Atomistic modelling of precipitation in Ni-base superalloys

Eric Schmidt

Department of Materials Science and Metallurgy
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Hughes Hall College

November 2017

This thesis is dedicated to my loving family.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit of 60,000 words, including summary/abstract, tables, footnotes and appendices, but excluding table of contents, photographs, diagrams, figure captions, list of figures/diagrams, list of abbreviations/acronyms, bibliography and acknowledgements.

Eric Schmidt
November 2017

Acknowledgements

First and foremost I wish to express my sincere gratitude to my advisor Dr. Paul Bristowe for enabling me to undertake a PhD in the materials science and metallurgy department at the University of Cambridge and his support. I greatly appreciate the freedom I was given under his supervision to develop and follow my scientific curiosity.

My sincere thanks also goes to all my colleagues and Dr. Jonathan Loh, Dr. Patrick Kiley, Andrew Fowler, Dr. Shafqat Shah, Zeyu Deng, Dr. Georg Schusteritsch, and Tianwei Wang for making the department a fun place to work the last couple of years. I would like to thank in particular the LAMMPS 'whizz' Dr. Jonathan Loh and my father Dr. Frank Schmidt for reading parts of this thesis and Dr. Patrick Kiley for many highly interesting discussion about Bayesian statistics and beyond. I am also very grateful to Andrew Fowler for working with me in exploring electron densities, empirical potentials, and Python. I am also very thankful to Andrew's supervisor Dr. James Elliott for making our collaboration possible.

Last but not least I wish to thank my loving sister Rita, and my loving parents Karen and Frank to whom I am eternally grateful for their unwavering support throughout my entire life.

This work was supported by Rolls-Royce plc and the EPSRC under the grants EP/J500380/1 and EP/K503009/1. Some of the calculations were done at the Cambridge University High Performance Computing Service (HPCS).

Abstract

The presence of the ordered γ' phase (Ni_3Al) in Ni-base superalloys is fundamental to the performance of engineering components such as turbine disks and blades which operate at high temperatures and loads. Hence for these alloys it is important to optimize their microstructure and phase composition. This is typically done by varying their chemistry and heat treatment to achieve an appropriate balance between γ' content and other constituents such as carbides, borides, oxides and topologically close packed phases. In this work we have set out to investigate the onset of γ' ordering in Ni-Al single crystals and in Ni-Al bicrystals containing coincidence site lattice grain boundaries (GBs) and we do this at high temperatures, which are representative of typical heat treatment schedules including quenching and annealing. For this we use the atomistic simulation methods of molecular dynamics (MD) and density functional theory (DFT).

In the first part of this work we develop robust Bayesian classifiers to identify the γ' phase in large scale simulation boxes at high temperatures around 1500 K. We observe significant γ' ordering in the simulations in the form of clusters of γ'-like ordered atoms embedded in a γ host solid solution and this happens within 100 ns. Single crystals are found to exhibit the expected homogeneous ordering with slight indications of chemical composition change and a positive correlation between the Al concentration and the concentration of γ' phase. In general, the ordering is found to take place faster in systems with GBs and preferentially adjacent to the GBs. The sole exception to this is the $\Sigma3$ (111) tilt GB, which is a coherent twin. An analysis of the ensemble and time lag average displacements of the GBs reveals mostly 'anomalous diffusion' behaviour. Increasing the Al content from pure Ni to Ni 20 at.% Al was found to either consistently increase or decrease the mobility of the GB as seen from the changing slope of the time lag displacement average. The movement of the GB can then be characterized as either 'super' or 'sub-diffusive' and is interpreted in terms of diffusion induced grain boundary migration, which is posited as a possible precursor to the appearance of serrated edge grain boundaries.

In the second part of this work we develop a method for the training of empirical interatomic potentials to capture more elements in the alloy system. We focus on the embedded atom method (EAM) and use the Ni-Al system as a test case. Recently, empirical potentials have been developed based on results from DFT which utilize energies and forces, but neglect the electron densities, which are also available. Noting the importance of electron densities, we propose a route to include them into the training of EAM-type potentials via Bayesian linear regression. Electron density models obtained for structures with a range of bonding types are shown to accurately reproduce the electron densities from DFT. Also, the resulting empirical potentials accurately reproduce DFT energies and forces of all the phases considered within the Ni-Al system. Properties not included in the training process, such as stacking fault energies, are sometimes not reproduced with the desired accuracy and the reasons for this are discussed. General regression issues, known to the machine learning community, are identified as the main difficulty facing further development of empirical potentials using this approach.

Contents

List of Figures

List of Tables

Nomenclature

Roman Symbols

(hkl)	Vector in the reciprocal space with the Miller indices *hkl* denoting a plane in real space.
$\langle uvw \rangle$	Family of real space vectors which are related by the symmetry operations of the given type of crystal, e.g. $\langle 100 \rangle = [100]$, $[010]$, and so on in a cubic crystal.
$[uvw]$	Vector with direction *uvw* in real space.
$\mathbb{E}[X]$	The expected value of random variable X.
$\mathbf{F}$	Force, $\mathbf{F} \in \mathbb{R}^3$, acting on an object.
$\mathbf{t}$	N target values, $\mathbf{t} \in \mathbb{R}^N$, for a model.
$\mathbf{w}$	Random variable, $\mathbf{w} \in \mathbb{R}^M$, for the weights of a linear model with M terms.
$\mathbf{y}$	Random variable, $\mathbf{y} \in \mathbb{R}^N$, containing the N predictions of a model.
$\partial_x f$	$\frac{\partial}{\partial x} f$ the partial derivative of f with respect to x.
$\mathrm{var}(X)$	The variance of random variable X.
$\{hkl\}$	Family of vectors in the reciprocal space with Miller indices *hkl*, denoting a family of planes in real space which are related due to symmetry operations available with the given type of crystal.
E	Potential Energy
$F1$ score	Measure of the quality of a binary classifier.

G	Free energy
H	Enthalpy
$KL\,(q \parallel p)$	Kullback-Leibler divergence
P	Pressure
R^2 score	Regression error defined as $R^2 \equiv 1 - \frac{\mathrm{var}(\mathbf{y}-\mathbf{t})}{\mathrm{var}(\mathbf{t})}$
S	Entropy
T	Temperature
U	Internal energy
V	Volume

Greek Symbols

β	Precision variable defined as $\beta \equiv \sigma^{-2}$.
$\mathbf{\Phi}$	Model input, $\mathbf{\Phi} \in \mathbb{R}^{N \times M}$, with N individual data points in an M-dimensional (feature) space.
$\mathbf{\Sigma}_{\boldsymbol{R}}^{-1}$	Covariance matrix for the prior of the model weights $\mathbf{w}$.
Δz	Displacement of the mean position of an approximately flat GB along the GB normal. The mean GB position is defined by averaging over the directions x and y which span the GB.
$\eta_{\mathrm{GS}}(\mathbf{r})$	Ground-state electron density at $\mathbf{r} \in \mathbb{R}^3$.
γ	Chemically disordered Ni-Al phase with fcc lattice.
γ'	$\mathrm{L1_2}$ ordered $\mathrm{Ni_3Al}$ phase.
$\left\langle (\Delta z)^2 \right\rangle$	Ensemble average of GB displacement which is defined as $\left\langle (\Delta z\,(t))^2 \right\rangle \equiv \int_{-\infty}^{\infty} (\Delta z\,(t'))^2\, p\,(\Delta z, t')\, dz$
$\mathscr{N}\left(\mu, \sigma^2\right)$	Normal distribution with mean μ and variance σ^2.
$\overline{\delta^2\,(\Delta)}$	Timelag average of GB displacement which is defined as $\overline{\delta^2\,(\Delta)} \equiv \frac{1}{t-\Delta} \int_0^{t-\Delta} [z\,(t'+\Delta) - z\,(t')]^2\, dt'$

$\phi(\mathbf{x})$	Non-linear transform of **x** from some input space $\mathbb{R}^{M_{in}}$ to feature space $\mathbb{R}^{M}$.
Σ	Value used to characterize special grain boundaries defined by coincidence site lattices.

Acronyms / Abbreviations

ANN	Artificial Neural Network
APB	Anti-phase boundary
at.%	Atom percent
bcc	Body centred cubic
BIC	Bayesian Information Criterion
BOP	Bond Order Parameter
CSL	Coincidence Site Lattice
CTRW	Continuous Time Random Walk
DFT	Density Functional Theory
DIGM	Diffusion Induced Grainboundary Migration
EAM	Embedded Atom Method
EM	Expectation Maximization
EMD	Earth Mover's Distance
ES-CMA	Evolution Strategy using the Covariance Matrix Adaptation algorithm.
fcc	Face centred cubic
GB	Grain boundarie
GGA	Generalized Gradient Approximation
GMM	Gaussian Mixture Model
GP	Gaussian Process

hcp	Hexagonal close packed
KRR	Kernel Ridge Regression
KS DFT	Kohn-Sham DFT
LDA	Local-Density Approximation
MAE	Mean Absolute Error
MAP	Maximum A Posteriori
MD	Molecular Dynamics
ML	Maximum Likelihood
NPT	Thermostat conserving the number of particles, the pressure and temperature.
OF DFT	Orbital-free DFT
PDF	Probability Density Function
RMSE	Root Mean Squared Error
RVM	Relevance Vector Machine
TCP	Topologically close packed

.

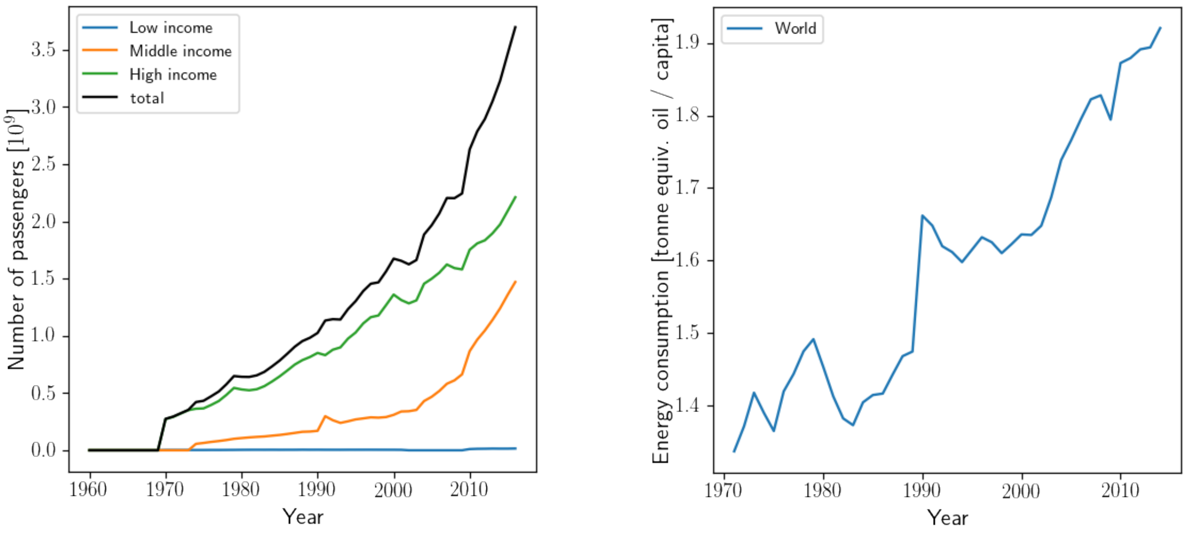
Chapter 1

Introduction

1.1 An Overview

This work is concerned with the understanding of phase transformations at the atomistic level in Ni-base superalloys. Improving the properties of Ni-base superalloys is of importance because they are the material of choice in high temperature, high load, regimes in gas turbine engines, which find use in air travel and gas power plants.

Air travel and gas power plants play important roles in the travel and power production industries. The importance of air travel and gas power plants can be seen from the data shown in figure 1.1 a) and b), respectively. The trends in both the number of air passengers per year and the energy consumption in equivalent tonnes of oil per capita per year are upwards, approaching 4 billion passengers and 2 tonnes respectively. Thus, these industries have increased by factors of about three and two, respectively, since 1970. Any improvement in the efficiency of gas engines following from improvement in the material properties could therefore have a significant impact for companies and the environment, effectively reducing the required amount of fuel.



(a) (b)

Figure 1.1: World Bank data on a) the total number of air passengers per year, as well as for specific income groups [11], and b) the total world wide energy consumption [12].

One way of improving the efficiency of gas engines is to increase the combustion temperature, which requires materials and component design, particularly of disks and blades, to withstand the load at the increased temperature. This necessitates the consideration of the microstructure and constituent phases of the disks and blades. Both the microstructure and the phase composition are linked in complex ways, depend on their initial state, and control the life time of the components.

The difficult task of optimizing superalloys is reflected in the complexity of the citation network from academia shown in fig. 1.2, where scientific articles are vertices in a graph connected by citations, accumulating 8,217 publications in the Web of Science's core collection at the time of writing of this thesis. The vertices are coloured by the detected citation community (area of research) and scaled by their overall importance. The research communities this analysis illustrates are centred around rafting (teal), element partitioning (purple), coating (red), processing (green), machining (yellow), fatigue (blue) and density functional theory (gold).

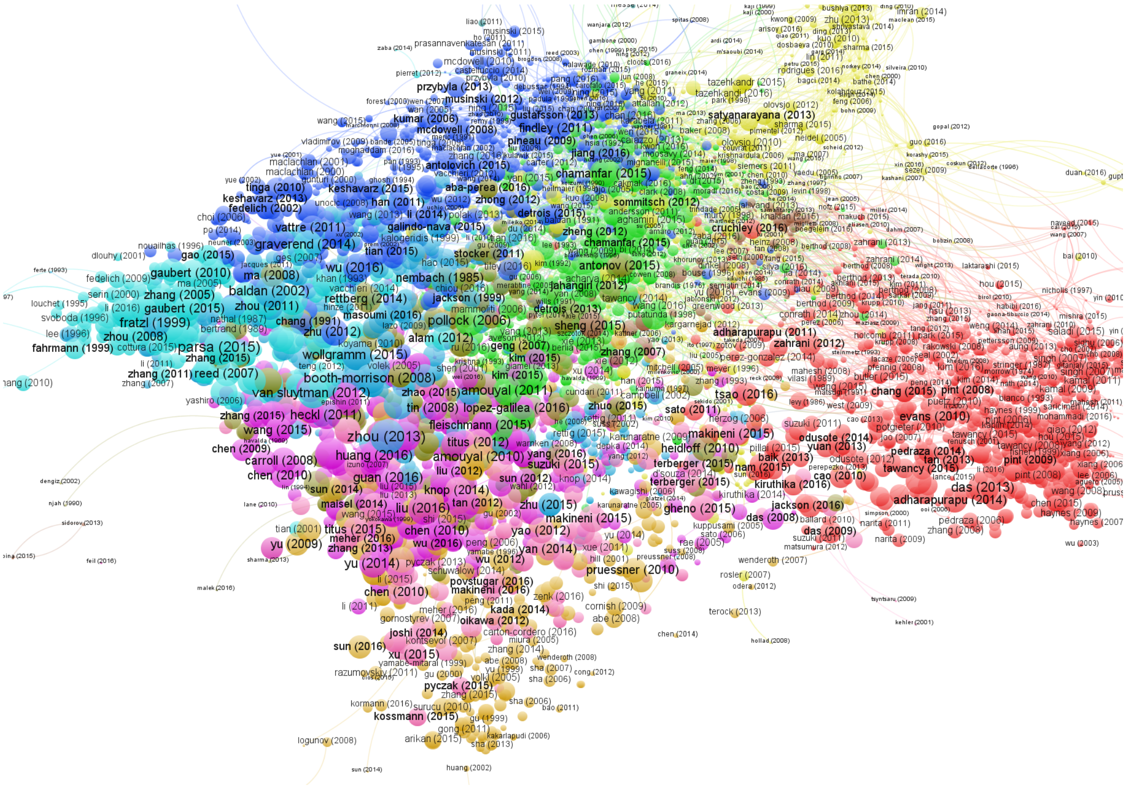


Figure 1.2: Citation network belonging to the topic ‘superalloys’ in Web of Science’s core collection showing 8,217 publications. The network is visualized with VOSviewer [247] which used the citation links alone to detect the following research groups indicated by colour: teal ~ rafting, purple ~ element partitioning, gold ~ DFT, red ~ coating, green ~ processing (solidification, welding, deformation), yellow ~ machining and blue ~ fatigue.

In this thesis we will use Density Functional Theory (DFT) and Molecular Dynamics (MD) to investigate some atomic-scale properties of Ni-base superalloys. The focus of this work is on the ordering phase transition which forms γ' (Ni_3Al) from γ (fcc matrix), particularly in the presence of grain boundaries and so-called serrated edge grain boundaries, which have the potential to improve the creep life of disk components. Another part of this work is the development of a systematic approach for constructing semi-empirical potentials which describe the formation of topologically close packed phases such as the σ (tetragonal) and δ (orthorhombic) phase.

We will develop a robust method for identifying chemically ordered configurations and grain boundaries at high temperatures. Employing a semi-empirical Al-Ni potential we will demonstrate that $\gamma \rightarrow \gamma'$ ordering can be observed in MD simulations. We will also show that the presence of solutes has a significant impact on the migration of some grain boundaries. This will be explained in terms of diffusion induced grain boundary migration

(DIGM). DIGM itself will be posited as a possible initiator of the grain boundary serration phenomenon.

In order to create glue-type semi-empirical potentials, which are based on host electron densities, we will introduce a method for modeling and predicting electron densities which solely depends on Bayesian linear regression. The proposed approach for the electron density regression will be shown to be applicable to materials with a range of different bonding types. Using electron density regression, we will develop semi-empirical potentials for the Al-Ni system that make accurate predictions of the energies and forces for test structures. The ability of these potentials to predict properties not included in the training set will be discussed in detail and a way forward for the systematic development of semi-empirical potentials for multi-component systems, including the δ and σ phases, will be presented.

1.2 Structure of This Thesis

In order to investigate $\gamma \rightarrow \gamma'$ phase transformations at an atomistic level, we have developed a plan to: i) develop the ability to classify structures on a per atom basis at high temperatures, ii) investigate the ordering behaviour in single alloy crystals and at grain boundaries, and iii) find a means of developing semi-empirical potentials, using Al-Ni as a test case, with readily available reference potentials.

This thesis is therefore organized as follows: Chapter 2 reviews the materials science aspects of this work and in particular phase transformations, grain boundaries, superalloys, and the phenomenon of grain boundary serration. This is followed in Chapter 3 by an overview of atomistic simulations, DFT, semi-empirical potentials and optimization algorithms. In Chapter 4, the Bayesian approach to regression and classification will be introduced and illustrated with an example of tracking diffusing Pt atoms on an Si surface as seen in STEM. In Chapters 5, 6, and 7, we will present our results on the $\gamma \rightarrow \gamma'$ transformation in single crystals, $\gamma \rightarrow \gamma'$ transformations in bicrystals containing migrating grain boundaries, and the development of semi-empirical potentials, respectively. In Chapter 8 we will give a summary and make some suggestions for future research.

Chapter 2

Superalloys

Superalloys are defined as “alloys based on Group VIIIA-base elements developed for elevated-temperature service, which demonstrate combined mechanical strength and surface stability.” [211]. They can be sub-divided into Fe-, Ni- or Co-based superalloys and are usually used in components operating under conditions of high homologous temperatures (temperature relative to the melting temperature) in turbine engines of mobile and stationary gas engines. The components are physically defined by their chemical composition, their phase constitution (in particular the γ' content) and the associated microstructure. The development of superalloys is closely connected with the history of processing techniques. In the first half of the twentieth century superalloys were mainly wrought and investment cast alloys until the advent of vacuum induction melting and directional solidification, allowing fine grained control of the microstructure and the impurity concentration leading ultimately to single crystals [186, 211].

In the following section we will give a brief introduction to gas engines and continue with an overview of the strengthening phases and mechanisms in Ni-base superalloys and their application to blades and disks in turbine engines. This is followed by a discussion of phase transformations in section 2.2, with focus on disorder $\rightarrow$ order reactions. After a short excursion to particle diffusion and random walks in section 2.3 we will discuss grain boundaries (GBs) and their connection to random walks, diffusion induced grain boundary migration (DIGM) and the peculiar effect of serrated (edge) GBs. At the end of section 2.3 we will also outline how the phenomena can be connected to the atomistic simulations which were undertaken in this work.

2.1 Superalloys in Gas Engines

Superalloys are found in stationary gas engines for electricity generation in power plants and in mobile gas engines for the jet propulsion of airplanes. Their application in blades and disks requires them to withstand high loads at high temperatures. This means that they need to fulfill a range of conditions. Among them are weight, shape consistency, fatigue, creep and hot corrosion (a form of corrosion at high temperatures in the presence of salts which destroy protective layers).

Depending on how the alloys are used, there is appreciable difference in the life cycles that they have to endure. In the case of power generation, they operate under longer sustained load and the corrosives are directly due to the fuel. For jet propulsion, on the other hand, the gas engines go through more operating stages depending on the take-off / landing frequency and they experience a variety of atmospheres / environments possibly containing volcanic ash, industry exhaust gasses, charged gasses and others.

However, both types of gas engine can be approximated to first order by the *Brayton cycle*, which describes gas engines with internal combustion. To understand the optimization process for superalloys we will now briefly discuss the thermodynamics of gas engines. This will be followed by a discussion of Ni-base alloys, their strengthening mechanisms and applications in turbine blades and disks.

2.1.1 Gas Engines

An approximation to real engines are *thermodynamic cycles* which break down the workings of the engine into idealized *thermodynamic processes* (isotherm, isobaric, isochoric, isentropic, adiabatic). Using thermodynamic cycles, one posits that the engine is in thermodynamic equilibrium at any given point, neglecting real world effects such as friction. Combinations of these thermodynamic processes result in thermodynamic cycles if, and only if, they lead to the same initial state.

These idealized thermodynamic cycles can be broken down into *internal* and *external combustion cycles*, with possibly the most well-known external combustion process being the Carnot cycle (isentropic compression, isothermal heat addition, isentropic expansion, isothermal heat rejection). For gas engines, the common approximation is the Brayton cycle (adiabatic compression, isobaric heat addition, adiabatic expansion, isobaric heat rejection). The work a thermodynamic process does is captured by the path integral $W = \oint PdV$. The work is therefore equal to the area encapsulated by the chosen path. In the Brayton cycle this

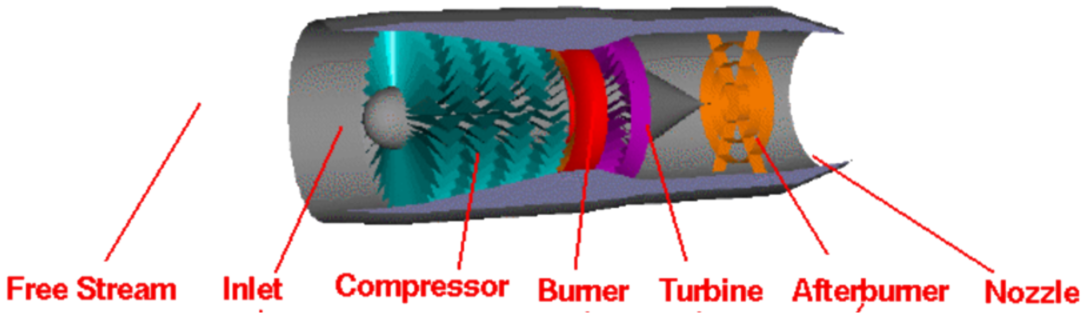


Figure 2.1: Schematic of a jet propulsion / mobile gas engine, taken from [160].

can be used to approximate the thermal efficiency with $1 - T_{\text{exhaust}}/T_{\text{entry}}$, where T_{entry} and T_{exhaust} are the temperatures at turbine entry (or after combustion) and upon exhaust respectively [211]. The different stages for gas engines are as follows 1) inlet of fuel and volume compression in the compressor, 2) combustion in combustor or burner, 3) inlet of combusted fluid into the turbine and volume expansion and finally 4) exhaust from the turbine engine. Thus one way to improve the Brayton engine is to increase the difference between T_{entry} and T_{exhaust} . Assuming that T_{exhaust} is already at the surrounding temperature the only option is to increase T_{entry}.

In the case of power generation almost all chemical energy is converted into mechanical energy by rotating the turbine shaft. For jet propulsion the chemical energy is mostly used for acceleration of the hot fluid out of the jet nozzle producing thrust. A schematic of a jet propulsion engine is shown below in fig. 2.1.

2.1.2 Strengthening Mechanisms and Phases

Ni-base superalloys are currently the preferred choice when it comes to critical components in gas engines. The use of Ni as an element in these alloys has been attributed to its close packed fcc structure and bonding with d electrons, resulting in tough and ductile crystals. Furthermore, Ni is stable in its fcc structure up to the melting temperature without a solid-solid phase transformation and exhibits low enough rates of diffusion thus impeding creep.

Ni-base superalloys commonly have around a dozen elements present with mostly 10-20 at.% Cr, 8 at.% Al and Ti, 5-10 at.% Co and traces of B, Zr and C for GB design along with Si, P, S, O and N from the production process. Other common additions are Mo, W, Nb, Ta and Hf. The main strengthening phase is the ordered γ' (Ni_3Al) phase which is mostly coherent with the disordered γ phase. Other phases present are carbides, mostly MC, $M_{23}C$ and M_6C, borides, oxides and so-called topologically close packed (TCP) phases such as the σ, μ and Laves phases. Grouping roughly by partitioning preference we find for γ: Ni, Co, Fe, Cr, Mo and W; for γ': Al, Ti, Nb, Ta and Hf; for GBs: B, C, Zr; for carbides Cr, Mo, W, Nb, Ta, Ti and Hf and for oxides: Al and Cr. TCP phases form mostly due to refractory elements. The design of superalloys, considering these phases, is a complex process but can be beneficial, for example by intentionally allowing borides and carbides to form to some degree in polycrystals leading to improved creep properties.

The co-existence of γ and γ' is quite fortunate because γ is ductile due to the availability of more than 5 slip systems in the fcc lattice without necessary thermal activation. It also allows the presence of a variety of solutes, such as Co, Fe, Cr, Mo, W, Ti and Al which add elastic strain fields impeding the motion of dislocations. γ furthermore forms passivating oxides with Cr and Al protecting the component surface. γ' on the other hand is an $L1_2$ ordered phase, which is of type A_3B, where element A is located at fcc face sites and element B at the corner site. Common compositions for γ' are $(\mathrm{Ni},\mathrm{Co})_3\,(\mathrm{Al},\mathrm{Ti})$. The commonly observed γ' growth sequence is: spherical (surface to volume optimization)$\rightarrow$ cuboidal (strain optimization) $\rightarrow$ dendritic [53, 188]. This sequence depends on the γ'/γ misfit requiring non-zero misfits for the typical cuboidal shape [188]. The commonly observed precipitate-size distribution is not unimodal but rather has three modes each referring to primary, secondary and tertiary γ' occurring during thermal treatment in that order.

A significant factor in the strengthening of Ni-base superalloys is precipitation hardening caused by ordered γ' precipitates. Although details are still under investigation it is established that the ordering has an effect on stacking faults and dislocations in γ' precipitates. Stacking faults are generally referred to as *intrinsic*, which equals the removal of a plane, or *extrinsic*, the addition of a plane. Different to γ is that in γ' the order leads to stacking faults defined by the type ordering violation, e.g. the *anti-phase boundary* (APB), *superlattice intrinsic stacking fault* (SISF), *superlattice extrinsic stacking fault* (SESF) and *complex stacking fault* (CSF). An atomistic model of an APB resulting from a displacement at $\frac{1}{2}[\bar{1}01]$ on the (010) is shown in fig. 2.2. Models for intrinsic and extrinsic stacking faults relative to the un-faulted crystal are shown in fig. 2.3. The difference between these stacking fauls is their respective shift vectors, e.g. in the $\{111\}$ plane: APB $\frac{a}{2}\left\langle\bar{1}01\right\rangle$, SISF $\frac{a}{3}\left\langle\bar{2}11\right\rangle$, SESF $\frac{a}{3}\left\langle\bar{2}\bar{1}1\right\rangle$ and CSF $\frac{a}{6}\left\langle\bar{1}\bar{1}2\right\rangle$ (leading to Al-Al nearest neighbours). APBs receive particular attention due to their relation to dislocations. The shortest lattice vector for dislocations in γ' is the $a\langle 100\rangle$ vector, which is not part of the densest slip system. But since γ' is still fcc, $\frac{a}{2}\left\langle 1\bar{1}0\right\rangle\{111\}$ remain the shortest displacements. This conflict is resolved by pairing up $\frac{a}{2}\left\langle 1\bar{1}0\right\rangle\{111\}$ dislocations forming *superdislocations*. Together they have a valid displacement. Each individual dislocation which belongs to a superdislocation is referred to as a *superpartial*. The pairing up results in the second dislocation removing the APB generated by the first one. When both superpartials are in the same γ' precipitate then they are referred to as 'strongly coupled', or 'weakly coupled' otherwise. Thus, the cutting of γ' by dislocations significantly depends on the APB energy, introducing a penalty for APB formation. To further complicate things it is found that dislocations themselves can dissociate. This is justified using Frank's rule, which scales the energy of a dislocation quadratically with its Burgers vector. Thus, partial dislocations with smaller respective Burgers vectors are favoured, but if added up they still produce the original dislocation's Burgers vector. Among these dissociations is the example: $\frac{a}{2}\langle 110\rangle = \frac{a}{6}\langle 211\rangle + \frac{a}{6}\left\langle 12\bar{1}\right\rangle$ which occurs on the $\{111\}$ plane producing two *Shockley* partials with a (S)ISF between them. The distance between these two partials is determined by a balance between the energy to form the stacking fault and the repulsive interaction between them. It can be shown that the separation distance for Shockley partials scales inversely with the stacking fault energy. The ordering in γ' leading to superdislocations and superpartials creates further mechanisms which can lock dislocation movement, i.e. the Kear-Wilsdorf lock. Dislocations tend to form complex dislocation networks in the γ phase around the γ' phase because of the energy penalty in cutting precipitates. This is done, for example, by Orowan bowing, where dislocations bow out so strongly that they join themselves again behind the precipitate and move on, leaving behind a dislocation ring around the precipitate. By limiting the way dislocations can move around precipitates significant strengthening can be achieved, as observed for single crystals with aligned cuboidal γ' precipitates which improve the creep life of the superalloy compared to unaligned irregular shaped precipitates [34, 175]. Thus, relevant factors for the precipitation hardening are the γ' volume fraction, the γ' radius, and the shape / orientation of the precipitates. A peculiarity of Ni-base superalloys is that they exhibit at high volume fractions of γ' an increase in critical resolved shear stress (CRSS), which is the stress necessary to initiate plastic shear deformation, with temperatures up to around 800 °C before falling off. This is anomalous behaviour because most alloys have a monotonically decreasing CRSS with increasing temperature. This effect is partially explained by the increased availability of $\{100\}$ slip planes with increasing temperature leading to additional locking of dislocations.

In iron rich alloys a phase resembling two stacked γ' unit cells is found and therefore referred to as γ'' with the composition $Ni_3\,(Nb, Al, Ti)$. This phase is an ordered body-centred tetragonal (BCT) phase of $D0_{22}$ type, with similar lattice constant a to γ' but $c_{\gamma''} \approx 2a_{\gamma'}$. It

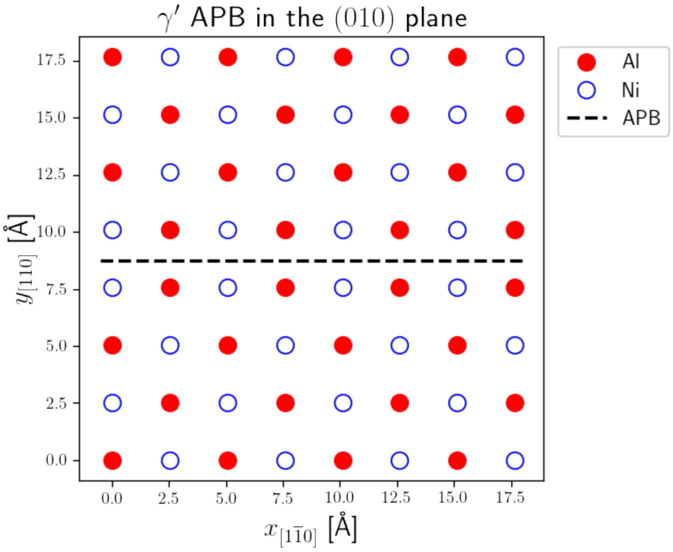


Figure 2.2: Anti-phase boundary in γ'on the (010) plane formed by a $\frac{1}{2}\left[\bar{1}01\right]$ displacement of the upper half against the lower half. The atoms shown correspond to $z = 0$Å.

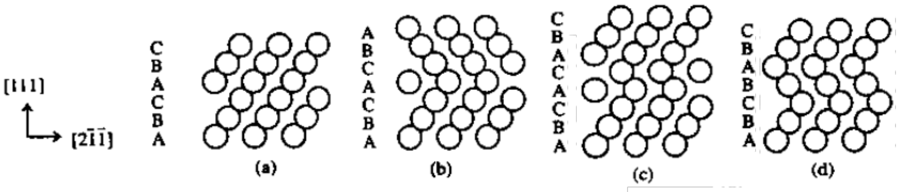


Figure 2.3: Stacking faults in fcc crystal, taken from Hammer *et al.* [92]: a) unfaulted crystal, b) twin, c) intrinsic stacking fault and d) extrinsic stacking fault.

is also ordered but has a more disk-shaped morphology.

Carbides can play a positive role in Ni-base alloys due to their preference to form at grain boundaries. In excess they may be detrimental but in small quantities carbides can improve creep life because they suppress GB sliding. Their morphologies depend on the carbide type and are found to be cubic for MC, blocky for $M_{23}C_6$, and blocky / Widmanstaetten for M_6C. MC commonly forms during the solidification of the alloy.

The final class of phases to mention are TCPs with representatives such as the μ and the σ phase. The μ phase has A_6B_7 stoichiometry while the σ phase has AB stoichiometry. TCPs have a high uniform packing density, include some directional bonding and, most importantly, can be characterized as stacked layers of Kasper coordination polyhedra [69, 70, 91], making them one of the more unusual phases. They are deemed generally detrimental since they are brittle and draw refractory elements, but investigation is still ongoing [261].

Thus, the aims of alloy design for Ni-base superalloys are, in general, to increase strength, creep resistance and resistance to surface degradation. If GBs are present, the GB size-distribution needs to be considered and the concentration of solutes known to interact with GBs. The size-distribution itself is also a contributor to the material's strength, as characterized by the famous Hall-Petch relation, which finds the that the yield stress is inversely proportional to the square root of the average grain diameter. Small grains as a rule result in increased tensile strength but decreased rupture strength and creep life, whereas large grains lead to more ductile behaviour. Hence it is desirable to design the grain size distribution with respect to the component under consideration to maximize the possible performance, which is of particular relevance for disks and polycrystalline blades.

2.1.3 Blades and Disks

The main materials aspects to consider for the degradation of blade and disk components are creep, fatigue and corrosion. Creep occurs under mechanical load at elevated temperatures aided by vacancy diffusion processes along GBs and through the bulk leading to incremental deformations. This can be problematic in cases where the geometric tolerance of the components has to be matched precisely. Fatigue is the decrease in mechanical resistance that a component can endure over its lifetime. One distinguishes low cycle fatigue (LCF), which occurs once per engine start or load change, and high cycle fatigue (HCF), which occurs continuously due to component vibrations and resonance. Corrosion is mainly an issue for LCF, particular at high temperature in the presence of salts causing hot corrosion. Blades and disks need to be designed with respect to these aspects. In the following we will first discuss blades and then disks.

Blades are designed as single and polycrystals, depending on their location in the engine [74], with single crystals near the combustor. Blades are fixed on disks with dovetail joints. Their objective is to convert the thermal energy of the hot fluid into mechanical energy. The design of the blades then also depends on the application process, i.e. there are larger blades in stationary gas engines to maximize the energy converted, with weight less of a concern. Blades rotate around 100,000 rotations per minute thus leading to severe centrifugal stresses. Therefore blades, particularly in the first rows, need to be temperature controlled to suppress creep. This is solved by designing them, so they are hollow to enable air cooling which limits homologous temperatures to 0.9. These blades are usually single crystals created with investment casting. During the casting process the solidification takes place with the formation of dendrites, which are a product of temperature gradient and withdrawal speed. The single crystal blades are commonly aligned along $\langle 100 \rangle$ because that is the preferred growth direction in fcc crystals and, fortunately, also leads to better LCF performances than the $\langle 111 \rangle$ orientation. Controlling the quality of the casts include identifying the presence of ceramic particles, the wall thicknesses of the hollow blade and the presence of high-angle grain boundaries. In terms of the alloy chemistry, high Al, Ti and Ta concentrations are used to promote γ' volume fractions of 0.7 to maximize the γ/γ' interface effects and to ensure that γ is the only stable phase for some solutioning window. Particular attention is given to refractory elements for solid solution strengthening while avoiding TCP phase formation, and the level of sulphur to minimize damage caused by hot corrosion.

Disks are polycrystalline, and their task is to transfer the kinetic energy from the blades to the shaft. They thus experience the pull of the blades and a significant variation in temperature between blades and shaft with the highest temperature approached being about 650 °C. Hence a major concern is the fracture of a blade or the failure of a disk itself, both of which are potentially fatal for the entire gas engine. Disks are produced by either an ingot casting or powder metallurgy route and then machine forging. The powder metallurgy is generally more complex but allows for direct control of the homogeneity of the concentration profiles and grain sizes by filtering and mixing of the powders. The aim for disks is to produce high yield stress, tensile strength, fracture toughness, fatigue crack initiation resistance and low fatigue crack propagation. This is achieved with 0.4 - 0.55 volume fractions of uniformly distributed γ' precipitates. The precipitate sizes are set through annealing, with supersolvus annealing leading to coarser grains and subsolvus leading to smaller grains due to the pinning of GBs by primary γ'. Subsolvus annealing results in larger primary γ' precipitates. Since both annealing routes have positive aspects for the component properties, a solution found which combines both is ‘dual’ heat treatment. This aims to develop larger, more creep resistant, grains at the rim and smaller grains, for higher strength, in the bore. Small quantities of boron and carbide are used as well to optimize the GBs further.

2.2 Phase Transformations

Solid state phase transformations can be roughly grouped into three categories: 1) *formation/precipitation or dissolution* of a second phase crossing one phase boundary, such as a specific temperature $\alpha \rightarrow \alpha + \beta$, 2) *complete transformation* of one phase to another crossing two phase boundaries maintaining the same concentration $\alpha \rightarrow \beta$ and 3) *decomposition* of a single phase into two new phases crossing three phase boundaries $\alpha \rightarrow \beta + \gamma$ [86]. The first case will be of particular interest because it is an example of the formation of γ'-phase from a supersaturated γ-phase host.

Further distinctions are also being made depending on the spatial distribution of precipitates, being either *heterogeneous* or *homogeneous*. Heterogeneous precipitation is most notable in cases of low supersaturation where generally fewer precipitates form and those that form have some advantage like reduced surface energy. Another distinction which is made relates to GBs explicitly, also referred to as *discontinuous* precipitation and will be discussed in more detail in section 2.4.2.4.

Phase stability, and possible phase transformation, is determined by the free energy G of a system. G depends on the enthalpy H, the temperature T and the entropy S by

$$G = H - TS, \tag{2.1}$$

where $H = U + PV$ with U the internal energy, P the pressure and V the volume. G considered here is in its simplified form focusing only on ordering, chemical composition and temperature. Generally if $G_2 - G_1 < 0$ then phase 2 forms preferentially over phase 1 for some given T, P and V.

In order to understand the results in Chapter 6 it is useful to discuss phase transformations assuming some knowledge about G for γ and γ'. We will see how *nucleation & growth* and *spinodal decomposition* arise from considerations of the G curves in the following. Note that all discussion considering thermodynamic quantities such as G is only valid for thermodynamic equilibria, which is an approximation to real observations.

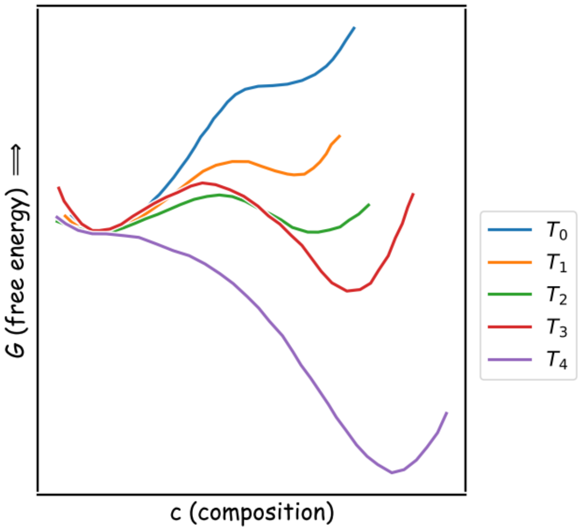


Figure 2.4: Free energy G as a function of the composition c for various temperatures T with: $T_4 < T_3 < T_2 < T_1 < T_0$. The points for the plot were taken from Soffa *et al.* [215].

2.2.1 Nucleation & Growth and Spinodal Decomposition

Nucleation & growth and spinodal decomposition are two extremes of the precipitation process. Paraphrasing Cahn [32], *nucleation & growth,* on one hand, is the case of a concentration change large in degree but small in extent, like dropping a small droplet (small extent) of a certain composition which is close to a more stable phase (large degree) into a metastable host. *Spinodal decomposition,* on the other hand, is a concentration change small in degree but large in extent and is described by small fluctuations throughout the host phase, becoming stronger with uphill diffusion and approaching more stable phases.

The actually observed precipitation process itself is usually a combination of both nucleation & growth and spinodal decomposition and may transition smoothly between them depending on the environment. The typical way to visualize the difference between spinodal decomposition and nucleation & growth is to imagine the former as uphill diffusion becoming more severe over time and the latter as the sudden appearance of ‘clusters’ with the correct stoichiometry and ordering [132]. The plots of G as a function of composition (c) and temperature (T) which follow are based on Soffa *et al.* [215] but generated by the author of this work.

An example of $G(c)$ at various T is illustrated in fig. 2.4. The curves shown characterize G for the same lattice type but different compositions assuming chemical disorder. An ordered phase would appear as a separate G curve.

Based on G curves one can construct phase diagrams. For this one uses the *tangent rule* combining the lowest stable phases. If at least two minima exist then we find the lowest possible G by considering the chemical potential $\mu_i \equiv \partial_{N_i} G$, where N_i is the mole of element i. This goes back to the original work of Gibbs. Equilibrium is achieved if $\mu_i = \mu_j$ for all i, j. Thus, equilibrium is on the shared tangent between both phases.

An indicator of how precipitation will proceed is $\partial_c^2 G = 0$. Compositions where this condition is true are also referred to as the *chemical spinodal* by Laughlin *et al.* [132]. When $\partial_c^2 G > 0$ then nucleation & growth is favoured, which commonly requires phase nuclei to exceed a threshold posed by the volume to surface energy balance. In the case of heterogeneous precipitation this threshold may be locally lower due to other pre-existing particles, lowering the surface energy. Hence this is a highly stochastic process which can be characterized by an incubation time required for precipitates to form. If, on the other hand, $\partial_c^2 G < 0$ then spinodal decomposition is favoured, which usually occurs quite homogeneously by increasing amplitudes of the concentration profile concomitantly or preceded by ordering [214].

A useful technique for understanding disorder → order transitions is the graph method developed by Soffa *et al.* [215] which captures other work such as the mean field and phase field approaches [32, 36, 100, 119]. Ordering can occur *homogeneously* or *heterogeneously*. They are also referred to as *continuous* and *congruent* ordering and either take place everywhere with uniform probability or with preferences [119, 215]. Both cases are illustrated in fig. 2.5 with G_d and G_o, the free energy curves for the disordered and ordered phases respectively. The connection $A \rightarrow B$ starts with a disordered configuration A, which is energetically *more* stable than a completely ordered phase of the same composition. But A is *metastable* with respect to a mixture of ordered and disordered phases. Ordering will occur but *heterogeneously* because of this metastability. Since G_d is concave upward the system decomposes via nucleation and growth until the lever rule is satisfied at B. The connection $C \rightarrow D$ on the other hand starts with a disordered configuration C which is energetically *less* stable than the ordered configuration and hence *unstable* with respect to a mixture of ordered and disordered phase leading to homogeneous ordering. Because the disordered phase is unstable it can be expected that ordering takes place until $G_o(c)$ is reached. From there on nucleation & growth takes over until the composition on the tangent is reached at D because $\partial_c^2 G_o(c) > 0$. If, on the other hand, $\partial_c^2 G_o(c) < 0$, as is the case in fig. 2.6, then homogeneous ordering ensues by spinodal decomposition. Because the spinodal decomposition is ‘conditional’ on the ordering it is also referred to as the *conditional spinodal* [2]. Thus, ordering is a fundamental step for both nucleation & growth and spinodal decomposition.

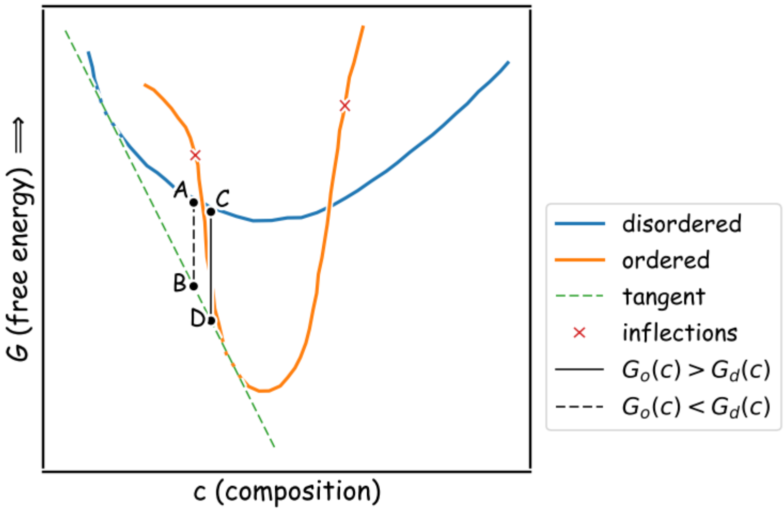


Figure 2.5: Free energy G versus composition c for an ‘ordered’, $G_o(c)$, and a ‘disordered’, $G_d(c)$, phase in orange and blue respectively. The black dashed and solid lines represent two hypothetical cases of precipitation with two initially disordered configurations A and C and two resulting equilibrium configurations B and D. The points for the plot were taken from Soffa *et al.* [215]

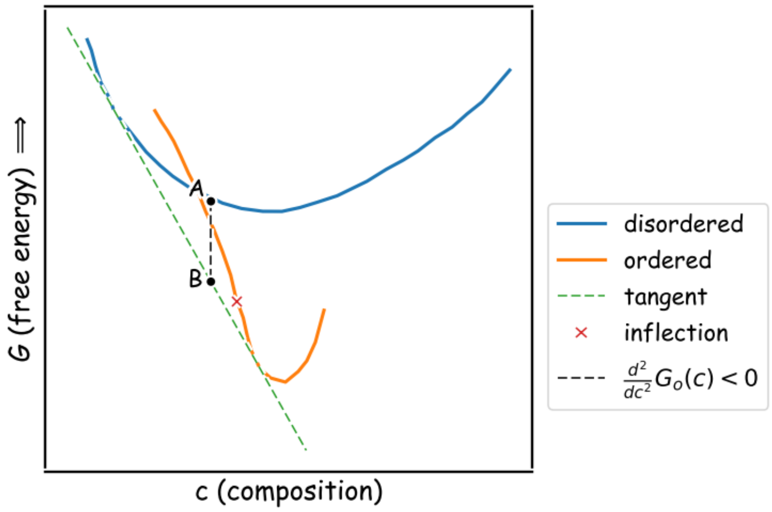


Figure 2.6: Free energy G versus composition c for an ‘ordered’ and a ‘disordered’ phase in orange and blue respectively. The black dashed line represents the hypothetical case of precipitation with an initially disordered configurations A and resulting equilibrium configurations B where c is chosen such that it is above the inflection point of the ordered free energy function, $G_o(c)$. The points for the plot were taken from Soffa *et al.* [215]

2.2.2 Formation of γ' in Ni-based Superalloys

Because disorder $\rightarrow$ order processes, and in particular the $\gamma \rightarrow \gamma + \gamma'$ reaction, are of significant technological importance their underlying theory has received significant attention principally by Soffa, Laughlin, Chen, Khachaturyan and Mishin [3, 36–39, 84, 119, 174, 215–217, 254]. The study of this process is aided by the development of atom probe techniques allowing insights into preferential solute sites in γ', γ/γ' interfaces, segregation, clustering and ordering [25, 26].

The general consensus regarding γ' formation seems to be that local ordering via vacancy jumps and chemical decomposition can occur concomitantly or subsequently and is usually the mechanism for precipitation of ordered phases from a disordered host in most situations [3, 36, 37, 216], where low supersaturations lead to nucleation & growth and high supersaturations lead to spinodal decompositions. However, the temperature, cooling rate, and other factors may cause nucleation & growth and spinodal decomposition to occur simultaneously. This has led to some researchers finding interconnected fluctuations of concentration profiles with concomitant ordering and spinodal decomposition [6, 224, 228, 229, 248] and others to find nucleation and growth [167, 169, 202]. It has also been claimed that the mechanism may be different for primary, secondary and tertiary γ' with primary γ' forming due to nucleation and growth and the other two by spinodal decomposition [212]. Most of the studies which investigate the quenching process also find that γ' formation takes place so rapidly in supersaturated conditions that it is difficult to carry out quenching rapid enough preventing γ' formation.

Although a lot of data has already been collected on γ' formation research is still ongoing both computationally and experimentally. Simulations, in particular, allow progressively deeper insight into the atomic scale mechanisms, especially with the development of more accurate semi-empirical potentials. Thus, to gain insight into the dynamics of this transition, or at least the initiation of it, we will study supersaturated the γ phase using MD. We will find that the simulations are consistent with the general picture of a rapid ordering process, including indications of chemical decomposition.

2.3 Diffusion

Different mechanisms which simulate random walks in 1D will be used to model GB migration along the GB plane normal. The name ‘diffusion’ is chosen as an umbrella name because these random walks belong to the field of ‘diffusion’. Later on, we will consider

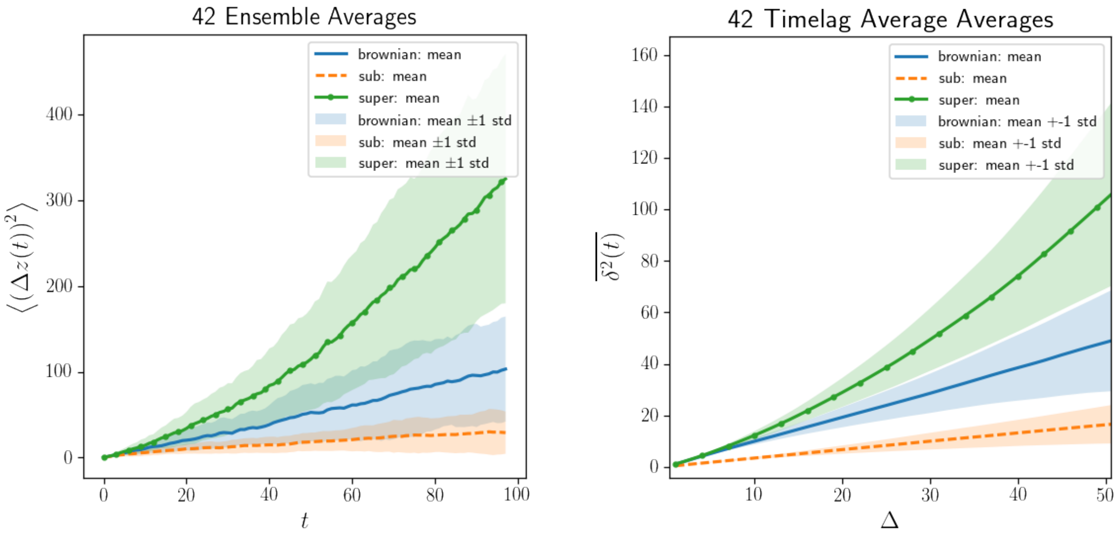
GBs as particles which are diffusing / migrating according to these mechanisms. Diffusion in this context should not be confused with 'grain boundary diffusion' which refers to atom motion within GBs. In what follows we will give a brief general introduction to diffusion. For a more GB-specific discussion see section 2.4.2.3.

Assuming 1) particles are independent, 2) particle displacements, $\Delta z = z(t) - z(t=0)$, are independent and 3) particle displacement distributions, $p(\Delta z, t)$, are symmetric centered at the origin, Einstein derived the diffusion equation $\partial_t p(\Delta z, t) = K_1 \partial^2_{\Delta z} p(\Delta z, t)$ using the central limit theorem, where K_1 is the diffusion constant and t represents time. The diffusion equation solves as $p(\Delta z, t) = \frac{1}{\sqrt{4\pi K_1 t}} \exp\left(-\frac{\Delta z^2}{4K_1 t}\right)$ [149, 150]. Using $p(\Delta z, t)$, the well-known variance $\left\langle (\Delta z(t))^2 \right\rangle = 2K_1 t$, which is referred to as the mean squared displacement or *ensemble average*. Later it was found that the time averaged MSD or *timelag average*,

$$\overline{\delta^2(\Delta)} = \frac{1}{t-\Delta} \int_0^{t-\Delta} \left[z(t'+\Delta) - z(t') \right]^2 dt', \tag{2.2}$$

also leads to $\overline{\delta^2(\Delta)} = 2K_1\Delta$ for Brownian motion, with Δ the length of a time lapse. Random walks for which $\overline{\delta^2(\Delta)} = \left\langle (\Delta z(t))^2 \right\rangle$ holds in the limit of long measurement times, as is true for Brownian motion, are called *ergodic*.

Sometime after Einstein's contribution continuous time random walks (CTRW) were introduced, as a more general form of random walks. The trajectory of CTRWs is the result of the sampling of the wait time and jump length from a joint probability density function $\psi(\Delta z, t)$. If both variances of wait time and jump length are finite, then we find Brownian motion. Hence Brownian motion can be regarded as a special case of the continuous time random walk (CTRW). If at least one of Einstein's assumptions is broken then anomalous diffusion occurs, characterized by $\left\langle (\Delta z(t))^2 \right\rangle \approx K_\alpha t^\alpha$ or $\overline{\delta^2(\Delta)} \approx K_\alpha \Delta^\alpha$ with $0 < \alpha < 1$ *sub-diffusion*, $\alpha > 1$ *super-diffusion* and K_α the corresponding diffusion coefficient. A case generally attributed to super-diffusion are Levy flights, which sample from heavier tailed spatial displacement distributions causing a divergence of the spatial variance [149, 150]. Examples of Brownian, sub- and super-diffusion are shown in fig. 2.7. An example for super-diffusion is passenger travel with plane flights representing infrequent very long displacements causing non-Gaussian tails in $p(\Delta z, t)$. Sub-diffusion, on the other hand, could be observed when particles are increasingly trapped, e.g. due to a decreasing variance of the underlying distribution.



(a) (b)

Figure 2.7: Characterization of 1D diffusion into ‘Brownian’ (solid blue line), ‘sub-diffusion’ (dashed orange line) and ‘super-diffusion’ (green line with stars) via a) ensemble average and b) timelag average. All trajectories are sampled for given uniformly distributed time steps. The Brownian and super diffusion examples were created sampling from the $\mathcal{N}(\mu = 0, \sigma = 1)$ and $\mathcal{N}(\mu = 0.15, \sigma = 1)$ normal distributions respectively. Sub-diffusive trajectories were generated sampling from $\mathcal{N}(\mu = 0, \sigma = f(t))$ with the time dependent function $f(t) = (1+t)^{-0.15}$.

As pleasing as these simple models may be it was noticed that they have an unphysical aspect, namely they imply infinite velocities, since for any passed time after the particle is released the probability of the particle being anywhere is non-zero, conflicting with the theory of relativity. A remedy for this is to include velocities. A modification like this for Levy flights involves sampling motion velocity and duration instead of displacement and wait time. This leads to Levy walks, introducing characteristic cone like trajectories, as shown for example in fig. 1 of the review of Zaburdaev *et al.* [270]. After each completed movement Levy walks choose a new direction at random.

We will use MD to track GBs and model them as random walkers in 1D as described in Chapter 6. To this end we have implemented Brownian motion, Brownian-like motions and variants of Levy walks. Each implemented random walk has specific parameters controlling the behaviour. These parameters can then be optimized to match observations from the MD. Brownian-like random walks are implemented in two different ways which can also be combined. One is the ‘discrete’ Brownian-like approach which is composed of 3 or 5 normal distributions symmetrically centred at zero with sampling of the normal distributions becoming more unlikely the further they are away from zero. The approach is referred to as ‘discrete’ because in the limit of zero variance of each of the normal distributions one samples from 3-5 allowed displacements. The other version of Brownian motion implemented it Brownian-like ‘resting’, which only adds a Markov process with the states ‘run’ and ‘stay’ to the sampling process. The Levy walk was modified similarly by adding a Markov process leading to the ‘resting’ Levy walk. Because some GB trajectories in Chapter 6 look like noisy straight lines we also implemented a ‘noisy’ Levy walk, where the straight sections of the Levy walk are perturbed with some normally distributed noise.

2.4 Grain Boundaries

Grain boundaries (GBs) have a significant influence on the thermomechanical properties of a material leading to ‘GB design’ being a field in its own right [31, 33, 154, 246, 274], as already mentioned in the context of polycrystalline superalloys. A particular grain boundary phenomenon that occurs in superalloys are ‘serrated edge’ GBs or serrated GBs, as discussed further in section 2.4.3. In an effort to understand the initiation of serrated GBs we have carried out MD simulations on several coincidence site lattice (CSL) GBs as described in Chapter 6. The CSL is be defined next and this is followed by a review of GB migration mechanisms including step motion, dislocation motion, correlated GB displacement, GB random walks and diffusion induced grain boundary motion (DIGM). Finally we will review serrated edge GBs and how their formation is related.

2.4.1 The Coincidence Site Lattice (CSL)

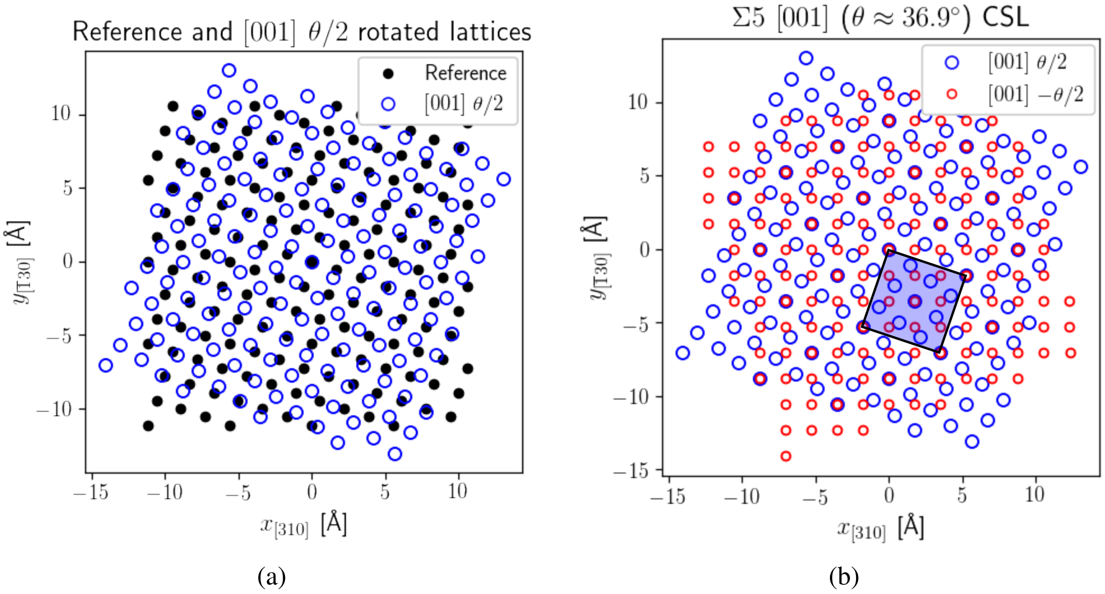
GBs are planar defects and are considered among the most complex imperfections in materials. They can be described by the relative misorientation of two grains in space and their common intersecting surface. The misorientation can be described by a rotation axis and rotation angle θ. Approximating the GB surface, at least locally, as a plane we can define CSL GBs for two grains which are of the same phase. The ‘coincidence sites’ refer to perfectly overlapping lattice sites of two lattices and can be visualized with the following example of a $\Sigma 5$ twist GB where the rotation is 36.9° around the $[001]$ axis on the (001) plane. For this construction it is useful to imagine, initially, the unphysical perfect overlap of two identical lattices. Choosing one rotation axis we rotate one lattice clockwise by $\theta/2$ relative to the reference lattice as in fig. 2.8 a) and then the other counter clockwise by $\theta/2$. Values of

$\theta \leq 15^\circ$ lead to GBs which are referred to as *low angle* GBs, whereas those exceeding 15° are referred to as *high angle* GBs. It turns out that there are discrete θ values specific to a rotation axes for a given lattice which lead to lattice sites of both lattices *coinciding*. These coinciding sites form a regular pattern referred to as the CSL, see fig. 2.8 b). The unit cell of the CSL is larger than the original lattice's unit cell. The ratio of their volumes is designated as

$$\Sigma = \frac{V_{\mathrm{CSL}}}{V_{\text{original lattice}}}, \tag{2.3}$$

where V_{CSL} is the volume of the CSL cell and $V_{\text{original lattice}}$ is the volume of the lattice's unit cell [86].

Alternatively, Σ can also be interpreted as every Σth original lattice cell site being a coincidence site. Σ is therefore the inverse density of coincidence sites. Continuing with our construction we need to remove the unphysical overlap of the clockwise and counter clockwise rotated lattices. This is done by choosing a plane in $\mathbb{R}^3$, designating a volume for either lattice and removing the lattice sites of the respective other lattice, resulting in two grains. The orientation of this plane relative to the rotation axis defines the GB. For this plane one distinguishes between planes which are perpendicular to the rotation axis leading to so-called *twist* CSL GBs and those planes which lie parallel to the rotation axis, *tilt* CSL GBs. These will be referred to in this work as twist and tilt GBs respectively. When choosing the plane for the tilt GB one has a variety of options, with one option producing an exact mirror symmetric pattern and hence a *symmetric* tilt GB. When referring to tilt GBs in the following we will consider only *symmetric* GBs, and not *inclined* or *asymmetric* GBs. Rotations other than these discrete rotation axis and angle combinations lead to *random* GBs. While *random* or general GBs are also important, CSL GBs are of interest here because of their periodicity, making them directly useful for periodic supercell simulations at the atomistic level.



(a)

(b)

Figure 2.8: Construction of the Σ5 [001] twist GB. Part a) shows rotation of the upper half (blue circles) of the bicrystal $\theta/2$ relative to the reference lattice (black dots). Part b) shows the superposition of the upper half and lower half (red circles) of the bicrystal after $\pm\theta/2$ rotations. The CSL sites are visible as overlapping red and blue circles. The CSL cell is indicated in b) with a blue square with a black border.

2.4.2 Grain Boundary Migration

GB migration is a complex phenomenon to which a significant part of Sutton and Balluffi's book [227] is devoted. With increasing computational power over the last decade, a significant number of atomistic simulations have been performed of GB migration under various conditions [154]. In this section we limit our review to the necessary background for understanding the results of the bicrystal simulations described in Chapter 6. We will first discuss driving forces for GB motion and the GB mobility. This is followed in sections 2.4.2.1, 2.4.2.2 and 2.4.2.3 by how GBs move and their migration. Each of these sections discusses specific aspects of GB migration: dislocation and step motion (section 2.4.2.1), correlated displacements (section 2.4.2.2) and random walks (section 2.4.2.3). Finally, we will review in section 2.4.2.4 diffusion induced grain boundary migration (DIGM), which may explain the initiation of serrated GB formation.

GB migration between adjacent grains A and B takes place when there is a net flux of atoms from grain A to grain B given a driving force. The driving force for GB migration is commonly one, or more, of the following: GB curvature / the capillarity effect [114], elastic strain (shear, normal or phase/solute misfit) [85, 203, 271, 272], chemical gradient [181], phase transitions [156] and magnetic / artificial [113, 245]. Artificial driving forces refer here to those induced in atomistic simulations by altering the potential of atoms adjacent to the GB. For the further discussion, the above driving forces are grouped into ‘external’ and ‘internal’ driving forces. ‘External’ driving forces contain externally applied elastic strain and magnetic / artificial driving forces, whereas ‘internal’ driving forces contain GB curvature, chemical gradient and phase transitions.

One quantity which has been studied is the ‘GB mobility’, M. Commonly one assumes the following linear relationship between for the GB velocity v and mobility:

$$v = Mp, \tag{2.4}$$

where p is the driving force [241]. p can be decomposed as $p = \kappa + \eta$, where κ is the pressure due to the curvature of the GB and η is the bulk force. The pressure is then approximated as $\kappa = \kappa_1 \left(\gamma + \partial^2_{\phi_1}\gamma\right) + \kappa_2 \left(\gamma + \partial^2_{\phi_2}\gamma\right)$ [240], where γ is the interfacial energy and ϕ_i are the angles defining deviations from the GB normal. When considering GB motion which is mostly one dimensional $\kappa \approx \kappa_1 \left(\gamma + \partial^2_{\phi_1}\gamma\right)$ where $\left(\gamma + \partial^2_{\phi_1}\gamma\right)$ is called the, ‘GB stiffness’, Γ.

While the explicit introduction of GB curvature can be used to study the GB migration for a given grain misorientation, as first done by Jhan and Bristowe [114], this does not allow a straight forward calculation of M for flat GBs. For this, two routes have emerged: i) the application of an external driving force, such as elastic strain or artificial force and ii) the analysis of GB fluctuations without an external driving force [50]. While method i) allows the direct evaluation of eq. 2.4, method ii) requires an analysis of the displacement of the GB (assuming a flat GB) [50, 51, 241]. Note that eq. 2.4 is a simplification and only holds in the limit of $p \to 0$. Non-linear behaviour can be observed particularly at low temperatures and high pressures [51, 88, 182, 271–273, 275]. The actual relationship between v and p depends on the active GB migration mechanisms, which will be discussed below. The computation of Γ can also be performed via the fluctuation method [240], requiring fluctuations of the GB along the GB plane.

The trajectories of the GBs themselves have been characterized as ‘stick-slip’, ‘jerky’ and ‘stop-and-go’ with GBs alternating between fluctuation around a resting position and directed movement. These kinds of trajectories have been observed in atomistic simulations [51, 88, 162, 244, 273] and experiments [5].

2.4.2.1 Migration via Dislocation and Step Motion

GBs can be constructed using arrays of dislocations and GBs may exhibit features called ‘steps’. Thus, the behaviour of both may give insights into the migration of GBs. According to Balluffi and Olson [8] dislocations may be assigned into a hierarchy of: i) primary interfacial dislocations (crystal lattice dislocations), ii) secondary interfacial dislocations (dislocations related to the CSL), iii) coherency dislocations and iv) translational interfacial dislocations. An example of a $\Sigma5$ [001] tilt GB constructed purely from primary dislocations, in the form of edge dislocations, is shown in fig. 2.9. Migration of the GB can then be described as a function of the dislocation configuration and mobility, thus involving dislocation glide and / or climb [213]. In the example shown in fig. 2.9 sideways movement would need the dislocations to glide and climb, thus requiring some activation temperature. Twist GBs, on the other hand are found to exhibit screw dislocations as in the case of the$\Sigma85$ twist GB shown in fig. 2.10. In fact Schoenfelder et al. [204] found that $\Sigma41$, $\Sigma85$ and $\Sigma181$ twist GBs all move via the motion of their respective screw dislocation networks. Both tilt and twist GBs with identifiable crystal lattice dislocations (i.e. usually low angle) each require the movement of the entire dislocation network for migration to proceed.

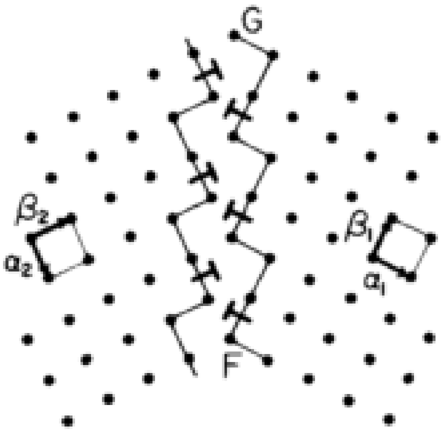


Figure 2.9: Dislocation model of a $\Sigma5$ [001] tilt GB in a cubic lattice taken from Balluffi and Olson [8].

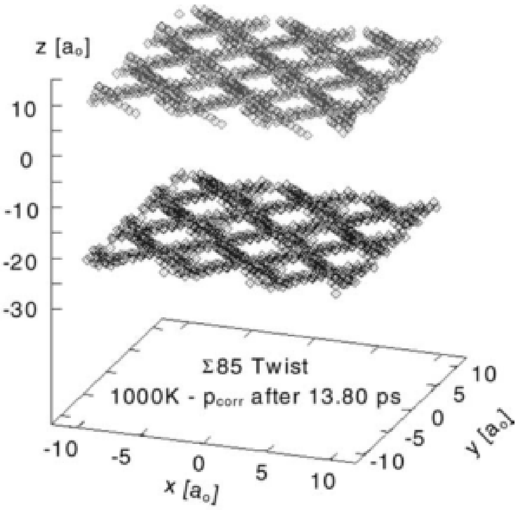


Figure 2.10: Screw dislocation core network of Σ85 [001] twist taken from Schoenfelder *et al.* [204].

Steps in the GB may be visualized as in fig. 2.11, where the red blocks represent CSL unit cells in the GB which were added to the lower grain (grey slab). The concept of GB steps goes back to Mott [158], who envisioned GB migration as the 'melting' of atoms off of one grain and their solidification in the form of an 'island' on the other grain. Observing low dislocation density areas of deformed Al-0.39 wt.% Cu, Gleiter [81] found GBs exhibiting a zigzag form and posited a GB migration mechanism in analogy to vapor deposition, with atoms being emitted from kinks in steps of the shrinking grain and absorbed in kinks of steps of the growing grain. Gleiter posited further that the steps would maintain a rigid grid throughout the migration.

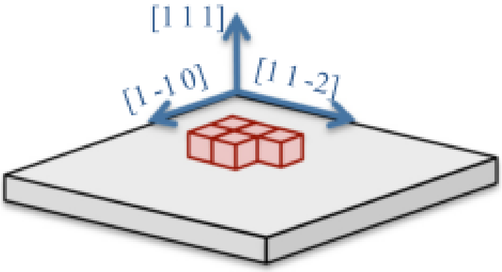


Figure 2.11: Cutaway view of a GB with an island of pure perfect steps (red) in a (111) plane (grey) taken from Hadian *et al.* [88].

The concept of GB migration via the growth of steps has recently received support from atomistic simulations of near symmetric $\Sigma7$ tilt GBs carried out by Race *et al.* [182] and Hadian *et al.* [88]. They found evidence that the rate limiting step in GB migration leading to ‘stick-slip’ behaviour is a nucleation event, for example the formation of islands of steps. While islands of steps were required for symmetric tilt GBs, asymmetric and mixed GBs required the formation of double kinks or merely the propagation of kinks. Since these nucleation processes are stochastic events in the GB plane, Race *et al.* and Hadian *et al.* concluded that the observed mobility in the simulations is also a function of the GB area itself and not completely an intrinsic property. The nucleation processes were furthermore found to be temperature dependent and, possibly not occurring at all at low temperatures for symmetric and asymmetric GBs, leading to practically immobile GBs.

A further process related to steps is the formation of facets, as observed by Hsieh *et al.* [109] for asymmetric tilt GBs. They studied asymmetric Al and Au $\Sigma3$ and $\Sigma11$ $[011]$ tilt boundaries and observed a transition between smooth GBs at high temperatures and zigzag shaped GBs at low temperatures. They referred to individual units of the zigzag shaped GBs as ‘facets’ and the entire process as (de-)facetting. The facetting was rationalized in terms of the minimization of the GB energy upon cooling leading to a decomposition of the asymmetric tilt GBs into lower energy, low index, GBs for both grains. This facetting reaction should be taken into account when analyzing room temperature micrographs of GBs after high temperature annealing.

2.4.2.2 Migration via Correlated Atomic Displacements

The motion of atoms themselves during GB migration may either be associated with ‘correlated’ or ‘uncorrelated’ displacements. Correlated displacements, also referred to as ‘shuffling’, means that the atoms move as a group in a deterministic way, whereas uncorrelated displacement means that they move in a stochastic manner. Due to the difficulty of investigating atom motion in GBs experimentally, most insights have been gained using atomistic simulations. One of the first indications of correlated displacements was found by Majid and Bristowe [114, 146] simulating a $\Sigma5$ $[001]$ twist GB. The displacements involved groups of 4 atoms rotation around $[001]$. Creating an atomistic model in which the GB took on the shape of a pyramid, thus utilizing the capillarity effect, Jhan and Bristowe [114] showed that the group rotation resulted in GB normal migration. This correlated shuffling was also confirmed by Schoenfelder *et al.* [204] who investigated a range of twist GBs, finding similar mechanisms for other geometries such as $\Sigma29$ and $\Sigma17$.

What could be understood as correlated displacements in tilt GBs was observed for the $\Sigma5$

[001] tilt system by Zhang *et al.* [271–273] and for the $\Sigma 7$ [111] tilt system by Ulomek and Mohles [244]. Using elastic bi-axial strain as the driving force, Zhang *et al.* identified three types of displacements: i) CSL site to CSL site jump without contribution to GB migration, ii) a lattice jump between the adjacent grains and iii) in plane jumps on sites of one grain only. In combination, type ii) and iii) were found to lead to 'string-like' atom motion during GB migration. Using an artificial driving force Ulomek and Mohles distinguish between the 'initiation' step, which includes the formation of a vacancy in the GB leading to a 'thickening' of the GB, creating a vacancy, and a subsequent shuffling step 'thinning' the GB. Interestingly Ulomek and Mohles also observed a plateauing of the GB velocity at high temperature, similar to Race *et al.* [182] and Hadian *et al.* [88], which thus could be related to their observed nucleation mechanisms.

2.4.2.3 Grain Boundary Random Walk

Trautt *et al.* [241] suggested that moving GBs could be approximated as 1D random walkers with displacements Δz along the GB normal. They then used the distribution $p(\Delta z, t)$ resulting from multiple MD simulations of Al to characterize the GB migration behaviour, e.g. the mobility M via $\left\langle (\Delta z(t))^2 \right\rangle$. This method is also referred to as the interface random walk method. To obtain $\left\langle (\Delta z(t))^2 \right\rangle$ they averaged the position of the GB over the other two perpendicular directions (x and y). Hence the GB itself does not necessarily have to be flat for this method to work. This is also the method used in the present work described in Chapter 6. Using this approach to analyse GB motion they found that impurities / solutes add significant drag without being required to use artificial driving forces. Note that this method requires that the GB mean position fluctuates, which requires high temperatures above the roughening temperature [50]. The roughening temperature itself depends on the bicrystallography [162].

Deng *et al.* [50, 51] suggested a modification to the interface random walk method of Trautt *et al.* [241], which they dubbed the adaptive interface random walk or 'AIRwalk' method. Comparison between this method and the work of Metzler *et al.* [149, 150] shows a high degree of similarity between AIRwalk and using the time lag average $\overline{\delta^2(\Delta)}$. With the AIRwalk method and an artificial driving force, Deng *et al.* [50, 51] analysed a set of $\Sigma 3$, $\Sigma 5$ and $\Sigma 7$ tilt GBs in either pure Al or Ni. Repeatedly simulating symmetric and inclined $\Sigma 5$ tilt GBs under different artificial driving forces and temperatures, they identified 'diffusional' and 'ballistic' motion regimes. While GBs in the 'diffusional' regime remain mostly in a 'locked' state, i.e. fluctuate around their initial position, GBs in the 'ballistic' regime exhibit predominantly directed movement, which Deng *et al.* refer to as the 'running' state. The transition between 'locked' and 'running' state was observed to become more frequent with increased temperature, which is reasonable in light of a possible bottleneck to GB migration such as the nucleation of kinks [88, 182].

Above, in section 2.3, we made the distinction between Brownian motion, super- and sub-diffusion. In the context of GB migration in pure Ni, Brownian motion would be expected in the limit of no active external and internal driving force at temperatures above the roughening temperature. In the case of an existing internal driving force we would find super-diffusion. If no driving force prevails and the GBs are increasingly trapped over time, i.e. due to solutes, one would expect sub-diffusive behaviour.

2.4.2.4 Diffusion Induced Grain Boundary Migration and Discontinuous Precipitation

Diffusion induced grain boundary migration (DIGM) is the particular way in which the diffusion of solutes, predominantly along the GB (not across from one grain to the next), induces GB migration, leaving a (de-)alloyed zone behind. The resulting GB topology may be stochastic or 'wavy' [121, 227], see fig. 2.12. A range of driving force models have been suggested for DIGM, overlapping with those mentioned above, among them: elastic coherency strain, chemical driving force on the bulk for (de-)alloying, chemical driving force on the GB itself (GB phase transformation), dislocation climb as a consequence of bulk diffusion and facet migration.

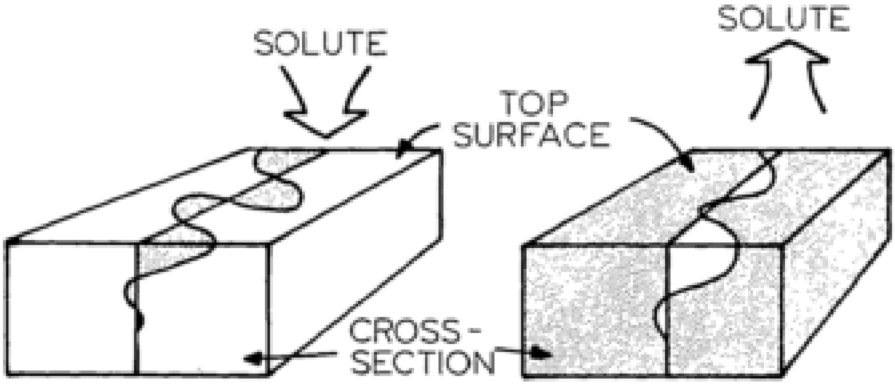


Figure 2.12: Schematic of diffusion induced GB migration (DIGM) causing (de-)alloying due to GB diffusion of solutes, taken from King [121].

The main opposing forces to DIGM include the energy required to increase GB area during migration and the deposition of defect clusters in the wake of the GB. An example system in which the elastic coherency strain was claimed to be the active driving force is Zn-Cd [139], where high solute concentration in the GB leaked into the adjacent crystals leading to thin strained slices of the lattice parallel to the GB plane. In regions where the strain was asymmetrical GB migration towards the more highly strained layer was observed. Liang *et al.* [139] noted that for symmetric GBs in particular the symmetry of the strain and concentration profile needed to be broken in order to allow for GB migration.

Investigating Cu/Ni/Cu diffusion couples above 873 K, Ma *et al.* [142, 143] found that the incubation period required for DIGM to occur depends on the temperature, decreasing in length with increasing temperature, a result also found by Liu *et al.* [140]. Tracking the GB, they found three stages of GB motion: i) the ‘initial’ stage (fast), ii) the ‘stationary’ stage (linear and slower) and iii) ‘mixed’ (DIGM and diffusion induced recrystallization). They also observed modes of GB migration with increasing temperature from unidirectional at low temperature, to bidirectional and high oscillatory at medium and high temperature. They rationalized their observations in terms of diffusion induced GB stresses which help to initiate GB migration, with GB roughening possibly playing a role. Additionally, Liu *et al.* [140] identified thin solute rich layers forming along the GB. Both Ma *et al.* and Liu *et al.* questioned whether elastic coherency strain alone would be a sufficient driving force. Complementary to their work, Rabkin *et al.* [180, 181] showed that even at high temperatures small discontinuities in solid solution fluctuations around the GB may be sufficient to initiate DIGM due to the contribution of the free energy gradient with increasing solute misfit.

Another process, potentially related to serrated GBs (see next section), is discontinuous precipitation. Discontinuous precipitation is particularly associated with GBs which are moving through a supersaturated grain leading to a $\alpha \rightarrow \alpha' + \beta$ reaction in their wake, where the resulting α is indicated as α' here because it has a different solute content not implying ordering. Discontinuous precipitation leads to the formation of lamellar structures which grow perpendicular to the GB. Examples of discontinuous precipitation can be found in the simultaneous formation of γ' and TCP phases [259].

Both DIGM and discontinuous precipitation are intricately connected, i.e. the α' channels leading to α'/α interfaces at the GB are driven by DIGM.

2.4.3 Serrated Edge Grain Boundaries

What is described as a 'serrated edge grain boundary' varies between quite wavy to faceted depending on the underlying alloy and heat treatment. An example of a wavy GB as observed in RR1000 is shown in fig. 2.13. In micrographs of serrated GBs, one usually finds precipitates on or near GBs that are coherent / semi-coherent with one of the two grains. The precipitates are found in plate-like, dendritic and fan-like shapes [46, 47, 156, 176]. GBs may exhibit serrations to different degrees depending on the heat treatment, with longer serration wavelengths and amplitudes for longer treatments at high temperatures, the same as observed by Ma *et al.* [142, 143] in the context of DIGM.

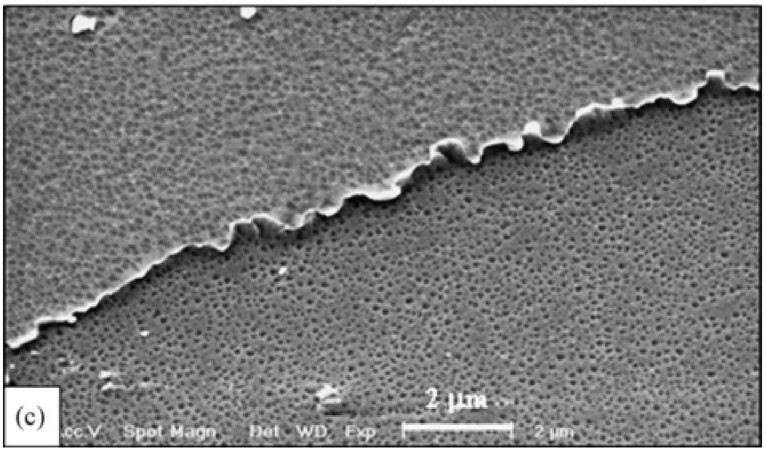


Figure 2.13: Scanning electron microscope micrograph of a serrated GB as observed in RR1000 by Mitchell *et al.* [156] for a cooling rate of 3.53 K s^{-1}. Grey dots indicate γ' precipiates and the wavy white line indicates the GB. Lower cooling rates were shown to lead to stronger GB serration. The serration wavelength is on the order of 1 μm.

Serrated GBs have been observed in a range of alloys. In alloys without γ' formation, a non-exhaustive list includes: alloy 718 [171, 268], Haynes [115, 269], Nimonic 263 [106], alloy 263 [107] and others [77, 105, 108, 230, 231]. Alloys with γ' include: Inconel [125, 130], Nimonic [125], Udimet 720 [155], Rene 104 [35], RR1000 [137, 156, 176], Rene 88DT [192], alloy 617 [135], astroloy [46, 136], modified waspalloy [263] and others [47, 99, 265, 266]. Properties which have been found to be affected by the presence of serrated edge GBs are creep life, crack growth, transgranular fracture, required stress intensity for creep and the onset of tertiary creep [130, 135, 137, 263]. The hypothesis is that the protruding GBs reduce grain boundary sliding and improve crack growth behaviour.

The common treatment leading to serrated edge GBs is slow cooling through a solvus temperature of either γ' or a carbide phase [46, 99, 115, 120, 125, 136, 156, 176, 265]. But there is also evidence for serration of GBs after slow cooling leading without precipitation [106, 131, 231].

A mechanism for serration proposed by Koul *et al.* [125] involving moving γ' particles caused by strain differentials and pushing GBs, is widely dismissed because it does not take shape evolution and low γ'/γ misfit into account [46, 99]. Another mechanism is the pushing of dendritic arms of γ' precipitates or discontinuous precipitation. This mechanism was put forward as an explanation for the increased frequency of γ' precipitates observed near serrated GBs and their sometimes dendritic and fan-like forms [46, 47, 99, 136, 156]. However, as there is evidence for serrated GB without precipitates it has also been proposed that GB serration may be due to solutes like B, Cr, Zr or C interacting with GBs [106, 131, 231]. Furthermore, serrations were also observed in the context of mechanical deformations and tempering [147].

Thus, GB serration is clearly a complex process. However, the initiation of the observed serrations, without mechanical deformations, could be interpreted in terms of DIGM using the following thought experiment. For the thought experiment we assume that the initial grain configuration is stable at the given temperature, no external deformation is present and thermal stresses are negligible. The latter should be reasonable particularly under slow cooling at which serrated GBs have been observed. Due to the cooling the solubility of the present initial will decrease leading to an increasing driving force for de-alloying, which could be achieved due to bulk and GB diffusion. This would be the classic case where DIGM could act, leading locally to GB migration where elastic strain and the chemical gradient are asymmetric. If we define the host phase to be a γ phase, which becomes supersaturated due to the cooling, we will have, in addition to the driving force for de-alloying, a driving force for the formation of the chemically ordered γ' phase, as well as an elastic strain due to the γ/γ' misfit. An asymmetry in this case could lead to an initiation of GB migration as well. Thus, in atomistic simulations one would expect to observe an elastic strain due to the phase transformation and a decrease in overall energy of the system. An asymmetry in these driving forces could then lead to GB migration in these simulations, if they are significant enough and depending on the GB area and bicrystallography.

The actual resulting GB topologies themselves would be a result of the local GB bicrystallography, which controls the available mechanisms, such as the nucleation of steps and kinks. The active mechanisms and resulting GB velocity should play a significant role in determining the formation frequency and distribution of precipitates in the wake of the GB. Accelerated ordering in the wake of GBs could lead to growth and coalescence of γ' particles resulting in ‘fan’-like shapes as observed in RR1000 [176]. The degree of ‘waviness’ as observed by Mitchell *et al.* [156] would then be a balance between the gain in interface energy, the reduction in potential energy due to precipitate growth, the resistance of the precipitates to dissolution by backtracking GBs and the present strain fields.

In order to shed some light on the initiation of serrations and the GB-γ' precipitate interaction we will describe in Chapter 6 bicrystal simulations of supersaturated γ at temperatures representing the conditions in experiments. Assuming that there is no severe systematic error inherent to the atomistic simulations then one should not really see any GB movement without γ' precipitates, assuming precipitate growth is the main cause of serration in supersaturated conditions. However, if ordering itself plays a relevant role we would expect to see GB migration with and without pre-existing precipitates. This would be expected in association with accelerated ordering in the vicinity of some GBs / in their wake accompanied by an overall decrease in volume and potential energy.

Chapter 3

Atomistic Simulation of Materials

Investigation of materials at the atomistic scale using computers and atomistic simulations has become one of the standard approaches to gain insights into the structure - property relationship of materials. The usefulness of atomistic simulations is founded in the complete control they allow over atom configurations, providing references for the interpretation of observations made in experiments. Experiments themselves oftentimes provide data of some ensemble properties but lack sufficient atomic resolution due to noise or other causes.

One branch of atomistic simulation methods we will discuss in section 3.1 is based on quantum mechanics and referred to as *density functional theory* (DFT). This *ab initio* theory leads to astonishingly accurate numerical methods and is thus widely applied, with 150,099 publications in the Web of Science's core collection at the date of writing this thesis. A reason for this popularity of DFT is its ability to predict ground state structures, lending itself to exploration and design of materials on the atomistic scale [89].

An alternative approach to simulating materials on the atomistic scale is to neglect the quantum mechanics and to assume classical mechanical behaviour. In this context we will discuss molecular dynamics (MD) and Markov chain Monte Carlo (MCMC or more commonly MC) in section 3.2. The combination of experiments, classical and quantum mechanical based simulations and machine learning has led to the rise of an interesting interdisciplinary field often referred to as 'materials informatics' [256]. Bridging length and timescales inherent to different simulation techniques both atomistic and continuum has led to the development of multiscale methods [58]. Classical mechanics simulations require an approximation for the interatomic interactions. These are captured in so-called empirical potentials, which will be discussed in section 3.3. A common problem encountered in the field of atomistic simulations is optimization, be it the optimization of atomic configurations, electronic structures, empirical potentials or other. Hence we will conclude this Chapter with section 3.4 giving a brief overview over available optimization algorithms.

3.1 Density Functional Theory

Describing materials at the atomistic scale using quantum mechanics means solving the time dependent Schrödinger equation

$$\hat{H}\Psi = i\frac{\partial}{\partial t}\Psi, \tag{3.1}$$

with $\hat{H}$ the Hamiltonian operator, containing the kinetic and Coulomb interactions of nuclei and electrons, and Ψ the many-body wave function. Ψ itself is a function of time, all electron and nuclei positions $\Psi(\mathbf{r}_0, \mathbf{r}_1, ...; \mathbf{R}_0, \mathbf{R}_1, ...; t)$, where $\mathbf{r}_i$ denotes the position of the ith of N electrons and $\mathbf{R}_i$ the ith of M nuclei. The challenge of equation 3.1 is to find $\Psi \in \mathbb{C}^{3\times(N+M)}$, which is a high dimensional function, and to remove the time dependency. In order to solve this a range of approximations are necessary for all but the smallest problems, which leads to DFT [79].

Considering that nuclei are much more massive than the surrounding electrons, a first approximation to make is to separate Ψ as

$$\Psi(\mathbf{r}_0, \mathbf{r}_1, ...; \mathbf{R}_0, \mathbf{R}_1, ...; t) = \psi(\mathbf{r}_0, \mathbf{r}_1, ...) \cdot \chi(\mathbf{R}_0, \mathbf{R}_1, ..; t) \tag{3.2}$$

since nuclei are practically immobile relative to the electrons and we are thus much more certain where the nuclei are. This is known as the adiabatic approximation. ψ are the electron wavefunction and χ are the nuclear wavefunction. The nuclei are then considered to act as an ‘external potential’ on the electrons. The problem reduces to solving the time-independent Schrödinger equation.

Now we are left to identify $\psi(\mathbf{r}_0, \mathbf{r}_1, ...)$, which is still a complicated function of multiple electrons. It would be much easier to treat this as the interaction of single electrons $\phi_i(\mathbf{r}_i)$, which is our next approximation. Assuming independent electrons we can construct $\psi(\mathbf{r}_0, \mathbf{r}_1, ...) = \prod_i \phi_i(\mathbf{r}_i)$. But this exact form violates the Pauli exclusion principle for Fermions, which requires a sign change upon position swap $\psi(\mathbf{r}_0, \mathbf{r}_1, ...) = -\psi(\mathbf{r}_1, \mathbf{r}_0, ...)$.

This is achieved by using the *Slater* determinant

$$\psi(\mathbf{r}_0, \mathbf{r}_1, ...) = \frac{1}{\sqrt{N}!} \begin{vmatrix} \phi_0(\mathbf{r}_0) & \phi_1(\mathbf{r}_0) & \cdots & \phi_{N-1}(\mathbf{r}_0) \\ \phi_0(\mathbf{r}_1) & \phi_1(\mathbf{r}_1) & & \\ \vdots & & \ddots & \vdots \\ \phi_0(\mathbf{r}_{N-1}) & \phi_1(\mathbf{r}_{N-1}) & \cdots & \phi_{N-1}(\mathbf{r}_{N-1}) \end{vmatrix} . \tag{3.3}$$

However, in real materials electrons are not independent. Thus, we may want to capture the effect all electrons have as a whole on individual electrons in the form of a *mean field approximation*. This leads to the Hartree potential, V_H.

Using a variational formulation, the total electronic energy

$$E = \langle \psi | \hat{H} | \psi \rangle \tag{3.4}$$

$$= \int \psi^* (\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_{N-1}) \hat{H} \psi (\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_{N-1}) d\mathbf{r}_0 d\mathbf{r}_1 d\mathbf{r}_{N-1} \tag{3.5}$$

(with $\langle\rangle$ the notation of the scalar product in quantum mechanics) Hartree and Fock [67, 96, 97] derived the Hartree-Fock equations with $\frac{\delta E}{\delta \phi_i^*} = 0$ and enforcing $\langle \phi_j | \phi_i \rangle = 0$. The Hartree-Fock equations include the electron exchange contribution and allow for an iterative numerical procedure which continues until self-consistency of the Hartree-Fock equations is reached.

Recognizing that the *ground state* electron density $\eta_{GS}(\mathbf{r})$ leads to a unique external potential which itself leads to a unique ψ and therefore E, Hohenberg and Kohn [102] postulated the ground state energy as a functional, F, of η_{GS},

$$E = F[\eta_{GS}] = T[\eta_{GS}] + U[\eta_{GS}] + \int \eta_{GS}(\mathbf{r}) V_n(\mathbf{r}) d\mathbf{r}, \tag{3.6}$$

where $T[\eta_{GS}]$ and $U[\eta_{GS}]$ are the functionals of the kinetic and Coulomb energy of the electrons, respectively. $V_n(\mathbf{r})$ is the external potential induced by the nuclei. Thus instead of having to find $\psi \in \mathbb{C}^{3N}$ we now need to find $\eta_{GS} \in \mathbb{R}^3$. This insight, which shifts the focus on to the electron density, resulted in naming this theory Density Functional Theory.

Building on this Kohn and Sham [123] decomposed the effective Hamiltonian for $F[\eta_{GS}]$ into the external potential, the kinetic energy, the Hartree potential and the exchange-correlation potential with $\delta F[\eta] / \delta \eta |_{\eta = \eta_{GS}} = 0$. Using, again, the independent electron approximation one finds the self-consistent numerical procedure to solve the so-called Kohn-Sham equations [123]

$$\left(V_n(\mathbf{r}) + \left(-\frac{\nabla_{\mathbf{r}}^2}{2}\right) + V_H[\eta] + V_{xc}[\eta]\right)\phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \tag{3.7}$$

obtained enforcing $\langle \phi_i \mid \phi_j \rangle = \delta_{ij}$. Hence there is some similarity to the Hartree-Fock approach but with η being the key quantity rather than the wave function. DFT based on the Kohn-Sham equations is also referred to as KS DFT. The exchange-correlation contribution which occurs in KS DFT is of the order of the kinetic energy and is therefore too large to be neglected. Common approximations for the exchange-correlation energy are the local density approximation (LDA)

$$E_{xc}^{LDA} = \int \varepsilon_{xc-LDA}(\eta_{GS})\,\eta_{GS}(\mathbf{r})\,d\mathbf{r} \tag{3.8}$$

and the generalized gradient approximation (GGA)

$$E_{xc}^{GGA} = \int \left(\varepsilon_{xc-LDA}(\eta_{GS}) + \varepsilon_{xc-GGA}(\eta_{GS}, \nabla_{\mathbf{r}}\eta_{GS})\right)\eta_{GS}(\mathbf{r})\,d\mathbf{r}, \tag{3.9}$$

with ε_{xc-LDA} and ε_{xc-GGA} representing energy functions which depend on the local electron density and its gradient. The LDA is derived for the case of a homogeneous electron gas yielding a functional [62, 102, 234]

$$\propto \int (\eta(\mathbf{r}))^{5/3}\,d\mathbf{r} \tag{3.10}$$

Instead of using both the wavefunction, in form of the Slater determinant, and the electron density, as in KS DFT, the Schrödinger equation can be cast entirely in explicit dependence of the electron density. This branch of DFT is referred to as orbital free DFT (OF DFT). Replacing Ψ with η_{GS} leaves most of the treatment unchanged but alters the description of the interaction of nuclei and electrons, requiring new pseudopotentials, and requiring another approach to the operator of the kinetic energy of the electrons. Four kinetic energy functionals, T, to mention in particular are the Thomas-Fermi (TF) [62, 234], von Weizsäcker (vW) [258], Wang-Teter (WT) [253] and, Wang-Govind-Carter (WGC) [255] functionals since they are implemented in Profess [101], an OF DFT code used later on in this work. The Thomas-Fermi, T_{TF}, functional is based on the homogeneous electron density approximation, which is also used for the LDA of the exchange-correlation functional. T_{TF} is given as

$$T_{TF}[\eta] \equiv C_F \int (\eta(\mathbf{r}))^{5/3} d\mathbf{r}, \tag{3.11}$$

where $C_F \equiv (3\pi)^{2/3} 3/10$. Since the T_{TF} functional only includes the contribution of the eigenfunction to the total kinetic energy, von Weizsäcker suggested to also add the contribution of the spatial derivatives of those eigenfunctions, introducing

$$T_{vW}[\eta] \equiv \int \sqrt{\eta(\mathbf{r})} \left(-\frac{1}{2}\nabla_{\mathbf{r}}^2\right) \sqrt{\eta(\mathbf{r})} d\mathbf{r}. \tag{3.12}$$

The combined TF and vW functional is then given as

$$T_{TF\lambda W} = T_{TF}[\eta] + \lambda T_{vW}[\eta], \tag{3.13}$$

where von Weizsäcker's original proposal equals T_{TF1W}. Considering that the vW and TF approximations could be further improved by incorporating information of η at some point $\mathbf{r}'$ in the vicinity of $\mathbf{r}$ using some function $f(\mathbf{r},\mathbf{r}')$ Wang and Teter proposed to put $T[\eta]$ as

$$T[\eta] = \int \eta(\mathbf{r}) \int f(\mathbf{r},\mathbf{r}')\, \eta(\mathbf{r}')\, d\mathbf{r}' d\mathbf{r}. \tag{3.14}$$

Enforcing Lindhard linear response to determine $f(\mathbf{r},\mathbf{r}')$, they recovered $T[\eta]$ as a sum of $T_{TF}[\eta]$, $T_{vW}[\eta]$ and the new term $T_{WT}[\eta]$

$$T_{WT}[\eta] \equiv \frac{48}{125}\left(3\pi^2\right)^{2/3} \int \eta^{5/6}(\mathbf{r})\, w_1(\mathbf{r}-\mathbf{r}')\, \eta^{5/6}(\mathbf{r}')\, d\mathbf{r}d\mathbf{r}', \tag{3.15}$$

where w_1 is a function of $\mathbf{r}-\mathbf{r}'$. Generalizing WT, WGC found that the 'kernel' w_1 can be made η dependent introducing

$$T_{WGC}[\eta] = C_{TF} \int \eta^{\alpha}(\mathbf{r})\, w_{\alpha,\beta}\left[\xi_\gamma(\mathbf{r},\mathbf{r}'), \mathbf{r}-\mathbf{r}'\right] \eta^{\beta}(\mathbf{r}')\, d\mathbf{r}d\mathbf{r}' \tag{3.16}$$

with $w_{\alpha,\beta}$ the new version of w_1, $\xi_\gamma(\mathbf{r},\mathbf{r}') = \left(\frac{k_F^\gamma(\mathbf{r})+k_F^\gamma(\mathbf{r}')}{2}\right)^{1/\gamma}$, $k_F(\mathbf{r}) = \left(3\pi^2\eta(\mathbf{r})\right)^{1/3}$ and α, β and γ parameters.

Equilibrium structures can then be found utilizing the convenient separation of Ψ into ψ and χ, allowing for an iterative approach by first computing $\eta(\mathbf{r})$ and then optimizing the nuclei positions $\{\mathbf{R}_i\}_{i\in[0,M-1]}$ in an 'effective' potential. The latter step uses the Born-Oppenheimer approximation that nuclei move much slower than the electrons and that electrons do not exchange energy with the nuclei, i.e. the adiabatic approximation. The forces on the nuclei are efficiently computable via the Feynman-Hellman theorem [65, 98] which describes them as electrostatic forces exerted by all the electrons and nuclei in the system, for a system in the ground state with known ground state electron density, $\eta_{GS}(\mathbf{r})$. Structure optimization can then be implemented either by minimizing the energy using conjugate gradients or discretizing Newton's equation of motion using the Verlet algorithm in a molecular dynamics approach. Aside from equilibrium structures, DFT can also be used to calculate vibrational properties, binding / cohesive energies, ionization potentials, electron affinities, band structures and defect formation energies. The capability of DFT algorithms to find ground state structures can also be exploited to discover unknown phases as with AIRSS [170] and USPEX [80] among others and high throughput methods such as AFLOW [45], AiiDA [172] and the materials project [111]. The electron densities themselves also contain plenty of information about the chemical bonding of atoms. This is utilized in Bader's atoms-in-molecules analysis and Laplacian based critical point analysis [165]. Contemporary implementations of DFT algorithms can be found in CASTEP [41] (plane wave basis set KS DFT), Profess [101] (orbital free DFT) and SIESTA [218] (local basis set KS DFT) among others.

3.2 Molecular Dynamics & Monte Carlo

MD is used to compute the natural time evolution of a system. This can either be done *ab initio* as mentioned above or in the classical sense, using an empirical potential which approximates the quantum mechanical interaction of atoms. MD is therefore the method of choice if one is interested in the study of long time dynamics and / or time averages of properties. An alternative approach is to use MC, which can also utilize empirical potentials, to calculate ensemble averages without recourse to time. Both methods are ergodic in their basic form, i.e. any state can be revisited but is revisited only aperiodically. Next, we will briefly discuss MC at first followed by MD and some modifications mostly based on Frenkel *et al.* [72].

MC is a technique to solve integrals such as $\langle A \rangle = \frac{1}{Z} \int A\left(\tilde{\mathbf{R}}\right) e^{-\beta U\left(\tilde{\mathbf{R}}\right)} d\tilde{\mathbf{R}}$ by sampling, where $A\left(\tilde{\mathbf{R}}\right)$ is some difficult function. In atomistic simulations we also encounter these integrals for ensemble properties, with $\tilde{\mathbf{R}}$ denoting the positions of all system particles, $\beta \equiv 1/k_B T$, U is the classical potential energy and $Z = \int A\left(\tilde{\mathbf{R}}\right) e^{-\beta U\left(\tilde{\boldsymbol{R}}\right)} d\tilde{\mathbf{R}}$ is the partition function. The Metropolis-Hastings algorithm [148] for sampling energy states is particularly significant for atomistic simulations. It uses the probability density function $p(x)$ of some position in phase space x and computes $\langle A \rangle$ by assuming the conditional probability density $p\left(x'|x\right)$, which contains the transition probability from x to x'. The difference between structures represented by x and x' could, for example, be the displacement of some atoms or swapped elements. Note that although the Metropolis-Hastings algorithm in the context of software implementations is referred to as MC, it is technically a Markov chain Monte Carlo (MCMC) algorithm because it makes use of Markov chains. The sufficient condition for Markov chains is that $p\left(x\right)p\left(x'|x\right) = p\left(x'\right)p\left(x|x'\right)$ is true, i.e. that *detailed balance* is upheld. If detailed balance is satisfied then $p\left(x\right)$ will converge to a stationary distribution. A more detailed introduction into probability theory can be found in Chapter 4.

Based on $p\left(x\right)$ and $p\left(x'|x\right)$ the Metropolis-Hastings algorithm contains an acceptance criterion $\mathrm{acc}\left(x \to x'\right)$ to accept x' if and only if $\min\left(1, \frac{p(x)p(x'|x)}{p(x')p(x|x')}\right) \geq 1$. These probabilities for atom configurations are then $p\left(x\right) = \frac{1}{Z}e^{-\beta U\left(\tilde{\mathbf{R}}\right)}$ and $p\left(x'|x\right) = \frac{1}{Z}e^{-\beta\left[U\left(\tilde{\mathbf{R}}'\right)-U\left(\tilde{\mathbf{R}}\right)\right]}$, resulting in $\frac{p(x)p(x'|x)}{p(x')p(x|x')} = e^{-\beta\left[U\left(\tilde{\mathbf{R}}'\right)-U\left(\tilde{\mathbf{R}}\right)\right]}$. Note that the acceptance criterion does not require knowledge of Z anymore. MC simulations therefore progress by assuming some initial x and then continuously propose alterations, which may be accepted or rejected, and by doing so sample the phase space using the potential energy surface inherent to $U\left(\tilde{\mathbf{R}}\right)$.

MD using empirical potentials can be considered a good approximation for larger atoms / molecules than H_2, He and D_2. The natural evolution is achieved by 1) initializing configurations by setting positions and velocities, 2) computing the forces on the nuclei, 3) integrating Newton's equation of motion followed by computing the new forces and so on. The part which tends to be the most computationally expensive is the calculation of the forces, requiring considerations of adjacent atoms. Which atoms are adjacent may not be known a priori and requires a query. The integration of Newton's equation can be done with a variety of algorithms, such as the popular Verlet algorithm $\mathbf{R}\left(t+\Delta t\right) \approx 2\mathbf{R}\left(t\right) - \mathbf{R}\left(t-\Delta t\right) + \frac{\mathbf{F}(t)}{m}\left(\Delta t\right)^2$, where Δt is the time increment, m the nuclei's mass, $\mathbf{R}\left(t\right)$ the time dependent position of a nucleus and $\mathbf{F}\left(t\right)$ the acting time dependent force. The general aim for integration algorithms is stability for large Δt to enable simulations of longer simulated time periods (not to be confused with the time it takes for the simulation to run). Other relevant aspects are energy conservation and time reversibility. Alternative integration algorithms are leap-frog, Beeman, velocity-corrected Verlet and predictor-corrector. The largest and longest MD simulations to date are of the order of $\sim 10^{12}$ atoms [57, 76], using the Lennard-Jones potential , and $\sim 10^{-6}s$ [71], using the CHARM22 potential, respectively. One of the common software choices nowadays for MD simulations is LAMMPS [173], enabling MD on high performance clusters.

Properties other than ensemble or time averages can be computed using MD and MC but require more effort. Examples are the free energy G and entropy S. An approach to compute free energy differences between two systems by biasing the MC sampling, thus leading to a larger volume of phase space being sampled, was introduced by Torrie *et al.* [238] and is referred to as Umbrella sampling. It has been applied, for example, by Auer et al. [4] to determine nucleation rates of hard sphere solids. An extension to Umbrella sampling was suggested by Kumar *et al.* [127] in form of the weighted histogram analysis method (WHAM). This method can use either MD or MC and by comparing multiple simulations and combining their discretized distributions (histograms) of state occupation by weighting, allow for calculation of entropy and free energy. While both methods have some sort of bias to keep the simulations sampling a certain volume in phase space another approach, called 'meta dynamics', actively pushes the simulated atom configuration out of equilibrium using so-called collective variables, which describe some important aspect of the atom configuration [128]. This pushing is made history dependent by modifying empirical potentials via 'filling up' the potential energy surface with for example Gaussians which depend on the collective variables. Meta dynamics has been applied, for example, by Quigley *et al.* [177, 178] who sampled water crystallization events. Another modification which increases sampling is 'hyperdynamics', which accelerates MD, particularly for low temperature simulations [250].

3.3 Empirical Potentials

Empirical potentials are a necessary ingredient for classical atomistic simulations and describe the interaction of atoms. They can be understood as a way of representing experimental and *ab initio* data concerning the interactions using some analytic form. As *ab initio* data has become the de facto standard data base for empirical potentials, one can also view empirical potentials as an attempt at transferring that accuracy in a computationally cheaper way to previously unseen atom configurations.

An assumption commonly made is that empirical potentials can be used to describe the total energy of a multi-atom configuration by superposition of per atom energy contributions. For these per atom contributions one generally has to consider the local neighbourhood $\{\mathbf{R}_i\}_{i\in\mathrm{neigh}(\mathbf{r})}$ of the nuclei at $\mathbf{r}$ using some criterion s which specifies the neighbourhood $\mathrm{neigh}(\mathbf{r})$. Alternatively one can also think in terms of 'bonds' describing the energy contributions with potentials V from one, to two up to m-body terms $E = \sum_i V_1(\mathbf{R}_i) + \sum_{i,j\neq i} V_2(\mathbf{R}_i, \mathbf{R}_j) + \ldots$ [232].

It follows that there are different types of potential forms which suit different types of atomic configurations / bonding characteristics best, such as the well-known Lennard-Jones potential $V = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{6} - \left(\frac{\sigma}{r}\right)^{12}\right]$ as well as more complex potentials such as the glue-type potentials. Glue-type potentials describe the interaction of atoms by incorporating an approximation of the many body interaction via a 'glue'. Examples of the glue-type potentials which will be discussed below are the Embedded Atom Method (EAM) [49], the Modified Embedded Atom Method (MEAM) [18] and the Angular Dependent Potential (ADP) [153]. All these types utilize a glue referred to as 'embedding density' to calculate an embedding energy as well as pair distances for the pair energy. The potential types will be discussed in the chronological order EAM, MEAM and ADP because they represent successive attempts in the refinement of glue-type potentials. The focus here is on these potentials since they are the common choice for metallic systems and also contain a description of the embedding density. Usually the embedding density is not taken as a physical property, but it will be shown in Chapter 4 how the valence electron density available from DFT can be regressed, removing one source of arbitrariness in the training of potentials. Both EAM and MEAM are implemented in LAMMPS as in Gullet *et al.* [87]. A general limitation of empirical potentials is the extrapolation to unseen atom configurations and material properties not included in the fitting process.

In the following section 3.3.1 we will discuss EAM and MEAM in more detail as well as other potential types and their relationships, if any, to EAM and MEAM in section 3.3.2. Machine learning has also become a factor for the development of empirical potentials and will be discussed in section 3.3.3.

3.3.1 The (Modified) Embedded Atom Method

Stott *et al.* [221] showed that one can develop an energy function for an atom embedded into a crystal with the function depending only on the local value of the crystals' electron density field, which they called the 'quasi atom' approach. It can therefore be argued that the electron density should be considered in order to account for many-body effects in empirical potentials. In fact, this idea is behind of the embedded atom method (EAM) which was first used by Daw *et al.* [48] to study H embrittlement in fcc Ni. The name directly relates to the idea of embedding an atom in a host electron density due to its surroundings.

In EAM the total energy E of an atomic configuration of N_a atoms is written as

$$E = \sum_{n=1}^{N_a} E_{\text{emb},\alpha_n}(\overline{\rho}_n) + \frac{1}{2}\sum_{n=1}^{N_a} E_{\text{pair},\alpha_n,\alpha_{n'}}\left(r_{n,n'}\right), \tag{3.17}$$

where E_{emb,α_n} is the embedding energy function for the nth atom of chemical element α which depends on $\overline{\rho}_n$, the embedding density for atom n. E_{pair} is the pair energy function for atom n and n' and their respective elements separated by the distance $r_{n,n'}$. The embedding density $\overline{\rho}_n$ itself is calculated using a 2-body electron density function $\eta_{\alpha_l}^{\text{2-body}}$ for neighbor l of atom n as

$$\overline{\rho}_n = \sum_{l=1}^{L} \eta_{\alpha_l}^{\text{2-body}}\left(r_{n,l}\right), \tag{3.18}$$

where L is the total number of neighbors considered.

Expanding on the EAM formalism Daw et al. [49] developed the embedding density function using Hartree-Fock wave functions obtained for s and d orbitals of single atoms by Clementi and Roetti [43] using $\rho(r) = N_s\rho_s(r) + (N - N_s)\rho_d(r)$, where N is the total number of outer electrons and N_s is the effective number of s electrons, fitting N_s.

Glue-type potential forces are ‘conservative’ forces obeying the three equivalent equations

$$\begin{aligned} \nabla \times \mathbf{F} &= \mathbf{0} \\ W = \oint_C \boldsymbol{F} \cdot d\boldsymbol{s} &= 0 \\ \mathbf{F} &= -\nabla_{\mathbf{r}} E, \end{aligned} \tag{3.19}$$

which define the way they are derived from the given empirical potential. Forces on the nth atom in a configuration are denoted as

$$\mathbf{F}_n = -\nabla_{\mathbf{R}_n} E. \tag{3.20}$$

There are different ways of choosing the functional forms with most papers suggesting their own representations, e.g. Mishin, Farkas, Duff [56, 60, 152] among others. A particularly interesting study to mention was done by Mitev *et al.* [157] who regressed the electron density $\eta_{GS}(\mathbf{r})$ directly to create embedding density functions and EAM potentials using the MERLIN package [166] for optimization. They proposed the following form of $\rho_{tot}(\mathbf{r})$

for the regression of $\eta_{GS}(\mathbf{r})$

$$\rho_{tot}(\mathbf{r}) = \left[\sum_{l=0,1} t^{(l)} \left(q^{(l)}(\mathbf{r})\right)^2\right]^{1/2}, \tag{3.21}$$

with $q^{(l)}$ representing s-type ($l = 0$) and p-type ($l = 1$) contributions

$$q^{(l)}(\mathbf{r}) = \begin{cases} l = 0, & \sum_i^{\mathrm{neigh}(\mathbf{r})} \rho^{(0)}(r_i) \\ l = 1, & \left[\sum_{\alpha \in \{x,y,z\}} \left(\sum_i^{\mathrm{neigh}(\mathbf{r})} \frac{\alpha_i}{r_i} \rho^{(l)}(r_i)\right)^2\right]^{1/2} \end{cases}, \tag{3.22}$$

where $\mathbf{r}_i = \mathbf{r} - \mathbf{R}_i \in \mathbb{R}^3$, $r_i = \|\mathbf{r}_i\|_2$ and $\mathbf{R}_i$ is the position of atom i. The $\rho^{(l)}(r)$ themselves are parameterized as

$$\rho^{(l)}(r) = \left(a_0^{(l)} + a_1^{(l)} r + a_2^{(l)} r^2\right)^2 \exp\left(-\frac{r}{c^{(l)}}\right), \tag{3.23}$$

with $a_i^{(l)}$ and $c^{(l)}$ some l-dependent constants. Using this analytic form they approximated $\eta_{GS}(\mathbf{r})$ for fcc Al with errors of $< 1\%$ for 'between' ions, and $< 2\%$ for 'near' ions.

Instead of using contributions from atoms near $\mathbf{r}$ as specified in eq. 3.21, they computed overlap integrals of the electron density as $\rho^{eff}(r_{ij}) = \frac{1}{2}\int \frac{Q_i(\mathbf{r})Q_j(\mathbf{r})}{Q_i(\mathbf{r})+Q_j(\mathbf{r})} d\mathbf{r}$, with Q_i and Q_j referring to respective functions as in eq. 3.21 for atom i and j respectively separated by r_{ij}. In order to not have to recompute the overlap integrals during simulations they parameterized ρ^{eff} as

$$\rho^{eff}(r) = \left(A_0 + A_1 r + A_2 r^2\right) \exp\left(-\frac{r}{C}\right), \tag{3.24}$$

where $A_n(a_0) = \kappa_n + \lambda_n \exp(-\mu_n a_0)$ and $C(a_0) = \kappa' + \lambda' a_0 + \mu' a_0^2$ are used for A_n and C in terms of the lattice constant a_0. The κ, λ and μ parameters are obtained fitting a range elastically distorted versions of the fcc unit cell.

They then calculated the embedding density $\bar{\rho}_i$ at atom i as $\bar{\rho}_i = \sum_{j \in \mathrm{neigh}(i)} \rho^{eff}(r_{ij})$ and the embedding energy using $F(\overline{\rho}_i) = \sum_n b_n (\overline{\rho}_i)^n$ with b_n as additional fitting parameters. Thus their procedure for determining ρ^{eff} was: 1) fit the $\rho^{(l)}(r)$ functions for describing η_{GS}, then 2) compute the integral version of ρ^{eff} and 3) fit this version using an analytic approximation. Though this seems quite convoluted they were able to produce a relatively accurate EAM potential for fcc Al.

Expanding on EAM, Baskes *et al.* [17, 19] introduced the modified EAM (MEAM) to describe Si. MEAM introduces angular dependencies into the embedding density function using $\rho_i = \sum_{j\neq i}\rho_j^a\left(r_{ij}\right) + \sum_{j\neq i,k\neq i}\left[a_j^1 a_k^1 \cos\theta_{jik} - a_j^2 a_k^2\left(1-3\cos^2\theta_{ijk}\right)\right]\ \rho_j^a\left(r_{ij}\right)\rho_k^a\left(r_{ik}\right)$. These angular dependencies were found useful for stabilizing non fcc lattice types. In a study of multiple different lattice types (fcc, bcc, diamond cubic) it was found that metals, semiconductors and even diatomic gases could be described adequately with MEAM [18].

3.3.2 Other Empirical Potential Types and Relationships with (M)EAM

Other popular glue-type potentials include the Finnis-Sinclair and the ADP potential. The Finnis-Sinclair potential was proposed by Finnis and Sinclair [66] to model bcc transition metals. It is similar analytically to EAM but the embedding energy is set equal to the square root of $\overline{\rho}_n$ as $-A\sqrt{\overline{\rho}_n}$. The embedding energy in the Finnis-Sinclair potential acts as an attractive term whereas the pair interaction is repulsive. The embedding energy was chosen as the square root to mimic tight binding theory results, where the cohesive energy per atom varies as the square root of the atom coordination.

In an approach similar to MEAM, Mishin *et al.* [153] suggested the angular dependent potential (ADP), which can be seen as a more general form of the EAM including angular terms directly in the energy function, instead of the embedding density alone. They give the energy function as $E = \frac{1}{2}\sum_{i\neq j}\Phi_{s_is_j}\left(r_{ij}\right) + \sum_i F_{s_i}\left(\overline{\rho}_i\right) + \frac{1}{2}\sum_{i,\alpha}\left(\mu_i^\alpha\right)^2 + \frac{1}{2}\sum_{i,\alpha,\beta}\left(\lambda_i^{\alpha\beta}\right)^2 - \frac{1}{6}\sum_i v_i^2$, where i and j enumerate atoms, α and β enumerate the Cartesian directions and s_i defines the chemical nature of atom i. μ_i^α and $\lambda_i^{\alpha\beta}$ both depend on pairwise functions $u_{s_is_j}\left(r_{ij}\right)$ and $w_{s_is_j}\left(r_{ij}\right)$ and v_i is the trace of λ_i.

Two other potential types to mention here are the Tersoff potential [232] and Stillinger-Weber potential [220] because of their relationship to (M)EAM. Both potentials explicitly contain angular terms and aim to capture two and three-body effects directly, with the Tersoff potential focusing on covalent systems by considering bonds as the central feature. Tersoff potentials combine repulsive and attractive distance-based functions, with the attraction depending on the ‘bond order’, i.e. the strength of bonds to neighbouring atoms, noting that the average bond strength decreases with increasing number of neighbours. Depending on the analytical choice it can be shown that the EAM and the Tersoff potential are analytically identical [29], which is also true for the Stillinger-Weber potential and the MEAM [233]. Noting these similarities Dongare et al. suggested A-EAM, a MEAM-SW hybrid potential type reproducing the Al-Si phase diagram [54, 55].

3.3.3 Machine Learning and Empirical Potentials

The recent application of machine learning in the development of empirical potentials has already lead to a variety of interesting approaches based on polynomial fitting, kernels, artificial neural networks, support vector machines and genetic programming [93].

Polynomial fitting uses atom distances transformed into 'Morse' variables expanding a polynomial. Similarly, but without the transform to Morse variables, *modified Shepard interpolation* develops second order polynomials for each structure and combines them for prediction as a linear superposition. The weights of that superposition are themselves functions of the difference between the structure under consideration and all references structures.

Kernel based regression directly relates a database of pairs of atomic configuration and potential energy. The Gaussian process regression variant was used by Bartok *et al.* [13–15] introducing the Gaussian approximation potential (GAP), which has found applications to C, Si, Ge, Fe, GaN and other systems. Ferre *et al.* [64] suggested a modification of GAP introducing graph theory by using random walk graph kernels instead of overlapping atom densities [64]. Using kernels Botu *et al.* [27, 28] suggested fitting force components directly as a function of the environment instead of fitting energies. A different application of use kernels was recently proposed by Bockherde et al. [30] who directly considered the ground state electron density $\eta_{GS}(\mathbf{r})$. They suggested representing an atomic structure via a custom 'external potential' $v(\mathbf{r})$, for which they used superimposed Gaussians centred on nuclei positions. This representation was then used to generate two machine learning (ML) 'maps', the Kohn-Sham (KS) map ML-KS: $v(\mathbf{r}) \mapsto E$ and the Hohenberg-Kohn map ML-HK: $v(\mathbf{r}) \mapsto \eta(\mathbf{r})$, utilizing the discrete Fourier transform for the latter.

Artificial neural networks (ANNs) underlie the recent surge of 'deep learning'. ANNs are really highly non-linear mathematical functions inspired by and visualized as a neural network, hence the name. ANNs are commonly represented in vertices which are connected by directed edges with certain weights. The vertices contain an 'activation function' (some analytical function) taking some input, either raw data or data transformed by a previous set of vertices. Distinctions are made between three types of ANNs: the feed forward ANN (no loops), the convolutional ANN (deep version of feed forward) and the recurrent ANN (with loops). Loops in this context are vertices connected by directed edges allowing paths which start and end at the same node. Schuett *et al.* [205] used deep tensor neural network to map molecule structures to energies using distance matrices and atomic charges, training on high symmetry structures as well as MD configurations.

Support vector machines (SVMs) are highly popular in the machine learning community, because of their high sparsity making them computationally efficient. They describe some hypersurface, usually for classification, by selecting samples from the data set as support vectors. But they can also be used for regression such as for empirical potentials.

Genetic programming, not to be confused with genetic optimization, is a way to find functional forms by providing a set of allowed mathematical operations changing and selecting them in a manner typical for genetic optimization. Genetic programming allows for the search of functional forms for empirical potentials themselves.

3.4 Optimization Algorithms

Optimization problems are described by some utility / fitness / objective function f. For this work it will be sufficient to consider $f : \mathbf{x} \in \mathbb{R}^N \mapsto \mathbb{R}$. f could, for example, be the Euclidean distance between a model function and some observed data to be regressed, i.e. between an empirical potential and DFT data. Without loss of generality we will assume that f needs to be minimized by $\mathbf{x}^*$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})\ \forall \mathbf{x} \in \mathscr{C}$, with $C \subseteq \mathbb{R}^N$. Depending on the problem one may distinguish between three different types: local minimization ($C = \{\mathbf{x} + \Delta\mathbf{x}\} \subset \mathbb{R}^N$ for some finite $\Delta\mathbf{x}$), global minimization (in the case of real numbers $C = \mathbb{R}^N$) and constrained minimization ($C = U \subset \mathbb{R}^N$). Global minimization is the most challenging case, for which, in virtually all cases, no guarantee exists that a value $\mathbf{x}^*$ indeed is the global minimum. Since commonly no ideal initial guess or constraint can be provided in the development of empirical potentials this problem belongs to the global optimization type. Depending on the problem, different requirements are presented to optimization algorithms, i.e. f may have discontinuities or may generally be non-smooth as happens when dealing with stochastic data. The smoothness has implications for the differentiability of f and thus the possible speed or chance deeper minima are found at all [118]. Examples of algorithms requiring continuous differentiability are Newton-Raphson, conjugate gradient and BFGS. A range of implemented optimization algorithms can be found in the Python package *scipy* [116].

Of particular relevance for this work are algorithms which can deal with partially stochastic f. Thus, we turn to what Miguel et al. calls ‘derivative free’ methods [189]. They distinguish derivative-free optimization algorithms further as ‘direct’ or ‘model based’, ‘local’ and ‘global’ and ‘stochastic’ or ‘deterministic’. ‘Direct’ and ‘model based’ specifies whether the algorithm uses the observed f values directly for directionality input or whether it creates a model of f itself for guidance. ‘Local’ and ‘global’ specifies whether an algorithm is able to refine and widen the scope of search or not, with ‘global’ algorithms being able to do so and ‘local’ algorithms not. An example of a ‘direct’ and ‘local’ algorithm is Nelder-Mead.

The response surface method, which fits a surrogate function to observed data and progressively refines the fit with new data, classifies as a ‘model based’ ‘global’ algorithm. To ‘stochastic’ ‘global’ algorithms belong the *hit and run* algorithm, *simulated annealing*, *basin hopping*, *genetic / evolutionary algorithms*. Hit and run is the simplest among these stochastic algorithms, updating a current solution $\mathbf{x}$ only with the randomly generated $\mathbf{x}_{\text{new}}$ if $f(\mathbf{x}) > f(\mathbf{x}_{\text{new}})$. Simulated annealing, inspired by annealing in metallurgy [122], always improves the current solution when a better one is available but also accepting a worse solution, by chance, controlled by the ‘annealing temperature’ T specified by $p(\mathbf{x}_{\text{new}}|\mathbf{x})$ as

$$p(\mathbf{x}_{\text{new}}|\mathbf{x}) = \begin{cases} \exp\left[-\frac{f(\mathbf{x}_{\text{new}})-f(\mathbf{x})}{T}\right], & \text{iff } f(\mathbf{x}_{\text{new}}) > f(\mathbf{x}) \\ 1, & \text{o.w.} \end{cases} . \tag{3.25}$$

Basin hopping models f with discrete segments corresponding to local minima in f removing intermediate states and modifying $\mathbf{x}$ using Monte Carlo [251, 252]. Genetic / evolutionary algorithms, introduced by Rechenberg [185] and Schwefel [207], are inspired by the theory of evolution, where individuals exchange genes and descendants are selected to improve the group’s ‘fitness’. Extending the family of genetic algorithms, Hansen *et al.* [94, 95] proposed the covariance matrix adaption algorithm (ES-CMA) which samples f and develops a covariance matrix to encourage mutation in directions which were beneficial in the past. Python implementations of genetic algorithms and ES-CMA can be found in the DEAP packages [68]. Recently it has been found that ES-CMA can be seen as part of the natural evolution strategies [260], which compromise genetic algorithms and iteratively update a search distribution using an estimated gradient of the distribution parameters.

Chapter 4

Machine Learning and Atomistic Simulations

Reflected in the name ‘machine learning’ is the ambition to endow computer programs / machines with the ability to improve themselves via ‘learning’. This idea was summarized by Tom Mitchell in 1997

> "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.”,

emphasizing the dynamic ability of computer programs to adapt. While it is far from obvious what each of those quantities should be in reality, it has been found that the amount of data itself has a significant impact, also referred to as the ‘unreasonable effectiveness of data’ [90, 226]. That is, increasing the amount of data available during the training phase of a program tends to increase the program’s performance more than using a more sophisticated program, leading to recent breakthroughs in computer vision. The range of application cases which are associated with machine learning is quite wide and includes, without being exhaustive, recommendation systems of Netflix and Amazon, visual and auditory recognition, identification of planets in neighbouring solar systems (as of 2018 the Kepler-90 is identified as the first system to also have 8 planets [209], solving games such as Go via reinforcement learning without human knowledge [210], prediction of stock prizes and chatbots which learn language from social media. Naturally, machine learning has also started to find its application in materials science as part of ‘materials informatics’ to improve the understanding of structure-property relationships [183], to search for new materials and to improve the simulations of materials, e.g. via the improvement of semi-empirical potentials for atomistic simulations [20].

Though machine learning algorithms are quite divers in their approaches they all can be described with a common recipe containing four ingredients [83]:

1. a data set with input $\mathbf{X}$ and response $\mathbf{t}$,
2. a model function $y(\mathbf{X})$ (and optionally an associated probability distribution p_{model}),
3. a cost function $L(y(\mathbf{X}), \mathbf{t})$ and,
4. an optimization procedure.

Points 2.-4. will be discussed in more detail in the following sections.

In general, machine learning algorithms can be described as ‘supervised’ or ‘unsupervised’. The difference between those two is whether $\mathbf{t}$ is known / available or not. Algorithms which lie somewhere in between are referred to as ‘semi-supervised’. Unsupervised methods include hierarchical clustering and principal component analysis. Supervised methods, which this chapter is devoted to, include regression methods such as linear regression, Bayesian ridge regression and others and classification methods such as (naive) Bayesian classification.

A further distinction made worth mentioning is ‘model-based’ versus ‘memory-based’, where ‘model-based’ algorithms are designed to learn an abstract concept of the observations but discard them for further predictions. An example is linear regression. ‘Memory-based’ algorithms on the other hand explicitly make use of prior observations and compare them to new observations for predictions. An example is k nearest neighbours, which makes a prediction for a new observation based on, for example, an average of the k most similar prior observations.

Why probabilities?

Often times when a system is modelled the system has some inherent stochasticity, not every necessary variable has been observed or the applied model itself is incomplete. All those effects add to increased uncertainty about the system being modelled and the predictions one can possibly make. One approach to quantify uncertainties is the probability measure. For the interpretation of the probability measure two points of view have developed, the Bayesian and the frequentist view. Inherent to the Bayesian view is that probability reflects a belief for a possible outcome, whereas the frequentist view sees probability as the number of times an outcome has occurred under identical repetitions of the same test.

Though a full discussion of frequentist methods versus Bayesian methods is beyond this work, the main objections to Bayesian statistics shall briefly be discussed. The objections are regarding the correctness of automated inference engines, e.g. variational inference, and the transferability of subjective priors [75]. Priors generally capture the belief of how likely outcomes for a certain observable are before making any observations. Thus, although different researchers may have different preferences regarding the analytical forms of their priors they do allow the inclusion of further knowledge about a system. Also, testing all kinds of priors has become very achievable with recent implementations, e.g. *PyMC3* [196] and *Edward* [239]. Current implementations also allow the tracking of the convergence of those automated inference engines to ascertain the correctness of the results.

For the regression and classification methods introduced in the following also exist non-Bayesian counterparts, e.g. Ridge regression is the non-Bayesian pendant of Bayesian ridge regression. In this thesis work the Bayesian approach is chosen over the non-probabilistic approach because it generally allows the treatment of uncertainty while also including the regularization of parameters and allowing model comparison, as outlined in MacKay's seminal work on regression [144].

Machine Learning in this Thesis

Analysing atomic configurations and developing models for interatomic potentials may seem conceptually like very different tasks. Adopting the Bayesian point of view, however, provides a foundation for tackling both problems. In the machine learning literature these problems fall into *regression* and *classification* categories.

Classification based on density estimation will play a crucial role in Chapter 5 where we will discuss the onset of a $\gamma \rightarrow \gamma + \gamma'$ phase transition observed in single and bi-crystals of γ Ni-20 at.% Al. Linear and non-linear Bayesian regression will play an important role in Chapter 7 where the regression of electron densities in space and the training of empirical potentials is investigated.

The remainder of this chapter covers the basic parts required for the machine learning algorithms in the result chapters. First random variables and their distributions will be introduced (section 4.1), followed by the discussion of models in general and the risk/loss/cost function (section 4.2), Bayesian regression (section 4.3) and Bayesian classification (section 4.4).

4.1 Random Variables & Distributions

In order to develop a probabilistic description, one defines *random variables* which are observed according to some *distribution*. Working with probabilities, defined by random variables and distributions, requires some rules which will be discussed in the following. Essentially anything that can be observed can be phrased in the form of a random variable, hence they can take all sizes and shapes. Scalar valued random variables can be either discrete or continuous, i.e. regression deals mostly with continuous random variables whereas in classification we have discrete random variables. Random variables are usually indicated as upper case Roman letters, e.g. X. Further, observed random variables taking on the value x are then usually denoted as $X = x$. In this section we will keep this notation but will drop it afterwards for notational simplicity, if not stated explicitly otherwise, interpreting x always $X = x$.

A deeper discussion of probability theory may require the discussion of *events*. This is deemed unnecessary for the applications in this work, but for thorough introductions we wish to refer to the excellent book by Koller *et al.* [124], in which so-called probabilistical graphical models are the particular focus. Probabilistical graphical models, which are a combination of graph theory and probability theory, enable very efficient algorithms to evaluate highly intricate relationships between random variables. Although not discussed explicitly in this work there are always probabilistic graphical model analogues for the following models.

Generally, distributions of random variables are referred to as the *probability density function* (PDF) in the continuous case and to *probability mass function* (PMF) in the discrete case. Observed values can be imagined as samples from underlying distributions of random variables. A single random variable X distributed according to a distribution, for example, the normal distribution $\mathcal{N}\left(\mu,\sigma^2\right)$, will be denoted here as $X \sim \mathcal{N}\left(\mu,\sigma^2\right)$. In order for a distribution $p(X)$ of random variable X to be a *proper* distribution it has, in the continuous case, to obey

$$\int_{\Omega} p(X=x)\,dx = 1,\, p(X=x) > 0 \forall x \in \Omega, \tag{4.1}$$

whereas in the discrete case the integral would be replaced by a sum. Once the distribution of a random variable is known it can be used to calculate useful characteristic quantities of the distribution itself, like the expectation value

$$\mathbb{E}[X] = \int_{\Omega_x} x p(X)\,dx \tag{4.2}$$

and the variance

$$\mathrm{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 . \tag{4.3}$$

In cases where more than one random variable occurs, for example X and Y we use a *joint distribution* $p(X,Y)$. In this case the distribution over a single random variable like $p(X)$ is also referred to as the *marginal distribution*. The marginal distribution can be obtained by summing, in the discrete case, or integrating, in the continuous case, the joint distribution. This process is then also referred to as marginalization $p(X) = \int_{\Omega_y} p(X, Y = y)\, dy$.

Two random variables X and Y may be dependent on each other. They are said to be independent of each other, denoted as $X \perp Y$, if and only if $p(X,Y) = p(X)\, p(Y)$ is satisfied. Generally the dependence behaviour, for example of X given Y, is described by the *conditional probability distribution* as

$$p(X|Y) = \frac{p(X,Y)}{p(Y)}, \tag{4.4}$$

where $p(Y)$ acts to normalize the joint probability distribution. This definition of the conditional probability distribution can be rewritten to obtain the chain rule $p(X,Y) = p(X|Y)\, p(Y)$ which can in turn be used to obtain the so-called *Bayes theorem*

$$p(X|Y) = \frac{p(Y|X)\, p(X)}{p(Y)}, \tag{4.5}$$

with a very useful modification using the chain rule again and marginalization leading to $\frac{p(Y|X)p(X)}{p(Y)} = \frac{p(Y|X)p(X)}{\int_{\Omega_x} p(Y|X)p(X)dx}$.

In the case of D, the observed ‘data’, and W the model weights, $p(W|D)$ is also referred to as the posterior $p(W|D) = \frac{p(D|W)p(W)}{p(D)} = \frac{\text{Likelihood}\cdot\text{Prior}}{\text{Evidence}}$. Having defined the above we have completed our set of concepts for the remainder of the chapter.

4.2 Models and the Risk/Cost/Loss Function

In the context of this thesis work it is of interest to predict some continuous or discrete quantity given some input, e.g. classify an atom or predict its force vector. The class of models which directly learn this connection without using probabilities are referred to as ‘discriminative’ models. If, however, the model is embedded in a probabilistic framework one differentiates between ‘probabilistic discriminative’ and ‘generative’ models. The former approximates the posterior distribution of the predicted quantity, i.e. $p(y \mid x)$, and the latter approximates the joint distribution, i.e. $p(x,y)$, capturing the distribution of the input data as well. Ordinary least squares, for example, is a discriminative model, Bayesian regression (see sections 4.3.1, 4.3.2 and 4.3.3) is probabilistic discriminative and density approximation using Gaussian mixture models (see section 4.3.4) is generative. Going from discriminative-only to generative models implies an increase in computational complexity as more parameters are required to capture the structure in the data. A recent advance in this field are the 'generative adversarial networks' proposed by Goodfellow *et al.* [82], having already led to the generation of highly photo-realistic images sampling from the approximating distribution [134].

In order to manipulate the chosen model to accurately make predictions one introduces a 'risk', 'cost' or 'loss' function, $L(y,t)$. Loss functions provide a measure of the deviation of the predicted value of y, given observation x, from the true reference values t. L can, for example, be the 0-1 loss for classification or the Euclidean distance for regression. 0-1 loss leads to a loss of 1 for each miss-classification and 0 otherwise. Hence one way of adjusting the models is to minimize the expected loss

$$\mathbb{E}[L] = \int L(y(x,\theta),t)\, p_{\text{data}}(x,t)\, dx dt, \tag{4.6}$$

with the predictive model $y(x,\theta)$, specified by model parameters θ. $p_{\text{data}}(x,t)$ is the joint distribution of the data, the true joint distribution of x and t.

The optimization of the model y is then carried out using $\nabla_\theta \mathbb{E}[L] \overset{!}{=} 0$. Since the true distribution of $p_{\text{data}}(x,t)$ usually is not available one uses the discrete version summing over all available observations

$$\mathbb{E}[L] = \frac{1}{N}\sum_{i=1}^{N} L(y(x_i,\theta),t_i) \tag{4.7}$$

leading to 'empirical risk minimization'. Also, note that $\mathbb{E}[L]$ can be re-written using

$$L(y(x,\theta),t) = -\log p_{\text{model}}(t \mid y(x,\theta)) \tag{4.8}$$

equating the minimization of $\mathbb{E}[L]$ with the minimization of the cross-entropy of p_{model} and p_{data}.

Employing probabilistic models has advantages and disadvantages. A significant disadvantage may be the computational cost, which until recently prohibited their application. Recent successful approaches use Markov chain Monte Carlo sampling and variational methods to describe intricate probabilistic models. Naive Bayes, an approximation which alleviates the computational cost, will be discussed in the section 4.4 in the context of classification. An advantage is that a characterization of a model in terms of a probability distribution allows for alterations of the loss function, without having to recompute the model. Also, one obtains a certainty of the prediction, which can be used to minimize false classification due to high uncertainty by applying a threshold flagging classification of high uncertainty. Furthermore, one may be handling data sets with $\mathbf{t}$ available only for a small fraction of the entire data set, but which must be approximated correctly. Using a probabilistic approach, models can be developed on adjusted data sets leading to useful models on the original data set. Having developed a probabilistic model, one can perform a refined analysis using the rules of probability.

4.3 Bayesian Regression

The Bayesian approach to regression is useful because it provides an underlying theory for the process, including the regularization of parameters and model comparison, as outlined in MacKay's seminal work [144]. In that study the regression and the model comparison are referred to as the first and second level of inference. The latter in particular can be useful when testing different hypotheses about observed data. In this section we differentiate between linear and non-linear models as well as parametric and non-parametric models. Lastly, we will discuss the Expectation Maximization (EM) algorithm for density estimation.

Even though linear models may sound like a very simplistic approach they turn out to be quite useful. In fact, the name is slightly misleading as linear models are linear maps from the feature space to a space of target values, but the map to the feature space may be non-linear. The focus in this section will be on Bayesian methods, even though there are a variety of others. This is partly because Bayesian methods were used in this work and also because non-explicit Bayesian methods are often very closely related to Bayesian methods and can be cast as such with appropriate priors. Examples of non-explicit Bayesian methods in regression are linear regression, LASSO and ElasticNet among others.

The following subsections are ordered such that the more general non-linear parametric models are discussed first, which is followed by linear parametric models, which are a subset. Parametric linear models are of interest because they allow for elegant analytical solutions, with Relevance Vector Machines (RVMs) being a prominent example. This discussion is followed by the non-parametric version of linear models discussing Kernel Ridge Regression (KRR) and Gaussian Processes (GPs). Linear regression and KRR can be considered non-Bayesian alternatives to RVMs and GPs respectively. Lastly, density estimation will be outlined which regresses density distributions, representing an unsupervised learning problem.

All the models in the following realize their loss function via their probabilistic model. That loss function is then minimized either by a generic optimizer, in the case of the parametric non-linear models, or by iterating analytically derived update equations.

4.3.1 Parametric Non-linear Models

The Bayesian framework for regression includes the Occam factor, which allows for the comparison of models / hypotheses, providing a theoretical underpinning for the calculation of model parameters and hyperparameters. Generally, hyperparameters regulate the obtained model parameters. How hyperparameters exactly relate to the model parameters will be explained below.

Here and in section 4.3.2, which deals with parametric linear models, we wish to describe some observed value t given some input vector $\mathbf{x}$ via a model $y(\mathbf{w},\mathbf{x})$ which depends on model parameters / weights $\mathbf{w}$ which are to be determined. Because of the dependence of the models on $\mathbf{w}$ they are dubbed ‘parametric’. The main difference here to section 4.3.2 is that we have no information about the analytic form of $y(\mathbf{w},\mathbf{x})$. Thus parametric linear models in section 4.3.2 form a special case of the parametric non-linear models.

Having specified some model $y(\mathbf{w},\mathbf{x})$ we wish to approximate N observed target values, $\mathbf{t}$, with respective input $\mathbf{x}_i$ for t_i, with $i \in [1,N]$ by modifying the M model parameters in $\mathbf{w}$. The collective form of all N individual observations t_i is here denoted as $\mathbf{t}$ and all inputs $\mathbf{x}_i$ as $\mathbf{X}$. Correspondingly, the values the model takes on for $\mathbf{X}$ is denoted as $\mathbf{y}$ such that $\mathbf{y} = y(\mathbf{w},\mathbf{X})$.

With a representation of the model at hand we can describe the deviation $\mathbf{y}-\mathbf{t}$ in the form of ‘energy’ such that better fits are lower in energy. Drawing from statistical physics we define the deviation of our model from the data as energy $E_D : (\mathbf{t}, y(\mathbf{w},\mathbf{X})) \mapsto \mathbb{R}$, where a popular choice is the euclidean/2-norm $E_D = \frac{1}{2}\|\mathbf{t}-\mathbf{y}\|_2^2$ with the corresponding probability $p(\mathbf{t}|\mathbf{w},\mathbf{X},\beta) = \frac{e^{-\beta E_D}}{Z_D(\beta)}$, where $Z_D(\beta) = \sqrt{\frac{\beta}{2\pi}}$. E_D may also be referred to as the *loss* or *risk* function, see section 4.2. This is equivalent to assuming a normal distribution of $E_D \sim \mathcal{N}\left(0,\beta^{-1}\right)$ or $t \sim \mathcal{N}\left(y(\mathbf{w},\mathbf{x}),\beta^{-1}\right)$ for our target values, with the noise level β being defined through the variance σ^2 as $\beta \equiv \sigma^{-2}$. In order to include the weights into our

Bayesian treatment we can map $\mathbf{w}$ to energy values as $E_R : (\mathbf{w}, \mathbf{\Sigma}_R) \mapsto \mathbb{R}$. For example $E_R = \mathbf{w}^T \mathbf{\Sigma}_R^{-1} \mathbf{w}$ with $\mathbf{\Sigma}_R \in \mathbb{R}^{M \times M}$. This implies the distribution $\mathbf{w} \sim \mathcal{N}(0, \mathbf{\Sigma}_R)$ as $p\left(\mathbf{w}|\mathbf{\Sigma}_R^{-1}\right) = \frac{e^{-E_R}}{Z_R(\mathbf{\Sigma}_R)}$ with $Z_R(\mathbf{\Sigma}_R) = \int e^{-E_R} d\mathbf{w}$. Using Bayes theorem we can write

$$p\left(\mathbf{w}|\mathbf{X}, \mathbf{t}, \mathbf{\Sigma}_R^{-1}, \beta\right) = \frac{p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \beta)\, p\left(\mathbf{w}|\mathbf{\Sigma}_R^{-1}\right)}{p\left(\mathbf{X}, \mathbf{t}|\mathbf{\Sigma}_R^{-1}, \beta\right)} \tag{4.9}$$

and keeping the previous notation this is equal to

$$p\left(\mathbf{w}|\mathbf{X}, \mathbf{t}, \mathbf{\Sigma}_R^{-1}, \beta\right) = \frac{e^{-E_M}}{Z_M\left(\mathbf{\Sigma}_R^{-1}, \beta\right)} \tag{4.10}$$

with $E_M = \beta E_D + E_R$ and $Z_M(\Sigma_R, \beta) = \int e^{-E_M} d\mathbf{w}$.

Finding the *maximum a posteriori* (MAP) point with $\mathbf{w}_{\text{MAP}}$, $\mathbf{\Sigma}_{R,\text{MAP}}$ and β_{MAP} then comes down to iteratively maximizing eq. 4.9, yielding $\mathbf{w}_{\text{MAP}}$, and calculating the hyperparameters. It is preferable to maximize $\ln p\left(\mathbf{w}|\mathbf{X}, \mathbf{t}, \mathbf{\Sigma}_R^{-1}, \beta\right)$ using the logarithm because of limited floating point precision on computers. A general approach for $p\left(\mathbf{w}|\mathbf{X}, \mathbf{t}, \mathbf{\Sigma}_R^{-1}, \beta\right)$ uses the approximation as a normal distribution $\mathcal{N}(\mathbf{w}|\mathbf{m}, \mathbf{S})$. This can be justified by the Bayesian Central Limit Theorem [144]. To find the hyperparameters $\mathbf{\Sigma}_R$ and β, one can use the *type-II maximum likelihood* or *evidence method* which maximizes

$$\ln p\left(\mathbf{\Sigma}_R^{-1}, \beta|\mathbf{t}, \mathbf{X}\right) = \ln \frac{p\left(\mathbf{t}|\mathbf{X}, \mathbf{\Sigma}_R^{-1}, \beta\right) p\left(\mathbf{\Sigma}_R^{-1}, \beta\right)}{p(\mathbf{t}, \mathbf{X})}, \tag{4.11}$$

which is equivalent to

$$\begin{aligned} \mathbf{\Sigma}_{R,\text{new}}, \beta_{\text{new}} &= \operatorname{argmax}_{\mathbf{\Sigma}_R, \beta} \ln\left\{p\left(\mathbf{t}|\mathbf{X}, \mathbf{\Sigma}_R^{-1}, \beta\right) p\left(\mathbf{\Sigma}_R^{-1}, \beta\right)\right\} \\ &= \operatorname{argmax}_{\mathbf{\Sigma}_R, \beta} \ln\left\{\int p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \beta)\, p\left(\mathbf{w}|\mathbf{\Sigma}_R^{-1}\right) d\mathbf{w}\right\}. \end{aligned} \tag{4.12}$$

Because both of the distributions under the integral are Gaussian the resulting distribution can be calculated exactly. The approximation of the MAP with a Gaussian is referred to as *Laplace method* and can also be useful as a first approximation where the actual distribution is not Gaussian [144]. In the case of $\mathbf{\Sigma}_R^{-1} = \alpha \mathbf{I}$, with $\mathbf{I}$ the identity matrix, α and β can be estimated as $\alpha_{\text{new}} = \frac{\gamma}{2E_R}$ and

$$\beta_{\text{new}} = \frac{N - \gamma}{2E_D}, \tag{4.13}$$

defining $\gamma \equiv k - \alpha \operatorname{trace} \mathbf{A}^{-1}$, $\mathbf{B} = \nabla_{\mathbf{w}} \nabla_{\mathbf{w}} E_D$, $\mathbf{C} = \nabla_{\mathbf{w}} \nabla_{\mathbf{w}} E_R$ and $\mathbf{A} = \alpha \mathbf{C} + \beta \mathbf{B}$.

Estimation of the hyperparameters can be performed in a similar fashion, by integrating out the model weights $\mathbf{w}$ and finding the most likely hypothesis $\mathscr{H}$. This hypothesis could contain the analytic form of the model chosen, the assumed distributions or other assumptions made. Analytically this can be represented, for example, as $\ln p\left(\mathbf{X},\mathbf{t}|\mathscr{H}_i\right) = \int p\left(\mathbf{t}|\mathbf{X},\boldsymbol{\Sigma}_R^{-1},\beta,\mathscr{H}_i\right)\, p\left(\boldsymbol{\Sigma}_R^{-1},\beta|\mathscr{H}_i\right) d\boldsymbol{\Sigma}_R d\beta$ where the last three factors make up the 'Occam factor' $\left(p\left(\boldsymbol{\Sigma}_R^{-1},\beta|\mathscr{H}_i\right)\, d\boldsymbol{\Sigma}_R d\beta\right)$.

As an example of regression we use here the function

$$y = m\cdot x + b + w_1 a_1 \sqrt{\frac{\beta_1}{2\pi}} e^{-\frac{\beta_1}{2}(x-\mu_1)^2} + (1-w_1)\, a_2 \sqrt{\frac{\beta_2}{2\pi}} e^{-\frac{\beta_2}{2}(x-\mu_2)^2} \tag{4.14}$$

and a noise level of 1 to generate t. Throughout this chapter the different approaches will be illustrated using this example, setting $m = 10$, $b = 0$, $w_1 = 7/10$, $a_1 = 5$, $\beta_1 = 10$, $\mu_1 = 1/2$, $a_2 = -5$, $\beta_2 = 200$, and $\mu_2 = 2.5$. When using non-linear parametric models we find $\mathbf{w}$ by minimizing eq. 4.9, calculating $\mathbf{w}_{MAP}$. Unlike in real cases where the ideal model to fit is unknown we have applied a BFGS minimizer directly to the generating model eq. 4.14, assuming $t \sim \mathscr{N}\left(y,\beta^{-1}\right)$. The generated data values t and correct y are shown in fig. 4.1 alongside the obtained fit. Unsurprisingly we find a very close match with high confidence in the model values, visible as the shaded green area.

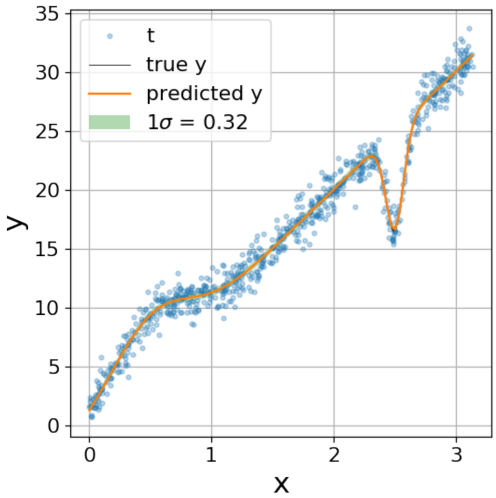


Figure 4.1: Regressing a test case specified by eq. 4.14 distorted with noise, normally distributed with $\beta = 1$, visible as blue dots. Applying the BFGS minimizer leads to $\mathbf{w}_{\mathrm{MAP}}$ yielding the fit shown by the orange line closely matching the true model values y before noise. The green shaded area represents 1 standard deviation around $y\left(\mathbf{x},\mathbf{w}_{\mathrm{MAP}}\right)$.

4.3.2 Parametric Linear Models

Linear models are specified by

$$y(\mathbf{w}, \mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x}, \tag{4.15}$$

where $\mathbf{w}$ are the model parameters and $\mathbf{x}$ are the input for the model, as before. Depending on whether or not $\mathbf{x}_n \perp \mathbf{x}_{n'}$ for $n' \neq n$ linear models can be expanded to describe a range of systems such as dynamical systems, hidden Markov models, Principal Component Analysis and others [193]. But here we assume that $\mathbf{x}_n \perp \mathbf{x}_{n'}$ for $n' \neq n$. Applying non-linear mappings ϕ to $\mathbf{x}$ can be used to transform the observation input $\boldsymbol{x}$ into the feature space using some basis function set, more suitable for the task at hand. Such transformations / basis function sets include polynomials $\phi_k : \mathbf{x} \mapsto \mathbf{x}^k$, cosine series $\phi_k : \mathbf{x} \mapsto \cos k\mathbf{x}$ and others, denoting the vector of all basis functions as $\boldsymbol{\phi}$. The size of $\boldsymbol{\phi}$ determines the size of $\mathbf{w}$. If $\boldsymbol{\phi} : \mathbb{R}^{M'} \mapsto \mathbb{R}^M$ then $\mathbf{w} \in \mathbb{R}^M$. Although the choice of $\boldsymbol{\phi}$ is not always obvious, an appropriate choice can prove very useful. The regression of the ground state electron density $\eta_{GS}(\mathbf{r})$ from DFT calculations described in Chapter 7 will rely heavily on this transformation into feature space. Applying the chosen basis set of M functions $\{\phi_k\}_{k \in [1,M]}$ to all N observations $\mathbf{x}_i$ creates the *design matrix* $\mathbf{\Phi}_{ik} = \phi_k(\mathbf{x}_i)$. The construction of the design matrix is the essential part to the application of methods based on linear models. Another approach based on kernel matrices, which can avoid the construction of a design matrix explicitly, will be discussed in the following section.

Assuming additive normally distributed noise $\varepsilon \sim \mathcal{N}\left(0, \beta^{-1}\right)$ we can include uncertainty into the linear model in eq. 4.15. These models are accordingly referred to as linear Gaussian models. Including the nonlinear mapping based on a chosen basis set we have

$$y(\mathbf{w}, \mathbf{x}) = \mathbf{w}^T \cdot \boldsymbol{\phi}(\mathbf{x}) + \varepsilon. \tag{4.16}$$

To fully treat the linear model with the Bayesian framework we choose a distribution for the weights, also referred to as the *prior* of the weights. Choosing that prior to also be of the exponential family is advantageous because it allows a complete analytical treatment. Particularly elegant solutions can be obtained with a normal prior for the weights, since products of normal distributions yield a normal distribution as well, leading to a normal posterior distribution of the weights. Generally, priors which lead to posteriors of the same family are also called *conjugate priors*. In order to specify the weight distribution given data we want to find $p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \mathbf{\Sigma}_R, \beta)$. We could compute the mean of that distribution by maximizing eq. 4.9. However, as indicated in the previous section, because of the choice of a linear model we can compute the weights posterior distribution exactly. With respect to the calculation of the hyper parameters $\mathbf{\Sigma}_{R,\text{new}}$ and β_new defined in the previous section we here only need to change how $\mathbf{\Sigma}_{R,\text{new}}$ is calculated.

Amazingly enough, in the case of the linear models by, using the normal weights prior $p\left(\mathbf{w}|\mathbf{\Sigma}_R^{-1}\right) = \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_R\right)$ and the normal likelihood $p\left(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta\right) = \mathcal{N}\left(\mathbf{t}|\mathbf{\Phi}\mathbf{w}, \beta^{-1}\right)$ we can find the posterior distribution for the weights as

$$p\left(\mathbf{w}|\mathbf{X}, \mathbf{t}, \mathbf{\Sigma}_R, \beta\right) = \mathcal{N}\left(\mathbf{w}|\mathbf{A}^{-1}\Phi^T\mathbf{t}, \mathbf{A}\right), \tag{4.17}$$

and hence

$$\mathbf{w}_\text{MAP/new} = \left(\mathbf{\Phi}^T\mathbf{\Phi} + \mathbf{\Sigma}_R^{-1}\right)^{-1}\mathbf{\Phi}^T\mathbf{t}. \tag{4.18}$$

with $\mathbf{A} = \mathbf{\Phi}^T\mathbf{\Phi} + \mathbf{\Sigma}_R^{-1}$ solely based on re-arranging of the exponents of the joint distribution $p\left(\mathbf{w}, \mathbf{t}|\mathbf{X}, \mathbf{\Sigma}_R, \beta\right)$ [24]. Now we are left to find the remaining hyperparameters, $\mathbf{\Sigma}_R$.

Relevance Vector Machines

The determination of $\mathbf{\Sigma}_R$ for independent weights can be done via so-called Relevance Vector Machines (RVMs), originally developed by Tipping and Faul [61, 235, 236]. This approach gets its name from the similarity to the so-called Support Vector Machines (SVMs), which are based on most informative observations / support vectors. Contrary to SVMs, the RVMs find the most relevant vector of basis functions by pruning. In RVMs $\mathbf{\Sigma}_R = \text{diag}\left(\boldsymbol{\alpha}\right)$ such that each weight is distributed independently normally as $w_i \sim \mathcal{N}\left(0, \alpha_i^{-1}\right)$. RVMs consider the problem of model complexity explicitly and make changes dynamically leading to high sparsity by optimizing the log evidence $\log p\left(\boldsymbol{\alpha}, \beta|\boldsymbol{t}, \mathbf{X}\right) = \mathcal{L}\left(\boldsymbol{\alpha}, \beta\right)$ for individual α_i. This relieves part of the problem in the model comparison because the basis set can be large encompassing various functions. In the end the approximations that will be preferred will maximize $\mathcal{L}\left(\boldsymbol{\alpha}, \beta\right)$. Recalling eq. 4.11 from the previous section we have

$$\mathcal{L}\left(\boldsymbol{\alpha}, \beta\right) = \ln p\left(\boldsymbol{\alpha}, \beta|\mathbf{t}, \mathbf{X}\right) = \frac{p\left(\mathbf{t}|\mathbf{X}, \boldsymbol{\alpha}, \beta\right)p\left(\boldsymbol{\alpha}, \beta\right)}{p\left(\mathbf{t}|\mathbf{X}\right)} \tag{4.19}$$

thus

$$\begin{aligned}\text{argmax}_{\boldsymbol{\alpha},\beta}\mathcal{L}\left(\boldsymbol{\alpha}, \beta\right) &\propto p\left(\mathbf{t}|\mathbf{X}, \boldsymbol{\alpha}, \beta\right)p\left(\boldsymbol{\alpha}, \beta\right) \\ &= \int p\left(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta\right)p\left(\mathbf{w}|\boldsymbol{\alpha}\right)d\mathbf{w}.\end{aligned} \tag{4.20}$$

Assuming that $p(\boldsymbol{\alpha},\beta)$ is uniform we find

$$\int p(\mathbf{t}|\mathbf{X},\mathbf{w},\beta)\,p(\mathbf{w}|\boldsymbol{\alpha})\,d\mathbf{w} = \mathscr{N}\left(\mathbf{t}\,\middle|\,\mathbf{0},\underbrace{\beta^{-1}\mathbf{I}+\mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^{T}}_{\equiv\mathbf{C}}\right), \tag{4.21}$$

with $\mathbf{A}^{-1}=\left(\beta\mathbf{\Phi}^{T}\mathbf{\Phi}+\mathbf{\Sigma}_{R}^{-1}\right)$. This can be found by re-arranging the exponent with completion of the square of $p(\mathbf{t}|\mathbf{X},\mathbf{w},\beta)\,p(\mathbf{w}|\boldsymbol{\alpha})$ followed by the application of the integral over $\mathbf{w}$ and the Woodbury inversion identity, see appendix A.

Crucial to the update of α_i in RVMs is the separation of $\mathscr{L}(\boldsymbol{\alpha},\beta)$ into a term that depends only on α_i and the rest. This is done using $\mathbf{\Phi}\mathbf{\Sigma}_{R}^{-1}\mathbf{\Phi}^{T}=\sum_i \boldsymbol{\phi}_i\alpha_i^{-1}\boldsymbol{\phi}_i^{T}$. One can rewrite $\mathbf{C}$ in two parts $\mathbf{C}_i$ and $\mathbf{C}_{-i}$ with the first part containing only the ith basis function contribution and the second all the others. This can be used to split

$$\mathscr{L}(\boldsymbol{\alpha},\beta)=\mathscr{L}(\boldsymbol{\alpha}_{-i},\beta)+l(\alpha_i,\beta). \tag{4.22}$$

The calculation of the first derivative $\partial_{\alpha_i}\mathscr{L}(\boldsymbol{\alpha},\beta)$ then leads to two extreme points, one at ∞ and one at

$$\alpha_{i,\text{new}}=\frac{S_i^2}{Q_i^2-S_i}, \tag{4.23}$$

with $Q_i\equiv\boldsymbol{\phi}_i^{T}\mathbf{C}_{-i}^{-1}\mathbf{t}$ the ‘quality’ factor and $S_i\equiv\boldsymbol{\phi}_i^{T}\mathbf{C}_{-i}^{-1}\boldsymbol{\phi}_i$ the ‘sparsity’ factor. Q_i can be seen as the part increasing the quality of the model by increasing $\mathscr{L}(\boldsymbol{\alpha},\beta)$ and S_i the part which ‘inflates’ the model, decreasing $\mathscr{L}(\boldsymbol{\alpha},\beta)$. If $Q_i^2>S_i$ the given basis function is to be updated if α_i is currently finite or introduced if $\alpha_i=\infty$. Otherwise for $Q_i^2\leq S_i$ the basis function is to be pruned from the active set of basis functions, as shown by Faul and Tipping [61]. β can be determined in the exact same way as done above for non-linear models (eq. 4.13). Condensing the above, the iterative algorithm for RVMs can be constructed as shown in appendix A.

In fig. 4.2 we apply a cosine based basis set $\{\cos(kx)\}_{k\in[0,30]}$ using RVMs to approximate our example data generated with eq. 4.14. We can see that a close approximation to the observed function can be achieved, although it also contains some periodic deviations. These deviations are due to two factors, the chosen basis functions and their number. As in real applications the size of the basis set and the chosen basis functions are important in order to closely approximate observed data and make adequate predictions. On the right side in fig. 4.2 we can see the weights of the individual basis functions and the uncertainty in those weights (black lines), reflecting the sparse solution. The ability to compute the uncertainty of individual weights distinguishes RVMs from other methods, particularly non-Bayesian methods.

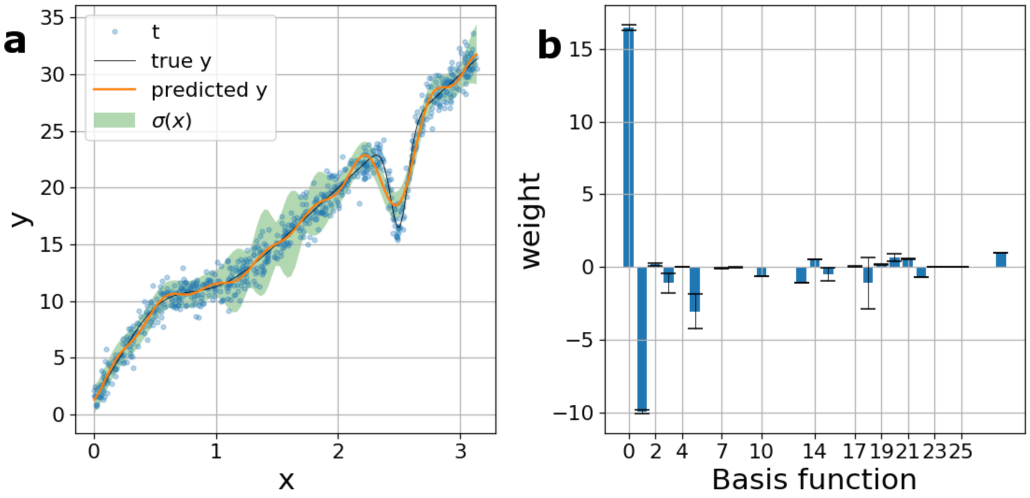


Figure 4.2: Observed data and models in a) and model weights in b). Part a) $y_{\text{true}}(\mathbf{x})$ (black line) as in eq. 4.14 is represented as t shown by blue dots. $t \sim \mathcal{N}\left(y, \beta^{-1} = 1\right)$. $y_{\text{predict}}(\mathbf{x})$ (orange line) and $\sigma(\mathbf{x})$ (green shaded area) are related to to the mean and variance of $p\left(\mathbf{t}^* | \mathbf{x}^*, \mathbf{X}, \mathbf{t}, \mathbf{\Sigma}_R, \beta\right)$ respectively as obtained by regressing with an RVM based on a $\{\cos(kx)\}_{k \in [0,30]}$ basis set. Part b) weights of basis functions vs the number of basis function. The black lines at the end of the blue bars represent $\text{var}(w_i)$.

As it turns out RVMs can be considered as a special case of the versatile EM-algorithm [24]. An application of the EM-algorithm will be demonstrated in section 4.3.4 discussing its application to the density estimation. Furthermore, RVMs and linear Bayesian models in general can be re-formulated in terms of a kernel based approach via the *equivalent kernel* as will be discussed in the following section 4.3.3, since it highlights the connection with Gaussian Processes [179]. Gaussian Processes have recently been shown to be useful in the development of empirical potentials [14]. The kernel-based approach, also referred to as non-parametric models, is a complementary view of regression and is included for completeness. Note that, although quite appealing due to their elegance, RVMs may not necessarily be the best algorithm for regression in high dimensional parameter space [112].

4.3.3 Non-parametric Models

Non-parametric models are based on kernel matrices. Non-parametric methods are an interesting alternative to the above because they do not require the specification of weights for a linear model and can thus be seen as complementary to the parametric models. But the choice of kernels for the construction of kernel matrices comes with its own difficulties and limitations.

Previously we found $\mathbf{w}_{\mathrm{MAP}}$ as a function of $\mathbf{\Phi}$, $\mathbf{\Sigma}_R$ and $\mathbf{y}$, eq. 4.18. Substituting $\mathbf{w}_{\mathrm{MAP}}$ into the linear model we find the *equivalent kernel*

$$\begin{aligned} y &= \boldsymbol{\phi}^T(\mathbf{x}) \cdot \mathbf{w}_{\mathrm{MAP}} = \boldsymbol{\phi}^T(\mathbf{x}) \cdot \left(\mathbf{\Sigma}_R + \mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \mathbf{y} \\ &= \sum_n \underbrace{\boldsymbol{\phi}^T(\mathbf{x}) \cdot \left(\mathbf{\Sigma}_R + \mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1} \boldsymbol{\phi}(\mathbf{x}_n)}_{\text{equivalent kernel(ek)}} y_n \\ &= \sum_n \mathrm{ek}(\mathbf{x}_n, \mathbf{x}) \cdot y_n \end{aligned} \tag{4.24}$$

with $\mathrm{ek}(\mathbf{x}, \mathbf{x}_n)$ the equivalent kernel allowing the estimation of y completely without model weights $\mathbf{w}$. But due to the presence of $\mathbf{\Phi}$ in $\mathrm{ek}(\mathbf{x}, \mathbf{x}_n)$ the finite basis set is still represented. However, there is an alternative to using a finite basis set. Making use of the *representer theorem* one can reformulate the linear model [184] as

$$y(\mathbf{x}) = \sum_i \alpha_i k(\mathbf{X}, \mathbf{x}). \tag{4.25}$$

Minimizing $J(\alpha) = \|\mathbf{y} - \sum_i \alpha_i k(\mathbf{X}, \mathbf{x}_i)\|^2 + \lambda \|\sum_i \alpha_i k(\mathbf{X}, \mathbf{x}_i)\|^2$ by solving $\partial_\alpha J(\alpha) \overset{!}{=} 0$ leads to $\alpha_i = (k(\mathbf{X}, \mathbf{X}) + \lambda \boldsymbol{I}_N)^{-1} y_i$, with $\mathbf{K} \equiv k(\mathbf{X}, \mathbf{X})$ also referred to as 'kernel matrix' or Gram matrix. Thus, the definition of a kernel enables one to circumvent explicit parameterization with a weights vector. Indeed it can be shown that kernels are equivalent to infinite size basis sets [145]. Regression based on this approach is also referred to as *Kernel Ridge Regression*.

The form of eq. 4.25 can also be obtained from finite basis models by rewriting $\boldsymbol{w}_{MP}$ using the matrix inversion identity[1]:

$$\begin{aligned} \mathbf{w} &= \left(\mathbf{\Phi}^T\mathbf{\Phi} + \mathbf{\Sigma}_R\right)^{-1} \mathbf{\Phi}^T \mathbf{y} \\ &= \mathbf{\Sigma}_R^{-1} \mathbf{\Phi}^T \left(\mathbf{\Phi} \mathbf{\Sigma}_R^{-1} \mathbf{\Phi}^T + \mathbf{I}_N\right)^{-1} \mathbf{y} \end{aligned} \tag{4.26}$$

with $\boldsymbol{\alpha} = \left(\mathbf{\Phi}\mathbf{\Sigma}_R^{-1}\mathbf{\Phi}^T + \mathbf{I}_N\right)^{-1} \boldsymbol{y}$ and $k(\mathbf{x}, \mathbf{X}) = \boldsymbol{\phi}(\mathbf{x}) \mathbf{\Sigma}_R^{-1} \mathbf{\Phi}^T$. From this follows that $k(\mathbf{x}, \mathbf{x}') = \boldsymbol{\phi}^T(\mathbf{x}) \mathbf{\Sigma}_R^{-1} \boldsymbol{\phi}(\mathbf{x}')$. Thus, kernels are symmetric.

In the simplest case $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \cdot \mathbf{x}'$ is called the linear kernel [24] with a symmetric and positive semidefinite kernel matrix $\mathbf{K}$. An interesting feature of the non-parametric approach is that kernel functions found to be valid can be combined with each other to construct new

---

[1] Matrix inversion identity $\left(A + UCU^T\right)^{-1} UC = A^{-1} U \left(U^T A^{-1} U + C^{-1}\right)^{-1}$.

valid kernel functions.

The non-parametric based approach can also be viewed from a Bayesian point of view, similar to linear regression with distributions over weights and the hyper parameters. An early example of a kernel based method is the Nadaraya-Watson estimator [159, 257] where $y = \mathbb{E}[T|X=x] = \int p(t|x)\,dt = \int \frac{p(t,x)t}{p(x)}dt$ is calculated explicitly, representing kernel density estimation.

Another example of Kernel based models are *Gaussian Processes* (GPs), where the comparison of previous observations with new observations (indicated with *) is performed using $p(\mathbf{y}^*|\mathbf{X}^*,\mathbf{t},\mathbf{X}) = \int p(\mathbf{y}^*|\mathbf{w},\mathbf{X}^*)\,p(\mathbf{w}|\mathbf{X},\mathbf{t})\,d\mathbf{w}$ interpreting the observed data $\mathbf{t}$ as a realization of the Gaussian Process [184], instead of relying on $\mathbf{w}$. Gaussian processes are *stochastic processes* which are defined by any subset of observed $y(\mathbf{x}_0)$, $y(\mathbf{x}_1)$, ..., $y(\mathbf{x}_N)$ being specified by a joint distribution. In the case of the Gaussian process this distribution is then a multivariate normal distribution. Particular to the Gaussian Process is that it is completely specified by its mean and covariance with the mean

$$\mathbb{E}[y(x)] = \mathbf{\Phi}\mathbb{E}_{\mathbf{w}}[\mathbf{w}] = 0 \tag{4.27}$$

and covariance as

$$\begin{aligned} \operatorname{cov}[y(x), y(x')] &= \operatorname{cov}_{\mathbf{w}}\left[\boldsymbol{\phi}(\mathbf{x})\mathbf{w}, (\boldsymbol{\phi}(\mathbf{x}')\mathbf{w})^T\right] \\ &= k(\mathbf{x},\mathbf{x}') \end{aligned} \tag{4.28}$$

or equivalently $\operatorname{cov}(\mathbf{y},\mathbf{y}) = \mathbf{\Phi}\mathbf{\Sigma}_R\mathbf{\Phi}^T = \mathbf{K}$.

However, the downside of non-parametric approaches is that they need to memorize all previous observations, at least to some extent to cover the range for prediction. Because the training data is ‘memorized’ the non-parametric models are also referred to as *memory-based* models since they require the storage of past observations to validate new observations. This memorization in form of the Gram-matrix leads to the kernel based algorithms having a storage complexity of $\mathcal{O}(N^2)$ making it difficult to store for large databases and computationally infeasible to invert since the inversion method of choice, the Cholesky decomposition, scales with $\mathcal{O}(N^3)$. An approximation for this problem are Bayesian Committee Machines [242] which produce a committee of kernels containing different parts of the training set and predicting via averaging over the committee. For a range of other approaches see [184].

An illustration of regression on our example data using Gaussian Process Regression with a radial basis function and white noise kernel is shown in fig. 4.3. We can see that compared to RVMs, as shown in fig. 4.2 a), the prediction is closer to the true function and the predicted uncertainties seem to fluctuate less on average. But like fig. 4.2 a) the prediction also shows a wavy behaviour, which is due to the chosen radial basis function kernel.

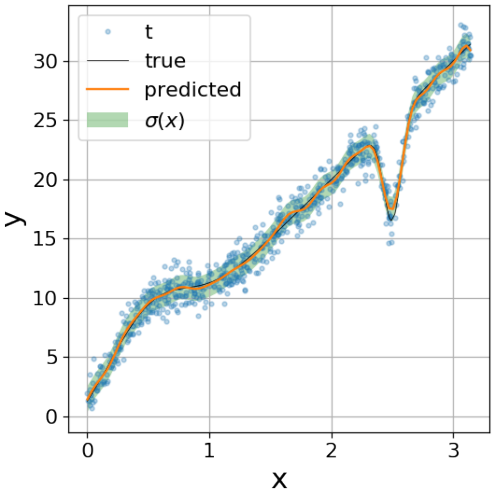


Figure 4.3: $y_{\text{true}}(\mathbf{x})$ (black line) from eq. 4.14 is represented as t shown by blue dots. $t \sim \mathcal{N}\left(y, \beta^{-1} = 1\right)$. $y_{\text{predict}}(\mathbf{x})$ (orange line) and $\sigma(\mathbf{x})$ (green shaded area) are related to the mean and variance of $p(\mathbf{t}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{y}, \boldsymbol{\Sigma}_R, \beta)$ respectively as obtained by regressing with a Gaussian Process and a radial basis function and white noise kernel (orange) The green shaded area is within 1 standard deviation of the fit.

4.3.4 Density Estimation & The EM-Algorithm

Contrary to the electron density regression which will be described in Chapter 7, where the electron density is known as a function of the spatial coordinate, density estimation here regresses the distributions of data points themselves. These distributions may not be directly available but can be thought of as ‘hidden’ in the data itself and therefore be approximated, via splines or analytical distributions such as Gaussian Mixture Models (GMMs). We discuss GMMs because they can be used to approximate data distributions using the Expectation Maximization (EM) algorithm and allow the setup of generative probabilistic classification, as will be discussed in section 4.4. We will make use of this property of EM-GMMs in Chapter 5 where we will discuss the classification of individual atoms based on their surroundings.

The Expectation-Maximization (EM) algorithm is an iterative approach to solve otherwise intractable parameter estimation problems. It proceeds by calculating the expected values of some hidden/latent variable Z, such as a class affiliation or model parameters, and then maximizes the log likelihood of $p(X,Z|\theta)$ by varying θ. The EM-algorithm, in special cases, can even be applied to stochastic non-linear dynamical systems [78]. Generally, the EM-algorithm is related to the MM-algorithm, which itself allows for the construction of optimization algorithms. The MM depends on the shape of the function, which is either to be maximized and then represents Minorize Maximization, or is to be minimized and then represents Majorize Minimization [110]. The EM-algorithm requires that the posterior $p(Z|X,\theta)$ and $\log p(X,Z|\theta)$ can be treated analytically. This is useful because in some cases it is easier to optimize $p(X,Z|\theta)$ than $p(X|\theta)$. For the derivation see appendix A.

Here we will demonstrate density estimation on the example shown in fig. 4.4. There we see clusters of coloured dots. Each dot was sampled from one of the two underlying multivariate normal distributions, selecting each distribution with its own probability. The colouring indicates the original affiliation. The task for the density estimation is to approximate this distribution using analytical models, with only the data points at hand and the information about the number of clusters.

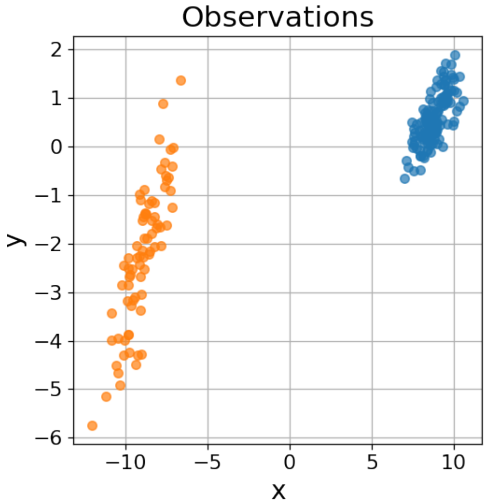


Figure 4.4: Data distributed according to two different multivariate normal distributions, indicated by the coloring.

Let X, θ and Z be our observed variables, the model parameters and the latent variables respectively. The latent variable then indicates the cluster affiliation. For the EM-algorithm the idea is to split $\ln p(X|\theta)$ into two parts, the lower bound $\mathscr{L}(q,\theta)$, which depends on some approximation distribution q to $p(Z|X,\theta)$, and the Kullback-Leibler divergence $KL(q||p)$. The EM-algorithm works by iteratively evaluating Z given θ and X (*E-step*) and then evaluating θ given the new Z (*M-step*).

Gaussian Mixture Models

GMMs combine multiple normal distributions with individual mean and covariance $p(X|Z=i,\boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$ weighting with the respective priors $p(Z=i)=\pi$ as

$$p\left(\mathbf{x}\,\middle|\,\{\boldsymbol{\mu}_i\}_{i\in\mathbb{N}},\{\boldsymbol{\Sigma}_i\}_{i\in\mathbb{N}},\{\pi_i\}_{i\in\mathbb{N}}\right) = \sum_{i=0}^{M} p(Z=i)\,p(X|Z=i,\boldsymbol{\theta}) = \sum_{i=0}^{M} \pi_i\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i), \tag{4.29}$$

with $\boldsymbol{\theta}$ as shorthand for all M Gaussian with $\{\boldsymbol{\mu}_i\}_{i\in M}$ and $\{\boldsymbol{\Sigma}_i\}_{i\in M}$.

In order to carry out the E-step we require $p(Z=i|\mathbf{x},\boldsymbol{\theta})$, which can be found applying Bayes Theorem $p(Z|\mathbf{x},\boldsymbol{\theta}) = \frac{p(\mathbf{x}|\boldsymbol{\theta},Z)p(Z)}{\sum_z p(\mathbf{x}|\boldsymbol{\theta},Z=z)p(Z=z)}$. Thus we find $p(Z=i|\mathbf{x},\boldsymbol{\theta})$ for GMMs as

$$p(Z=i|\mathbf{x}_n,\boldsymbol{\theta}) = \frac{p(\mathbf{x}_n|Z=i,\theta)\,p(Z=i)}{\sum_j p(\mathbf{x}_n|Z=j,\theta)\,p(Z=j)} = \frac{\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\,\pi_i}{\sum_j \mathcal{N}\left(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j\right)\pi_j} \equiv \gamma(z_{ni}). \tag{4.30}$$

In the context of density estimation $\gamma(z_{ni})$ is also referred to as the *responsibility*, in the sense that normal distribution i is responsible for data point $\mathbf{x}_n$ with $p(Z=i|\mathbf{x}_n,\boldsymbol{\theta})$. For the M-step on the other hand we require updated rules for $\{\pi_i\}_{i\in M}$, $\{\boldsymbol{\mu}_i\}_{i\in M}$ and $\{\boldsymbol{\Sigma}_i\}_{i\in M}$. These can be found solving $\partial_{\pi_i,\lambda}\ln p(\mathbf{X}|\boldsymbol{\theta}) - \lambda\left(\sum_j^M \pi_j - 1\right) \overset{!}{=} 0$, $\partial_{\boldsymbol{\mu}_k}\ln p(\mathbf{X}|\boldsymbol{\theta}) \overset{!}{=} 0$ and $\partial_{\boldsymbol{\Sigma}_k}\ln p(\mathbf{X}|\boldsymbol{\theta}) \overset{!}{=} 0$ respectively leading to

$$\boldsymbol{\mu}_i = \frac{1}{\sum_n^N \gamma(z_{ni})}\sum_n^N \mathbf{x}_n\gamma(z_{ni}), \tag{4.31}$$

$$\boldsymbol{\Sigma}_i = \frac{1}{\sum_n^N \gamma(z_{ni})}\sum_n^N \gamma(z_{ni})(\mathbf{x}_n-\boldsymbol{\mu}_i)(\mathbf{x}_n-\boldsymbol{\mu}_i)^T \tag{4.32}$$

and

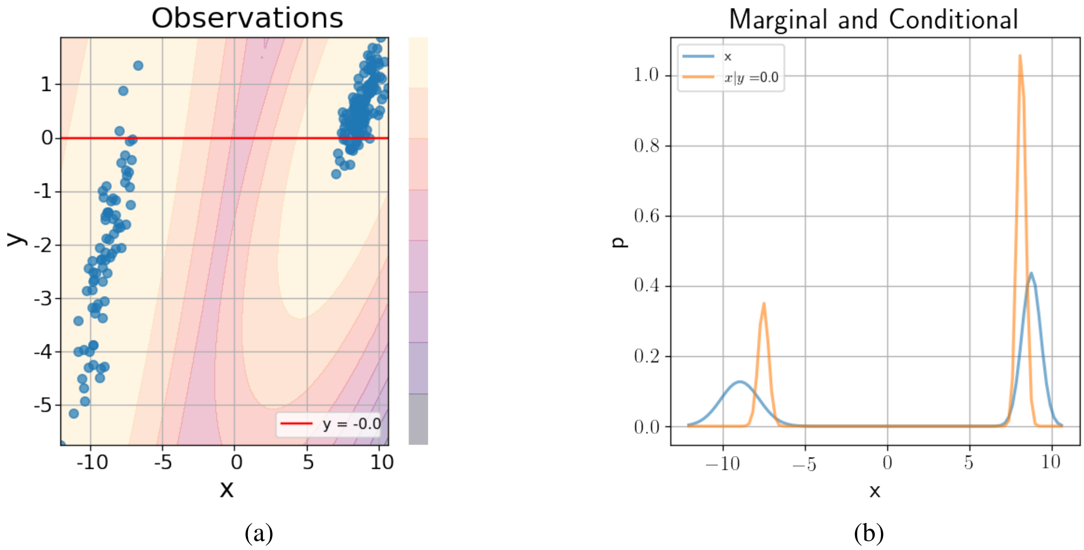
$$\pi_i = \frac{\sum_n^N \gamma(z_{ni})}{N}. \tag{4.33}$$

For the EM-algorithm with GMMs see appendix A.

In our current example we have N observations $\mathbf{x}_n$ each representing a tuple $\mathbf{x}_n = (x,y)$. Applying the EM-algorithm for GMMs for our example we find distributions. The corresponding $\ln p(X|\theta)$ is shown as contours in fig. 4.5 a) alongside the original data points. For this example we are interested in approximating the probability density function in order to analyze the observed data, i.e. the conditional distribution of x given some value $y=a$,

$p(x|y=a)$, and the marginal distribution of $p(x)$.

Conditioning on for example $y=0$ equals normalizing the cross section shown as the red line in fig. 4.5 a). The corresponding conditional distribution $p(x|y=0)$ is shown in orange in fig. 4.5 b). The marginal distribution for x can be obtained by integrating over the sum of distributions. For the derivation see appendix A.



(a)

(b)

Figure 4.5: Estimating the density distribution for given data points. a) The original distribution of observations (dots) and fitted $\log p(x,y)$ (shaded areas). The red line in a) is a cross section for a given y value. The corresponding conditional $p(x|y)$ and marginal $p(x)$ are shown in b).

4.4 Classification

The distribution $p(\mathbf{x},t)$ can be written as $p(x_0|x_1,\ldots x_M,t)\,p(x_1|x_2,x_3,\ldots x_M,t)\ldots p(x_M|t)\,p(t)$. Note that in this expansion all factors, except $p(t)$, are conditioned on t and usually at least one feature x. Capturing these feature-feature dependencies quickly becomes intractable. Hence one simplification is to neglect these dependencies, which is also referred to as *naive Bayes,* leading to $p(\mathbf{x},t)=p(t)\prod_i p(x_i|t)$. The full classification process is then

$$t=\text{argmax}_{t\in\text{classes}}\,p(t|\mathbf{x})\,. \tag{4.34}$$

Using the Bayes theorem and assuming uniform probability of all classes we find

$$t = \mathrm{argmax}_{t\in\mathrm{classes}} p\left(\mathbf{x}|t\right). \tag{4.35}$$

Now with naive Bayes one has

$$t = \mathrm{argmax}_{t\in\mathrm{classes}} \prod_i p\left(x_i|t\right). \tag{4.36}$$

The remaining conditional distributions themselves can then be approximated analytically or via discretization, for example using splines. Both are also possible for non-naive Bayes but may quickly become computationally expensive with increases in the dimensionality of the feature space. The discretized models are affected by the bias-variance trade-off in terms of the number of size of the discrete intervals [267].

In order to illustrate the difference that the independence assumptions in distributions can make, which will then influence the classifier itself, we will discuss a small example of data samples from a multivariate normal distribution with a two dimensional feature space with features named x_0 and x_1. The independence assumption of the features can be expressed as $p\left(x_0,x_1\right) = p\left(x_0\right)p\left(x_1\right)$. The sampled points are shown as crosses in fig. 4.6. For the independence assumption the two marginal distributions are regressed with normal distributions and then used to compute $p\left(x_0,x_1\right)$. The model without the independence assumption was regressed using a full multivariate normal distribution, which was then used to compute the marginals $p\left(x_0\right)$ and $p\left(x_1\right)$. The joint distributions and marginal distributions for both assumptions are shown in fig. 4.6, with the marginals along the axes and the joint distribution in the centre as contour plots. The contour plot with the shaded areas corresponds to assuming feature dependence whereas the lines correspond to assuming independence. Although the marginals (shown as log plots) seem relatively similar, the joint distributions show significant difference, with the distribution assuming independence forming circles.

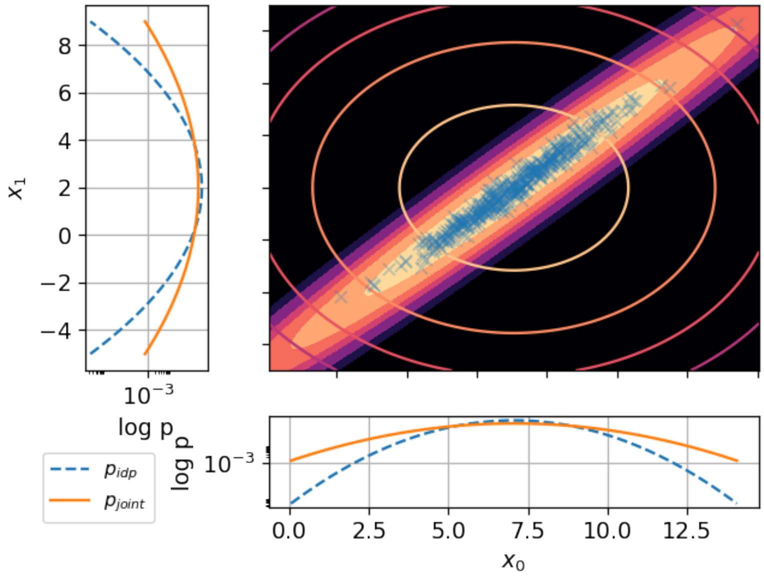


Figure 4.6: Comparing the influence of the independence assumption $p(x,y) = p(x)\,p(y)$ approximating the data (crosses). Left and bottom showing $\log p(y)$ for both the independence assumption (p_{ipd}) and the dependence assumption (p_{joint}). The center of the plot represents the $p(x,y)$ where the shaded areas correspond to assuming dependence and the lines for the independence assumption. Both contour plots were created using the same $\log p$ levels.

One aspect of the discussion of classifiers are the ‘decision boundaries’, which are surfaces partitioning feature space. An example for such a decision boundary is shown in red in fig. 4.7. In this example two multivariate normal distributions were used to generate samples for the respective classes. EM-GMM was used here to obtain all the model parameters, just like the development of classifiers in Chapter 5. An interesting aspect of this approach is that the labels were completely unknown to the algorithm making it only a semi-supervised learning algorithm because it knew about the number of classes. In fig. 4.7 we also see shaded areas which indicate the probability difference $\Delta p = p(X|\text{class} = 1) - p(X|\text{class} = 2)$. As indicated before one may be interested in minimizing risk as well by preventing classification in areas with high uncertainty, indicated with ‘x’ hatches in fig. 4.7.

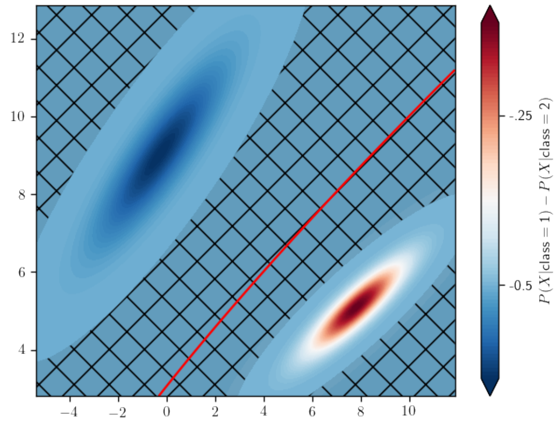


Figure 4.7: Illustration of the decision boundaries (red) induced by the two normal distributions for each class from the density estimation example. Areas are shaded according to $p(X|\text{class}=1) - p(X|\text{class}=2)$. Each of the classes is represented by a multivariate normal distribution obtained with the EM-Algorithm from given data distributions.

Chapter 5

Phase Transformations in Single Crystals

The aim of this and the following chapter is to study the beginning of the $\gamma \rightarrow \gamma + \gamma'$ phase transformation in Ni-based superalloys from an initially supersaturated disordered host (γ-phase) focusing on single crystals and then the influence of grain boundaries [200]. For high levels of supersaturation this phase transformation has been found to take place rapidly, even forming γ' during quenching down from super solvus temperatures, as shown for example, by Tan *et al.* [228] for Ni-Al-Cr alloys using a 3D atom probe. In other cases with slow cooling through γ' solvus temperatures it was found that for a range of alloys, such as RR1000, so-called *edge serration* of grain boundaries can take place [137, 156, 176]. Both the phase transformation and the edge serration of grain boundaries are of technical significance affecting the thermo-mechanical properties of blades and disks in turbine engines.

A common sequence for disorder $\rightarrow$ order transitions, like γ' formation is the following: initial chemical ordering followed by, or concomitantly with, chemical decomposition either in the form of nucleation and growth or spinodal decomposition [224, 228, 229, 248]. This agrees with computational investigations summarized by the graphical method proposed by Soffa et al. [215]. A detailed understanding of this transition and its dependency on the chemical and thermal environment requires studying the material at the atomistic level. Because γ' ordering depends on vacancy defect jumps, which are potentially much faster than chemical decomposition, we thought it possible to see ordering on the atomistic scale during short annealing times using Molecular Dynamics (MD) simulations (~100 ns) with currently available computational resources and codes which scale well over multiple CPUs. To understand chemical ordering on the atomistic scale we need to be not only able to run the simulations but also analyze their configurations and how they are changing over time. Therefore, a key part is to use *features*, as introduced in Chapter 4, to describe local atomic arrangements. A wide variety of these features have been developed by the computational physics community and a brief overview will be given in section 5.1. Building on Steinhardt / bond order parameters in particular and Bayesian classification we show results of MD simulations in section 5.2 of annealed supersaturated Ni-Al single crystals for up to 100 ns using an EAM potential. The simulated single crystals were found to exhibit γ' ordering and clustering, the evolution of which can be reasonably described using Becker-Doehring theory. We also found a positive correlation of the c_{Al} and $c_{\gamma'}$ concentration profiles, possibly indicating the onset of a chemical decomposition. Furthermore we will show in Chapter 6 that MD simulations of bi-crystals containing coincidence site lattice grain boundaries (CSL GBs) with varying solute content exhibit an interplay between γ' ordering and GB motion. There we observed that the presence of Al solutes does alter the GB behaviour changing the frequency of transitions between 'locked' and 'running' states, which is also observed by Deng *et al.* and Ulomek *et al.* for single element systems using artificial driving forces [51, 244]. We also find that the volume in the vicinity of the GBs shows accelerated γ' ordering with respect to the remaining volume and relative to single crystals.

5.1 Atom Configurations - Features & Classification

The classification of atomic structures does require a classifier, induced by some training set, but also an appropriate choice of features to capture local neighbourhoods of atoms. Over the last few decades a variety of inventive approaches have been made for extracting as much of the necessary information as possible from the 3D atom configurations, while minimizing the computational cost to allow for scaling up to hundreds of thousands of atoms. The approaches found in the literature during this work are based on one or more of the following: root mean square distances (RMSDs) between structures, spherical harmonics, graph theory, discrete distributions of distances and angles and, more recently, kernels. In a recent review by Stukowski [223] a range of features were discussed for the classification of atoms, among them centro symmetry parameters (CSP), bond order or Steinhardt parameters (BOPs), common neighbour analysis (CNA) and Voronoi tessellation. During the discussion of these parameters Stukowski also introduced two additional analyses: the adaptive CNA (ACNA) and neighbour distance analysis (NDA). A selection of these parameters is implemented in Atomeye and Ovito [138, 222].

5.1.1 Feature Overview

Going through the above-mentioned features, roughly in chronological order, we start with *BOP*, which were originally suggested for classification features by Steinhardt *et al.* [219]. BOPs describe the distribution of geometrical bonds, not to be confused with chemical bonds. They represent the profile of neighbouring atoms around a chosen atom on a unit sphere via spherical harmonics, which depend on θ and ϕ of the spherical coordinate system. BOPs depend on the integer parameter l which for $l = 4$ and 6 identifies cubic and hexagonal symmetries particularly well. BOPs were originally used to distinguish between fcc, bcc, hcp and icosaheders in liquid to solid transition simulations using a Lennard-Jones potential [219]. By modifying the BOPs used during this work we found features capable of capturing the γ' ordering transition. BOPs themselves will be discussed in some detail in section 5.1.2.

CNA parameters were developed from graph theoretical considerations by Honeycutt *et al.* [104], studying the melting and freezing of Lennard-Jones clusters. CNA parameters are index based and had originally four indices $ijkl$ with: i indicating whether or not two atoms are nearest neighbours or not, j the number of common neighbours of two atoms, k the number of bonds between the common neighbours and l some further detail to allow additional distinction. The version of CNA parameters used by Faken *et al.* [59] dropped the i and specified l to indicate the longest continuous chain of k bonds. Faken *et al.* also showed how CNA parameters can be used in conjunction with element specific radial distribution functions to compute further values such as effective numbers of atoms in 1st, 2nd and high order shells. CNA parameters have since then been used in studies with EAM, such as in the melting of Au [44], and have been modified for ordered structures by Luemmen et al. [141] and extended to cope with multiple different phases as ACNA [223].

Considering fcc structures only, Kelchner *et al.* [117] developed *CSP*s to study defect structures due to the indentation of an Au (111) surface. They defining CSPs as $P = \sum_{i=1}^{6} |\boldsymbol{R}_i + \boldsymbol{R}_{i+6}|^2$ with $\boldsymbol{R}_i$ and $\boldsymbol{R}_{i+6}$ being the 6 vector pairs of the 12 closest atoms. Combining CSPs with CNA parameters Tsuzuki *et al.* [243] developed a hybrid parameter they called *CNP*.

Voronoi tessellations have also found their way into atom features since they were first introduced in 1908 by Voronoi [249]. They are of interest because they allow the partitioning of space using atom centres, e.g. Wigner-Seitz cells are results of Voronoi tesselations. Recently Schablitzki *et al.* [198] introduced a ‘topological fingerprint’ designed to distinguish topologically close packed phases which consist of characteristic coordination polyhedra.

These coordination polyhedra form duals of Voronoi cells and were used jointly with information about possible n order jumps (jumps of length n from the current atom to its neighbours and back again) and the element and coordination polyhedra information of atoms jumped over. Larsen *et al.* [129] also used Voronoi cells to sort neighbouring atoms by distance to construct a convex hull. They compared these with a database including $L1_2$ ordered structures.

A possibly simpler but useful approach was taken by Ackland *et al.* [1] who built a classifier upon discrete distributions of bond distances r and bond angles θ allowing them to distinguish between fcc, bcc and hcp lattice types.

A more direct way of relating *RMSD*s to features of the classification process was suggested by Stukowski *et al.* [223] using the *NDA*. In the NDA the distances of an atom to its neighbours are computed, sorted and checked to determine whether or not they lie somewhere within the expected intervals for structures in the provided database. Also using RMSDs Sadeghi *et al.* [194] and Zhu *et al.* [276] suggested metrics by making use of Gaussian type orbitals computing a *kernel* or overlap matrix. They then used the corresponding sorted eigenvectors to compute distances to reference structures. The computational cost for this is the rotation of the structure and the permutation of the atom indices. A similar approach which does not require rotation and permutation was taken by Bartok *et al.* [16] who introduced the smooth overlap of atomic positions (SOAP) approach using a field $\rho(\boldsymbol{r})$ to represent an atomic configuration. These fields are then used for a similarity measure of pairs of structures $S(\rho,\rho') = \int \rho(\boldsymbol{r})\rho'(\boldsymbol{r})\,d\boldsymbol{r}$. Ferre *et al.* [63] extended this approach by directly considering these fields as probability fields introducing a distance metric, in the mathematical sense, $\|\rho_1 - \rho_2\|^2 = S(\rho_1,\rho_1) - 2S(\rho_1,\rho_2) + S(\rho_2,\rho_2)$.

5.1.2 Steinhardt / Bond Order Parameters

BOPs were originally used for atom classification by Steinhardt *et al.* [219]. Using spherical harmonics they are a 3D extension of the 2D ψ_6 parameter, $\psi_6 \equiv \left|\frac{1}{M}\sum_{m=1}^{M}\exp(6i\theta_m)\right|$, as for example used by Schwanzer *et al.* [206] to study agglomeration of particles into clusters, with θ_m the angle between the bond of the current particle and its mth neighbour particle (out of M neighbours) and a reference frame.

BOPs have the advantage that they are translation and rotation independent and one set of them is defined as

$$q_l \equiv \sqrt{\frac{2\pi}{2l+1}\sum_{m=-l}^{l}\sum_{n=1}^{N}\left|Y_m^l(\theta_n,\phi_n)\right|^2} \in [0,1], \tag{5.1}$$

where $Y_m^l(\theta,\phi)$ are spherical harmonic functions with $l \in \mathbb{N}$, $m \in [-l,l]$, the azimuthal $\theta_n \in [0,2\pi]$ and polar angle $\phi_n \in [0,\pi]$. Although information is 'thrown away' with rotational and translational invariant parameters this invariance property is important since the orientation of crystals relative to the simulation box or reference frame cannot be guaranteed to be identical for all simulations. The distributions of the q_l parameters depend on the number of neighbouring atoms considered, i.e. the characteristic distributions or q_l values commonly considered are for the 12 nearest neighbours. For the purpose of this work we consider a fixed number $N = 18$ of neighbouring atoms since 18 neighbours for fcc includes the first and second nearest neighbours. Including the first two fcc shells was found to be useful for distinguishing disordered from $L1_2$ ordered configurations.

BOPs lead to characteristic distributions for fcc, hcp, bcc and icosahedras using different l values, where, in particular, for $l = 4$ and 6 identify cubic and hexagonal structures. Another BOP $\hat{w}_l$ is not used in the present work due to its computational cost in naive implementations which is greater than for q_l and was found prohibitive for more than 100,000 atoms. The q_l distributions themselves are temperature dependent and broaden with increasing temperature around their original delta distribution-like centres. This broadening behaviour will be taken advantage of in section 5.2 using splines and in Chapter 6 with a mixture of Gaussians. This property of characteristic temperature dependent distributions for fcc, bcc and other phases was also used by Auer et al. [4] to decompose distributions of phase mixture systems with distributions obtained for single phases.

The definition of q_l in eq. 5.1 considers each geometric bond equally and applies uniform weights. Non-uniform weights are an interesting option particularly for highly distorted structures but have not been tested in this work. An example of non-uniformly weighted bonds is the work by Dietz *et al.* [52] who used Voronoi face areas as the weights for individual bonds. However, we do modify q_l by considered three 'filters': 'all', 'like' and 'unlike' to represent the same neighbourhood of an atom in three different ways in order to achieve chemical sensitivity. In the case of an atom in a neighbourhood of the same element we find $q_{l,\text{all}} = q_{l,\text{like}}$ whereas $q_{l,\text{unlike}} = 0$. For the other extreme case of an atom that is embedded in atoms with only unlike elements we find $q_{l,\text{all}} = q_{l,\text{unlike}}$ whereas $q_{l,\text{like}} = 0$. Furthermore, we multiply q_l by 10 to ensure numerical stability of the classifiers.

Although it was found that additional averaging of q_l, once already computed over the local neighbourhood, can be advantageous for bunching classes of atoms spatially this also leads to a coarsening of the visible q_l profile [133]. Therefore, in the following approach we will not apply this kind of averaging. Note that this singular averaging could also be interpreted in an iterative Bayesian framework to update the classification of atoms having a similar effect.

A common critique of BOPs is that they are computationally too expensive. However, according to a recent study by Winczewski *et al.* [262] this is not necessarily true. With their approach they even find the computational cost of BOPs to be of the same order as CSP, CNA parameters and CNPs.

An alternative to explicitly choosing features in the way we have done here would be to use deep neural nets, as carried out recently for image classification using the famous AlexNet [126]. Deep neural nets have been found to be able to generate features by themselves. However, this may require larger computational resources for the training and a large labelled data set and would require retraining when adding a phase. A further alternative would be the recently developed Generative Adversarial Neural Nets (GANs) which are trained on smaller training sets but generate data and labelling as part of the process [82].

5.1.3 Classifying atomic local structures

Hard coded or manually tuned classifiers seem to be a common way of studying atom configurations in contrast to the approach taken in this work. One other work which made use of a probabilistic framework for atom configurations using BOPs is by Gades *et al.* [73]. We here use eq. 4.35 from section 4.4 to classify phases according to

$$\text{Phase} = \text{argmax}_{\text{Phase} \in \text{Phases}} p\left(\mathbf{x} | \text{Phase}\right) \tag{5.2}$$

Note that the full Bayesian treatment finds the ‘Phase’ value for a given atom by maximizing $p\left(\text{Phase}|\mathbf{x}\right)$, with $\mathbf{x}$ representing the feature vector for the current atom. However, for the purpose of classification one can neglect $p\left(\mathbf{x}\right)$ and we do assume that all phases are equally probable, i.e. we assume a uniform prior over $p\left(\text{Phase}\right)$. The uniform prior is reasonable when it is not known *a priori* which phases should be more likely in general. In the following we will separate $\mathbf{x}$ into parts related to BOPs with ‘all’, ‘like’ and ‘unlike’ filters and generate probability density functions (PDFs) for each filter. We will use these different filters to classify each atom for chemical environment separately and subsequently interpret the joint result in terms of a phase in both section 5.2 and Chapter 6. This separation into PDFs for chemically insensitive $q_{l,\mathrm{all}}$ on the one hand and chemically sensitive $q_{l,\mathrm{like}}$ and $q_{l,\mathrm{unlike}}$ on the other was done to allow for abstraction of general lattice symmetries, with $q_{l,\mathrm{like}}$ and $q_{l,\mathrm{unlike}}$ further specifying the phase based on the symmetry of the arrangement of neighbouring elements. Separating into separate PDFs for $q_{l,\mathrm{like}}$ and $q_{l,\mathrm{unlike}}$ was mainly done to a first approximation to decrease computational cost by decreasing the dimensionality of that feature space and for interpretability of the resulting PDFs. Though the combination of both 'like' and 'unlike' features would be too expensive for spline-based approaches it could potentially be very useful for functional approximations to the PDFs for example with mixtures of Gaussians. An example for $q_{l,\mathrm{all}}$ PDFs is shown in fig. 5.1, illustrating PDFs for fcc, bcc and hcp configurations with $q_{4,\mathrm{all}}$ and $q_{6,\mathrm{all}}$ as the x- and y-axes.

We found that constructing a classifier this way by splitting into 'all', 'like' and 'unlike' is particularly useful for distinguishing chemically ordered from disordered arrangements. Sections 5.2.1.1 and 6.1.1 will go into some detail about how these PDFs are obtained using splines and Gaussian mixtures respectively. Although there are various classifiers available which could also potentially be used, as indicated at in section 4.4, a probabilistic approach is taken here.

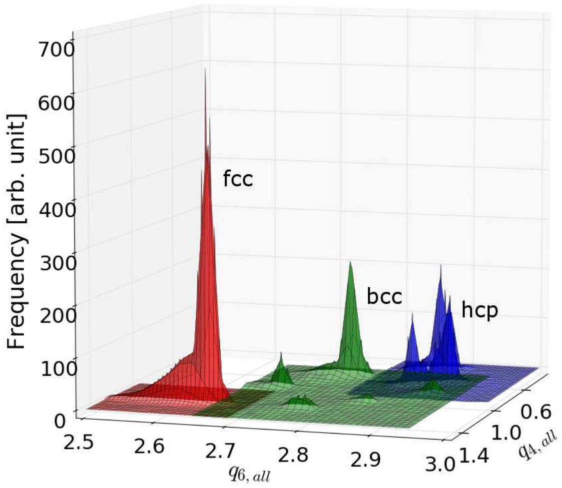


Figure 5.1: Conditional probability density functions for the fcc, bcc and hcp structures as a function of the Steinhardt parameters q_l ($l = 4$ and $l = 6$) shown in red, green and blue respectively.

5.2 Early Stage Precipitation in Single Crystals

In this section we will discuss how to create a spline based probabilistic classifier to distinguish the γ from γ' phase, among others. As mentioned we will make use of BOPs evaluating large simulations of 108 000 atoms over long time periods of up to 100 ns to investigate early stages of precipitation as occur during quenching in experiments. Thus temperatures just below the solvus temperatures are chosen for supersaturated compositions of γ to promote precipitation of γ' .

We will show that spline based PDFs can successfully be used to create classifiers robustly identifying γ' even at high temperatures and embedded in γ hosts. Applying the developed classifier to annealing simulations we find that the Al concentration profile, c_{Al}, positively correlates with the γ'-like atom concentration profile, $c_{\gamma'}$. This possibly indicates chemical decomposition since γ' ordering seems locally accelerated particularly in above average c_{Al} regions. The simulations are interpreted as demonstrating *homogeneous ordering* which could be followed by spinodal decomposition or nucleation and growth, where spinodal decomposition would be expected based on experimental observations.

5.2.1 Method

Next, we will describe more precisely how configurations were generated to obtain PDFs from MD simulations as well as artificial distortions, and cross-validate the classifier. The ability of the classifier to distinguish γ' from γ will be demonstrated on a configuration annealed at 1500 K with a γ' sphere embedded in a block of γ Ni-20 at.% Al.

5.2.1.1 Obtaining PDFs

In order to generate distorted configurations of the ideal crystal configurations we used NPT MD simulations and artificial distortions, sampled from a multivariate normal distribution. The latter method was found to be particularly useful due to its low cost in generating configurations, not requiring the steps necessary in the process for MD. For a stable classifier to distinguish γ' from γ we found it sufficient to create one PDF for γ' and one for γ for each filter. A given degree of distortion, here denoted as η, induces a multivariate normal distribution $\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathrm{diag}(\Delta x_0, \Delta x_1, \Delta x_2)$,$\Delta x_i \equiv \eta \cdot a_i$ and a_i the length of the ith cell vector. Artificially distorting a structure this way was found to be sufficient for the present problem because all axes of γ and γ' expand in the same way for both phases respectively. This definition of Δx_i leads to a broadening of the normal distribution as η increases thus causing larger atom displacements, emulating higher temperatures.

Comparing artificial distortions and those obtained with NPT MD we find that $\eta = 10^{-5}$ and 10^{-3} correspond to 0.1 K and 1500 K approximately, as shown in table 5.1. The resulting histograms were then approximated using splines. Using $\eta = 10^{-5}$, 10^{-4} and 10^{-3} to emulate a broad temperature range for a given phase we decided to obtain three PDFs ($N = 3$) and combine them in a linear superposition, as in eq. 5.3,

$$p\left(\mathbf{x}|\text{Phase}\right) = \sum_{i=1}^{N} w_{\eta_i} p\left(\mathbf{x}|\text{Phase}, \eta_i\right) \tag{5.3}$$

using some weighting w_{η_i} into one PDF for the given phase and filter. Experimenting with a variety of weightings we found an equal contribution leads to a robust classifier for the high temperatures of interest here. The extreme case of using only high distortion PDFs leads to very poor overall accuracy whereas using only low distortion PDFs leads to good accuracy at low temperatures but low accuracy at high temperatures. We found that using either an equal contribution of exponentially distributed distortions as in table 5.1 leads to good PDFs or an exponential weighting, decaying with increasing temperature, when using linearly distributed distortions.

Having obtained the PDFs we interpret an atom classified as γ' or γ when *both* ‘like’ and ‘unlike’ indicate the respective phase as more probable than all the others. Mixed classifications are not considered as either γ or γ'.

artificial		NPT MD	
η	RMSD	T (K)	RMSD
10^{-5}	0.00170	0.1	0.00073
10^{-4}	0.00541	500	0.00699
10^{-3}	0.01688	1500	0.01700

Table 5.1: Comparison of distortions induced artificially or via annealing with NPT MD in a γ' 10x10x10 supercell. The root mean square displacement (RMSD) is averaged over all frames. The RMSD is calculated for fractional coordinates with respect to the supercell.

5.2.1.2 Cross-Validation

In order to cross-validate the classifier we created atomic models embedding a spherical γ' precipitate with a radius of 10 Å in pure fcc Ni and γ Ni-20 at.% Al. These configurations were then artificially distorted with the same η values and used to create the PDFs. To quantify the performance of the classifier classifying into γ' and 'not γ'' the *F1* score was used, a binary classification measure. The F1 score $\equiv \frac{2RP}{R+P}$ is defined via the 'Recall' $R \equiv \frac{t_p}{t_p+f_n}$ and the 'Precision' $P \equiv \frac{t_p}{t_p+f_p}$ therefore taking on values between 0 (everything classified incorrectly) and 1 (everything classified correctly). R and P in turn are based on t_p (true positive), t_n (true negative), f_p (false positive), f_n (false negative). The four latter quantities can be understood in the following way. The first term 'true' or 'false' indicates if the classifier identified the class correctly or not, whereas 'positive' and 'negative' indicates the class. For example, an Al atom sitting on a corner site of the unit cell in γ' classified as γ-like would constitute a 'false negative' because it actually should be classified as γ'-like. An Al atom in a γ environment classified as γ'-like on the other hand would constitute a 'false positive' because it actually should be classified as γ-like. Proceeding this way computing the *F1* score we find that the classifier has an easier time identifying γ' correctly in a pure Ni environment, decaying in quality with increasing distortions, than for the γ Ni-20 at.% Al environment as shown in fig. 5.2. Note that for this cross-validation we manually assigned atoms with γ/γ' labels.

Applying the classifier to the last snapshot of an atomic model with a γ' sphere of 10 Å radius embedded in a $14\times14\times14$ supercell of γ Ni-20 at.% Al annealed for 20 ps at 1500 K with an NPT barostat using LAMMPS led to the configuration shown in fig. 5.3, with only atoms identified as γ'-like being illustrated. There one sees random single atoms identified as γ'-like (false positives or actual random γ') and the sphere, as expected. The radius found for the sphere is slightly smaller than 10 Å, which we understand to be due to the interface, which is not classified as γ' even without high temperature annealing, due to the dependence of the classification on the BOPs which represent the local neighbourhood (in this case for the 18 nearest atoms). Investigating individual atoms from the centre of the γ' sphere outwards to the host phase one can find that the 'like' and 'unlike' BOPs change and the probability of γ' decreases whereas the probability of γ increases. Note that due to the presence of the different filters we require for a clear identification of γ or γ', both 'like' and 'unlike' classifications to agree with each other. The classified atoms in the interface are found to be a mix of γ and γ' , i.e. classified as γ for the 'like' filter and γ' for the 'unlike' filter or vice versa.

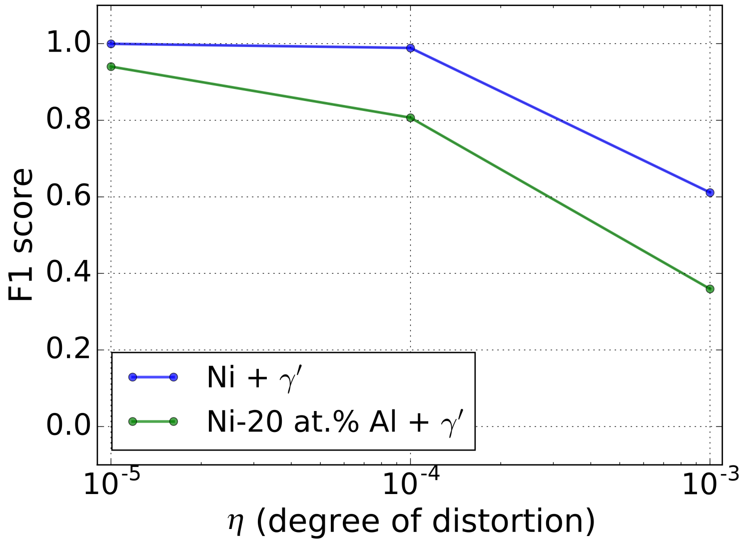


Figure 5.2: *F1* score measuring the ability of the classifier to correctly identify γ' within pure Ni and γ' Ni-20 at.% Al, for three different degrees of lattice distortions η. The *F1* score is computed comparing two frames of the same degree of distortion.

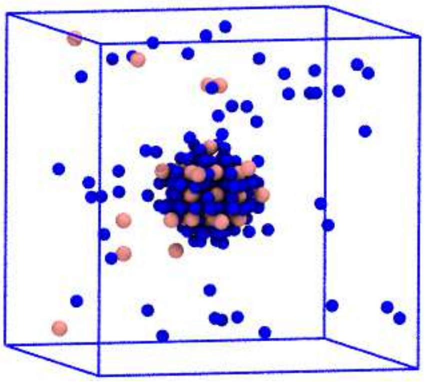


Figure 5.3: Atoms (Ni blue and Al pink) identified as γ'-like in a model of γ' Ni-20 at.% Al containing a pre-existing spherical γ' precipitate. The sytem has been annealed using MD at 1500 K for 20 ps.

5.2.2 Results

Emulating the initial stages of ordering during quenching from supersolvus temperatures we performed simulated annealing of γ Ni-20 at.% Al single crystals at 1300 K, 1400 K and 1500 K using NPT MD as implemented in LAMMPS. The simulations were carried out up to 100 ns at each temperature using a Ni-Al EAM potential [152]. The models included up to 108 000 atoms and thermal vacancies. No pre-existing γ' were deliberately inserted. The chemically disordered γ matrix was obtained by first constructing a pure fcc Ni crystal and then randomly removing atoms to create thermal vacancies followed by random substitution of Ni with Al atoms. Note that one cannot easily calculate the number of vacancies in equilibrium within a disordered alloy of a given composition, it depends on the mutual interaction of the components. In this case, since Al has the larger atomic radius, it probably favours the introduction of more rather than less vacancies than there would be in pure Ni. For simplicity, the formula for a single element was used, replacing the Gibbs energy vacancy formation by its weighted average for the pure elements.

The number of thermal vacancies, N_{vac}, was calculated based on literature values of thermodynamic quantities ($H_{f,\mathrm{Ni}} = 1.56$ eV, $H_{f,\mathrm{Al}} = 0.775$ eV, $S_{f,\mathrm{Ni}} = 3.3\,k_B$, $S_{f,\mathrm{Al}} = 2.6\,k_B$) [190] as

$$N_{\mathrm{vac}} = N_{\mathrm{atoms}} \cdot \exp\left(-\frac{c_{\mathrm{Al}} H_{f,\mathrm{Al}} + c_{\mathrm{Ni}} H_{f,\mathrm{Ni}}}{k_B T} + c_{\mathrm{Al}} S_{f,\mathrm{Al}} + c_{\mathrm{Ni}} S_{f,\mathrm{Ni}}\right),$$

where H_f and S_f are vacancy formation enthalpies and entropies respectively for pure fcc Al and fcc Ni. If multiple values for H_f or S_f were found, they were averaged. This resulted in 9, 23 and 49 vacancies at 1300 K, 1400 K and 1500 K respectively. After completion of the simulations q_l were calculated for all three filters.

In fig. 5.4 the concentration of γ'-like atoms, $c_{\gamma'}$, is shown as a function of time for annealing at 1300 K and 1500 K to gain an insight into *global* trends. There we note three effects: 1) a clear trend of increasing $c_{\gamma'}$ is visible, i.e. frame to frame noise is significantly below the systematic $c_{\gamma'}$ shift over time, 2) we see a noisy but otherwise monotonic increase for both 1300 K and 1500 K and 3) $c_{\gamma'}$ is larger for 1500 K than for 1300 K after the same annealing time. Note that both curves start with $c_{\gamma'} > 0$, which can be explained by γ' appearing at random in γ as initiated. The exact thermal contribution to false positives is not known but we think it is a fair assumption that it would not lead to a systematic increase or decrease over time and rather lead to white noise, which would explain the observed noisy behaviour in fig. 5.4 from time step to time step.

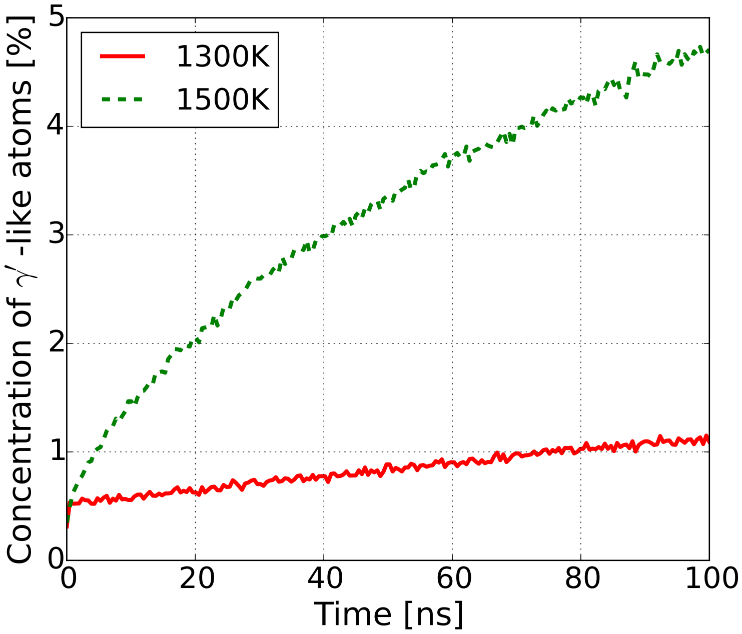
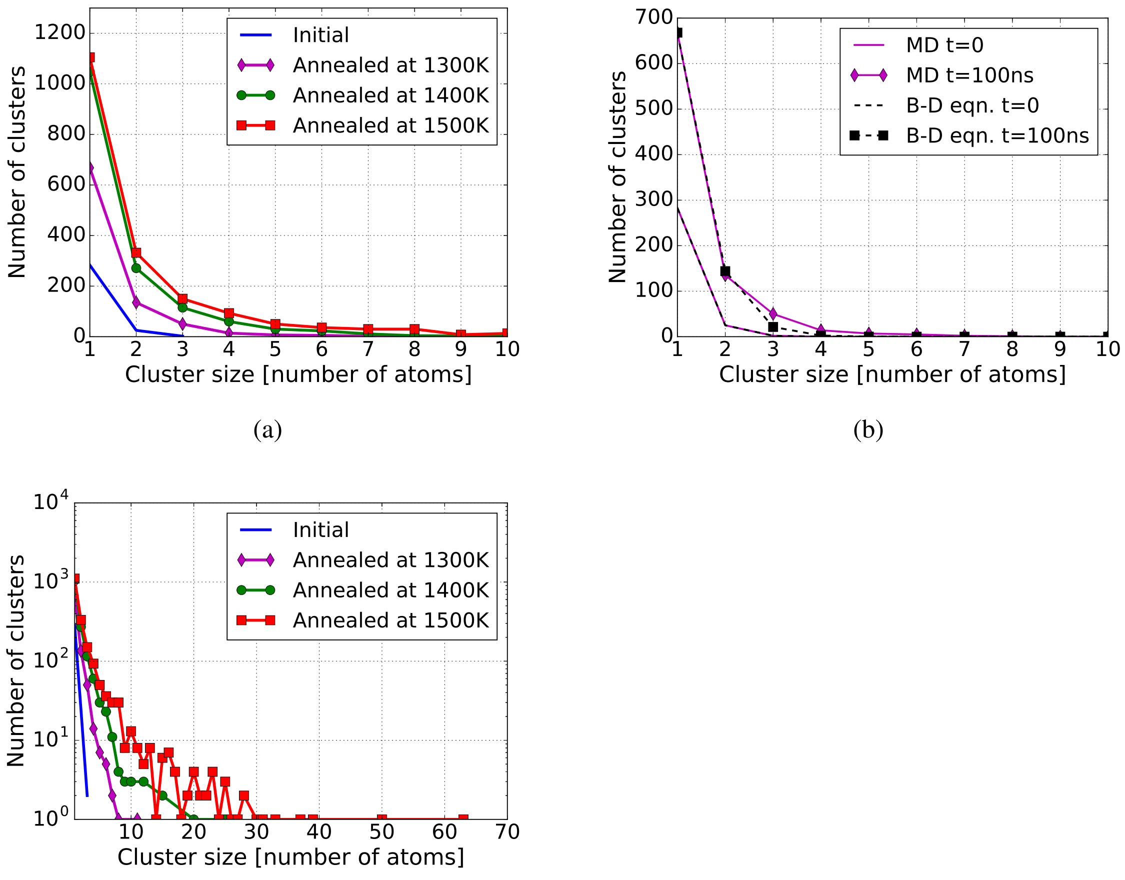


Figure 5.4: Concentration of atoms identified as γ'-like for annealed single crystals of 108 000 atoms over a period of 100 ns at 1300 and 1500 K.

Having classified individual atoms, we are also interested in the *semi-local* properties, such as the agglomeration of γ' and concentration profiles. If the globally observed ordering is physical, we should find clusters of γ' atoms which evolve over time. To this end we applied hierarchical clustering, linking γ' atoms which were less than 4 Å apart, including potential 2nd nearest neighbours of ideal γ'. Note that the choice of other cut-offs may be useful as well. Generally, we found that by decreasing the cut-off distance the frequency of smaller clusters increased, and the frequency of larger clusters decreased and vice versa for an increase in cut-off distance. Changes around 4 Å though did not significantly change the characteristics of the cluster-size distributions described below. The cluster-size distribution before and after annealing is shown for all three temperatures in fig. 5.5 a). The cluster-size curves appear to follow near exponential distributions, see fig. 5.5 c). The evolution of the cluster-size distribution can be characterized with the diffusion equation as derived using the Becker-Doehring theory [40]. Modelling the cluster-size distribution in fig. 5.5 a) we find reasonable fits, as shown for the distribution obtained for 1300 K in fig. 5.5 b). Results for 1400 K and 1500 K look very similar and are therefore not shown. Solutions of the diffusion equation were fitted using a constant diffusion coefficient and a diffusion coefficient which depends linearly on the cluster size. The constant diffusion coefficient was found to give better fits with 0.0029 atoms ns^{-1}, 0.0052 atoms ns^{-1} and 0.0057 atoms ns^{-1} at 1300, 1400 and 1500 K respectively. It can be seen in fig. 5.5 b) that although the modelled cluster-size distribution shape closely matches the overall distribution, the frequencies of larger cluster sizes are underestimated. A plausible reason for this could be the simple model for the diffusion coefficient as well as the optimization criterion: the Euclidean distance between true distribution and prediction. In particular the latter emphasizes the fit of high frequencies. Further insights into the ordering and chemical profiles on the semi-local scale can be gleaned from fig. 5.6, where a projection of the c_{Al} and $c_{\gamma'}$ onto the x-axis of the simulation box are shown as functions of time. These concentration profiles were obtained by creating histograms of 2 Å wide bins along the x-axis collecting all Al or γ' atoms and computing the percentage for the respective bin. Fig. 5.6 a) and b) show $c_{\mathrm{Al}}(t)$ and $c_{\gamma'}(t)$ for 1300 K whereas fig. 5.6 c) and d) show $c_{\mathrm{Al}}(t)$ and $c_{\gamma'}(t)$ for 1500 K. Note how c_{Al} in fig. 5.6 a) and c) shows hill and valley type behaviour, induced by the initial random assignment of Al slightly changing over time. Fig. 5.6 b) and d) give a similarly slowly changing profile in $c_{\gamma'}$ showing a localized increase in $c_{\gamma'}$, particularly pronounced for 1500 K. Visually comparing the c_{Al} hills and valleys with the $c_{\gamma'}$ profile it seems that above average Al concentrations are related to regions with increased $c_{\gamma'}$ values. An attempt to validate this is shown in fig. 5.6 e) which shows the Pearson correlation coefficient $\rho(t) \equiv \mathrm{cov}\left(c_{\mathrm{Al}}(t=0), c_{\gamma'}(t)\right)/\mathrm{std}\left(c_{\mathrm{Al}}(t=0)\right)\mathrm{std}\left(c_{\gamma'}(t)\right)$ at 1500 K. We see slight positive correlation of c_{Al} and $c_{\gamma'}$. That γ' would occur preferentially in higher Al concentrations is plausible since γ' has a stoichiometry of 3 Ni and 1 Al, and thus the global average concentration of Al is below the ideal stoichiometry. A similar trend as in fig. 5.6 e) can be found when calculating $\rho(t)$ without fixing c_{Al} to the first time step.



(a)

(b)

(c)

Figure 5.5: (a) linear and (c) log scale size distributions of γ' clusters before and after 100 ns annealing. Shown are the initial distribution (blue) and the final distributions at 1300 K (magenta), 1400 K (green) and 1500 K (red). (b) Linear cluster-size distribution as shown in a) for the single crystal of γ' Ni-20 at.% Al annealed at 1300 K using MD together with fitted functions which solve the Becker-Doehring (B-D) equation.

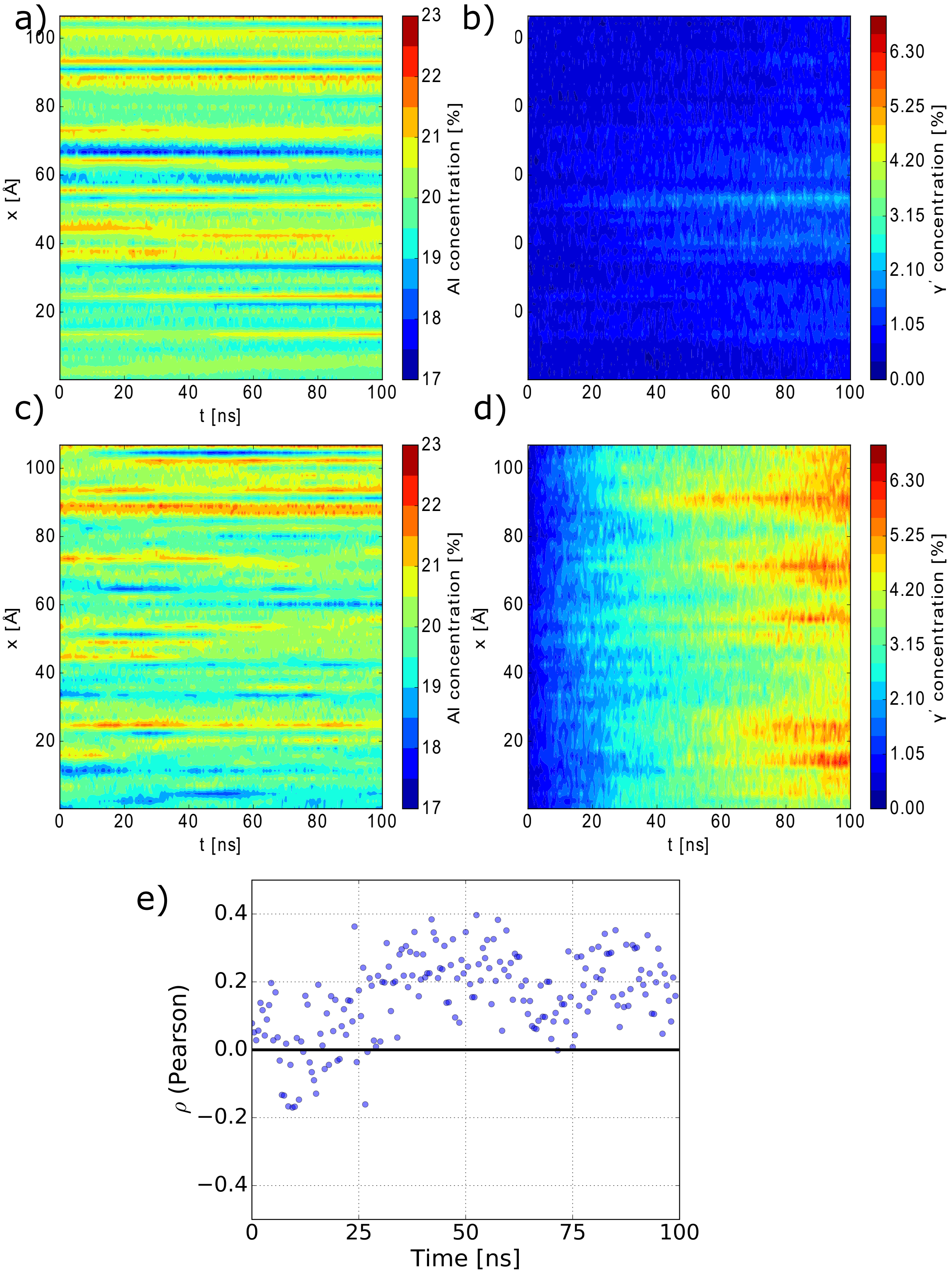


Figure 5.6: c_{Al} (left (a) and (c)) and $c_{\gamma'}$ (right (b) and (d)) concentration profiles as a function of annealing time along the x-axis at 1300 K (top (a) and (b)) and 1500 K (middle (c) and (d)). Part (e) displays the Pearson correlation coefficient between the initial c_{Al} and the $c_{\gamma'}$ at 1500 K, with $c_{\gamma'}$ varying over the simulation time.

Finally, as a representation of *local* measures we show identified γ'-like atoms forming clusters such as in fig. 5.7. The observed clusters contain the typical $\langle 100 \rangle$ row of Al atoms characteristic for γ'. This configuration of two clusters is shown in particular because one can see that they are in antiphase configuration. This was observed quite frequently in the observed clusters.

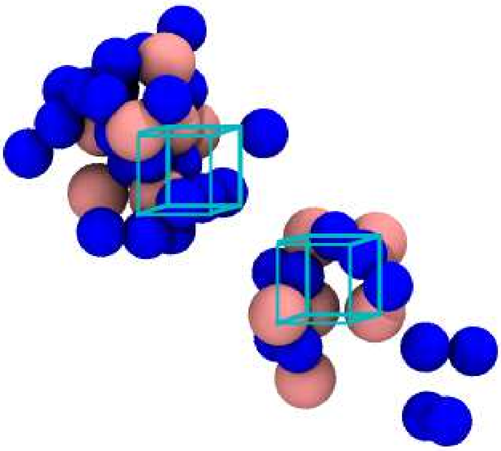


Figure 5.7: Two clusters of atoms are shown which are classified as γ'-like after annealing with MD for 100 ns at 1500 K. Pink atoms represent Al and blue Ni. The superimposed cubes indicate γ' unit cells which are in an antiphase orientation.

5.2.3 Discussion

In this section we applied supervised learning to the problem of observing the disorder $\rightarrow$ order phase transition for γ'. This approach qualifies as supervised because we explicitly defined for the training of the classifier which configurations were γ' and which γ. This approach was chosen among a variety of others available because a probability based approach was required for simplified anomaly detection and for future extensions, as discussed in section 4.2. For this approach we made the simplifying assumption that all phases are equally probable and furthermore that classifications of individual atoms are independent of their neighbourhood. Having tested this with an iterative Bayesian treatment we found it led to an unwanted ‘homogenization’ effect removing desired effects such as clusters. The disadvantage of the current method to generate PDFs for the classifier is that all atoms in an atom configuration are assumed to belong to one kind of phase. This is not true for effects such as defects or phase mixtures. This problem is addressed in the following Chapter 6. A further technical disadvantage of the current approach using splines is that they quickly become computationally unfeasible for features of increasing dimension. However, despite this naive approach and the strong assumptions made we created a relatively robust classifier up to 1500 K. It is thus conceivable that this general probabilistic approach could also be used for more complex phases such as topologically close packed ones.

It was furthermore demonstrated that vacancy assisted ordering indeed takes place in the simulated annealing of Ni-20 at.% Al at 1300 K to 1500 K, as visible in fig. 5.6 b) and d). Using a semi-grand canonical ensemble method as implemented by Sadigh et al. [195] we estimate the γ + γ' two phase region to lie between 17.6 and 25.5 at.% Al at 1500 K for the given EAM potential. This places our Ni-20 at.% Al composition in the middle under a high driving force for chemical ordering and decomposition. Given the typical Ni-Al phase diagram, one would expect to find less γ' for the same composition at higher temperature in equilibrium. Thus, finding a larger acceleration to high $c_{\gamma'}$ values for 1500 K over 1300 K may be surprising. But the MD simulations are hardly already at equilibrium and a higher temperature does allow for a faster sampling of configuration space making it more likely for possible ordering to occur. Though this may lead to a faster initial rise we expect that for longer annealing times the $c_{\gamma'}$ for 1300 K will exceed that for 1500 K. For 1400 K we would expect 40.6 at.% γ' in equilibrium using the lever rule for Ni-20 at.% Al. Additionally as evidence of ordering a comparison of fig. 5.6 b) and d) with a) and c) shows that a positive, if stochastic, correlation of above average c_{Al} and $c_{\gamma'}$ likely exists as illustrated in fig. 5.6 d). This stochastic behaviour is probably due to the stochastic motion of the thermal vacancies in the single crystal. Although the vacancy-assisted ordering leads to a heterogeneous profile on the microscopic scale there is no other spatial preference, e.g. a GB. Hence on a macroscopic scale this ordering would likely appear homogeneous. This would agree with prior findings indicating that homogeneous ordering can precede or appear concomitantly as c_{Al} changes, ordering the phase without decomposition. Also, the positive correlation of $c_{\gamma'}$ and c_{Al} is found to increase from 1300 K to 1500 K. Possible longer time scale behaviour could be spinodal decomposition (conditional or not, depending on the position of the $G_{\gamma'}(c)$ inflection point relative to Ni-20 at.% Al) [2] or nucleation and growth [36, 119, 215]. However, a spinodal decomposition of the simulated γ Ni-20 at.% Al phase would be most consistent with recent experimental observations [6, 228, 248] since it lies at the centre of the $\gamma + \gamma'$ phase region and is thus very likely between the spinodal curves of both γ and γ'. Regarding the evolution of the cluster-size distribution we note that Becker-Doehring theory assumes that all clusters have a similar shape and are characterized with an attachment/detachment frequency for a given size. Furthermore we note that the definition of cluster is not straightforward, i.e. depending on the definition it may vary from quite spongy as found for B2 clusters by Athenes *et al.* [3] to more compact as found by Auer *et al.* and ten Wolde *et al.* [4, 187]. Also, the equations used here for the diffusion coefficient were quite simple and more accurate functions may be possible as well as other fitness measures used to fit the cluster-size distribution which prioritize for example, larger size clusters. Creating good descriptions of the cluster-size models based on the clustering observed in MD could provide a useful tool for the prediction of the longer time scale precipitation mechanism, connecting MD with observations from experiments. Visual inspection of the growth of the clusters seen in the present work seem to depend on the formation of Al rows, possibly forming the rate limiting step. For further ordering we expect the interaction of clusters, particularly adjacent clusters in antiphase orientation, to play an increasingly important role slowing down the ordering.

5.2.4 Conclusion

In this section we used bond order parameters (BOPs) / Steinhardt parameters in a supervised learning scheme with a Bayesian classifier to investigate ordering in highly chemically disordered γ Ni-20 at.% Al solid solution at high temperatures emulating the initial stages in the quenching processes of Ni-based superalloy components. In order to identify L1_2 ordered γ' configurations we proposed BOPs which were made sensitive to chemical information. A method for simplifying the sample generation via distortions taken from a multivariate normal distribution was suggested.

Clear evidence for ordering was found for γ' at temperatures up to 1500 K, which was interpreted as homogeneous ordering, likely already accompanied by spinodal decomposition. Furthermore, modelling the evolution of the cluster-size distributions using Becker-Doehring theory yielded reasonable predictions which slightly underestimated the frequency of larger clusters.

Chapter 6

Phase Transformations in Bicrystals

In this chapter we investigate the interplay between the solutes, the ordering and the kinetics of CSL grain boundaries in Ni, γ Ni-10 at.% Al and Ni-20 at.% Al environments. This is done to gain some understanding of the atomistic processes involved in the formation of serrated GBs from a supersaturated solid solution. To this end we propose the 'decomposition algorithm', which is a method for 'learning' PDFs in a semi-supervised fashion by decomposing the distributions in feature space using mixtures of Gaussians. This allows PDFs to be generated by groups of trajectories, e.g. configurations of pure Ni with and without GBs.

For GBs to migrate, a sufficient driving force is required, where 'sufficient' depends on the GB type, the temperature, and the time scale under consideration. In this thesis research no artificial driving force, such as a modified interatomic potential, was imposed on the system. However, we will show that GB migration can be simulated, leading to trajectories that resemble those observed by others in elemental systems [50, 51, 204, 244], which did use artificial driving forces, exhibiting 'locked' and 'running' states (i.e. stationary and dynamic behaviour).

After the introduction of the decomposition algorithm and our technique for tracking GBs, we will show that the driving force in our calculations relates to the elastic strain and the reduction of the system's enthalpy accompanying a local γ' ordering. We will show that the solute concentration affects the properties of the GBs. While the relationship between GB mobility and ordering is relatively complex, we find that the symmetric tilt GBs studied in this work are less mobile than the symmetric twist GBs.

Lastly, we will discuss the potential implications of our results for understanding the initial stages of GB serration. We propose that diffusion induced grain boundary migration (DIGM) is a possible alternative to previous models, which require pre-existing precipitates actively pushing the GBs.

6.1 Method

Employing the method described in the previous chapter, we make use of BOPs and the three element filters to set up our feature spaces. However, now we develop conditional PDFs, $p(\mathbf{x}|\text{Phase})$, by making use of EM-GMMs (see section 4.4) approximating the distribution of the atoms in feature space using multivariate normal distributions, implemented using *scikit-learn* [168]. In order to find $p(\mathbf{x}|\text{Phase})$ we start by modelling the feature space distributions for individual trajectories giving $p(\mathbf{x}|\text{Trajectory})$. This is sensible because we found that the individual phases and defects have characteristic distributions in the form of collections of peaks. The approximation with GMMs is then that we assume these peaks can be approximated by Gaussians, and that we know how many Gaussians to use.

To produce $p(\mathbf{x}|\text{Phase})$ from $p(\mathbf{x}|\text{Trajectory})$, we have developed a decomposition algorithm, which will be discussed in more detail in section 6.1.1. In order to use EM-GMMs we have to determine the number of Gaussians to use. This is done using the Bayesian Information Criterion (BIC) $\ln p(\mathbf{X}|\text{Phase}) \approx -\ln p(\mathbf{X}|\boldsymbol{\theta}, \text{Phase}) + \frac{M}{2} \ln N$ with M the number of Gaussians and N the number of observations. The implemented code therefore performs searches increasing the number of components and terminates when the expected change in BIC is larger than zero indicating a worsening of the fit. The model with the number of Gaussians that produced the smallest BIC is then used. Then we decompose $p(\mathbf{x}|\text{Trajectory})$ into several $p(\mathbf{x}|\text{Phase})$. The term ‘Phase’ is used generically, since we now actually refer to any set of Gaussians as a ‘Phase’ which cannot be further reduced using the decomposition algorithm for the given trajectories. When one is studying trajectories, which contain only ideal fcc or bcc Ni configurations, for example, this is perfectly fine. However, when studying trajectories with fcc Ni and point defects we will get models for fcc Ni and for point defects, where point defects, strictly speaking, may not be considered a phase.

A bonus of being able to develop PDFs this way is that it allows for the capturing of atoms as GB atoms. This itself opens up the possibility of tracking GB positions (i.e. planes). While it will prove difficult with the PDFs obtained here to perfectly identify all aspects of GBs at high temperatures, one is definitively able to calculate their mean position over some volume. We will thus track GBs along the z-axis (the normal to the GB plane) as described in section 6.1.2.

6.1.1 The Decomposition Algorithm

The aim is to achieve a fine grained basis set of models $p(\mathbf{x}|\text{Phase})$. The algorithm proposed in this work therefore has the following three stages: 1) the identification of *duplicate* Gaussians in models of different trajectories, 2) the *decomposition* of models into a new set of models and 3) the *updating* of the weights of the new set of models.

Duplicates are the same, or very similar, Gaussians which appear more than once throughout all $p(\mathbf{x}|\text{Trajectory})$. This can easily happen when distributions for two or more trajectories are approximated containing among them the same phase, such as fcc Ni. The fcc configuration in those trajectories then leaves very similar peaks in the $p(\mathbf{x}|\text{Trajectory})$ leading to duplicate Gaussians. In order to find duplicate positions one can, for Gaussians, compare their mean vector $\boldsymbol{\mu}$ and / or covariance matrix $\mathbf{\Sigma}$. We found it sufficient to compare $\boldsymbol{\mu}$ only using the Euclidic distance as the similarity measure. In order to find all Gaussians which are duplicates we used hierarchical clustering as implemented in *scipy* [116], accepting as duplicate any two Gaussians which are closer than a threshold of 0.01. These Gaussians are then combined into an average version by computing the means of their $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$.

For the decomposition we identified three conceivable cases: the *identity* case (two $p(\mathbf{x}|\text{Phase})$ for different trajectories are identical), the *superset* case ($p(\mathbf{x}|\text{Phase})$ for one of the trajectories is a superset for $p(\mathbf{x}|\text{Phase})$ of another trajectory) and the *intersecting* case (two $p(\mathbf{x}|\text{Phase})$ for different trajectories share some Gaussians but not all). The corresponding actions to be taken are as follows: identity case: remove one of the two identical $p(\mathbf{x}|\text{Phase})$ from the database; superset case: remove the shared Gaussians from the $p(\mathbf{x}|\text{Phase})$ which forms the superset; intersecting case: remove the shared Gaussian from both $p(\mathbf{x}|\text{Phase})$ and create a new one. This algorithm is outlined step by step in alg. 6.1. The decomposition algorithm thus ensures that the smallest model contains at least one Gaussian.

After the decomposition algorithm has produced a basis set of models $\{p(\mathbf{x}|\text{Phase})\}_{\text{Phase}\in\text{Phases}}$ we re-adjust the weights for each $p(\mathbf{x}|\text{Phase})$. This is done by regressing over the distributions of all trajectories.

Algorithm 6.1 The decomposition algorithm takes the J models $p\left(\mathbf{x}|\text{Trajectory}\right)$ obtained for each of the J trajectories and compares them using their M Gaussians and returns K new phase related models $p\left(\mathbf{x}|\text{Phase}\right)$. Note that this algorithm also generates models which describe only part of a phase if that phase shares at least one Gaussian with another phase.

```
procedure DECOMPOSITION(trajectory models $\{p(\mathbf{x}, \text{Trajectory=j})\}_{j\in[1,J]}$)
    phase models ← $\emptyset$
    for Gaussian $m$ in $[1,M]$ do
        selected models ← all $p(\mathbf{x}|\text{Trajectory=j})_{j\in[1,J]}$ which contain $m$
        if number of models in selected models $\geq 2$ then
            I ← intersections between all selected models
            new model ← intersection with fewest Gaussians in I
            selected models ← update selected models by removing the new model's components
            phase models ← add selected models and new model
        else
            phase models ← selected models
        end if
    end for
    return phase models $\{p(\mathbf{x}|Phase = k)\}_{k\in K}$
end procedure
```

6.1.2 Tracking Grain Boundaries

The algorithm is illustrated for GBs in fig. 6.1 and has 4 stages. In the *first stage*, the atoms are classified and grouped as GB/non-GB atoms, upper left in fig. 6.1, allowing the determination of the z-coordinate of the expected GB position (the z-axis is taken to be normal to the GB plane). In the *second stage* the expected GB positions are collected for each time frame, upper right in fig. 6.1. In the *third stage* the GB positions are linked temporally using a ranking scheme based on the sigmoid of $p\left(\Delta z|\text{link} = \text{True}\right)$ and $p\left(\Delta z|\text{link} = \text{False}\right)$, lower right in fig. 6.1. Lastly, the *fourth stage* yields the GB trajectories, lower left in fig. 6.1.

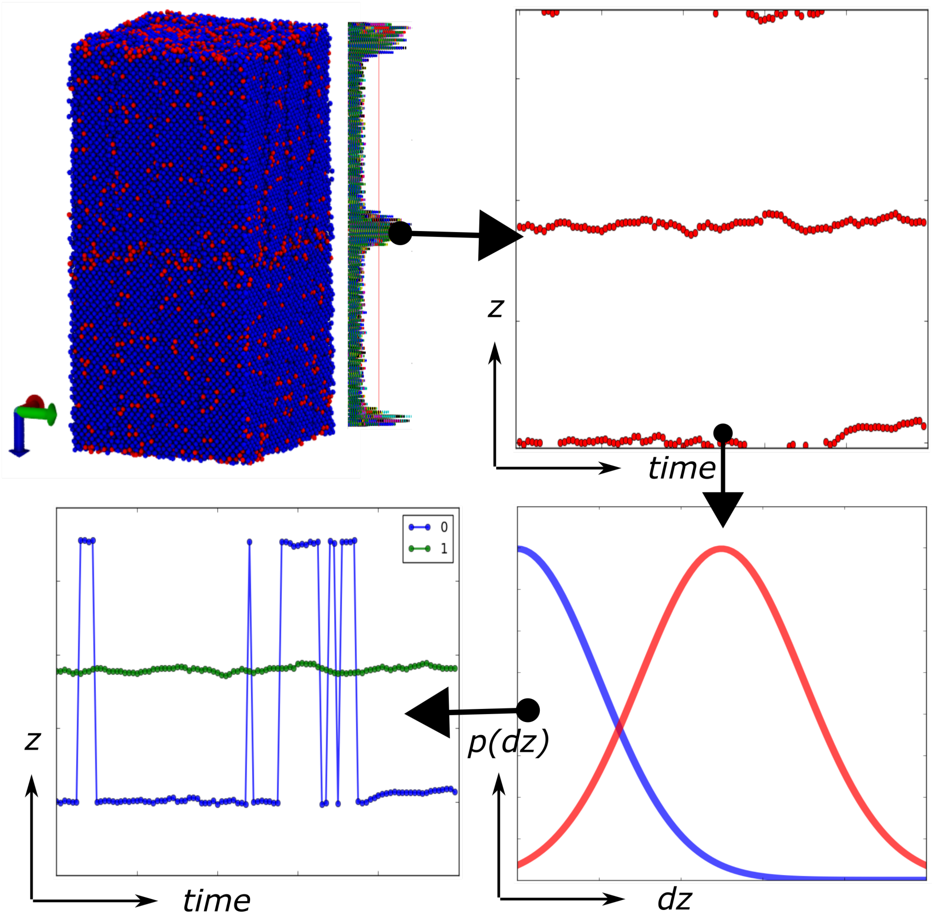


Figure 6.1: Schematic illustration of the grain boundary tracking algorithm. The arrows highlight the sequence of steps in the algorithm. First (upper left) the atoms are classified into GB (red) and non-GB (blue) atoms, here for a $\Sigma 85$ (001) twist GB in fcc Ni at 1500 K. Note that the model contains two GBs because of the periodic boundary conditions. The concentration profile of non-GB atoms along z is illustrated by the histogram in the top left also drawn along z (the normal direction to the plane containing the GB). The resulting peaks are then collected for each time step (upper right). In order to connect the collected peak positions two conditional PDFs are created for $p\left(\Delta z | \mathrm{link} = \mathrm{True}\right)$ and $p\left(\Delta z | \mathrm{link} = \mathrm{False}\right)$ (lower right). Together they generate the trajectories of the grain boundaries (lower left).

6.2 Results

The symmetric CSL tilt and twist GBs studied in this research were simulated using LAMMPS with an NPT barostat and Mishin's Ni-Al EAM potential [152]. The CSL geometries were generated using *GBstudio* [161], although new Python based alternatives have recently become available developed by Banadaki *et al.* [9, 10]. In this thesis research, the following

GB systems were studied: $\Sigma 3$ tilt, $\Sigma 5$ tilt, $\Sigma 7$ tilt, $\Sigma 9$ tilt, $\Sigma 27a$ tilt, $\Sigma 27b$ tilt, $\Sigma 5$ twist, $\Sigma 13$ twist, $\Sigma 25$ twist and $\Sigma 85$ twist. The geometrical details of each GB are given in table 6.1. These GBs were chosen because they have been experimentally observed or calculated before [191, 197]. The GBs were simulated under the following model conditions: pure fcc Ni, pure fcc Ni + ppt, γ Ni-10 at.% Al, γ-Ni-20 at.% Al and γ-Ni-20 at.% Al + ppt, where 'ppt' denotes 'precipitate'. If a precipitate was pre-existing, then it was separated by 5 Å from one of the GBs and was of 10 Å radius.

The following sections discuss: 1. The cross-validation of the atomistic classifier; 2. The driving forces in the bicrystal simulations; 3. The impact of the GBs on the ordering; and 4. The impact of the Al solutes on the GB migration.

6.2.1 Generation and Cross-Validation of the Classifiers

The PDFs necessary to form the classifier were created by training on trajectories of two bicrystal and two single crystal systems: $\Sigma 3$ tilt (in fcc Ni), $\Sigma 5$ twist (in fcc Ni), bulk γ' and γ Ni-20 at.% Al. The simulations of these systems were run twice for a range of temperatures, one for the training and one for the testing. The γ and γ' configurations were set up as $5 \times 5 \times 5$ unit cells and annealed at 0.1, 300, 1000 and 1500 K for 50 ps with snapshots taken every 2.5 ps. The $p(\mathbf{x}|\text{Trajectory})$ for γ and γ' respectively were created and GMMs fitted over all temperatures and snapshots. The two GBs, on the other hand, were annealed at 0.1 and 300 K for 50 ps with snapshots taken every 5 ps created using $3 \times 5 \pm 5$ CSL cells for $\Sigma 3$ tilt and $6 \times 6 \pm 35$ CSL cells for $\Sigma 5$ twist. For both types of GB, $p(\mathbf{x}|\text{Trajectory})$ was constructed at each temperature individually, which led to stabler classifiers. This was followed by applying the decomposition algorithm.

In order to avoid the previous uncertainties about the labelling encountered in the γ/γ' interfaces, we cross-validated the classification of the γ and γ' directly on a distorted version of single phase configurations. We again made use of the *F1* score (see section 5.2.1.2) and also the 'Probability' score $\equiv \frac{t_p}{N_p}$, where t_p means 'true positive' and N_p is the number of actual positives, to gain a feeling for how many atoms were correctly classified. Both scores are shown in fig. 6.2 with the *F1* score in a) and the probability score in b). In fig. 6.2 a) we see that the F score for γ' increases with increasing temperature from 0.25 at 0.1 K to ~0.97 at 1500 K. The *F1* score for γ, on the other hand, first increases to ~0.97 at 300 K and then decreases to ~0.64 at 1500 K. Fig. 6.2 b) shows a similar trend with γ having a more shallow overall curve, remaining within the range of 0.6 to 0.7 for all temperatures. The poor performance at 0.1 K is somewhat surprising but could be explained by the general behaviour of Gaussians, for which outliers that were introduced by adding configurations from 1500 K have a disproportionate influence on the parameters. However, these outliers seem to have a positive effect on the stability of the classifier at high temperatures. A remedy for this problem would be to approximate $p(\boldsymbol{x}|\text{Trajectory})$ with Student's t distributions instead, which are known to be less sensitive to outliers.

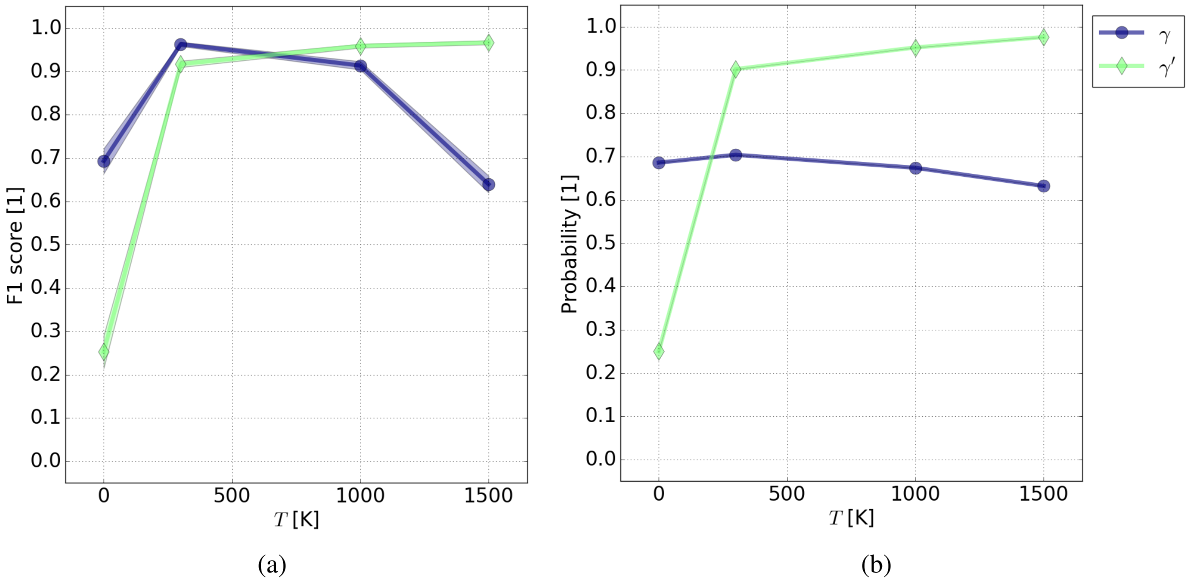


Figure 6.2: Robustness against thermal noise of the γ/γ' classification illustrated in terms of a) the *F1* score and b) the probability of correct classification.

6.2.2 Atomistic Models

After creating the PDFs using the decomposition algorithm, we apply them to the CSL GB configurations annealed for 5 ns at 1500 K using the NPT barostat, taking snapshots every 0.05 ns. Each of these simulations was repeated three times. The dimensions of the simulation boxes for each GB type are shown in tables 6.1 and 6.2 and were chosen by simulating prototype fcc Ni versions of each GB system for 1 ns at 1500 K. The box sizes were also chosen to maximize the GB cross-sectional area, and so that no unexpected GB movement for the prototype occurred. Depending on the solute content, the atomic models also contained thermal vacancies, as previously described in section 5.2.2.

tilt CSL Σ	rot. axis	rot. angle (°)	$\mathbf{z}$ (GB normal)	$\mathbf{x}$	$\mathbf{y}$
3	$[1\,1\,1]$	60	$[1\,1\,1]$	$[1\,1\,\bar{2}]$	$[\bar{1}\,1\,0]$
5	$[1\,0\,0]$	36.9	$[0\,1\,2]$	$[1\,0\,0]$	$[0\,2\,\bar{1}]$
7	$[1\,1\,1]$	38.2	$[3\,\bar{2}\,\bar{1}]$	$[1\,2\,\bar{1}]$	$[2\,1\,4]$
9	$[1\,1\,0]$	38.9	$[2\,\bar{2}\,\bar{1}]$	$[\bar{1}\,1\,\bar{4}]$	$[1\,1\,0]$
27a	$[1\,1\,0]$	31.6	$[1\,\bar{1}\,5]$	$[1\,1\,0]$	$[\bar{5}\,5\,2]$
27b	$[2\,1\,0]$	35.4	$[1\,\bar{2}\,7]$	$[5\,\bar{1}\,\bar{1}]$	$[1\,4\,1]$

(a)

tilt CSL Σ	$N_x \times N_y \pm N_z$	x [Å]	y [Å]	z [Å]
3	$9 \times 15 \pm 30$	77.6	74.8	365.8
5	$6 \times 9 \pm 60$	21.12	70.8	944.5
7	$4 \times 4 \pm 30$	34.5	64.5	790.2
9	$4 \times 4 \pm 60$	59.74	19.9	1267.2
27a	$3 \times 3 \pm 30$	14.9	77.6	1097.4
27b	$2 \times 2 \pm 20$	36.6	29.9	1034.7

(b)

Table 6.1: Dimensions of the simulation boxes for tilt CSL GB types. Part a) specifies the CSL GB geometry; b) specifies the simulation box size.

twist CSL Σ	rot. axis	rot. angle (°)	**z** (GB normal)	**x**	**y**
5	[0 0 1]	36.9	[0 0 1]	[3 1 0]	$[\bar{1}\ 3\ 0]$
13	[0 0 1]	22.6	[0 0 1]	[5 1 0]	$[\bar{1}\ 5\ 0]$
25	[0 0 1]	16.3	[0 0 1]	[7 1 0]	$[\bar{1}\ 7\ 0]$
85	[0 0 1]	8.8	[0 0 1]	[13 1 0]	$[\bar{1}\ 13\ 0]$

(a)

twist CSL Σ	$N_x \times N_y \pm N_z$	x [Å]	y [Å]	z [Å]
5	$6 \times 6 \pm 35$	66.8	66.8	246.4
13	$5 \times 5 \pm 30$	89.7	89.7	211.2
25	$3 \times 3 \pm 30$	74.7	74.7	211.2
85	$2 \times 2 \pm 25$	91.8	91.8	176.0

(b)

Table 6.2: Dimensions of the simulation boxes for twist CSL GB types. Part a) specifies the CSL GB geometry; b) specifies the simulation box size.

6.2.3 Driving Forces

In order for GBs to migrate, a sufficient driving force is required. Hence, for a directed GB migration to occur, as in the case of GB serration, the system's free energy would have to be able to decrease due to the GB migration. Thus, in simulations without a driving force and for flat symmetric GBs, no directed GB migration is to be expected. Since the computation of the free energy is expensive, the potential energy is used here as an approximation instead.

To compare the different GB types and single crystals, we define the potential energy difference to be

$$\Delta E\left(t,t'\right) = \Delta E_{\text{condition}}\left(t,t'\right) - \Delta E_{\text{pure Ni}}\left(t,t'\right), \tag{6.1}$$

where $E\left(t\right)$ always refers to the per atom mean potential energy over all trajectories, t is the elapsed simulation time, $\Delta E_{\text{condition}}\left(t,t'\right) = E_{\text{condition}}\left(t\right) - E_{\text{condition}}\left(t'\right)$ and $\Delta E_{\text{pure Ni}}\left(t,t'\right) = E_{\text{pure Ni}}\left(t\right) - E_{\text{pure Ni}}\left(t'\right)$, with $t' = 50\,\text{ps}$. To simplify the notation, $\Delta E\left(t,t'\right)$ will be abbreviated to ΔE. This measure, ΔE, therefore relates the potential energy for a GB type to that of its pure Ni equivalent, removing structural changes specific to its type of GB and shifting the potential energy to the origin by t'. ΔE is also applied to single crystals, whose $\Delta E_{\text{pure Ni}} = 0$. Hence ΔE should relate mostly to changes related to solutes, such as ordering (which is time and temperature dependent), for both single crystals and bicrystals.

The ΔE of the kinetic energy was also calculated. Although the kinetic energy itself is not negligible, it did stay relatively constant for all simulations. Hence the ΔE of the kinetic energy always fluctuated around the origin, with the largest values being about two orders of magnitude smaller than the values of ΔE for potential energy. The contribution of the kinetic energy will therefore be neglected.

The other measure is the relative volume change

$$\Delta V\left(t,t'\right) = \frac{V\left(t\right) - V\left(t'\right)}{V\left(t'\right)}, \tag{6.2}$$

where $V\left(t\right)$ is the volume of the simulation box at t and $t' = 50\,\mathrm{ps}$. To simplify the notation, we will denote $\Delta V\left(t,t'\right)$ by ΔV. If the γ' ordering occurs, one would expect a change in the volume of γ Ni-20 at.% Al. Using the linear thermal expansion coefficients given by Mishin [152] for this Ni-Al potential, we would expect that the difference in lattice constants between γ' and γ Ni-20 at.% Al (γ-alloy) at $T = 1500\,\mathrm{K}$ would be $\left(a_{\gamma'}\left(T\right) - a_{\gamma\text{-alloy}}\left(T\right)\right)/a_{\gamma\text{-alloy}}\left(T\right) = -2.13\%$ and similarly for the volume $\frac{V_{\gamma'}(T)}{V_{\gamma\text{-alloy}}(T)} - 1 = -6.26\%$. Although this assumes a complete transformation to γ', we would still expect a significant volume shrinkage for the transition from γ Ni-20 at.% Al to the equilibrium $\gamma + \gamma'$ system during annealing at 1500 K.

Fig. 6.3 shows ΔE for all GB types and single crystals in γ Ni-20 at.% Al. The black curves indicate the single crystals and show a noisy but monotonic decrease throughout the simulations, which agrees with the observed ordering in fig. 5.4. This shows a noisy but monotonic increase in the overall concentration of γ'-like atoms. The coloured lines in fig. 6.3 indicate ΔE for all GB types in this study, with the shaded areas representing the respective standard deviations.

The twist GB systems were observed to equilibrate quickly, within 50 ps, and show an initial phase with a steep slope in ΔE until ~2.5 ns. The steepest initial slopes in fig. 6.3 are exhibited by the $\Sigma 13$ and $\Sigma 25$ twist GBs. These steep initial slopes for the twist GBs are related to the accelerated ordering observed in the wake of GBs exhibiting directed migration. An example of accelerated ordering is shown for the $\Sigma 13$ twist GBs in fig. 6.6 d), displaying the GB positions and the concentration profile of γ'-like atoms with a clear increase of the concentration of γ' in the wake of the GBs. The apparently discontinuous transition from the steep ΔE slope to the shallower slope of the $\Sigma 13$ twist GBs is linked with the annihilation of some of the GBs.

The tilt GB systems were found, in some cases, to take up to about 200 ps to equilibrate, with the $\Sigma7$ and $\Sigma9$ tilt GBs taking the longest. This was caused by the chosen initial CSL configurations of the individual GB types, with the $\Sigma7$ and $\Sigma9$ tilt GBs containing minimum atom–atom distances of 0.94 Å and 1.174 Å, which would not be observed in experiments. The other tilt GBs exhibited larger, more reasonable minimum atom–atom distances. Consequently, the values of ΔE and ΔV during the initial periods for the $\Sigma7$ and $\Sigma9$ GBs were disregarded. The other tilt GB types show an overall decrease in ΔE similar to the single crystal annealed at 1500 K, with a slight offset.

Fig. 6.4 shows that for the systems studied (twist and tilt GBs and single crystals) there were overall decreases in ΔV. These decreases in ΔV are within the bounds expected for a $\gamma \rightarrow \gamma' + \gamma$ reaction. The decreases in both ΔV and ΔE depend on the rate of ordering, which in turn depends on the GB type, with twist GBs showing a faster decrease than tilt GBs. This is related to the twist GBs being more mobile, with an accelerated ordering in their wake, and tilt GBs being more stationary, with also an accelerated ordering but mostly in thin slabs in the vicinity of the GBs. Note that accelerated ordering is understood relative to the reference single crystal annealed at 1500 K. Systems of all GB types order at least as rapidly as the single crystal. The link between GB migration and Al solutes will be investigated in sections 6.2.4 and 6.2.5.

Thus, the driving forces for GB migration due to ordering are definitively present in the simulated γ Ni-20 at.% Al systems, but depending on the GB type, they may not be sufficient to drive the GBs during the MD simulations.

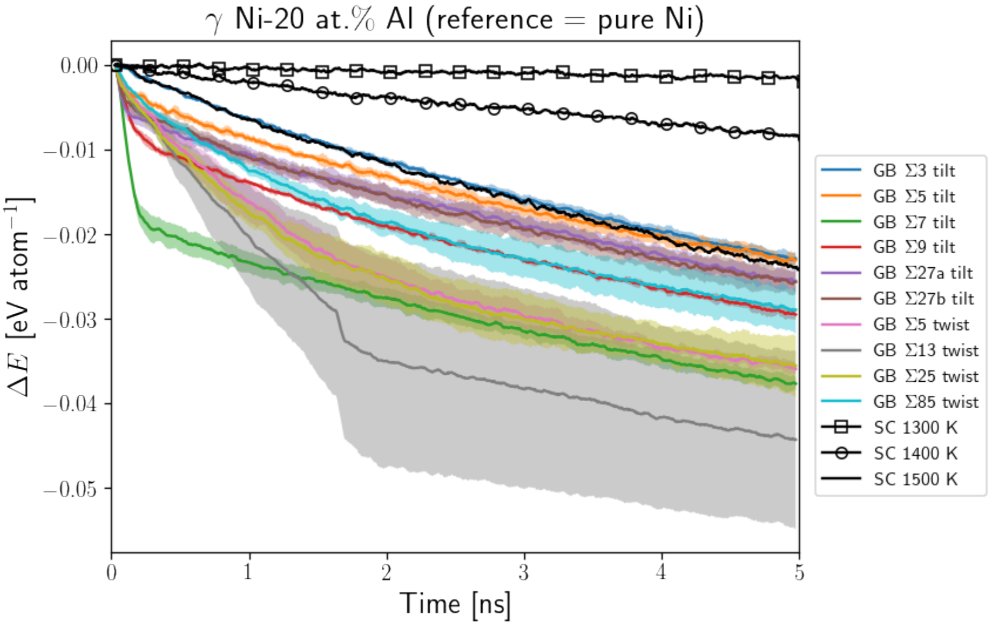


Figure 6.3: Potential energies throughout the simulation of single crystals (SC) and GBs. The lines for the GBs are the mean over all three repetitions of the MD simulation at 1500 K and the shaded areas indicate one standard deviation. The mean potential energies for all GBs at any time step of the pure Ni condition were subtracted.

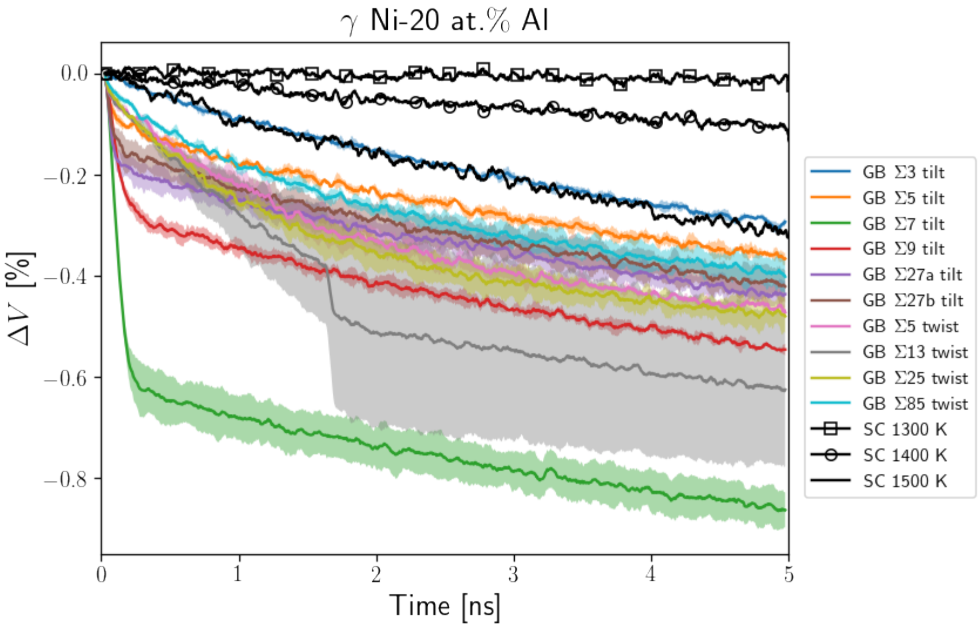
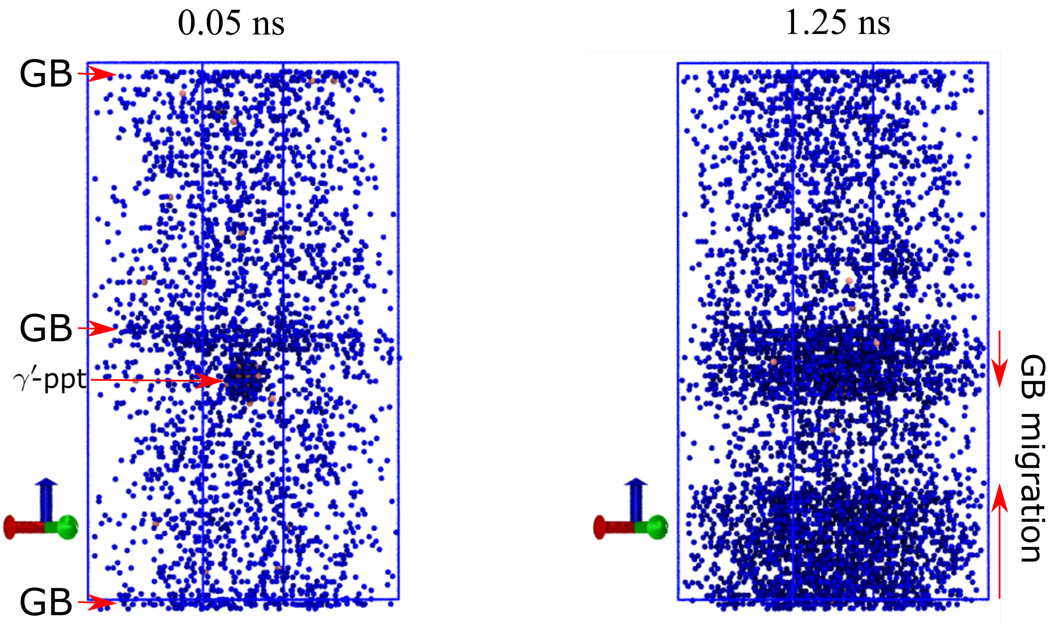


Figure 6.4: Relative simulation box volume changes throughout the simulation of single crystals (SC) and GBs. The lines for the GBs are the mean over all three repetitions of the MD simulation at 1500 K whereas the shaded areas indicate one standard deviation. The relative simulation box volume is computed with respect to the volume after 50 ps annealing.

6.2.4 Impact of GBs on Ordering

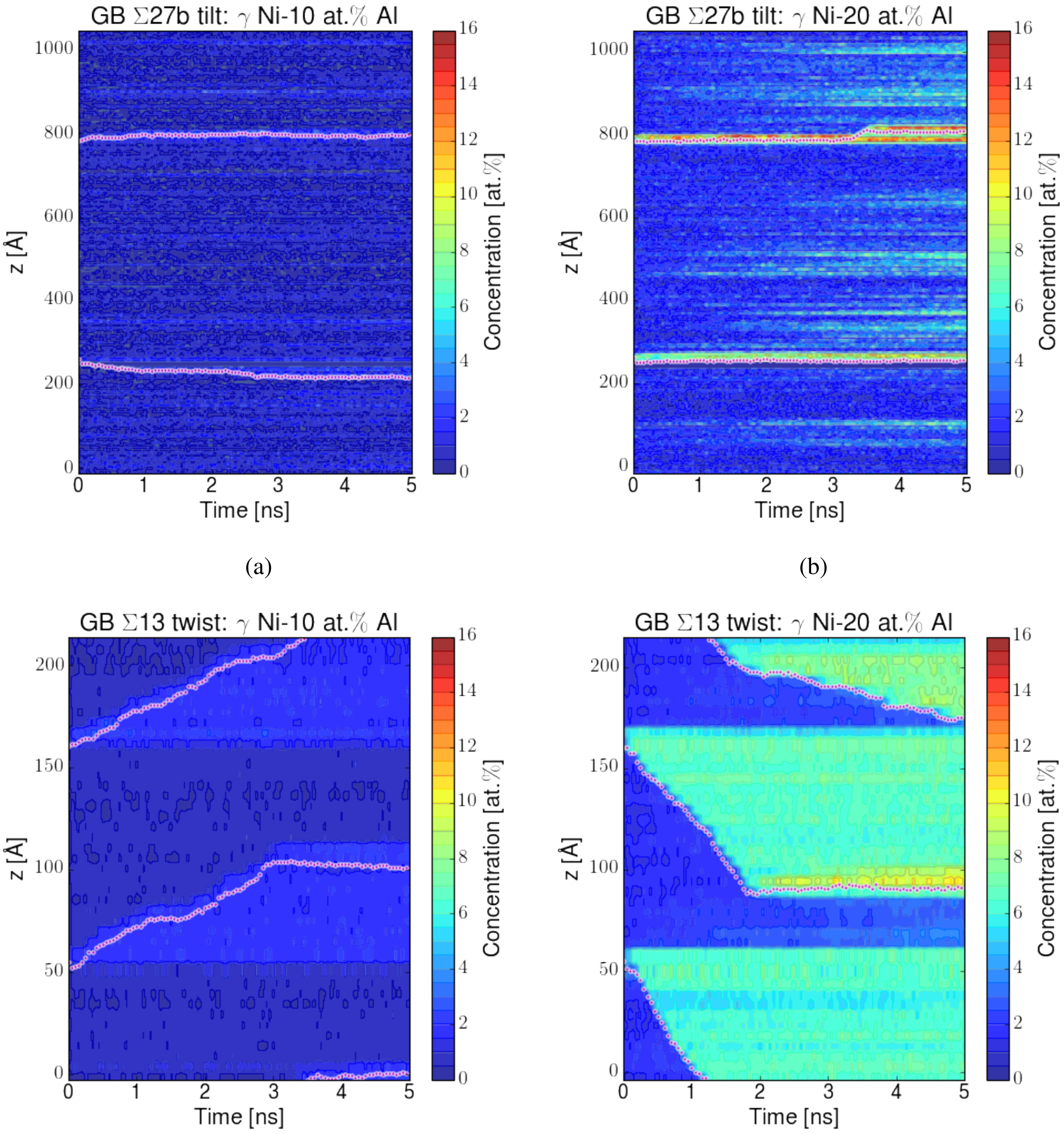
The following discussion is focused on the $\Sigma 27b$ tilt and $\Sigma 13$ twist GBs because they exhibit transitioning behaviour between a ‘running’ state and a ‘locked’ state. To give an example of the interplay between ordering and GB motion, albeit an extreme one, γ'-like atoms are shown in fig. 6.5 for a $\Sigma 13$ twist GB system contained in a γ Ni-20 at.% Al solid solution with a nearby pre-existing γ' precipitate. Fig. 6.5 shows the configuration after annealing for a) 0.05 ns and b) 1.25 ns. After 0.05 ns one can see the random appearance of γ'-like atoms throughout the volume of the supercell with increased density where the original precipitate was inserted and also where the GBs are positioned (there are two GBs due to the periodicity of the simulation box). After 1.25 ns the overall density of γ'-like atoms throughout the volume has increased, particularly in the region swept by the GBs as they move (the GBs themselves are not explicitly visible here). We also see that one GB has approached the position of the inserted precipitate. The GB actually remains in this position for some time before dissolving the precipitate and annihilating with the other GB. As we mentioned, this is a unique case, which is by no means representative of all GB behaviour observed in our simulations. While the other GBs are also found to have a profound impact on the ordering, they do not all dissolve the γ' precipitates or annihilate the other GB.



(a) (b)

Figure 6.5: Σ13 twist GB in Ni-20 at.% Al simulated at 1500 K for (a) 0.05 ns and (b) 1.25 ns. Blue and pink atoms represent Al and Ni respectively, both classified as γ'-like. The vertical direction is [0 0 1] and the supercell contains two GBs.

In fig. 6.6, the projection onto the the z-axis (taken to be normal to the plane of the GB) of the γ' concentration profiles and GB positions are shown and plotted against the simulation time. Each timestep corresponds to a configuration similar to that in fig. 6.5. The bins used to create the projection of the profile had a width of 5Å. The value for each bin was computed by counting the number of γ'-like atoms belonging to that bin and dividing by the sum over all bins for each time step. In fig. 6.6 a) and b) we see the results for the Σ27b tilt GB and in c) and d) for the Σ13 twist GB. In fig. 6.6 a) and c) both GBs are in a γ Ni-10 at.% Al environment whereas they are in a γ Ni-20 at.% Al environment in b) and d). One can see that the Σ27b tilt GBs in 6.6 a) and b) are mostly 'locked' for both environments, but do experience a short 'running' state in b). The Σ13 twist GBs, on the other hand, are found to experience a 'running' state in both environments. In γ Ni-10 at.% Al, they move about 50 Å in 3 ns before returning to a locked state. In this environment, the Σ13 twist GBs are also found to move from the beginning, but with one GB getting locked earlier than the other, which appeared to slow down after around 1.5 ns but continued moving for up to 5 ns. For both GB types we find little ordering in the γ Ni-10 at.% Al environment, as seen in fig. 6.6 a) and c), which makes sense since the driving force for ordering should be smaller for this composition (i.e. the solid solution is not supersaturated). On the other hand, we find significant ordering in the γ Ni-20 at.% Al environment, particularly in the wake of the moving GBs, see fig. 6.6 b) and d). Summarizing the results from the corresponding plots of the other GB types, including those with pre-existing precipitates, we found that the 10 Å spherical γ' precipitate had no apparent influence on the tendency of the GB to move in a particular direction but did slow it down if the GB tried to move past it. In all other simulations where the GBs were in a γ Ni-20 at.% Al environment, we consistently observed accelerated γ' ordering, particularly near GBs, whether they moved or not.



(a) (b)

(c) (d)

Figure 6.6: The γ' concentration along z (the GB normal) from 0 to 5 ns indicating GB positions with white circles for the Σ13 twist and Σ27b tilt GBs. Parts a) and b) represent the Σ27b tilt GB in a Ni-10 at.% Al and Ni-20 at.% Al environment respectively. Parts c) and d) represent the Σ13 twist GB in a Ni-10 at.% Al and Ni-20 at.% Al environment respectively. The concentrations are shown in the form of heat maps.

To quantify the interplay between the ordering and the characteristics of the migration of the GBs, we compare the median and maximum gain in γ' concentration, $\Delta c_{\gamma'}$, from the initial configuration to after 5 ns annealing, against the distance between the displacement distributions $p(\Delta z)$. In tables 6.3 and 6.4 the GB systems are ranked by maximum and median $\Delta c_{\gamma'}/A_{\mathrm{GB}}$ showing that tilt GB systems lead to an overall larger maximum and median value of $\Delta c_{\gamma'}/A_{\mathrm{GB}}$ (except for the coherent Σ3 tilt/twin boundary).

In fig. 6.7 the global $c_{\gamma'}$ is shown as a function of time comparing single and bicrystal systems for the first 5 ns of annealing. We see that single crystals have an overall lower rate of ordering and reach smaller $c_{\gamma'}$ values after the end of the simulation than the GB systems, which is consistent with the $\Delta c_{\gamma'}/A_{\mathrm{GB}}$ analysis.

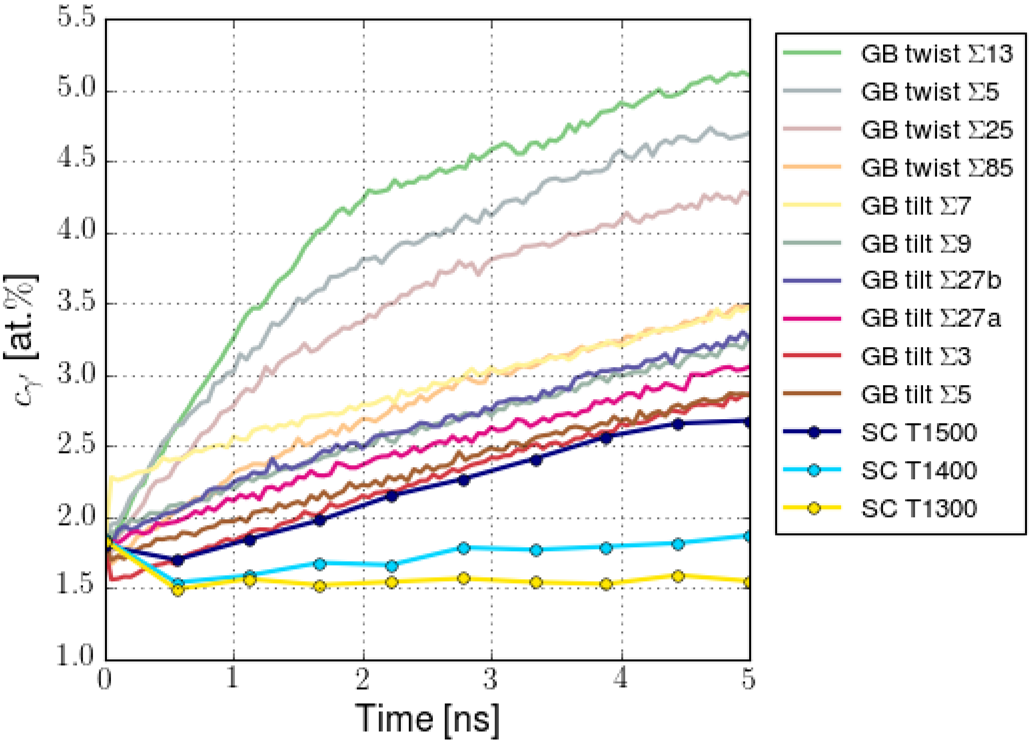


Figure 6.7: Total $c_{\gamma'}$ as a function of time for a Ni-20 at.% Al solid solution showing single crystals annealed at 1300, 1400 and 1500 K and GBs annealed at 1500 K only.

GB	Ti 27b	Ti 9	Ti 27a	Ti 5	Ti 7	Tw 5	Tw 25	Tw 13	Ti 3	Tw 85
$\Delta c_{\gamma'}/A_{\mathrm{GB}}$	3.52	2.36	2.32	1.85	1.16	0.537	0.337	0.246	0.187	0.15

(a)

GB	Ti 27b	Ti 9	Ti 27a	Ti 5	Ti 7	Tw 5	Tw 25	Tw 13	Tw 85	Ti 3
$\Delta c_{\gamma'}/A_{\mathrm{GB}}$	9.23	8.11	8.1	5.13	3.48	2.2	1.43	1.14	0.58	0.42

(b)

GB	Ti 27b	Ti 27a	Ti 9	Ti 5	Ti 7	Tw 5	Tw 25	Tw 13	Tw 85	Ti 3
$\Delta c_{\gamma'}/A_{\mathrm{GB}}$	16.9	16.1	15.2	7.67	4.29	1.77	1.44	1.11	0.71	0.63

(c)

Table 6.3: *Maximum* values of $\Delta c_{\gamma'}/A_{\mathrm{GB}}$ after 5 ns at 1500 K for all GB systems in different environments: a) Ni-10 at.% Al, b) Ni-20 at.% Al and c) Ni-20 at.% Al + ppt. ‘Ti’ and ‘Tw’ denote tilt and twist GB respectively. The columns are sorted by decreasing max $\Delta c_{\gamma'}$ and A_{GB} stands for the GB area. ‘ppt’ denotes a pre-existing γ' precipitate. Values of $\Delta c_{\gamma'}/A_{\mathrm{GB}}$ are [$10^{-3}\cdot$ at.% Å^{-2}].

GB	Ti 27b	Ti 9	Ti 27b	Ti 5	Ti 7	Tw 13	Tw 25	Tw 5	Ti 3	Tw 85
$\Delta c_{\gamma'}/A_{\mathrm{GB}}$	0.46	0.43	0.31	0.26	0.2	0.013	0.012	0.01	0.008	0.006

(a)

GB	Ti 9	Ti 27b	Ti 27a	Ti 7	Ti 5	Tw 13	Tw 5	Tw 25	Ti 3	Ti 85
$\Delta c_{\gamma'}/A_{\mathrm{GB}}$	0.82	0.71	0.65	0.46	0.45	0.37	0.31	0.3	0.15	0.14

(b)

GB	Ti 27a	Ti 27b	Ti 9	Ti 5	Ti 7	Tw 13	Tw 25	Ti 3	Tw 5	Tw 85
$\Delta c_{\gamma'}/A_{\mathrm{GB}}$	0.83	0.81	0.73	0.58	0.46	0.35	0.24	0.15	0.14	0.13

(c)

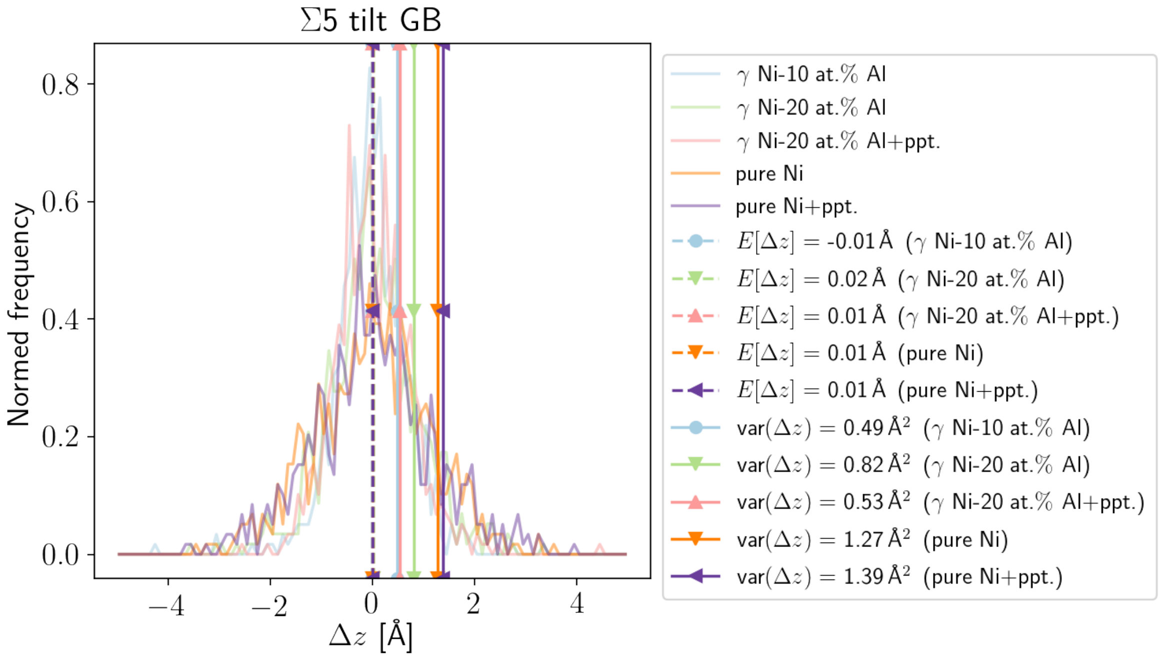
Table 6.4: *Median* $\Delta c_{\gamma'}/A_{\mathrm{GB}}$ after 5 ns at 1500 K for all GB systems in different environments: a) Ni-10 at.% Al, b) Ni-20 at.% Al and c) Ni-20 at.% Al + ppt. ‘Ti’ and ‘Tw’ denote tilt and twist GB respectively. The columns are sorted by decreasing max $\Delta c_{\gamma'}$ and A_{GB} stands for the GB area. ‘ppt’ denotes a pre-existing γ' precipitate. Values of $\Delta c_{\gamma'}/A_{\mathrm{GB}}$ are [$10^{-3}\cdot$ at.% Å^{-2}].

6.2.5 Impact of Al Solutes on GB Motion

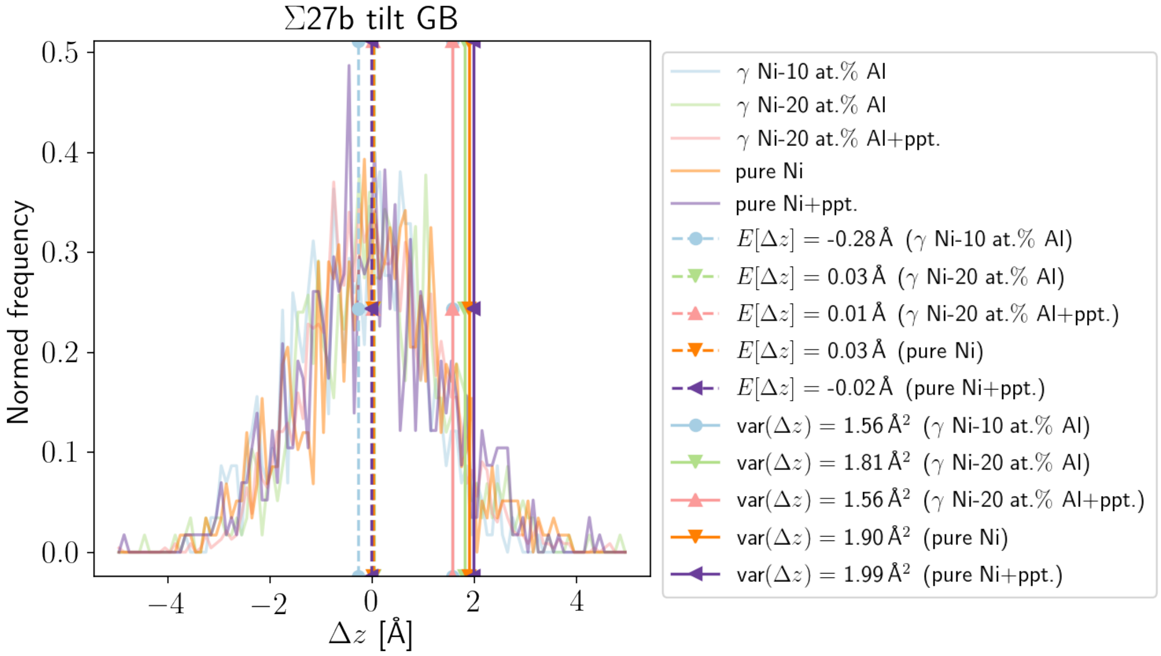
After finding that the degree of γ' ordering depends on the presence of the GBs we will discuss the impact of the chemical environment / solute concentration on the GB migration. For this we proceed in three steps. First, we investigate the displacement distribution of entire simulations, $p\left(\Delta z\right)$, for all trajectories of a GB type and for one chemical environment. For this we use, in particular, the mean, $\mathbb{E}\left[\Delta z\right]$, and variance, $\mathrm{var}\left(\Delta z\right)$, of $p\left(\Delta z\right)$. $\mathbb{E}\left[\Delta z\right]$ will indicate whether an overall directed movement of $\Delta z < 0$ or > 0 is observed. $\mathrm{var}\left(\Delta z\right)$, on the other hand, characterizes the size of the displacements, i.e. an increase in $\mathrm{var}\left(\Delta z\right)$ for one GB type with increased solute content would indicate an increasing mobility. Second, we will illustrate the simulation of GB motion using continuous time random walk (CTRW) models, in the form of variants of the Brownian motion and Levy walk models, focussing on the $\Sigma 27b$ tilt and $\Sigma 13$ twist systems. Third and lastly, we will analyse the GB displacement ensemble averages, $\left\langle \left(\Delta z\left(t\right)\right)^2 \right\rangle$, and timelag averages, $\overline{\delta^2\left(\Delta\right)}$, comparing $\overline{\delta^2\left(\Delta\right)}$ for all GB types in pure Ni and γ Ni-20 at.% Al.

Analysis of $\mathbb{E}\left[\Delta z\right]$ and var $\left(\Delta z\right)$

First one may ask what is the impact of the solutes on $\mathbb{E}\left[\Delta z\right]$ and $\mathrm{var}\left(\Delta z\right)$. In figs. 6.8 a), 6.8 b) , and 6.9 both are illustrated as vertical lines alongside $p\left(\Delta z\right)$ for $\Sigma 5$ tilt, $\Sigma 27b$ tilt and $\Sigma 13$ twist GBs respectively. Comparing all GB types we find that $\mathbb{E}\left[\Delta z\right]$ seems to be mostly zero for all GBs with the exception of slight deviations for the $\Sigma 9$ tilt, $\Sigma 27b$ tilt (for γ Ni-10 at.% Al only), $\Sigma 13$ twist, $\Sigma 25$ twist and $\Sigma 85$ twist GBs implying some directionality. The largest $\mathbb{E}\left[\Delta z\right]$ are observed for the $\Sigma 13$ twist GB, certainly the result of the sustained directed movements shown previously in fig. 6.6 c) and d). For $\mathrm{var}\left(\Delta z\right)$ we find a mixed behaviour, with all GB types, except $\Sigma 13$ twist and $\Sigma 25$ twist, showing a decreasing value for $\mathrm{var}\left(\Delta z\right)$ with increasing c_{Al}. While the decrease is probably due to GB–solute interaction, the origin of the increase is less clear, but may be related to the γ' ordering, particularly on one side of the GB, as visible for the $\Sigma 13$ twist GB in fig.6.6 d).



(a)



(b)

Figure 6.8: $p(\Delta z)$ for a) Σ5 tilt and b) Σ27*b* tilt GBs for the compositions pure Ni, pure Ni + ppt., γ Ni-10 at.% Al, γ Ni-20 at.% Al and γ Ni-20 at.% Al + ppt. The first and second moments of $p(\Delta z)$ for each composition are also shown, as dotted and solid vertical lines, respectively. The colouring and markers indicate the chemical environment. ‘ppt’ denotes a pre-existing γ' precipitate.

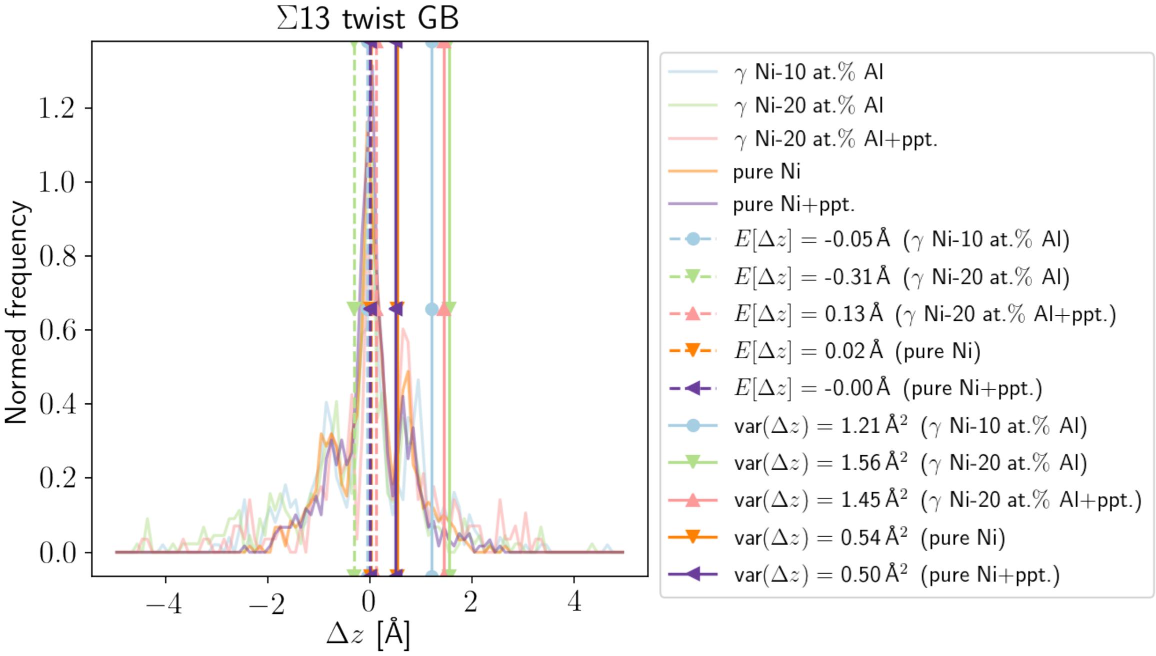


Figure 6.9: $p\,(\Delta z)$ for Σ13 twist GBs for the compositions pure Ni, pure Ni + ppt., γ Ni-10 at.% Al, γ Ni-20 at.% Al and γ Ni-20 at.% Al + ppt. The first and second moments of $p\,(\Delta z)$ for each composition are also shown, as dotted and solid vertical lines, respectively. The colouring and markers indicate the chemical environment. ‘ppt’ denotes a pre-existing γ' precipitate.

CTRW Models

The GB motion as observed in our MD simulations was modelled using variations of the methods based on Brownian motion and Levy walks, as described in section 2.3, optimizing with an evolution strategy based algorithm, called covariance matrix adaption (CMA) [94, 95], as implemented in *deap* [68]. In the following discussion, we will refer to the Brownian motion and Levy walk based mechanisms as continuous time random walks (CTRWs). Since we are trying to approximate the MD displacement distribution, $p\,(\Delta z)$, by displacement distributions from diffusion models simulating diffusion trajectories, $q\,(\Delta z)$, we use the Kullback–Leibler divergence $KL\,(q||p) = \int q\,(\Delta z) \ln \frac{q(\Delta z)}{p(\Delta z)} d\Delta z \geq 0$, with $KL\,(q||p) = 0$ if and only if $q\,(\Delta z) = p\,(\Delta z)$, which is not a distance metric since it is not symmetric, but converges to 0 from above once the approximating distribution $q\,(\Delta z)$ approaches the true distribution $p\,(\Delta z)$.

Both Brownian-like motions and Levy walks describe the displacement distribution, with the difference being that Levy walks are more heavily tailed than Brownian motion. We chose Brownian motion since it is commonly used to describe diffusion processes, although it implicitly assumes infinite displacement velocities and samples displacements from the normal distribution. Levy walks, on the other hand, can be generated by sampling the movement direction, velocity, and duration, which seems a plausible model for what is observed during the ‘running’ state of a GB as seen, for example, in fig. 6.6 d). The CTRW models and their modifications are given in table 6.5.

The regression then varies the parameters of the model. To approximate, on average, the MD displacement distribution, $p(\Delta z)$, by the displacement distributions of CTRW models, $q(\Delta z)$, each CTRW model parameter setting was used 100 times to generate 6 trajectories each (the same number of the GB trajectories) with 100 timesteps each. The mean of the resulting distributions was then compared with $p(\Delta z)$. The results of these regressions for the Σ27b tilt and Σ13 twist GBs in pure Ni and γ Ni-20 at.% Al are shown in figs. 6.10 – 6.12 and figs. 6.13 – 6.15 respectively. Figs. 6.10 and 6.13 show the best 4 CTRWs approximating the pure Ni case (black line). The shaded areas indicate one standard deviation whereas the dash-dotted lines represent the corresponding means of the approximations. The CTRWs are ranked in the legend in descending order of the value of $KL(q||p)$. We see that the $\Sigma 27b$ tilt GB in pure Ni is quite well approximated and could justifiably be characterized as a Brownian-like motion. But while the Σ13 twist GB is also well described by a Brownian-like motion, we see that there are secondary peaks which are not really captured. Since Levy walks were included as possible models, we interpret this as either (i) the optimization algorithm being unable to find useful parameter combinations for Levy walks which better approximate the distribution, or (ii) a combination of the mechanisms is actually active, with Brownian-like motion being the closer approximation.

Model / modifier	Distributions
model: Brownian	$\Delta z \sim \mathcal{N}\left(\mu, \sigma^2\right)$, $\Delta t \sim \Gamma(k, \theta)$
model: Levy walk	$\Delta z \sim \mathcal{N}\left(\mu, \sigma^2\right)$, $v \sim \Gamma(k_v, \theta_v)$ and $\Delta t \sim \Gamma(k_t, \theta_t)$ with new sampling of z and v once $v \cdot \Delta t \geq \Delta z$, direction $= \operatorname{sign}(\Delta z)$
modifier: discrete	$\Delta z \sim$ discrete $\left(N_{\text{discrete}}, \Delta z_{\text{spacing}}, \mu, \sigma^2\right)$
modifier: resting	'moving' ↔ 'resting' state transition: $\Delta z \sim \begin{cases} p_{\text{resting}\to\text{resting}}, & 0 \\ \text{o.w.}, & \mathcal{N}\left(\mu, \sigma^2\right) \end{cases}$
modifier: noisy	adding to Δz: $\Delta z_{\text{noise}} \sim \mathcal{N}\left(\mu_{\text{noise}}, \sigma^2_{\text{noise}}\right)$

Table 6.5: CTRW models as implemented for the approximation of MD GB trajectories. CTRW trajectories were generated with 100 iterations and 'observed' after discrete time intervals. Trajectories were generated combining one of the specified 'models' with 'modifiers'. The 'discrete' modifier allows N_{discrete} displacements symmetric around the origin, all separated by $\Delta z_{\text{spacing}}$. The 'resting' modifier includes a state transition using a coin toss with probabilities $p_{\text{resting}\to\text{resting}}$ or $p_{\text{moving}\to\text{moving}}$ and their complements depending on the previous state.

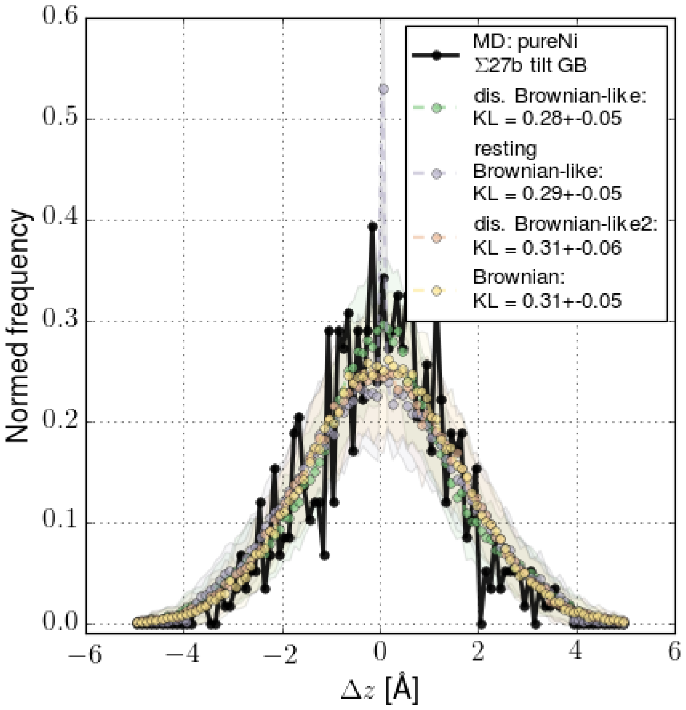
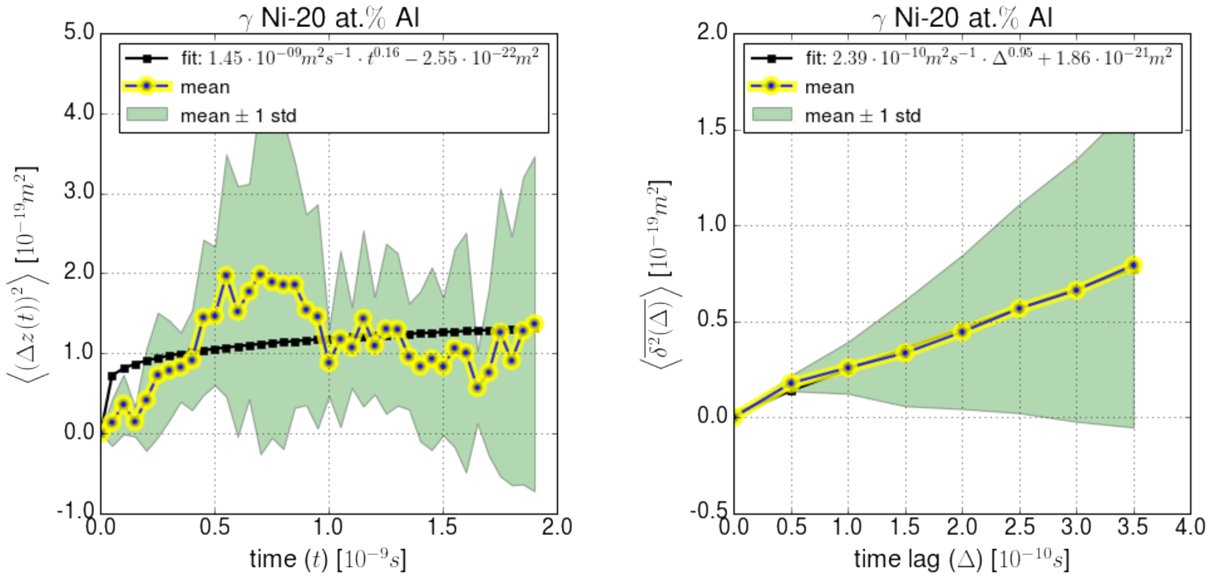


Figure 6.10: Σ27*b* tilt GB: the displacement distribution $p(\Delta z)$ for pure Ni (black) and modeled trajectories. The abbreviation 'dis.' refers to the 'discrete' modifier in table 6.5.



(a)

(b)

Figure 6.11: $\Sigma 27b$ tilt GB: the ensemble average for Ni-20 at.% Al, c) the time lag average for Ni-20 at.% Al.

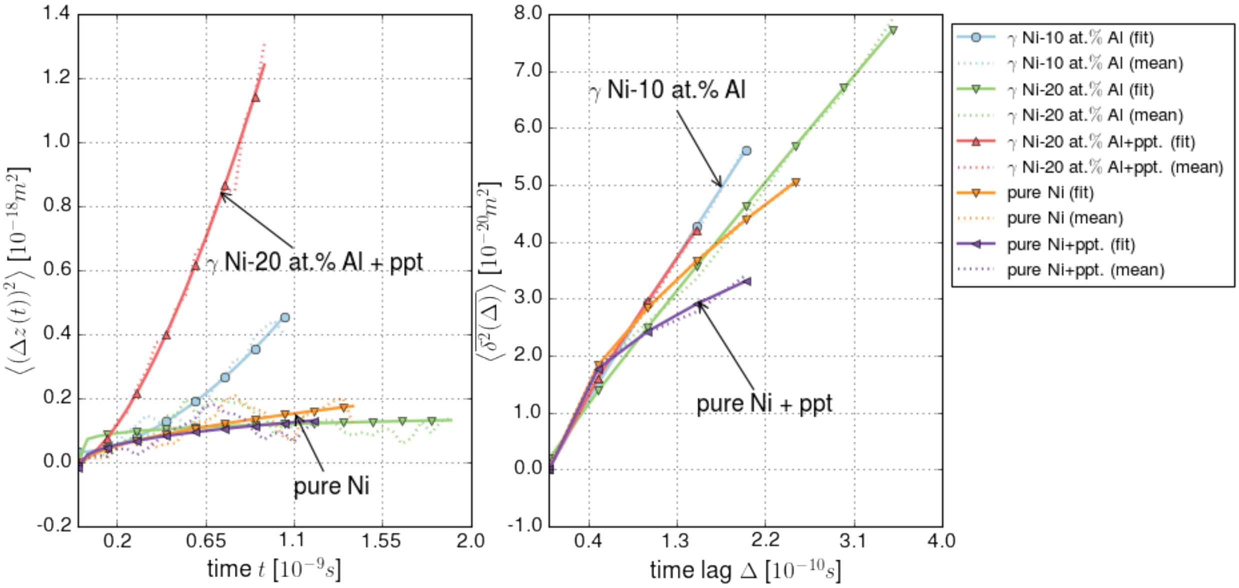


Figure 6.12: $\Sigma 27b$ tilt GB: fitted ensemble and time lag averages for all conditions.

Modelling all $p(\Delta z)$ for the different GB types we found that each one was best described with either ‘Brownian-like resting’ or ‘Brownian-like discrete’ models, except for the Σ3 tilt GB. This boundary barely moved and thus exhibited a very sharp distribution which the optimizer best described using a Levy walk. Note that although these models are characterized as ‘Brownian-like’, this does not mean that they cannot describe super- or sub-diffusive behaviour. We furthermore did not find a clear difference in the CTRWs due to the presence of solutes or a precipitate (ppt.), which may be because this is too subtle for the statistics of the trajectories available.

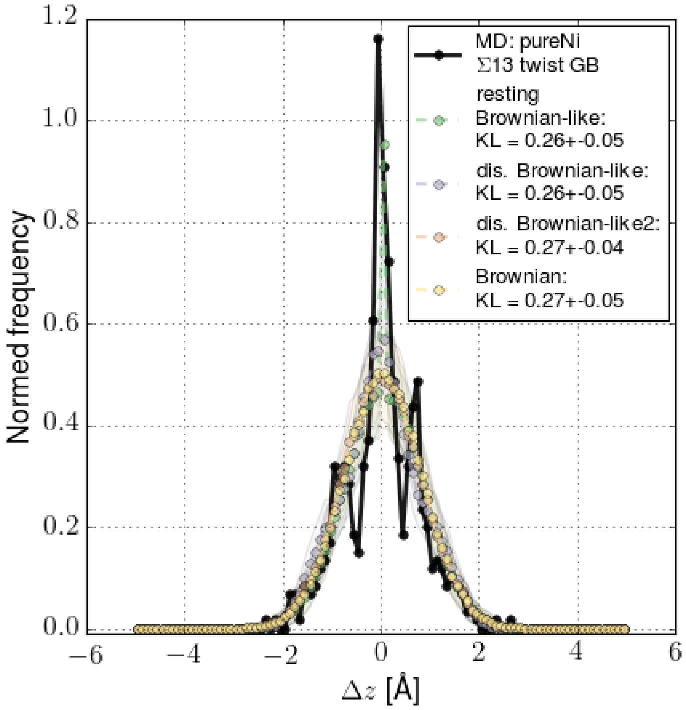
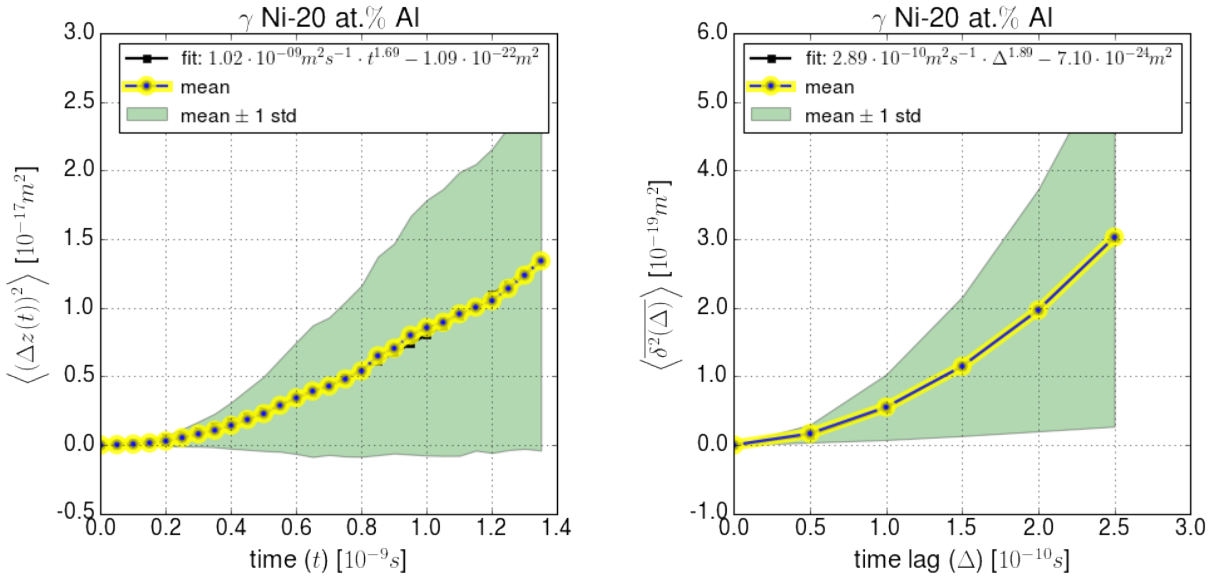


Figure 6.13: Σ13 twist GB: the displacement distribution $p(\Delta z)$ for pure Ni (black) and modeled trajectories. The abbreviation ‘dis.’ refers to the ‘discrete’ modifier in table 6.5.



(a) (b)

Figure 6.14: Σ13 twist GB: the ensemble average for Ni-20 at.% Al, c) the time lag average for Ni-20 at.% Al.

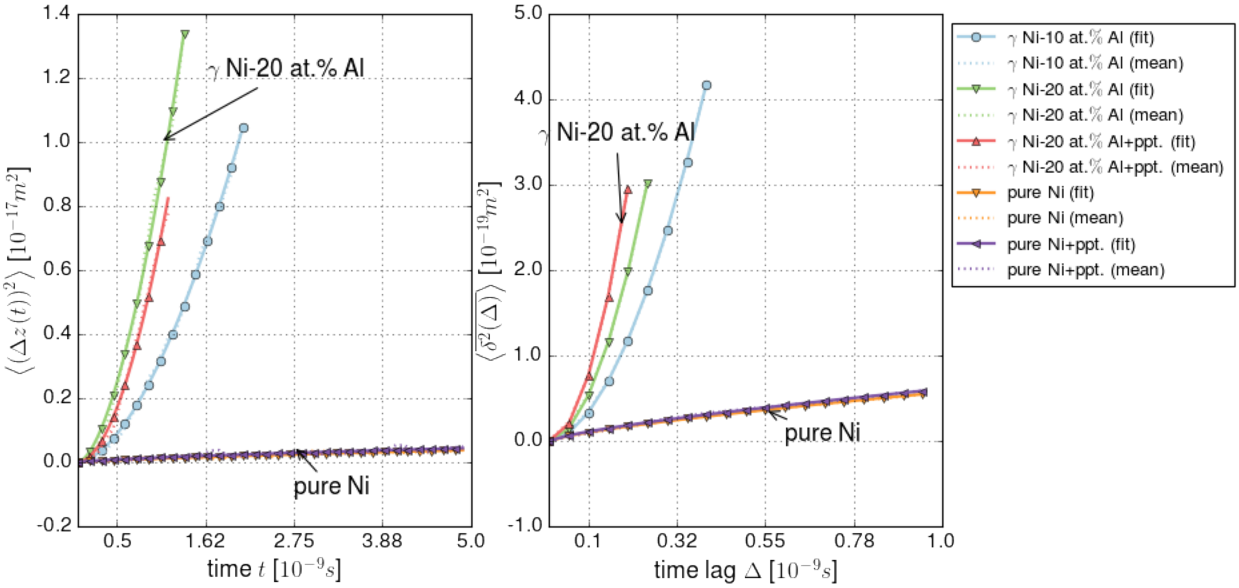


Figure 6.15: Σ13 twist GB: fitted ensemble and time lag averages for all conditions.

Analysis of Ensemble and Timelag Averages

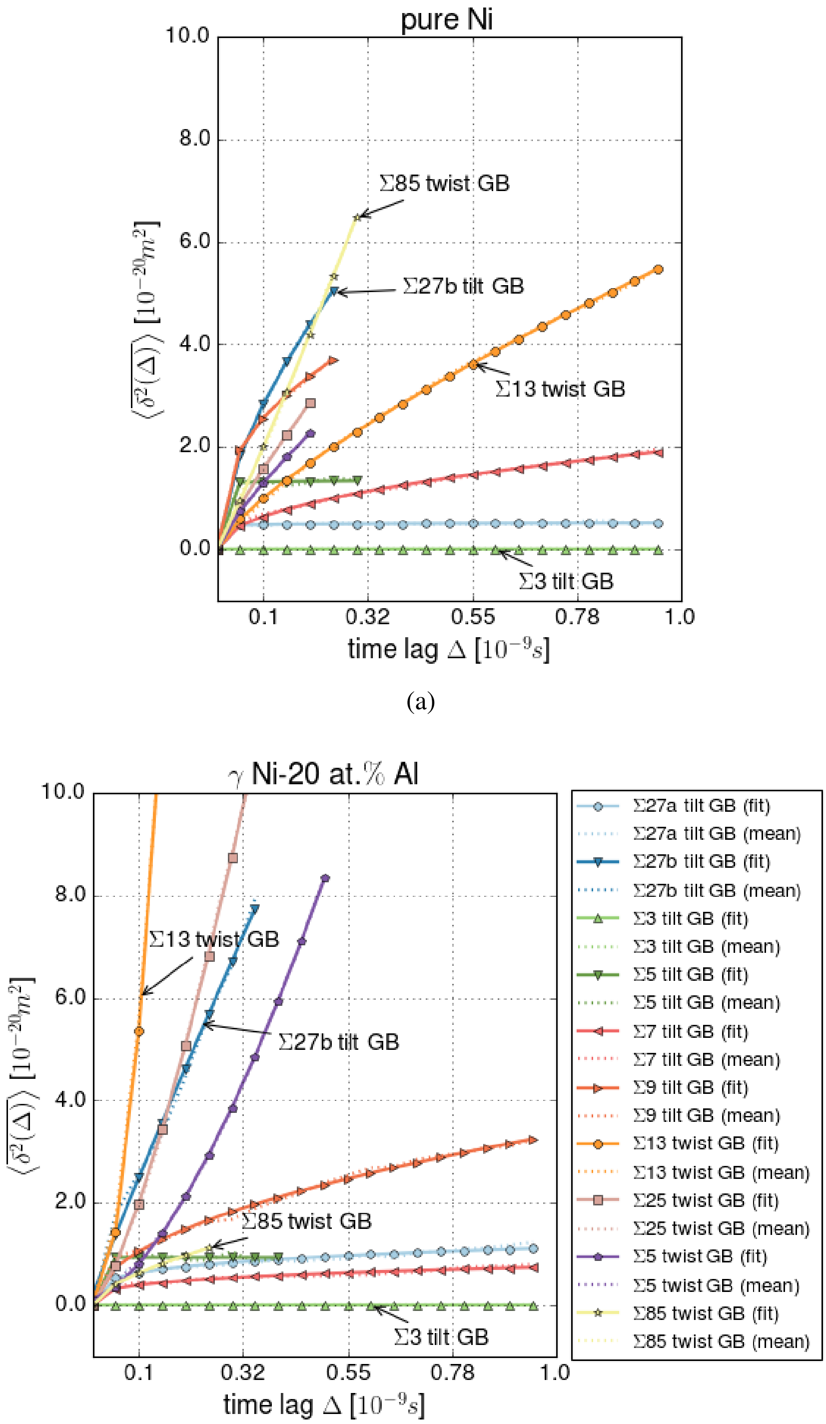
Another approach to jointly describe all GB trajectories for a given chemical condition and GB type is to use the ensemble, $\left\langle (\Delta z(t))^2 \right\rangle$, and timelag average $\overline{\delta^2(\Delta)}$, as introduced in section 2.3. The resulting functions may be straight, concave upwards, or concave downwards, and are then referred to as 'Brownian' / 'normal', 'super-diffusive', and 'sub-diffusive', respectively. Super- and sub-diffusive behaviour are observed when the ergodicity is broken one way or the other, for example, when there is an underlying drift (the displacement distribution of individual particles is not symmetric around zero due to some internal driving force) or aging occurs (decay of the variance of the displacement distribution). Since $\overline{\delta^2(\Delta)}$ is for a single trajectory, we will make use of $\left\langle \overline{\delta^2(\Delta)} \right\rangle$, the ensemble average of the the timelag averages for a specific GB type and chemical condition.

In figs. 6.11 and 6.14 a) and b) $\left\langle (\Delta z(t))^2 \right\rangle$ and $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ are shown for the $\Sigma 27b$ tilt and $\Sigma 13$ twist GBs respectively, in γ Ni-20 at.% Al environments. In these figures the mean (blue line with circles outlined in yellow) with one standard deviation indicated as the green shaded area are shown. Overlapping the average in both images is a black line with black square markers representing a fitted curve of the form $a \cdot t^b + c$ (for $\left\langle (\Delta z(t))^2 \right\rangle$) or $a \cdot \Delta^b + c$ (for $\left\langle \overline{\delta^2(\Delta)} \right\rangle$). We find that the $\Sigma 27b$ tilt GB obeys a normal / slightly sub-diffusive trend with mean $\left\langle (\Delta z(t))^2 \right\rangle$ being approximately straight and mean $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ being concave downwards. The $\Sigma 13$ twist GB, on the other hand, appears normal / weakly super-diffusive with both mean $\left\langle (\Delta z(t))^2 \right\rangle$ and $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ being slightly concave upwards. The means obtained for all the chemical environments are shown in figs. 6.12 and 6.15. We find that the $\Sigma 27b$ tilt GB follows a sub-diffusive or Brownian trend whereas the $\Sigma 13$ twist GB shows an increase in slope which looks like a transition from sub-diffusive to super-diffusive for increasing c_{Al}. Note the $\left\langle (\Delta z(t))^2 \right\rangle$ and $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ here gave partially conflicting information. In the following interpretation, we focus more on $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ because more steps were available than trajectories, i.e. for good statistics with $\left\langle (\Delta z(t))^2 \right\rangle$ one may want hundreds or thousands of trajectories. Also, the trajectories obtained from the MD were partially interrupted due to the linking process, which is why not all curves shown have the same length.

Summarizing how the chemical composition affects the characteristics of the migration of a single GB type, we observed the following behaviour with increasing c_{Al}: increasingly sub-diffusive ($\Sigma 3$ tilt, $\Sigma 5$ tilt, $\Sigma 7$ tilt, $\Sigma 9$ tilt), slightly sub-diffusive becoming increasingly Brownian ($\Sigma 27a$ tilt, $\Sigma 27b$ tilt), increasingly super-diffusive ($\Sigma 13$ twist, $\Sigma 25$ twist), slightly sub-diffusive except for γ Ni-20 at.% Al ($\Sigma 5$ twist), and Brownian with changing slope ($\Sigma 85$

twist).

Fig. 6.16 shows $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ as a function of the time lag Δ for all GB types in two ‘extreme’ chemical environments: a) pure Ni and b) γ Ni-20 at.% Al. We see that solutes mostly lead to a decrease in $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ for a given Δ for most GB types (particularly $\Sigma 85$ twist and $\Sigma 9$ tilt), with the notable exceptions of the $\Sigma 13$ and $\Sigma 25$ twist GBs, which show a significant increase in slope. The $\Sigma 27b$ tilt GB on the other hand seems relatively unaffected, with approximately the same behaviour for γ Ni-20 at.% Al as for pure Ni. Listing the 4 types of GBs with the largest slopes, in decreasing order we find for γ Ni-20 at.% Al: $\Sigma 13$ twist, $\Sigma 25$ twist, $\Sigma 27b$ tilt and $\Sigma 5$ twist, whereas for pure Ni we have $\Sigma 85$ twist, $\Sigma 27b$ tilt, $\Sigma 9$ tilt and $\Sigma 25$ twist.



(a)

(b)

Figure 6.16: $\overline{\delta^2(\Delta)}$ for a) pure Ni and b) γ Ni-20 at.% Al comparing all studied types of GB. Each type is represented by its mean $\overline{\delta^2(\Delta)}$ and fits using $a \cdot \Delta^b + c$, both of which have a high overlap throughout. The fits are indicated for each type by colour and marker.

6.3 Discussion

In section 5.2, we presented a spline-based classification of single crystals to identify early stage γ' precipitates. In this chapter we expanded on this and applied EM-GMMs to obtain the distributions for individual 'phases' observed in given training trajectories, and used these distributions to distinguish γ from γ'-ordered configurations in bicrystals with GBs. In doing so we have observed a rich interplay between the GB motion, the solutes, and the γ' ordering, in a variety of CSL GB geometries.

6.3.1 Gaussian Mixture Model & Decomposition Algorithm

The proposed decomposition algorithm was applied successfully to decompose $p\left(\mathbf{x}|\text{Trajectory}\right)$ for multiple trajectories into sensible $p\left(\mathbf{x}|\text{Phase}\right)$ for a range of 'phases'. This approach could be generally useful for developing classifiers for a range of phases or effects provided that the chosen features in $\mathbf{x}$ are sufficient to decompose the desired phases / effects. For the approximation of $p\left(\mathbf{x}|\text{Trajectory}\right)$, we made use of multivariate Gaussians optimized using the EM algorithm. We think that choosing Gaussians was partly responsible for the classifier's poorly identifying γ' at low temperatures. This is because of the sensitivity of Gaussians to outliers. This poor low temperature performance may also be partly because of the training set bias, with the training set being explicitly designed to capture γ' and γ at high temperatures, since we aimed to carry out simulations at 1500 K and not 0.1 K. The resulting classifier found was therefore quite robust at high temperatures of 300 K to 1500 K with *F1* scores for γ' of at least 0.9, where *F1* scores of 1 correspond to perfect classification.

We also note that there is a slight bias towards γ' over γ, since the *F1* score and 'Probability score' are better for γ' than for γ. But the scores certainly are also affected by the assumption that all atoms in the simulation of the γ phase actually are not by chance in γ' configurations. This may not be true, because γ is by definition a disordered phase, which could by chance contain γ' ordered configurations.

An interesting side result is that all GBs were studied at high temperatures by tracking a single class, indicating that GBs at high temperatures become quite geometrically disordered, at least in terms of BOPs. However, inspecting the obtained classifications for 0 K GB configurations, we found a variety of different classifications in GBs, capturing differences in atomic neighbourhoods in great detail.

6.3.2 GB Motion and Phase Ordering

The presence of GBs in the simulations of this research consistently led to an increase in ordering and decrease in potential energy, $\Delta E < 0$, and supercell volume , $\Delta V < 0$, with respect to single crystals. The sole exceptions were the $\Sigma 3$ tilt GBs, which showed trends which were about the same as for the single crystals. Besides the $\Sigma 3$ tilt GBs, the ordering was found to increase over time, particularly in the vicinity of GBs, forming thin slabs of ordered areas along the GB, particularly for tilt GBs. While the tilt GBs were mostly stationary, the twist GBs exhibited sustained directed movement for at least ~2 ns into the simulations, see fig. 6.6. This sustained directed movement of twist GBs was accompanied by an accelerated decrease in the potential energy and the volume of the supercell due to an accelerated ordering in the wake of the GBs, see fig. 6.3 and 6.4, leading to an overall higher concentration of γ'-like configurations in twist GB systems compared to tilt GB systems, see fig. 6.7.

Since there are no external driving forces in the simulations performed in the present study, the GBs were driven solely by a reduction in their free energy with respect to elastic misfit strains and supersaturation. This is reflected in decreases in the potential energy for the simulated γ Ni-20 at.% Al GB systems during 5 ns of annealing, decreasing by as much as 0.05 eV atom^{-1} (for the $\Sigma 13$ twist GB).

All GBs simulated in this research exhibited some motion, with the exception of the coherent $\Sigma 3$ tilt/twin GB annealed at 1500 K. This motion, even in pure Ni, can be explained in terms of GB roughening above the GB-specific roughening temperature. Simulating $\Sigma 5$ [001] symmetric and asymmetric tilt GBs, Olmsted *et al.* [162] found that the asymmetric tilt GB roughened at lower temperatures than the symmetric GB. This is reasonable since the symmetric GB is a more stable configuration, which would explain why the $\Sigma 3$ tilt GB in the present simulation was immobile, being the most stable of all GB configurations studied. In fact, investigating a series of GB types using atomistic simulations and an artificial driving force, Olmsted *et al.* [163, 164] found a zero mobility for coherent $\Sigma 3$ tilt/twin GBs.

We have analysed the impact of the solute concentration on the GB migration using $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ and have found the behaviour to be complex and dependent on the type of GB. However, it can be roughly summarized using the slope of $\left\langle \overline{\delta^2(\Delta)} \right\rangle$ which tends to: i) decrease for tilt GBs with increasing solute level (normal to sub-diffusive), but ii) increase for twist GBs with increasing solute level (normal to super-diffusive). A particularly strong increase was observed for $\Sigma 13$ twist GBs. Sub-diffusive behaviour can be understood in terms of an aging behaviour where the variance of the displacement distribution is time dependent and decreases over time, so that the GBs are increasingly immobile. This could be caused by an increasingly symmetric strain field. Super-diffusive behaviour, on the other hand, can be caused by heavy-tailed displacement distributions [150] or an underlying drift.

Trajectories which fit to an underlying drift have been observed by others using physical and artificial external driving forces, unlike in this thesis research, and have there been referred to as 'stick-slip' and 'locked'/'running' behaviour, among others, [51, 204, 244]. Analysing simulations of symmetric and inclined GBs with AIRwalk, Deng *et al.* [51] found that the GBs transitioned between a 'locked' and a 'running' state as a function of temperature and artificial driving force. They found that the artificial driving force required to transition from the 'locked' to the 'running' state decreased with increasing temperature. However, since the GBs in this work appear to be driven by internal forces it remains to be shown how anomalous diffusion relates to the 'ballistic' regime precisely.

Furthermore, Race *et al.* [182] and Hadian *et al.* [88] have shown that the GB area itself also plays a key role in simulated GB mobilities, demonstrating that the mobility is linked to the ability of the GB to nucleate steps or kinks. This fits into the observations of others claiming that GB migration involves correlated atom displacements in several stages with an incubation time for at least one of the stages involved [114, 244, 271–273]. Mechanisms reported using atomistic simulations are 'string-like' [271, 272] and rotational shuffling [114] for some tilt and twist GBs respectively. However, the 50 ps intervals between frames in the current simulations were found to be too large to adequately identify the underlying mechanisms.

Dislocations in the GBs were only identified for the $\Sigma 25$ and $\Sigma 85$ twist GBs in the form of a network of screw dislocations. GB profiles with local shapes resembling steps were observed for twist GBs, but this was not clear for tilt GBs. Since the GB area was larger in the simulations for twist GBs than for tilt GBs, a larger observed mobility for the former would be consistent with the concept of step nucleation.

Modelling the behaviour of the GB motion with a range of CTRWs we found that tilt GBs, without regard to their Σ value, were best described by 'resting' or 'discrete' Brownian-like motion. While the best approximations for twist GBs were also with Brownian-like motion particularly for $\Sigma 5$, $\Sigma 13$ and $\Sigma 25$, apparent second order peaks in $p(\Delta z)$ were not captured, although they could have resulted from a combination of migration mechanisms or possibly an artifact.

Also, the presence of the pre-existing precipitate did not have a clear effect, which may have been too subtle due to the size and separation of the precipitate from the GB as well as an insufficient number of repeat simulations.

6.3.3 Relevance for GB Serration

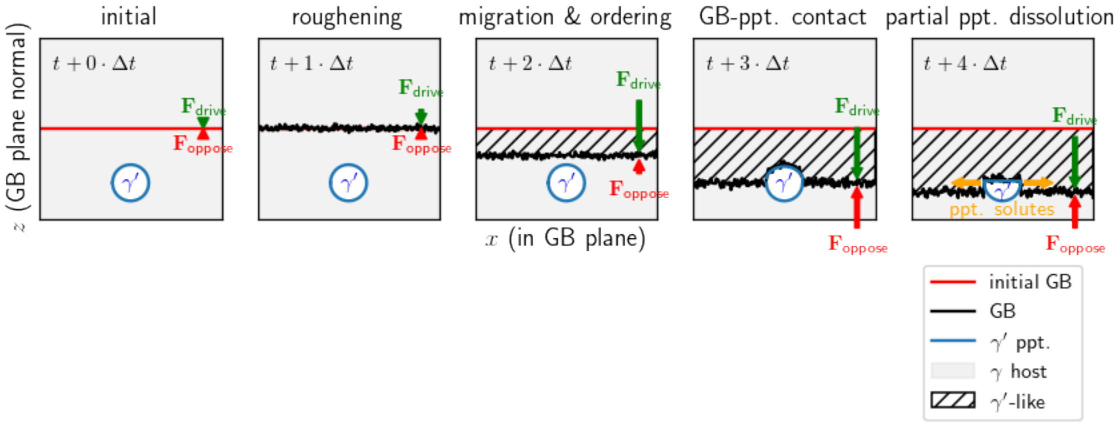


Figure 6.17: Schematic illustration of a GB and a spherical γ' precipitate ('ppt.') embedded in a γ host (grey background) representing the overall driving (green) and opposing (red) internal forces, $\mathbf{F}_{\text{drive}}$ and $\mathbf{F}_{\text{oppose}}$, as a function of time. The time advances from the leftmost image to the rightmost image by Δt as indicated in the top left of each image. The red line represents the initial position of the GB, whereas the black line represents the current position of the GB. The accelerated γ' ordering is indicated by the '/' hatched area. Upon encountering the precipitate, the GB starts to dissolve it, which is indicated as a flattening of the blue circle representing the precipitate. The orange arrows indicate the direction in which the solutes move after they are released from the precipitate. The length of the force arrows indicates the strength. Initially the forces are 'symmetric', having the same magnitude ($F \equiv \|\mathbf{F}\|$), but become 'asymmetric' when F_{drive} exceeds F_{oppose}.

In these simulations, GB migration was observed without pre-existing precipitates, as shown in fig. 6.6 d). The presence of pre-existing precipitates also did not clearly alter the GB migration direction. Hence precipitates, at least in the current systems, did not impose a clear driving force on the GBs. However, a significant driving force was found to be due to supersaturation, leading to ordering and GB migration, depending on the GB type. A noteworthy effect was observed for the $\Sigma 13$ twist GB in γ Ni-20 at.% Al (which included a precipitate), where one GB migrated towards the precipitate and was temporarily locked, see fig. 6.5 b). Note that pre-existing precipitates were not observed to consistently attract their adjacent GBs.

A schematic illustration of this effect involving GB roughening, migration, γ' ordering, and GB–precipitate interaction is shown as a time series in fig. 6.17. The slowing down of the

GB encountering the precipitate is reasonable since the size of the simulation is too small to allow the GB to completely bow around the precipitate. Furthermore the precipitate is in a more stable configuration than the γ Ni-20 at.% Al host, therefore making dissolution unfavourable. Thus, one way to imagine the GB–γ' interaction is to use a GB displacement distribution, which captures all possible displacements in the next timestep and their probabilities. Then a net driving force 'biases' the distribution, e.g. by changing its mean. When GB migration across a volume of γ' becomes less likely compared to migration across γ, then there is a driving force on the GB. The net driving force would be a time dependent function of the driving and opposing forces as illustrated in fig. 6.17. Hence, initially, when the system is highly symmetric, relative to the GB plane, the magnitudes of both internal forces F_{drive} and F_{oppose} are equal and the forces are thus symmetric. The resulting displacement distribution is then centred at the origin. However, once the GB roughens, leading to γ' ordering in the vicinity of the GB, the symmetry of the internal forces is broken, they become asymmetric, and the displacement distribution becomes biased, resulting eventually in movement of the GB. This movement occurs only provided the GB has sufficient mobility.

Hence one could imagine that on a larger length scale, the net driving force varies in sign along the GB, due to ordering on either side, with the sequence in fig. 6.17 representing a smaller section along the GB. This varying sign of the net driving force could lead to the migration of the GB in two directions. If this bidirectional movement is sustained for longer times, one could observe the initiation of GB serration followed by preferential precipitation in the GB's wake leading eventually to fan-like γ' precipitates or γ' 'fans' via discontinuous precipitation or coagulation [47, 176].

DIGM is also observed when there is an asymmetric driving force on GBs due to solute diffusion along them leading to asymmetric coherency strain fields and (de-)alloying free energy gradients [121, 181]. DIGM can produce wavy GBs and arguably occurs alongside discontinuous precipitation. The latter effect has been claimed by Danflou *et al.* [46] in the case of GB serration which leads to γ' fans. They also reported an increase in GB serration with increased grain misorientation and no GB serration for coherent $\Sigma 3$ twins. As GB serrations have been observed to become more wavy with decreased cooling rate [156], so have GBs due to DIGM at higher annealing temperatures [143].

Hence we think that the super-diffusive behaviour observed in our simulations driven by the ordering process could reasonably be described as DIGM with the already observed heterogeneous ordering becoming heterogeneous / discontinuous precipitation thus causing fan-like precipitate structures to appear, as observed by Mitchell *et al.* [156] and Qiu *et al.* [176]. The selection of GB types with low and high mobilities should therefore enable the distribution and degree of serrated GBs to be controlled. For a more thorough understanding, further study is required to improve the statistics and to include asymmetric and general GBs as well as artificial driving forces and various GB areas.

6.4 Conclusion

In this chapter we have:

1. extended a spline based classifier method by developing the ‘decomposition algorithm’ to identify γ, γ' and GB phases,
2. simulated a range of CSL GB bicrystals under several alloying conditions at 1500 K with MD and observed significant γ'-ordering in a supersaturated γ Ni-20 at.% Al host, particularly around and in the wake of the GBs,
3. observed GB migration which could be quantitatively classified into normal or sub-diffusive behaviour for tilt GBs and normal or super-diffusive behaviour for twist GBs with increasing Al content,
4. explained super-diffusive GB migration behaviour as a consequence of heterogeneous ordering, which itself could result in heterogeneous / discontinuous precipitation at longer times with the subsequent serration of GBs.

Chapter 7

Empirical Potential Training

The achievement of the work of Hohenberg and Kohn with their famous theorems was to show that the ground state electron density, $\eta_{GS}(\mathbf{r})$, is of fundamental importance for the description of condensed matter [102]. These theorems (see Chapter 3) provide the basis for density functional theory (DFT). Although DFT has come a long way since 1964, the treatment of materials on the basis of electrons beyond thousands of atoms is computationally still prohibitive and thus requires approximations for the interatomic interactions referred to as empirical potentials.

Commonly empirical potentials do not try to explicitly capture $\eta_{GS}(\mathbf{r})$ but a variety of other properties, such as potential energies and forces. In order to systematically develop empirical potentials which contain two or more elements, noting the importance of $\eta_{GS}(\mathbf{r})$ in DFT, we decided to include $\eta_{GS}(\mathbf{r})$ in the process of developing empirical potentials. In section 7.1 we will show how the problem of approximating $\eta_{GS}(\mathbf{r})$, in form of the valence electron densities obtained from the DFT codes CASTEP [41] and Profess [101], can be interpreted as a linear regression problem [201]. The linear models are formulated using per atom contributions in form of 2 and 3-body terms. Using this approach we were able to predict $\eta_{GS}(\mathbf{r})$ with reasonable accuracy for high symmetry structures under different strains and for a database of the technologically important Al-Ni alloy system, which included a range of different phases and deformations.

For high symmetry structures we use atomic configurations of phases not containing thermal distortions or crystal defects. The range of high symmetry structures spans Al (fcc, bcc, hcp), single walled carbon nanotubes, α-W & W_3O, graphite, a H_2-crystal, NaCl and $BaTiO_3$. In section 7.2 we will make use of the 2-body approximations for an alloy database of Al-Ni to develop Embedded Atom Method (EAM) potentials reproducing energies and forces to within 0.01 meV atom^{-1} and 0.1 eV Å^{-1} respectively.

The work contained in this chapter is the result of a collaboration with Andrew Fowler of the Macro Molecular Materials Group, Department of Materials and Metallurgy, University of Cambridge. The code which forms the basis of this chapter was developed jointly. While Andrew initially suggested the regression of $\eta_{GS}(\mathbf{r})$ itself and performed the Profess calculations, the author of this thesis developed the approach of approximating $\eta_{GS}(\mathbf{r})$ using Bayesian linear regression and carried out the CASTEP calculations and steps to develop the EAM potentials presented in this work.

7.1 Electron Density Regression

Even though $\eta_{GS}(\mathbf{r})$ is crucial to DFT it has not received much attention in the literature as an object of interest itself to be modelled. To the best of the author's knowledge only two recent works by Brockherde *et al.* [30] and Mitev *et al.* [157] incorporate explicit modelling of $\eta_{GS}(\mathbf{r})$. However in the original EAM papers [48, 49] empirically mixed analytical s and d orbital functions were used, which were obtained from Clementi *et al.* [42], thus introducing an additional fitting parameter which was interpreted as effective s-like electron content.

Complementary to the present work, where parametric linear models are applied, Brockherde *et al.* [30] used a non-parametric approach is which weights are 'learnt' for basis function representations of $\eta_{GS}(\mathbf{r})$ for H_2, H_2O and small hydrocarbons. To this end they use an 'external potential' in the form of squared exponentials $v(\mathbf{r}) = \sum_i Z_i e^{-\|\mathbf{r}-\mathbf{r}_i\|^2/2\gamma}$ with γ, a free parameter, and Z_i the nuclear charge for atom i, capturing neighbourhoods at $\mathbf{r}$. The $\eta_{GS}(\mathbf{r})$ is then developed in terms of that external potential $v(\mathbf{r})$.

Also using a parametric model, albeit a non-linear one, Mitev *et al.* [157] regressed s and p type atomic orbitals. They used these regressed 'orbitals' as input for a further modification to represent local bonding properties in functional form which was used as the embedding density function $\rho(r)$ for an EAM potential for Al.

The approach we will follow here is based on Bayesian linear regression to obtain functions from $\eta_{GS}(\mathbf{r})$. This is accomplished by decomposing $\eta_{GS}(\mathbf{r})$ into 2-body and 3-body contributions. The developed method can be extended beyond 3-body contributions but was not found to be necessary here because 3-body contributions were found to be versatile enough to capture even highly anisotropic densities such as those present in graphite or carbon nanotubes.

In order to estimate the usefulness of our proposed $\eta_{GS}(\mathbf{r})$ regression method we will, in the following, quantify the accuracies of our approximations to $\eta_{GS}(\mathbf{r})$ by studying the ability to simultaneously describe fcc, bcc and hcp Al, computing density and OF DFT energy-based errors using Profess. This is followed by an application of the method to a range of phases computed with CASTEP and the inspection of 2 and 3-body functions for the γ' Ni_3Al phase in particular.

7.1.1 Bayesian Linear Regression of $\eta_{GS}(\mathbf{r})$

In the following we discuss how the design matrix $\mathbf{\Phi}$ for $\eta_{GS}(\mathbf{r})$ can be set up to regress electron densities using parametric linear models. Since the Bayesian treatment of regression provides a theoretical background to regularization and allows for model comparison, as discussed in section 4.3, it is used here as our method of choice. The relevance vector machine (RVM) [235] stood out among the Bayesian linear regression methods due to its sparsity, see section 4.3.2 for further details, and will be discussed here as well.

At the core of using linear regression for $\eta_{GS}(\mathbf{r})$ is the *assumption* that $\eta_{GS}(\mathbf{r})$ can be sufficiently approximated in the form of a linear superposition of atom contributions, with 2-body (1 atom & 1 density point), 3-body (2 atoms & 1 density point) up to n-body ($n-1$ atoms & 1 density point). This assumption is a core part of the EAM, where embedding densities are computed using 2-body contributions. This assumption of linear superposition is quite strong but nonetheless it will be shown that it produces quite accurate predictions for various phases in high symmetry configurations. This assumption of linear superposition allows the re-ordering of the sum of 2- and 3-body contributions for individual density points allowing for the application of linear regression in the first place. An advantage of describing $\eta_{GS}(\mathbf{r})$ in this way is that the resulting functions can later be used, for example, for empirical potentials such as EAM, where the 2-body approximations can used as the embedding density functions. To predict $\eta_{GS}(\mathbf{r})$ by atom contributions we assume that $\eta_{GS}(\mathbf{r}) = y(\mathbf{x},\mathbf{w}) + \varepsilon$ holds, where $\mathbf{x}$ is the information about neighboring atoms at $\mathbf{r}$ (such as distance vectors and atom species), $\mathbf{w}$ the model parameters and $\varepsilon \sim \mathcal{N}(0,\beta)$ the prediction error/noise. β is the noise level. This becomes a linear regression problem assuming $y(\mathbf{x},\mathbf{w}) = \mathbf{\Phi}\cdot\mathbf{w}$ with $\mathbf{\Phi}$ the *design matrix*. In the following we will show how $\mathbf{\Phi}$ needs to be set up to allow for linear regression to be applied.

Generally we can write the predictive electron density function $\eta_{\text{pred}}(\mathbf{r})$ as an m-body function

$$\eta_{\text{pred}}(\mathbf{r}) = \underbrace{\sum_{i=1}^{L} f_i(\Delta\mathbf{r}_i)}_{\text{2-body}} + \underbrace{\sum_{i=1}^{L}\sum_{j>i}^{L} g_{i,j}\left(\Delta\mathbf{r}_i, \Delta\mathbf{r}_j\right) + ...}_{\text{3-body}} \tag{7.1}$$

with $\Delta\mathbf{r}_i \equiv \mathbf{r} - \mathbf{r}_i$, L the number of neighbouring atoms considered and f_i and $g_{i,j}$ some 2- and 3-body functions respectively. In order to maintain translation and rotation invariance of $\eta_{\text{pred}}(\mathbf{r})$ we use the magnitude r_i of $\Delta\mathbf{r}_i$ for the 2-body approximation and $\left(r_i, r_j, \theta_{ij}\right)$ for the 3-body approximation with $\theta_{ij} = \cos^{-1}\left(\frac{\Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j}{r_i r_j}\right)$. Neglecting 4-body and higher order terms we find

$$\eta_{\text{pred}}(\mathbf{r}) \approx \underbrace{\sum_{i=1}^{L} \eta_i^{\text{2-body}}(r_i)}_{\rho^{\text{2-body}}} + \underbrace{\sum_{i=1}^{L}\sum_{j>i}^{L} \eta_{i,j}^{\text{3-body}}\left(r_i, r_j, \theta_{ij}\right)}_{\rho^{\text{3-body}}}. \tag{7.2}$$

$\rho^{\text{2-body}}$ and ρ^{3-body} are the predicted 2- and 3-body components for $\eta_{\text{pred}}(\mathbf{r})$. With these approximations we are one step closer to forming our design matrix. Now we must choose a form for $\eta_i^{\text{2-body}}(r_i)$ and $\eta_{i,j}^{\text{3-body}}\left(r_i, r_j, \theta_{ij}\right)$. For both we choose a cosine series expansion:

$$\eta_i^{\text{2-body}}(r_i) = \sum_{k=0}^{M_{\text{el},i}-1} w_{\text{el},k} \cos\left(\frac{\pi k}{r_{\text{cut}}} r_i\right) \tag{7.3}$$

and

$$\eta_{i,j}^{\text{3-body}}\left(r_i, r_j, \theta_{ij}\right) = \sum_{k_0,k_1,k_2=0}^{M_{\text{el},i}-1} w_{\text{el},k_0,k_1,k_2} \cos\left(\frac{\pi k_0}{r_{\text{cut}}} r_i\right) \cos\left(\frac{\pi k_1}{r_{\text{cut}}} r_j\right) \cos\left(k_2 \theta_{ij}\right) \tag{7.4}$$

with k up to $M_{el,i} - 1$ and r_{cut} the cut-off distance.

Because the re-arrangement of the terms of the 2-body and 3-body parts to construct $\mathbf{\Phi}$ are almost identical we demonstrate it for the 2-body part only. Inserting eq. 7.3 into eq. 7.2 we find the outer sum to be the sum over the neighbouring atoms and the inner sum to be the sum over neighbouring element dependent basis function terms. Re-arranging this by grouping the neighbouring information by elements for the 2-body component leads to

$$\rho^{\text{2-body}} = \sum_{\text{el}\in\text{element types}} \sum_{k=0}^{M_{\text{el}}-1} w_{\text{el},k} \underbrace{\sum_{i\in\text{neigh(el)}} \cos\left(\frac{\pi k}{r_{\text{cut}}} r_i\right)}_{\equiv \phi_k(\mathbf{x}_{\text{el}})}. \tag{7.5}$$

We thus recover the 'linear' model

$$\rho^{\text{2-body}} = \sum_{\text{el}\in\text{element types}} \mathbf{w}_{\text{el}}^{T} \cdot \boldsymbol{\phi}\left(\mathbf{x}_{\text{el}}\right). \tag{7.6}$$

The non-linear mapping occurring here with $\boldsymbol{\phi}$ is therefore a combination of the basis functions and the sum over neighbouring atoms of specific species. Having recovered the linear form, we can directly apply any linear regression method, such as RVMs. The same derivation holds for approximations of higher order than 2-body. An interesting result of this way of setting up the design matrix is that the design matrix is partitioned into blocks corresponding to specific elements for the 2-body approximation and different degrees of n-body approximations. Once the design matrix is set up one can obtain the weights via a method such as RVMs. Because the design matrix is partitioned the corresponding weights vector is also partitioned encoding the 2-body, 3-body and higher order functions for the different species and species pairs. In this section and the empirical potential section, we denote tapering of a function and its derivatives above point c with the notation $y^{(k)}(x \geq c) = 0$. Tapering below c is indicated analogously. To perform the tapering we use $\psi(\tilde{x}) = \frac{\tilde{x}^4}{1+\tilde{x}^4}$ with $\tilde{x} = \frac{x-c}{0.1}$,where x will be the spatial distance or embedding density.

In order to cope with periodic boundary conditions we create 'ultracells' of the original crystal structures as input from the DFT calculations for both the density approximation and the training of empirical potentials. This is done using Algorithm 7.1 developed in the work for this purpose. By generating an 'ultracell' the algorithm assures that the neighbourhood searched for a given point in space really includes all possible neighbours, which is not necessarily the case when the cut-off radius exceeds the size of the supercell.

Algorithm 7.1 Algorithm to generate an ultracell for neighborhood searches in a periodic simulation box. In order to generate an ultracell the given cell/supercell is copied and its copy displaced. The algorithm is written with fractional coordinates in mind for generality. The aim is to form the ultracell as a sphere around the given cell/supercell such that each point in the cell/supercell has all required periodic neighbors available given a cutoff distance r_{cut}.

```
procedure ULTRACELL(atoms, $r_{\text{cut}}$, cell, maxiter)
    center ← center of cell
    $r_{\text{search}} \leftarrow (\max(\text{distance}(\text{cell corners} - center)) + r_{\text{cut}}) \cdot 1.5$
    allowed moves ← $[-1,0,1] \times [-1,0,1] \times [-1,0,1]$
    start ← starting displacements, i.e. [0,0,0]
    accepted ← start                    ▷ displacements to realize for new positions
    test ← start                        ▷ known positions to check
    saturated ← $\emptyset$             ▷ positions for which all displacements are checked
    while not exceeding maxiter do
        possible ← combination of all test positions and allowed moves which are not yet in saturated
        possible ← filter possible for positions within $r_{\text{search}}$
        if possible == $\emptyset$ then
            exit loop
        end if
        accepted ← add possible
        saturated ← add test
        test ← $\emptyset$
        test ← add possible
    end while
    ultracell atoms ← $\emptyset$
    for cell position ∈ accepted do
        ultracell atoms ← add atoms for cell position
    end for
    return ultracell atoms
end procedure
```

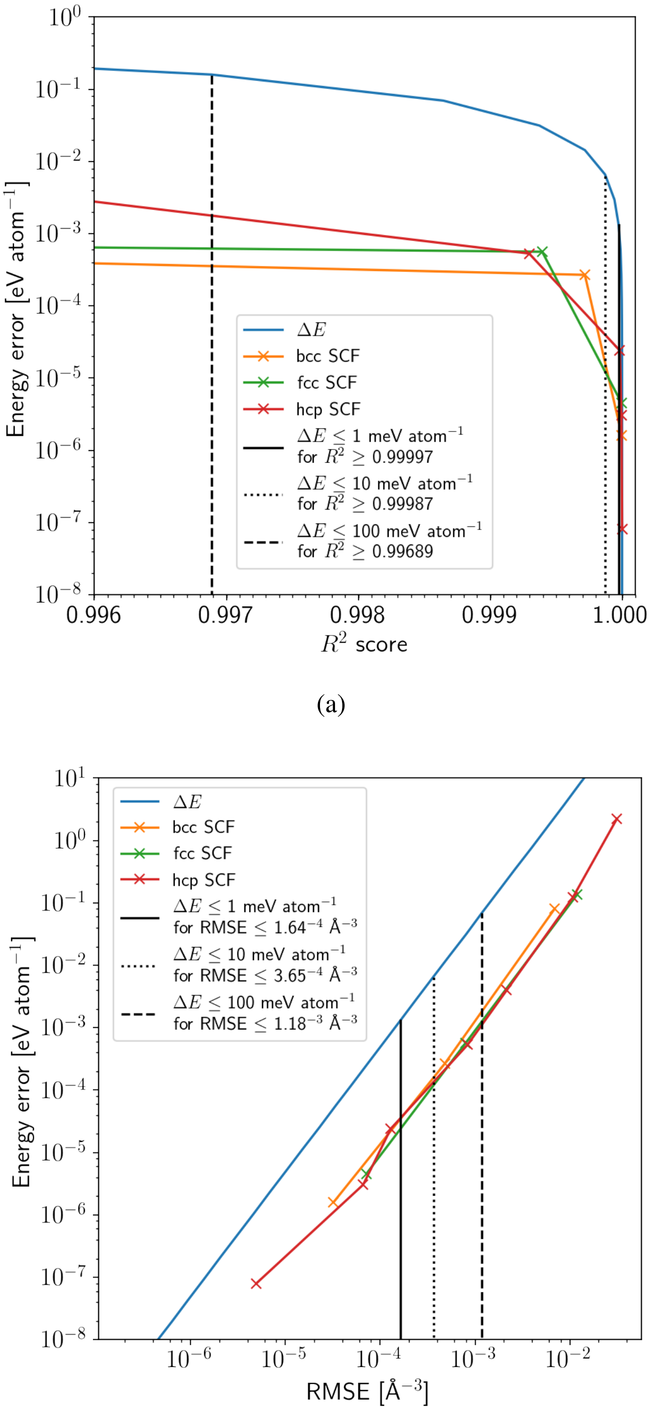
7.1.2 Results

In this section we will first present two measures for quantifying the accuracy of the 2- and 3-body approximations obtained from a joint regression of Al phases at different strains and then applying the density regression to a variety of other phases. We will illustrate the 2- and 3-body approximations on the given phase and show that 2-body approximations provide quite accurate descriptions of ionic systems but have difficulties with covalent bonds. As can be expected, these more anisotropic effects in the electron density are captured more accurately employing 3-body approximations.

Quantification of Approximation Accuracy

The accuracy of the chosen approximation is investigated using three measures: 1) the energy difference in OF DFT energies between $\eta_{GS}(\mathbf{r})$ and $\eta_{\text{pred}}(\boldsymbol{r})$, $\Delta E = E\left[\eta_{\text{pred}}(\mathbf{r})\right] - E\left[\eta_{GS}(\mathbf{r})\right]$, 2) the ‘coefficient of determination’ or R^2 score$= 1 - \frac{\text{var}(\eta_{\text{GS}} - \eta_{\text{pred}})}{\text{var}(\eta_{\text{GS}})}$ where $\text{var}\left(\eta_{\text{GS}} - \eta_{\text{pred}}\right)$ is the prediction error variance and $\text{var}\left(\eta_{\text{GS}}\right)$ the sample variance and 3) the $\text{RMSE}\left(\eta_{\text{GS}} - \eta_{\text{pred}}\right)$. The RMSE, R^2 score and ΔE quantities are chosen because ΔE directly relates to physically important regions of the electron density surface which are mostly contributing to the energy, the R^2 score because it directly relates to the distribution of the electron density values as a whole and provides a unit-less measure between 1 for a perfect fit and $-\infty$ and the RMSE because it is a default absolute measure of distance. The R^2 score and RMSE allow the comparison of approximation accuracies between different phases for which the electron density based energy cannot be directly computed. By the end of this section we establish the $\Delta E\left(\text{R}^2\right)$ and $\Delta E\,(\text{RMSE})$ bounds using Profess for the OF DFT calculations on Al which are then used to interpret the density errors obtained for a range of phases using CASTEP.

To establish these bounds we choose fcc Al and compute $E\left[\eta(\mathbf{r})\right]$ for different $\eta(\mathbf{r})$. We used two methods to obtain the different $\eta(\mathbf{r})$, one directly using the intermediate densities, $\eta_{\text{SCF}}(\mathbf{r})$, produced during the SCF cycles of Profess itself and the other by an artificial distortion of $\eta_{GS}(\mathbf{r})$, $\eta_{\text{normal}}(\mathbf{r})$, with density deviations sampled from a normal distribution. We use a normal distribution because our model for the $\eta_{GS}(\boldsymbol{r})$ already assumes a normal noise, as outlined at the beginning of section 7.1.1. Thus we use $\eta_{\text{normal}}(\mathbf{r}) = \eta_{GS}(\mathbf{r})\left(1 + f \cdot \mathcal{N}(0,1)\right)$ with f being varied between 0.001 and 0.01 to simulate different noise levels. The resulting bounds are shown in fig. 7.1 a) for the R^2 score and b) for the RMSE, with $\eta_{\text{normal}}(\mathbf{r})$ as the blue line, $\eta_{\text{SCF}}(\mathbf{r})$ for bcc Al, fcc Al and hcp Al and the dashed and dotted black lines indicating the ΔE for a given R^2 score, e.g. $\Delta E \leq 10$ meV atom^{-1} for an R^2 score ≥ 0.99972 or RMSE $\leq 3.65\ 10^{-4}$ Å^{-3} (black dashed line). Furthermore we observe two effects, 1) ΔE is highly sensitive to the R^2 score and RMSE for both $\eta_{\text{normal}}(\mathbf{r})$ and $\eta_{\text{SCF}}(\mathbf{r})$ and 2) $\eta_{\text{normal}}(\mathbf{r})$ has larger ΔE for the same R^2 score / RMSE than $\eta_{\text{SCF}}(\mathbf{r})$. The reason for this difference in ΔE for effect 2) is likely due to the prioritization during the SCF cycles which optimizes $\eta(\mathbf{r})$ preferentially in regions with significant gradients and lead to large energy contributions. Hence $\eta_{\text{random}}(\mathbf{r})$ with random perturbations of $\eta_{GS}(\mathbf{r})$ leads to non-smooth electron density fields and derivatives and therefore to larger energy differences. This interpretation is supported by the impact of the *error* in the density on the resulting distributions of density derivatives, which are shown in fig. 7.2, fig. 7.3 and fig. 7.4 for the 0th, 1st and 2nd order density derivatives respectively. In these figures the colour of the distributions for all derivatives indicate the R^2 scores of the 0th derivative, using the same colour coding for all three figures. Comparing these three figures one can see that the distributions of the density errors for the 1st and 2nd derivatives are larger for $\eta_{\text{normal}}(\mathbf{r})$ than for $\eta_{\text{SCF}}(\mathbf{r})$ explaining why $\eta_{\text{normal}}(\mathbf{r})$ leads to higher energies for the same R^2 score as $\eta_{\text{SCF}}(\mathbf{r})$. Since the linear regression models are likely to lead to smoother $\eta_{\text{pred}}(\mathbf{r})$ than $\eta_{\text{normal}}(\mathbf{r})$ we think it is reasonable to use the $\Delta E\left(\text{R}^2\right)$ and $\Delta E\,(\text{RMSE})$bounds for $\eta_{\text{normal}}(\mathbf{r})$ and $\eta_{\text{SCF}}(\mathbf{r})$ as upper and lower bounds respectively.



(a)

(b)

Figure 7.1: Relationships between of the orbital free DFT energy deviation, $\Delta E = E\left[\eta_{\text{normal/SCF}}(\mathbf{r})\right] - E\left[\eta_{GS}(\mathbf{r})\right]$ and a) the R^2 score and b) the root mean squared error (RMSE), due to deviation of the given electron density. η_{normal} (blue) represents energy deviations for artificially distorted $\eta_{GS}(\mathbf{r})$ for fcc Al and is shown alongside energy deviations during SCF cycles for bcc Al (orange), fcc Al (green) and hcp Al (red). The solid, dashed and dotted lines indicate energy-density error regions based on the η_{normal} curve with $\Delta E \leq 1$ meV atom^{-1}, $\Delta E \leq 10$ meV atom^{-1} and $\Delta E \leq 100$ meV atom^{-1} in both a) and b).

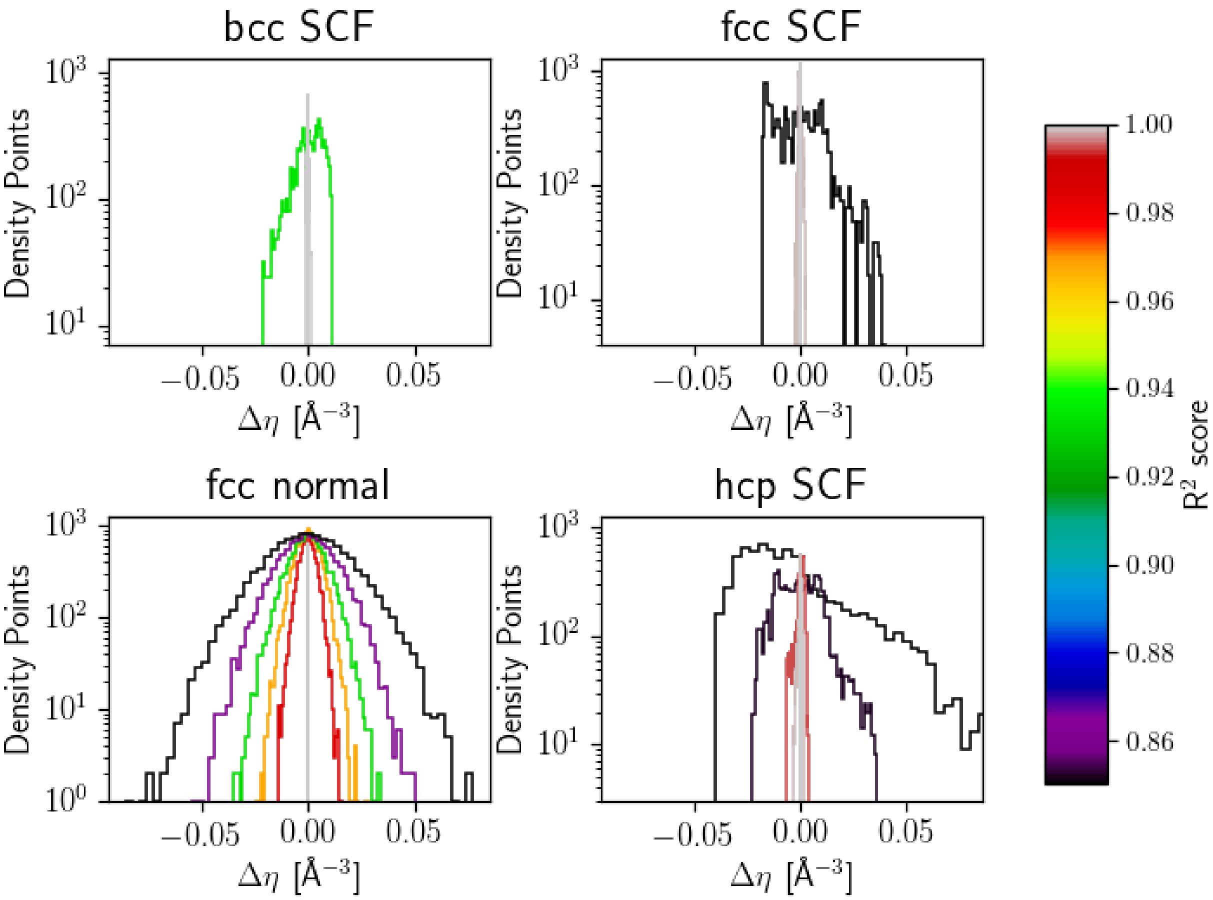


Figure 7.2: Histograms of *0th* derivative values of electron density points $\eta_{\text{SCF}}(\mathbf{r})$ and $\eta_{\text{normal}}(\mathbf{r})$ at different degrees of perturbation from $\eta_{GS}(\mathbf{r})$. Shown are the results for the different phases of Al. The histograms are color coded with colorbar shown to the right assigning R^2 scores to the respective degree of perturbation.

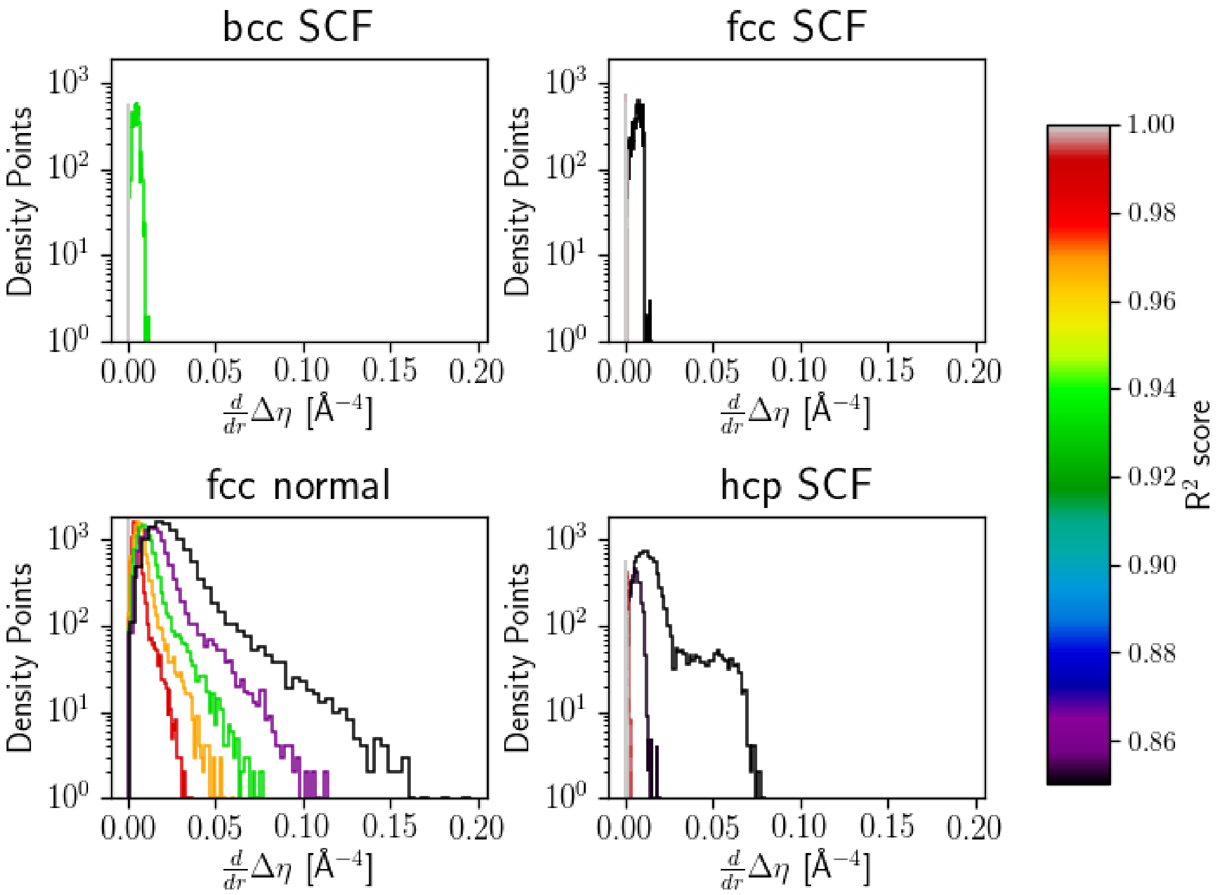


Figure 7.3: Histograms of *1st* derivative values, $\|\nabla_{\mathbf{r}}\eta(\mathbf{r})\|_2$, of electron density points $\eta_{\text{SCF}}(\mathbf{r})$ and $\eta_{\text{normal}}(\mathbf{r})$ at different degrees of perturbation from $\eta_{GS}(\mathbf{r})$. Shown are the results for the different phases of Al. The histograms are color coded with colorbar shown to the right assigning R^2 scores to the respective degree of perturbation.

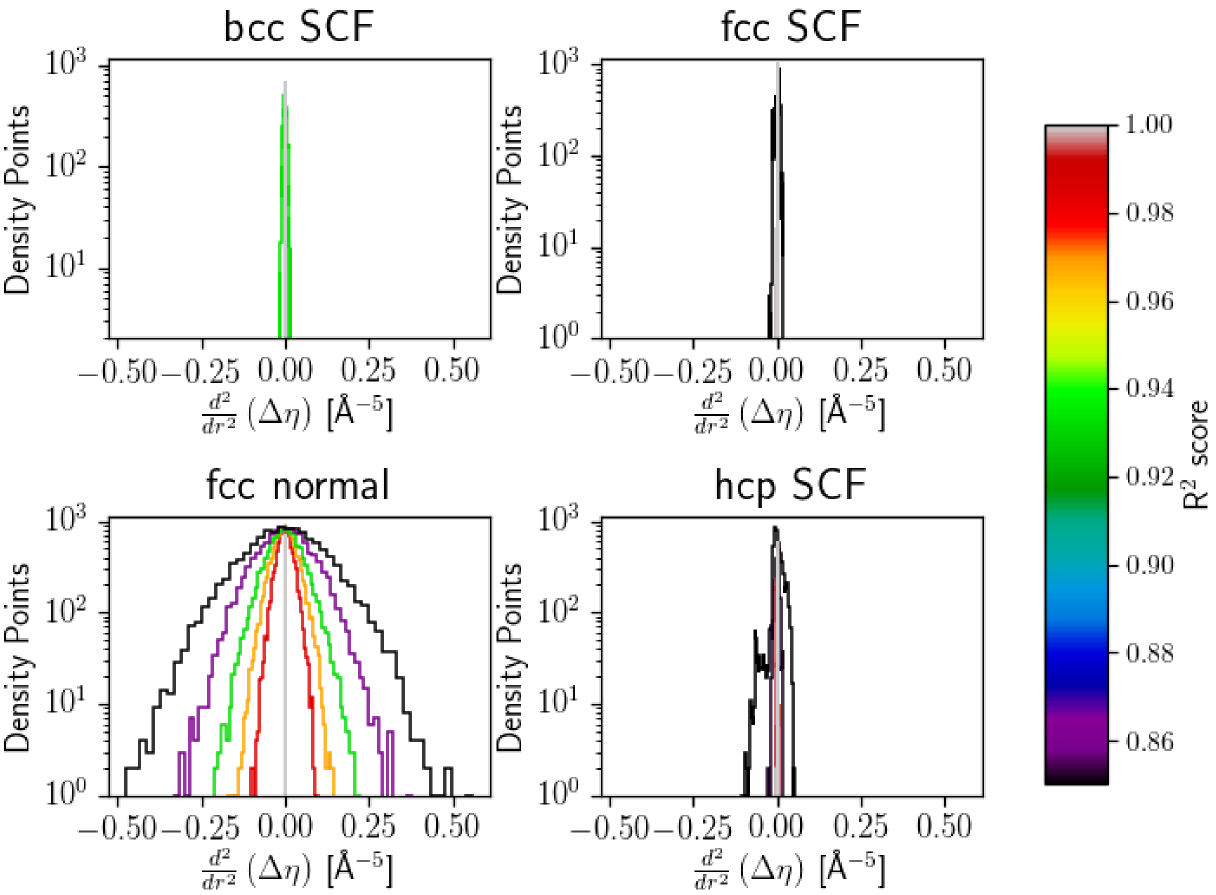


Figure 7.4: Histograms of *2nd* derivative values, trace$(\nabla_{\mathbf{r}}\nabla_{\mathbf{r}}\eta(\mathbf{r}))$, of electron density points $\eta_{\text{SCF}}(\mathbf{r})$ and $\eta_{\text{normal}}(\mathbf{r})$ at different degrees of perturbation from $\eta_{GS}(\mathbf{r})$. Shown are the results for the different phases of Al. The histograms are color coded with colorbar shown to the right assigning R^2 scores to the respective degree of perturbation.

In order to verify the estimates for the ΔE bounds and to test whether $\eta_{GS}(\mathbf{r})$ of more than a single structure can be captured appropriately with this approach, we trained the 2- and 2&3-body approximations on high symmetry fcc Al, bcc Al and hcp Al structures describing all three phases simultaneously. The Al structures were deformed equiaxially between -1 and 1 %. To obtain our 2 and 2&3-body functions we split the data randomly into test and training set with 50% of the structures for test and training set. Calculating the OF DFT energy given the $\eta_{\text{pred}}(\mathbf{r})$ we find that the 2&3-body approximations match more closely the actual energy than the 2-body approximation alone, as can be seen in fig. 7.5, although the 2-body approximation itself is accurate to within ~50 meV atom^{-1}. The simultaneous approximation of fcc, bcc and hcp Al in this case seems to favour bcc Al. This is the result of taking equal numbers of density points for the training from each Al phase and is related to the resulting error distributions of the density derivatives, not shown here, which are slightly wider for fcc and hcp than for bcc. The number of density points taken from the phases was found to influence ΔE. ΔE could, for example, be improved for fcc Al by sampling more density points from fcc Al structures relative to the others. A more detailed comparison of R^2 scores and the median absolute energy difference and the maximum absolute energy difference are shown in tab. 7.1. There it can be seen that the $\Delta E \approx 1$ meV atom^{-1} rule of thumb for an R^2 score ≈ 0.99997 is still valid as, for example, for 'Fcc-2&3' which has a $\mathrm{MAE}_{\mathrm{Train}}$ of 2.086 meV atom^{-1} for an R^2 score of $\eta\left(\boldsymbol{r}\right)$ of 0.99998. We also note that the test and training set errors shown in tab. 7.1 are very similar and the obtained functions thus showing some transferability to hitherto unseen structures. Furthermore one can see that the R^2 scores worsen with increasing degree of differentiation with respect to $\boldsymbol{r}$, which is plausible since only $\eta_{GS}\left(\mathbf{r}\right)$ was provided as target data for the regression. Despite the derivatives not explicitly being part of the training, their R^2 scores are relatively high which we interpret as a result of the choice of smooth basis functions.

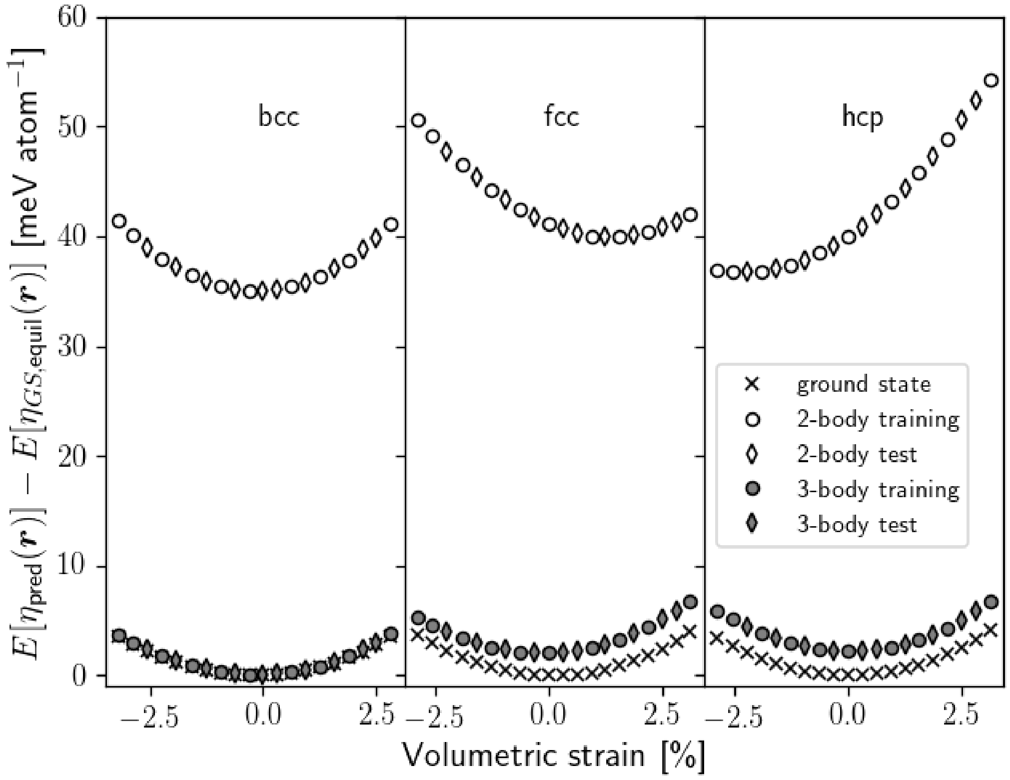


Figure 7.5: $E\left[\eta\left(\mathbf{r}\right)\right]$ for ground state (hatches), 2-body approximation (empty markers) and 2&3-body approximation (filled markers) electron densities versus equiaxial deformations between -3 and 3 % volumetric strain. Shown are energies for bcc, fcc and hcp Al. The functions for the approximations were trained over all three lattice types indicating the training set with circles and test set with diamonds.

	$\mathrm{R^2}_{\mathrm{Test}}$ $(\eta(\boldsymbol{r}))$	$\mathrm{R^2}_{\mathrm{Test}}$ $(\|\nabla_{\mathbf{r}}\eta(\mathbf{r})\|_2)$	$\mathrm{R^2}_{\mathrm{Test}}$ $(\triangle_{\mathbf{r}}\eta(\mathbf{r}))$	$\mathrm{MAE}_{\mathrm{Train}}$ $[\frac{\mathrm{meV}}{\mathrm{atom}}]$	$\mathrm{MAE}_{\mathrm{Test}}$ $[\frac{\mathrm{meV}}{\mathrm{atom}}]$	$\mathrm{MAX}_{\mathrm{Train}}$ $[\frac{\mathrm{meV}}{\mathrm{atom}}]$	$\mathrm{MAX}_{\mathrm{Test}}$ $[\frac{\mathrm{meV}}{\mathrm{atom}}]$
Fcc	0.998	0.994	0.97	42.057	40.989	47.044	45.523
Fcc-2&3	0.99998	0.9998	0.998	2.086	2.1684	2.8524	2.7529
Bcc	0.998	0.995	0.97	36.223	35.847	37.925	37.077
Bcc-2&3	0.999995	0.99999	0.997	0.1333	0.1106	0.3205	0.2293
Hcp	0.997	0.993	0.97	40.281	41.681	50.153	49.072
Hcp-2&3	0.99998	0.99996	0.998	2.3824	2.33	2.6489	2.5324

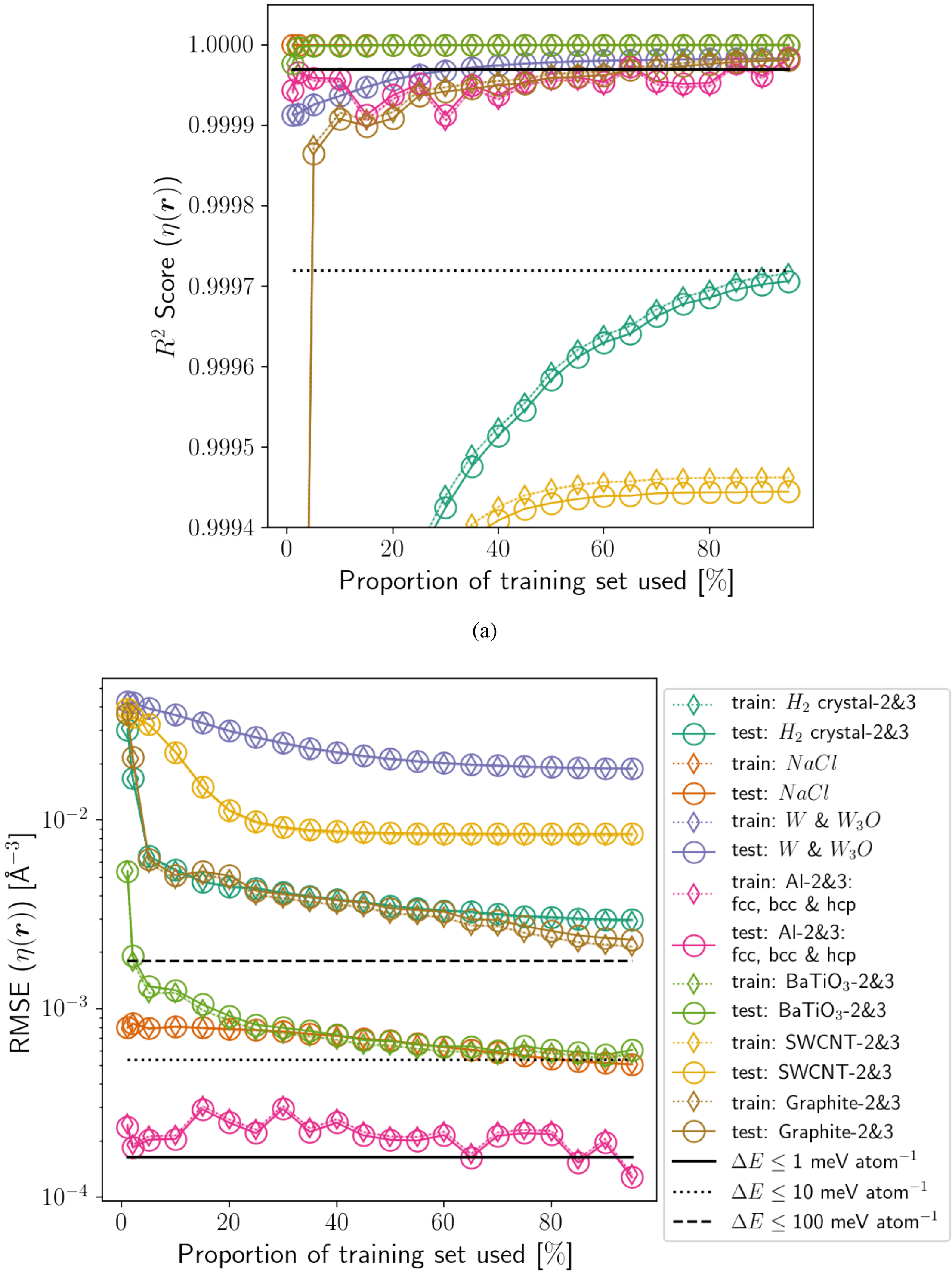
Table 7.1: Correlation of R^2 scores for 0th, 1st and 2nd derivative of $\eta_{\mathrm{pred}}(\mathbf{r})$ and OF DFT energies in the form of the mean absolute energy (MAE) and maximum absolute energy error (MAX) for 2-body and 2&3-body approximations to $\eta_{GS}(\mathbf{r})$. The rows contain 2- and 2&3-body approximations in alternating order for the three Al phases, with '-2&3' indicating the 2&3-body approximation.

Modelling Metallic, Ionic and Covalent Phases

Having established the approximate ΔE bounds for Al we will now test the regression method over a range of phases or groups of phases computed with CASTEP (also for the three Al phases). For all phases we consider as the neighbourhood of a single density point everything within 6 Å. This may not be ideal for every phase and thus allows for further optimization for specific phases of interest. One immediate question for the regression of $\eta_{GS}(\mathbf{r})$ is how severe the overfitting is and how the prediction depends on the size of the training set. This is answered in fig. 7.6 which shows a) the R^2 score in a) and the RMSE in b), both versus the percent of the training set utilized. Both the RMSE and the R^2 score are computed over all training and test set structures respectively for a group of structures, but only a percentage of the density points in the training set structures are actually used for the training.

The overfitting taking place seems to be minor as the training and testing errors are very similar, with the training error being only slightly better. Although not shown in fig. 7.6, we also note that the gain in using 3-body terms depends on the system, e.g. NaCl did not benefit much from 3-body terms whereas graphite did. Thus for phases like NaCl where $\eta_{GS}(\mathbf{r})$ is dominated by isotropic features, a simpler basis set is sufficient for reaching a good prediction error compared to phases with strongly anisotropic features. Hence the ability of the 2-body approximation to capture $\eta_{GS}(\mathbf{r})$ can be seen as an implicit way of judging the general anisotropy of $\eta_{GS}(\mathbf{r})$ of a given system. Comparing the different phases in fig. 7.6

a) we find that most crystals have an R^2 score of above 0.9999, including graphite, except the H_2-crystal and the single wall-carbon nanotubes (SWCNTs), which have R^2 scores of 0.9994 and 0.9997 respectively, at 95% utilization of the training set. We understand this difference for these two phases in terms of the larger vacuum present in the structures providing less information for our model, possibly requiring a change of the cut-off radius. Comparing fig. 7.6 a) and b) we find that only the Al system, NaCl and $BaTiO_3$ are within $\Delta E \leq 10$ meV atom^{-1}, whereas graphite, the H_2 crystals, the SWCNT and the tungsten system are in the range of hundreds of meV atom^{-1}, with the tungsten system having the largest RMSEs. This shift is due to the sample variance taking on different values, varying from about $9.9\,10^{-4}$ Å^{-3} for the Al systems to about 20.6 Å^{-3} for the tungsten system. This is reasonable since for the same RMSE but larger sample variance the R^2 score will be larger than for a smaller sample variance. An increase in atomic number and number of electrons present in the calculations was observed to lead to an increase in the variance in electron density, explaining the re-ordering. Also note that the tungsten system was regressed with 2-body terms only.



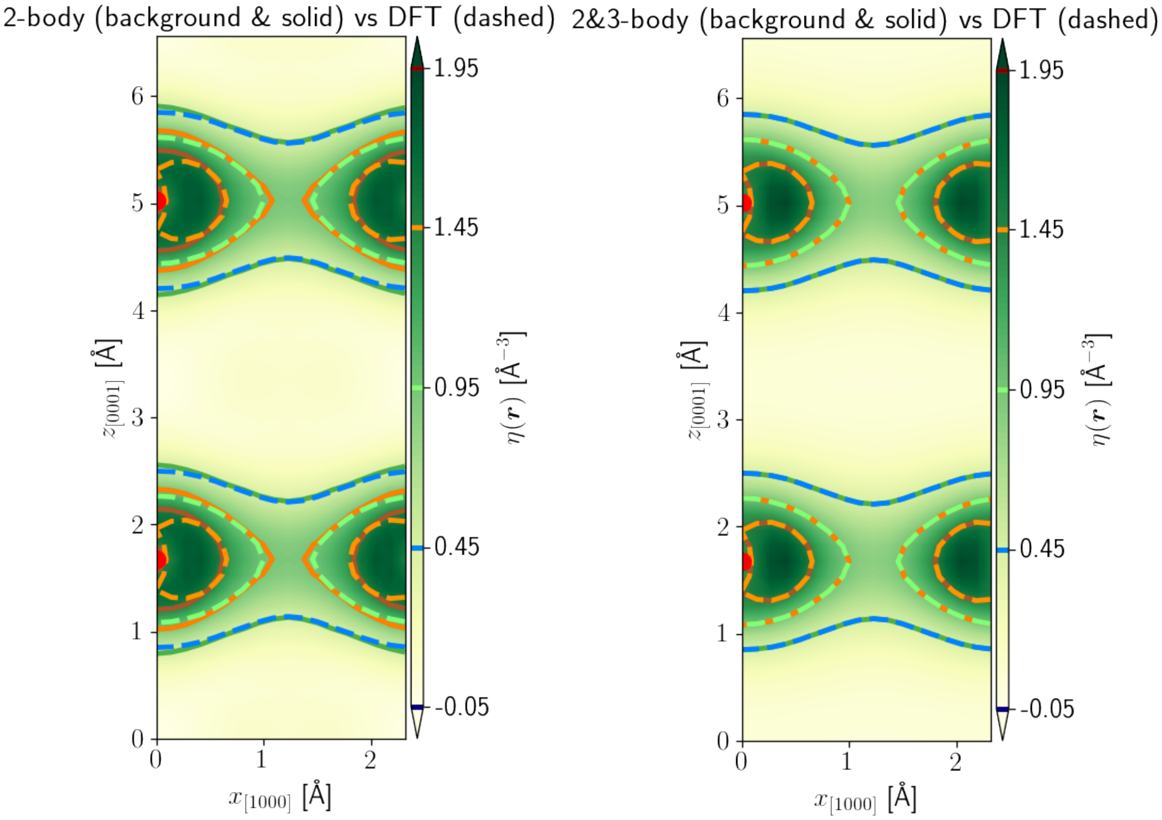
(a)

(b)

Figure 7.6: Learning curves of the model approximating $\eta_{GS}(\mathbf{r})$ for groups of strained crystals obtained with CASTEP. All structures are approximated with at least the 2-body approximation, 2&3-body approximations are indicated with '-2&3' where used. The accuracy of the trained functions are shown as a) R^2 scores and b) RMSEs versus the percentage of density points in the training set used. The R^2 scores and RMSEs for both test (circles) and training (diamonds) were computed over all density points in their respective sets. The black curves indicate the ΔE bounds from fig. 7.1.

In order to illustrate the interplay of anisotropic features and the 2- and 3-body approximations, cross-sections of $\eta_{GS}(\mathbf{r})$ and $\eta_{\text{pred}}(\mathbf{r})$ are shown for both graphite and the L1_2 (γ') Ni$_3$Al phase in fig. 7.7 and fig. 7.8 respectively. In these figures $\eta_{GS}(\mathbf{r})$ is shown as dashed lines with $\eta_{\text{pred}}(\mathbf{r})$ as solid lines and area shading. In both figures part a) shows the 2-body approximation whereas part b) shows the 2&3-body approximation using contour and solid lines for each. While for graphite the x-z-plane is shown for $y = 0$, the γ' phase is shown in the x-y-plane at $z = 1/4$ along the c-axis. The γ' phase exhibits a density profile which is closer to what one would see for ionic systems where the valence electrons are arranged spherically around the atom cores. Graphite on the other hand exhibits clear anisotropies. Interestingly enough we find that the 2-body approximations in fig. 7.7 a) and fig. 7.8 a) do capture the characteristics of $\eta_{GS}(\mathbf{r})$ and is missing only finer details.

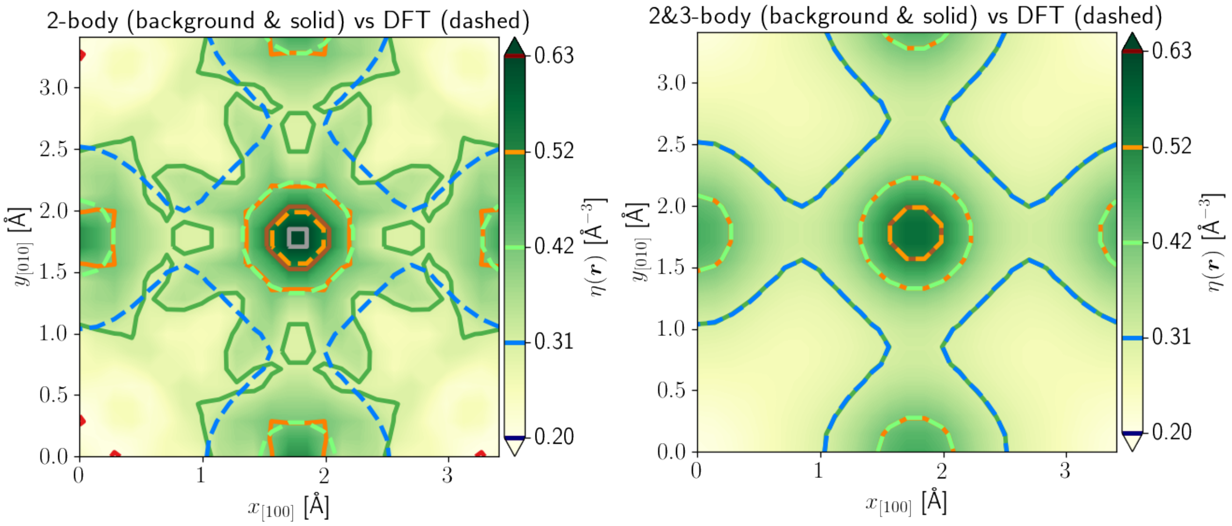
Examples of resulting 2-body and 3-body functions from our regression are shown in fig. 7.9 for the γ' phase, fitting either only with the 2-body approximation or with both 2&3-body. The functions are shown for both elements Al and Ni in fig. 7.9 a). One can see how the curves for the same element have similar characteristics. We interpret this different baseline with the underlying anisotropy of γ'. The 3-body function for γ' is shown in fig. 7.9 b) for $r = r'$ and $\theta \in [0, \pi]$. There we see that contributions of the 3-body function vary approximately between -2 and 4 Å^{-3} for $r \leq 5$ Å. The contribution above 5 Å vanishes, which is related to the distance based tapering, which is also visible for the 2-body functions in fig. 7.9 a).



(a)

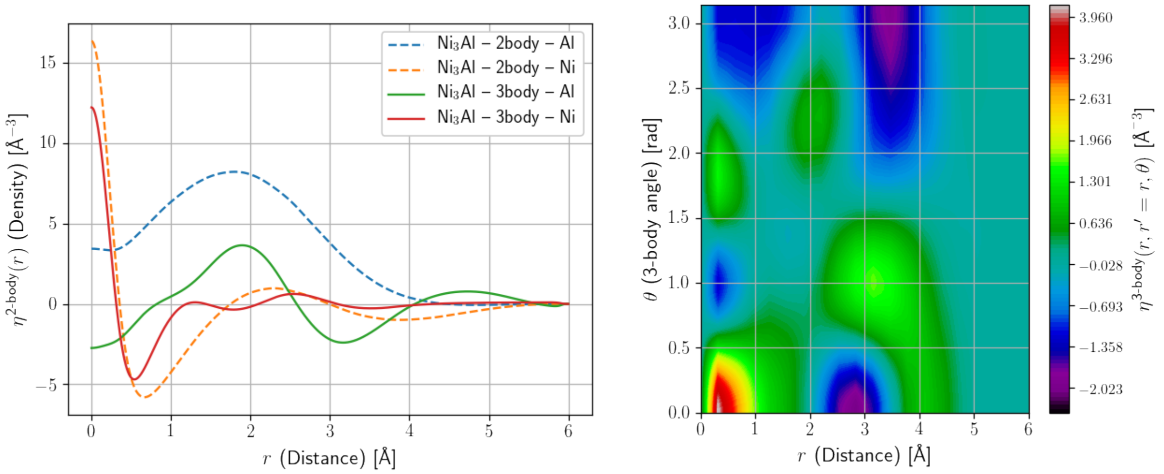
(b)

Figure 7.7: Regression of $\eta_{GS}(\mathbf{r})$ (dashed lines) for graphite using : a) 2-body and b) 2&3-body approximations. The contour and solid lines represent these approximations in both a) and b) illustrating the cross-section within the x-z-plane for $y = 0$. Red dots represent the C-atoms within 0.1 Å of the plane. The unit cell specifications are: a = 2.456 Å, b = 2.456 Å, c = 6.7 Å, $\alpha = 90°$, $\beta = 90°$ and $\gamma = 120°$



(a) (b)

Figure 7.8: Regressing $\eta_{GS}(\boldsymbol{r})$ (dashed lines) of Ni₃Al using: a) 2-body and b) 2&3-body approximations. Both contour and solid lines represent the approximation in both a) and b) illustrating the cross-section within the *x*-*y*-plane at $z \approx 1/4\,\mathrm{a}$, where a=3.553 Å. The atoms are not shown since they are separated by more than 0.1 Å from the plane, but can be imagined in the corners (Al) and along the edges and the centre (Ni).



(a) (b)

Figure 7.9: Functions obtained for the regression of $\eta_{GS}(\mathbf{r})$ for the L1$_2$ (γ') phase Ni$_3$Al: a) comparing with (solid lines) and without (dashed lines) 3-body approximations for both elements Al and Ni and b) the 3-body function for $r = r' \in [0, 6]$ Å and $\theta \in [0, \pi]$.

7.2 Empirical Potentials for Al-Ni alloys

As mentioned earlier in section 3.3.1, glue type potentials, particularly EAM potentials, incorporate many-body features referred to as the embedding density. The embedding density was originally motivated by $\eta_{GS}(\boldsymbol{r})$, [48, 49] but recently is interpreted as a further layer adding flexibility without requiring a physical foundation. We will show here that EAM potentials can indeed be developed systematically to accurately reproduce energies and forces obtained with DFT, based on our previously developed method of regressing $\eta_{GS}(\boldsymbol{r})$. To this end we study the Al-Ni alloy system. We will also discuss the weaknesses of the developed potentials in terms of 'emergent' properties, such as the stiffness constant C_{11}, and show a way for overcoming these and develop potentials which fully describe binary systems and beyond.

7.2.1 EAM Energy and Force Functions

As introduced in section 3.3, potentials of the EAM type determine crystal energies per atom based on pair and embedding energy functions (eq. 3.17). Similar to the previous section we chose to develop both energy function types as cosine series to allow for high flexibility. We find the embedding energy functions for atom n as

$$E_{\text{emb},\alpha_n}(\overline{\rho}_n) = \psi(\tilde{x}_\rho) \sum_{k=0}^{M_{\text{emb},\alpha_n}-1} w_{\text{emb},\alpha_n,k} \cos\left(\frac{\pi}{\rho_c} k \cdot \overline{\rho}_n\right) \tag{7.7}$$

where ψ is the tapering function $\psi(\tilde{x}) = \frac{\tilde{x}^4}{1+\tilde{x}^4}$ with $\tilde{x}_\rho = \frac{\overline{\rho}_n}{f}$, ρ_c some normalizing constant, $w_{\text{emb},\alpha_n,k}$ is the embedding weight for element α_n for the given k, $M_{\text{emb},\alpha_n} - 1$ is the number of terms in the series and f is a factor for controlling how sharp the tapering is.

Due to the limitation of the *setfl* file format enforcing $\overline{\rho} \geq 0$ for the embedding energy functions we modify $\eta^{\text{2-body}}(r)$ which were obtained from regressing $\eta_{GS}(\boldsymbol{r})$ in order to produce $\rho(r)$. We do this by shifting the 2-body functions for all elements by the smallest found function value and tapering towards r_{cut} as $\rho^{(k)}(r \geq r_{\text{cut}}) = 0$.

The pair energy functions are defined here for element pairs $(\alpha_n, \alpha_{n'})$ with atom-atom distances $r_{n,n'}$ as

$$E_{\text{pair},\alpha_n,\alpha_{n'}}(r_{n,n'}) = \psi(\tilde{x}_{r,n,n'}) \sum_{k=0}^{M_{\text{pair},\alpha_n,\alpha_{n'}}-1} w_{\text{pair},\alpha_n,\alpha_{n'},k} \cos\left(\frac{\pi}{r_{\text{cut}}} k \cdot r_{n,n'}\right), \tag{7.8}$$

where $\tilde{x}_r = \frac{r - r_{\mathrm{cut}}}{f}$, r_{cut} is a normalizing constant for r ($r = \|\vec{r}\|_2$), $w_{\mathrm{pair},\alpha_n,\alpha_{n'},k}$ is the pair weight for element pair (α_n, α_l) for the given k and $M_{\mathrm{pair},\alpha_n,\alpha_l} - 1$ is the number of terms in the series.

The choice of cosines not only leads to smooth functions but also they show friendly behaviour for differentiation, which is of benefit for the calculation of forces. The forces on the target atom n' for structure i is defined using the structure's potential energy E_i as

$$\mathbf{F}_{n'} \equiv -\boldsymbol{\nabla}_{\mathbf{r}_{n'}} \cdot E_i. \tag{7.9}$$

Because the energy contribution for each atom, n', depends on its *neighbourhood*, $\mathrm{neigh}(n')$, the same is true for the forces. To simplify the notation we introduce the *group* of atoms $\mathrm{group}(n') \equiv \mathrm{neigh}(n') \cup n'$, which in contrast to $\mathrm{neigh}(n')$ contains the current target atom n'. For both embedding and pair energy functions we can split the derivation into two types of sources each contributing to $\mathbf{F}_{n',\mathrm{emb}}$ and $\mathbf{F}_{n',\mathrm{pair}}$ respectively by considering as source either the *current* atom $n' = n$ or the corresponding neighbourhood $n \in \mathrm{neigh}(n')$, $n \neq n'$. These two groups of sources are abbreviated as $n' = n$ and $n' \neq n$. This split is conceptually sensible because we have to consider changes in the energy of atom n' itself (the case of $n' = n$) but also changes in the energy of other atoms which are considered neighbours of n' and contribute to that atom's energy (the case of $n' \neq n$). In the following we will use the shorthand $\nabla_{n'}$ for $\boldsymbol{\nabla}_{\mathbf{r}_{n'}}$.

Starting with the embedding energy contribution to the force by inserting eq. 7.7 into eq. 7.9 we find

$$\begin{aligned}
\mathbf{F}_{\mathrm{emb},n'} &= -\nabla_{n'} \cdot \sum_{n \in \mathrm{group}(n')} E_{\mathrm{emb},\alpha_n}\left(\rho_n, \mathbf{w}_{\mathrm{emb},\alpha_n}\right) \\
&= -\sum_{n \in \mathrm{group}(n')} \nabla_{n'} \cdot \psi\left(\tilde{x}_{\rho,n}\right) \sum_{k=0}^{M_{\alpha_{n'}}-1} w_{\mathrm{emb},\alpha_n,k} \cos\left(\frac{\pi}{\rho_c} k \cdot \rho_n\right).
\end{aligned}$$

In the case of $n' = n$, adding another index to the above definition of $\tilde{x}_\rho$ to indicate the current atom as the source under consideration, we find: $\nabla_{n'} \psi\left(\tilde{x}_{\rho,n'}\right) = 4\tilde{x}^3_{\rho,n'} / \left(1 + \tilde{x}^4_{\rho,n'}\right)^2$ $\sum_{n \in \mathrm{neigh}(n')} \left(\frac{d}{dr}\rho_{\alpha_n}\left(r_{n'n}\right)\right) \frac{\mathbf{r}_{n'} - \mathbf{r}_n}{r_{n'n}}$ applying the chain rule[1], where $\rho_{\alpha_n}\left(r_{n'n}\right)$ the embedding

---

[1] $n' = n$: $\nabla_{n'} \psi\left(\tilde{x}_{\rho,n'}\right) = \psi'\left(\tilde{x}_{\rho,n'}\right) \nabla_{n'} \tilde{x}_{\rho,n'}$ with leads with $\psi \equiv \frac{x^4}{1+x^4}$ to $\frac{d}{dx}\psi = \psi' = \frac{4x^3}{\left(1+x^4\right)^2}$ and $\nabla_{n'} \tilde{x}_{\rho,n'} = \sum_{n \in \mathrm{neigh}(n')} \left(\frac{d}{dr}\rho_{\alpha_n}\left(r_{n'n}\right)\right) \frac{\mathbf{r}_{n'} - \mathbf{r}_n}{r_{n'n}} \frac{1}{f}$.

density function for element α_n. Similarly[2] we find for $n' \neq n$ with $n \in \text{neigh}\,(n')$: $\nabla_{n'} \psi\left(\tilde{x}_{\rho,n}\right) = 4\tilde{x}^3_{\rho,n} / \left(1+\tilde{x}^4_{\rho,n}\right)^2 \left(\frac{d}{dr}\rho_{\alpha_{n'}}\left(r_{nn'}\right)\right) \frac{\boldsymbol{r}_n - \boldsymbol{r}_{n'}}{r_{n,n'}} \frac{1}{f}$. Thus we obtain the embedding energy contribution to the force for $n' = n$ as

$$\begin{aligned}\mathbf{F}_{\text{emb},n',n'=n} \quad = \quad & -\textstyle\sum_{k=0}^{M_{\alpha_{n'}}} w_{\text{emb},\alpha_{n'},k} \left(\psi'\left(\tilde{x}_{\rho,n'}\right) \cos\left(\frac{\pi}{\rho_c} k \cdot \rho_{n'}\right)\right. \\ & \left. -\psi\left(\tilde{x}_{\rho,n'}\right) \frac{\pi}{\rho_c} k \sin\left(\frac{\pi}{\rho_c} k \cdot \rho_{n'}\right)\right) \cdot \\ & \left(\textstyle\sum_{n \in \text{neigh}(n')} \left(\frac{d}{dr}\rho_{\alpha_n}\left(r_{n'n}\right)\right) \frac{\mathbf{r}_n - \mathbf{r}_l}{r_{nl}}\right)\end{aligned} \tag{7.10}$$

with $\psi'(x) \equiv \frac{d}{dx}\psi(x)$ and for $n \neq n'$ as

$$\begin{aligned}\mathbf{F}_{\text{emb},n',n' \neq n} \quad = \quad & -\textstyle\sum_{n \in \text{neigh}(n')} \left[\left(\frac{d}{dr}\rho_{\alpha_{n'}}\left(r_{n'n}\right)\right) \left(-\frac{\mathbf{r}_n - \mathbf{r}_{n'}}{r_{nn'}}\right) \cdot \right. \\ & \textstyle\sum_{k=0}^{M_{\alpha_n}} w_{\text{emb},\alpha_n,k} \left(\psi'\left(\tilde{x}_{\rho,n}\right) \cos\left(\frac{\pi}{\rho_c} k \cdot \rho_n\right)\right. \\ & \left.\left. -\psi\left(\tilde{x}_{\rho,n}\right) \frac{\pi}{\rho_c} k \sin\left(\frac{\pi}{\rho_c} k \cdot \rho_n\right)\right)\right].\end{aligned} \tag{7.11}$$

Combining both contributions to $F_{\text{emb},n'}$ from eq. 7.10 and 7.11 we find the embedding force as

$$\mathbf{F}_{\text{emb},n'} = \mathbf{F}_{\text{emb},n',n'=n} + \mathbf{F}_{\text{emb},n',n' \neq n}. \tag{7.12}$$

Inserting the definition for E_{pair} from eq. 7.8 into eq. 7.9 we find

$$\begin{aligned}\mathbf{F}_{\text{pair},n'} &= -\nabla_{n'} \cdot \frac{1}{2} \sum_{n \in \text{group}(n')} E_{\text{pair}}\left(\mathbf{R}_n, \mathbf{w}_{\text{pair}}\right) \\ &= -\nabla_{n'} \cdot \frac{1}{2} \sum_{n \in \text{group}(n')} \sum_{l \in \text{group}(n)} \psi\left(\tilde{x}_{r,n',n}\right) \cdot \\ & \sum_{k=0}^{M_{\text{pair}\alpha_n,\alpha_l}} w_{\text{pair},\alpha_{n'},\alpha_n,k} \cos\left(\frac{\pi}{r_{\text{cut}}} k \cdot r_{n'n}\right).\end{aligned}$$

---

[2] $n' \neq n$: $\nabla_{n'}\psi\left(\tilde{x}_{\rho,n}\right) = \psi'\left(\tilde{x}_{\rho,n}\right) \nabla_{n'}\tilde{x}_{\rho,n}$ with $\nabla_{n'}\tilde{x}_{\rho,n} = -\left(\frac{d}{dr}\rho_{\alpha_{n'}}\left(r_{nn'}\right)\right) \frac{\boldsymbol{r}_{n'} - \boldsymbol{r}_n}{r_{nn'}} = \left(\frac{d}{dr}\rho_{\alpha_{n'}}\left(r_{nn'}\right)\right) \frac{\boldsymbol{r}_n - \boldsymbol{r}_{n'}}{r_{n,n'}} \frac{1}{f}$.

Splitting this into our two sources of contributions we find with the chain rule[3] for $n' = n$

$$\begin{aligned} \mathbf{F}_{\text{pair},n',n'=n} &= -\tfrac{1}{2}\cdot\textstyle\sum_{n\in\text{neigh}(n')}\left\{\frac{\mathbf{r}_{n'}-\mathbf{r}_n}{r_{n'n}}\textstyle\sum_{k=0}^{M_{\text{pair}\alpha_{n'},\alpha_n}}\left[w_{\text{pair},\alpha_{n'},\alpha_n}\right.\right.\\ &\quad\cdot\left(\psi'\left(\tilde{x}_{r,n',n}\right)\cos\left(\frac{\pi}{r_\text{cut}}k\cdot r_{n'n}\right)\right.\\ &\quad\left.\left.\left.-\psi\left(\tilde{x}_{r,n',n}\right)\frac{\pi}{r_c}k\sin\left(\frac{\pi}{r_\text{cut}}k\cdot r_{n'n}\right)\right)\right]\right\} \end{aligned} \tag{7.13}$$

and for $n' \neq n$

$$\begin{aligned} \mathbf{F}_{\text{pair},n',n'\neq n} &= -\tfrac{1}{2}\cdot\textstyle\sum_{n\in\text{neigh}(n')}\left\{\left(-\frac{\mathbf{r}_n-\mathbf{r}_{n'}}{r_{n'n}}\right)\left[\textstyle\sum_{k=0}^{M_{\text{pair}\alpha_{n'},\alpha_n}}w_{\text{pair},\alpha_n,\alpha_{n'}}\right.\right.\\ &\quad\cdot\left(\psi'\left(\tilde{x}_{r,n,n'}\right)\cos\left(\frac{\pi}{r_\text{cut}}k\cdot r_{nn'}\right)\right. \qquad\cdot\\ &\quad\left.\left.\left.-\psi\left(\tilde{x}_{r,n,n'}\right)\frac{\pi}{r_\text{cut}}k\sin\left(\frac{\pi}{r_\text{cut}}k\cdot r_{nn'}\right)\right)\right]\right\} \end{aligned} \tag{7.14}$$

Fortunately, because the pair energy contribution is symmetric with respect to the element pair ($\boldsymbol{w}_{\text{pair},\alpha_{n'},\alpha_n} = \boldsymbol{w}_{\text{pair},\alpha_n,\alpha_{n'}}$), $r_{n'n} = r_{nn'}$ and $-\boldsymbol{r}_n - \boldsymbol{r}_{n'} = \boldsymbol{r}_{n'} - \boldsymbol{r}_n$ and thus $\mathbf{F}_{\text{pair},n',n'=n} = \mathbf{F}_{\text{emb},n',n'\neq n}$, we can write $\mathbf{F}_{\text{pair},n'}$simplified as

$$\mathbf{F}_{\text{pair},n'} = \mathbf{F}_{\text{pair},n',n'=n} = \mathbf{F}_{\text{pair},n',n'\neq n}. \tag{7.15}$$

Having computed the forces due to the pair and embedding energy contributions we finally can calculate the force on atom n' as $\mathbf{F}_{n'} = \mathbf{F}_{\text{pair},n'} + \mathbf{F}_{\text{emb},n'}$ making use of eq. 7.12 and 7.15.

7.2.2 Results for the Regression of an Al-Ni alloy database

To develop EAM potentials for the Al-Ni alloy system using the method developed to approximate $\eta_{GS}(\mathbf{r})$ and modify the obtained $\eta^{\text{2-body}}(r)$ as described to yield $\rho(r)$, we regress energies and forces in a three step procedure:

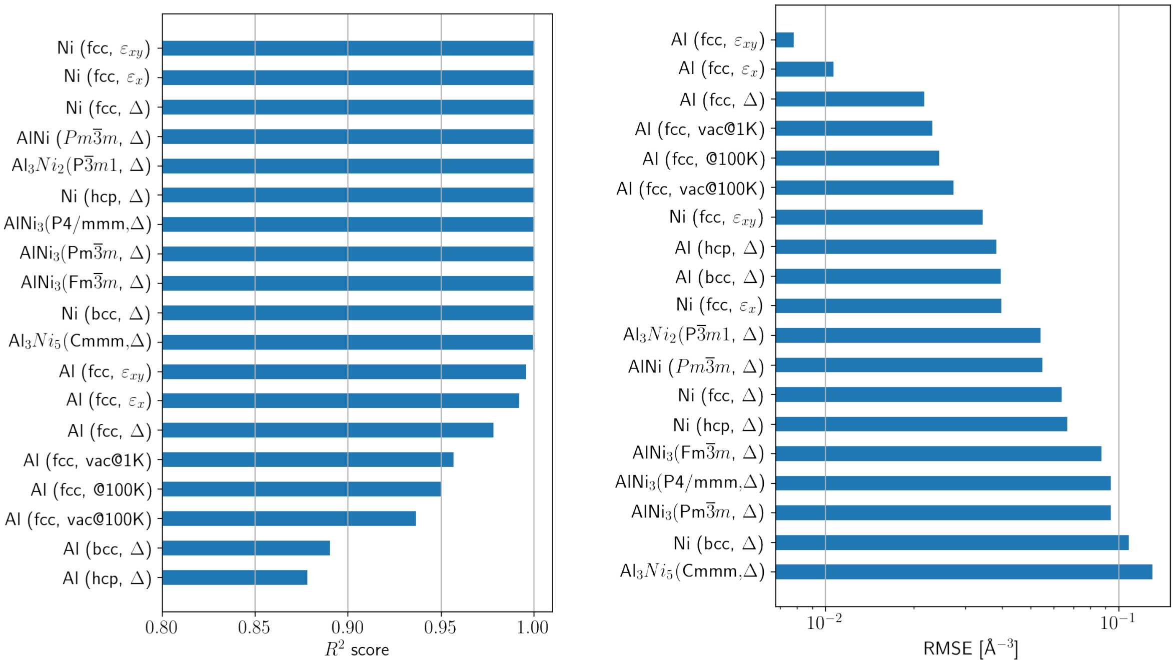
1. regress $\eta_{GS}(\boldsymbol{r})$ within 0.5 Å of all atom cores for all crystals in the alloy database,
2. optimize per atom energies and forces by adjusting the parameters of $E_{\text{emb},\alpha_n}(\overline{\rho}_n)$ and $E_{\text{pair},\alpha_n,\alpha_{n'}}(r_{nn'})$ where $\alpha_n = \alpha_{n'}$ for single element structures, and finally
3. optimize per atom energies and forces for single and multiple element structures adjusting $E_{\text{pair},\alpha_n,\alpha_{n'}}(r_{nn'})$ for $\alpha_n \neq \alpha_{n'}$ as well as $E_{\text{emb},\alpha_n}(\overline{\rho}_n)$.

---

[3] $n' = n$: $\nabla_{n'}\psi\left(\tilde{x}_{r,n',n}\right) = \psi'\left(\tilde{x}_{r,n',n}\right)\frac{\mathbf{r}_{n'}-\mathbf{r}_n}{r_{n'n}}$ and for $n' \neq n$: $\nabla_{n'}\psi\left(\tilde{x}_{r,n,n'}\right) = -\psi'\left(\tilde{x}_{r,n,n'}\right)\frac{\mathbf{r}_{n'}-\mathbf{r}_n}{r_{n',n}} = \psi'\left(\tilde{x}_{r,n,n'}\right)\frac{\mathbf{r}_n-\mathbf{r}_{n'}}{r_{n',n}}$

The database for the alloy consists of Al and Ni phases, fcc Al configurations (2×2×2 supercells) sampled from *ab initio* molecular dynamics at 1 K and 100K with and without vacancy defects, and binary phases with B2 and $L1_2$ structures. All high symmetry configurations were strained equiaxially between ±1 %. Furthermore, normal strain along the x-cell vector and shear strains in the x-y-plane were applied to fcc Al as well as fcc Ni.

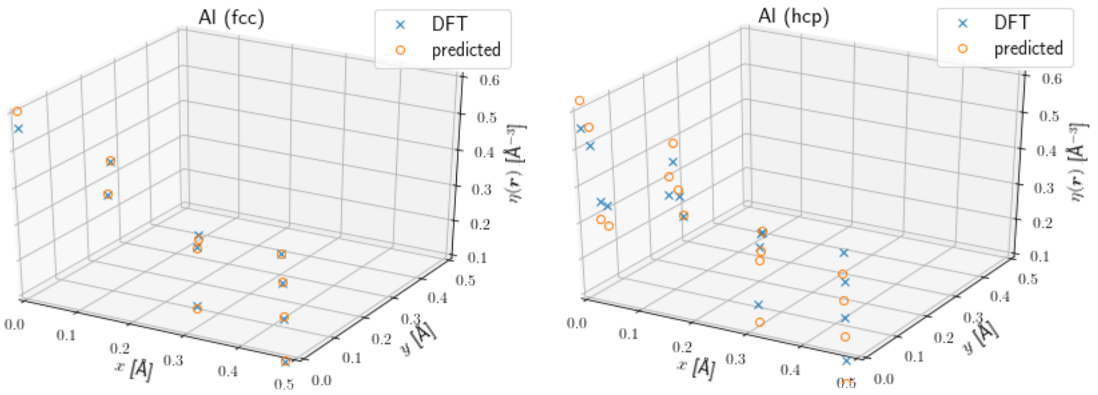
The prediction errors of $\eta_{GS}(\mathbf{r})$ within 0.5 Å of the atom cores using $\eta^{\text{2-body}}(r)$ obtained for Al and Ni over the entire Al-Ni alloy system are shown in fig. 7.10 in form of a) R^2 scores and b) RMSEs. In this figure atom configurations are grouped by phase and deformation and sorted by the respective error measure. It can be seen that the approximations have R^2 scores of at least 0.87 for all structures, with hcp Al having the lowest R^2 score. The RMSE re-arranges these groups showing roughly that Al crystals have the smallest RMSEs, and Ni-containing structures have the largest RMSEs. A similar argument applies as before. Because Ni has a larger atomic number than Al and therefore more valence electron, the sample variance, $\mathrm{var}(\eta_{\mathrm{GS}})$, of Ni-containing crystals is larger than for pure Al crystals, leading to larger R^2 scores for Ni-containing structures for a similar RMSE.



(a) (b)

Figure 7.10: All phases, deformations and configurations included in the Al-Ni alloy database are shown. The R^2 scores in a) and RMSEs in b) correspond to the 2-body density approximation of $\eta_{GS}(\mathbf{r})$ within 0.5 Å of all atom cores. The volumetric strain, the normal strain along the x-cell vector and the shear strain parallel to the x-z-plane are indicated as Δ, ε_x and ε_{xy} respectively. The groups are sorted from ‘best’ at the top to ‘worst’ at the bottom for both R^2 score and RMSE.

The predicted and DFT electron densities for Al crystals with the worst (fcc) and best (hcp) approximations are plotted on the x-y-plane ((001) plane normal) at $z = 0$ for 0 % strain in fig. 7.11 a) and b) respectively. We find that fcc Al is closely approximated everywhere except for a slight overestimation of $\eta_{GS}(\boldsymbol{r})$ on top of the atom. The predictions for hcp Al, on the other hand, over- or underestimate $\eta_{GS}(\boldsymbol{r})$ in more areas. However, even for hcp Al the characteristics of $\eta_{GS}(\boldsymbol{r})$ are still visible in the predictions. Proceeding to the fitting of the E_{emb} and E_{pair} functions, the embedding density functions are obtained, as described above, by shifting and tapering leading to the functions for Al and Ni as shown in fig. 7.12.



(a)

(b)

Figure 7.11: Representation of $\eta_{GS}(\mathbf{r})$ (hatches, 'DFT') and $\eta_{\text{pred}}(\mathbf{r})$ (circles, 'predicted') for $z = 0$ Å. $\eta_{\text{pred}}(\mathbf{r})$ was obtained regressing $\eta_{GS}(\mathbf{r})$ within 0.5 Å of all atom cores for the Al-Ni alloy database. Parts a) and b) show fcc and hcp Al respectively.

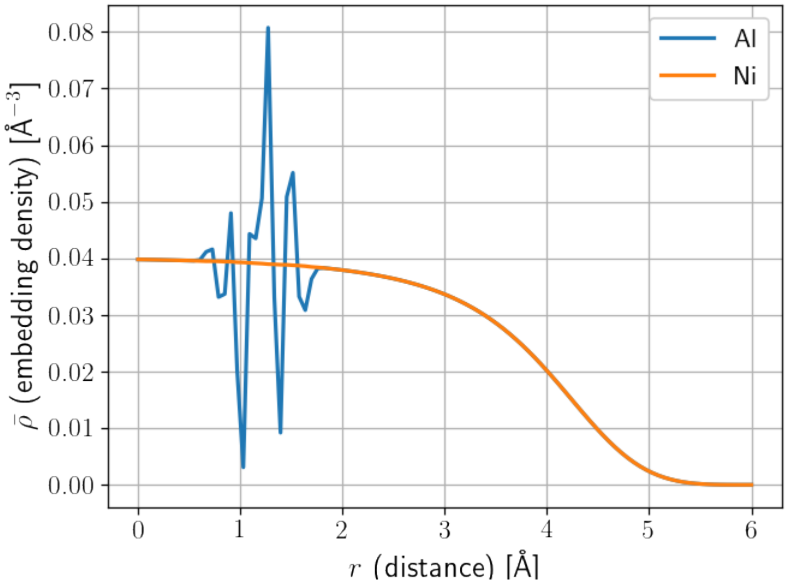


Figure 7.12: Embedding density functions for Al and Ni after applying a transformation to preserve $\bar{\rho}(r = r_{\text{cut}}) = 0$ and $\bar{\rho}(r < r_{\text{cut}}) \geq 0$.

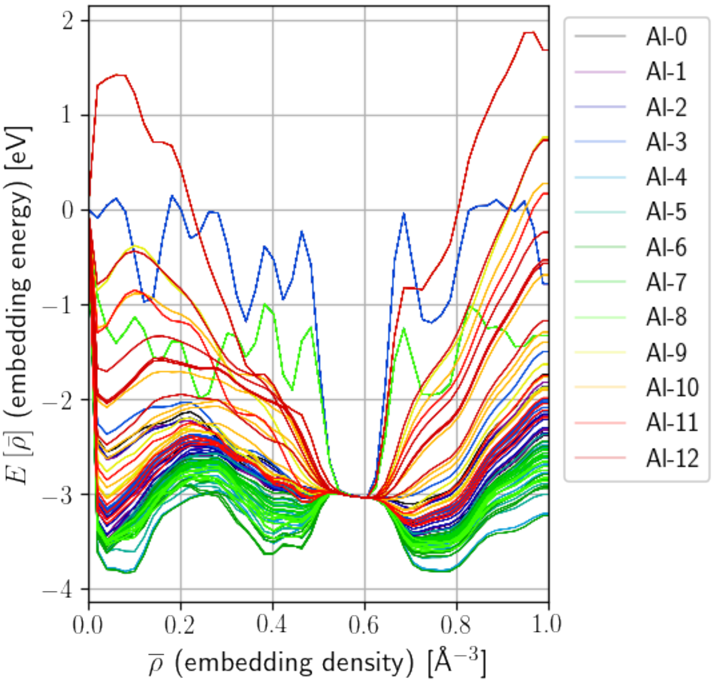
The embedding and pair energy functions are set up using the following number of weights for each: $M_{\alpha_n} = 30$ and $M_{\alpha_n,\alpha_{n'}} = 40$ leading to 180 model parameters $\mathbf{w}$ to be optimized. The regression was carried out fitting to energies and forces with a Bayesian loss function $\mathscr{L}$ to make the approach general enough to allow for the optimization of the hyperparameters, which is preferred to grid searches since even random choices are better than grid searches [23]. $\mathscr{L}$ is given as

$$\mathscr{L} = \prod_{i=1}^{N} p\left(E_i, \mathbf{F}_i | \mathbf{w}, \beta_E, \beta_F\right) p\left(\mathbf{w} | \boldsymbol{\alpha}\right) \tag{7.16}$$

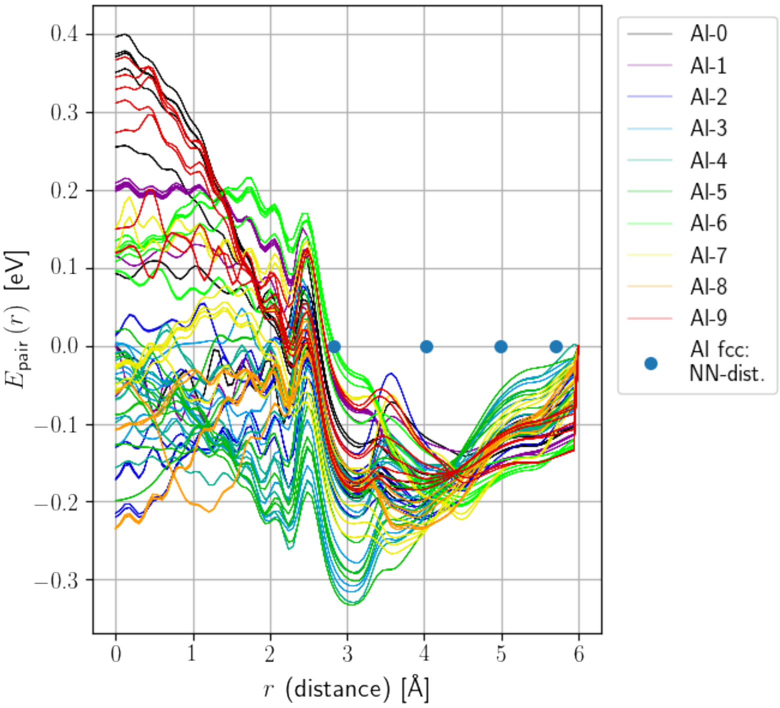
where N is the number of structures, E_i is the energy for a given structure, $\mathbf{F}_i$ are the forces for that structure, and $\boldsymbol{\alpha}$, β_E and β_F are the hyperparameters. The noise levels for energy deviations, ΔE, and magnitudes of force deviations, $||\Delta\mathbf{F}||_2$, are assumed to be normally distributed as $\Delta E \sim \mathscr{N}\left(0, \beta_E\right)$ and $||\Delta\mathbf{F}||_2 \sim \mathscr{N}\left(0, \beta_F\right)$. Also the model weights are assumed to be drawn from individual normal distributions $w_i \sim \mathscr{N}\left(0, \alpha_i\right)$. This approach allows for the constraint of individual weights by optimizing $\boldsymbol{\alpha}$ and the automated scaling of the contribution of observed energies and forces.

The method then proceeds by maximizing $\log\mathscr{L}$, by alternately optimizing $\mathbf{w}$ and the hyperparameters. Testing various optimization methods for $\mathbf{w}$ it was found that the covariance matrix adaption evolution strategy (CMA-ES) [95] as implemented in *deap* [68] is best suited to maximize $\log\mathscr{L}$. Other optimizers tested such as BFGS, simulated annealing and Nelder-Mead among others also found maxima but stalled more easily whereas CMA-ES consistently found well defined maxima. A general advantage of evolution strategies is that they are embarrassingly parallel, with *deap*'s implementation of CMA-ES fortunately already designed for parallel usage with a distributed work manager *scoop* [103]. The optimizations of $\mathbf{w}$ were carried out with a maximum of 1000 iterations for $20 \cdot \lambda$ potentials or 'individuals', where λ is the number of model parameters. Convergence was achieved if the standard deviation of the fitness among the potentials fell below 10^{-5}. After $\mathbf{w}$ was optimized the hyperparameters were adjusted using BFGS as implemented in *scipy* [116]. This loop was repeated 7 times for each optimization job. In order to use different initial seeds multiple optimization jobs were carried out. This was done for step 2) and once sufficient single element potentials were identified then also for step 3, fixing $E_{\mathrm{pair},\alpha_n,\alpha_n}$ as found in step 2 but optimizing the remaining energy functions freely over the entire alloy database.

Optimizing the energy functions we observed a 'pinching' effect where most of the energy functions obtained during the optimization process take on similar values as shown in fig. 7.13 for a) $E_{\mathrm{emb,Al}}$ and b) $E_{\mathrm{pair,Al,Al}}$. The pinching can be observed for $E_{\mathrm{emb,Al}}$ at $\overline{\rho}$ around 0.6 and for $E_{\mathrm{pair,Al,Al}}$ at $r \gtrsim 2.7$ Å. The probable reason for this effect is indicated in 7.13 b), which illustrates the nearest neighbour distances for Al's equilibrium fcc structure. We interpret this effect by viewing the regression process as the sampling of energy functions from a function posterior distribution, i.e. $p\left(\text{functions}|\text{database}\right)$, with the database containing pair distances, embedding densities, energies and forces. Thus in regions for $r_{n'n}$, visible in 7.13 b), where possible pair energy contributions for given distances are constrained by the structures in the training set, the resulting functions sampled during the optimization process have highly similar values but take on pretty much any value with uniform probability outside these regions.



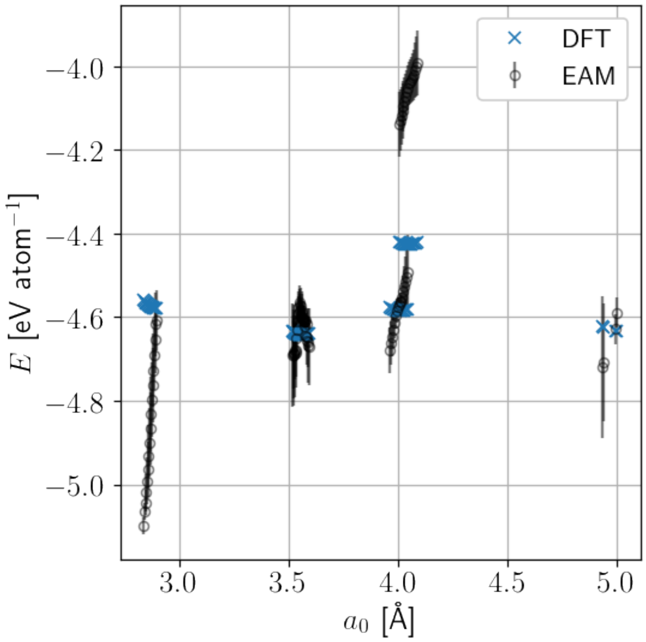
(a)



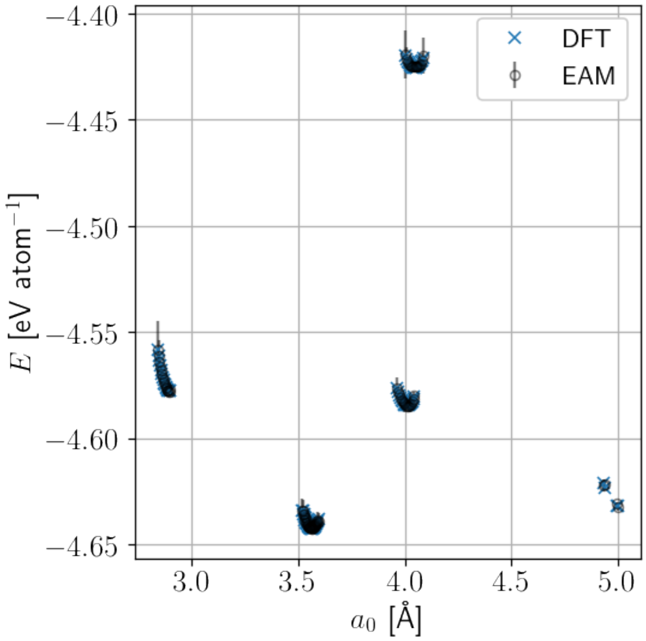
(b)

Figure 7.13: (a) Embedding energy $E_{\text{emb,Al}}\,(\overline{\rho})$ and (b) pair energy $E_{\text{pair,Al,Al}}\,(r)$ functions for 10 repeated optimization jobs updating model- and hyperparameters displaying the 70 best potentials each. $E_{\text{emb,Al}}\,(\overline{\rho})$ was obtained fitting the entire Al-Ni database. The blue dots in b) indicate nearest neighbor distances in equilibrium fcc Al.

Commonly the energy functions are obtained by fixing $E_{\mathrm{pair},\alpha_n,\alpha_n}$ and $E_{\mathrm{emb},\alpha_n}$ and fitting to single element structures whereas $E_{\mathrm{pair},\alpha_n,\alpha_{n'}}$ is optimized for structures with more than one element [49, 152]. Trying this, however, we found that while single element energies were approximated well the two element energies were significantly off as shown in fig. 7.14 a) for the present database of the Al-Ni alloy system. Note that the black circles represent the mean energy predicted for all potentials collected during the optimization process with the black lines representing one standard deviation. Fixing both single element E_{pair} and E_{emb} can be made to work if further mixing parameters are introduced for adjustment as in [152]. Since in this work no mixing parameters are assumed we found it necessary (although computationally more expensive) to include the E_{emb} functions into the fitting of the structures with more than one element. This leads to close energy force fits of all structures included in the database, see 7.14 b) for the binary Al-Ni structures. We deem it sensible with the way the EAM potential is formulated to fit E_{emb} to all structures since the E_{emb} depends on $\overline{\rho}$ which can depend on more than one element. Using a basis set for the E_{emb} function allows for great flexibility but also leads to essentially random values outside of the pinching areas, which means for potentials where E_{emb} are only fitted to single element systems that the contribution for two or more element systems may become partially random. This randomness then needs to be buffered by the remaining free $E_{\mathrm{pair},\alpha_n,\alpha_{n'}}$ leading to limitations in the predictive power of the potential.



(a)



(b)

Figure 7.14: Comparison of predicted energies for binary Al-Ni structures versus the x-component of the x-cell vector, here denoted as a_0. Part a) shows the result of fixing all $E_{\mathrm{emb},\alpha}(\overline{\rho})$ on single element structures and b) adjusting all $E_{\mathrm{emb},\alpha}(\overline{\rho})$ against all structures in the alloy database for the Al-Ni alloy system. Blue hatches represent DFT reference energies whereas black dots with black lines indicate the mean and standard deviation for 910 generated potentials.

To complement fig. 7.14 b), table 7.2 shows the root mean squared errors (RMSEs) of the energy deviation for a) Al, b) Al-Ni and c) Ni structures. In this table the distribution of RMSEs is captured for potentials obtained with the first 4 optimization runs for all structures in the respective element group by the minimum RMSE, maximum RMSE and mean RMSE. It can be seen that the systems are approximated in the order of increasing mean RMSE: Ni, Al-Ni and Al. The smallest RMSE is 0.07 meV atom^{-1} and the largest RMSE is 12.92 meV atom^{-1}, with the majority of potentials having an accuracy of 10 meV atom^{-1} or better.

RMSE $\left[\frac{\text{meV}}{\text{atom}}\right]$	Min	Max	Mean
job 1	0.99	12.92	2.7
job 2	0.79	7.3	1.82
job 3	0.8	7.3	1.79
job 4	0.8	7.3	1.79

(a)

RMSE $\left[\frac{\text{meV}}{\text{atom}}\right]$	Min	Max	Mean
job 1	0.39	10.6	1.89
job 2	0.36	10.6	1.94
job 3	0.39	10.59	1.88
job 4	0.27	4.51	0.96

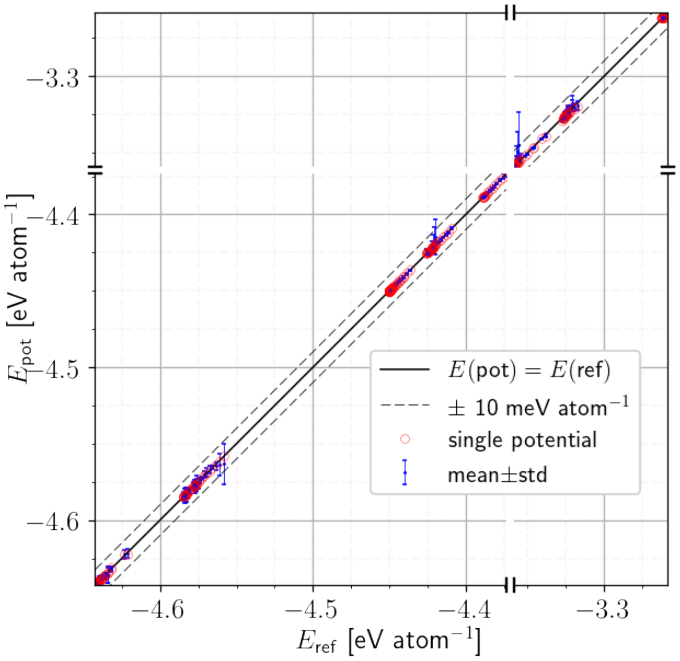
(b)

RMSE $\left[\frac{\text{meV}}{\text{atom}}\right]$	Min	Max	Mean
job 1	0.09	0.89	0.21
job 2	0.08	0.82	0.37
job 3	0.08	0.83	0.35
job 4	0.07	0.69	0.18

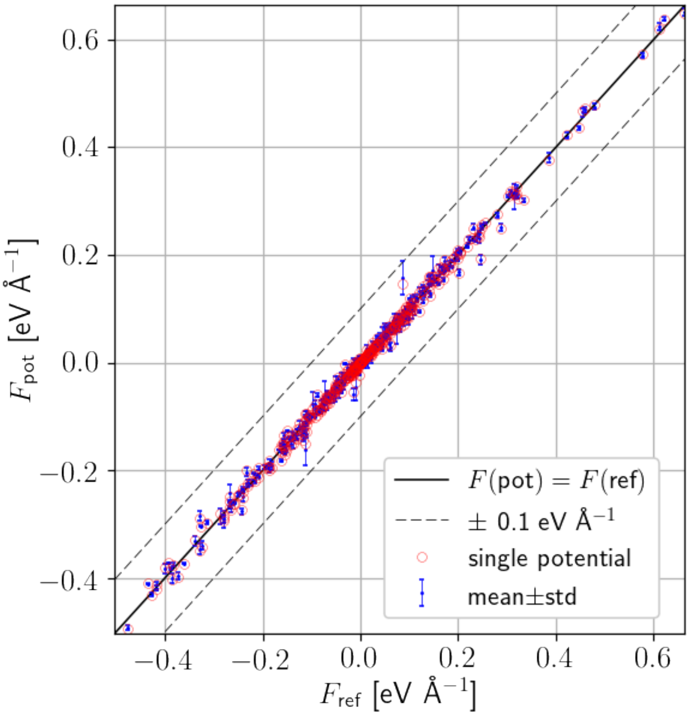
(c)

Table 7.2: a), b) and c) show the RMSEs over all Al, Al-Ni and Ni structures respectively. Computing root mean square errors (RMSEs) for individual potentials over all structures yields a RMSE distribution when performed for all potentials. ‘Min’ and ‘Max’ refer to the potentials found with the smallest and largest RMSE for the given job whereas ‘Mean’ is the mean over the RMSE for all potentials of that job.

In fig. 7.15 the predictions of the 910 generated EAM potentials for energies and forces (x-components) are shown for all structures in the alloy database relative to the reference energies and forces. The energies and forces are predicted generally well within 10 meV atom^{-1} and 0.1 eV atom^{-1} respectively. The ‘best’ potential highlighted in red has the smallest mean absolute energy error of 0.28 meV atom^{-1}, with the mean over all structures in the database.



(a)



(b)

Figure 7.15: Predictions of 910 EAM potentials versus DFT reference values for a) energies and b) x-components of forces. Red circles represent predictions for individual structures using the potential which has the smallest mean absolute energy error of 0.28 meV atom^{-1} for all potentials over all structures. The mean energy and force values over all collected potentials are shown as blue dots with error bars representing one standard deviation.

Another test for empirical potentials are emergent properties such as vacancy formation energies $E_{f,\text{vac}}$, elastic constants and other quantities not explicitly included in the training or $\mathscr{L}$. Because only energies and forces are directly fitted in this process these emergent properties are found to be intricately linked to the training set itself as is shown in figs. 7.16 to 7.19. In these figures we see histograms for each of the emergent properties showing the number of potentials corresponding to values in certain bins. In fig. 7.16 this relationship between training set and 'emergent' properties is studied for Al only. The Al structures were partitioned into the following groups: 'high symmetry' (only contains high symmetry structures), 'MD no vac' (structures from MD DFT at 1 K and 100 K without a vacancy defect), 'MD vac' (MD DFT with a vacancy defect), 'MD all' (all MD DFT structures) and 'all' (all high symmetry and MD structures). Here $\rho_{Al}(r)$, E_{pair} and E_{emb} were obtained by regressing those groups of Al structures only. Naively one would expect that 'all' clearly outperforms all other groups but this does not seem to be the case with 'high symmetry' having more potentials near reference values than other groups, except for $E_{f,\text{vac}}$, which is not surprising since it did not contain any structures with vacancy defects. Another aspect to note is that the strained structures present in the 'high symmetry' group seem to have a positive impact on the prediction of the elastic constants. Although not shown here, the inclusion of structures from MD DFT generally leads to improvements in the predicted phonon spectrum. The effect of the strained structures in the training set on the predicted elastic constants is also visible when comparing the elastic constants for fcc Al, fcc Ni and γ' Ni_3Al in fig. 7.17, 7.18 and 7.19 respectively. It can be seen that the predicted C_{ij} for γ' are relatively poor in comparison to fcc Al and fcc Ni. This is probably because no uniaxial or simple shear deformations are present in the training set for γ'. Contrary to fig. 7.16 the predictions of $E_{f,\text{vac}}$ are relatively poor for fcc Al and fcc Ni, shown in fig. 7.17 and 7.18, with the most significant difference in the fitting process being how the embedding density and the energy functions were obtained. Comparing fcc Al and fcc Ni with γ' we also find that the ground state energies and lattice constants for the elements are spot on whereas they are slightly over- and underestimated respectively for γ'. Also, we find that the vacancy formation energies for γ' are much closer to the reference values than for the elements. This is surprising since the vacancy formation energies for fcc Al and Ni are relatively poor. The antiphase boundary (APB) and intrinsic stacking fault (ISF) energies are not well predicted except in the latter case for fcc Al. This is reasonable since they represent quite significant changes in the configuration. Thus, it seems that emergent properties which are implicitly included in the training set are more likely to predicted properly.

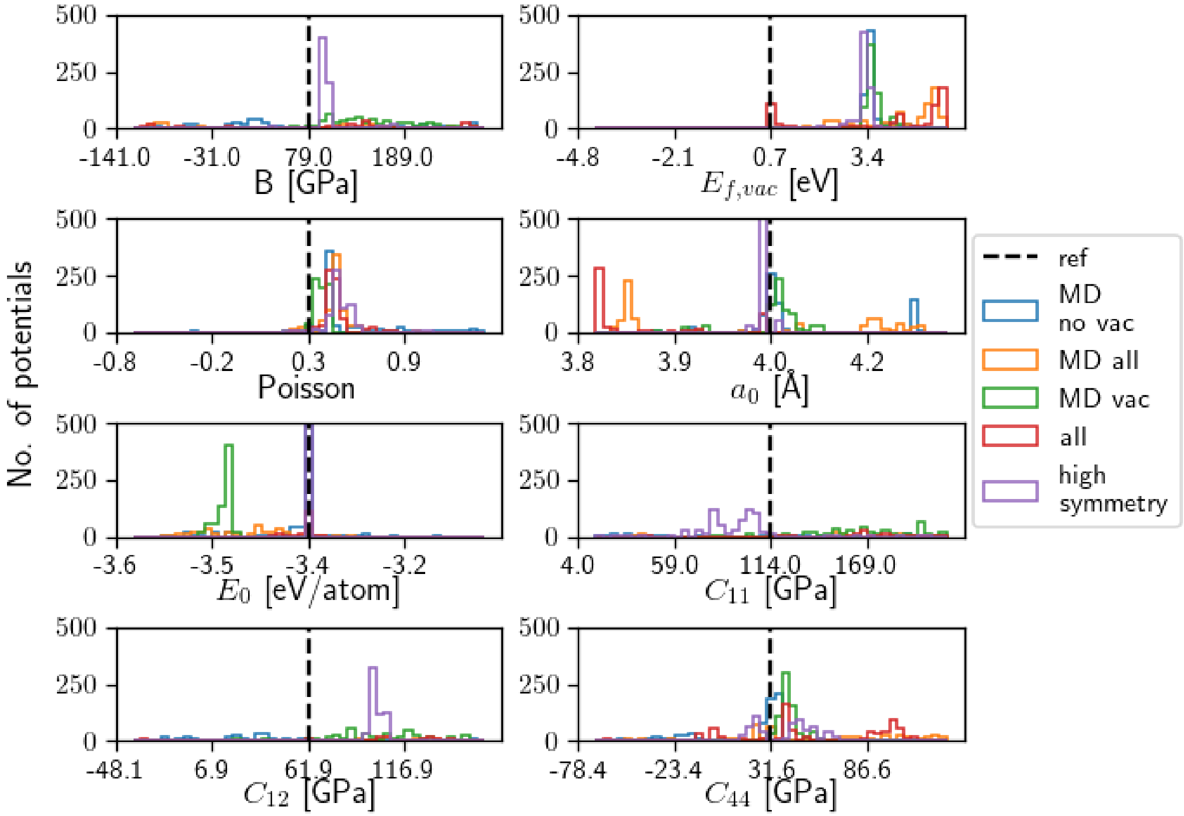


Figure 7.16: Comparison of the influence of different training sets on the EAM potential properties for fcc Al. All histograms display the number of potentials on the y-axis and the respective property on the x-axis. Shown are training sets with configurations obtained by annealing 2x2x2 fcc cells with NPT MD at 100 K (‘MD’) and ground state calculations of fcc, bcc and hcp Al at 0 K (‘high symmetry’). The high symmetry configurations contain normal and shear strained cells with in total 85 configurations whereas ‘MD no vac’ and ‘MD vac’ contain 5 configurations respectively. ‘MD all’ combines both MD training sets representing structures with one vacancy and without a vacancy. The histograms labeled ‘all’ represent training sets which contain all MD and high symmetry configurations.

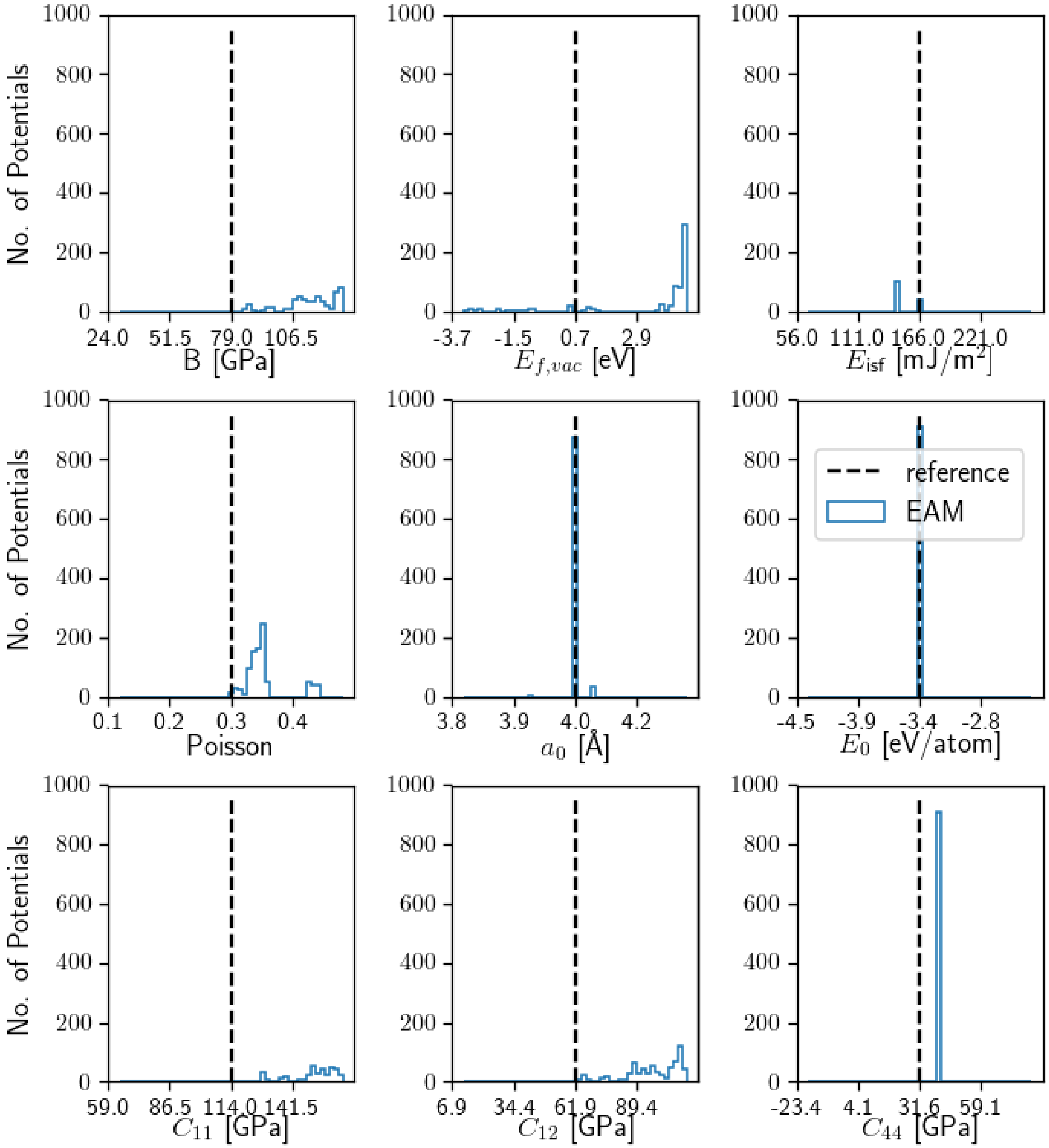


Figure 7.17: Emergent properties of fcc Al collected from 910 EAM potentials obtained after fitting the entire Al-Ni alloy database. The reference values were obtained from [152]. a_0 represents the relaxed lattice constant, $E_{f,vac}$ the vacancy formation energy and E_{isf} the intrinsic stacking fault energy.

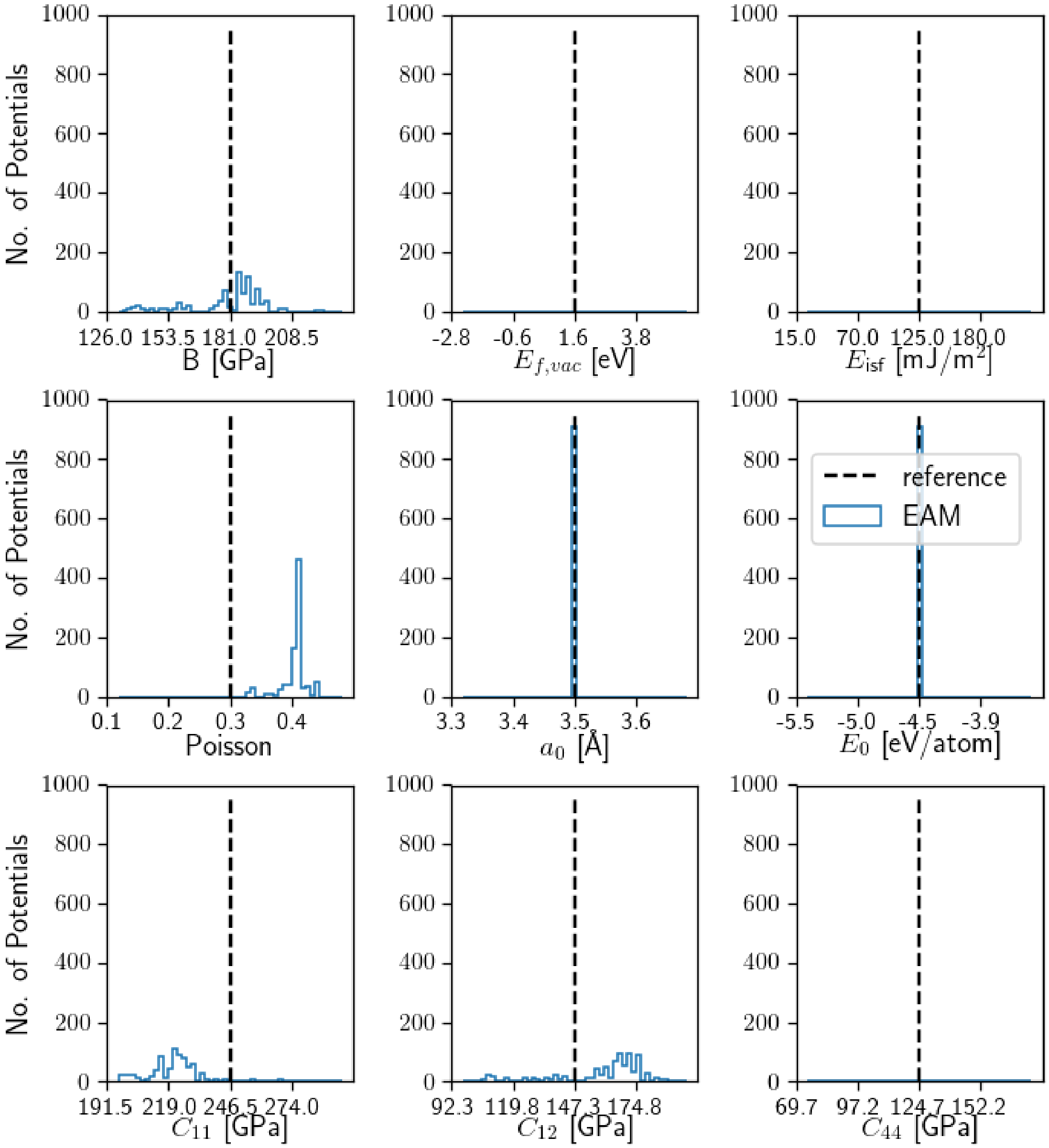


Figure 7.18: Emergent properties of fcc Ni collected from 910 EAM potentials obtained after fitting the entire Al-Ni alloy database. The reference values were obtained from [152]. a_0 represents the relaxed lattice constant, $E_{f,vac}$ the vacancy formation energy and E_{isf} the intrinsic stacking fault energy.

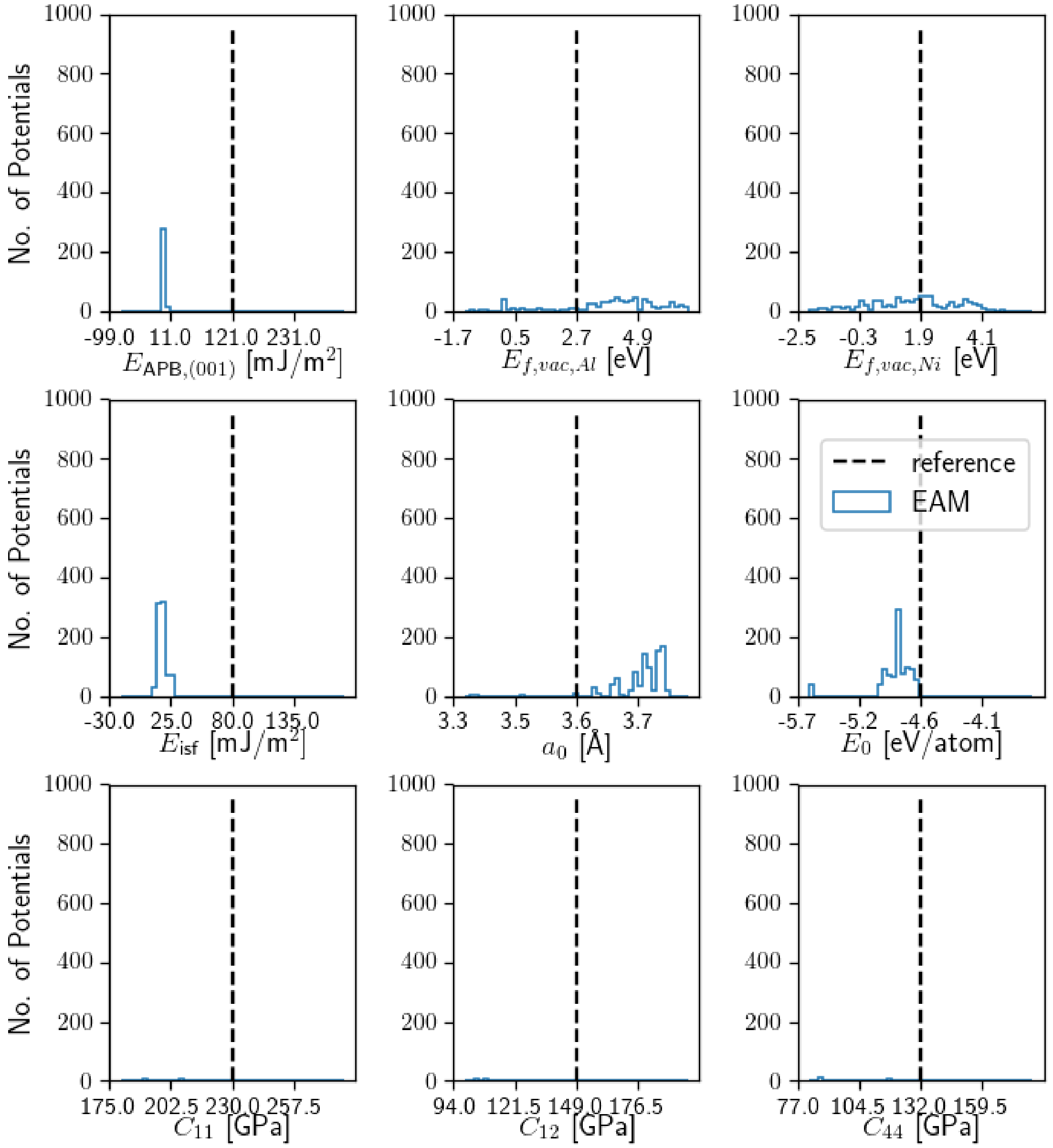


Figure 7.19: Emergent properties of γ' ($L1_2$ Ni_3Al ordered phase) collected from 910 EAM potentials obtained after fitting the entire Al-Ni alloy database. The reference values were obtained from [152]. a_0 represents the relaxed lattice constant, $E_{f,vac}$ the vacancy formation energy, E_{isf} the intrinsic stacking fault energy and $E_{\mathrm{APB},[001]}$ the [001] anti-phase boundary energy.

7.3 Discussion

Ground State Electron Densities

The linear regression method presented here was employed successfully to reproduce $\eta_{GS}(\mathbf{r})$ and its derivatives using 2- and 3-body approximations despite the relatively strong assumption of linear superposition of contributions from single atoms and pairs of atoms. Because no one optimization method is superior at solving all optimization problems, which is the essence of the 'no free lunch' theorem [264], a range of methods were tried including Bayesian Ridge, linear and ElasticNet regression among others. RVMs were implemented for this problem by the author but the other methods were used as provided by scikit-learn [168]. Among these methods RVMs were always among the top three with the lowest error, which were usually very similar. But RVMs have an additional advantage which is high sparsity leading to a runtime complexity of $\mathcal{O}(N)$ with $M \ll N$.

Complementary methods to this approach are based on non-parametric models, such as Gaussian processes, which currently are limited in the number of observations they can treat but potentially provide accurate predictive variances. An approach using kernels was suggested by Brockherde *et al.* [30] who coupled $\eta_{\text{pred}}(\mathbf{r})$ with a map for $E\left[\eta_{\text{pred}}(\mathbf{r})\right]$ to predict ground state energies with high accuracies in the range of 2.5 - 212.5 meV for H_2O and 3.7 - 125.8 meV for H_2.

This accuracy compares with the OF DFT energy errors found in the present work for high symmetry structures with the smallest and largest of the MAX values for the test found to be 0.2293 meV atom^{-1} and 50.153 meV atom^{-1} for Al. Assuming the $\Delta E\left(\mathrm{R}^2\right)$ bounds hold using KS DFT for all high symmetry structures studied, then all structures except SWCNT would have an error smaller than 10 meV atom^{-1}. If the ΔE (RMSE) bounds hold we would expect energy errors of 10 meV atom^{-1} or better for the Al system, NaCl and $BaTiO_3$, whereas the H_2-crystal, graphite, the SWCNT and the tungsten system would have errors in the range of hundreds of meV atom^{-1}. We think that taking these bounds as a guideline is reasonable because the basis set for our model is highly differentiable and therefore less likely to lead to density perturbations as used to generate the upper bounds themselves.

Improvements could be made in $\eta_{\text{pred}}(\mathbf{r})$ by sampling density points which include higher derivatives into the weighting for the sampling or directly including density derivatives into the regression. Including derivatives could significantly reduce the OF DFT energy error since they play an important role in the kinetic energy calculations. Furthermore, one could extend the basis with higher order many body terms or other functional forms, such as Chebyshev polynomials or Gaussians. However, independent of the degree of the many body information used and independent of the functional forms chosen to represent that information in feature space, the current approach is inherently local due to the usage of a cut-off radius and the neglect of $\eta_{\text{GS}}(\mathbf{r}')$ in the vicinity of $\mathbf{r}$ for prediction. Assuming that the locality of the η_{GS} approximation introduces only small errors but otherwise reproduces η_{GS} correctly, it seems likely that the Thomas-Fermi functional, depending explicitly only on the local electron density, would show the smallest $\left|E_{\text{pred}} - E_{\text{GS}}\right|$ deviation. For the current proposal to compute $\eta_{\text{pred}}(\mathbf{r})$, excluding $\eta_{\text{GS}}(\mathbf{r}')$ and $\nabla_{\mathbf{r}}\eta_{GS}(\mathbf{r})$, this deviation would be expected to increase with increasing usage of gradient and non-local information in the kinetic energy operators, i.e. von Weizsäcker, Wang-Teter and Wang-Govind-Carter, as implemented in Profess.

EAM Potentials for Al-Ni alloys

The entire Al-Ni alloy database was represented using the 2-body approximation for $\rho(r)$ assuming that the embedding density, $\overline{\rho}$, equals $\eta_{GS}(\boldsymbol{r})$. A limitation preventing the direct use of the obtained $\eta^{\text{2-body}}(r)$ for $\rho(r)$ is the requirement that $\overline{\rho} \geq 0$ in the EAM *setfl* file format leading to necessary modifications to $\eta^{\text{2-body}}(r)$. However, the density approximations themselves near the atom cores were found to have an R^2 score of at least 0.87 (for hcp Al) and an RMSE of mostly less than 10^{-1} Å^{-3}. The electron density prediction errors are below what is desirable for accurate OF DFT energies but acceptable for EAM since the characteristics near the core are still captured, particularly considering that only a 2-body approximation was used which was shown earlier to lead to coarser approximations than when the 3-body approximation is included. The resulting potentials optimized with the CMA-ES and BFGS methods using $\log\prod_{i=1}^{N} p(E_i, \mathbf{F}_i, \mathbf{w}|\boldsymbol{\alpha}, \beta_E, \beta_F)$ as a fitness measure accurately reproduce energies and forces to within 10 meV atom^{-1} and 0.1 eV atom^{-1} respectively. In this process all structures available from DFT were used in the training set. This is partly due to the cost of generating more structures and because the testing can be done by calculating the emergent properties of the potentials. Also, the extrapolation and interpolation properties can be tested and were found to lead to smooth behaviour using volumetric strains between -4.5 and 4.5 %, exceeding the training set strains.

The emergent properties were found to depend on the training set itself in intricate ways. Ground state energy and cell size obtained by performing geometry optimization with the developed potentials, for example, were found to accurately reproduce the reference values in the case of fcc Al and fcc Ni and reasonably well for γ' Ni_3Al. ISF and APB energies were poorly reproduced. The elastic constants on the other hand, were quite close to the reference values for fcc Ni and fcc Al but almost distributed uniformly at random for γ'. Understanding the relationship between the training set and emergent properties is reminiscent of the credit assignment problem where action and reward are separated sufficiently enough in time to make optimization difficult [151]. Comparing the different training sets for Al itself and the differences between Al, Ni and Al-Ni we think it is likely that introduction of further deformed structures for γ' would improve the predicted mechanical properties and the addition of configurations with ISFs and APBs would improve their predicted formation energies.

7.4 Conclusion

In this chapter we proposed a method for describing ground state electron densities in solids in terms of linear models using Bayesian regression. The output of those linear models are functions which may be used to systematically develop empirical potentials of two or more elements, e.g. in the form of an embedding density function, or to predict $\eta_{\mathrm{GS}}(\mathbf{r})$. The error predicting $\eta_{\mathrm{GS}}(\mathbf{r})$ was quantified using the R^2 score, the RMSE and the energy difference $E\left[\eta_{\mathrm{pred}}(\mathbf{r})\right] - E\left[\eta_{\mathrm{GS}}(\mathbf{r})\right]$ using orbital free DFT for various phases of Al. The maximum test set error was found to be within 0.2293 meV atom^{-1} and 50.153 meV atom^{-1} for 2&3-body and 2-body approximations respectively with R^2 scores better than 0.999 in both cases. The method was successfully tested on solids with metallic, covalent and ionic bonds and different lattice types: H_2 crystal, Al (fcc, bcc & hcp jointly), NaCl, $BaTiO_3$, single-wall carbon nanotubes and W (bcc) & W_3O.

EAM potentials were developed based on 2-body approximations to $\eta_{GS}(\mathbf{r})$ for a database of the Al-Ni alloy system which accurately reproduced energies and forces, with one potential having a mean absolute energy error of only 0.28 meV atom^{-1}. The emergent properties of 910 resulting EAM potentials were found to be sensitive to the choice of structures in the training set. These structures implicitly contained information about emerging properties which when included in the training set led to a narrowing of the predicted properties around reference values. Suggestions were made for systematic improvement to these properties.

Chapter 8

Summary and Outlook

In this chapter we will begin by summarizing the results regarding the simulation of the $\gamma \rightarrow \gamma'$ ordering and the GB motion in superalloys. This will be followed by the development of semi-empirical potentials for these and other materials. Suggestions for future research aimed at an improved understanding of GB motion and phase transformations in superalloys at the atomistic level will be presented. Lastly, we will outline the parallels between the development of semi-empirical potentials and aspects of artificial intelligence research and suggest ways to develop these potentials for multi-component systems, with the EAM as the current example.

8.1 Summary

8.1.1 Ordering and GB Motion in Superalloys

The first major part of this thesis research was the development of robust Bayesian classifiers to investigate $\gamma \rightarrow \gamma'$ ordering reactions in supersaturated binary Ni-Al alloys. Robust, in this context, means reliably identifying γ' in a highly supersaturated γ phase, such as Ni-20 at.% Al as well as GBs in the fcc lattice at temperatures up to 1500 K.

To this end, we employed bond order parameters (BOP), which we made chemically sensitive 'like' and 'unlike' species environments, denoted by $q_{\mathrm{like},l}$ and $q_{\mathrm{unlike},l}$. For the classification of chemical order, BOP distributions of $q_{\mathrm{like},l}$ and $q_{\mathrm{unlike},l}$ were split into separate PDFs for the classifier. This was done to reduce the dimension of the feature space as well as the computational cost of the splines. Splined PDFs were created in a supervised fashion directly over entire BOP distributions for specific phases, leading to PDFs for γ and γ'.

Using Gaussians as analytic approximations to the peaks of BOP distributions together with the EM algorithm enabled the development of the decomposition algorithm. The decomposition algorithm is motivated by the observation that the lattice types, ordered structures, and defects themselves, exhibit characteristic BOP distributions, which are then superposed in complex systems. The application of the decomposition algorithm allowed learning multiple classes by comparing the distributions obtained from multiple trajectories. Note that with this algorithm, more classes can be learned than there are trajectories available, i.e. comparing trajectories for fcc single crystals, a bcc / fcc bicrystal, and a bcc bicrystal, would allow the identification of fcc, bcc and both GBs in both bicrystals (the GBs do not have to be of the same type). This decomposition approach allows the generation of classes for defects which enabled us to track the motion of GBs as well as the ordering at the same time.

Investigating γ phase supersaturated single crystals of Ni-20 at.% Al at temperatures between 1300 K and 1500 K using MD and an EAM potential we observed a significant, relatively homogeneous, vacancy-assisted γ' ordering. The single crystals were simulated for times of up to 100 ns. The ordering was found to occur at a higher rate with increased temperature and showed a positive correlation between $c_{\gamma'}(z,t)$ and $c_{\mathrm{Al}}(z,t)$ in space and time, with barely noticeable changes in composition. The cluster-size distributions of γ'-like atoms, determined via hierarchical clustering, were modelled as functions of time, exhibiting the characteristics of the diffusion equation from Becker-Doehring theory. We furthermore observed neighbouring clusters with [001] APBs relative to each other.

In order to investigate the relation between the solutes, GBs and γ' ordering, we simulated fcc Ni coincidence site lattice (CSL) bicrystals at 1500 K under different alloying conditions using MD with the same EAM potential as for the single crystals. The systems were set up in such a way that the GBs were normal to z in the computational cell. The chemical conditions included pure Ni, Ni-10 at.% Al, Ni-20 at.% Al and the same systems including a 10 Å radius γ' precipitate for a range of symmetric CSL GBs including both twist and tilt configurations. The GBs were tracked along the longest axis of the simulation box, the z-axis, using the mean position, since they exhibited significant fluctuation along directions of the x and y axes, increasing in amplitude with increasing temperature. Modelling the GB motion as a continuous time random walk (CTRW), we found that most GB types were best approximated with Brownian-like motion. However, there was a difference between the tilt and twist GBs: the tilt GBs were much better approximated with this type of motion than were the twist GBs, possibly due to the presence of several active mechanisms in the twist GBs.

The solutes were found to have a significant effect on the GB behaviour, depending on the

GB type. While the range of GB types studied is deemed too small to identify a trend with respect to misorientation, we found that most of the tested tilt GBs, except $\Sigma 27b$, showed mostly normal to sub-diffusive behaviour whereas the twist GBs, except $\Sigma 85$, showed more normal to super-diffusive behaviour. A particularly dramatic change was observed for the $\Sigma 13$ twist GB, where an increase in Al concentration led to a transition from sub-diffusive behaviour to super-diffusive behaviour. Tilt GBs mostly fluctuated around their original positions; with the exception of the immobile $\Sigma 3$ tilt / twin GB, twist GBs showed frequent directed movement.

The presence of pre-existing γ' precipitates did not have a significant impact on the direction of the GB movement other than functioning as an apparent obstacle to their passage. More importantly, the ordering observed in the vicinity of the GBs and in their wake, was significantly increased relative to single crystals for both tilt and twist GBs, an early indicator of heterogeneous precipitation. Because of the similarity of the ordering and (de-)alloying driving forces in our simulations and DIGM, as well as the structural similarity of the ordering adjacent to / in the wake of GBs and (de-)alloyed zones in DIGM, we suggest that our observations are a form of DIGM. DIGM itself is associated with the formation of wavy GBs and discontinuous precipitation, both of which are exhibited by serrated GBs in Ni-base superalloys and could therefore be part of the initiation of GB serration.

8.1.2 Semi-Empirical Potentials

Noting the importance of $\eta_{GS}(\mathbf{r})$ for the properties of crystals in their ground state, we developed a method to accurately represent $\eta_{GS}(\mathbf{r})$ in order to develop EAM potentials for superalloys. The method approximates $\eta_{GS}(\mathbf{r})$ for multiple structures directly as a function of the relative position and chemical species of the atoms in given configurations, using linear regression. Comparing a range of linear regression methods with cosine basis sets, we identified relevance vector machines (RVMs) as the preferred choice. RVMs consistently led to among the best three fits to the DFT data, but also allow a high model sparsity. The functions obtained with the RVMs were found to accurately reproduce the training and test set data $\eta_{GS}(\mathbf{r})$ of configurations with metallic, covalent and ionic contributions. The use of up to 3-body approximations led to highly accurate predictions of $\eta_{GS}(\mathbf{r})$ with predicted OF DFT energy deviations of around 10 meV atom^{-1} for the Al system, $BaTiO_3$ and NaCl, and energy deviations of around 100 meV atom^{-1} for single wall carbon nanotubes, H_2-crystals and the tungsten system. The reason for the larger deviations for the H_2-crystals and carbon nanotubes is thought to be the lower packing density of their configurations used for the computations, relative to the cut-off radius. The reason for the larger deviations for the tungsten system is the use of only 2-body terms.

In order to develop EAM potentials for superalloys, we tested the approach on the Al-Ni system using a potential from the literature [152], which was also used to simulate the single and bicrystal systems, as a reference. For the EAM potentials, $\eta_{GS}(\mathbf{r})$ near-core density values were regressed using the 2-body approximation for all structures in the Al-Ni alloy database. The accuracy of the predictions was surprisingly good for all structures, given the naive assumption of 2-body approximations. Using the embedding density functions obtained from the 2-body approximations, we fit the EAM potentials against the DFT energies and forces using a genetic algorithm. Independent of the structures included in the training set, the DFT energies and forces were always found to be reproduced to high accuracy. The 'best' EAM potential found for the Al-Ni system had a mean absolute energy error over all structures of 0.28 meV atom^{-1} and force errors of less than 0.1 eV atom^{-1}.

However, properties not included in the training process, which we refer to as emergent properties, such as the elastic constants and stacking faults, were predicted to varying levels of accuracy depending on the chosen training set. Interestingly, we found that including strained crystals in the training set improved the prediction of the elastic constants for the resulting potentials. The inclusion of structures from *ab initio* MD similarly improved the predicted phonon spectrum. Thus the regressed potentials do 'learn' physical properties as inherent patterns of the provided data to some extent, which certainly is limited by the potential type itself. As a rule of thumb, we find that including structures related to the properties of interest improves the accuracy of the findings relevant to those properties. Below, we make some suggestions for a systematic approach to quantifying the impact of the structures on the properties of the resulting potential, based on our findings.

8.2 Outlook

8.2.1 Ordering & GB motion in Superalloys

The advantage of the Bayesian classification approach using the decomposition algorithm is its high flexibility, allowing a problem-based development of classifiers, directly using the targeted structures. Thus, we think it is likely that transformations of more complex phases should be observable with the methods developed here. The approach itself is also not limited to BOPs as features. It would be of interest to apply these methods to study the following systems in single and bicrystals.

For single crystals it would be worth studying the vacancy–cluster and the cluster–cluster interactions during continued ordering and at finer time resolutions. This should be feasible with the proposed classification method if steps are taken to include the identification of vacancies. A natural next step would be to add additional elements as solutes and study their impact on the ordering, e.g. Nb, which could potentially slow down the process.

For bicrystals as well, there is a range of possible aspects worthy of study. From a statistical point of view, one would want to increase the number of GB systems simulated as well as include inclined and general GBs, ideally using multiple empirical potentials. Of particular interest for GB migration would be the study of artificial driving forces, changes in the chemical composition, and pre-existing precipitates with different solutes that alter the γ/γ' lattice misfit.

8.2.2 Semi-Empirical Potentials

In the course of this research we encountered a number of problems in training EAM potentials which have not (or only partially) been solved so far. As two main issues to be solved for any given semi-empirical potentials, we see:

- predictive uncertainties in the energies, forces, and emergent properties (which increase with increasing distance from the training set); and
- relation between training set and resulting properties of semi-empirical potentials.

Knowledge of the predictive uncertainties would be useful for the estimation of the significance of the predictions of semi-empirical potentials and would therefore be useful in the selection process before the simulation of physical systems. Insight into the relation between the training set and the semi-empirical potentials would be useful for the estimation of the limitations of the semi-empirical potential types themselves and for the optimization of the size of the training set.

In general, the training of semi-empirical potentials is, at its core, a regression problem, albeit a peculiar one. Hence, we think it is likely that future solutions to these problems are related to tools and approaches developed in the artificial intelligence community, which is also working on these two issues, among others. Artificial neural nets (ANNs), for example, have already been successfully applied as empirical potentials [205]. As an alternative to creating new types of potential, one could also frame well-established semi-empirical potentials in ANN-like terms, as in fig. 8.1 for EAM. The ANN framework could possibly lead to the discovery of new modifications of semi-empirical potentials using genetic programming and reinforcement learning [7, 225]. ANNs appear a promising approach because they are able to deal with more complex problems or identify higher order concepts by increasing the number of their layers [21, 22, 199] and they can be used to compute uncertainties when used in a Bayesian framework.

The study of the impact of selected training samples on the resulting empirical potentials and their properties could be approached via active learning / experimental design [208], where the training set is dynamically updated by actively posing queries for samples which are expected to contain valuable insights for further improvements of the model. Tong *et al.* [237], for example, actively queried a database of samples which trained support vector machines (SVMs) to classify texts, with the objective of using as few samples as possible so as to lead to the fastest possible convergence of the classifier. The application of active learning to semi-empirical potentials could be made by calculating the uncertainties of the energies, forces, and emergent properties, and the changes of these uncertainties as samples are added to the training set.

The combination of both predictive uncertainties and active learning could therefore be supposed to improve the resulting semi-empirical potentials, their interpretability, and the training process.

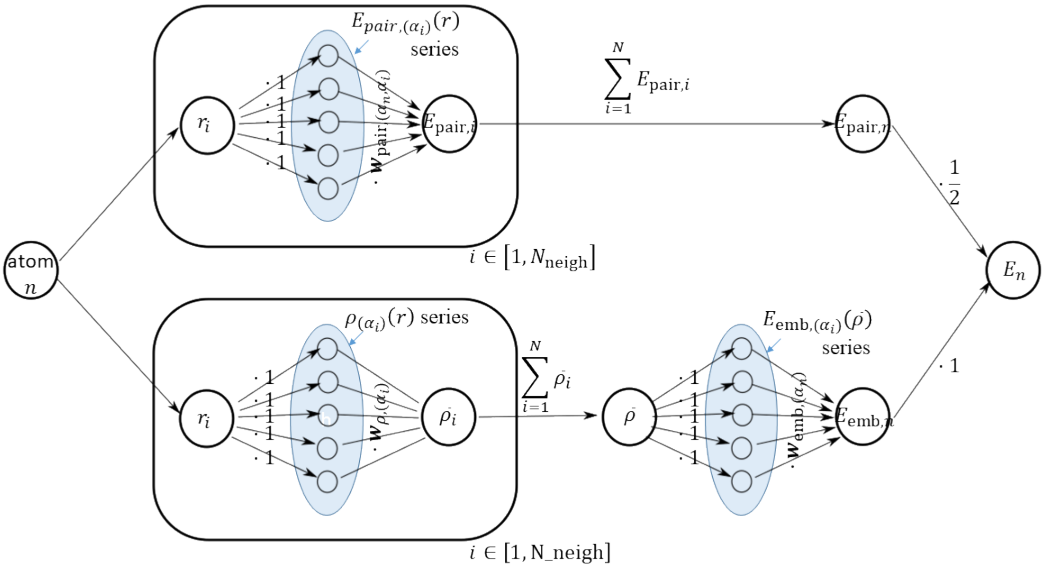


Figure 8.1: Representation of the calculation of the energy contribution E_n of atom n to the total energy for the Embedded Atom Method potential type in the form of an artificial neural network. Using the neighbouring information, the pair and embedding energy contributions are represented as two paths with hidden layers (blue ellipsoids).

References

[1] G Ackland and A Jones. Applications of local crystal structure measures in experiment and simulation. *Physical Review B*, 73(5):054104, feb 2006. ISSN 1098-0121. doi: 10.1103/PhysRevB.73.054104.

[2] S Allen and J Cahn. Mechanisms of phase transformations within the miscibility gap of Fe-rich Fe-Al alloys. *Acta Metallurgica*, 24(5):425–437, may 1976. ISSN 00016160. doi: 10.1016/0001-6160(76)90063-8.

[3] M Athènes, P Bellon, G Martin, and F Haider. A Monte-Carlo study of B2 ordering and precipitation via vacancy mechanism in b.c.c. lattices. *Acta Materialia*, 44(12): 4739–4748, 1996. ISSN 1359-6454. doi: 10.1016/S1359-6454(96)00135-8.

[4] S Auer and D Frenkel. Numerical prediction of absolute crystallization rates in hard-sphere colloids. *Journal of Chemical Physics*, 120:3015–3029, 2004. ISSN 00219606. doi: 10.1063/1.1638740.

[5] S Babcock and R Balluffi. Grain boundary kinetics - II. In situ observations of the role of grain boundary dislocations in high-angle boundary migration. *Acta Metallurgica*, 37(9):2367–2376, sep 1989. ISSN 00016160. doi: 10.1016/0001-6160(89)90034-5.

[6] S Babu, M Miller, J Vitek, and S David. Characterization of the microstructure evolution in a Nickel base superalloy during continuous cooling conditions. *Acta Materialia*, 49:4149–4160, 2001. doi: 10.1016/S1359-6454(01)00314-7.

[7] B Baker, O Gupta, N Naik, and R Raskar. Designing Neural Network Architectures using Reinforcement Learning. In *Proceedings of the 5th International Conference on Learning Representations*, pages 1–16, 2017. ISBN 2857825749.

[8] R Balluffi and G Olson. On the hierarchy of interfacial dislocation structure. *Metallurgical Transactions A*, 16(4):529–541, apr 1985. ISSN 0360-2133. doi: 10.1007/BF02814227.

[9] A Banadaki and S Patala. An efficient algorithm for computing the primitive bases of a general lattice plane. *Applied Crystallography*, 48:585–588, 2015. doi: 10.1107/S1600576715004446.

[10] A Banadaki and S Patala. A simple faceting model for the interfacial and cleavage energies of Σ 3 grain boundaries in the complete boundary plane orientation space. *Computational Materials Science*, 112:147–160, 2016. ISSN 0927-0256. doi: 10.1016/j.commatsci.2015.09.062.

[11] World bank. world wide air travel. http://data.worldbank.org/indicator/IS.AIR.PSGR, 2017.

[12] World bank. world wide energy consumption. http://data.worldbank.org/indicator/EG.USE.PCAP.KG.OE, 2017.

[13] A Bartók and G Csányi. Gaussian Approximation Potentials: A Brief Tutorial Introduction. *International Journal of Quantum Chemistry*, 115:1051–1057, 2015. doi: 10.1002/qua.24927.

[14] A Bartók, M Payne, R Kondor, and G Csányi. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Physical Review Letters*, 104:136403, 2010. doi: 10.1103/PhysRevLett.104.136403.

[15] A Bartók, M Gillan, F Manby, and G Csányi. Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water. *Physical Review B*, 88:054104, 2013. ISSN 10980121. doi: 10.1103/PhysRevB.88.054104.

[16] A Bartók, R Kondor, and G Csányi. On representing chemical environments. *Physical Review B*, 87:184115, 2013. ISSN 10980121. doi: 10.1103/PhysRevB.87.184115.

[17] M Baskes. Application of the Embedded-Atom Method to Covalent Materials: A Semiempirical Potential for Silicon. *Physical Review Letters*, 59(23):2666–2669, 1987. doi: 10.1103/PhysRevLett.59.2666.

[18] M Baskes. Modified embedded-atom potentials for cubic materials and impurities. *Physical Review B*, 46(5):2727–2742, 1992. ISSN 01631829. doi: 10.1103/PhysRevB.46.2727.

[19] M Baskes, J Nelson, and F Wright. Semiempirical modified embedded-atom potentials for silicon and germanium. *Physical Review B*, 40(9):6085–6100, 1989. doi: 10.1103/PhysRevB.40.6085.

[20] J Behler. Perspective: Machine learning potentials for atomistic simulations. *The Journal of chemical physics*, 145:170901, 2016. doi: 10.1063/1.4966192.

[21] Y Bengio. *Learning Deep Architectures for AI*, volume 2. 2009. ISBN 2200000006. doi: 10.1561/2200000006.

[22] Y Bengio, A Courville, and P Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8):1798–1828, 2013. ISSN 01628828. doi: 10.1109/TPAMI.2013.50.

[23] J Bergstra and Y Bengio. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13:281–305, 2012. ISSN 1532-4435. doi: 10.1162/153244303322533223.

[24] C Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, Cambridge, 1st edition, 2009. ISBN 978-0387310732.

[25] D Blavette, E Cadel, and B Deconihout. Role of the Atom Probe in the Study of Nickel-Based Superalloys. *Materials Characterization*, 44(1):133–157, 2000. ISSN 10445803. doi: 10.1016/S1044-5803(99)00050-9.

[26] D Blavette, E Cadel, C Pareige, B Deconihout, and P Caron. Phase Transformation and Segregation to Lattice Defects in Ni-Base Superalloys. *Microscopy and microanalysis*, 13:464–483, 2007. ISSN 1431-9276. doi: 10.1017/S143192760707078X.

[27] V Botu and R Ramprasad. Learning scheme to predict atomic forces and accelerate materials simulations. *Physical Review B*, 92:94306, 2015. ISSN 1550235X. doi: 10.1103/PhysRevB.92.094306.

[28] V Botu and R Ramprasad. Adaptive Machine Learning Framework to Accelerate Ab initio Molecular Dynamics. *International Journal of Quantum Chemistry*, 115: 1074–1083, 2015. ISSN 1097461X. doi: 10.1002/qua.24836.

[29] D Brenner. Relationship between the Embedded-Atom Method and Tersoff Potentials. *Physical Review Letters*, 63(9):1022, 1989. doi: 10.1103/PhysRevLett.63.1022.

[30] F Brockherde, L Vogt, L Li, M Tuckerman, K Burke, and K Müller. Bypassing the Kohn-Sham equations with machine learning. *Nature Communications*, 8(1):872, dec 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-00839-3.

[31] N Burbery, R Das, and W Ferguson. Thermo-kinetic mechanisms for grain boundary structure multiplicity , thermal instability and defect interactions. *Materials Chemistry and Physics*, 179:254–265, 2016. ISSN 0254-0584. doi: 10.1016/j.matchemphys.2016.05.037.

[32] J Cahn. On spinodal decomposition. *Acta Metallurgica*, 9:795–801, 1961. ISSN 00016160. doi: 10.1016/0001-6160(61)90182-1.

[33] J Cahn, Y Mishin, and A Suzuki. Coupling grain boundary motion to shear deformation. *Acta Materialia*, 54:4953–4975, 2006. ISSN 13596454. doi: 10.1016/j.actamat.2006.08.004.

[34] P Caron and T Khan. Improvement of Creep Strength in a Nickel-base Single-crystal Superalloy by Heat Treatment. *Materials Science and Engineering*, 61:173–184, 1983. doi: 10.1016/0025-5416(83)90199-4.

[35] J Carter, N Zhou, J Sosa, P Shade, A Pilchak, M Kuper, Y Wang, H Fräser, M Uchic, and M Mills. CHARACTERIZATION OF STRAIN ACCUMULATION AT GRAIN BOUNDARIES OF NICKEL-BASED SUPERALLOYS. *Superalloys*, pages 43–52, 2012.

[36] L Chen and A Khachaturyan. Computer simulation of structural transformations during precipitation of an ordered intermetallic phase. *Acta Metallurgica et Materialia*, 39(11):2533–2551, nov 1991. ISSN 09567151. doi: 10.1016/0956-7151(91)90069-D.

[37] L Chen and A Khachaturyan. Computer simulation of decomposition reactions accompanied by a congruent ordering of the second kind. *Scripta Metallurgica et Materialia*, 25(1):61–66, jan 1991. ISSN 0956716X. doi: 10.1016/0956-716X(91)90354-4.

[38] L Chen and A Khachaturyan. Kinetics of virtual phase formation during precipitation of ordered intermetallics. *Physical Review B*, 46(10):5899–5905, sep 1992. ISSN 0163-1829. doi: 10.1103/PhysRevB.46.5899.

[39] L Chen and A Khachaturyan. Dynamics of simultaneous ordering and phase separation and effect of long-range Coulomb interactions. *Physical Review Letters*, 70(10): 1477–1480, mar 1993. ISSN 0031-9007. doi: 10.1103/PhysRevLett.70.1477.

[40] J Christian. *Theory of Transformation in Metals and Alloys*. Pergamon Press, Oxford, New York, 2nd edition, 1975. ISBN 0080180310.

[41] S Clark, M Segall, C Pickard, P Hasnip, M Probert, K Refson, and M Payne. First principles methods using CASTEP. *Zeitschrift für Kristallographie - Crystalline Materials*, 220(5/6):567–570, jan 2005. ISSN 2196-7105. doi: 10.1524/zkri.220.5.567.65075.

[42] E Clementi and D Raimondi. Atomic Screening Constants from SCF Functions. *The Journal of Chemical Physics*, 38(11):2686–2689, jun 1963. ISSN 0021-9606. doi: 10.1063/1.1733573.

[43] E Clementi and C Roetti. Roothaan-Hartree-Fock atomic wavefunctions. *Atomic Data and Nuclear Data Tables*, 14(3-4):177–478, sep 1974. ISSN 0092640X. doi: 10.1016/S0092-640X(74)80016-1.

[44] C Cleveland, W Luedtke, and U Landman. Melting of Gold Clusters: Icosahedral Precursors. *Physical Review Letters*, 81(10):2036–2039, sep 1998. ISSN 0031-9007. doi: 10.1103/PhysRevLett.81.2036.

[45] S Curtarolo, W Setyawan, G Hart, M Jahnatek, R Chepulskii, R Taylor, S Wang, J Xue, K Yang, O Levy, M Mehl, H Stokes, D Demchenko, and D Morgan. AFLOW: An automatic framework for high-throughput materials discovery. *Computational Materials Science*, 58:218–226, jun 2012. ISSN 09270256. doi: 10.1016/j.commatsci.2012.02.005.

[46] H Danflou, M Marty, and A Walder. Formation of Serrated Grain Boundaries And Their Effect On The Mechanical Properties In a P/M Nickel Base Superalloy. *TMS Superalloys*, pages 63–72, 1992.

[47] H Danflou, M Macia, T Sanders, and T Khan. Mechanisms of formation of serrated grain boundaries in Nickel base superalloys. *TMS Superalloys*, (5):119–127, 1996.

[48] M Daw and M Baskes. Semiempirical, Quantum Mechanical Calculation of Hydrogen Embrittlement in Metals. *Physical Review Letters*, 50(17):1285–1288, 1983. ISSN 00319007. doi: 10.1103/PhysRevLett.50.1285.

[49] M Daw and M Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical Review B*, 29(12):6443–6453, jun 1984. ISSN 0163-1829. doi: 10.1103/PhysRevB.29.6443.

[50] C Deng and C Schuh. Atomistic Simulation of Slow Grain Boundary Motion. *Physical Review Letters*, 106(4):045503, jan 2011. ISSN 0031-9007. doi: 10.1103/PhysRevLett.106.045503.

[51] C Deng and C Schuh. Diffusive-to-ballistic transition in grain boundary motion studied by atomistic simulations. *Physical Review B*, 84(21):214102, dec 2011. ISSN 1098-0121. doi: 10.1103/PhysRevB.84.214102.

[52] C Dietz and M Thoma. Investigation and improvement of three-dimensional plasma crystal analysis. *Physical Review E*, 94(3):033207, sep 2016. ISSN 2470-0045. doi: 10.1103/PhysRevE.94.033207.

[53] M Doi, T Miyazaki, and T Wakatsuki. The effect of elastic interaction energy on the morphology of γ' precipitates in nickel-based alloys. *Materials Science and Engineering*, 67(2):247–253, nov 1984. ISSN 00255416. doi: 10.1016/0025-5416(84)90056-9.

[54] A Dongare, M Neurock, and L Zhigilei. Angular-dependent embedded atom method potential for atomistic simulations of metal-covalent systems. *Physical Review B*, 80 (18):184106, nov 2009. ISSN 1098-0121. doi: 10.1103/PhysRevB.80.184106.

[55] A Dongare, B LaMattina, D Irving, A Rajendran, M Zikry, and D Brenner. An angular-dependent embedded atom method (A-EAM) interatomic potential to model thermodynamic and mechanical behavior of Al/Si composite materials. *Modelling and Simulation in Materials Science and Engineering*, 20(3):035007, apr 2012. ISSN 0965-0393. doi: 10.1088/0965-0393/20/3/035007.

[56] A Duff, M Finnis, P Maugis, B Thijsse, and M Sluiter. MEAMfit: A reference-free modified embedded atom method (RF-MEAM) energy and force-fitting code. *Computer Physics Communications*, 196:439–445, nov 2015. ISSN 00104655. doi: 10.1016/j.cpc.2015.05.016.

[57] W Eckhardt and A Heinecke. SuperMUC boosts the largest molecular dynamics simulation by 4X in Number of Particles. *inSide*, 11(1), 2013.

[58] J Elliott. Novel approaches to multiscale modelling in materials science. *International Materials Reviews*, 56(4):207–225, jul 2011. ISSN 0950-6608. doi: 10.1179/1743280410Y.0000000002.

[59] D Faken and H Jónsson. Systematic analysis of local atomic structure combined with 3D computer graphics. *Computational Materials Science*, 2(2):279–286, mar 1994. ISSN 09270256. doi: 10.1016/0927-0256(94)90109-0.

[60] D Farkas. Interatomic potentials for Ti-Al with and without angular forces. *Modelling and Simulation in Materials Science and Engineering*, 2(5):975–984, sep 1994. ISSN 0965-0393. doi: 10.1088/0965-0393/2/5/003.

[61] A Faul and M Tipping. Analysis of sparse Bayesian learning. *Advances in Neural Information Processing Systems*, 14:383–389, 2002.

[62] E Fermi. Statistical method to determine some properties of atoms. *Rend. Accad. Naz. Lincei*, 6(December):602–607, 1927. ISSN 1434-6001. doi: 10.1007/BF01351576.

[63] G Ferré, J Maillet, and G Stoltz. Permutation-invariant distance between atomic configurations. *The Journal of Chemical Physics*, 143(10):104114, sep 2015. ISSN 0021-9606. doi: 10.1063/1.4930541.

[64] G Ferré, T Haut, and K Barros. Learning molecular energies using localized graph kernels. *The Journal of Chemical Physics*, 146(11):114107, mar 2017. ISSN 0021-9606. doi: 10.1063/1.4978623.

[65] R Feynman. Forces in Molecules. *Physical Review*, 56(4):340–343, aug 1939. ISSN 0031-899X. doi: 10.1103/PhysRev.56.340.

[66] M Finnis and J Sinclair. A simple empirical N-body potential for transition metals. *Philosophical Magazine A*, 50(1):45–55, jul 1984. ISSN 0141-8610. doi: 10.1080/01418618408244210.

[67] V Fock. Naeherungsmethode zur Loesung des quantenmechanischen Mehrkoerperproblems. *Zeitschrift fuer Physik*, 61(1-2):126–148, jan 1930. ISSN 1434-6001. doi: 10.1007/BF01340294.

[68] F Fortin, F De Rainville, M Gardner, M Parizeau, and C Gagne. DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research*, 13(Jul):2171–2175, 2012.

[69] F Frank and J Kasper. Complex alloy structures regarded as sphere packings. I. Definitions and basic principles. *Acta Crystallographica*, 11(3):184–190, mar 1958. ISSN 0365110X. doi: 10.1107/S0365110X58000487.

[70] F Frank and J Kasper. Complex alloy structures regarded as sphere packings. II. Analysis and classification of representative structures. *Acta Crystallographica*, 12 (7):483–499, jul 1959. ISSN 0365110X. doi: 10.1107/S0365110X59001499.

[71] P Freddolino, F Liu, M Gruebele, and K Schulten. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophysical Journal*, 94(10): L75–L77, may 2008. ISSN 00063495. doi: 10.1529/biophysj.108.131565.

[72] D Frenkel and B Smit. *Understanding Molecular Simulation*. Academic Press, San Diego, San Francisco, New York, Boston, London, Sydney, Tokyo, 2nd edition, 2012. ISBN 9780122673511.

[73] H Gades and A Mitus. Local Structure in Computer-generated liquids and glasses: Classification of three-dimensional patterns. *Physica A*, 176:297–324, aug 1991. ISSN 03784371. doi: 10.1016/0378-4371(91)90291-J.

[74] M Gell, D Duhl, and A Giamei. The development of single crystal superalloy turbine blades. *Superalloys*, pages 205–214, 1980.

[75] A Gelman. Objections to Bayesian statistics. *Bayesian Analysis*, 3(3):445–450, 2008. ISSN 19360975. doi: 10.1214/08-BA318.

[76] T Germann and K Kadau. TRILLION-ATOM MOLECULAR DYNAMICS BECOMES A REALITY. *International Journal of Modern Physics C*, 19(09):1315–1319, sep 2008. ISSN 0129-1831. doi: 10.1142/S0129183108012911.

[77] P Geslin, Y Xu, and A Karma. Morphological Instability of Grain Boundaries in Two-Phase Coherent Solids. *Physical Review Letters*, 114(10):105501, mar 2015. ISSN 0031-9007. doi: 10.1103/PhysRevLett.114.105501.

[78] Z Ghahramani and S Roweis. Learning Nonlinear Dynamical Systems using an EM Algorithm. *Advances in neural information processing systems*, 11(1):431–437, 1999.

[79] F Giustino. *Materials Modelling using Density Functional Theory*. Oxford Press, Oxford, 1st edition, 2014. ISBN 9780199662432.

[80] C Glass, A Oganov, and N Hansen. USPEX–Evolutionary crystal structure prediction. *Computer Physics Communications*, 175(11-12):713–720, dec 2006. ISSN 00104655. doi: 10.1016/j.cpc.2006.07.020.

[81] H Gleiter. The mechanism of grain boundary migration. *Acta Metallurgica*, 17(5): 565–573, 1969. ISSN 00016160. doi: 10.1016/0001-6160(69)90115-1.

[82] I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, A Courville, and Y Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680, 2014. ISBN 1406.2661.

[83] I Goodfellow, Y Bengio, and A Courville. *Deep Learning*. MIT Press, Cambridge, MA; London, England, 1st edition, 2016. ISBN 9780262035613.

[84] V Gorentsveig, P Fratzl, and J Lebowitz. Kinetics of joint ordering and decomposition in binary alloys. *Physical Review B*, 55(5):2912–2919, feb 1997. ISSN 0163-1829. doi: 10.1103/PhysRevB.55.2912.

[85] T Gorkaya, D Molodov, and G Gottstein. Stress-driven migration of symmetrical <100> tilt grain boundaries in Al bicrystals. *Acta Materialia*, 57(18):5396–5405, oct 2009. ISSN 13596454. doi: 10.1016/j.actamat.2009.07.036.

[86] G Gottstein. *Physikalische Grundlage der Materialkunde*. Springer-Verlag, Heidelberg, Berlin, 3rd edition, 2007. ISBN 978-3-540-71104-9.

[87] P Gullet, G Wagner, and A Slepoy. Numerical Tools for Atomistic Simulations. Technical report, Sandia National Laboratories, 2004.

[88] R Hadian, B Grabowski, C Race, and J Neugebauer. Atomistic migration mechanisms of atomically flat, stepped, and kinked grain boundaries. *Physical Review B*, 94(16):165413, oct 2016. ISSN 2469-9950. doi: 10.1103/PhysRevB.94.165413.

[89] J Hafner, C Wolverton, and G Ceder. Toward Computational Materials Design: The Impact of Density Functional Theory on Materials Research. *MRS Bulletin*, 31(09): 659–668, sep 2006. ISSN 0883-7694. doi: 10.1557/mrs2006.174.

[90] A Halevy, P Norvig, and F Pereira. The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*, 24(2):8–12, 2009. ISSN 1541-1672. doi: 10.1109/MIS.2009.36.

[91] E Hall and S Algie. The Sigma Phase. *Metallurgical Reviews*, 11(1):61–88, jan 1966. ISSN 0076-6690. doi: 10.1179/mtlr.1966.11.1.61.

[92] B Hammer, K Jacobsen, V Milman, and M Payne. Stacking Fault Energy in Aluminium. *J. Phys. Condens. Matter*, 4:10453–10460, 1992.

[93] C Handley and J Behler. Next generation interatomic potentials for condensed systems. *The European Physical Journal B*, 87(7):152, jul 2014. ISSN 1434-6028. doi: 10.1140/epjb/e2014-50070-0.

[94] N Hansen. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In *Proceedings of the 11th annual conference companion on Genetic and evolutionary computation conference - GECCO '09*, page 2389, New York, New York, USA, 2009. ACM Press. ISBN 9781605585055. doi: 10.1145/1570256.1570333.

[95] N Hansen and A Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, jun 2001. ISSN 1063-6560. doi: 10.1162/106365601750190398.

[96] D Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(1):89–110, 1928. ISSN 0305-0041. doi: 10.1017/S0305004100011919.

[97] D Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(1):89–110, 1928. ISSN 0305-0041. doi: 10.1017/S0305004100011919.

[98] H Hellmann. *Einfuehrung in die Quantenchemie*. F. Deuticke, Leipzig, 1937.

[99] M Henry, Y Yoo, D Yoon, and J Choi. The dendritic growth of γ' precipitates and grain boundary serration in a model nickel-base superalloy. *Metallurgical Transactions A*, 24A(August):1733–1743, 1993.

[100] M Hillert. A solid-solution model for inhomogeneous systems. *Acta Metallurgica*, 9 (6):525–535, jun 1961. ISSN 00016160. doi: 10.1016/0001-6160(61)90155-9.

[101] G Ho, V Lignères, and E Carter. Introducing PROFESS: A new program for orbital-free density functional theory calculations. *Computer Physics Communications*, 179 (11):839–854, dec 2008. ISSN 00104655. doi: 10.1016/j.cpc.2008.07.002.

[102] P Hohenberg and W Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B): 864–871, 1964. ISSN 01631829. doi: 10.1103/PhysRevB.7.1912.

[103] Y Hold-Geoffroy, O Gagnon, and M Parizeau. Once you SCOOP, no need to fork. In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment - XSEDE '14*, pages 1–8, New York, New York, USA, 2014. ACM Press. ISBN 9781450328937. doi: 10.1145/2616498.2616565.

[104] J Honeycutt and H Andersen. Molecular dynamics study of melting and freezing of small Lennard-Jones clusters. *The Journal of Physical Chemistry*, 91(19):4950–4963, sep 1987. ISSN 0022-3654. doi: 10.1021/j100303a014.

[105] H Hong and S Nam. The occurrence of grain boundary serration and its effect on the M23C6 carbide characteristics in an AISI 316 stainless steel. *Materials Science and Engineering: A*, 332(1-2):255–261, jul 2002. ISSN 09215093. doi: 10.1016/S0921-5093(01)01754-3.

[106] H Hong, I Kim, B Choi, M Kim, and C Jo. The effect of grain boundary serration on creep resistance in a wrought nickel-based superalloy. *Materials Science & Engineering A*, 517:125–131, 2009. doi: 10.1016/j.msea.2009.03.071.

[107] H Hong, H Jeong, I Kim, B Choi, Y Yoo, and C Jo. Significant decrease in interfacial energy of grain boundary through serrated grain boundary transition. *Philosophical Magazine*, 92(22):2809–2825, aug 2012. ISSN 1478-6435. doi: 10.1080/14786435.2012.676212.

[108] H Hong, I Kim, B Choi, Y Yoo, and C Jo. ON THE MECHANISM OF SERRATED GRAIN BOUNDARY FORMATION IN NI-BASED SUPERALLOYS WITH LOW γ' VOLUME FRACTION. *Superalloys*, pages 53–61, 2012.

[109] T Hsieh and R Balluffi. Observations of roughening/de-faceting phase transitions in grain boundaries. *Acta Metallurgica*, 37(8):2133–2139, aug 1989. ISSN 00016160. doi: 10.1016/0001-6160(89)90138-7.

[110] D Hunter and K Lange. A Tutorial on MM Algorithms. *The American Statistician*, 58(1):30–37, feb 2004. ISSN 0003-1305. doi: 10.1198/0003130042836.

[111] A Jain, S Ong, G Hautier, W Chen, W Richards, S Dacek, S Cholia, D Gunter, D Skinner, G Ceder, and K Persson. Commentary: The Materials Project: A materials

genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, jul 2013. ISSN 2166-532X. doi: 10.1063/1.4812323.

[112] T Jamil and C ter Braak. Selection properties of type II maximum likelihood (empirical Bayes) in linear models with individual variance components for predictors. *Pattern Recognition Letters*, 33(9):1205–1212, jul 2012. ISSN 01678655. doi: 10.1016/j.patrec.2012.01.004.

[113] K Janssens, D Olmsted, E Holm, S Foiles, S Plimpton, and P Derlet. Computing the mobility of grain boundaries. *Nature Materials*, 5(2):124–127, feb 2006. ISSN 1476-1122. doi: 10.1038/nmat1559.

[114] R Jhan and P Bristowe. A molecular dynamics study of grain boundary migration without the participation of secondary grain boundary dislocations. *Scripta Metallurgica et Materialia*, 24(7):1313–1318, jul 1990. ISSN 0956716X. doi: 10.1016/0956-716X(90)90348-K.

[115] L Jiang, R Hu, H Kou, J Li, G Bai, and H Fu. The effect of M23C6 carbides on the formation of grain boundary serrations in a wrought Ni-based superalloy. *Materials Science and Engineering: A*, 536:37–44, feb 2012. ISSN 09215093. doi: 10.1016/j.msea.2011.11.060.

[116] E Jones, T Oliphant, and P Peterson. SciPy: Open source scientific tools for Python, 2001.

[117] C Kelchner, S Plimpton, and J Hamilton. Dislocation nucleation and defect structure during surface indentation. *Physical Review B*, 58(17):11085–11088, nov 1998. ISSN 0163-1829. doi: 10.1103/PhysRevB.58.11085.

[118] C Kelley. *Iterative Methods for Optimization*. Society for Industrial and Applied Mathematics, jan 1999. ISBN 978-0-89871-433-3. doi: 10.1137/1.9781611970920.

[119] A Khachaturyan, T Lindsey, and J Morris. Theoretical investigation of the precipitation of δ' in Al-Li. *Metallurgical Transactions A*, 19(2):249–258, feb 1988. ISSN 0360-2133. doi: 10.1007/BF02652533.

[120] K Kim, H Hong, and S Nam. A study on the mechanism of serrated grain boundary formation in an austenitic stainless steel. *Materials Chemistry and Physics*, 126(3):480–483, apr 2011. ISSN 02540584. doi: 10.1016/j.matchemphys.2010.12.025.

[121] A King. Diffusion induced grain boundary migration. *International Materials Reviews*, 32(1):173–189, jan 1987. ISSN 0950-6608. doi: 10.1179/095066087790150304.

[122] S Kirkpatrick. Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5-6):975–986, mar 1984. ISSN 0022-4715. doi: 10.1007/BF01009452.

[123] W Kohn and L Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, nov 1965. ISSN 0031-899X. doi: 10.1103/PhysRev.140.A1133.

[124] D Koller and N Friedman. *Probabilistic Graphical Models - Principles and Techniques*. The MIT Press, London, England, 2009.

[125] A Koul and G Gessinger. On the mechanism of serrated grain boundary formation in Ni-based superalloys. *Acta Metallurgica*, 31(7):1061–1069, jul 1983. ISSN 00016160. doi: 10.1016/0001-6160(83)90202-X.

[126] A Krizhevsky and G Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *NIPS Proceedings*, pages 1–9, 2012.

[127] S Kumar, J Rosenberg, D Bouzida, R Swendsen, and P Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8):1011–1021, oct 1992. ISSN 0192-8651. doi: 10.1002/jcc.540130812.

[128] A Laio and M Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, oct 2002. ISSN 0027-8424. doi: 10.1073/pnas.202427399.

[129] P Larsen, S Schmidt, and J Schiotz. Robust structural identification via polyhedral template matching. *Modelling and Simulation in Materials Science and Engineering*, 24(5):055007, jun 2016. ISSN 0965-0393. doi: 10.1088/0965-0393/24/5/055007.

[130] J Larson and S Floreen. Metallurgical factors affecting the crack growth resistance of a superalloy. *Metallurgical Transactions A*, 8(1):51–55, jan 1977. ISSN 0360-2133. doi: 10.1007/BF02677263.

[131] F Latief, H Hong, T Blanc, I Kim, B Choi, C Jo, and J Lee. Influence of chromium content on microstructure and grain boundary serration formation in a ternary Ni-xCr-0.1C model alloy. *Materials Chemistry and Physics*, 148(3):1194–1201, dec 2014. ISSN 02540584. doi: 10.1016/j.matchemphys.2014.09.047.

[132] D Laughlin and W Soffa. Spinodal Structures. In *ASM Handbook*, pages 652–654. 9th edition, 1985.

[133] W Lechner and C Dellago. Accurate determination of crystal structures based on averaged local bond order parameters. *The Journal of Chemical Physics*, 129(11): 114707, sep 2008. ISSN 0021-9606. doi: 10.1063/1.2977970.

[134] C Ledig, L Theis, F Huszár, J Caballero, A Cunningham, A Acosta, A Aitken, A Tejani, J Totz, Z Wang, and W Shi Twitter. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *Cvpr*, pages 4681–4690, 2017. ISSN 0018-5043. doi: 10.1109/CVPR.2017.19.

[135] J Lee, D Kim, and H Hong. A new approach to strengthen grain boundaries for creep improvement of a Ni–Cr–Co–Mo superalloy at 950°C. *Materials Science and Engineering: A*, 625:164–168, feb 2015. ISSN 09215093. doi: 10.1016/j.msea.2014.12.010.

[136] L Letellier, S Chambreland, P Duval, and D Blavette. Grain boundary segregation in nickel base superalloys Astroloy: an atom-probe study. *Applied Surface Science*, 67 (1-4):305–310, apr 1993. ISSN 01694332. doi: 10.1016/0169-4332(93)90330-E.

[137] H Li, J Sun, M Hardy, H Evans, S Williams, T Doel, and P Bowen. Effects of microstructure on high temperature dwell fatigue crack growth in a coarse grain PM nickel based superalloy. *Acta Materialia*, 90:355–369, may 2015. ISSN 13596454. doi: 10.1016/j.actamat.2015.02.023.

[138] J Li. AtomEye: an efficient atomistic configuration viewer. *Modelling and Simulation in Materials Science and Engineering*, 11(2):173–177, mar 2003. ISSN 0965-0393. doi: 10.1088/0965-0393/11/2/305.

[139] L Liang and A King. Diffusion induced grain boundary migration in the zinc-cadmium system. *Acta Materialia*, 44(7):2983–2998, jul 1996. ISSN 13596454. doi: 10.1016/1359-6454(95)00391-6.

[140] D Liu, W Miller, and K Aust. Diffusion induced grain boundary migration in Ni-Cu Diffusion couples. *Acta Metallurgica*, 37(12):3367–3378, dec 1989. ISSN 00016160. doi: 10.1016/0001-6160(89)90209-5.

[141] N Lümmen and T Kraska. Common neighbour analysis for binary atomic systems. *Modelling and Simulation in Materials Science and Engineering*, 15(3):319–334, apr 2007. ISSN 0965-0393. doi: 10.1088/0965-0393/15/3/010.

[142] C Ma and W Gust. Kinetic behavior of diffusion induced grain boundary migration. *Scripta Metallurgica et Materialia*, 30(4):509–514, feb 1994. ISSN 0956716X. doi: 10.1016/0956-716X(94)90612-2.

[143] C Ma, E Rabkin, W Gust, and S Hsu. On the kinetic behavior and driving force of diffusion induced grain boundary migration. *Acta Metallurgica Et Materialia*, 43(8): 3113–3124, 1995. ISSN 09567151. doi: 10.1016/0956-7151(95)00011-J.

[144] D MacKay. Bayesian Interpolation. *Neural Computation*, 4(3):415–447, may 1992. ISSN 0899-7667. doi: 10.1162/neco.1992.4.3.415.

[145] D Mackay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, 2003. ISBN 978-0-521-64298-9.

[146] I Majid and P Bristowe. Dynamical simulations of structural multiplicity in grain boundaries. *Scripta Metallurgica*, 21(8):1153–1157, 1987. doi: 10.1016/0036-9748(87)90268-7.

[147] H McQueen. Formation and application of grain boundary serrations. *Canadian Metallurgical Quarterly*, 34(3):219–229, sep 1995. ISSN 00084433. doi: 10.1016/0008-4433(95)00001-E.

[148] N Metropolis, A Rosenbluth, M Rosenbluth, A Teller, and E Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6): 1087–1092, jun 1953. ISSN 0021-9606. doi: 10.1063/1.1699114.

[149] R Metzler and J Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, 339(1):1–77, dec 2000. ISSN 03701573. doi: 10.1016/S0370-1573(00)00070-3.

[150] R Metzler, J Jeon, A Cherstvy, and E Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single

particle tracking. *Phys. Chem. Chem. Phys.*, 16(44):24128–24164, 2014. ISSN 1463-9076. doi: 10.1039/C4CP03465A.

[151] M Minsky. Steps toward Artificial Intelligence. *Proceedings of the IRE*, 49(1):8–30, jan 1961. ISSN 0096-8390. doi: 10.1109/JRPROC.1961.287775.

[152] Y Mishin. Atomistic modeling of the γ and γ'-phases of the Ni–Al system. *Acta Materialia*, 52(6):1451–1467, apr 2004. ISSN 13596454. doi: 10.1016/j.actamat.2003.11.026.

[153] Y Mishin, M Mehl, and D Papaconstantopoulos. Phase stability in the Fe–Ni system: Investigation by first-principles calculations and atomistic simulations. *Acta Materialia*, 53(15):4029–4041, sep 2005. ISSN 13596454. doi: 10.1016/j.actamat.2005.05.001.

[154] Y Mishin, M Asta, and J Li. Atomistic modeling of interfaces and their impact on microstructure and properties. *Acta Materialia*, 58(4):1117–1151, feb 2010. ISSN 13596454. doi: 10.1016/j.actamat.2009.10.049.

[155] R Mitchell, M Preuss, S.Tin, and M Hardy. The influence of cooling rate from temperatures above the γ' solvus on morphology, mismatch and hardness in advanced polycrystalline nickel-base superalloys. *Materials Science and Engineering: A*, 473 (1-2):158–165, jan 2008. ISSN 09215093. doi: 10.1016/j.msea.2007.04.098.

[156] R Mitchell, H Li, and Z Huang. On the formation of serrated grain boundaries and fan type structures in an advanced polycrystalline nickel-base superalloy. *Journal of Materials Processing Technology*, 209(2):1011–1017, jan 2009. ISSN 09240136. doi: 10.1016/j.jmatprotec.2008.03.008.

[157] P Mitev, G Evangelakis, and E Kaxiras. Embedded atom method potentials employing a faithful density representation. *Modelling and Simulation in Materials Science and Engineering*, 14(4):721–731, jun 2006. ISSN 0965-0393. doi: 10.1088/0965-0393/14/4/013.

[158] N Mott. Slip at Grain Boundaries and Grain Growth in Metals. *Proceedings of the Physical Society*, 60(4):391–394, 1948. doi: 10.1088/0959-5309/60/4/309.

[159] E Nadaraya. On Estimating Regression. *Theory of Probability & Its Applications*, 9 (1):141–142, jan 1964. ISSN 0040-585X. doi: 10.1137/1109020.

[160] NASA Glenn Research Center. Gas Turbine Schematic. https://www.grc.nasa.gov/www/K-12/airplane/turbdraw.html.

[161] H Ogawa. GBstudio: A Builder Software on Periodic Models of CSL Boundaries for Molecular Simulation. *MATERIALS TRANSACTIONS*, 47(11):2706–2710, 2006. ISSN 1345-9678. doi: 10.2320/matertrans.47.2706.

[162] D Olmsted, S Foiles, and E Holm. Grain boundary interface roughening transition and its effect on grain boundary mobility for non-faceting boundaries. *Scripta Materialia*, 57(12):1161–1164, dec 2007. ISSN 13596462. doi: 10.1016/j.scriptamat.2007.07.045.

[163] D Olmsted, S Foiles, and E Holm. Survey of computed grain boundary properties in face-centered cubic metals – I: Grain boundary energy. *Acta Materialia*, 57(13):3694–3703, 2009. ISSN 1359-6454. doi: 10.1016/j.actamat.2009.04.007.

[164] D Olmsted, E Holm, and S Foiles. Survey of computed grain boundary properties in face-centered cubic metals – II: Grain boundary mobility. *Acta Materialia*, 57(13):3704–3713, 2009. ISSN 1359-6454. doi: 10.1016/j.actamat.2009.04.015.

[165] A Otero-de-la Roza, E Johnson, and V Luaña. Critic2: A program for real-space analysis of quantum chemical interactions in solids. *Computer Physics Communications*, 185(3):1007–1018, mar 2014. ISSN 00104655. doi: 10.1016/j.cpc.2013.10.026.

[166] D Papageorgiou, I Demetropoulos, and I Lagaris. Merlin-3.0 A multidimensional optimization environment. *Computer Physics Communications*, 109(2-3):227–249, apr 1998. ISSN 00104655. doi: 10.1016/S0010-4655(98)00005-8.

[167] C Pareige-Schmuck, F Soisson, and D Blavette. Ordering and phase separation in low supersaturated Ni–Cr–Al alloys: 3D atom probe and Monte Carlo simulation. *Materials Science and Engineering: A*, 250(1):99–103, jul 1998. ISSN 09215093. doi: 10.1016/S0921-5093(98)00543-7.

[168] F Pedregosa and G Varoquaux. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12. Oct.:2825–2830, 2011. ISSN 1532-4435.

[169] A Picasso, A Somoza, and A Tolley. Nucleation, growth and coarsening of γ'-precipitates in a Ni–Cr–Al-based commercial superalloy during artificial aging. *Journal of Alloys and Compounds*, 479(1-2):129–133, jun 2009. ISSN 09258388. doi: 10.1016/j.jallcom.2008.12.068.

[170] C Pickard and R Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5):053201, feb 2011. ISSN 0953-8984. doi: 10.1088/0953-8984/23/5/053201.

[171] E Pickering, H Mathur, A Bhowmik, O Messé, J Barnard, M Hardy, R Krakow, K Loehnert, H Stone, and C Rae. Grain-boundary precipitation in Allvac 718Plus. *Acta Materialia*, 60(6-7):2757–2769, apr 2012. ISSN 13596454. doi: 10.1016/j.actamat.2012.01.042.

[172] G Pizzi, A Cepellotti, R Sabatini, N Marzari, and B Kozinsky. AiiDA: automated interactive infrastructure and database for computational science. *Computational Materials Science*, 111:218–230, jan 2016. ISSN 09270256. doi: 10.1016/j.commatsci.2015.09.013.

[173] S Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*, 117(1):1–19, mar 1995. ISSN 00219991. doi: 10.1006/jcph.1995.1039.

[174] R Poduri and L Chen. Computer simulation of atomic ordering and compositional clustering in the pseudobinary Ni3Al–Ni3V system. *Acta Materialia*, 46(5):1719–1729, mar 1998. ISSN 13596454. doi: 10.1016/S1359-6454(97)00335-2.

[175] T Pollock and A Argon. Creep resistance of CMSX-3 nickel base superalloy single crystals. *Acta Metallurgica et Materialia*, 40(1):1–30, jan 1992. ISSN 09567151. doi: 10.1016/0956-7151(92)90195-K.

[176] C Qiu and P Andrews. On the formation of irregular-shaped gamma prime and serrated grain boundaries in a nickel-based superalloy during continuous cooling. *Materials Characterization*, 76:28–34, feb 2013. ISSN 10445803. doi: 10.1016/j.matchar.2012.11.012.

[177] D Quigley and P Rodger. Metadynamics simulations of ice nucleation and growth. *The Journal of Chemical Physics*, 128(15):154518, apr 2008. ISSN 0021-9606. doi: 10.1063/1.2888999.

[178] D Quigley and P Rodger. A metadynamics-based approach to sampling crystallisation events. *Molecular Simulation*, 35(7):613–623, jun 2009. ISSN 0892-7022. doi: 10.1080/08927020802647280.

[179] J Quinonero-Candela and C Rasmussen. A Unifying View of Sparse Approximate Gaussian Process Regression. *Journal of Machine Learning Research*, 6(6):1939–1959, 2005.

[180] E Rabkin. Gradient and coherency strain energies as driving forces for DIGM. *Scripta Metallurgica et Materialia*, 30(11):1443–1448, jun 1994. ISSN 0956716X. doi: 10.1016/0956-716X(94)90243-7.

[181] E Rabkin, L Shvindlerman, and W Gust. Theory of grain boundary motion during high-temperature DIGM. *Interface Science*, 1(2):133–137, 1993. ISSN 0927-7056. doi: 10.1007/BF00203602.

[182] C Race, J von Pezold, and J Neugebauer. Role of the mesoscale in migration kinetics of flat grain boundaries. *Physical Review B*, 89(21):214110, jun 2014. ISSN 1098-0121. doi: 10.1103/PhysRevB.89.214110.

[183] K Rajan. Materials Informatics: The Materials "Gene" and Big Data. *Annual Review of Materials Research*, 45(1):153–169, 2015. ISSN 1531-7331. doi: 10.1146/annurev-matsci-070214-021132.

[184] C Rasmussen and C Williams. *Gaussian Processes for Machine Learning*. 2006.

[185] I Rechenberg. *Evolutionsstrategie*. Friedrich Frommann Verlag, Stuttgart-Bad Cannstatt, 1973.

[186] R Reed. *The Superalloys*. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, 1st edition, 2006. ISBN 0-511-24546-7.

[187] P Rein ten Wolde, M Ruiz-Montero, and D Frenkel. Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. *The Journal of Chemical Physics*, 104(24):9932–9947, jun 1996. ISSN 0021-9606. doi: 10.1063/1.471721.

[188] R Ricks, A Porter, and R Ecob. The growth of γ' precipitates in Nickel-base superalloys. *Acta Metall*, 31:43–53, 1983. doi: 10.1016/0001-6160(83)90062-7.

[189] L Rios and N Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3): 1247–1293, jul 2013. ISSN 0925-5001. doi: 10.1007/s10898-012-9951-y.

[190] J Rogal, S Divinski, M Finnis, A Glensk, J Neugebauer, J Perepezko, S Schuwalow, M Sluiter, and B Sundman. Perspectives on point defect thermodynamics. *physica status solidi (b)*, 251(1):97–129, jan 2014. ISSN 03701972. doi: 10.1002/pssb.201350155.

[191] G Rohrer, D Saylor, B El Dasher, B Adams, A Rollet, and P Wynblatt. The distribution of internal interfaces in polycrystals. *Zeitschrift für Metallkunde*, 95(4):197–214, 2004. ISSN 0044-3093. doi: 10.3139/146.017934.

[192] T Rojhirunsakool, A Singh, S Nag, J Hwang, J Tiley, and R Banerjee. Temporal evolution of non-equilibrium γ' precipitates in a rapidly quenched nickel base superalloy. *Intermetallics*, 54:218–224, nov 2014. ISSN 09669795. doi: 10.1016/j.intermet.2014.06.011.

[193] S Roweis and Z Ghahramani. A Unifying Review of Linear Gaussian Models. *Neural Computation*, 11(2):305–345, feb 1999. ISSN 0899-7667. doi: 10.1162/089976699300016674.

[194] A Sadeghi, S Ghasemi, B Schaefer, S Mohr, M Lill, and S Goedecker. Metrics for measuring distances in configuration spaces. *The Journal of Chemical Physics*, 139 (18):184118, nov 2013. ISSN 0021-9606. doi: 10.1063/1.4828704.

[195] B Sadigh, P Erhart, A Stukowski, A Caro, E Martinez, and L Zepeda-Ruiz. Scalable parallel Monte Carlo algorithm for atomistic simulations of precipitation in alloys. *Physical Review B*, 85:184203, 2012. ISSN 10980121. doi: 10.1103/PhysRevB.85.184203.

[196] J Salvatier, T Wiecki, and C Fonnesbeck. Probabilistic Programming in Python using PyMC. *PeerJ Computer Science*, 2:e55, 2016. ISSN 2376-5992. doi: 10.7717/peerj-cs.55.

[197] M Sangid, H Sehitoglu, H Maier, and T Niendorf. Grain boundary characterization and energetics of superalloys. *Materials Science and Engineering: A*, 527(26):7115–7125, oct 2010. ISSN 09215093. doi: 10.1016/j.msea.2010.07.062.

[198] T Schablitzki, J Rogal, and R Drautz. Topological fingerprints for intermetallic compounds for the automated classification of atomistic simulation data. *Modelling and Simulation in Materials Science and Engineering*, 21(7):075008, oct 2013. ISSN 0965-0393. doi: 10.1088/0965-0393/21/7/075008.

[199] J Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, jan 2015. ISSN 08936080. doi: 10.1016/j.neunet.2014.09.003.

[200] E Schmidt and P Bristowe. Identifying early stage precipitation in large-scale atomistic simulations of superalloys. *Modelling and Simulation in Materials Science and Engineering*, 25(3):035005, apr 2017. ISSN 0965-0393. doi: 10.1088/1361-651X/aa5c53.

[201] E Schmidt, A Fowler, J Elliott, and P Bristowe. Learning Models for Electron Densities with Bayesian Regression. *Computational Materials Science*, 149(March):250–258, 2017. ISSN 0927-0256. doi: 10.1016/j.commatsci.2018.03.029.

[202] C Schmuck, P Caron, A Hauet, and D Blavette. Ordering and precipitation of γ' phase in low supersaturated Ni-Cr-Al model alloy: An atomic scale investigation. *Philosophical Magazine A*, 76(3):527–542, sep 1997. ISSN 0141-8610. doi: 10.1080/01418619708214021.

[203] B Schönfelder, D Wolf, S Phillpot, and M Furtkamp. Molecular-Dynamics Method for the Simulation of Grain-Boundary Migration. *Interface Science*, 5(4):245–262, 1997. ISSN 09277056. doi: 10.1023/A:1008663804495.

[204] B Schönfelder, G Gottstein, and L Shvindlerman. Comparative study of grain-boundary migration and grain-boundary self-diffusion of [001] twist-grain boundaries in copper by atomistic simulations. *Acta Materialia*, 53(6):1597–1609, apr 2005. ISSN 13596454. doi: 10.1016/j.actamat.2004.12.010.

[205] K Schütt, F Arbabzadah, S Chmiela, K Müller, and A Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature Communications*, 8: 13890, jan 2017. ISSN 2041-1723. doi: 10.1038/ncomms13890.

[206] D Schwanzer, D Coslovich, and G Kahl. Two-dimensional systems with competing interactions: dynamic properties of single particles and of clusters. *Journal of Physics: Condensed Matter*, 28(41):414015, oct 2016. ISSN 0953-8984. doi: 10.1088/0953-8984/28/41/414015.

[207] H Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie. Mit einer vergleichenden Einführung in die Hill-Climbing- und Zufallsstrategien.* Birkhaeuser Verlag, Basel-Stuttgart, 1977.

[208] B Settles. Active Learning Literature Survey. Technical report, University of Wisconsin-Madison, 2010.

[209] C Shallue and A Vanderburg. Identifying Exoplanets with Deep Learning: A Five Planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *The Astronomical Journal*, 155(2):94, 2018. ISSN 1538-3881. doi: 10.3847/1538-3881/aa9e09.

[210] D Silver, J Schrittwieser, K Simonyan, I Antonoglou, A Huang, A Guez, T Hubert, L Baker, M Lai, A Bolton, Y Chen, T Lillicrap, F Hui, and L Sifre. Mastering the game of Go without human knowledge. *Nature Publishing Group*, 550(7676):354–359, 2017. ISSN 0028-0836. doi: 10.1038/nature24270.

[211] C Sims, N Stoloff, and W Hagel. *Superalloys II*. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1st edition, 1987. ISBN 0-471-01147-9.

[212] A Singh, S Nag, S Chattopadhyay, Y Ren, J Tiley, G Viswanathan, H Fraser, and R Banerjee. Mechanisms related to different generations of γ' precipitation during continuous cooling of a nickel base superalloy. *Acta Materialia*, 61(1):280–293, jan 2013. ISSN 13596454. doi: 10.1016/j.actamat.2012.09.058.

[213] D Smith, V Vitek, and R Pond. Computer simulation of symmetrical high angle boundaries in aluminium. *Acta Metallurgica*, 25(5):475–483, may 1977. ISSN 00016160. doi: 10.1016/0001-6160(77)90187-0.

[214] W Soffa and D Laughlin. Recent experimental studies of continuous transformations in alloys: An overview. *Proceedings of an International Conference on Solid-Solid Phase Transformations, AIME*, pages 159–183, 1982.

[215] W Soffa and D Laughlin. Decomposition and ordering processes involving thermodynamically first-order order → disorder transformations. *Acta Metallurgica*, 37(11): 3019–3028, nov 1989. ISSN 00016160. doi: 10.1016/0001-6160(89)90338-6.

[216] W Soffa, W Pueschl, and W Pfeiler. Homo or hetero? How vacancy motion brings discreteness into phase transformations. *Intermetallics*, 11:161–167, 2003.

[217] W Soffa, D Laughlin, and N Singh. Interplay of ordering and spinodal decomposition in the formation of ordered precipitates in binary fcc alloys: Role of second nearest-neighbor interactions. *Philosophical Magazine*, 90(1-4):287–304, jan 2010. ISSN 1478-6435. doi: 10.1080/14786430903127520.

[218] J Soler, E Artacho, J Gale, A García, J Junquera, P Ordejón, and D Sánchez-Portal. The SIESTA method for ab initio order- N materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745–2779, mar 2002. ISSN 0953-8984. doi: 10.1088/0953-8984/14/11/302.

[219] P Steinhardt, D Nelson, and M Ronchetti. Bond-orientational order in liquids and glasses. *Physical Review B*, 28(2):784–805, jul 1983. ISSN 0163-1829. doi: 10.1103/PhysRevB.28.784.

[220] F Stillinger and T Weber. Computer simulation of local order in condensed phases of silicon. *Physical Review B*, 31(8):5262–5271, apr 1985. ISSN 0163-1829. doi: 10.1103/PhysRevB.31.5262.

[221] M Stott and E Zaremba. Quasiatoms: An approach to atoms in nonuniform electronic systems. *Physical Review B*, 22(4):1564–1583, aug 1980. ISSN 0163-1829. doi: 10.1103/PhysRevB.22.1564.

[222] A Stukowski. Visualization and analysis of atomistic simulation data with OVITO - the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering*, 18(1):015012, 2010. ISSN 0965-0393. doi: 10.1088/0965-0393/18/1/015012.

[223] A Stukowski. Structure identification methods for atomistic simulations of crystalline materials. *Modelling and Simulation in Materials Science and Engineering*, 20(4): 045021, jun 2012. ISSN 0965-0393. doi: 10.1088/0965-0393/20/4/045021.

[224] C Sudbrack, K Yoon, Z Mao, R Noebe, D Isheim, and D Seidman. Temporal evolution of nanostructures in a model nickel-base superalloy: experiments and simulations. *The Mike Meshii Symposium*, pages 43–50, 2003.

[225] M Suganuma, S Shirakawa, and T Nagao. A Genetic Programming Approach to Designing Convolutional Neural Network Architectures. In *GECCO Proceedings*, pages 1–8, Berlin, 2017. ISBN 9781450335492. doi: 10.1145/nnnnnnn.nnnnnnn.

[226] C Sun, A Shrivastava, S Singh, and A Gupta. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-Octob:843–852, 2017. ISSN 15505499. doi: 10.1109/ICCV.2017.97.

[227] A Sutton and R Balluffi. Interfaces in Crystalline Materials. In R J Brook, A Cheetham, A Heuer, Peter Hirsch, T J Marks, J Silcox, D A Smith, M V Tirrell, and V Vitek, editors, *Interfaces in Crystalline Materials*, pages 514–518. Clarendon Press, Oxford, 1 edition, 1995. ISBN 0198513852.

[228] X Tan, D Mangelinck, C Perrin-Pellegrino, L Rougier, C Gandin, A Jacot, D Ponsen, and V Jaquet. Spinodal Decomposition Mechanism of γ' Precipitation in a Single Crystal Ni-Based Superalloy. *Metallurgical and Materials Transactions A*, 45(11): 4725–4730, oct 2014. ISSN 1073-5623. doi: 10.1007/s11661-014-2506-8.

[229] X Tan, D Mangelinck, C Perrin-Pellegrino, L Rougier, C Gandin, A Jacot, D Ponsen, and V Jaquet. Atom probe tomography of secondary γ' precipitation in a single crystal Ni-based superalloy after isothermal aging at 1100°C. *Journal of Alloys and Compounds*, 611:389–394, oct 2014. ISSN 09258388. doi: 10.1016/j.jallcom.2014.05.132.

[230] B Tang, L Jiang, R Hu, and Q Li. Correlation between grain boundary misorientation and M23C6 precipitation behaviors in a wrought Ni-based superalloy. *Materials Characterization*, 78:144–150, apr 2013. ISSN 10445803. doi: 10.1016/j.matchar.2013.02.006.

[231] M Terner, H Hong, J Lee, and B Choi. On the role of alloying elements in the formation of serrated grain boundaries in Ni-based alloys. *International Journal of Materials Research*, 107(3):229–238, mar 2016. ISSN 1862-5282. doi: 10.3139/146.111332.

[232] J Tersoff. New empirical approach for the structure and energy of covalent systems. *Physical Review B*, 37(12):6991–7000, apr 1988. ISSN 0163-1829. doi: 10.1103/PhysRevB.37.6991.

[233] B Thijsse. Relationship between the modified embedded-atom method and Stillinger-Weber potentials in calculating the structure of silicon. *Physical Review B*, 65(19): 195207, may 2002. ISSN 0163-1829. doi: 10.1103/PhysRevB.65.195207.

[234] L Thomas. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(05):542, jan 1927. ISSN 0305-0041. doi: 10.1017/S0305004100011683.

[235] M Tipping. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of machine learning research*, 1(Jun):211–244, 2001.

[236] M Tipping and A Faul. Fast Marginal Likelihood Maximisation for Sparse Bayesian Models. *Proc. 9th Int. Workshop Artificial Intelligence and Statistics*, 2003.

[237] S Tong and D Koller. Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*, pages 45–66, 2001. ISSN 15324435. doi: 10.1162/153244302760185243.

[238] G Torrie and J Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2): 187–199, feb 1977. ISSN 00219991. doi: 10.1016/0021-9991(77)90121-8.

[239] D Tran, A Kucukelbir, A Dieng, M Rudolph, D Liang, and D Blei. Edward: A library for probabilistic modeling, inference, and criticism. *Arxiv*, 2016.

[240] Z Trautt and M Upmanyu. Direct two-dimensional calculations of grain boundary stiffness. *Scripta Materialia*, 52(11):1175–1179, jun 2005. ISSN 13596462. doi: 10.1016/j.scriptamat.2004.12.033.

[241] Z Trautt, M Upmanyu, and A Karma. Interface Mobility from Interface Random Walk. *Science*, 314(5799):632–635, oct 2006. ISSN 0036-8075. doi: 10.1126/science.1131988.

[242] V Tresp. A Bayesian Committee Machine. *Neural Computation*, 12(11):2719–2741, nov 2000. ISSN 0899-7667. doi: 10.1162/089976600300014908.

[243] H Tsuzuki, P Branicio, and J Rino. Structural characterization of deformed crystals by analysis of common atomic neighborhood. *Computer Physics Communications*, 177(6):518–523, sep 2007. ISSN 00104655. doi: 10.1016/j.cpc.2007.05.018.

[244] F Ulomek and V Mohles. Separating grain boundary migration mechanisms in molecular dynamics simulations. *Acta Materialia*, 103:424–432, jan 2016. ISSN 13596454. doi: 10.1016/j.actamat.2015.10.021.

[245] F Ulomek, C O'Brien, S Foiles, and V Mohles. Energy conserving orientational force for determining grain boundary mobility. *Modelling and Simulation in Materials Science and Engineering*, 23(2):025007, mar 2015. ISSN 0965-0393. doi: 10.1088/0965-0393/23/2/025007.

[246] R Unocic, G Viswanathan, P Sarosi, S Karthikeyan, J Li, and M Mills. Mechanisms of creep deformation in polycrystalline Ni-base disk superalloys. *Materials Science and Engineering: A*, 483-484:25–32, jun 2008. ISSN 09215093. doi: 10.1016/j.msea.2006.08.148.

[247] N van Eck and L Waltman. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523–538, aug 2010. ISSN 0138-9130. doi: 10.1007/s11192-009-0146-3.

[248] G Viswanathan, R Banerjee, A Singh, S Nag, J Tiley, and H Fraser. Precipitation of ordered phases in metallic solid solutions: A synergistic clustering and ordering process. *Scripta Materialia*, 65(6):485–488, sep 2011. ISSN 13596462. doi: 10.1016/j.scriptamat.2011.06.002.

[249] G Voronoi. Nouvelles applications des parametres continus a la theorie des formes quadratiques. *J. Reine Angew. Math*, 134:97–178, 1908.

[250] A Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Chemical Engineering Journal*, 78(20):3908–3911, 1997. ISSN 13858947. doi: 10.1016/S1385-8947(99)00055-8.

[251] D Wales. Surveying a complex potential energy landscape: Overcoming broken ergodicity using basin-sampling. *Chemical Physics Letters*, 584:1–9, oct 2013. ISSN 00092614. doi: 10.1016/j.cplett.2013.07.066.

[252] D Wales and J Doye. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, jul 1997. ISSN 1089-5639. doi: 10.1021/jp970984n.

[253] L Wang and M Teter. Kinetic-energy functional of the electron density. *Physical Review B*, 45(23):196–220, 1992. doi: 10.1103/PhysRevB.45.13196.

[254] Y Wang and A Khachaturyan. Effect of antiphase domains on shape and spatial arrangement of coherent ordered intermetallics. *Scripta Metallurgica et Materialia*, 31(10):1425–1430, nov 1994. ISSN 0956716X. doi: 10.1016/0956-716X(94)90130-9.

[255] Y Wang, N Govind, and E Carter. Orbital-free kinetic-energy density functionals with a density-dependent kernel. *Physical Review B*, 60(24):350–358, 1999. doi: 10.1103/PhysRevB.60.16350.

[256] L Ward and C Wolverton. Atomistic calculations and materials informatics: A review. *Current Opinion in Solid State and Materials Science*, 21(3):167–176, jun 2017. ISSN 13590286. doi: 10.1016/j.cossms.2016.07.002.

[257] G Watson. Smooth regression analysis. *The Indian Journal of Statistics*, 26(4):359–372, 1964. ISSN 0581572X. doi: 10.2307/25049340.

[258] C Weizsäcker. Zur Theorie der Kernmassen. *Zeitschrift für Physik*, 96(7-8):431–458, jul 1935. ISSN 1434-6001. doi: 10.1007/BF01337700.

[259] D Welton, N D'Souza, J Kelleher, S Gardner, Z Dong, G West, and H Dong. Discontinuous Precipitation in Ni-Base Superalloys During Solution Heat Treatment. *Metallurgical and Materials Transactions A*, 46(9):4298–4315, sep 2015. ISSN 1073-5623. doi: 10.1007/s11661-015-3046-6.

[260] D Wierstra, T Schaul, J Peters, and J Schmidhuber. Natural Evolution Strategies. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, volume 15, pages 3381–3387. IEEE, jun 2008. ISBN 978-1-4244-1822-0. doi: 10.1109/CEC.2008.4631255.

[261] A Wilson. Formation and effect of topologically close-packed phases in nickel-base superalloys. *Materials Science and Technology*, 33(9):1108–1118, jun 2017. ISSN 0267-0836. doi: 10.1080/02670836.2016.1187335.

[262] S Winczewski, J Dziedzic, and J Rybicki. A highly-efficient technique for evaluating bond-orientational order parameters. *Computer Physics Communications*, 198:128–138, jan 2016. ISSN 00104655. doi: 10.1016/j.cpc.2015.09.009.

[263] A Wisniewski and J Beddoes. Influence of grain-boundary morphology on creep of a wrought Ni-base superalloy. *Materials Science and Engineering: A*, 510-511:266–272, jun 2009. ISSN 09215093. doi: 10.1016/j.msea.2008.04.130.

[264] D Wolpert and W Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, apr 1997. ISSN 1089778X. doi: 10.1109/4235.585893.

[265] K Wu, G Liu, B Hu, F Li, Y Zhang, Y Tao, and J Liu. Formation mechanism and coarsening behavior of fan-type structures in a new Ni–Cr–Co-based powder metallurgy superalloy. *Journal of Materials Science*, 47(11):4680–4688, jun 2012. ISSN 0022-2461. doi: 10.1007/s10853-012-6336-z.

[266] W Yang, G Liu, K Wu, and B Hu. Influence of sub-solvus solution heat treatment on γ' morphological instability in a new Ni–Cr–Co-based powder metallurgy superalloy. *Journal of Alloys and Compounds*, 582:515–521, jan 2014. ISSN 09258388. doi: 10.1016/j.jallcom.2013.07.045.

[267] Y Yang and G Webb. On Why Discretization Works for Naive-Bayes Classifiers. In T.D. Gedeon and L.C.C. Fung, editors, *AI 2003: Advances in Artificial Intelligence. AI 2003. Lecture Notes in Computer Science*, volume 2903, pages 440–452. Springer, Heidelberg, Berlin, 2003. ISBN 3540206469. doi: 10.1007/978-3-540-24581-0_37.

[268] A Yeh, K Lu, C Kuo, H Bor, and C Wei. Effect of serrated grain boundaries on the creep property of Inconel 718 superalloy. *Materials Science and Engineering: A*, 530:525–529, dec 2011. ISSN 09215093. doi: 10.1016/j.msea.2011.10.014.

[269] J Yoon, H Jeong, Y Yoo, and H Hong. Influence of initial microstructure on creep deformation behaviors and fracture characteristics of Haynes 230 superalloy at 900°C. *Materials Characterization*, 101:49–57, mar 2015. ISSN 10445803. doi: 10.1016/j.matchar.2015.01.002.

[270] V Zaburdaev, S Denisov, and J Klafter. Lévy walks. *Reviews of Modern Physics*, 87 (2):483–530, jun 2015. ISSN 0034-6861. doi: 10.1103/RevModPhys.87.483.

[271] H Zhang and D Srolovitz. Simulation and analysis of the migration mechanism of $\Sigma 5$ tilt grain boundaries in an fcc metal. *Acta Materialia*, 54(3):623–633, feb 2006. ISSN 13596454. doi: 10.1016/j.actamat.2005.10.001.

[272] H Zhang, D Srolovitz, J Douglas, and J Warren. Characterization of atomic motion governing grain boundary migration. *Physical Review B*, 74(11):115404, sep 2006. ISSN 1098-0121. doi: 10.1103/PhysRevB.74.115404.

[273] H Zhang, D Srolovitz, J Douglas, and J Warren. Atomic motion during the migration of general [001] tilt grain boundaries in Ni. *Acta Materialia*, 55(13):4527–4533, aug 2007. ISSN 13596454. doi: 10.1016/j.actamat.2007.04.015.

[274] L Zhang, C Lu, and K Tieu. A review on atomistic simulation of grain boundary behaviors in face-centered cubic metals. *Computational Materials Science*, 118:180–191, jun 2016. ISSN 09270256. doi: 10.1016/j.commatsci.2016.03.021.

[275] J Zhou and V Mohles. Towards realistic molecular dynamics simulations of grain boundary mobility. *Acta Materialia*, 59(15):5997–6006, sep 2011. ISSN 13596454. doi: 10.1016/j.actamat.2011.06.008.

[276] L Zhu, M Amsler, T Fuhrer, B Schaefer, S Faraji, S Rostami, S Ghasemi, A Sadeghi, M Grauzinyte, C Wolverton, and S Goedecker. A fingerprint based metric for measuring similarities of crystalline structures. *The Journal of Chemical Physics*, 144(3): 034203, jan 2016. ISSN 0021-9606. doi: 10.1063/1.4940026.

Appendix A

Machine Learning

The Relevance Vector Machine

Evaluation of $\mathscr{L}(\boldsymbol{\alpha},\beta)$

The exponent of $p(\mathbf{t}|\mathbf{X},\mathbf{w},\beta)\,p(\mathbf{w}|\boldsymbol{\alpha})$ can be re-arranged as

$$\begin{aligned} (\mathbf{t}-\boldsymbol{\Phi}\mathbf{w})^T\beta(\mathbf{t}-\boldsymbol{\Phi}\mathbf{w})+\mathbf{w}^T\boldsymbol{\Sigma}_R^{-1}\mathbf{w} &= \mathbf{w}^T\left(\beta\boldsymbol{\Phi}^T\boldsymbol{\Phi}+\boldsymbol{\Sigma}_R^{-1}\right)\mathbf{w}-2(\boldsymbol{\Phi}\mathbf{w})^T\beta\mathbf{t}+\mathbf{t}^T\beta\mathbf{t}\\ &= (\mathbf{w}-\mathbf{w}_{MP})^T\mathbf{A}^{-1}(\mathbf{w}-\mathbf{w}_{MP})\\ &\quad +\mathbf{t}^T\beta\mathbf{t}-\mathbf{w}_{MP}^T\mathbf{A}^{-1}\mathbf{w}_{MP} \end{aligned} \tag{A.1}$$

when integrating over $\mathbf{w}$ the first term is integrated out as we are left with

$$\begin{aligned} \mathbf{t}^T\beta\mathbf{t}-\mathbf{w}_{MP}^T\mathbf{A}^{-1}\mathbf{w}_{MP} &= \mathbf{t}^T\beta\mathbf{t}-\left(\beta\mathbf{A}\boldsymbol{\Phi}^T\mathbf{t}\right)^T\boldsymbol{A}^{-1}\left(\beta\mathbf{A}\boldsymbol{\Phi}^T\mathbf{t}\right)\\ &= \mathbf{t}^T\beta\mathbf{t}-\beta^2\mathbf{t}^T\boldsymbol{\Phi}\mathbf{A}^T\boldsymbol{\Phi}^T\mathbf{t}\\ &= \mathbf{t}^T\beta\mathbf{t}-\beta^2\mathbf{t}^T\boldsymbol{\Phi}\left(\boldsymbol{\Phi}^T\beta\boldsymbol{\Phi}+\boldsymbol{\Sigma}_R\right)^{-1}\boldsymbol{\Phi}^T\mathbf{t}\\ &= \mathbf{t}^T\left(\beta-\beta^2\boldsymbol{\Phi}\left(\boldsymbol{\Phi}^T\beta\boldsymbol{\Phi}+\boldsymbol{\Sigma}_R\right)^{-1}\boldsymbol{\Phi}^T\right)\mathbf{t}\\ &= \mathbf{t}^T\left(\beta^{-1}\boldsymbol{I}+\boldsymbol{\Phi}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{\Phi}^T\right)^{-1}\mathbf{t}. \end{aligned} \tag{A.2}$$

The step from $\mathbf{t}^T\left(\beta-\beta^2\boldsymbol{\Phi}\left(\boldsymbol{\Phi}^T\beta\boldsymbol{\Phi}+\Sigma_R\right)^{-1}\boldsymbol{\Phi}^T\right)\mathbf{t}$ to $\mathbf{t}^T\left(\beta^{-1}\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{\Phi}^T\right)^{-1}\mathbf{t}$ uses the Woodbury inversion identity[1].

---

[1] The Woodbury inversion identity: $(A+UCV)^{-1}=A^{-1}-A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}$

Algorithm by Tipping and Faul

Algorithm A.1 Relevance Vector Machine algorithm returning model weights given the design matrix $\mathbf{\Phi}$, target values $\mathbf{t}$ and initial hyperparameters $\boldsymbol{\alpha}$ and β.

```
procedure RVM($\mathbf{\Phi}$, $\boldsymbol{t}$, $\boldsymbol{\alpha}$, $\beta$)
    while not converged do
        $\boldsymbol{A} \leftarrow \left(\beta \mathbf{\Phi}^T \mathbf{\Phi} + \mathbf{\Sigma}_R^{-1}\right)$
        $\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{t},\boldsymbol{X}} \leftarrow \boldsymbol{A}^{-1} \Phi^T \boldsymbol{t}$
        $\mathbf{\Sigma}_{\boldsymbol{w}|\boldsymbol{t},\boldsymbol{X}} \leftarrow \boldsymbol{A}$
        for $i \in [1, M]$ do
            compute $Q_i$ and $S_i$
            if $Q_i^2 > S_i^2$ then
                update $\alpha_i$ with $\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{t},\boldsymbol{X}}$ and $\mathbf{\Sigma}_{\boldsymbol{w}|\boldsymbol{t},\boldsymbol{X}}$ using eq. 4.23
            else
                prune: $\alpha_i \leftarrow \infty$
            end if
            update $\beta$ with $\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{t},\boldsymbol{X}}$ using eq. 4.13
        end for
    end while
    return $\boldsymbol{\mu}_{\boldsymbol{w}|\boldsymbol{t},\boldsymbol{X}}$ as weights
end procedure
```

The Expectation-Maximization Algorithm

Derivation

The EM-algorithm is derived as

$$\begin{aligned} \ln p(X|\theta) &= \ln \textstyle\sum_z p(X, Z|\theta) \\ &= \ln \textstyle\sum_z q(Z)\, p(X|\theta) \end{aligned} \tag{A.3}$$

$\ln p(X|\theta) \sum_z q(Z) = \ln p(X|\theta)$ for any legitimate distribution $q(Z)$ without any approximation.

Using the Jensen inequality[2] we find

$$\begin{aligned}\ln \textstyle\sum_z q(Z)\, p(X|\theta) &\geq \textstyle\sum_z q(Z) \ln p(X|\theta) \\ &= \textstyle\sum_z q(Z) \ln \frac{p(X,Z|\theta)}{p(Z|X,\theta)} \\ &= \textstyle\sum_z q(Z) \ln \frac{p(X,Z|\theta) q(Z)}{p(Z|X,\theta) q(Z)}\end{aligned} \tag{A.4}$$

$$\ln p(X|\theta) = \underbrace{\sum_z q(Z) \ln \frac{p(X,Z|\theta)}{q(Z)}}_{\mathscr{L}(q,\theta)} - \underbrace{\sum_z q(Z) \ln \frac{p(Z|X,\theta)}{q(Z)}}_{KL(q||p)} \tag{A.5}$$

Having derived the central equation for the EM-algorithm, eq. A.5, we start the EM-algorithm by first guessing initial model parameters θ_{old}. Then we perform the *E-step* such that $q \to p(Z|X,\theta_{\text{old}})$. This corresponds to $\text{argmax}_q \mathscr{L}(q,\theta_{\text{old}})$ which leads to $\mathscr{L}(q,\theta_{\text{old}}) = \ln p(X|\theta_{\text{old}})$ because $KL(q||p)$ goes to zero. Alternatively we could use $\text{argmin}_q KL(q||p)$ which is always greater equal zero. After we have obtained q and fixed it we maximize (*M-step*) $\mathscr{L}(q,\theta)$ with respect to θ. Due to the fixed q we find that $\mathscr{L}$ increases as well as $KL(q||p)$, leading to an overall maximization $\ln p(X|\theta)$. This can be seen substituting $q(Z) = p(Z|X,\theta_{\text{old}})$ in eq. A.5. One finds

$$\ln p(X|\theta) = \sum_z p(Z|X,\theta_{\text{old}}) \ln p(X,Z|\theta) - \underbrace{\sum_z p(Z|X,\theta_{\text{old}}) \ln p(Z|X,\theta_{\text{old}})}_{-H(q)}, \tag{A.6}$$

where only $p(X,Z|\theta)$ changes with θ to find θ_{new}. The EM-algorithm thus converges because every iteration repeating the E and M-steps always increases $\ln p(X|\theta)$.

Conditionals of GMMs

The marginal and conditional can be calculated analytically. Writing x,y explicitly instead of $\mathbf{x}$ we have

$$p(x,y|\theta) = \sum_k \pi_k \mathscr{N}\left(\left[\begin{array}{c} x \\ y \end{array} \right] \middle| \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \right). \tag{A.7}$$

First we want to consider each normal distribution individually to find $p(x|y,\theta,k)$. Once we have done that we will determine $p(k|y)$. Using completion of the square and inversion of

---

[2] Jensen's inequality for ln is $\ln \mathbb{E}[X] \geq \mathbb{E}[\ln(X)]$ since ln is convex, instead of $\ln \mathbb{E}[X] \leq \mathbb{E}[\ln(X)]$ as is true for concave.

partitioned matrices[3] $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Lambda}_k = \boldsymbol{\Sigma}_k^{-1}$ we find

$$p(x|y,k) = \mathcal{N}\left(x \middle| \mu_{k,x|y}, \Lambda_{k,xx}^{-1}\right) \tag{A.8}$$

with $\mu_{k,x|y} = \mu_{k,x} - \Lambda_{k,xx}^{-1}\Lambda_{k,xy}\left(y - \mu_{k,y}\right)$. Using Bayes' theorem we find $p(k|y)$ as

$$p(z|y) = \frac{p(y|z)\,p(z)}{p(y)} = \frac{\mathcal{N}\left(y\,|\,\mu_{y,k}, \Sigma_{yy,k}\right)\pi_k}{\sum_j \mathcal{N}\left(y\,|\,\mu_{y,j}, \Sigma_{yy,j}\right)\pi_j}. \tag{A.9}$$

Combining the obtained conditionals for individual normal distributions and their conditional weights for the GMM we can fill in

$$p(x|y) = \sum_k \pi_{k,x|y}\mathcal{N}\left(x \middle| \mu_{x|y,k}, \Lambda_{xx,k}^{-1}\right). \tag{A.10}$$

For an individual normal distribution we have $p(x) = \mathcal{N}(x\,|\,\mu_x, \Sigma_{xx})$ and hence for the complete GMM $p(x) = \sum_k \pi_k \mathcal{N}\left(x\,|\,\mu_{x,k}, \Sigma_{xx,k}\right)$

The Algorithm

---

[3] Inversion of a partitioned matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1}CMBD^{-1} \end{pmatrix}$ with the *Schur-complement* $M \equiv \left(A - BD^{-1}C\right)^{-1}$.

Algorithm A.2 Expectation Maximization algorithm for Gaussian Mixture Models with M Gaussians.

```
procedure EM-GMM($\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \pi_i\}_{i\in[1,M]}$, $\{\boldsymbol{x}_n\}_{n\in[1,N]}$)
    while not converged do
        for $n \in [1, N]$ do                                              ▷ E-step
            for $i \in [1, M]$ do
                $\gamma(z_{ni}) \leftarrow$ eq. 4.30 with $\boldsymbol{\mu}_i$ & $\boldsymbol{\Sigma}_i$
            end for
        end for
        for $i \in [1, M]$ do                                              ▷ M-step
            $\boldsymbol{\mu}_i \leftarrow$ eq. 4.31 with $\{\gamma(z_{ni})\}_{n\in[1,N]}$
            $\boldsymbol{\Sigma}_i \leftarrow$ eq. 4.32 with $\{\gamma(z_{ni})\}_{n\in[1,N]}$ and $\boldsymbol{\mu}_i$
            $\pi_i \leftarrow$ eq. 4.33 with $\{\gamma(z_{ni})\}_{n\in[1,N]}$
        end for
    end while
    return $\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \pi_i\}_{i\in[1,M]}$
end procedure
```

Appendix B

Grain Boundary Models

All coincidence site lattice (CSL) grain boundary (GB) models used in this thesis are shown for fcc Ni in their smallest form constructed with *GBstudio* Ogawa [161]. The smallest possible form is one CSL cell in each x and y and two cells in z for tilt GBs. Twist GBs are shown for 2 cells in each direction.

The cell directions are indicated for each GB type with an coordinate system to the left with black horizontal and vertical arrows representing normal and parallel directions of the atomistic model respectively. The red arrows in the representation of the coordinate system is a projection of that vector onto the two black vectors in the paper plane and thus indicates the orientation of the rotation axis of the GB type, as provided by GBstudio, relative to the orientations indicated by the black arrows.

The atoms of the atomistic model represented as a sequence of colored symbols with the x- or y-axis normal to the paper plane, depending on the GB type. The position of the atoms on the axis normal to the paper plane is indicated by color, size and symbol. The atoms in the layer within the paper plane (0 Å) are indicated as deep blue and those in the layer furthest away from the paper plane are indicated as yellow, as illustrated by the color bar. The size of the symbols increases also from within the paper plane on wards such that the atoms closest to the reader are represented by the largest symbols. The symbols were chosen to illustrate the alteration of the layers and repeatedly iterate over circle, square,pentagon and octagon, in that order starting from the layer in the paper plane. Red lines in the 2d representation of the atomistic model indicate the GB planes themselves. Each atomic model has two GB planes due to periodic boundary conditions.

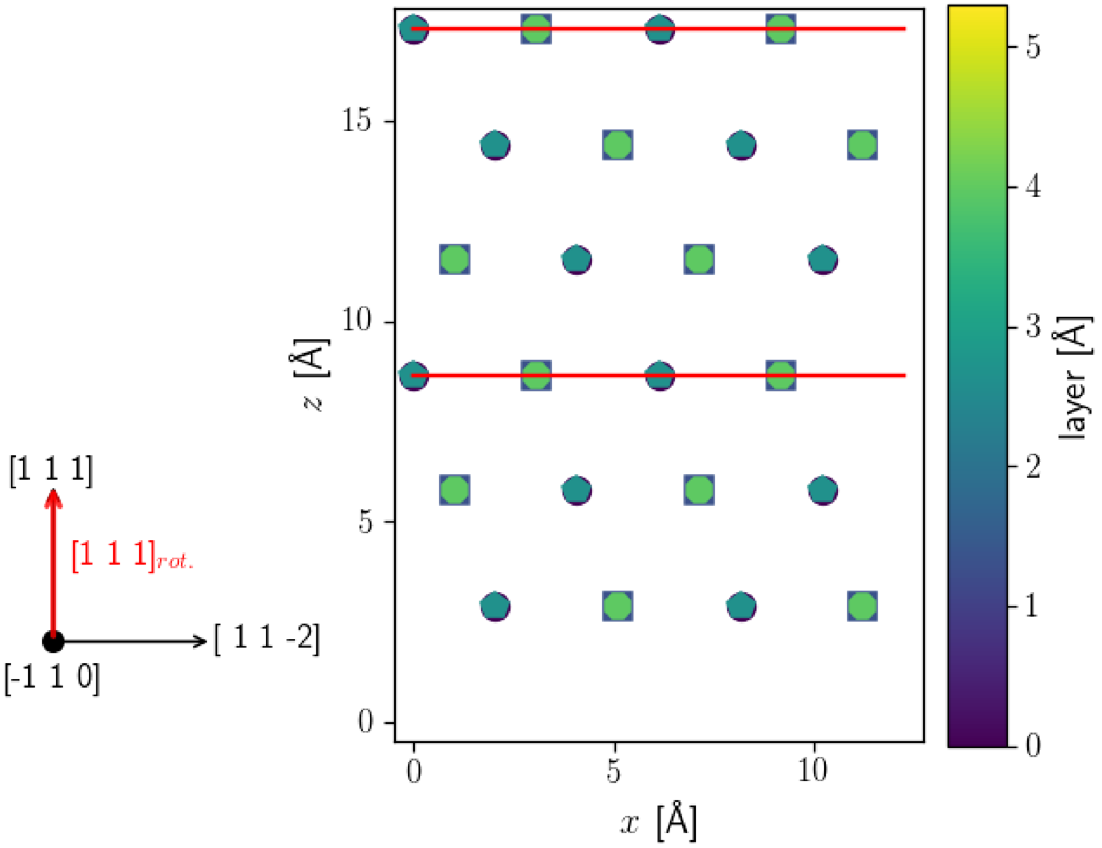


Figure B.1: The Σ3 tilt CSL GB with two layers along $[-1\,1\,0]$.

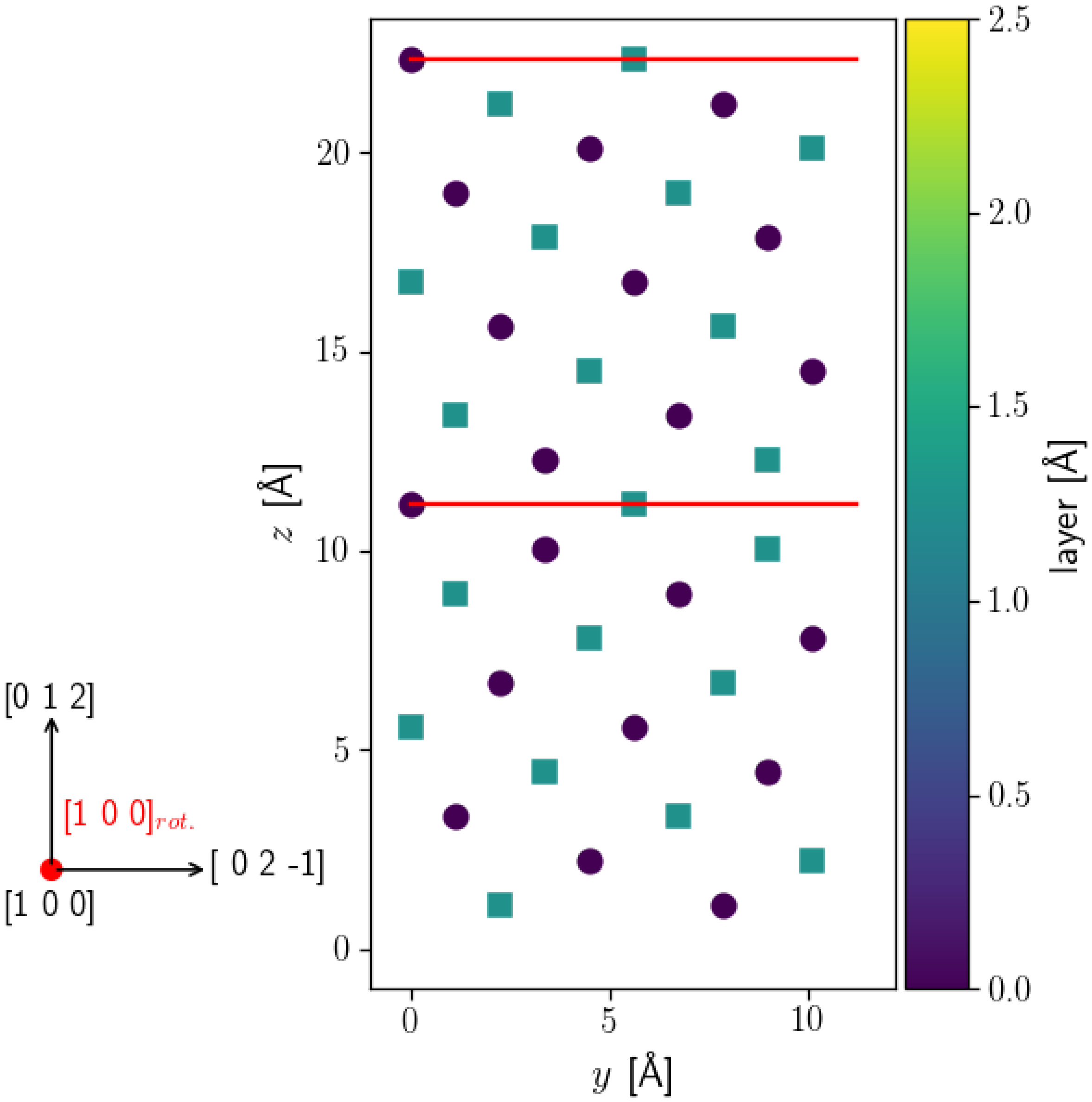


Figure B.2: The Σ5 tilt CSL GB with two layers along $[1\,0\,0]$.

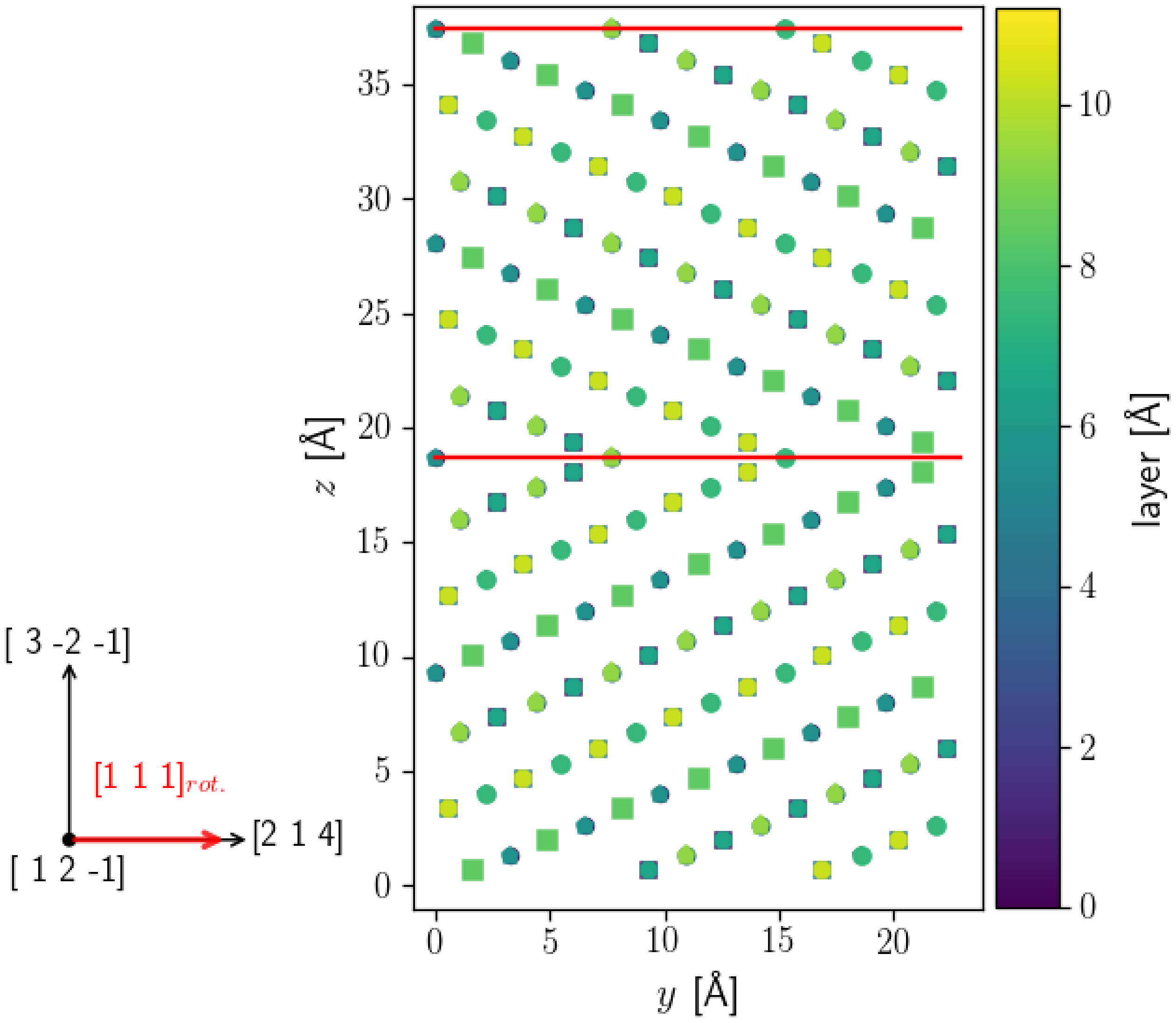


Figure B.3: The Σ7 tilt CSL GB with 12 layers along $[1\,2\,-1]$.

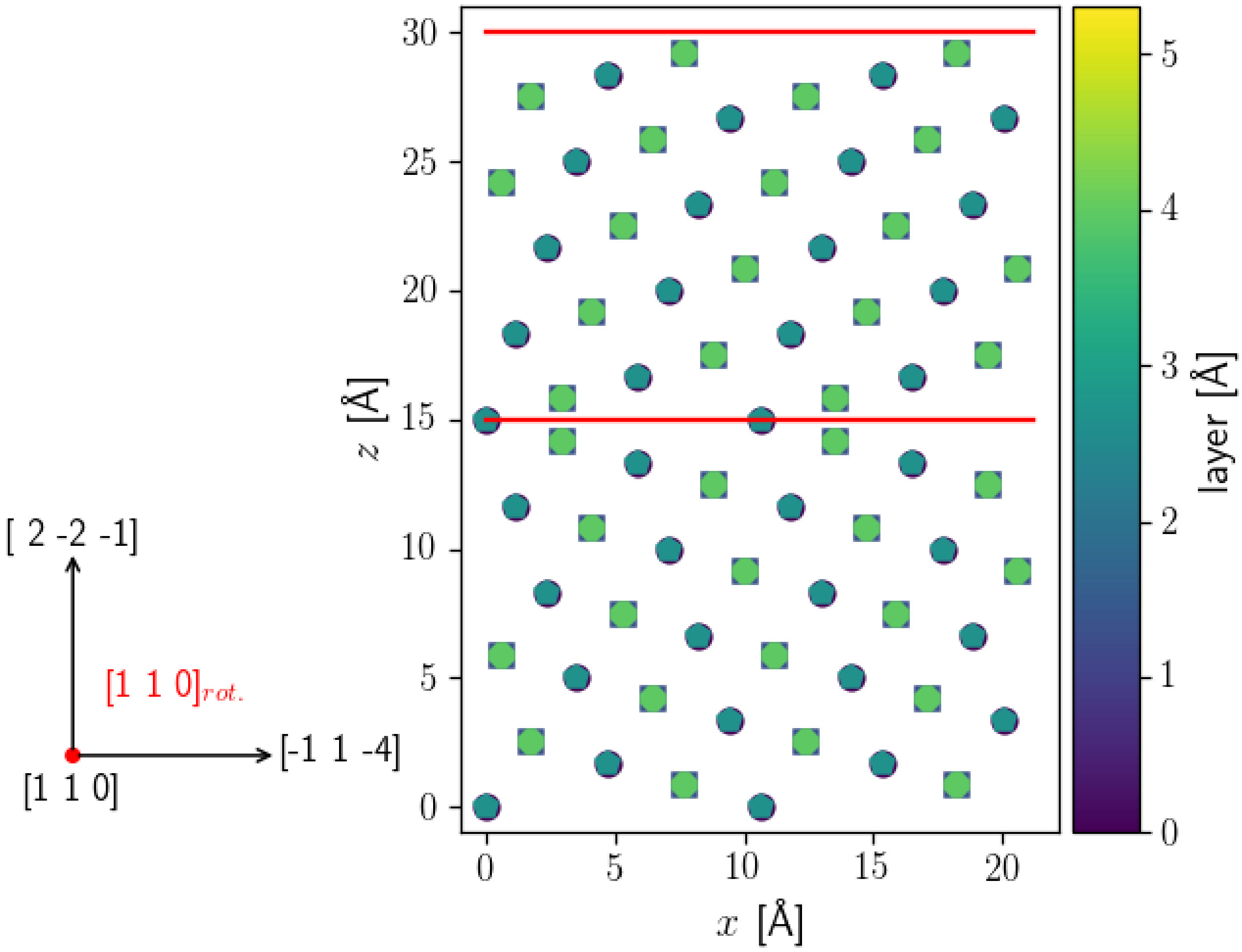


Figure B.4: The Σ9 tilt CSL GB with 4 layers along $[1\,1\,0]$.

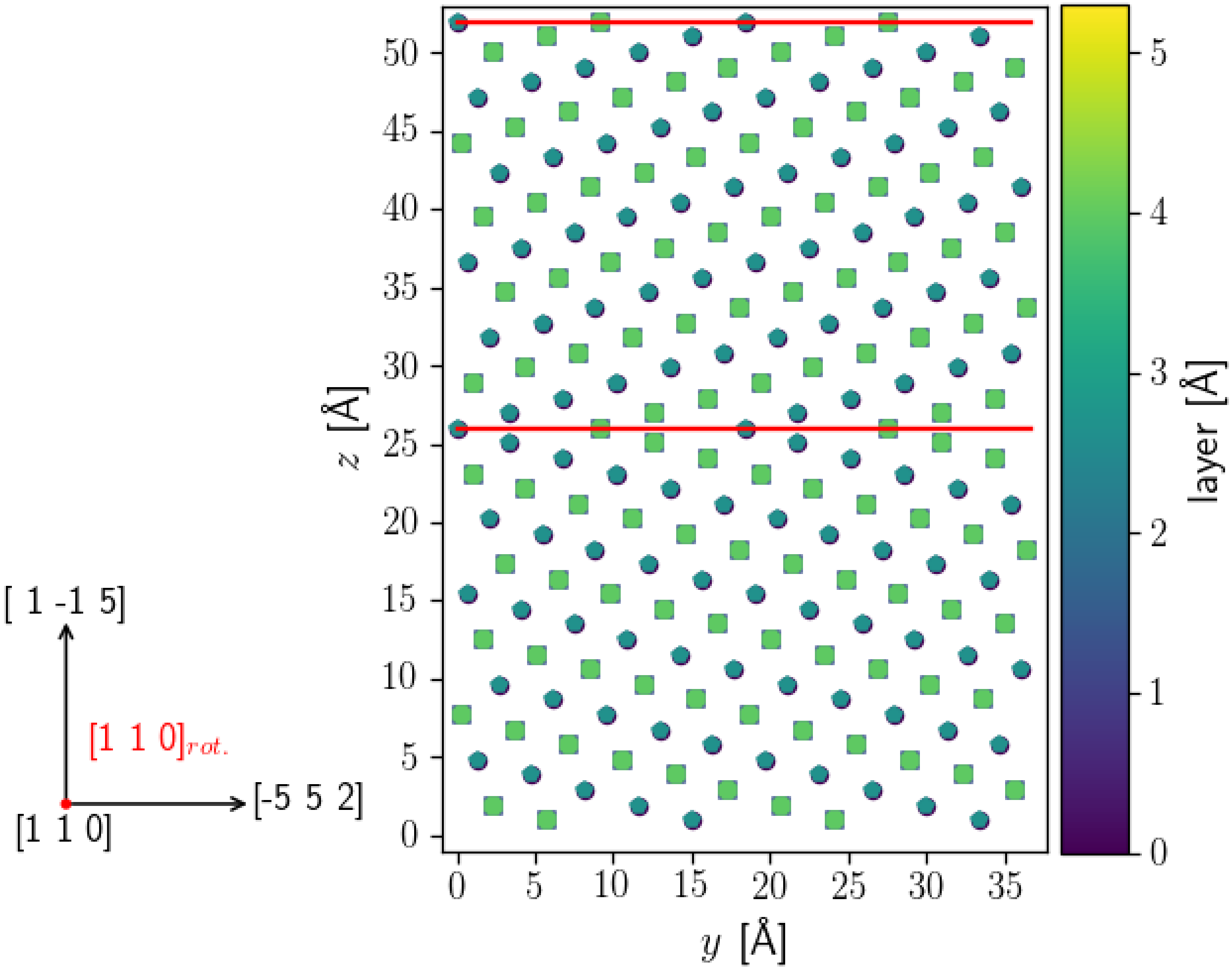


Figure B.5: The Σ27*a* tilt CSL GB with 4 layers along [1 1 0].

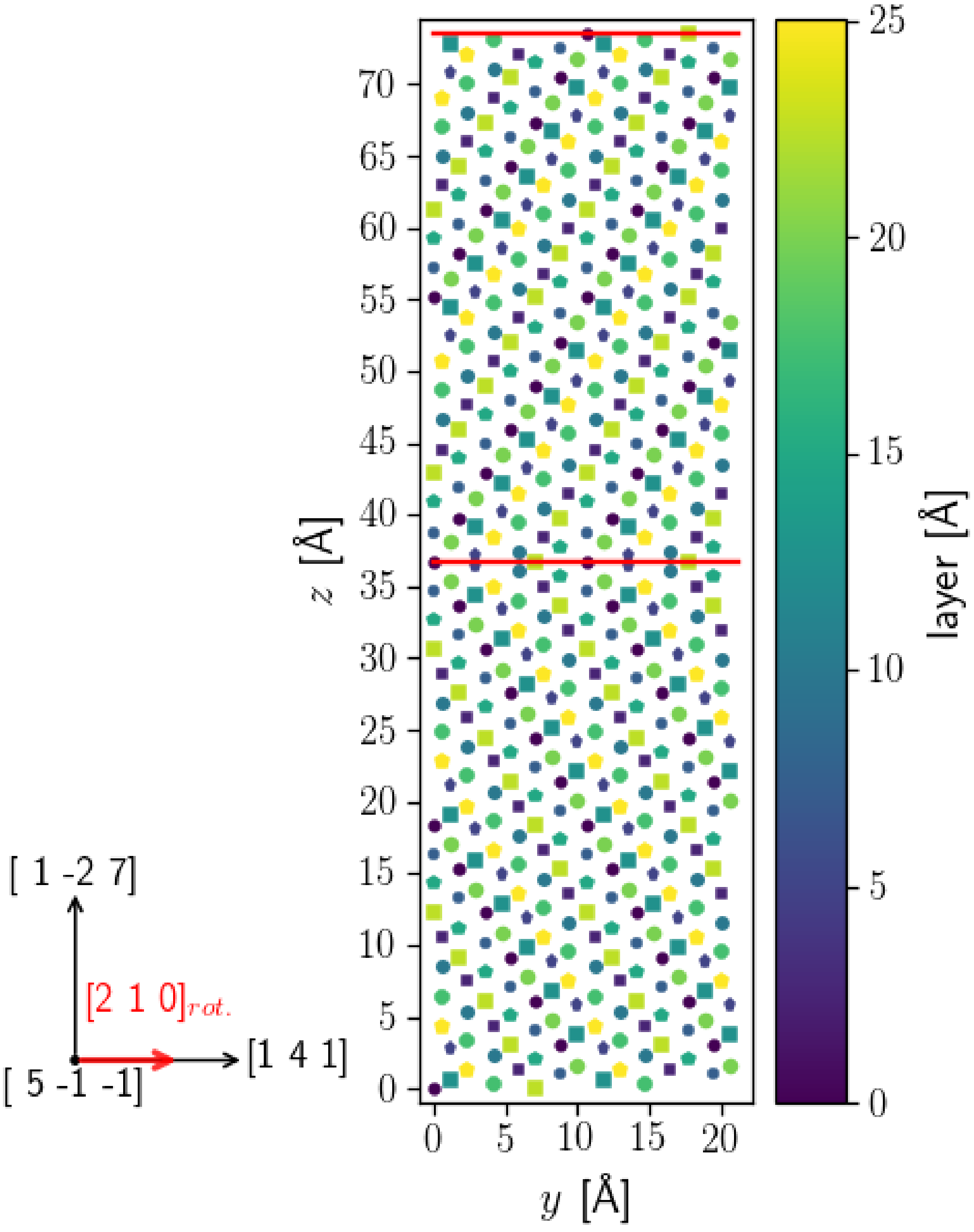


Figure B.6: The $\Sigma 27b$ tilt CSL GB with 30 layers along $[5\ -1\ -1]$.

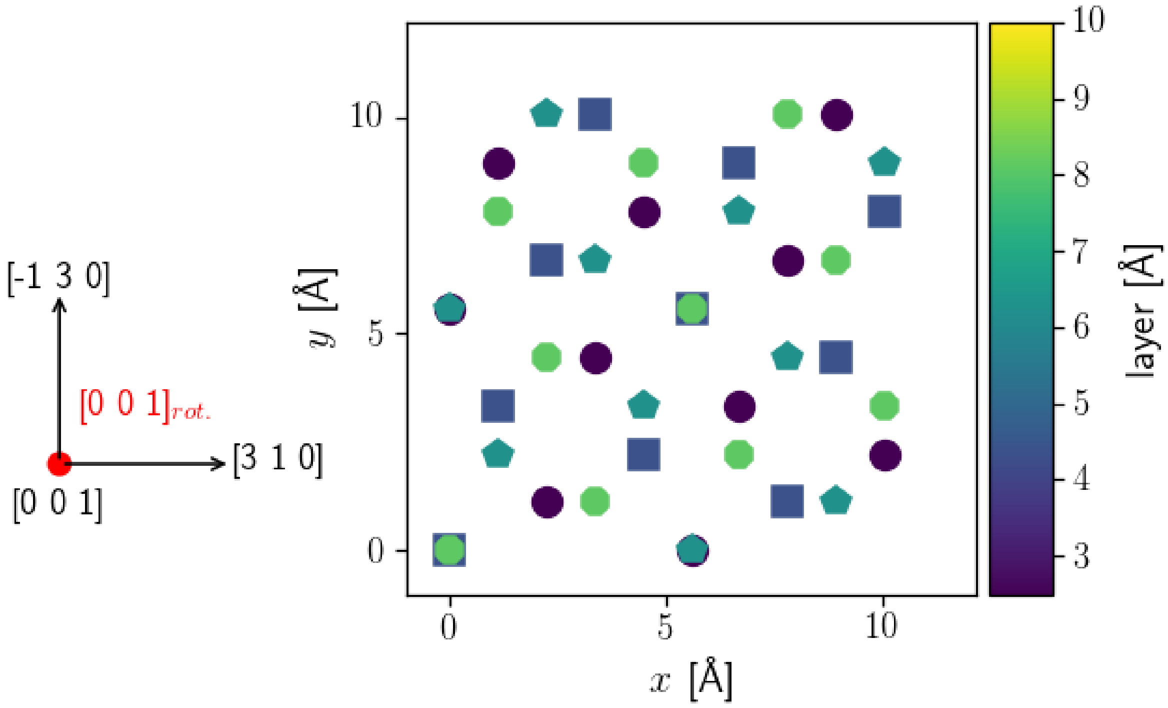


Figure B.7: The Σ5 tilt CSL GB with 4 layers along $[0\,0\,1]$.

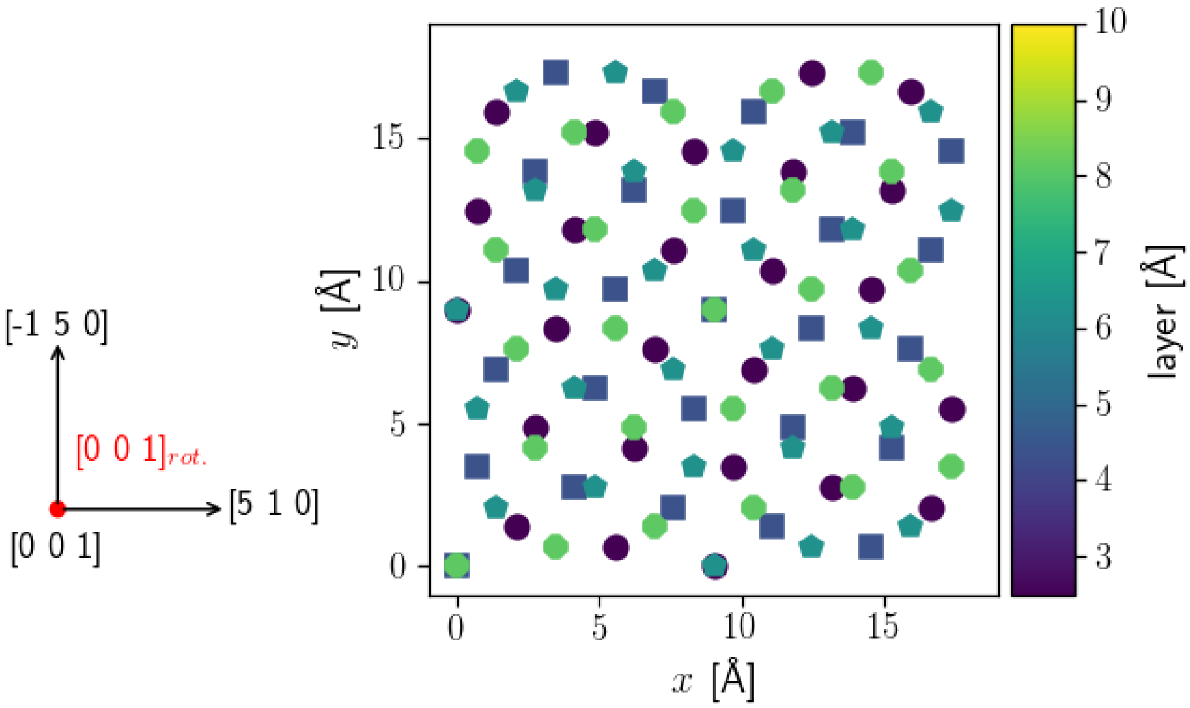


Figure B.8: The Σ13 tilt CSL GB with 4 layers along $[0\,0\,1]$.

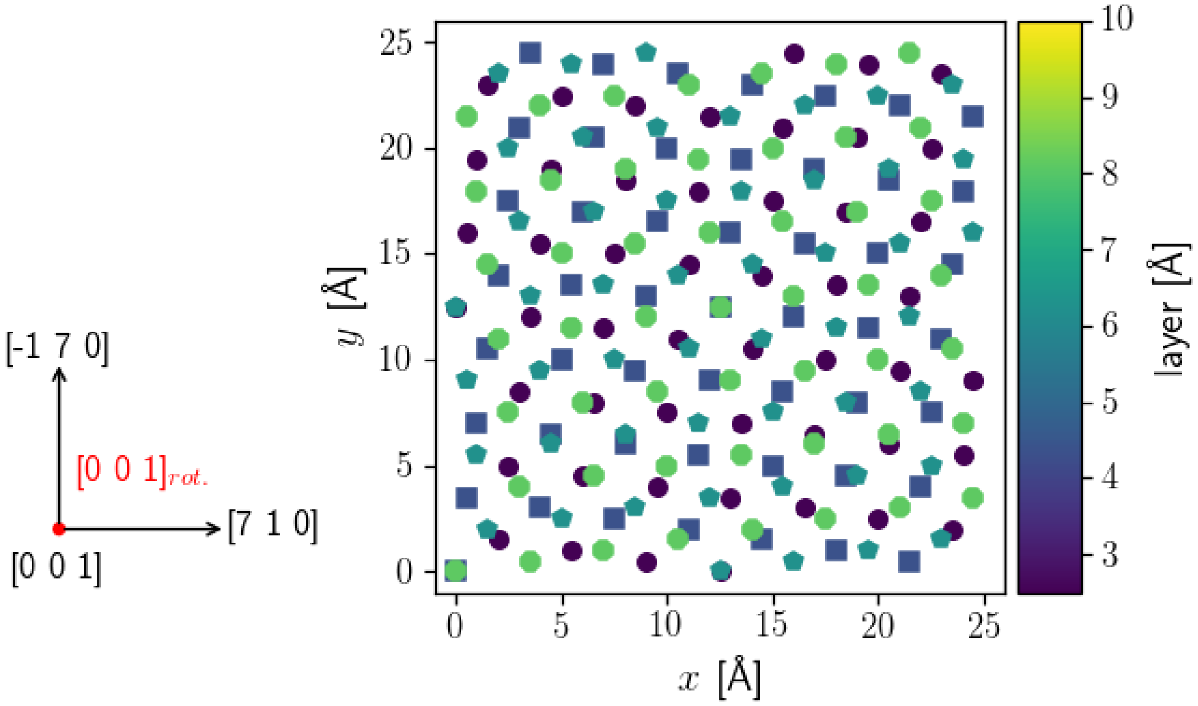


Figure B.9: The Σ25 tilt CSL GB with 4 layers along [0 0 1].

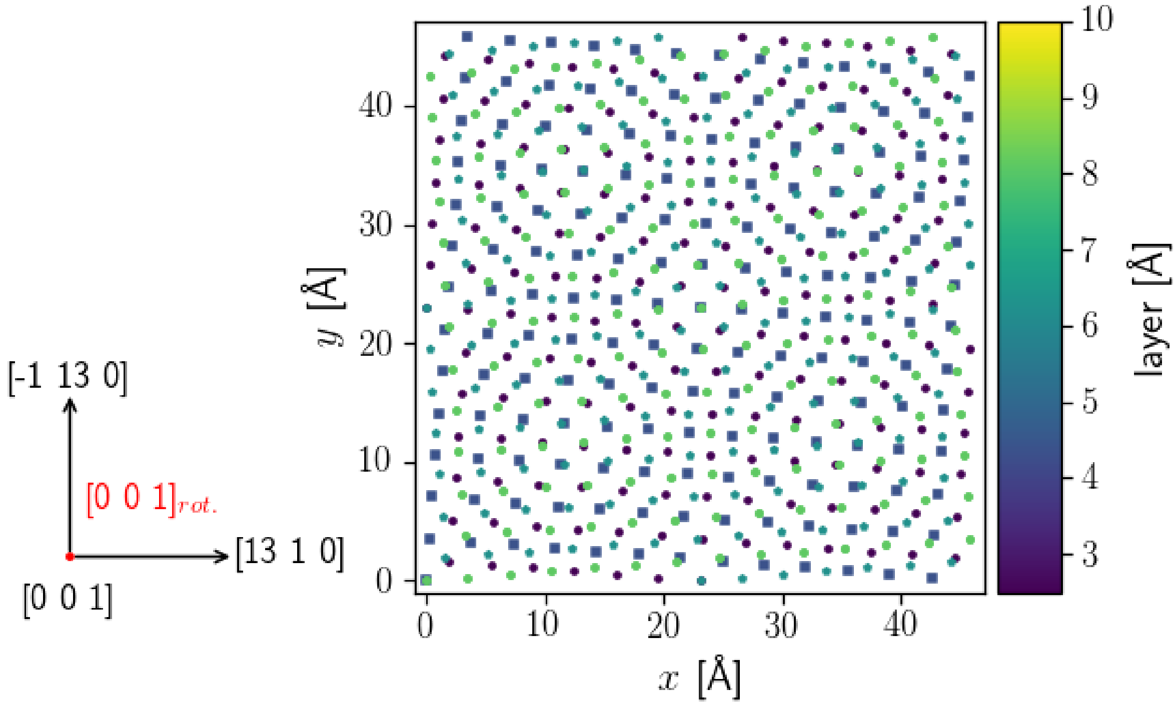


Figure B.10: The Σ85 tilt CSL GB with 4 layers along [0 0 1].