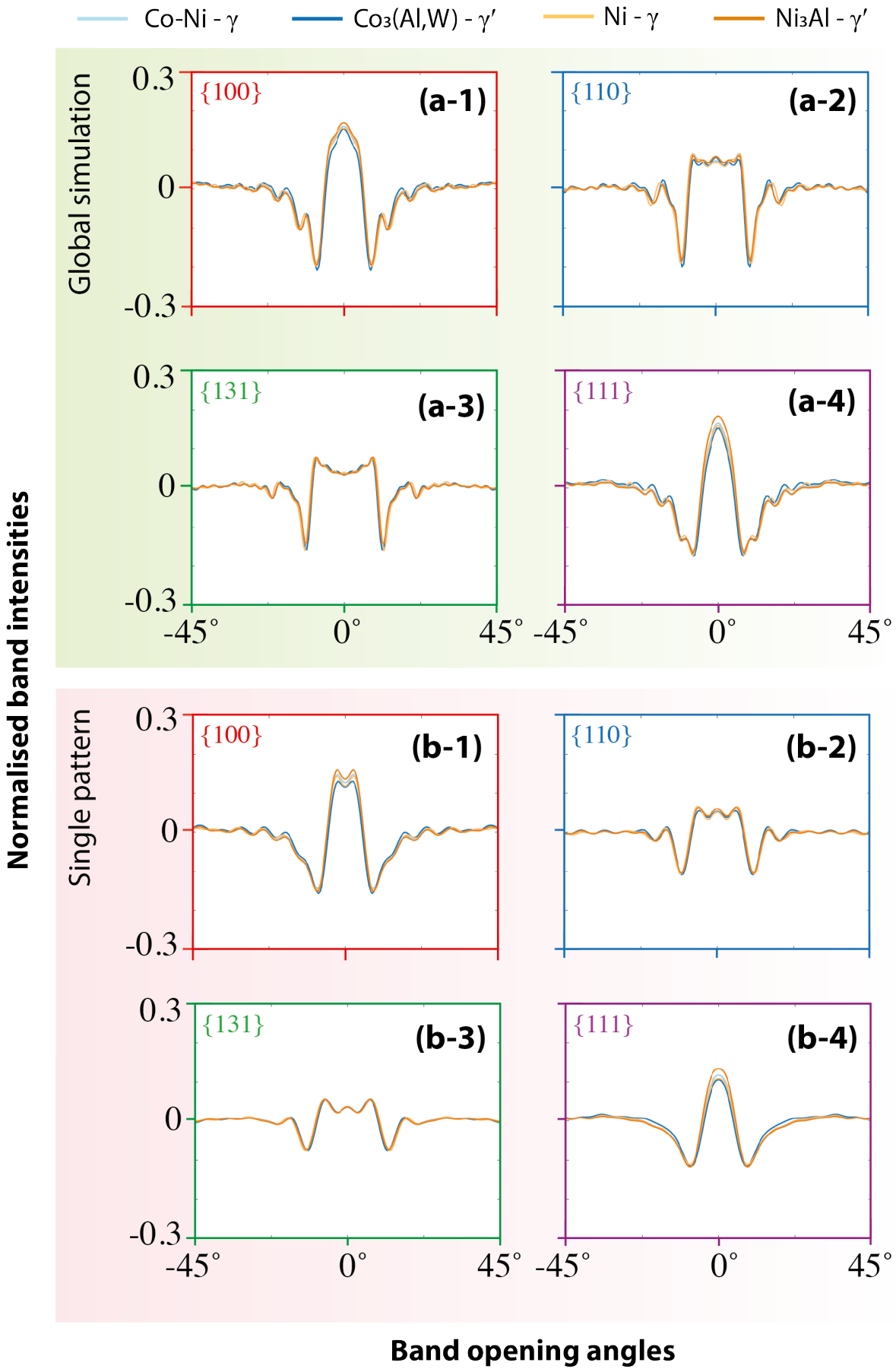

Figure 3-8: Comparison between Kikuchi band profiles obtained from (a) the global simulation (full diffraction sphere) and (b) a re-projected single template, for both a CoNi-$Co_3(Al,W)$ and Ni-Ni_3Al ɣ - ɣ′ simulated systems.

4.3 Spherical-angular dark field imaging

In order to directly observe the ɣ - ɣ′ intensity polarisation seen in the cluster-average EBSPs, the profile sums within one Bragg angle of the plane projection (integrating +/- the Bragg angle from 0˚ in the profile scheme of **Figure 3-5**) are calculated for every scan point and plane family. Specifically:

$$S_{ij}^{hkl} = \int_{\theta=0-\theta_{bragg}}^{\theta=0+\theta_{bragg}} \Phi_{ij}^{hkl}\, d\theta \qquad 3\text{-}7$$

This is evaluated at scan point *i,j,* with spherically projected profile Φ_{ij}^{hkl}, plane family{hkl}, and band opening angle θ. This enables generation of virtual 2D microstructural images based from specific diffraction conditions, similarly to 'virtual dark field' analyses commonly performed in the TEM community [**147**], [**148**], [**269**]. This is a more advanced approach than previous dark field EBSD-based methods [**270**], [**271**]. Analysis 'on-the-sphere', accounting for hyperbolic divergence of the Kikuchi bands in the gnomonic projection, enables windowing of specific diffraction based contrast variations. This is only possible as the profiles from all integrated {hkl} Kikuchi bands are included to increase signal to noise, and improving confidence in the crystallographic origin of the contrast. This contrast is verified through analysis of the dynamical simulations.

Analysis of the {100}, {110}, {131}, and {111} conditions is presented in **Figure 3-9**. There is a wider spread in intensity for the {100} and {111} integrations. Correspondingly, the magnitudes of the peaks in **Figure 3-5** (**b-1**, **b-4**) are greater than those in (b-2, b-3), in turn agreeing with dynamical simulation. The probability density function, **Figure 3-9e**, accordingly displaying a greater spread of intensities for these high-contrast conditions than {110} and {131}. In

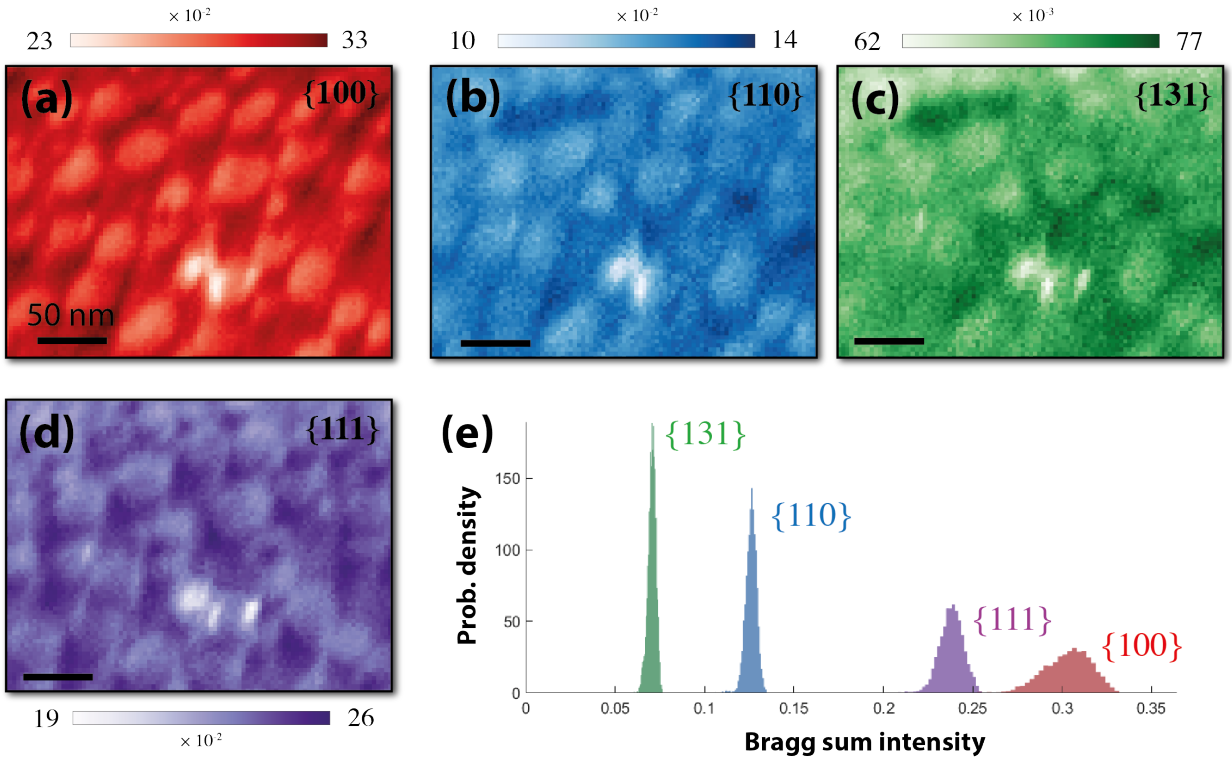

Figure 3-9: Spherical-angular dark field imaging (Bragg summations) of the dataset: (a-d) integrated intensities of the corresponding band profiles , and (e) probability density distribution (histogram normalised by number of observations and bin width) of the calculated intensities for each of the diffraction conditions.

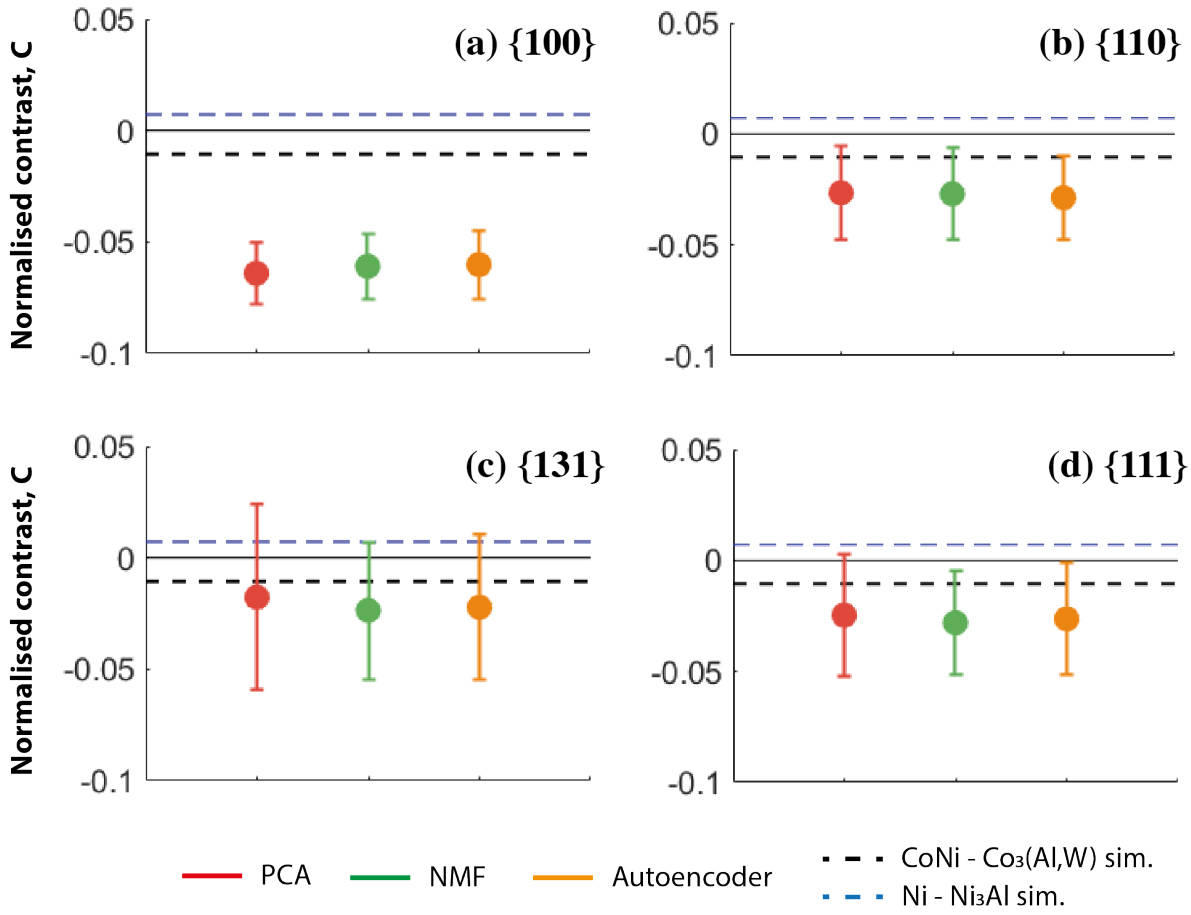

Figure 3-10: Normalised contrast metrics calculated for each of the diffraction conditions observed, compared to CoNi-Co_3(Al,W) and Ni-Ni_3Al ɣ - ɣ′ pairing dynamical simulations.

Figure 3-9e there are two superposed intensity distributions (which are not resolved here) for each diffraction condition, corresponding to signals from the matrix and precipitate. The wider spread in intensity for {100} and {111} virtual crystallographic images is a result of better separated average intensities, as observed in the cluster-average profiles of **Figure 3-5**.

4.4 Consistency and discussion of clustering approaches

In order to compare our clustering approaches for separation of precipitate from matrix, a simple ɣ - ɣ′ normalised contrast metric, *C*, is calculated for each diffraction condition family and clustering approach:

$$C = \frac{S_{\gamma'}^{hkl} - S_{\gamma}^{hkl}}{S_{\gamma'}^{hkl} + S_{\gamma}^{hkl}} \qquad 3\text{-}8$$

For S_{phase}^{hkl} in the first instance corresponding to the mean value of ɣ and ɣ′ segmented regions of S_{ij}^{hkl}. *C* is calculated for each of the segmentation approaches (spatially resolved in **Figure 3-6**). *C* is also calculated for the simulated CoNi-Co_3(Al,W) and Ni-Ni_3Al ɣ - ɣ′ pairings. These, along with errors propagated from standard error in the means of the corresponding plane-specific integrated intensities, are presented in **Figure 3-10**. These contrast metrics show that all three dataset segmentations capture the negative contrast predicted for the CoNi-Co_3(Al,W) system. It can be concluded

that the ɣ′ in this alloy has crystallographic behaviour closer to the $Co_3(Al,W)$ archetype than Ni_3Al, a characteristic that has historically required TEM to observe.

Of the presented approaches, NMF seems to achieve the best consistency in segmentation (resulting in smaller error-bars in **Figure 3-10**), and broadly the only approach able to present consistently negative contrast across all diffraction conditions fully within error. This is due to the algorithm's convergence to a basis that explicitly includes a factor corresponding to a ɣ - ɣ′ vector in variable space. This was not the case for PCA, which did not identify a component as well aligned to this crystallographic difference.

The autoencoder identified a similar latent factor to NMF, well aligned to a crystallographic difference vector. However, the segmentation was not as well spatially resolved. This is likely due to the activation function we employ, which due to the extreme similarity between matrix and precipitate EBSPs experienced a challenging task in where to correctly assign the domain of the logistic sigmoid function to observe the necessitated (Kullback-Leibler regularised) variation in latent score. The autoencoder implementation with one hidden layer and no convolutional filters compares favourably to PCA and NMF. With supervised structure classification using deep neural networks beginning to see application to EBSD data, it is valuable to investigate specifically what can be learned by simpler architectures. After 300 epochs of (over-) training the latent representation the simple network has learned is noisy, despite very accurate reconstructions. The ability of the network to learn EBSP-specific features (such as zone axis contrast, band intensity profiles) likely further diminishes with max-pooling and convolution operations.

Spherical-angular dark field imaging presents exciting possibilities for microstructural analysis. Conventional micro-scale SEM imaging modes (usually backscatter or secondary electron) are naïve to detected electron energy and geometry. Superalloys, due to their structural and chemical similarity in total inelastic scattering propensity, therefore often require chemical etching and surface modification to generate suitable contrast for evaluation. This results in highly subjective analysis, as knowledge of how the surface is modified can be extremely complicated and vary. Direct use of scattering and diffraction data reduces this uncertainty and it has been shown that collecting intensity 'on-the-sphere' generates contrast between precipitate and matrix at specific diffraction conditions. This contrast is due to intensity differences derived from chemical segregation. Averaging over ML-derived segmentations of the dataset amplifies the signal to noise for these subtle variations and provide a comparison metric between microstructural constituents. Such an approach may prove useful for similarly challenging crystallographic similarity problems in EBSD, such as martensite characterisation and carbide type differentiation.

5. Conclusions

In this work segmentation of ɣ matrix from ɣ′ precipitate in a Co/Ni-base superalloy has been investigated, using unsupervised machine learning. EBSPs were successfully clustered using principal component analysis, non-negative matrix factorisation, and an autoencoder neural network. The following conclusions are drawn:

- All three approaches are suitable for identification of very subtle differences in EBSP band contrast resulting from superlattice ordering. NMF provides the most physically justifiable basis, including a factor that explicitly aligns with region ordering. The autoencoder finds a similar feature, but with worse spatial fidelity and noisy latents. PCA finds a basis that includes reasonable superlattice contrast in the strongest principal components, with higher order terms appearing to represent sample topography.

- Segmentations from all three approaches explicitly show less intense diffraction at the band cores in superlattice ($L1_2$) ɣ′ than in matrix (FCC) ɣ. This agrees with simulations of a CoNi-Co_3(Al,W) system, and is opposite in sense to simulations of a Ni-Ni_3Al system. This behaviour is attributed to reduced elastic scattering in ɣ′ where heavier elements (such as W) tend to segregate.
- Virtual crystallographic imaging of the area of interest (summing intensity within one Bragg angle of the plane projection, accounting for hyperbolic band divergence) shows greater normalised superlattice/matrix contrast for {100} and {111} diffraction conditions than {110} and {131} for all three segmentation approaches. NMF provides consistently the lowest standard error in this contrast metric for the bands that were integrated.

6. Chapter Acknowledgements

We are grateful to the developers of the *MTEX* orientation analysis package and Hielscher *et al* [**145**] for making their code available open access under an MIT license at github.com/mtex-toolbox/mtex-paper.

TPM, DD, and TBB acknowledge support from the Rolls-Royce plc - EPSRC Strategic Partnership in Structural Metallic Systems for Gas Turbines (EP/M005607/1), and the Centre for Doctoral Training in Advanced Characterisation of Materials (EP/L015277/1) at Imperial College London. TBB additionally acknowledges the Royal Academy of Engineering for funding his research fellowship. DD acknowledges funding from the Royal Society for his industry fellowship with Rolls-Royce plc. We thank the Harvey Flower Microscopy Suite at Imperial College London for access to microscopy facilities.

Chapter 4 - Variation in physical properties with Nb and Ti additions to a Co/Ni-base superalloy

1. Chapter Summary

Additions of Nb and Ti to a series of Co/Ni-base superalloys derived from previous work were found to improve room and high temperature strength. Ti was observed to have a stronger strengthening effect than Nb. Both alloying additions raised the ɣ′ solvus, but even +2at.% Ti maintained a solvus of < 1100˚C. Ti reduced the density while Nb raised it. Nb was found to be beneficial to oxidation performance, in agreement with previous findings in Ni-base superalloys, while Ti was shown to be deleterious. In this the physical properties of these alloys are compared to the state-of-the-art in gas turbine (polycrystalline) disc materials, and it is shown that they are competitive.

2. Introduction

The V208 alloy series [**1**], [**12**], [**261**] has seen improvements in yield strength, creep resistance, oxidation resistance and microstructural stability in the past 6 years [**13**], [**272**], [**273**]. Novel characterisation methods have been built to facilitate rapid phase identification for alloy development [**93**]. In this work, years of research into Co/Ni-base superalloys is leveraged to achieve unprecedented strength, oxidation resistance, and density reductions, while maintaining a low solvus (seen to be the principal advantage of Co-superalloys as compared to their typically stronger Ni-base cousins. To facilitate this, the alloy series presented by Reynolds [**272**] (which builds upon V208C by including Mo for creep resistance) is improved by including additions of Nb and Ti. These additions are expected to stabilise ɣ′ and improve strength, *via* raising the antiphase-boundary (APB) energy and lowering the stacking fault energy (SFE) [**251**], [**274**], [**275**]. APB creation and stabilisation by successive a/2<110> superpartial dislocations in the ɣ′ is a well known phenomenon [**6**], [**21**], [**32**]. It is typically a yield mechanism, but has been observed during creep [**32**]. It is only observed in ɣ′, because only here will APBs be created and require dislocation pairing. Thus, segregation of Nb and Ti to the ɣ′ phase, and interaction with propagating Shockley partial dislocations, has capacity to improve yield strength through inhibition of dislocation motion.

Alloy	Co	Ni	Mo	Cr	Al	W	Ta	C	B	Zr	Ti	Nb
(0Ti, 0Nb)	~36	~36	2	12	10	2.75	1.25	0.3	0.085	0.04	0	0
Base: (0.3Ti, 0Nb)	~36	~36	2	12	10	2.75	1.25	0.1	0.042	0.04	0.3	0
Base + 0.5Nb	~36	~36	2	12	10	2.75	1.25	0.1	0.042	0.02	0.3	0.5
Base + 1Nb	~35	~35	2	12	10	2.75	1.25	0.1	0.042	0.02	0.3	1
Base + 2Nb	~35	~36	2	12	10	2.75	1.25	0.1	0.042	0.02	0.3	2
Base + 1 Ti	~35	~35	2	12	10	2.75	1.25	0.1	0.042	0.02	1.3	0
Base + 2Ti	~35	~36	2	12	10	2.75	1.25	0.1	0.042	0.02	2.3	0

Table 4-1: Alloy compositions discussed in this chapter (all at.%).

3. Methods

Seven approximately 415 g ingots were made by vacuum arc melting under a back-filled argon atmosphere. Compositions are presented in **Table 4-1**: Nb and Ti additions were made by substituting for equal parts Co and Ni. The alloy labelled (0Ti, 0Nb) is the same as that labelled Alloy 1 in Chapter 2. 'Base' is the same as Alloy 13. Each was vacuum homogenised for 48 h at 1200℃. Ingots were then hot rolled at nominally 1200℃, above the solvus temperature, from an initial thickness of 23 mm to 12 mm, using successive 12-15% reductions. Samples for testing were electrical discharge machined from the rolled bars, and encapsulated in back-filled argon quartz tubes for heat treatment. They were brought to the solution heat treatment temperature of 1100˚C at 4˚C/min (above 500˚C), soaked for 1 hour; cooled at 20˚C/min to 800˚C, and aged for 4 h; cooled at 20˚C/min to 500˚C, and finally air cooled. The solution stage aims to dissolve all the ɣ′ precipitated during hot rolling and uncontrolled cooling, in order to 'reset' the alloy and generate a controlled ɣ′ distribution upon ageing.

Tensile testing was performed using a Zwick-Roell 5kN rig, with 'dogbone' shaped samples of 19 mm gauge section and approximately 2-by-1.5 mm (varying but measured) cross-section. Crosshead displacement was used for measuring extension. All stresses and strains reported are engineering-type. Yield stress was taken as 0.2% offset, and ultimate tensile strength was taken as the maximum achieved stress. A *METTLER-TOLEDO* differential scanning calorimeter (DSC) was employed to determine T_{solvus} at a 10°C/minute scan rate under argon atmosphere. A *Zeiss-Auriga* scanning electron microscope (SEM), equipped with a focussed ion beam (FIB) was used to examine oxide-cross sections. Average oxide depths were inferred from sampling 20 points across the tilt-corrected micrograph, observing the total damage depth, and taking the mean.

Electron backscatter diffraction (EBSD) using a *Zeiss Gemini Sigma300* SEM, with a Bruker e$^-$Flash^{HD} EBSD detector was used to determine grain size. An accelerating voltage of 20 kV, step size of 3 μm, and pattern resolution of 160-by-120 pixels was used to measure crystallographic orientation. Grain boundaries were defined where a pixel-neighbour misorientation of over 5˚ was measured. EBSD was combined with energy-dispersive X-ray spectroscopy (EDS) using an *XFlash 6160* detector for Phase-ID, as discussed in **Chapters 1** and **2**. *Bruker DynamicS* was used to dynamically simulate library EBSPs for each candidate structure [**132**], [**247**]. From within the fundamental zone for each phase, a SO(3) sampling frequency of 7˚ was employed for generation of an EBSP template library in the detector reference frame (pattern centre selected with *Bruker Esprit 2.1* from the well-indexed matrix regions and simulated patterns sampled as 200 by 200 px). RC-spectra were quantified with *Bruker Esprit 2.1* using a P/B ZAF correction algorithm accommodating the 70˚ sample tilt required for EBSD.

Density measurements were taken using the Archimedes method, by comparing the measured weight both in air and suspended in deionised water.

4. Results

All melted alloys except for +2Nb were successfully hot rolled above the ɣ′ solvus. The +2Nb alloy exhibited 'alligatoring' upon the final pass through the hot rolls. This is likely due to its high solvus, which may have been transitioned through due to heat flow from the hot ingot to the cold rolls. Backscatter SEM images of 0Ti, 0Nb and the Base+ 1 Ti alloys are presented in **Figure 4-1**, with the latter representative of others' microstructures. For further comparison see **Appendix 4**, **Figure Ap4-1**.

Figure 4-2 presents high temperature tensile testing results from the characterised alloys. It is observed that additions of both Nb and Ti increase substantively increase the room temperature strength of the developed alloys, permitting over 1 GPa in yield strength at room temperature in the +2Ti alloy. Nb appears to have less of a strengthening effect than Ti. Furthermore, Ti additions raise the high temperature yield stress more substantially than Nb. Neither Nb nor Ti appear to have a certain effect on ultimate (peak) strength, but the +2Ti alloy exhibits the highest ultimate strength at 700˚C, exceeding 1.2 GPa. In **Table 4-2** grain sizes, lognormal distribution parameters, and phse presence are presented. The natural log of the grain size is fitted to a Gaussian distribution, with sample mean $\hat{m}$ and sample variance $\hat{s}^2$. Formally:

$$\log(G) \sim N(m, s^2) \qquad 7\text{-}1$$

Where G is grain size, m and s are the true population mean and standard deviations, and N is the normal distribution. Comparisons of the observed distributions both with and without twins shows strong similarity. The values of $\hat{m}$ lie within much less than a single $\hat{s}$; it can be concluded that there is no evidence that the grain sizes are strongly affected

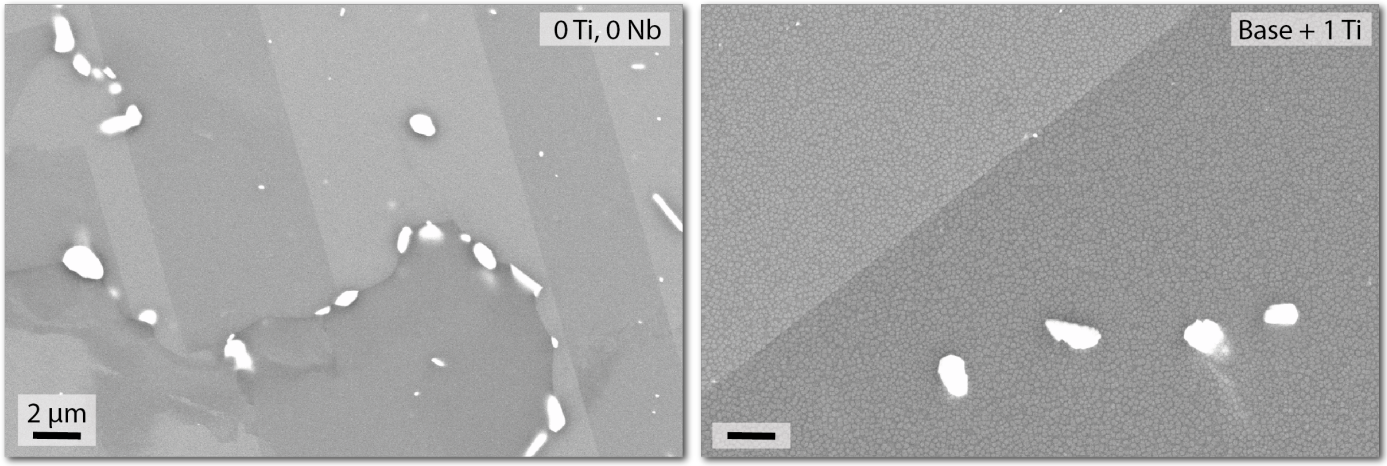

Figure 4-1: Backscatter SEM imagse from 0Ti, 0Nb and Base + 1 Ti alloys. The latter is representative of all other alloy microstructures, for further comparison see **Appendix 4**.

	Alloy	Median / µm	$\hat{m}$ / log(µm)	$\hat{s}$ / log(µm)	Precipitates
Without twins	0 Ti, 0 Nb	13.74	2.72	0.630	M_2B, M_6C, MC
	Base	64.4	3.99	0.746	MC
	Base + 1 Ti	107.7	4.48	1.07	MC
	Base + 2 Ti	72.4	4.10	0.865	MC
	Base + 0.5 Nb	54.6	3.81	0.830	MC
	Base + 1 Nb	59.4	3.91	0.753	MC
With twins	0 Ti, 0 Nb	16.8	2.93	0.752	M_2B, M_6C, MC
	Base	28.7	3.28	0.754	MC
	Base + 1 Ti	45.7	3.71	1.00	MC
	Base + 2 Ti	29.5	3.31	0.820	MC
	Base + 0.5 Nb	21.8	3.07	0.747	MC
	Base + 1 Nb	25.5	3.21	0.783	MC

Table 4-2: Grain sizes and distribution parameters after Nb and Ti additions to the alloys.

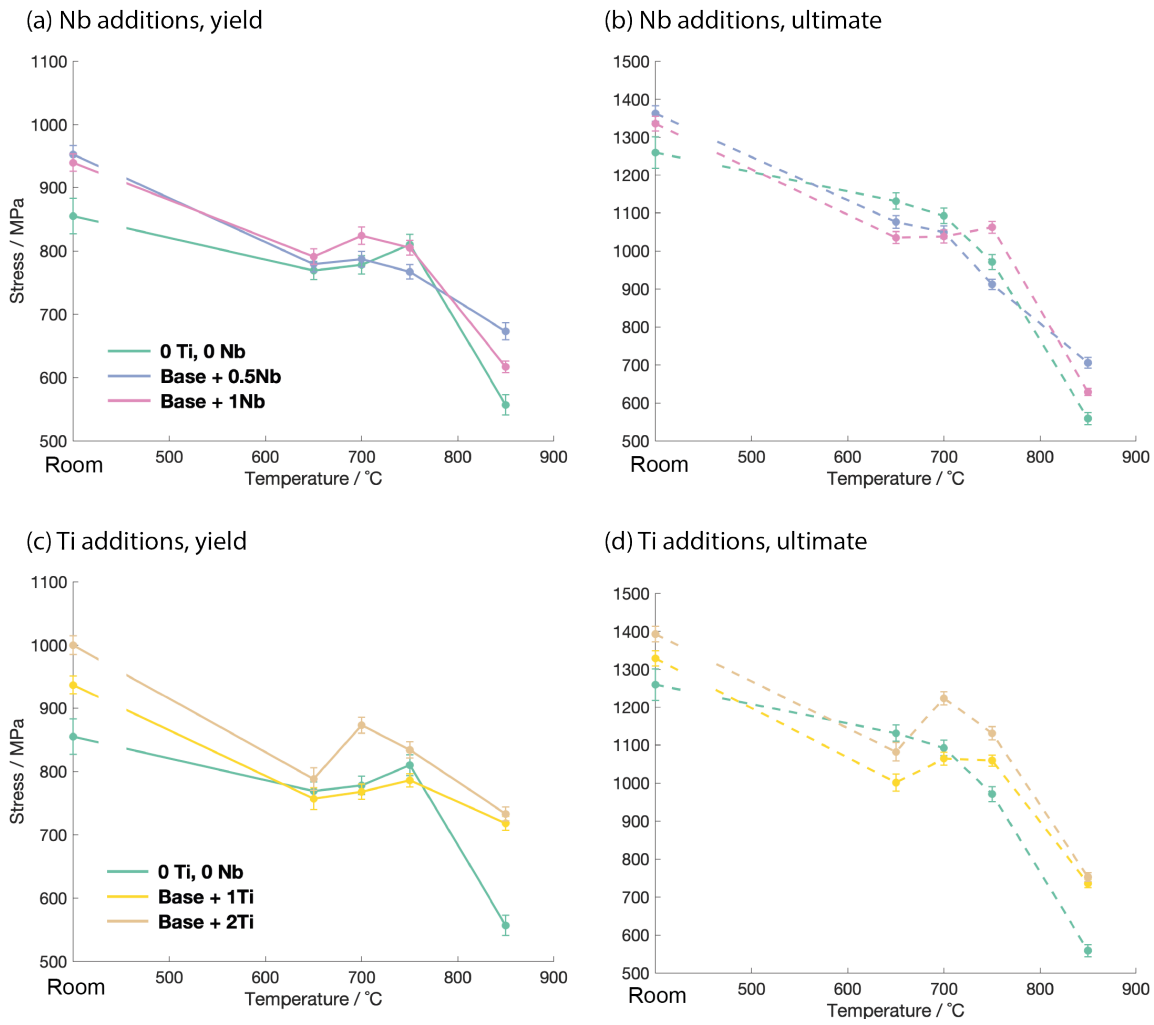

Figure 4-2: Yield and ultimate stress variation with temperature, for each of the candidate alloys.

by the alloying additions. Instead, it seems that the strength improvements observed across temperature regimes are derived from chemical interaction with dislocation propagation mechanisms.

As presented in **Figure 4-3**, additions of Nb and Ti both act to raise the ɣ′ solvus. Fitting a linear trend to the measured values gives an increase of 33˚C / at.% for Ti additions (with an R^2 of 0.96), and 16˚C / at.% for Nb additions (R^2 of 0.97). This is expected – Nb [**251**] and Ti [**276**] are both known to be ɣ′ stabilisers, and their addition will permit the superlattice to be more stable at higher temperatures, implicitly raising the solvus.

The oxidation performance of the considered alloys is examined by measuring the damage penetration depth after exposure at 800˚C for 100 h. Oxide cross-section micrographs are presented in **Figure 4-4**, obtained by slowly polishing the surface at tilt with a FIB. It is observed that Ti additions appear detrimental to oxide performance, and lead to significant increases in damage depth, especially with regard to ingress depth of Al_2O_3 'fingers'. Nb improves oxidation performance in our alloys, similarly to in Ni-base superalloys as discussed by Christofidou *et al* [**251**].

Finally, the density of each of the alloys was measured, and results are presented in **Table 4-3**. Additions of Ti reduce the density, while Nb raises it. This agrees with expectation, as equal parts of (heavier than Ti, lighter than Nb) Ni and Co were replaced in the nominal atomic compositions examined. All physical properties are tabulated in **Table Ap4-2**.

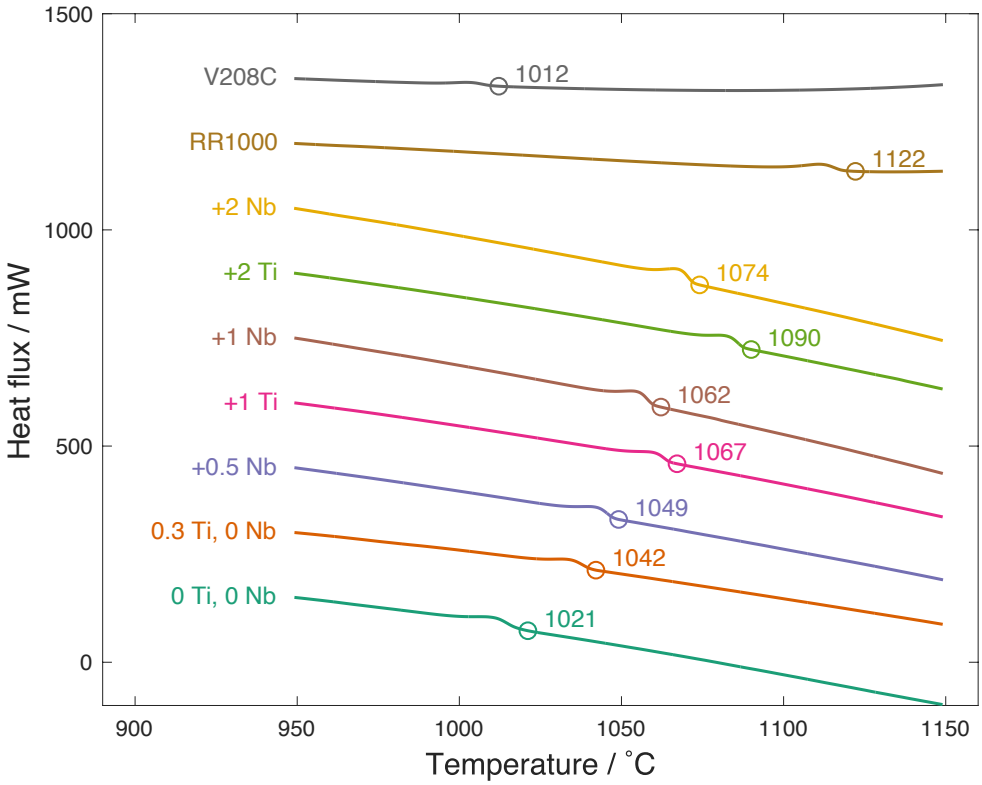

Figure 4-3: Solvus measurements from DSC cooling curves for each of the candidate alloys, with comparison to previously published V208C [**1**] and coarse grained RR1000 [**2**].

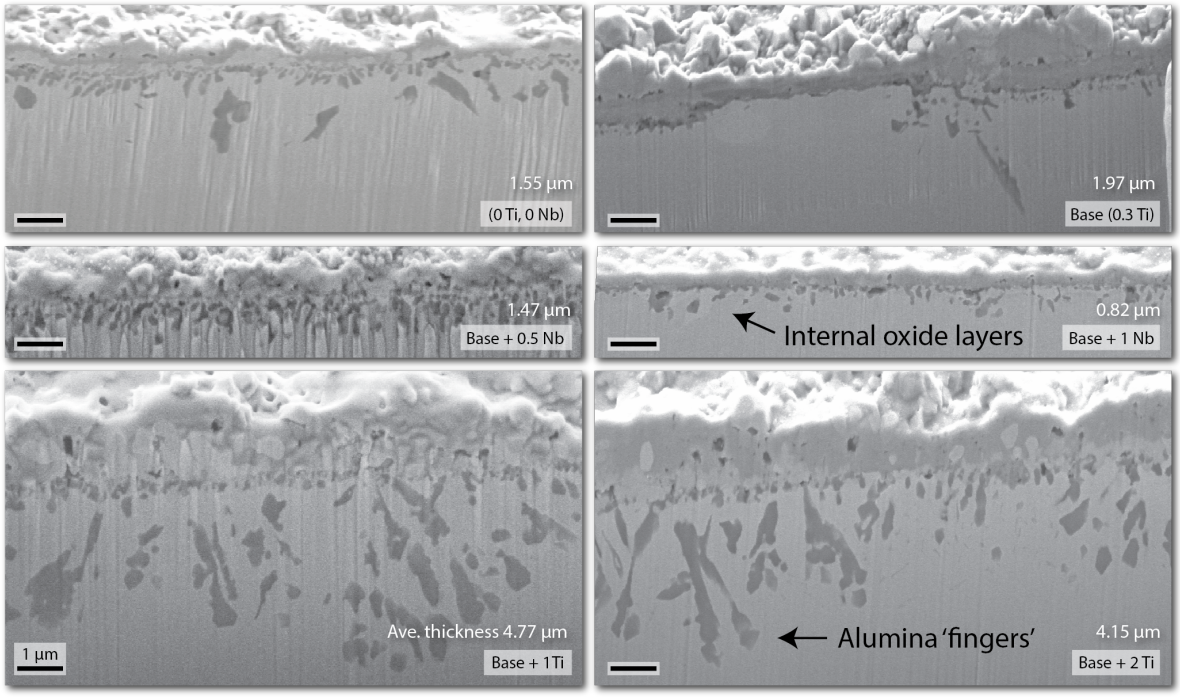

Figure 4-4: Cross-sections of oxide scale after 100 h exposure at 800˚C.

Alloy:	(0 Ti, 0 Nb)	Base (0.3 Ti, 0 Nb)	Base + 0.5 Nb	Base + 1 Nb	Base + 2 Nb	Base + 1 Ti	Base + 2 Ti
Density: / g cm^{-3}	8.67	8.61	8.69	8.74	8.79	8.61	8.53

Table 4-3: Measured densities for each of the compared alloys.

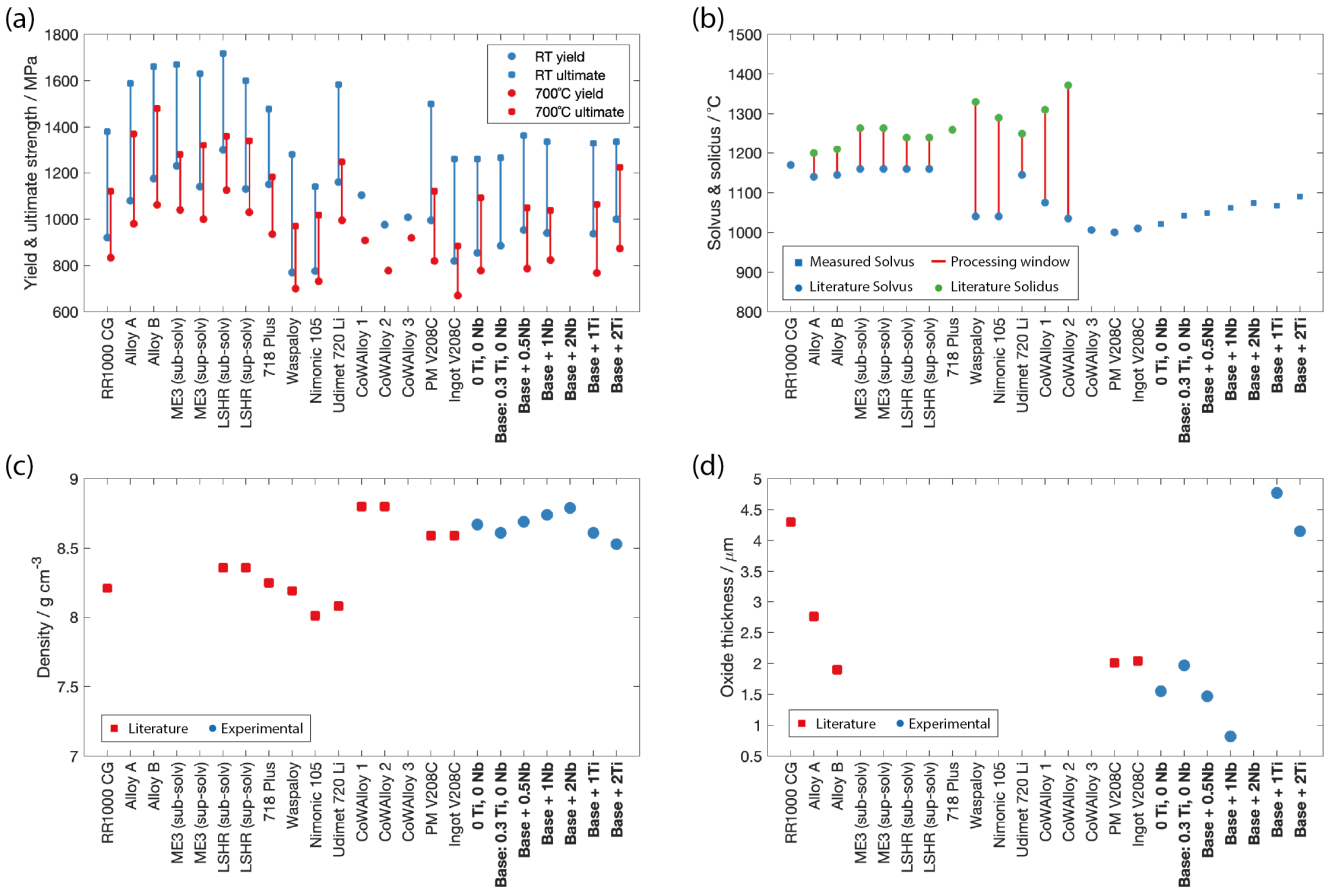

Figure 4-5: High-temperature tensile (HTT) testing of candidate alloys (a), solvus and solidus (b), density (c) and oxide damage depth at 100 h / 800˚C (d), referenced with respect to literature where comparable measurements are available.

5. Discussion

Figure 4-5 presents a comparison to the evaluated physical properties of this alloys series in comparison to state of the art superalloys used in gas turbine applications, by various engine manufacturers. The compositions presented in this work maintain a low solvus (< 1100˚C), significantly lower than RR1000 and the candidate replacements Alloys A and B [**251**]. The densities we observe are comparable to previous Co-base alloys, such as the CoWAlloy series [**10**] and V208C [**1**], but typically higher than the Ni-base alloys such as RR1000 [**2**] and LSHR [**277**]. The oxidation resistance of the alloys appears to be comparable, and in the case of the Nb-containing alloys slightly better than, the competitor alloys. As shown by Ismail *et al* [**273**], Co is known to detriment the oxidation properties of superalloys as it has been shown to promote the formation of a Co_3O_4 outermost layer, which is mechanically unstable and non-protective, as it spalls off the surface. Additions of Ti, as observed here, tend to detriment oxidation performance of superalloys. This is usually attributed to a destabilisation and/or addition of vacancies to stable, protective Cr_2O_3 layers [**278**], which have

Addition:	(0 Ti, 0 Nb)	Base (0.3 Ti, 0 Nb)	Base + 0.5 Nb	Base + 1 Nb	Base + 2 Nb	Base + 1 Ti	Base + 2 Ti
Al	0.71	0.70	0.69	0.67	0.63	0.67	0.63
W	0.20	0.19	0.19	0.18	0.17	0.18	0.17
Ti	0.00	0.02	0.00	0.00	0.00	0.07	0.13
Nb	0.00	0.00	0.03	0.07	0.13	0.00	0.00
Ta	0.09	0.09	0.09	0.08	0.08	0.08	0.08

Table 4-4: Nominal ɣ′ Al site occupancies for the considered alloys, given full partitioning to the preferred phase.

been shown to be present in the V208 series of Co-base superalloys [**272**], [**273**]. In contrast, this work verifies that the unexpected improvement in oxidation resistance provided by Nb, as observed by Christofidou *et al* [**251**], holds true for the V208 Co/Ni-base superalloy series. This was attributed to stabilisation of the Cr_2O_3 protective layer, and a hypothesised reduction in the solubility of O in the matrix. Limited diffusion of O at the metal/oxide interface in this way was used to rationalise the much reduced ingress depth of Al_2O_3 'fingers', an observation observed in the results presented here as well.

Finally, the room and high temperature yield strengths of the presented alloys are comapred to the state-of-the-art. As previously discussed, additions of Nb and Ti improve the yield and ultimate strengths at both room and high temperature, with the +2Ti alloy achieving > 1 GPa RT yield strength. The comparison to the CoWAlloy series, in which compression testing was published, was performed by using compression/tension asymmetry ratio from equivalent temperature comparisons of Udimet720-Li polycrystalline superalloy, published in the same paper [**10**].

The results presented here agree with the findings of Shang *et al* [**279**], who adopt a first-principles approach to compute the influence of 26 alloying elements to Ni-base superalloys at finite temperature. All decrease the SFE, which will lead to improved forest strengthening of the matrix phase. Additions of Ti are predicted to have the greatest effect, and the effect of Ti is observed to be greater than that of Nb. After Collins & Stone [**280**] the gains in strength can be considered to be a function of enhanced weak/strong pair coupling between propagating dislocations and ɣ′ precipitates. An increase in APB energy corresponds to more strongly coupled superpartial dislocation pairs, and a subsequent increase in the required critical resolved shear stress (CRSS) to propagate the system. Chandran & Sondhi [**281**] have shown using density functional theory methods that both Nb and Ti raise the APB energy on the {111} plane. In the pure Ni_3Al case, an APB energy of 181 mJ m^{-2} was calculated, rising to over 600 mJ m^{-2} for a 0.5 Ti occupation of the atomic site. Nb saw similar, but slightly lower, enhancements. Notably, it appears that the rate of increase in APB energy for Ti additions is greater than for Nb at low (< 0.2) Al site fractional occupancies. As presented in **Table 4-4**, the addition of +2X to the Al site, all else remaining equal, leads to an occupancy of 0.13. At this level Ti is expected to have a greater effect than Nb on the {100} APB energy, at 0K. Correspondingly the +2Ti alloy sees the greatest strength improvements. At higher temperature it is known that the {100} slip system begins to activate, initially leading to Kear-Wilsdorf locking but eventually strength loss. It seems likely that a similar trend exists for increased APB energy in Ti and Nb containing systems, given the trends in strength improvements at high temperature that have been observed.

6. Conclusions

In this work seven new V208-derivative Co/Ni-base superalloys are presented, with systematically varied Nb and Ti content. Their physical properties have been characterised: ɣ′ solvus, density, oxidation resistance, and room/high-temperature yield strength. The alloys are comparable to state-of-the-art gas turbine Ni-base superalloys, but with a substantially lower ɣ′ solvus. Specifically:

- Nb and Ti improve the room and high-temperature yield strength. The alloy with +2Ti achieved > 1 GPa for room temperature yield. Ti appears to raise the strength to a greater extent than Nb, likely due to its stronger effect on APB energy.
- Nb and Ti raise the ɣ′ solvus temperature, but even the most significant additions maintain the desirable processing window provided by Co-base superalloys (< 1100˚C).
- Ti is detrimental to oxidation performance, while Nb improves it. There is much reduced Al_2O_3 intrusion in the Nb-containing alloys.
- Ti additions reduce the density, while Nb additions increase it.

7. Chapter Acknowledgements

This work was a collaborative effort, involving Dr Ioannis Bantounas, Beiya Wang, Xinyue Zhang, as well as TPM and supervised by DD.

TPM, IB, and DD would like to acknowledge support from the Rolls-Royce plc - EPSRC Strategic Partnership in Structural Metallic Systems for Gas Turbines (EP/M005607/1). TPM and DD acknowledge the Centre for Doctoral Training in Advanced Characterisation of Materials (EP/L015277/1) at Imperial College London. DD acknowledges funding from the Royal Society for his industry fellowship with Rolls-Royce plc.

Thesis discussion

This thesis had dual primary objectives, as introduced in the **Preamble**: (1) to develop and validate modern, correlative approaches to microstructural characterisation, and (2) to employ them to develop materials with improved thermal capability. **Chapters 1** and **3** focus on (1), while **2** and **4** mainly consider (2). The microstructural optimisations presented in **Chapters 2** and **4** would not have been possible without the rapid and robust characterisation developed earlier in the thesis. A detailed discussion follows.

Using modern electron microscopy techniques, such as EBSD and EDS, we are able to collect large amounts of data over wide AOIs that detail the structural and chemical properties of microstructural constituents. Such techniques allow rapid validation of alloy design strategies, and can be compared to thermodynamic modelling to adjust nominal composition, heat treatment, and processing conditions to achieve the desired microstructure. This thesis has focussed on the development of post-processing methodologies to facilitate rapid phase identification (in **Chapters 1**, **3**, as well as **Appendices 1**, **2**), as well as employing them to optimise grain boundary character (**Chapter 2**), and screen the microstructure of a new set of high strength alloys (**Chapter 4**).

Unsupervised machine learning is invaluable for efficient access to latent features in microscopy data. Regardless of the specific signals that have been measured, indpendent microstructural constituents interact differently with the rastered electron beam. A latent information space contains the identity and engineering-relevant characteristics (structure, chemistry) of these features. The methods in this work demonstrate how to access this latent space, and how it can be leveraged for data reduction and efficient composition screening. The information channels of particular interest in this work are of course the electron backscatter diffraction pattern, a function of structure and orientation (see **Section LR.3.1**), and stimulated X-ray spectra. These are correlated to each other due to both being functions of the latent space (structure, orientation, chemistry) that we with to access. For this reason, PCA is a natural approach to identifying uncorrelated (to first order) factors. As discussed in **Section LR.4.2** and **Chapter 1**, this approach infers the unit eigenvectors of the (symmetric) covariance matrix, which are necessarily orthogonal, which is equivalent to the SVD of an observation-by-feature data matrix. When combining EBSD and EDS signals, careful treatment of the intra-observation variance is essential, otherwise the EBSD signal (due to having many more EBSP pixels than EDS energy bins) will dominate the covariance matrix and EDS will be essentially irrelevant. This effect can be further leveraged to bias the decomposition in favour of EDS information, which can lead to improved signal to noise. In **Chapter 1** this is explored in depth, and it is observed that spatial resolution is lost as the decomposition is biased in favour of EDS due to the larger interaction volume of electrons for this measurement than EBSP formation. Subsequently, a VARIMAX rotation is employed as a boundary condition for which set of orthogonal EBSP vectors should be regarded ‘characteristic’ of the microstructure. Rather than taking the set which lead to maximal variance for a given reconstruction rank (after the Eckart-Young theorem, see **Section LR.4.2**), finding the set which are most similar in variance to each other yields physically meaningful results, as this is condition is expected. This is because all EBSPs have many (and approximately similar numbers) of overlapping band, and measured patterns are normalised to unit variance. For a visual comparison of PCA, PCA+VARIMAX, and the NLPAR approach of Brewick *et al* [**129**], see **Figure LR-19**.

Having identified these rotated characteristic EBSPs (RC-EBSPs), each representative of a microstructural domain and coupled with a characteristic spectrum (RC-Spectra), modern post-processing methods for structure and orientation

determination can be employed. The data reduction performed by PCA facilitates faster refined template matching, as developed by Foden *et al* [**137**], but with a smaller set of patterns to index. The compute time gained by reducing the set of patterns to index can be repurposed to template matching with a much greater number of phases. In **Chapter 2**, the RC-EBSPs are template matched with : FCC Co/Ni matrix; NiAl and Co_3W intermetallics; M_6C, $M_{23}C_6$ and MC carbides; MB, M_3B_2 and M_5B_3 borides; eta (Ni_3Ti), sigma, mu and P topologically close packed phases. Template matching to such a range of structures (some of very low symmetry) is only possible on the PCA reduced dataset. The improvement in signal-to-nosie facilitated by PCA also permits template matching of microstructural regions that only weakly diffract, such as borides. In **Chapter 2**, this approach was used to screen grain boundary precipitation over a wide region of composition space, to identify the point at which M_6C carbides would no longer precipitate. This point was then used as the 'base' composition for subsequent alloying to improve thermomechanical performance in **Chapter 4**, through Nb and Ti additions. The PCA approach was also used in **Chapter 4** to validate that no additional grain boundary phases had precipitated, other than the benign MC carbide. Improvements in strength and oxidation resistance were achieved, leading to the submission of a patent for a new Co/Ni-base superalloy.

It is hoped that the robust microstructural characterisation pipeline developed in this thesis can be used for similar analyses to those presented in this work as a matter of routine, as an easy to use tool for understanding phase character (structure and chemistry) in a host of engineering applications. For example, Dessolier *et al* have leveraged the approach for understanding carbide and G-phase precipitation in alloys relevenat to the oil & gas industry [**282**].

A particularly challenging task in EBSD is the separation of $L1_2$ γ' precipitates from the FCC γ matrix. Separating these precipitates, which are of extreme engineering significance (see **Section LR.2**), is impossible by eye and requires a very sensitive analysis. In **Chapter 3** three unsupervised machine learning approaches are employed to tackle this problem: PCA, NMF, and autoencoder neural networks. Unlike the polycrystalline case of the previously discussed analyses, here a single crystal region was analysed. For this reason a VARIMAX rotation was *not* employed, as we were trying to draw out the very subtle differences in band contrast between the two structures, and a maximally variance-efficient decomposition was required. Performing an inverse gnomonic projection (as described at **Section LR.3.2.4**) permits easier visualisation of the differences and features present in the latent factors. All three approaches were able to access a sensible latent space, with NMF doing the best job of explicitly identifying a factor representative of the difference in crystal structure, as required. This correlated with its minimal error in a derived contrast metric for quantifying the difference in class-average diffraction pattern for specific crystallographic plane projection / Kikuchi band profiles. In this chapter 'spherical-angular dark field imaging' is also presented, which enables imaging and visualisation of specific crystallographic Kikuchi features.

In the appendices to this thesis, specific aspects of the analyses presented in the preceding chapters are discussed. These are included to briefly expand on interesting topics not deserving of a full chapter. A significant effort of the work for Chapter 3 was in developing an algorithm to very accurately obtain the pattern centre and crystal orientation, which are needed for the inverse gnomonic transform and band integration. **Appendix 1** details this algorithm in detail, and presents some refinement tests and analysis of its convergence in a supervised setting. **Appendix 2** develops a Bayesian representation of the cross-correlation peak height, in which the statistic is treated as a member of the population of all cross-correlations in an SO(3) library search. Modelling the distribution function as lognormal is validated, and an approach is developed that considers how likely it is that a given XCF peak height is *randomly* as high as it is. It is

suggested that such an approach may be helpful for determining a cut-off value for acceptability of RTM indexing, as cross-correlation will always provide a solution and its quality is not immediately apparent.

Finally, a slightly different microstructural characterisation is conducted in **Appendix 3**, as the lattice strain state surrounding a TWIP-steel twin is examined. The 4D-STEM approach identifies transmission electron diffraction spots and tracks their movement across the AOI, which permits evaluation of the local strain state and its variation across microstructural features of interest. In **Appendix 3** the lattice strain resolved in crystallographically significant directions is evaluated and compared to an analytical model.

Thesis conclusions

The dual objectives of this thesis, as discussed in the **Preamble**, were: (1) to develop new approaches for the quantitative characterisation of metal alloy microstructures, and (2) employ them to produce new alloys with optimised grain boundary character and high temperature strength. Given this context, the following conclusions are drawn from the presented work:

- In **Chapter 1** a new characterisation pipeline, combining EBSD and EDS, leveraging unsupervised machine learning to learn latent features (each generally aligned with a microstructural grain or sub-grain), was developed.. The approach enables observation of fine grain boundary phases, such as carbides, and an investigation into weighting between EBSD and EDS information yielded insight into optimal signal-to-noise for subsequent phase identification with template matching. The number of principal components to retain, an extremely important parameter, was identified by specifying a lower limit to dataset variance contribution (equivalently, eigenvalue of the covariance matrix).
- In **Chapter 2**, this method was employed to better understand grain boundary precipitation of carbides and borides in Co/Ni-base superalloys, and compared to ThermoCalc modelling. Subpixel cross-correlation of latent EBSPs with simulated templates, sampled from the SO(3) orientation group, permit accurate determination of phase and orientation. This was applied to thirteen alloys with varied C, B, Zr and Ti content, as well as two with lower Cr:Al ratio, V208C, and RR1000. The main alloy series saw variation in M_6C and MC carbide precipitation, and M_2B borides at higher B levels. NiAl was observed in the reduced Cr:Al alloys. RR1000 only exhibited precipitation of MC and M_3B_2. The carbides in the main alloy series were primarily enriched in Mo and W, but with Mo exhibiting a preference for M_2B when that precipitate was present.
- **Chapter 3** saw application of similar ideas to the difficult task of characterising $L1_2$ γ′ precipitates. Unsupervised machine learning (PCA, NMF, autoencoders) is capable of separating the two structures on the fine-grained differences in band contrast. These can be visualised after applying an inverse gnomonic projection to generate a spherical representation of the EBSPs, from which cross-sectional band profiles can be integrated. This also permits calculation of a form of virtual crystallographic imaging we call ‘spherical-angular dark field’ microscopy to map intensity of specific Kikuchi band features. All three ML approaches are capable of finding a latent that explicitly aligns with crystallographic ordering. We derive a normalised contrast metric to observe the difference between γ and γ′ classifications. Of the three approaches, NMF provides the lowest error in this metric, providing an argument for superior classification capability, arguably rationalised by its physical (detector counting) justification as a non-negative basis.
- In **Chapter 4**, the alloys developed in **Chapter 2** were taken further with additions of Nb and Ti for high temperature strength. The physical properties of these were characterised using the techniques developed in **Chapter 1**, and compared to the state-of-the-art. This new generation of Co/Ni-base superalloys saw improvements in strength, oxidation resistance, and thermal processing window.

Conclusions from complementary work in the appendices may also be drawn:

- In **Appendix 1**, the algorithm used in **Chapter 3** for EBSP pattern centre refinement was explicitly highlighted and evaluated further. It was found to be able to converge to an accurate orientation / PC solution at 15˚ of initial Euler angle misorientation.
- **Appendix 2** presents a novel normalisation procedure for the cross-correlation function used in refined template matching was presented. This permits evaluation of the suitability of an orientation solution, but it was found that the raw cross-correlation peak height is a better metric for absolute template similarity.
- In **Appendix 3**, the micromechanical stress state surrounding a TWIP steel twin was characterised using a novel TEM method. This was compared to an analytical model, and the implications for future steel design were considered.

Aside from papers and preprints, this thesis has resulted in several other tangible outputs. The alloys developed in **Chapter 4** lead to the filing of a patent in collaboration with Rolls-Royce PLC, for a new Co/Ni-base superalloy composition range.The PCA/RTM pipeline developed for **Chapter 1** and **Chapter 2** is now publicly available as open source software, implemented in *MATLAB*, as part of the *AstroEBSD* structure and orientation analysis package - github.com/benjaminbritton/AstroEBSD. The *Python* package *ebspy* was developed for handling, preparation, and normalisation of EBSD data for further machine learning, available at github.com/tmcaul/ebspy. Spherical-angular dark field imaging, as used in **Chapter 3**, is publicly available at github.com/tmcaul/SphericalAngularDF. Finally, a series of generic *mtex* orientation analysis codes and pipelines is available at github.com/tmcaul/tpm_EBSD

References

[**1**] M. Knop *et al.*, "A new polycrystalline Co-Ni superalloy," *JOM*, vol. 66, no. 12, pp. 2495–2501, 2014.

[**2**] M. C. C. Hardy, B. Zirbel, G. Shen, R. Shenkar, and R. Shankar, "Developing damage tolerance and creep resistance in a high strength nickel alloy for disc applications," in *Superalloys 2004*, 2004, pp. 83–90.

[**3**] T. J. A. Slater *et al.*, "Multiscale correlative tomography: An investigation of creep cavitation in 316 stainless steel," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.

[**4**] F. Ram *et al.*, "On the origin of creep dislocations in a Ni-base, single-crystal superalloy: An ECCI, EBSD, and dislocation dynamics-based study," *Acta Mater.*, vol. 109, pp. 151–161, 2016.

[**5**] D. S. Gianola, T. Ben Britton, and S. Zaefferer, "New techniques for imaging and identifying defects in electron microscopy," pp. 450–458, 2019.

[**6**] R. C. Reed, *The Superalloys: Fundamentals and Applications*. 2006.

[**7**] R. W. Kozar, A. Suzuki, W. W. Milligan, J. J. Schirra, M. F. Savage, and T. M. Pollock, "Strengthening mechanisms in polycrystalline multimodal nickel-base superalloys," *Metall. Mater. Trans. A*, vol. 40, no. 7, pp. 1588–1603, 2009.

[**8**] J. Sato, T. Omori, K. Oikawa, I. Ohnuma, R. Kainuma, and K. Ishida, "Cobalt-base high-temperature alloys," *Science (80-.).*, vol. 312, no. April, pp. 90–92, 2006.

[**9**] A. Suzuki, G. C. DeNolf, and T. M. Pollock, "Flow stress anomalies in γ/γ' two-phase Co-Al-W-base alloys," *Scr. Mater.*, vol. 56, pp. 385–388, 2007.

[**10**] S. Neumeier, L. P. Freund, and M. Göken, "Novel wrought γ/γ' cobalt base superalloys with high strength and improved oxidation resistance," *Scr. Mater.*, vol. 109, pp. 104–107, 2015.

[**11**] M. Knop, "Development of Co-Ni Superalloys," Imperial College London, 2015.

[**12**] M. Knop, V. A. Vorontsov, M. C. Hardy, and D. Dye, "High-temperature γ (FCC)/ γ ′ (L1 2) Co-Al-W based superalloys," *MATEC Web Conf.*, vol. 14, p. 18003, 2014.

[**13**] P. Mulvey, "Alloying Effects and the Development of Co-Ni Superalloys," Imperial College London, 2017.

[**14**] V. A. Vorontsov, J. S. Barnard, K. M. Rahman, H. Y. Yan, P. A. Midgley, and D. Dye, "Coarsening behaviour and interfacial structure of γ′ precipitates in Co-Al-W based superalloys," *Acta Mater.*, vol. 120, pp. 14–23, 2016.

[**15**] H.-Y. Yan, V. A. Vorontsov, and D. Dye, "Alloying effects in polycrystalline γ′ strengthened Co–Al–W base alloys," *Intermetallics*, vol. 48, pp. 44–53, 2014.

[**16**] H. Y. Yan, V. A. Vorontsov, J. Coakley, N. G. Jones, H. J. Stone, and D. Dye, "Quaternary alloying effects and the prospects for a new generation of Co-base superalloys," in *Superalloys 2012*, 2012, pp. 705–714.

[**17**] H. Y. Yan, J. Coakley, V. A. Vorontsov, N. G. Jones, H. J. Stone, and D. Dye, "Alloying and the micromechanics of Co-Al-W-X quaternary alloys," *Mater. Sci. Eng. A*, vol. 613, pp. 201–208, 2014.

[**18**] C. H. Zenk *et al.*, "Intermediate Co/Ni-base model superalloys - Thermophysical properties, creep and oxidation," *Scr. Mater.*, vol. 112, pp. 83–86, 2016.

[**19**] A. Bauer, S. Neumeier, F. Pyczak, and M. Göken, "Creep strength and microstructure of polycrystalline γ′ - strengthened cobalt-base superalloys," in *Superalloys 2012*, 2012, pp. 695–703.

[**20**] B. Décamps *et al.*, "On the shearing mechanism of gamma prime precipitates by a single a/6[112] Shockley partial in Ni-based superalloys," *Philos. Mag.*, vol. 84, no. 1, pp. 91–107, 2004.

[**21**] T. M. Pollock and A. S. Argon, "Intermediate temperature creep deformation in CMSX-3 single crystals," in *Superalloys 1988*, 1988, p. 285.

[**22**] V. A. Vorontsov, C. M. F. Rae, C. Shen, Y. Wang, and D. Dye, "Shearing of γ' precipitates by a<112> dislocation ribbons in Ni-base superalloys: A phase field approach," *Acta Mater.*, vol. 58, no. 12, pp. 4110–4119, 2010.

[**23**] B. H. Kear, J. M. Oblak, and A. F. Giamei, "Stacking faults in gamma prime Ni$_3$(Al,Ti) precipitation hardened nickel-base alloys," *Metall. Trans.*, vol. 1, no. 9, pp. 2477–2486, 1970.

[**24**] B. H. Kear, A. F. Giamei, G. R. Leverant, and J. M. Oblak, "On intrinsic/extrinsic stacking fault pairs in the L1$_2$ lattice," *Scr. Metall.*, vol. 3, no. 5, pp. 123–130, 1969.

[**25**] W. W. Milligan and S. D. Antolovich, "The mechanisms and temperature-dependence of superlattice stacking-

fault formation in the single-crystal superalloy PWA-1480," *Metall. Trans. A*, vol. 22, no. 10, pp. 2309–2318, 1991.

[**26**] Q. Z. Chen and D. M. Knowles, "Mechanism of <112>/3 slip initiation and anisotropy of γ phase in CMSX-4 during creep at 750°C and 750 MPa," *Mater. Sci. Eng. A*, vol. 356, no. 1–2, pp. 352–367, 2003.

[**27**] P. Zhang, Y. Yuan, B. Li, S. W. Guo, G. X. Yang, and X. L. Song, "Tensile deformation behavior of a new Ni-base superalloy at room temperature," *Mater. Sci. Eng. A*, vol. 655, pp. 152–159, 2016.

[**28**] C. M. F. Rae and R. C. Reed, "Primary creep in single crystal superalloys: Origins, mechanisms and effects," *Acta Mater.*, vol. 55, pp. 1067–1081, 2007.

[**29**] G. R. Leverant and B. H. Kear, "The mechanism of creep in gamma prime precipitation-hardened nickel-base alloys at intermediate temperatures," *Metall. Mater. Trans. B*, vol. 1, no. 2, pp. 491–498, 1970.

[**30**] C. M. F. Rae, N. Matan, and R. C. Reed, "The role of stacking fault shear in the primary creep of [001]-oriented single crystal superalloys at 750°C and 750 MPa," *Mater. Sci. Eng. A*, vol. 300, no. 1–2, pp. 125–134, Feb. 2001.

[**31**] W. W. Milligan and S. D. Antolovich, "Yielding and deformation behavior of the single crystal superalloy PWA-1480," *Metall. Trans. A*, vol. 18, no. 1, pp. 85–95, 1987.

[**32**] R. R. Unocic, G. B. Viswanathan, P. M. Sarosi, S. Karthikeyan, J. Li, and M. J. Mills, "Mechanisms of creep deformation in polycrystalline Ni-base disk superalloys," *Mater. Sci. Eng. A*, vol. 483–484, no. 1-2 C, pp. 25–32, 2008.

[**33**] L. Kovarik, R. R. Unocic, J. Li, and M. J. Mills, "The intermediate temperature deformation of Ni-based superalloys: Importance of reordering," *JOM*, vol. 61, no. 2, pp. 42–48, 2009.

[**34**] C. Rae, V. Vorontsov, L. Kovarik, and M. Mills, "Dislocations in a Ni-based superalloy during low temperature creep," in *MATEC Web of Conferences*, 2014, vol. 14, p. 01006.

[**35**] M. S. Titus, Y. M. Eggeler, A. Suzuki, and T. M. Pollock, "Creep-induced planar defects in L1$_2$-containing Co- and CoNi-base single-crystal superalloys," *Acta Mater.*, vol. 82, pp. 530–539, 2015.

[**36**] G. B. Viswanathan, P. M. Sarosi, D. H. Whitis, and M. J. Mills, "Deformation mechanisms at intermediate creep temperatures in the Ni-base superalloy René 88DT," *Mater. Sci. Eng. A*, vol. 400–401, pp. 489–495, 2005.

[**37**] P. M. Sarosi, G. B. Viswanathan, and M. J. Mills, "Direct observation of an extended complex stacking fault in the γ' phase of a Ni-base superalloy," *Scr. Mater.*, vol. 55, no. 8, pp. 727–730, 2006.

[**38**] L. P. Freund, O. M. D. M. Messé, J. S. Barnard, M. Göken, S. Neumeier, and C. M. F. Rae, "Segregation assisted microtwinning during creep of a polycrystalline L12-hardened Co-base superalloy," *Acta Mater.*, vol. 123, pp. 295–304, Jan. 2017.

[**39**] R. R. Unocic *et al.*, "Deformation mechanisms in Ni-base disk superalloys at higher temperatures," in *Superalloys 2008*, 2008, pp. 377–385.

[**40**] M. Kolbe, "The high temperature decrease of the critical resolved shear stress in nickel-base superalloys," *Mater. Sci. Eng. A*, vol. 319–321, pp. 383–387, 2001.

[**41**] L. Kovarik *et al.*, "Microtwinning and other shearing mechanisms at intermediate temperatures in Ni-based superalloys," *Prog. Mater. Sci.*, vol. 54, no. 6, pp. 839–873, 2009.

[**42**] T. M. Smith, Y. Rao, Y. Wang, M. Ghazisaeidi, and M. J. Mills, "Diffusion processes during creep at intermediate temperatures in a Ni-based superalloy," *Acta Mater.*, vol. 141, pp. 261–272, 2017.

[**43**] Y. M. Eggeler, M. S. Titus, A. Suzuki, and T. M. Pollock, "Creep deformation-induced antiphase boundaries in L1$_2$-containing single-crystal cobalt-base superalloys," *Acta Mater.*, vol. 77, pp. 352–359, 2014.

[**44**] N. Matan, D. C. Cox, P. Carter, M. a. Rist, C. M. F. Rae, and R. C. Reed, "Creep of CMSX-4 superalloy single crystals: effects of misorientation and temperature," *Acta Mater.*, vol. 47, no. 5, pp. 1549–1563, 1999.

[**45**] H. Y. Yan, V. A. Vorontsov, and D. Dye, "Effect of alloying on the oxidation behaviour of Co-Al-W superalloys," *Corros. Sci.*, vol. 83, pp. 382–395, 2014.

[**46**] K. Tanaka, M. Ooshima, N. Tsuno, A. Sato, and H. Inui, "Creep deformation of single crystals of new Co-Al-W-based alloys with fcc/L1$_2$ two-phase microstructures," *Philos. Mag.*, vol. 92, no. 32, pp. 4011–4027, 2012.

[**47**] J. Coakley, R. C. Reed, J. L. W. Warwick, K. M. Rahman, and D. Dye, "Lattice strain evolution during creep in single-crystal superalloys," *Acta Mater.*, vol. 60, pp. 2729–2738, 2012.

[**48**] J. W. W. Edington, K. N. N. Melton, and C. P. P. Cutler, "Superplasticity," *Prog. Mater. Sci.*, vol. 21, no. 1–2, pp. 61–170, 1976.

[**49**] J. Don and S. Majumdar, "Creep cavitation and grain boundary structure in type 304 stainless steel," *Acta Metall.*, vol. 34, no. 5, pp. 961–967, 1986.

[**50**] I. W. Chen and A. S. Argon, "Creep cavitation in 304 stainless steel," *Acta Metall.*, vol. 29, no. 7, pp. 1321–1333, 1981.

[**51**] A. Ma, D. Dye, and R. C. Reed, "A model for the creep deformation behaviour of single-crystal superalloy CMSX-4," *Acta Mater.*, vol. 56, pp. 1657–1670, 2008.

[**52**] F. A. Mohamed, "Micrograin superplasticity: Characteristics and utilization," *Materials (Basel).*, vol. 4, no. 7, pp. 1194–1223, 2011.

[**53**] I. M. Lifshitz, "On the theory of diffusion-viscous flow of polycrystalline bodies," *Sov. Phys. JETP*, vol. 17, no. 44, pp. 1349–1367, 1963.

[**54**] W. A. Rachinger, "Relative grain translations in the plastic flow of aluminium," *J. Inst. Met.*, vol. 81, pp. 33–41, 1952.

[**55**] O. D. Sherby and J. Wadsworth, "Superplasticity-Recent advances and future directions," *Prog. Mater. Sci.*, vol. 33, no. 3, pp. 169–221, 1989.

[**56**] T. G. Langdon, "Grain boundary sliding as a deformation mechanism during creep," *Philos. Mag.*, vol. 22, no. 178, pp. 689–700, 1970.

[**57**] M. F. Ashby and R. A. Verrall, "Diffusion-accommodated flow and superplasticity," *Acta Metall.*, vol. 21, no. 2, pp. 149–163, 1973.

[**58**] T. G. Langdon, "A unified approach to grain boundary sliding in creep and superplasticity," *Acta Metall. Mater.*, vol. 42, no. 7, pp. 2437–2443, 1994.

[**59**] R. C. Gifkins, "Grain-boundary sliding and its accommodation during creep and superplasticity," *Metall. Trans. A*, vol. 7, no. 7, pp. 1225–1232, 1976.

[**60**] A. Ball and M. M. Huchinson, "Superplastic behaviour of rapidly solidified Al-5Mg-1.2Cr alloy," *J. Mater. Sci.*, vol. 3, no. 1, 1969.

[**61**] A. Mukherjee, J. Bird, and J. Dorn, "Experimental correlations for high-temperature creep," *ASTM Trans. Q.*, vol. 62, no. 1, pp. 155–179, 1968.

[**62**] R. C. Gifkins, "A Mechanism for the Formation of Intergranular Cracks when Boundary Sliding Occurs," *Acta Metall.*, vol. 4, pp. 98–99, 1956.

[**63**] M. E. Kassner and T. A. Hayes, "Creep cavitation in metals," *Int. J. Plast.*, vol. 19, no. 10, pp. 1715–1748, 2003.

[**64**] P. Kontis, A. Kostka, D. Raabe, and B. Gault, "Influence of composition and precipitation evolution on damage at grain boundaries in a crept polycrystalline Ni-based superalloy," *Acta Mater.*, vol. 166, pp. 158–167, 2019.

[**65**] P. Kontis *et al.*, "On the effect of boron on grain boundary character in a new polycrystalline superalloy," *Acta Mater.*, vol. 103, pp. 688–699, 2016.

[**66**] P. Kontis, S. Pedrazzini, Y. Gong, P. A. J. Bagot, M. P. Moody, and R. C. Reed, "The effect of boron on oxide scale formation in a new polycrystalline superalloy," *Scr. Mater.*, vol. 127, pp. 156–159, 2017.

[**67**] M. Kolb *et al.*, "On the grain boundary strengthening effect of boron in γ/γ′ Cobalt-base superalloys," *Acta Mater.*, vol. 145, pp. 247–254, 2018.

[**68**] T. J. Garosshen, T. D. Tillman, and G. P. Mccarthy, "Effects of B , C , and Zr on the Structure and Properties of a P/M Nickel Base Superalloy," *Metall. Trans. A*, vol. 18A, no. January, pp. 69–77, 1987.

[**69**] A. K. Koul and G. H. Gessinger, "On the mechanism of serrated grain boundary formation in Ni-based superalloys," *Acta Metall.*, vol. 31, no. 7, pp. 1061–1069, 1983.

[**70**] Y. T. Tang, A. J. Wilkinson, and R. C. Reed, "Grain Boundary Serration in Nickel-Based Superalloy Inconel 600: Generation and Effects on Mechanical Behavior," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 49, no. 9, pp. 4324–4342, 2018.

[**71**] S. Xu, J. I. Dickson, and A. K. Koul, "Grain growth and carbide precipitation in superalloy, UDIMET 520," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 29, no. 11, pp. 2687–2695, 1998.

[**72**] S. Karagöz, I. Liem, E. Bischoff, and H. F. Fischmeister, "Determination of carbide and matrix compositions in

high-speed steels by analytical electron microscopy," *Metall. Trans. A*, vol. 20, no. 12, pp. 2695–2701, 1989.

[**73**] E. S. Lee, W. J. Park, J. Y. Jung, and S. Ahn, "Solidification microstructure and M2C carbide decomposition in a spray-formed high-speed steel," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 29, no. 5, pp. 1395–1404, 1998.

[**74**] M. M. Serna and J. L. Rossi, "MC complex carbide in AISI M2 high-speed steel," *Mater. Lett.*, vol. 63, no. 8, pp. 691–693, 2009.

[**75**] M. M. Serna, E. R. B. Jesus, E. Galego, L. G. Martinez, H. P. S. Corrêa, and J. L. Rossi, "An Overview of the Microstructures Present in High-Speed Steel -Carbides Crystallography," *Mater. Sci. Forum*, vol. 530–531, no. January 2014, pp. 48–52, 2009.

[**76**] A. S. Wilson, "Formation and effect of topologically close-packed phases in nickel-base superalloys," *Mater. Sci. Technol.*, vol. 0836, no. May, pp. 1–11, 2016.

[**77**] R. J. Mitchell, M. Preuss, S.Tin, and M. C. Hardy, "The influence of cooling rate from temperatures above the γ′ solvus on morphology, mismatch and hardness in advanced polycrystalline nickel-base superalloys," *Mater. Sci. Eng. A*, vol. 473, no. 1–2, pp. 158–165, 2008.

[**78**] L. Jiang *et al.*, "M2C and M6C carbide precipitation in Ni-Mo-Cr based superalloys containing silicon," *Mater. Des.*, vol. 112, pp. 300–308, 2016.

[**79**] J. Safari and S. Nategh, "On the heat treatment of Rene-80 nickel-base superalloy," *J. Mater. Process. Technol.*, vol. 176, no. 1–3, pp. 240–250, 2006.

[**80**] C. H. Lund and H. J. Wagner, "Identification of microconstituents in superalloys," *Def. Met. Inf. Cent. Memo.*, vol. 160, pp. 1–21, 1962.

[**81**] W. H. Jiang, X. D. Yao, H. R. Guan, and Z. Q. Hu, "Secondary M6C precipitation in a cobalt-base superalloy," *J. Mater. Sci. Lett.*, vol. 18, pp. 303–305, 1999.

[**82**] A. K. Koul and R. Thamburaj, "Serrated grain boundary formation potential of Ni-based superalloys and its implications," *Metall. Trans. A*, vol. 16, no. 1, pp. 17–26, 1985.

[**83**] R. J. Mitchell, H. Y. Li, and Z. W. Huang, "On the formation of serrated grain boundaries and fan type structures in an advanced polycrystalline nickel-base superalloy," *J. Mater. Process. Technol.*, vol. 209, no. 2, pp. 1011–1017, 2009.

[**84**] J. L. W. Carter *et al.*, "Characterization of strain accumulation at grain boundaries of nickel based superalloys," *Superalloys 2012*, no. April 2017, pp. 43–52, 2012.

[**85**] O. Miyagawa, M. Yamamoto, and M. Kobayashi, "Zig-Zag Grain Boundaries and Strength of Heat Resisting Alloys," in *Superalloys 1976*, 1976, pp. 245–254.

[**86**] J. M. Larson, "Carbide morphology in p/m IN-792," *Metall. Trans. A*, vol. 7, no. 9, pp. 1497–1502, 1976.

[**87**] M. Yamazaki, "The Effect of Two-Step Solution Treatment on the Creep Rupture Properties of a High Carbon 18Cr-12Ni Stainless Steel," *J. Japan Inst. Met.*, vol. 30, no. 11, pp. 1032–1036, 1966.

[**88**] J. L. W. Carter, M. W. Kuper, M. D. Uchic, and M. J. Mills, "Characterization of localized deformation near grain boundaries of superalloy René-104 at elevated temperature," *Mater. Sci. Eng. A*, vol. 605, pp. 127–136, 2014.

[**89**] J. L. Walley, R. Wheeler, M. D. Uchic, and M. J. Mills, "In-Situ Mechanical Testing for Characterizing Strain Localization During Deformation at Elevated Temperatures," *Exp. Mech.*, vol. 52, no. 4, pp. 405–416, Apr. 2012.

[**90**] S. T. Wlodek, M. Kelly, and D. A. Alden, "The Structure of Rene' 88DT," pp. 129–136, 2012.

[**91**] M. Gao and R. P. Wei, "Grain boundary niobium carbides in Inconel 718," *Scr. Mater.*, vol. 37, no. 12, pp. 1843–1849, 1997.

[**92**] T. P. Gabb, A. Garg, D. L. Ellis, and K. M. O'Connor, "Detailed Microstructural Characterization of the Disk Alloy ME3," no. May 2004, 2004.

[**93**] T. P. McAuliffe, A. Foden, C. Bilsland, D. Daskalaki-Mountanou, D. Dye, and T. B. Britton, "Advancing characterisation with statistics from correlative electron diffraction and X-ray spectroscopy, in the scanning electron microscope," *Ultramicroscopy*, no. c, p. 112944, Jan. 2020.

[**94**] H. F. Fischmeister, S. Karagöz, and H. O. Andrén, "an Atom Probe Study of Secondary Hardening in High Spped Steels," vol. 36, no. 4, pp. 817–825, 1988.

[**95**] Y. Shen *et al.*, "Sigma phases in an 11\%Cr ferritic/martensitic steel with the normalized and tempered condition," *Mater. Charact.*, vol. 122, pp. 113–123, 2016.

[**96**] Z. W. Huang, H. Y. Li, G. Baxter, S. Bray, and P. Bowen, "Characterization of the Weld Line Zones of an Inertia Friction Welded Superalloy," *Adv. Mater. Res.*, vol. 278, pp. 440–445, 2011.

[**97**] K. L. More and M. K. Miller, "Microstructural Characterization of Udimet 720 : a Nickel-Base Alloy," *Le J. Phys. Colloq.*, vol. 49, no. C6, pp. C6-391-C6-396, 1988.

[**98**] O. A. Ojo and H. R. Zhang, "Analytical electron microscopy study of boron-rich grain boundary microconstituent in directionally solidified RENE 80 superalloy," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 39, no. 12, pp. 2799–2803, 2008.

[**99**] E. J. Pickering, R. Muñoz-Moreno, H. J. Stone, and N. G. Jones, "Precipitation in the equiatomic high-entropy alloy CrMnFeCoNi," *Scr. Mater.*, vol. 113, pp. 106–109, 2016.

[**100**] J. X. Yang, Q. Zheng, X. F. Sun, H. R. Guan, and Z. Q. Hu, "Topologically close-packed phase precipitation in a nickel-base superalloy during thermal exposure," *Mater. Sci. Eng. A*, vol. 465, no. 1–2, pp. 100–108, 2007.

[**101**] H. R. Zhang, O. A. Ojo, and M. C. Chaturvedi, "Nanosize boride particles in heat-treated nickel base superalloys," *Scr. Mater.*, vol. 58, no. 3, pp. 167–170, 2008.

[**102**] R. Vincent, "Precipitation around welds in the nickel-base superalloy, Inconel 718," *Acta Metall.*, vol. 33, no. 7, pp. 1205–1216, 1985.

[**103**] A. García-Bórquez, "TEM studies on the borides formed in an austenitic steel with 38 ppm boron," *Scr. Metall.*, vol. 19, pp. 3257–62117, 1985.

[**104**] I. Goldfarb, W. D. Kaplan, S. Ariely, and M. Bamberger, "Fault-induced polytypism in (Cr, Fe)2B," *Philos. Mag. A Phys. Condens. Matter, Struct. Defects Mech. Prop.*, vol. 72, no. 4, pp. 963–979, 1995.

[**105**] F. X. Kayser and G. F. Kayser, "Re-examination of the Kraft and Flinn diffraction data for Cr2B, (Cr,Fe)2B, and the boride phase in Fe + 18.5 wt % Ni + 20 wt % Cr + B alloys," *J. Mater. Sci.*, vol. 34, no. 6, pp. 1271–1275, 1999.

[**106**] T. Phillips, "X-ray Crystallography and the Fourier Transform," *American Mathematical Society*, Oct. 2011.

[**107**] C. B. Carter and D. B. Williams, *Transmission Electron Microscopy*, Second. Springer, 2009.

[**108**] "Structure factor," *Online dictionary of crystallography*. IUCR, 2017.

[**109**] S. Zaefferer, "A critical review of orientation microscopy in SEM and TEM," *Cryst. Res. Technol.*, vol. 46, no. 6, pp. 607–628, 2011.

[**110**] A. S. Eggeman, R. Krakow, and P. A. Midgley, "Scanning precession electron tomography for three-dimensional nanoscale orientation imaging and crystallographic analysis," *Nat. Commun.*, vol. 6, pp. 1–7, 2015.

[**111**] B. L. Adams, S. I. Wright, and K. Kunze, "Orientation imaging: The emergence of a new microscopy," *Metall. Trans. A*, vol. 24, no. 4, pp. 819–831, 1993.

[**112**] A. J. Wilkinson and T. Ben Britton, "Strains, planes, and EBSD in materials science," *Mater. Today*, vol. 15, no. 9, pp. 366–376, 2012.

[**113**] S. G. Thomson, "High-angle Kikuchi patterns," *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.*, vol. 221, no. 1145, pp. 224–242, Jan. 1954.

[**114**] G. Nolze and A. Winkelmann, "Crystallometric and projective properties of Kikuchi diffraction patterns," *J. Appl. Crystallogr.*, vol. 50, no. 1, pp. 102–119, 2017.

[**115**] H. Yu, J. Liu, P. Karamched, A. J. Wilkinson, and F. Hofmann, "Mapping the full lattice strain tensor of a single dislocation by High Angular Resolution Transmission Kikuchi Diffraction (HR-TKD)," *Scr. Mater.*, vol. 1, no. 1, pp. 36–41, 2019.

[**116**] J. Liu, S. Lozano-Perez, A. J. Wilkinson, and C. R. M. Grovenor, "On the depth resolution of transmission Kikuchi diffraction (TKD) analysis," *Ultramicroscopy*, vol. 205, no. May, pp. 5–12, 2019.

[**117**] T. B. Britton *et al.*, "Tutorial: Crystal orientations and EBSD - Or which way is up?," *Mater. Charact.*, vol. 117, pp. 113–126, 2016.

[**118**] N. C. K. Lassen, D. J. Jensen, and K. Conradsen, "Image processing procedures for analysis of electron back scattering patterns," *Scanning Microsc.*, vol. 6, no. 1, pp. 115–121, 1992.

[**119**] T. B. Britton, V. S. Tong, J. Hickey, A. Foden, and A. J. Wilkinson, "AstroEBSD : exploring new space in

pattern indexing with methods launched from an astronomical approach," *J. Appl. Crystallogr.*, vol. 51, no. 6, pp. 1525–1534, Dec. 2018.

[**120**] S. I. Wright and B. L. Adams, "Automatic analysis of electron backscatter diffraction patterns," *Metall. Trans. A*, vol. 23, no. 3, pp. 759–767, 1992.

[**121**] B. L. Adams, K. Kunze, D. J. Dingley, and S. I. Wright, "Orientation Imaging Microscopy: New possibilities for microstructural investigations using automated BKD analysis," in *International conference on textures of materials*, 1993.

[**122**] A. J. Schwartz, M. Kumar, B. L. Adams, and D. P. Field, *Electron Backscatter Diffraction in Materials Science*. Boston, MA: Springer US, 2009.

[**123**] D. J. Dingley and S. I. Wright, "Determination of crystal phase from an electron backscatter diffraction pattern," *J. Appl. Crystallogr.*, vol. 42, no. 2, pp. 234–241, 2009.

[**124**] L. Li and M. Han, "Determining the Bravais lattice using a single electron backscatter diffraction pattern," *J. Appl. Crystallogr.*, vol. 48, no. 1, pp. 107–115, 2015.

[**125**] A. Winkelmann, G. Nolze, G. Cios, and T. Tokarski, "Mapping of local lattice parameter ratios by projective Kikuchi pattern matching," *Phys. Rev. Mater.*, vol. 2, no. 12, pp. 1–7, 2018.

[**126**] R. Birch, S. Wang, V. S. Tong, and T. B. Britton, "The effect of cooling rate and grain size on hydride microstructure in Zircaloy-4," *J. Nucl. Mater.*, vol. 513, pp. 221–225, 2019.

[**127**] T. B. Britton, "A High Resolution Electron Backscatter Diffraction Study of Titanium and its Alloys," 2010.

[**128**] S. I. Wright, M. M. Nowell, S. P. Lindeman, P. P. Camus, M. De Graef, and M. A. Jackson, "Introduction and comparison of new EBSD post-processing methodologies," *Ultramicroscopy*, vol. 159, no. P1, pp. 81–94, 2015.

[**129**] P. T. Brewick, S. I. Wright, and D. J. Rowenhorst, "NLPAR: Non-local smoothing for enhanced EBSD pattern indexing," *Ultramicroscopy*, vol. 200, no. December 2018, pp. 50–61, 2019.

[**130**] L. N. Brewer, P. G. Kotula, and J. R. Michael, "Multivariate statistical approach to electron backscattered diffraction," *Ultramicroscopy*, vol. 108, no. 6, pp. 567–578, 2008.

[**131**] A. J. Wilkinson, D. M. Collins, Y. Zayachuk, R. Korla, and A. Vilalta-Clemente, "Applications of Multivariate Statistical Methods and Simulation Libraries to Analysis of Electron Backscatter Diffraction and Transmission Kikuchi Diffraction Datasets," *Ultramicroscopy*, vol. 196, no. June 2018, pp. 88–98, 2018.

[**132**] A. Winkelmann, "Dynamical Simulation of Electron Backscatter Diffraction Patterns," *Electron Backscatter Diffr. Mater. Sci.*, 2009.

[**133**] A. Winkelmann and M. Vos, "The role of localized recoil in the formation of Kikuchi patterns," *Ultramicroscopy*, vol. 125, pp. 66–71, 2013.

[**134**] P. G. Callahan and M. De Graef, "Dynamical electron backscatter diffraction patterns. Part I: Pattern simulations," pp. 1255–1265, 2013.

[**135**] A. Winkelmann, "Dynamical effects of anisotropic inelastic scattering in electron backscatter diffraction," *Ultramicroscopy*, vol. 108, no. 12, pp. 1546–1550, 2008.

[**136**] Y. H. Chen *et al.*, "A Dictionary Approach to Electron Backscatter Diffraction Indexing," *Microsc. Microanal.*, vol. 21, no. 3, pp. 739–752, 2015.

[**137**] A. Foden, D. M. Collins, A. J. Wilkinson, and T. B. Britton, "Indexing electron backscatter diffraction patterns with a refined template matching approach," *Ultramicroscopy*, vol. 207, p. 112845, Dec. 2019.

[**138**] P. E. Anuta, "Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques," *IEEE Trans. Geosci. Electron.*, vol. 8, no. 4, pp. 353–368, 1970.

[**139**] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Math. Comput.*, vol. 19, no. 90, p. 297, Apr. 1965.

[**140**] A. J. Wilkinson, G. Meaden, and D. J. Dingley, "High resolution mapping of strains and rotations using electron backscatter diffraction," *Mater. Sci. Technol.*, vol. 22, no. 11, pp. 1271–1278, 2006.

[**141**] A. J. Wilkinson, G. Meaden, and D. J. Dingley, "High-resolution elastic strain measurement from electron backscatter diffraction patterns: New levels of sensitivity," *Ultramicroscopy*, vol. 106, no. 4–5, pp. 307–313, 2006.

[**142**] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Opt.*

Lett., vol. 33, no. 2, p. 156, Jan. 2008.

[**143**] F. Ram, S. Wright, S. Singh, and M. De Graef, "Error analysis of the crystal orientations obtained by the dictionary approach to EBSD indexing," *Ultramicroscopy*, vol. 181, pp. 17–26, 2017.

[**144**] F. Ram and M. De Graef, "Phase differentiation by electron backscatter diffraction using the dictionary indexing approach," *Acta Mater.*, vol. 144, pp. 352–364, 2018.

[**145**] R. Hielscher, F. Bartel, and T. B. Britton, "Gazing at crystal balls: Electron backscatter diffraction pattern analysis and cross correlation on the sphere," *Ultramicroscopy*, vol. 207, p. 112836, Dec. 2019.

[**146**] A. P. Day, "Spherical EBSD," *J. Microsc.*, vol. 230, no. 3, pp. 472–486, Jun. 2008.

[**147**] C. Ophus, "Four-Dimensional Scanning Transmission Electron Microscopy (4D-STEM): From Scanning Nanodiffraction to Ptychography and Beyond," *Microsc. Microanal.*, vol. 25, no. 2019, pp. 563–582, 2019.

[**148**] B. H. Savitzky *et al.*, "py4DSTEM: a software package for multimodal analysis of four-dimensional scanning transmission electron microscopy datasets," pp. 1–32, 2020.

[**149**] P. R. Chidambaram, C. Bowen, S. Chakravarthi, C. Machala, and R. Wise, "Fundamentals of silicon material properties for successful exploitation of strain engineering in modern CMOS manufacturing," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 944–964, 2006.

[**150**] F. Uesugi, A. Hokazono, and S. Takeno, "Evaluation of two-dimensional strain distribution by STEM/NBD," *Ultramicroscopy*, vol. 111, no. 8, pp. 995–998, 2011.

[**151**] P. F. Rottmann and K. J. Hemker, "Nanoscale elastic strain mapping of polycrystalline materials," *Mater. Res. Lett.*, vol. 6, no. 4, pp. 249–254, 2018.

[**152**] T. C. Pekin, C. Gammer, J. Ciston, C. Ophus, and A. M. Minor, "In situ nanobeam electron diffraction strain mapping of planar slip in stainless steel," *Scr. Mater.*, vol. 146, pp. 87–90, 2018.

[**153**] T. C. Pekin, C. Gammer, J. Ciston, A. M. Minor, and C. Ophus, "Optimizing disk registration algorithms for nanobeam electron diffraction strain mapping," *Ultramicroscopy*, vol. 176, pp. 170–176, May 2017.

[**154**] B. H. Savitzky *et al.*, "py4DSTEM: Open Source Software for 4D-STEM Data Analysis," *Microsc. Microanal.*, vol. 25, no. S2, pp. 124–125, 2019.

[**155**] A. Béché, J. L. Rouvière, L. Clément, and J. M. Hartmann, "Improved precision in strain measurement using nanobeam electron diffraction," *Appl. Phys. Lett.*, vol. 95, no. 12, p. 123114, Sep. 2009.

[**156**] V. B. Ozdol *et al.*, "Strain mapping at nanometer resolution using advanced nano-beam electron diffraction," *Appl. Phys. Lett.*, vol. 106, no. 25, p. 253107, 2015.

[**157**] S. E. Zeltmann *et al.*, "Patterned Probes for High Precision 4D-STEM Bragg Measurements," 2019.

[**158**] B. D. Conduit, N. G. Jones, H. J. Stone, and G. J. Conduit, "Design of a nickel-base superalloy using a neural network," *Mater. Des.*, vol. 131, pp. 358–365, Oct. 2017.

[**159**] "Severstal: Steel Defect Detection," *Kaggle*, 2019. https://www.kaggle.com/c/severstal-steel-defect-detection (accessed May 26, 2020).

[**160**] H. Okuyucu, A. Kurt, and E. Arcaklioglu, "Artificial neural network application to the friction stir welding of aluminum plates," *Mater. Des.*, vol. 28, no. 1, pp. 78–84, 2007.

[**161**] R. Mishra, J. Malik, I. Singh, and J. P. Davim, "Neural network approach for estimating the residual tensile strength after drilling in uni-directional glass fiber reinforced plastic laminates," *Mater. Des.*, vol. 31, no. 6, pp. 2790–2795, 2010.

[**162**] D. B. Miracle, J. D. Miller, O. N. Senkov, C. Woodward, M. D. Uchic, and J. Tiley, "Exploration and development of high entropy alloys for structural applications," *Entropy*, vol. 16, no. 1, pp. 494–525, 2014.

[**163**] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013.

[**164**] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge: MIT Press, 2016.

[**165**] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer New York, 2001.

[**166**] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay," pp. 1–21, 2018.

[**167**] J. Howard, "fastai - a layered API for deep learning," 2019. https://docs.fast.ai (accessed May 28, 2020).

[**168**] C. Lemaréchal, “Cauchy and the Gradient Method,” *Doc. Math.*, vol. ISMP, pp. 251–254, 2012.

[**169**] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical Methods for Physics and Engineering*, Third. Cambridge University Press, 2006.

[**170**] L. Bottou, *Large-Scale Machine Learning with Stochastic Gradient Descent*. 2010.

[**171**] N. N. Schraudolph, J. Yu, and S. Gunter, “A Stochastic Quasi-Newton Method for Online Convex Optimization,” in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, 2007, vol. 2, pp. 436–443.

[**172**] N. Schraudolph and T. Graepel, “Combining conjugate direction methods with stochastic approximation of gradients,” *Proc. Ninth Int. Work. Artif. Intell. Stat. AISTATS 2003*, pp. 2–7, 2003.

[**173**] S. ichi Amari, “Backpropagation and stochastic gradient descent method,” *Neurocomputing*, vol. 5, no. 4–5, pp. 185–196, 1993.

[**174**] A. Botev, G. Lever, and D. Barber, “Nesterov’s accelerated gradient and momentum as approximations to regularised update descent,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2017-May, no. 2, pp. 1899–1903, 2017.

[**175**] J. C. Duchi, E. Hazan, and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, 2011.

[**176**] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[**177**] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” no. NeurIPS, 2019.

[**178**] M. Abadi *et al.*, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” 2015.

[**179**] M. Abadi *et al.*, “TensorFlow: A system for large-scale machine learning,” 2015.

[**180**] M. Belkin, D. Hsu, S. Ma, and S. Mandal, “Reconciling modern machine-learning practice and the classical bias–variance trade-off,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 32, pp. 15849–15854, 2019.

[**181**] S. A. A. Kohl *et al.*, “A probabilistic U-net for segmentation of ambiguous images,” *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, no. NeurIPS, pp. 6965–6975, 2018.

[**182**] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, “Algorithms and applications for approximate nonnegative matrix factorization,” *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 155–173, 2007.

[**183**] A. Krizhevsky, “ImageNet Classification with Deep Convolutional Neural,” *Adv. Neural Inf. Process. Syst.*, 2012.

[**184**] R. W. Fleming and K. R. Storrs, “Learning to see stuff,” *Curr. Opin. Behav. Sci.*, vol. 30, pp. 100–108, 2019.

[**185**] M. Leshno and S. Schocken, “Multilayer Feedforward Networks with Non-Polynomial Activation Functions Can Approximate Any Function,” no. 21, pp. 1–16, 1991.

[**186**] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Comput.*, vol. 18, pp. 1527–1554, 2006.

[**187**] N. Srivastava, G. Hinton, A. Sutskever, K. Ilya, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[**188**] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.

[**189**] A. Krogh and J. A. Hertz, “A simple weight decay can improve generalization,” in *Advances in Neural Information Processing Systems 4*, 1991.

[**190**] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” vol. 58, no. 1, pp. 267–288, 1996.

[**191**] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” pp. 1–31, 2016.

[**192**] G. W. Lindsay, “Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future,” 2020.

[**193**] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8689 LNCS, no. PART 1, pp. 818–833, 2014.

[**194**] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *Br. Mach. Vis. Conf. 2016, BMVC 2016*, vol. 2016-Septe, pp. 124.1-124.14, May 2016.

[**195**] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[**196**] J. Donahue *et al.*, “DeCAF: A deep convolutional activation feature for generic visual recognition,” *31st Int. Conf. Mach. Learn. ICML 2014*, vol. 2, pp. 988–996, 2014.

[**197**] K. Pearson, “ LIII. On lines and planes of closest fit to systems of points in space ,” *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.

[**198**] G. E. Hinton and R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks,” *Science (80-.).*, vol. 313, no. July, pp. 504–507, 2006.

[**199**] I. T. Jolliffe and J. Cadima, “Principal component analysis : a review and recent developments,” *Philos. Trans. R. Soc. A*, vol. 374, p. 16, 2016.

[**200**] G. Raîche, T. A. Walls, D. Magis, M. Riopel, and J. G. Blais, “Non-graphical solutions for Cattell’s scree test,” *Methodology*, vol. 9, no. 1, pp. 23–29, 2013.

[**201**] R. Thomas, “Computational Linear Algebra,” *fastai*, 2017. https://www.fast.ai/2017/07/17/num-lin-alg/ (accessed May 31, 2020).

[**202**] S. Roweis, “EM Algorithms for PCA and SPCA,” in *Advances in neural information processing systems*, 1998, vol. 10, pp. 626–632.

[**203**] E. Anderson *et al.*, *LAPACK Users ’ Guide*. OSTI, 1992.

[**204**] G. Golub and W. Kahan, “Calculating the singular values and pseudo-inverse of a matrix,” *J. Soc. Ind. Appl. Math. Ser. B Numer. Anal.*, vol. 2, no. 2, pp. 205–224, 1965.

[**205**] J. Lambers, “Lecture 6 - The SVD Algorithm,” 2010.

[**206**] M. Gu and S. C. Eisenstat, “A Divide-and-Conquer Algorithm for the Bidiagonal SVD,” *SIAM J. Matrix Anal. Appl.*, vol. 16, no. 1, pp. 79–92, Jan. 1995.

[**207**] J. C. Nash and S. Shlien, “Simple Algorithms for the Partial Singular Value Decomposition,” *Comput. J.*, vol. 30, no. 3, pp. 268–275, 1987.

[**208**] D. D. Lee and H. S. Seung, “Learning the parts of objects by nonnegative matrix factorization,” *Nature*, vol. 401, pp. 788–791, 1999.

[**209**] C. Eckart and G. Young, “The approximation of one matrix by another of lower rank,” *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep. 1936.

[**210**] N. Gillis, “The Why and How of Nonnegative Matrix Factorization,” pp. 1–25, 2014.

[**211**] H. Valpola, “From neural PCA to deep unsupervised learning,” *Adv. Indep. Compon. Anal. Learn. Mach.*, pp. 143–171, 2015.

[**212**] I. Higgins *et al.*, “Early Visual Concept Learning with Unsupervised Deep Learning,” 2016.

[**213**] S. Wold, K. Esbensen, and P. Geladi, “Principal Component Analysis,” *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, 1987.

[**214**] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, no. Ml, pp. 1–14, 2014.

[**215**] A. Bhattacharya, C. M. Parish, J. Henry, and Y. Katoh, “High throughput crystal structure and composition mapping of crystalline nanoprecipitates in alloys by transmission Kikuchi diffraction and analytical electron microscopy,” *Ultramicroscopy*, vol. 202, no. March, pp. 33–43, Jul. 2019.

[**216**] C. J. Kuehmann and G. B. Olson, “Computational materials design and engineering,” *Mater. Sci. Technol.*, vol. 25, no. 4, pp. 472–478, 2009.

[**217**] I. Toda-Caraballo, E. I. Galindo-Nava, and P. E. J. Rivera-Díaz-Del-Castillo, “Unravelling the materials genome: Symmetry relationships in alloy properties,” *J. Alloys Compd.*, vol. 566, pp. 217–228, 2013.

[**218**] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, “Machine learning in materials informatics: Recent applications and prospects,” *npj Comput. Mater.*, vol. 3, no. 1, 2017.

[**219**] M. Joo, J. Ryu, and H. K. D. H. Bhadeshia, “Domains of steels with identical properties,” *Mater. Manuf. Process.*, vol. 24, no. 1, pp. 53–58, 2009.

[**220**] G. Papadopoulos, P. J. Edwards, and A. F. Murray, “Confidence estimation methods for neural networks: A

practical comparison," *IEEE Trans. Neural Networks*, vol. 12, no. 6, pp. 1278–1287, 2001.

[**221**] C. Wen *et al.*, "Machine learning assisted design of high entropy alloys with desired property," *Acta Mater.*, vol. 170, pp. 109–117, 2019.

[**222**] "Automatic FLOW for Materials Discovery," *Duke University*. http://www.aflowlib.org (accessed Jun. 02, 2020).

[**223**] A. Jain *et al.*, "Commentary: The materials project: A materials genome approach to accelerating materials innovation," *APL Mater.*, vol. 1, no. 1, 2013.

[**224**] "NOMAD," *Fritz-Haber-Institut der Max-Planck-Gesellschaft*. https://nomad-coe.eu (accessed Jun. 02, 2020).

[**225**] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, "New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design," *Acta Crystallogr. Sect. B Struct. Sci.*, vol. 58, no. 3, pp. 364–369, Jun. 2002.

[**226**] S. Gražulis *et al.*, "Crystallography Open Database – an open-access collection of crystal structures," *J. Appl. Crystallogr.*, vol. 42, no. 4, pp. 726–729, Aug. 2009.

[**227**] A. van de Walle and G. Ceder, "Automating first-principles phase diagram calculations," *J. Phase Equilibria*, vol. 23, no. 4, pp. 348–359, 2002.

[**228**] T. Mueller and G. Ceder, "Bayesian approach to cluster expansions," *Phys. Rev. B - Condens. Matter Mater. Phys.*, vol. 80, no. 2, pp. 1–13, 2009.

[**229**] M. Watanabe *et al.*, "Improvements in the X-ray analytical capabilities of a scanning transmission electron microscope by spherical-aberration correction," *Microsc. Microanal.*, vol. 12, no. 6, pp. 515–526, 2006.

[**230**] X. Zhang, X. Qi, M. Zou, and F. Liu, "Rapid authentication of olive oil by Raman spectroscopy using principal component analysis," *Anal. Lett.*, vol. 44, no. 12, pp. 2209–2220, 2011.

[**231**] S. Wold and O. H. J. Christie, "Extraction of mass spectral information by a combination of autocorrelation and principal components models," *Anal. Chim. Acta*, vol. 165, no. C, pp. 51–59, 1984.

[**232**] C. M. Parish, G. L. Brennecka, B. A. Tuttle, and L. N. Brewer, "Quantitative X-Ray spectrum imaging of lead lanthanum zirconate titanate PLZT thin-films," *J. Am. Ceram. Soc.*, vol. 91, no. 11, pp. 3690–3697, 2008.

[**233**] C. M. Parish and L. N. Brewer, "Multivariate statistics applications in phase analysis of STEM-EDS spectrum images," *Ultramicroscopy*, vol. 110, no. 2, pp. 134–143, 2010.

[**234**] P. G. Kotula and M. R. Keenan, "Application of multivariate statistical analysis to STEM X-ray spectral images: Interfacial analysis in microelectronics," *Microsc. Microanal.*, vol. 12, no. 6, pp. 538–544, 2006.

[**235**] J. Li, J. M. Bioucas-DIas, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectra unmixing (R-CONMF)," *Work. Hyperspectral Image Signal Process. Evol. Remote Sens.*, vol. 2015-June, no. 10, pp. 6076–6090, 2015.

[**236**] J. Spiegelberg, S. Muto, M. Ohtsuka, K. Pelckmans, and J. Rusz, "Unmixing hyperspectral data by using signal subspace sampling," *Ultramicroscopy*, vol. 182, pp. 205–211, 2017.

[**237**] J. F. Einsle *et al.*, "Nanomagnetic properties of the meteorite cloudy zone," *Proc. Natl. Acad. Sci.*, vol. 115, no. 49, pp. E11436–E11445, 2018.

[**238**] R. Liu, A. Agrawal, W. K. Liao, A. Choudhary, and M. De Graef, "Materials discovery: Understanding polycrystals from large-scale electron patterns," *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016*, pp. 2261–2269, 2016.

[**239**] Y. F. Shen, R. Pokharel, T. J. Nizolek, A. Kumar, and T. Lookman, "Convolutional neural network-based method for real-time orientation indexing of measured electron backscatter diffraction patterns," *Acta Mater.*, vol. 170, pp. 118–131, 2019.

[**240**] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[**241**] A. Foden, A. Previero, and T. B. Britton, "Advances in electron backscatter diffraction," in *40th Risoe International Symposium: Metal Microstructures in 2D, 3D and 4D*, Aug. 2019, vol. 22, no. 11, p. 1261.

[**242**] K. Kaufmann *et al.*, "Crystal symmetry determination in electron diffraction using machine learning," *Science (80-.).*, vol. 568, no. January, pp. 564–568, 2020.

[**243**] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.

[**244**] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *SAE Int. J. Mater. Manuf.*, vol. 7, no. 3, pp. 1251–1258, Oct. 2016.

[**245**] K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, H. Wang, and K. S. Vecchio, "Phase Mapping in EBSD Using Convolutional Neural Networks," *Microsc. Microanal.*, pp. 1–11, 2020.

[**246**] G. B. Viswanathan, P. M. Sarosi, M. F. Henry, D. D. Whitis, W. W. Milligan, and M. J. Mills, "Investigation of creep deformation mechanisms at intermediate temperatures in René 88DT," *Acta Mater.*, vol. 53, no. 10, pp. 3041–3057, 2005.

[**247**] A. Winkelmann, G. Nolze, M. Vos, F. Salvat-Pujol, and W. S. M. Werner, "Physics-based simulation models for EBSD: Advances and challenges," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 109, no. 1, 2016.

[**248**] P. G. Callahan *et al.*, "Transmission scanning electron microscopy: Defect observations and image simulations," *Ultramicroscopy*, vol. 186, pp. 49–61, 2018.

[**249**] J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. M. Ritchie, J. H. J. Scott, and D. C. Joy, *Scanning electron microscopy and X-ray microanalysis*, Fourth. New York, NY: Springer, 2018.

[**250**] R. Jiang *et al.*, "Strain accumulation and fatigue crack initiation at pores and carbides in a SX superalloy at room temperature," *Int. J. Fatigue*, 2018.

[**251**] K. A. Christofidou *et al.*, "On the Effect of Nb on the Microstructure and Properties of Next Generation Polycrystalline Powder Metallurgy Ni-Based Superalloys," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 49, no. 9, pp. 3896–3907, 2018.

[**252**] A. Ilin, "Practical approaches to Principal Component Analysis in the presence of missing values," *J. Mach. Learn. Res.*, vol. 11, pp. 1957–2000, 2010.

[**253**] V. S. Tong, A. J. Knowles, D. Dye, and T. Ben Britton, "Rapid electron backscatter diffraction mapping: Painting by numbers," *Mater. Charact.*, vol. 147, no. September 2018, pp. 271–279, 2019.

[**254**] A. Campbell, P. Murray, E. Yakushina, S. Marshall, and W. Ion, "New methods for automatic quantification of microstructural features using digital image processing," *Mater. Des.*, vol. 141, pp. 395–406, 2018.

[**255**] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, 1994.

[**256**] S. Beucher and C. Lantejoul, "Use of watersheds in contour detection," in *International workshop on image processing*, 1979.

[**257**] H. F. Kaiser, "The Application of Electronic Computers to Factor Analysis," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 141–151, Apr. 1960.

[**258**] A. Winkelmann, T. Ben Britton, and G. Nolze, "Constraints on the Effective Electron Energy Spectrum in Backscatter Kikuchi Diffraction," vol. 064115, pp. 1–13, 2018.

[**259**] M. R. Keenan and P. G. Kotula, "Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images," *Surf. Interface Anal.*, vol. 36, no. 3, pp. 203–212, 2004.

[**260**] M. R. Keenan, "Exploiting spatial-domain simplicity in spectral image analysis," *Surf. Interface Anal.*, vol. 41, no. 2, pp. 79–87, 2009.

[**261**] D. Dye, M. Hardy, H. Yan, M. Knop, and H. Stone, "EP2821519B1," 2017.

[**262**] V. A. Vorontsov, L. Kovarik, M. J. Mills, and C. M. F. Rae, "High-resolution electron microscopy of dislocation ribbons in a CMSX-4 superalloy single crystal," *Acta Mater.*, vol. 60, pp. 4866–4878, 2012.

[**263**] R. C. Reed, N. Matan, D. C. Cox, M. A. Rist, and C. M. F. Rae, "Creep of CMSX-4 superalloy single crystals: Effects of rafting at high temperature," *Acta Metall.*, vol. 47, no. 12, pp. 3367–3381, 1999.

[**264**] T. McAuliffe, L. Reynolds, I. Bantounas, T. Britton, and D. Dye, "The Use of Scanning Electron Beam-based Phase Classification as a Crucial Tool in Alloy Development for Gas Turbine Engine Applications," *Microsc. Microanal.*, vol. 25, no. S2, pp. 2402–2403, 2019.

[**265**] M. Shiga *et al.*, "Sparse modeling of EELS and EDX spectral imaging data by nonnegative matrix factorization," *Ultramicroscopy*, vol. 170, pp. 43–59, 2016.

[**266**] W. Wang, Y. Huang, Y. Wang, and L. Wang, "Generalized Autoencoder: A Neural Network Framework for Dimensionality Reduction," *CVPR Work.*, pp. 490–497, 2014.

[**267**] R. Guo, M. Ahn, and H. Z. Hongtu Zhu, "Spatially Weighted Principal Component Analysis for Imaging Classification," *J. Comput. Graph. Stat.*, vol. 24, no. 1, pp. 274–296, Jan. 2015.

[**268**] M. Inokuti and S. T. Manson, "Cross sections for inelastic scattering of electrons by atoms: selected topics related to electron microscopy," 1982.

[**269**] P. A. Midgley and A. S. Eggeman, "Precession electron diffraction - A topical review," *IUCrJ*, vol. 2, pp. 126–136, 2015.

[**270**] T. Ben Britton, D. Goran, and V. S. Tong, "Space rocks and optimising scanning electron channelling contrast," *Mater. Charact.*, vol. 142, no. May, pp. 422–431, 2018.

[**271**] S. Zhao, R. Zhang, T. Pekin, and A. M. Minor, "Probing Crystalline Defects Using an EBSD-Based Virtual Dark-Field Method," *Microsc. Microanal.*, vol. 25, no. S2, pp. 1992–1993, 2019.

[**272**] L. R. Reynolds, "Development of Cobalt-based Superalloys," Imperial College London, 2019.

[**273**] F. B. Ismail, "Investigation Of The Oxidation Behaviour Of The Polycrystalline Co-Ni Base Superalloys," Imperial College London, 2017.

[**274**] C. H. Zenk, S. Neumeier, H. J. Stone, and M. Göken, "Mechanical properties and lattice misfit of γ/γ' strengthened Co-base superalloys in the Co-W-Al-Ti quaternary system," *Intermetallics*, vol. 55, pp. 28–39, 2014.

[**295**] S. Kobayashi, Y. Tsukamoto, and T. Takasugi, "Phase equilibria in the Co-rich Co-Al-W-Ti quaternary system," *Intermetallics*, vol. 19, no. 12, pp. 1908–1912, 2011.

[**276**] Y. Gu, H. Harada, C. Cui, D. Ping, A. Sato, and J. Fujioka, "New Ni-Co-base disk superalloys with higher strength and creep resistance," *Scr. Mater.*, vol. 55, no. 9, pp. 815–818, 2006.

[**277**] T. P. Gabb, J. Gayda, J. Telesman, and P. T. Kantzos, "Thermal and Mechanical Property Characterization of the Advanced Disk Alloy LSHR - Technical Memorandum," no. June, p. 82, 2005.

[**278**] S. Cruchley, H. E. Evans, M. P. Taylor, M. C. Hardy, and S. Stekovic, "Chromia layer growth on a Ni-based superalloy: Sub-parabolic kinetics and the role of titanium," *Corros. Sci.*, vol. 75, pp. 58–66, 2013.

[**279**] S. L. Shang, C. L. Zacherl, H. Z. Fang, Y. Wang, Y. Du, and Z. K. Liu, "Effects of alloying element and temperature on the stacking fault energies of dilute Ni-base superalloys," *J. Phys. Condens. Matter*, vol. 24, no. 50, 2012.

[**280**] D. M. Collins and H. J. Stone, "A modelling approach to yield strength optimisation in a nickel-base superalloy," *Int. J. Plast.*, vol. 54, pp. 96–112, 2014.

[**281**] M. Chandran and S. K. Sondhi, "First-principle calculation of APB energy in Ni-based binary and ternary alloys," *Model. Simul. Mater. Sci. Eng.*, vol. 19, no. 2, 2011.

[**282**] T. Dessolier, T. McAuliffe, W. J. Hamer, C. G. M. Hermse, and T. Ben Britton, "Effect of high temperature service on the complex through-wall microstructure of centrifugally cast HP40 reformer tube," pp. 1–21, Aug. 2020.

[**283**] K. M. Rahman, N. G. Jones, and D. Dye, "Micromechanics of twinning in a TWIP steel," *Mater. Sci. Eng. A*, vol. 635, pp. 133–142, 2015.

[**284**] M. A. Meyers, A. Mishra, and D. J. Benson, "Mechanical properties of nanocrystalline materials," *Prog. Mater. Sci.*, vol. 51, no. 4, pp. 427–556, 2006.

[**285**] J. L. W. Warwick, N. G. Jones, K. M. Rahman, and D. Dye, "Lattice strain evolution during tensile and compressive loading of CP Ti," *Acta Mater.*, vol. 60, no. 19, pp. 6720–6731, 2012.

[**286**] A. K. Ackerman *et al.*, "Interface characteristics in an <math> <mrow> <mi>α</mi> <mo>+</mo> <mi>β</mi> </mrow> </math> titanium alloy," *Phys. Rev. Mater.*, vol. 4, no. 1, p. 013602, Jan. 2020.

[**287**] T. W. J. Kwok *et al.*, "Design of a High Strength, High Ductility 12 wt% Mn Medium Manganese Steel With Hierarchical Deformation Behaviour," *arXiv.org*, Aug. 2019.

[**288**] K. M. Rahman, "Mechanical Behaviour of Twinning Induced Plasticity (TWIP) Steels," Imperial College London, 2013.

[**289**] O. Grässel, L. Krüger, G. Frommeyer, and L. W. Meyer, "High strength Fe-Mn-(Al, Si) TRIP/TWIP steels development - properties - application," *Int. J. Plast.*, vol. 16, no. 10, pp. 1391–1409, 2000.

[**290**] K. M. Rahman, V. A. Vorontsov, and D. Dye, "The dynamic behaviour of a twinning induced plasticity steel,"

Mater. Sci. Eng. A, vol. 589, pp. 252–261, 2014.

[**291**] M. Koyama, E. Akiyama, K. Tsuzaki, and D. Raabe, "Hydrogen-assisted failure in a twinning-induced plasticity steel studied under in situ hydrogen charging by electron channeling contrast imaging," *Acta Mater.*, vol. 61, no. 12, pp. 4607–4618, 2013.

[**292**] P. Müllner, "On the ductile to brittle transition in austenitic steel," *Mater. Sci. Eng. A*, vol. 234–236, pp. 94–97, 1997.

[**293**] P. Müllner, C. Solenthaler, and M. O. Speidel, "Second order twinning in austenitic steel," *Acta Metall. Mater.*, vol. 42, no. 5, pp. 1727–1732, 1994.

[**294**] P. Müllner, C. Solenthaler, P. J. Uggowitzer, and M. O. Speidel, "Brittle fracture in austenitic steel," *Acta Metall. Mater.*, vol. 42, no. 7, pp. 2211–2217, 1994.

[**295**] J. Johansson, M. Odén, and X. H. Zeng, "Evolution of the residual stress state in a duplex stainless steel during loading," *Acta Mater.*, vol. 47, no. 9, pp. 2669–2684, 1999.

[**296**] P. Hedström, T. S. Han, U. Lienert, J. Almer, and M. Odén, "Load partitioning between single bulk grains in a two-phase duplex stainless steel during tensile loading," *Acta Mater.*, vol. 58, no. 2, pp. 734–744, 2010.

[**297**] J. P. Chateau, A. Dumay, S. Allain, and A. Jacques, "Precipitation hardening of a FeMnC TWIP steel by vanadium carbides," *J. Phys. Conf. Ser.*, vol. 240, 2010.

[**298**] H. W. Yen, M. Huang, C. P. Scott, and J. R. Yang, "Interactions between deformation-induced defects and carbides in a vanadium-containing TWIP steel," *Scr. Mater.*, vol. 66, no. 12, pp. 1018–1023, 2012.

[**299**] Y. Li *et al.*, "Segregation stabilizes nanocrystalline bulk steel with near theoretical strength," *Phys. Rev. Lett.*, vol. 113, no. 10, pp. 1–5, 2014.

[**300**] J. Pokluda, M. Černý, P. Šandera, and M. Šob, "Calculations of theoretical strength: State of the art and history," *J. Comput. Mater. Des.*, vol. 11, no. 1, pp. 1–28, 2004.

[**301**] K. H. So, J. S. Kim, Y. S. Chun, K. T. Park, Y. K. Lee, and C. S. Lee, "Hydrogen delayed fracture properties and internal hydrogen behavior of a Fe-18Mn-1.5AI-0.6C TWIP steel," *ISIJ Int.*, vol. 49, no. 12, pp. 1952–1959, 2009.

[**302**] B. Clausen, C. N. Tomé, D. W. Brown, and S. R. Agnew, "Reorientation and stress relaxation due to twinning: Modeling and experimental characterization for Mg," *Acta Mater.*, vol. 56, no. 11, pp. 2456–2468, 2008.

Appendix 1 - Pattern Centre & Orientation Refinement

1. Chapter Summary

Very accurate knowledge of an electron backscatter pattern (EBSP) pattern centre, (PC), the centroid of the inelastically scattered electron intensity, is required in order for robust template matching and inverse gnomonic projection for 'on-the-sphere' analysis.

2. Methods

Given an *a priori* known crystallographic phase, a coupled orientation / PC solution can be iteratively updated *via* the RTM procedure. The peak height of the cross-correlation function (XCF) of the refined solution and the experimental pattern can be used as an objective function to be maximised. In this paradigm, well established gradient ascent algorithms, such as those discussed in **Section LR.4.1** can be employed. The algorithm takes the following form:

Initialise parameters:

- Running orientation solution = $\mathbf{\Phi}_n$ (a vector of three Euler angles)
- Running PC solution = $\boldsymbol{p}_n$ (a vector - [PCX, PCY, DD])
- Running best PH = χ_n (a scalar)
- Initial orientation solution = $\mathbf{\Phi}_1$
- Initial PC solution = $\boldsymbol{p}_1$
- Experimental pattern = $\boldsymbol{\psi}$
- Initial XCF peak height = $\chi_1 = 0$
- Crystal phase = C
- Number of iterations to perform = *n_iterations*

---

for n = 1:*n_iterations*

1. Simulate EBSP, $\mathbf{\Psi}_\mathrm{n} = \mathrm{Generate}(\Phi_n, p_n\ ;\ \mathrm{C})$.
2. Calculate and refine XCF($\mathbf{\Psi}_\mathrm{n}$, $\boldsymbol{\psi}$) to generate new orientation, $\boldsymbol{\phi}$, with a corresponding peak height, *x*.
3. **if** previous iteration resulted in an update, estimate ∇_{p_n}XCF ($\boldsymbol{\psi}$, $\mathbf{\Psi}_\mathrm{n}$) with a finite difference method.
4. **if** x > χ_{n-1}:
   - $\boldsymbol{p}_n$ = **p**
   - $\mathbf{\Phi}_n = \boldsymbol{\phi}$
   - χ_n = x
   - $\boldsymbol{p} = \boldsymbol{p}_n + \alpha\ \nabla_{p_n}\mathrm{XCF}\ (\boldsymbol{\psi}, \mathbf{\Psi}_\mathrm{n})$
5. **else**:
   - $\alpha = 0.9\ \alpha$
   - $\boldsymbol{p}_n = \boldsymbol{p}_{n-1}$
   - $\mathbf{\Phi}_n = \mathbf{\Phi}_{n-1}$
   - $\chi_n = \chi_{n-1}$
   - $\boldsymbol{p} = \boldsymbol{p}_{n-1} + 0.1\ \alpha\ rand$

Algorithm Ap1-1: Gradient descent for orientation and PC refinement

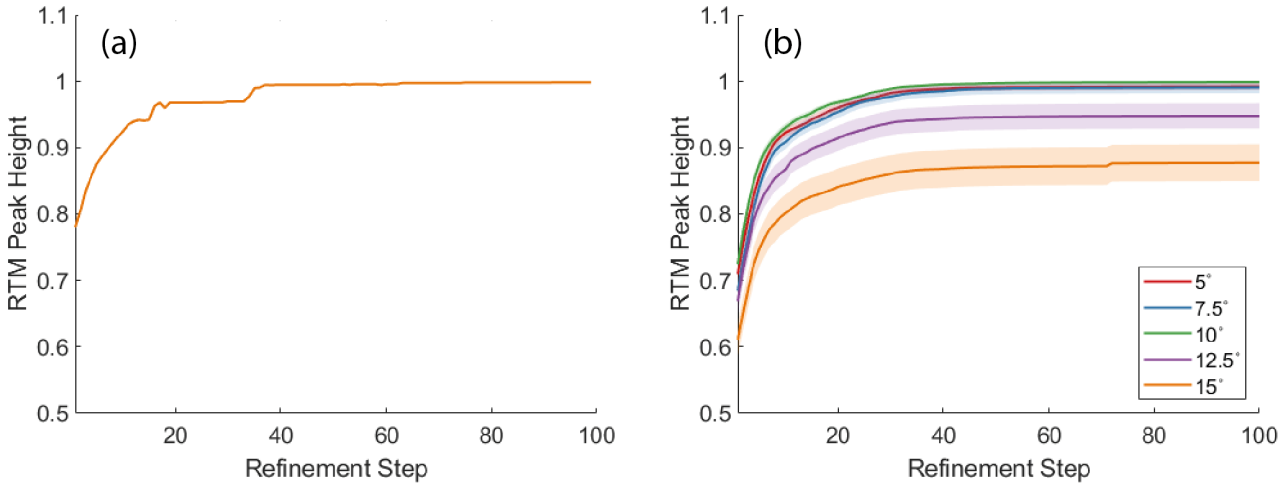

Figure Ap1-1: (a) Increase of the RTM peak height, χ_n, as refinement proceeds, (b) average across 100 tests, for each of five initial misorientations, with standard error in the mean shaded.

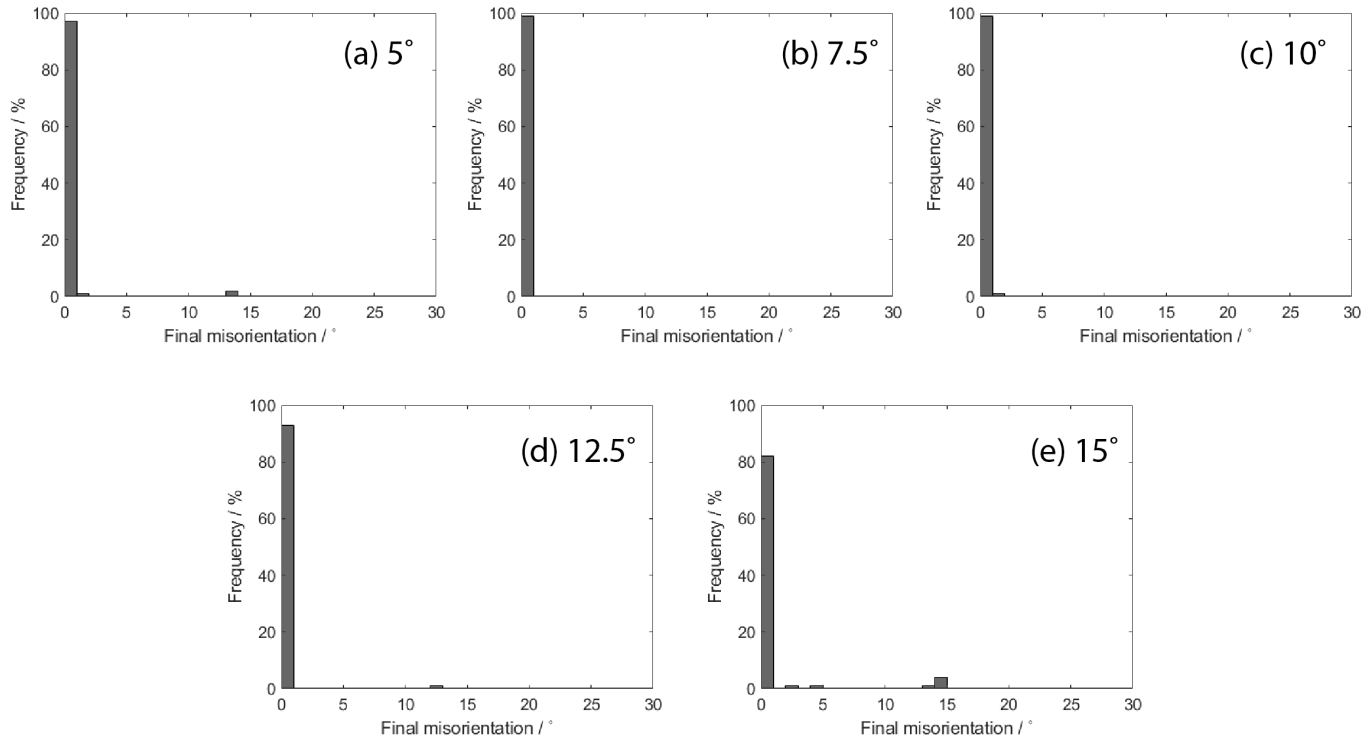

Figure Ap1-2: Histograms of final misorientation after refinement, with initial misorientations given (a)-(e).

Initial misori.	Proportion of tests with final misorientation less than: **0.25°**	**0.5°**	**1°**	**2°**
5°	0.33	0.82	0.97	0.98
7.5°	0.32	0.81	0.99	0.99
10°	0.21	0.82	0.99	1.00
12.5°	0.27	0.81	0.93	0.93
15°	0.25	0.57	0.82	0.82

Table Ap1-1: Proportion of tests that resulted in misorientations below four specified levels, for each initial misorientation.

To evaluate the procedure, simulated patterns of known, random orientation and PC were used as experimental patterns. Specifically, the true PC, $q = [0.45, 0.4, 0.65] + [0.1\ rand, 0.1\ rand, 0]$. This is perturbed by [0.15 *rand*, 0.15 *rand*, 0] to provide the initial solution p_1. A true orientation is randomly sampled using the *mtex* crystallographic analysis package. Random candidates for the initial orientation solution are generated until one is sampled with misorientation to the true orientation of less than 1°. This orientation is then set as Φ_1. These p_1, Φ_1 were then refined using the algorithm presented. A PC deviation parameter is defined:

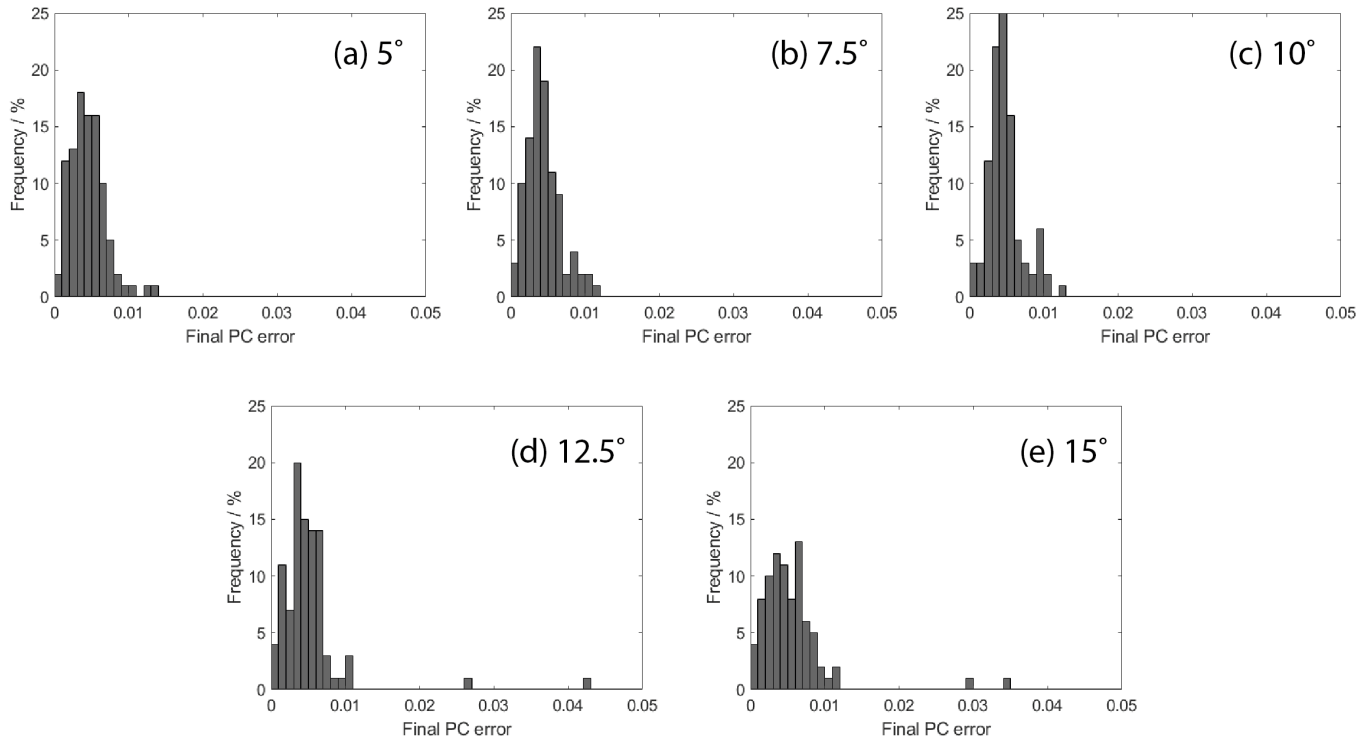

Figure Ap1-3: Histograms of final PC error, Δ, after refinement, with initial misorientations given (a)-(e).

	Proportion of tests with final PC error less than:			
Initial misori.	**0.0025**	**0.005**	**0.010**	**0.050**
5°	0.23	0.61	0.95	0.98
7.5°	0.25	0.68	0.96	0.99
10°	0.10	0.65	0.97	1.00
12.5°	0.18	0.57	0.90	0.95
15°	0.19	0.45	0.79	0.84

Table Ap1-2: Proportion of tests that resulted in PC errors below four levels, for each initial misorientation.

PC designation	[PCX, PCY, DD]
True	[0.4694, 0.4251, 0.6500]
Start	[0.5723, 0.4697, 0.6500]
Final	[0.4715, 0.4307, 0.6507]
Difference (True – Final)	[-0.0021, -0.0055, -0.0007], or Δ = 0.0060

Table Ap1-3: Breakdown of PC errors for an example 15° misorientation test.

$$\Delta = \sqrt{\sum_{i=1}^{3} (q_i - p_i)^2}$$

Which is the Euclidean distance between q and p. Final misorientation and Δ were measured upon completion. This evaluation procedure was repeated for 100 randomised start conditions, and 5 levels of misorientation to the true solution: 5°, 7.5°, 10°, 12.5°, 15°. The parameter *n_iterations* was set to 100.

3. Results & Discussion

A typical plot of the objective function as refinement proceeds is presented in **Figure Ap1-1a**. As is typical in gradient based methods, an initial rapid improvement is followed by a slow ascent towards a local maximum. Evaluating the arithmetic mean average of this objective profile across 100 tests for each of the given initial misorientations yields the profiles presented in **Figure Ap1-1b**. The presented approach performs better when the initial conditions are closer to the true values.

In the 15° misorientation case, many tests result in an accurate solution, but the mean and standard deviation are magnified by a larger number of tests failing to converge. Histograms of the final misorientations after refinement are presented in **Figure Ap1-2**. These can be quantitatively summarised as in **Table Ap1-1**, by considering the proportion

of tests that result in final misorientations less than specified levels. It is evident that the highest levels of initial misorientation lead to a lower likelihood of successful refinement. However, many tests at the highest degree of misorientation were successful to within a small margin of error. The principal difference across test regimes seems to be the higher initial misorientations are slightly less likely to successfully converge, within the experiment paradigm of 100 iterations. Surprisingly, it seems that the best performance is achieved when the initial misorientation is 10˚. This is possibly due to the choice of step size α, and insufficiently fast learning rate annealing.

The same error analysis was conducted to consider the final PC error, with results summarised in **Figure Ap1-3***:* Histograms of final PC error, Δ, after refinement, with initial misorientations given (a)-(e). and **Table Ap1-2**. The same trend as for final misorientation is observed; a higher degree of initial misorientation leads to a slightly reduced likelihood of successful refinement, with the optimal performance seen for initial misorientation of 10˚.

Finally, for illustration, the final PC errors are expanded upon for an example refinement of 14.89˚ initial misorientation, and 0.49˚ final misorientation after refinement. This is presented in **Table Ap1-3**. The algorithm results in a difference in PCX of 0.0021, for a 300 px square image corresponding to 0.63 px error, 1.65 px error in PCY, and 0.21 px error in DD.

4. Conclusions

This mini-chapter has demonstrated the coupled orientation / PC refinement possible using a finite difference gradient based method. The algorithm was tested on dynamical simulations of known orientation, with randomised initial orientation and PC (to within a specified tolerance of misorientation), and the final misorientation and PC error measured.

Appendix 2 - Measuring the quality of cross-correlation

1. Chapter Summary

The distribution of XCF peak heights for refined template matching are evaluated, and a new metric for determining quality of matching presented. This standardised score, Z, corresponds to confidence that the solution was not effectively sampled randomly from the SO(3) search template library. This score is suitable for determining whether a solution should be considered or regarded as noise. The unnormalised cross-correlation peak height remains a better metric of absolute similarity, and should be used to determine the structure and orientation solution from acceptable (non-random) candidates.

2. Method and Results

The cross-correlation function (XCF) peak height allows inference of the displacements (lag) in multidimensional signals that maximise similarity [**137**]. In the refined template matching (RTM) procedure, the measured horizontal and vertical displacements between a measured electron backscatter pattern (EBSP) and the most similar template (with a known orientation) allows rotations about two axes to be corrected, improving the determined orientation. This procedure always gives a result, which is problematic in cases where the identified crystal (with the largest XCF peak height) either has no solution (the EBSP is noise), or the correct crystal structure was not simulated. For this reason we seek a metric of absolute match quality, to quantify how reasonable the RTM solution is.

It is sensible to normalise the XCF peak heights with respect to the distribution from which they are generated. The normalised XCF peak heights, Z are then given:

$$Z = \frac{\log\left(\chi_{max}\right) - \mu}{\sigma} \qquad \textit{Ap2-1}$$

For SO(3) search maximum XCF peak height χ_{max} (for a given phase). Further, μ and σ are the arithmetic mean and standard deviation of the XCF peak heights for the SO(3) library search:

$$\mu = \frac{1}{n}\sum_{i}^{n} \log\left(\chi_i\right) \qquad \textit{Ap2-2}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i}^{n}\left(\log\left(\chi_i\right) - \mu\right)^2} \qquad \textit{Ap2-3}$$

Where χ_i are the XCF peak heights in the library search of n templates. The normalisation routine assumes a lognormal distribution of peak heights within an SO(3) library search for a test EBSP. An example of such distributions are presented in **Figure Ap2-1**, with test pattern known *a priori* to be an FCC Ni EBSP [**1**].

Ni	M6C	ZrC	Ferrite
1.1516×10^{-34}	2.548×10^{-12}	2.123×10^{-24}	8.825×10^{-31}

Table Ap2-1: p-values for Chi-square hypothesis tests comparing log(XCF peak height) to a normal distribution.

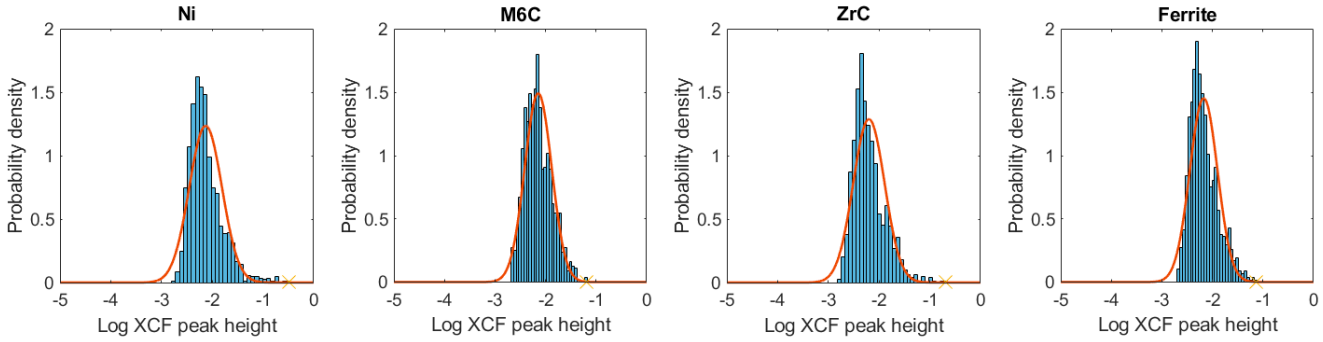

Figure Ap2-1: Histograms and fitted probability distributions for SO(3) searches against Ni, M6C, ZrC, and Ferrite dynamical simulations. The 'x' correspond to the measured maximum peak height for the corresponding library search.

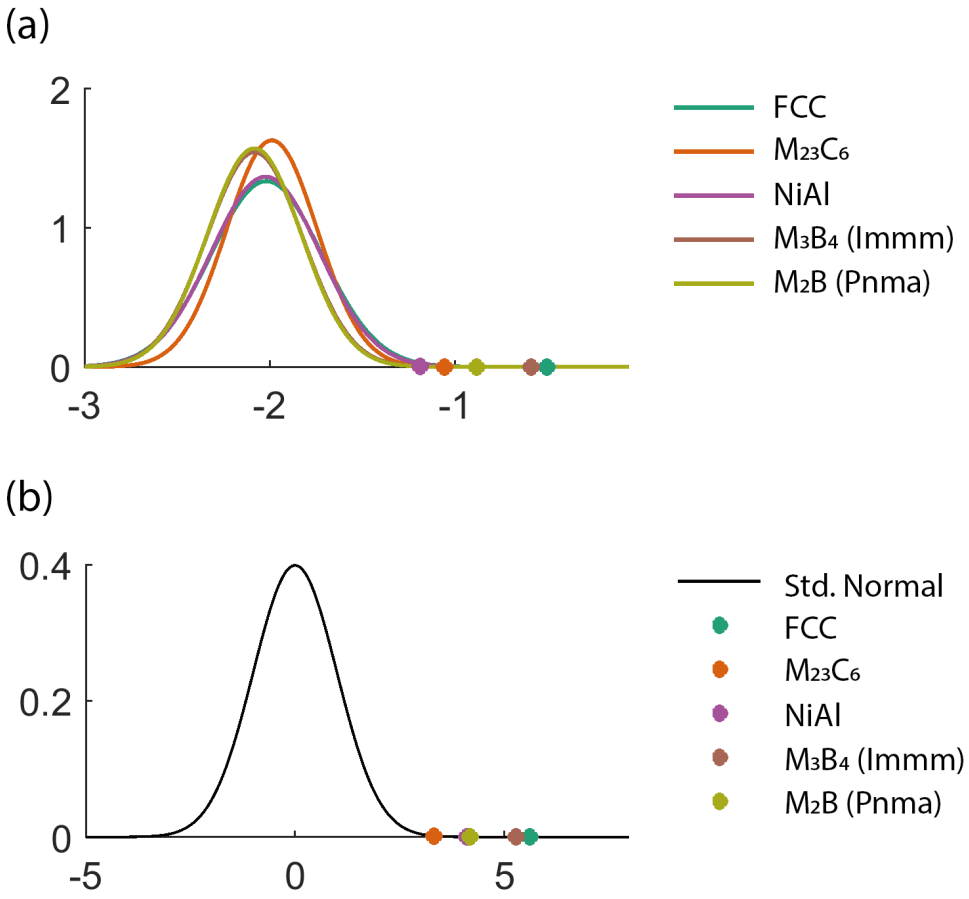

Figure Ap2-2: (a) Comparison of fitted probability distributions for each phase, for a Ni EBSP, and (b) XCF peak heights normalised with respect to a standard data-generating distribution.

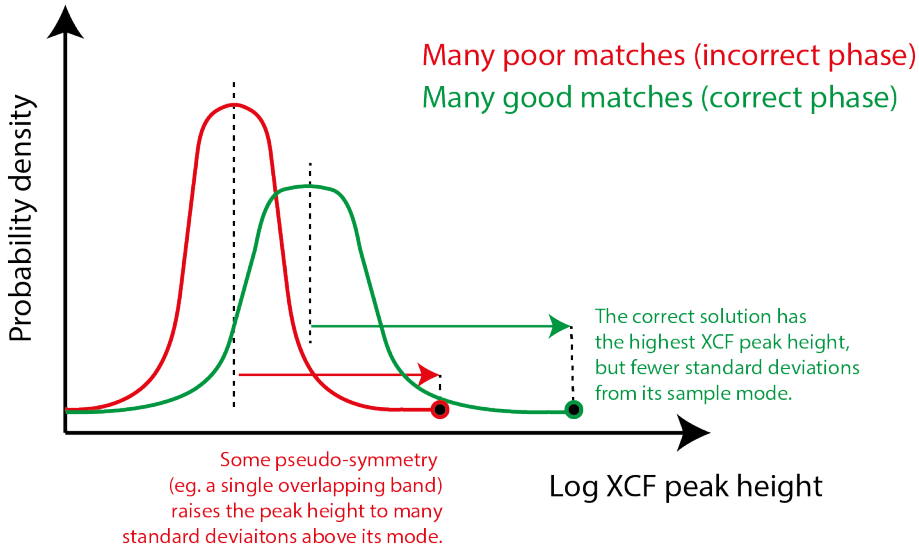

Figure Ap2-4: Demonstration of how a poor solution can present a large *Z*.

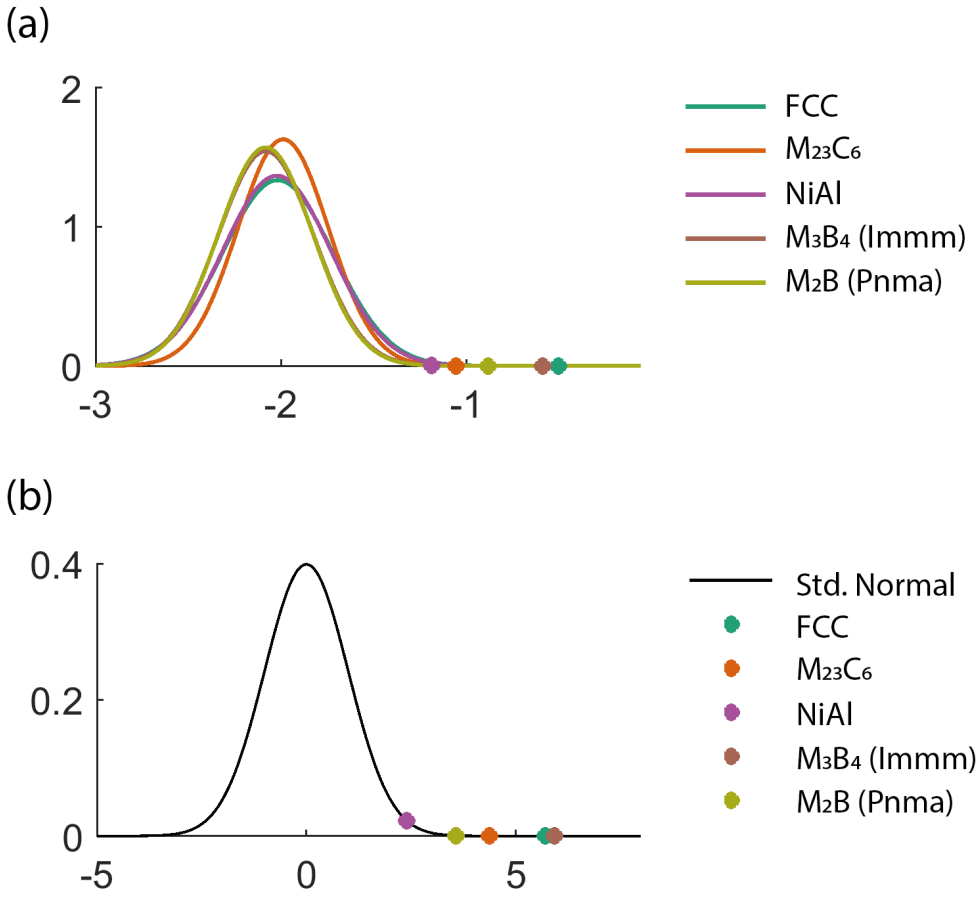

Figure Ap2-3: The correct solution (Ni) does not exhibit the maximal *Z*, due to the relatively large standard deviation of its Gaussian parameterisation. It does exhibit the maximal XCF peak height.

In order to validate the veracity of this prior, a Chi-square goodness of fit hypothesis test is performed to determine whether the null hypothesis (that the measured XCF peak heights are not a sample from a lognormal distribution) can be rejected. Rejection of the null hypothesis corresponds to validating that the log of the peak height is lognormally distributed. The p-values for the Chi-square tests performed are presented in **Table Ap2-1**.

The p-values correspond to the probability that the log of the peak heights fitting a normal distribution just as well as

has been observed given that the null hypothesis is true. Their small value means that we reject the null hypothesis, and validate the prior that that XCF peak heights are lognormally distributed.

Figure Ap2-2a presents a comparison of the log of the XCF peak heights for a Ni EBSP to the corresponding fitted lognormal distributions. Clearly the maximum peak height from the Ni SO(3) search is greater than the other maxima, and the FCC ZrC phase is a runner up.

Now applying the standardisation routine, **Figure Ap2-2b** shows the normalised XCF peak heights in **Figure Ap2-2a**, interpreted in terms of a standard normal data-generating distribution. Importantly, in this case the ordering is preserved.

The principal advantage of this metric is that Z corresponds to a number of standard deviations away from the expected value of the data-generating distribution. This allows direct interpretation of Z in terms of a probability that the measurement was not randomly sampled from the distribution. For example, $Z > 2$ would correspond to approximately a > 97.8% probability this came from a purely random sample from the SO(3) library. Therefore, it is suitable as a metric for determining whether a solution is acceptable (a pattern of pure noise is expected to have ≤ the modal XCF peak height). Then, for instance, one can only accept solutions with Z > 4, which implies an extreme degree of confidence that the solution is better than any random sample from the SO(3) group.

In **Figure Ap2-2b**, the ordering between candidate phases was preserved with respect to their unnormalised XCF peak heights. Unfortunately, due to strong variation in the standard deviation of $\log(\chi_i)$ between candidate phases, the largest value of Z does not always correspond to the best solution. This is illustrated in **Figure Ap2-3**.

This effect can result in the incorrect phase being assigned to a solution, if the maximal Z is taken. An experimental example is presented in **Figure Ap2-4**. Here, all the candidate phases have a *better-than-random* template with at least moderate similarity to the measured pattern. Thus, it can be concluded that none should be discarded and all are valid solutions. The optimal solution should of course be that with the absolute best similarity (most shared pixels) to the experimental pattern, independent of its class's sampled distribution, which corresponds to the maximal value of the unnormalised XCF peak height.

3. Conclusions

From the analysis presented in this chapter the following straightforward conclusions can be drawn:

- A lognormal distribution is suitable for approximating the population probability density function for XCF peak heights.
- Standardisation of the inferred parameters permits quantifying confidence that a peak height value was not randomly selected from its population. This provides a metric for determining whether to retain a structure/orientation solution.
- The standardised peak heights are not suitable for selection of absolute best phase from a group of acceptable (non-random) candidates. This is best achieved with the absolute (unnormalised) XCF peak height.

Appendix 3 - 4D-STEM elastic stress state characterisation of a TWIP steel nanotwin

1. Chapter Summary

The stress state in and around a deformation nanotwin in a twinning-induced plasticity (TWIP) steel is measured. Using four dimensional scanning transmission electron microscopy (4D-STEM), the elastic strain field in a 68.2-by-83.1 nm area of interest is measured with a scan step of 0.36 nm and a diffraction limit resolution of 0.73 nm. The stress field in and surrounding the twin matches the form expected from analytical theory and is on the order of 15 GPa, close to the theoretical strength of the material. It is inferred that the measured back-stress limits twin thickening, providing a rationale for why TWIP steel twins remain thin, continuously dividing grains to give substantial work hardening. These results support modern mechanistic understanding of the influence of twinning on crack propagation and embrittlement in TWIP steels.

2. Introduction

Understanding the mechanical behaviour of engineering alloys at the nanoscale is critical to improving alloy design and processing, and hence performance [**283**]–[**286**]. Steel is the most common man-made material after concrete, at around 1.6 billion tonnes per year production, seeing critical application in gas turbines, nuclear plant, construction, and automotive. Plastic deformation by twinning or shear-associated martensite transformation is commonly used in alloy design strategies, from Mg to Zr, steels and BCC Ti alloys, as well as functional intermetallics such as NiTi. Thus, better understanding of the stress state and back stress around such features will open up new avenues for improving performance, by manipulation of composition and processing to achieve desirable nanoscale behaviour and therefore bulk properties. In the case of face-centred cubic (FCC) crystal twinning, as presented in **Figure Ap3-1**, Shockley partial <112>{111} dislocations, formed from the dissociation of a full <110>{111} lattice dislocation, propagate on successive {111} planes. These impart a local strain field, and their propagation (and therefore accumulation of plastic strain) is limited by interaction with other lattice defects (interstitial or substitutional solutes, additional twins, dislocations, *etc*).

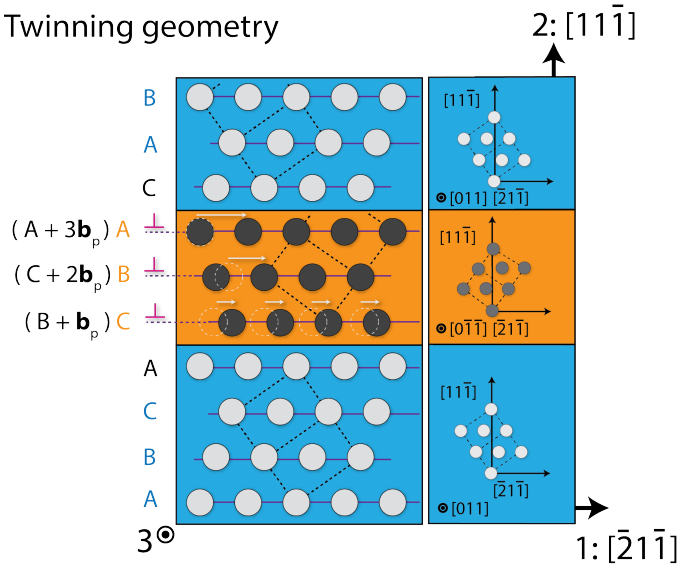

Figure Ap3-1: Twinning in FCC materials, with habit plane (11$\bar{1}$). A series of Shockley partial [$\bar{2}$1$\bar{1}$](11$\bar{1}$) dislocations transform the crystal plane by plane, building up the twin.

Four dimensional scanning transmission electron microscopy (4D-STEM) is a relatively new technique in which an electron diffraction spot pattern is acquired at every point in a scan grid. In this regard it is similar to electron backscatter diffraction (EBSD), now a routine method for microscale structural analysis [**109**], [**117**], but a much finer 'pencil-beam' probe permits sub-nm spatial resolution. The trade-off is that a zone-axis generally must be identified and aligned with the transmitted beam, inherently limiting knowledge of the

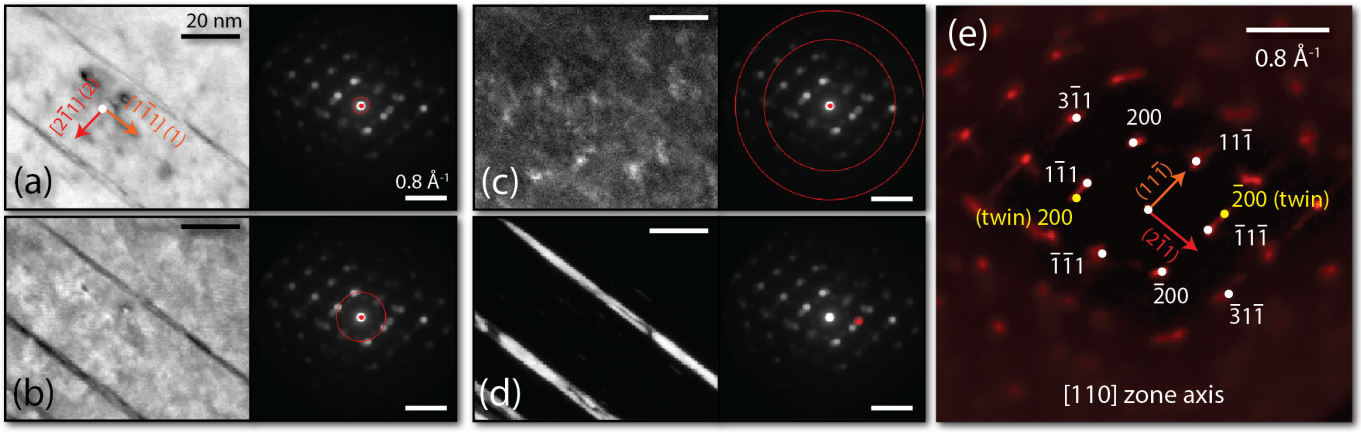

Figure Ap3-2: Virtual bright (a,b) and dark (c,d) field images of the steel nanotwin, as well as average Bragg vector maps (e) - sums over real space of the identified Bragg peak locations, weighted by intensity. The red circles in (a-d) diffraction patterns correspond to the limits of the virtual aperture.

reciprocal lattice to two coplanar vectors. A comprehensive review of 4D-STEM and its applications in strain mapping, imaging, and ptychography is available in ref [**147**]. In this work the *py4DSTEM* open source software package, developed by Savitzky *et al* [**148**], is employed.

Lattice strain measurement with this approach is becoming fairly routine. It has been used to investigate resistivity in semiconductors [**149**], [**150**], and more recently begun to be applied to polycrystalline materials [**151**]. Pekin *et al* [**152**], [**153**] have measured the strain field around austenitic (FCC) stainless steel features. They observed a ~4% variation in strain across their area of interest, which included dislocations and an annealing twin boundary. In this study we investigate the elastic strain fields in a similar FCC Fe material, and additionally calculate the stress fields directly from the elastic strain measurements. Here twinning-induced plasticity (TWIP) steel is examined, with a focus on thin deformation nanotwins. Twinning-induced plasticity in these systems can result in large ductility of up to 95% [**287**]–[**289**]. It is believed that continuous subdivision of grains by ongoing twin nucleation, without significant thickening, leads to a dynamic 'Hall-Petch' effect with sufficient twin back-stress to inhibit propagation of dislocations at these barriers [**290**]. This leads to pile-ups, hardening, and ductility. Deformation twins have also been explored in the context of crack initiation in TWIP steels by Koyama *et al* [**291**]. In this study we provide a direct measurement of the nanoscale stress state for comparison to the increasingly cited analytical model of Müllner *et al* [**292**]–[**294**]. In this chapter it is qualitatively observed that the analytical form of this model corresponds well to measurements of local stress.

3. Methods

An ingot of TWIP-steel (Fe – 16.4Mn – 0.9C – 0.5Si – 0.05Nb – 0.05 V wt%) was produced by vacuum arc melting in an Ar atmosphere. It was cast, homogenised at 1300˚C for 24 h, hot rolled with a 50% reduction, cold rolled with a 67% reduction, and annealed at 1000˚C for 5 min. This gave a fully austenitic microstructure. A 'dogbone' tensile specimen with 1 x 1.5 mm cross section and 19 mm gauge was deformed to 6% plastic strain at a nominal stroke rate of 10^{-3} s^{-1}.

A <110> zone was selected for this experiment, defined such that the [011] direction is out of the plane. EBSD was used to identify a grain with a <110> crystallographic direction in the vertical axis of the scanning electron microscope field of view. An electron transparent sample was then prepared using the focussed ion beam (FIB) lift-out technique, extracting a thin lamella with its plane aligned to the horizontal axis of the scan, thus yielded a thin foil of the {110} plane, with the <110> direction its normal.

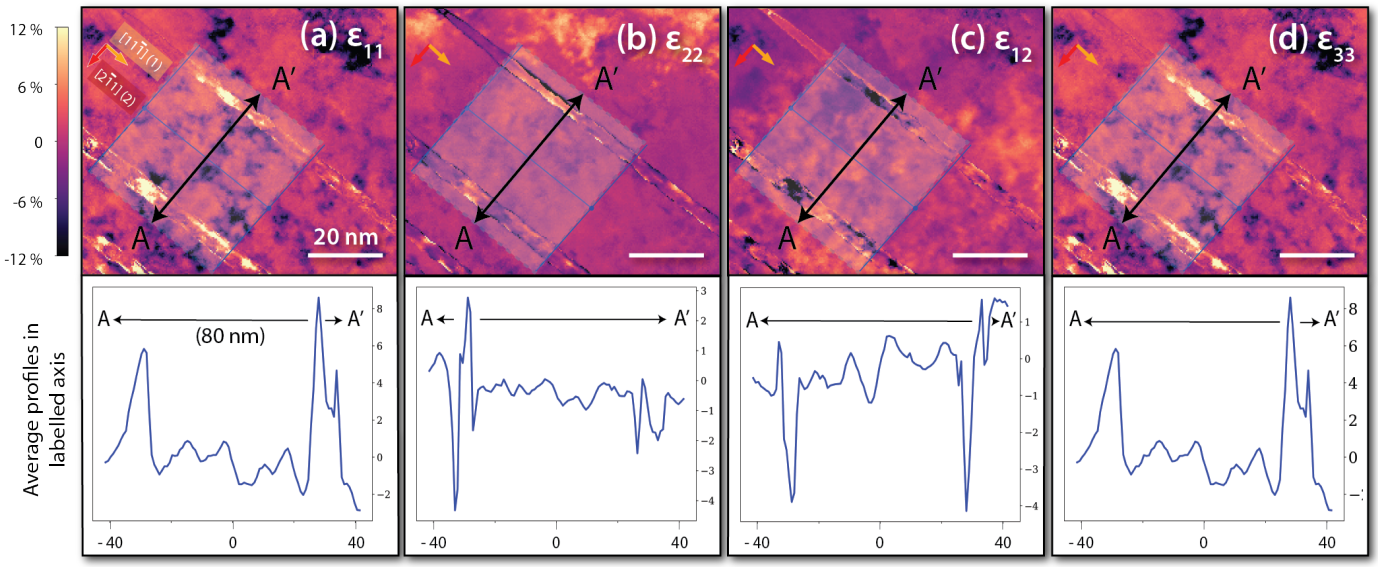

Figure Ap3-3: Elastic strain maps resolved in the 11 (a), 22 (b), 12 (c) and 33 (d) directions. An average 1-direction profile, perpendicular to the twin's length, was calculated by integrating all points in the 2-direction in the highlighted region.

4D-STEM was performed on the electron transparent foil using the probe and image spherical aberration-corrected *JEOL* ARM300CF TEM at *ePSIC*. The pencil beam was set-up by turning off the probe corrector hexapoles and using the condenser and transfer lens pairs to reach a small convergence semi-angle. Diffraction patterns were collected with a Merlin (MediPix) direct electron detector. An accelerating voltage of 200 kV and a camera length of 9 cm was used, with a 10 µm condenser lens aperture. Calibration diffraction data was gathered from evaporated gold on amorphous carbon sample and using the 10 µm aperture a 2.1 mrad convergence semi-angle and 0.0157 $Å^{-1}$ detector pixel size were measured. This gives a diffraction-limited spatial resolution of 0.73 nm. A 68.2-by-83.1 nm area of interest was scanned in 188-by-229 real space pixels, and with a 256-by-256 pixel diffraction pattern captured at each of these scan positions, with a 1 ms dwell time per pattern.

The in-plane elastic strain tensor was calculated from the electron diffraction pattern at each scan location. Bragg peak identification, dataset calibration including elliptical distortion correction and diffraction shift correction, and elastic strain calculation were performed with the open source py4DSTEM analysis package [**148**], [**154**]. For locating Bragg peaks a correlation power of 1 was used, corresponding to cross-correlation [**153**], and estimate the subpixel Bragg disc positions with local Fourier up-sampling by a factor of 16 [**142**]. The details of this method are considered further in **Section LR.3.3**.

The in-plane elastic stress was determined from the measured strains and Hooke's law. The axis system presented in **Figure Ap3-1** is used: *1*,*2*,*3* refer to the twinning system, with the *23* plane at the interface (origin at the twin centre). In this scheme the $[11\bar{1}]$ twin plane normal is aligned with the *2* direction, with *1* along $[\bar{2}1\bar{1}]$. To calculate the stresses, a reference stiffness tensor, $\mathbf{C}^{ref}$ is rotated to the *1*, *2*, *3* axis system. Reference stiffness tensor components C_{11}^{ref} = 197.5 GPa, C_{12}^{ref} = 124.5 GPa, C_{44}^{ref} = 122.0 GPa are used, with subscripts referring to tensor components rather than our axis systems, as measured for a similar austenite by Johansson *et al* [**295**], [**296**] with respect to a <100> system. Having previously calculated the in-plane elastic strains resolved in the $[1\bar{2}1]$ and $[10\bar{1}]$ directions, the complete elastic stress state can be inferred by assuming plane strain: the unknown strain components ε_{33}, ε_{13}, and ε_{23} are found by assuming

the stress components σ_{33}, σ_{13}, and σ_{23} are zero. The full strain state and compliance tensor are then used to calculate the full stress tensor, and the unkonwn, non-zero, σ_{11}, σ_{22}, and σ_{12}. This procedure is discussed in **Section LR.3.3**.

4. Results & Discussion

4.1 Virtual imaging and strain mapping

Presented in **Figure Ap3-2** are 'virtual' bright and dark field (VBF, VDF respectively) images reconstructed from intensity collected from the highlighted digital apertures. VBF images (a) and (b) clearly distinguish the twin from the matrix. Given the axis system adopted, [011] out of plane, and the diffraction vectors as indexed in **Figure Ap3-2**, it is inferred that these twins have habit plane $(11\bar{1})$.

A large amount of structure-dependent information is contained in the direct beam. Traditional bright field imaging (using an objective aperture to isolate the direct beam) uses electron wave phase information as well as intensity to re-interfere and reconstruct an image [**107**]. In VBF one only has access to electron intensity in the diffraction plane, so it is likely that the contrast observed between twin and matrix is derived from local strain, lattice rotation, or dynamical effects which will alter the ratio of diffracted to direct intensity. Two VBF images are presented in **Figure Ap3-2** (**a**,**b**) to show the presence of diffraction contrast in the direct beam as well as in the first order diffraction spots. The high angle VDF image in (**c**) is akin to conventional high-angle annular dark field (HAADF) imaging, which indicates propensity to scatter electrons to high angles, and suggests that there is no detectable variation in local chemistry between the twin and the matrix. The minimal contrast observed in (**c**) could be attributed to local ordering as a precursor to the formation of coherent V-rich carbides [**297**], [**298**]. The twin is explicitly highlighted in (**d**) by reconstructing the spatial image from the $\bar{2}00$ reciprocal lattice point for the twinned region only, analogous to a traditional TEM dark field image. In (**e**) we present a Bragg vector map (after Savitzky *et al* [**154**]), showing the distribution of all measured Bragg peak locations for the untwinned region. It is noted that a small amount of intensity is observed at the twin reciprocal lattice points even for the untwinned class. This is possibly due to the geometry of the specimen, dynamical diffraction, and through-thickness sampling of both untwinned and twinned material near the interface.

The strain components ε_{11}, ε_{22}, ε_{12} were calculated from the relative movements of the diffraction spots. This operation was performed independently for the twin and the matrix regions. A set of reference reciprocal lattice vectors were obtained by averaging the untwinned region's reciprocal lattice basis vectors. The twin basis vectors magnitudes were normalised to this unstrained length. As such, elastic strains are given in reference to this 'unstrained' state. The measurement could alternatively be considered as the elastic strain variation across the area of interest.

Maps of measured (11, 22, 12) and inferred (33) strain components across the area of interest are presented in **Figure Ap3-3**. Included is an integration along the 2-direction in a highlighted area, to obtain an average line profile in the 1 direction. It is observed that ε_{22} and ε_{12} remain fairly level in the matrix region between the twins, while ε_{11} exhibits more deviation. The twinned regions exhibit an increase in ε_{11} and ε_{22}, and a reduction in ε_{12} relative to the matrix.

Maps of the stress tensor are presented in **Figure Ap3-4**. The same line profile integration as for the strain maps is performed. The components σ_{11} and σ_{22} exhibit large positive stress rises across the interface. The component σ_{12} is much smaller in magnitude, but generally observes a negative 'sense' shear stress profile.

4.2 Comparison to micromechanical model

For comparison to an analytical micromechanical model, the formulation of Müllner *et al* [**292**]–[**294**] is followed. This considers the elastic stress field around the twin to follow that of a disclination dipole. Given a distance *x* along the 1-direction, and *y* along the 2-direction, in this scheme:

$$\sigma_{11} = D\,\omega \left\{ \frac{1}{2} \log \left[\frac{x^2 + (y-a)^2}{x^2 + (y+a)^2} \right] + \frac{x^2}{x^2 + (y-a)^2} - \frac{x^2}{x^2 + (y+a)^2} \right\} \quad \textit{Ap3-1}$$

$$\sigma_{22} = D\,\omega \left\{ \frac{1}{2} \log \left[\frac{x^2 + (y-a)^2}{x^2 + (y+a)^2} \right] + \frac{x^2}{x^2 + (a)^2} - \frac{x^2}{x^2 + (y+a)^2} \right\} \quad \textit{Ap3-2}$$

$$\sigma_{12} = D\,\omega\, x \left\{ \frac{y+a}{x^2 + (y+a)^2} - \frac{y-a}{x^2 + (y-a)^2} \right\} \quad \textit{Ap3-3}$$

Using a natural logarithm, where *a* is the twin half-thickness, with pre-factor *D* given:

$$D = \frac{C_{44}^{ref}}{2\pi\,(1-\nu)} \quad \textit{Ap3-4}$$

Using the shear modulus, *ν* Poisson's ratio, and parameter ω the 'power' of the disclination:

$$\omega = 2\,\tan^{-1}\frac{b}{2h} = 38.94^{\circ} \text{ for FCC materials [293]}$$

With *b* as the magnitude of the Shockley partial Burgers vector and *h* the separation between twinning planes. The co-ordinates *x*, *y*, *z* are taken in reference to the centre of the disclination dipole from which the stress field is derived [**293**]. Models of σ_{11}, σ_{22}, and σ_{12} along the **A-A'** profile are included in **Figure Ap3-4**. Here *y*, *z* = 0 at the centre of the twins, and *y* is varied along the profile **A-A'**. It is found that an initial *x* value of ~ -5 nm leads to stress profiles with similar form to those observed. This corresponds to the sample representing an average of 5 nm along the *x* direction away from the disclination dipole, which is the basis of the considered model. A Poisson's ratio of 0.31 is used, and the model considers a twin thickness of 10 px (3.63 nm) [**296**]. Note that in order to qualitatively compare discretely sampled positions in the analytical model to the experiments, in which some degree of beam overlap is expected, we apply a Chebyshev windowing function to the integration. This accelerates the function's descent towards zero in the limit, which was necessary for superposition of the two twin stress fields in our small area of interest.

Our strain measurements are significantly larger than previously seen in most 4D-STEM experiments, for example the ~4% range observed by Pekin *et al* [**153**]. Stresses of the magnitude we have measured are rarely observed under standard loading, but under conditions of severe plastic deformation in a drawn wire these levels are reached macroscopically [**299**].

4.3 Interpretation

The sample has seen significant plastic strain. The tolerance for very high defect density is precisely what makes steels (especially those exhibiting the TWIP effect) amongst the strongest and readily work-hardening engineering materials available. It is observed that the forms and sense of the stress profiles measured are similar to those predicted by the model. The (uniaxial) theoretical strength of a material can be approximated as Young's Modulus / 10 in the absence of

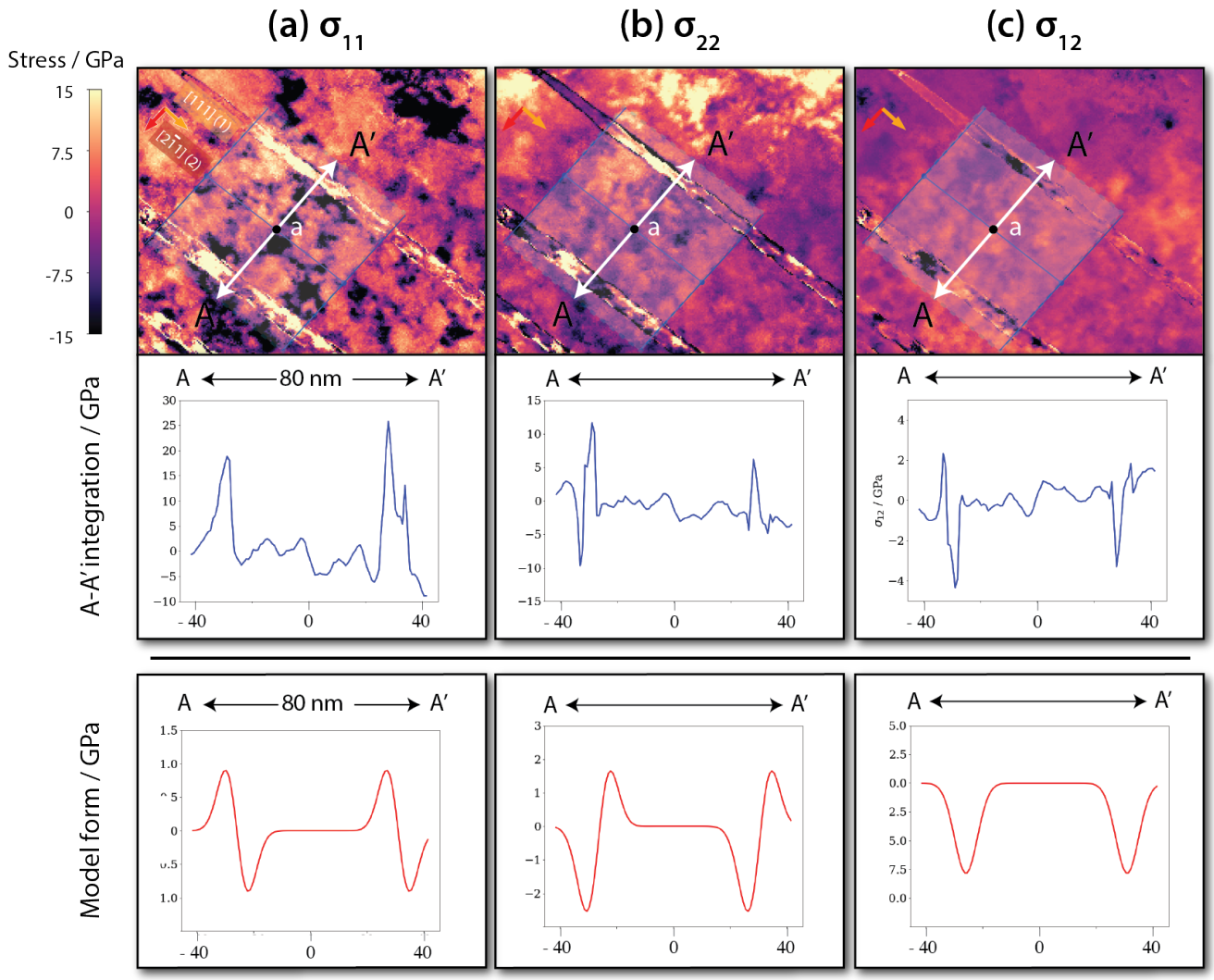

Figure Ap3-4: Elastic stress maps of the *11*, *22*, and *12* components. As in Figure 3, an integrated 1-direction profile (over the shaded region), perpendicular to the twin's length, was calculated by integrating all points in the 2-direction in the highlighted region.

accurate potentials [**300**]. This is around 20 GPa in our case. The presented measurements lie below this threshold, but are fairly close. The stresses predicted by the model appear to exceed our approximate threshold, but follow the same trend as the experimental results.

Koyama et al [**291**] have used electron channelling contrast imaging to investigate hydrogen embrittlement initiation at deformation twin boundaries. They observe that transgranular cracks always propagate along twin boundaries; in this work σ_{11} is measured to be locally very high in a similar material, in accordance with their observation. Furthermore, in this work evidence of significant strain at twin boundaries is provided, which will attract the hydrogen and embrittle the steel, which Koyama *et al* set out as a softening mechanism. As they suggest, this is despite the coherence of the Σ3 boundary, and explains observed hydrogen trapping at such features in similar materials [**301**].

Finally, it is suggested that the significant stress intensity observed parallel and perpendicular to their boundaries controls their thickening. Twin thickness is determined by the tendency to minimise elastic strain energy. Clausen *et al* [**302**] have described the twin internal back-stress generated by matrix constraint of the transformed twin in Mg: the plastic shear provided by the twinning transformation is spread over the incorporating grain, resulting in an equal and opposite elastic strain being contained within the twin. This elastic back-strain (leading to internal back-stress) is what has been observed. The twin thickness is controlled by accommodation of the transformation strain, as there is an energy

penalty to this back-stress. The stress intensity surrounding our nanotwins thus prevents their thickening. This allows for a large number density of fine twins, enhanced grain sub-division, and therefore a greater work hardening rate. As deformation progresses, plastic strain accumulation increases, resulting in a greater elastic back-stress. Equivalently, in larger grains where the twinning strain can be more widely distributed, the back stress is lower and twins are able to grow thicker, as suggested by Rahman *et al* [**293**], [**290**].

5. Conclusions

The following conclusions are drawn:

- Measurement of the elastic strain and stress state around deformation nanotwins in a TWIP steel reveals a significant polarisation, with stresses close to, but less than, the approximate theoretical strength of the material.
- The profiles and sense of the stresses follow those predicted by the analytical model of Müllner et al [**292**]–[**294**].
- The magnitude of the stress state surrounding the twin could explain the observations and provide evidence for hydrogen embrittlement mechanisms associated with twins set out by Koyama *et al* [**291**]. The observed stress field also likely plays a critical role in controlling twin thickness. This determines the rate of grain subdivision and the alloy work hardening rate [**302**].

6. Chapter Acknowledgements

This work was a collaborative effort, involving Dr Abigail Ackerman, Thomas Kwok, Dr Mohsen Danaie, Dr Benjamin Savitzky, and Dr Colin Ophus, as well as TPM, supervised by DD.

TPM and DD would like to acknowledge support from the Rolls-Royce plc - EPSRC Strategic Partnership in Structural Metallic Systems for Gas Turbines (EP/M005607/1), and the Centre for Doctoral Training in Advanced Characterisation of Materials (EP/L015277/1) at Imperial College London. AA acknowledges EPSRC grant IAA EP/R511547/1. TK is grateful for support from A*STAR We thank Diamond Light Source for access and support in use of the electron Physical Science Imaging Centre (Instrument E02 and proposal number EM18770) that contributed to the results presented here. BHS and CO acknowledge funding from the Toyota Research Institute, and that work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

We are grateful to TBB for insight into electron diffraction and micromechanics, and Ben Poole (Imperial) for guidance with basis rotation. We also thank Vassili Vorontsov (Strathclyde) and Alexander Knowles (Birmingham) for helpful conversations when planning the project.

TPM co-acquired the data and drafted the initial manuscript. AKA co-acquired the data and FIB prepared the foil. TPM and BHS performed the analysis, using the py4DSTEM Python package developed by BHS and CO. TK developed the TWIP steel and plastically strained the sample. MD aligned, calibrated and supervised operation of the ePSIC TEM. CO provided critical microscopy and 4D-STEM insight. DD conceived the project and supervised the work. All the authors edited the chapter's corresponding manuscript.

Appendix 4 – supplementary data

1. Results

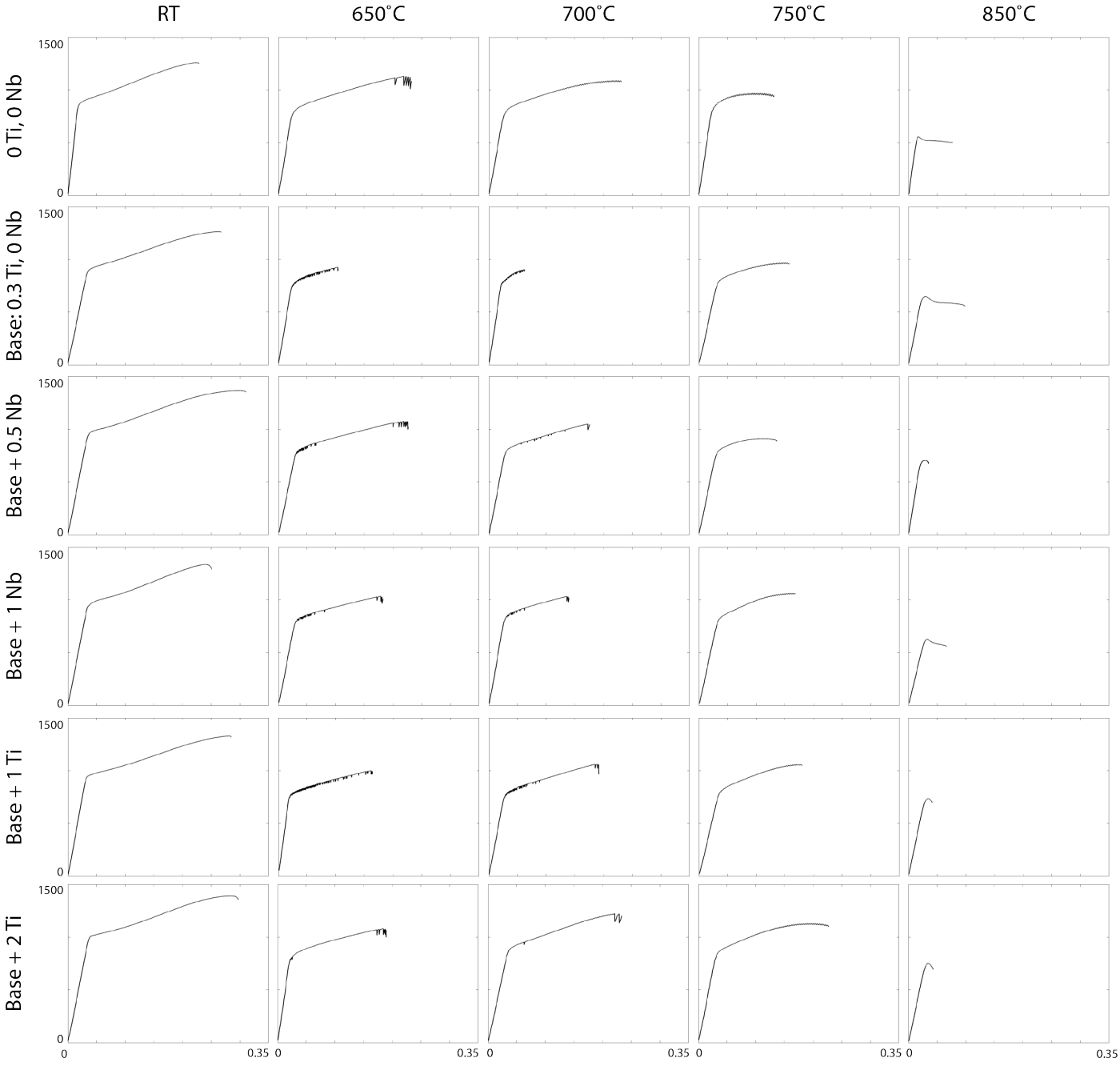

Figure Ap4-1: Full tensile testing (engineering stress against strain) for alloys presented in **Chapter 4**.

Alloy designation	Co	Ni	Mo	Cr	Al	W	Ta	C	B	Zr	Ti	Nb	
(0Ti, 0Nb)	35.7875	35.7875	2	12	10	2.75	1.25	0.3	0.085	0.04	(at%)		
Base: (0.3Ti, 0Nb)	35.759	35.909	2	12	10	2.75	1.25	0.1	0.042	0.04	0.3	(at%)	
Base + 0.5Nb	35.919	35.919	2	12	10	2.75	1.25	0.1	0.042	0.02	0.3	0.5	(at%)
Base + 1Nb	35.269	35.269	2	12	10	2.75	1.25	0.1	0.042	0.02	0.3	1	(at%)
Base + 2Nb	34.769	35.919	2	12	10	2.75	1.25	0.1	0.042	0.02	0.3	2	(at%)
Base + 1 Ti	35.269	35.269	2	12	10	2.75	1.25	0.1	0.042	0.02	1.3	0	(at%)
Base + 2Ti	34.769	35.919	2	12	10	2.75	1.25	0.1	0.042	0.02	2.3	0	(at%)

Alloy designation	Density / g cm^{-3}	Solvus / ˚C	RT Yield / MPa	RT Ult / MPa	650˚C Yield / MPa	650˚C Ult / MPa	700˚C Yield / MPa	700˚C Ult / MPa	750˚C Yield / MPa	750˚C Ult / MPa	850˚C Yield / MPa	850˚C Ult / MPa
(0Ti, 0Nb)	8.67	1021	855	1260	769	1131	778	1092	810	971	557	559
Base: (0.3Ti, 0Nb)	8.61	1042	885	1265	752	927	774	905	777	967	628	650
Base + 0.5Nb	8.69	1049	953	1363	779	1076	787	1049	767	912	673	706
Base + 1Nb	8.74	1062	940	1336	791	1035	824	1038	805	1062	617	629
Base + 2Nb	8.79	1074										
Base + 1Ti	8.61	1067	937	1329	757	1001	768	1064	786	1059	718	736
Base + 2Ti	8.53	1090	1000	1336	788	1082	824	1038	834	1131	733	753

Table Ap4-1: Summary of physical properties of alloys developed in **Chapter 4**.