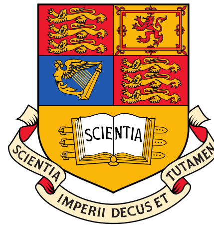


On the properties of point defects in silicon nanostructures from *ab initio* calculations

Fabiano Corsetti

A thesis submitted for the degree of

Doctor of Philosophy

and the

Diploma of Imperial College

Thomas Young Centre

Imperial College London

February 2012

[. . .] devant cette nuit chargée de signes et d'étoiles, je m'ouvrais pour la première fois à la tendre indifférence du monde.

Albert Camus, *L'Étranger*

Preface

The work presented in this thesis was carried out between October 2008 and December 2011 at Imperial College London, under the supervision of Dr. Arash A. Mostofi. I certify that I am the sole author of this thesis and that the work herein is my own, except where explicitly referenced. Part of this work has been published in the following paper: *System-size convergence of point defect properties: The case of the silicon vacancy*, F. Corsetti and A. A. Mostofi, Phys. Rev. B **84**, 035209 (2011).

Fabiano Corsetti
London, February 2012

Abstract

In this thesis we apply a variety of computational methods based on density-functional theory (DFT) to the study of defect centres in bulk silicon and silicon nanostructures.

Firstly, we discuss the system-size convergence of point defect properties in the supercell method for deep-level defects in bulk silicon; we consider both the vacancy and gold impurity.

For the vacancy, we investigate systematically the main contributions to the finite size error that lead to the well-known slow convergence with respect to system size of defect properties, and demonstrate that different properties of interest can benefit from the use of different k-point sampling schemes. We also present a simple and accurate method for calculating the potential alignment correction to the valence band maximum of charged defect supercells by using maximally-localised Wannier functions, and show that the localised view of the electronic structure provided by them gives a clear description of the nature of the electronic bonding at the defect centre.

For the gold impurity, we show that the system becomes a non-spin-polarised negative-U centre due to the effect of Jahn-Teller distortion, thus providing a simple explanation for the absent electron paramagnetic resonance signal for gold in silicon. The calculated transition levels are found to be in excellent agreement with experimental measurements.

We then investigate the segregation of arsenic impurities in silicon close to an interface with amorphous silica. We employ a multiscale approach, generating a realistic disordered interface structure from Monte Carlo simulation, with a continuous random network model of the system parametrised from DFT. We calculate the segregation energy using DFT for a large number of

substitutional sites encompassing all the oxidation states of silicon, and show that the results can be understood with a minimal model based only on the local strain and volume of the defect site.

Acknowledgements

Firstly, I would like to thank my supervisor, Arash Mostofi, for his help and support throughout my time at Imperial, and for patiently listening to and answering all my questions, even the really weird ones.

I am grateful to Nick Hine for his guidance in all manner of queries physical and computational. Thanks also go to my second supervisor, Matthew Foulkes, for several illuminating discussions on condensed matter physics, and James Spencer, for sorting out my Linux problems on innumerable occasions.

I would like to thank all the members of my research group, past and present, for making my PhD such an enjoyable, interesting and convivial experience. I would also like to thank the Thomas Young Centre for creating a stimulating environment in which to carry out my research, and for providing the free food that fuelled it.

Most of all (sorry Arash!), my thanks go to Jo Sarsam and Laura Ratcliff, for wading with me through the waters of understanding and fighting side by side against the tide of segmentation faults and mysterious minus signs. To everyone else, I say: "You don't know man, you weren't there!"

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC). The calculations were performed on cx1/cx2 (Imperial College London High Performance Computing Service) and HECToR (UK National Supercomputing Service). Access to HECToR was partly provided by the UK's HPC Materials Chemistry Consortium and Car-Parrinello Consortium. Finally, I would like to thank Martti Puska and Risto Nieminen for helpful discussions on the silicon vacancy, and Matt Probert, Phil Hasnip, and Keith Refson for their guidance in the art of using CASTEP.

Contents

1 Introduction **21**

1.1 Theory and simulation of materials 21

1.2 First-principles defect calculations 24

1.3 Outline of thesis 25

2 Many-body quantum mechanics **27**

2.1 The Schrödinger equation 27

2.1.1 The variational principle 29

2.1.2 Non-interacting particles 33

2.2 The Born-Oppenheimer approximation 39

2.3 Hartree theory 43

2.4 Hartree-Fock theory 47

2.5 Density-functional theory 51

2.5.1 The two theorems 51

2.5.2 The Kohn-Sham formalism 53

2.5.3 Exchange and correlation 55

3 Computational methods for DFT **59**

3.1 The pseudopotential approximation 60

3.2 Plane waves and periodic systems 64

3.3 Localised orbitals 70

3.3.1 The tight-binding approach 71

3.3.2 Maximally-localised Wannier functions 74

3.3.3 The density matrix formalism and linear-scaling DFT . . 78

4 Point defects in crystalline materials **83**
4.1 Classification of defects . . . 83
4.2 Defect concentration at equilibrium . . . 85
4.3 Doping in semiconductors . . . 89
4.4 Simulation of defect properties . . . 92
4.4.1 Formation energies . . . 96
4.4.2 Chemical potentials . . . 98
4.4.3 Transition levels . . . 100
4.4.4 The band gap problem in DFT . . . 103
4.4.5 Band gap correction schemes . . . 106

5 The silicon vacancy **111**
5.1 Previous studies . . . 113
5.2 Watkins' vacancy model . . . 115
5.3 Computational methodology . . . 117
5.3.1 Supercell calculations within DFT . . . 117
5.3.2 Linear-scaling DFT calculations . . . 120
5.3.3 Maximally-localised Wannier functions . . . 120
5.4 Determining the electronic chemical potential . . . 121
5.4.1 The real-space Voronoi cell method . . . 123
5.4.2 The MLWF method . . . 124
5.5 Results . . . 126
5.5.1 The neutral unrelaxed vacancy . . . 126
5.5.2 The neutral relaxed vacancy . . . 131
5.5.3 Charged vacancies . . . 135
5.5.4 Transition levels . . . 137
5.5.5 Visualisation using MLWFs . . . 139
5.6 Conclusions . . . 142

6 Gold in silicon **145**
6.1 Previous studies . . . 146
6.2 Computational methodology . . . 148
6.3 Electrostatic image charge corrections . . . 150

6.4 Results . . . 153
6.4.1 The neutral defect centre . . . 153
6.4.2 Charged defect centres . . . 157
6.4.3 Transition levels . . . 159
6.4.4 Watkins' model . . . 164
6.4.5 MLWF analysis . . . 166
6.5 Conclusions . . . 169

7 Doping at the Si–SiO_2 interface 171
7.1 Simulating the interface . . . 173
7.1.1 Previous studies . . . 173
7.1.2 Ideal relaxed interfaces . . . 177
7.1.3 Disordered interfaces with the CRN model . . . 179
7.2 Arsenic segregation at the interface . . . 188
7.2.1 Previous studies . . . 188
7.2.2 Computational methodology . . . 190
7.2.3 Results . . . 192
7.2.4 Conclusions . . . 202

8 Conclusions 205
8.1 Future work . . . 207

A Phonon calculations in ONETEP 211

B The CASTEP-TO-WANNIER90 code 219

C Wannier representations of bulk silicon 225

D The extended LCAO vacancy model 233

List of Figures

1.1 Hierarchy of materials modelling techniques 22

2.1 Rayleigh-Ritz scheme applied to a 1D harmonic oscillator potential 32

3.1 True and pseudised wavefunctions and potentials for the 2s orbital of carbon 61
3.2 Schematic illustration of the use of a plane-wave basis in a finite 1D system 66
3.3 Wannier functions for the 1D free electron model 75
3.4 NGWF optimisation in ONETEP for a carbon p orbital in ethene 81

4.1 Density of states for a doped semiconductor 92
4.2 Charge transition levels for the N_i defect in wurtzite GaN 102

5.1 LCAO model for the vacancy in silicon 116
5.2 Potential alignment correction for v^{2-} using the MLWF and real-space Voronoi cell methods 124
5.3 Convergence of k-point volume for the neutral unrelaxed vacancy supercell 127
5.4 Convergence of the vacancy formation energy with system size for different supercell geometries 129
5.5 Dispersion of the unrelaxed defect level 130
5.6 Relaxation effects for the neutral vacancy 131
5.7 Band structures for the 256-atom BCC supercell with a vacancy and aluminium impurity 132

5.8 Transition levels for the silicon vacancy 138
5.9 Contour-surface plots of the MLWFs most strongly associated with the vacancy defect centre 140
5.10 Split vacancy configuration for the doubly negative charge state of the defect centre 141

6.1 Electrostatic corrections of the defect formation energy for the Au^{3-} centre 152
6.2 Convergence of the gold substitutional formation energy with system size for different supercell geometries 153
6.3 Relaxation effects for the neutral gold substitutional defect . . . 155
6.4 Band structure for the 256-atom BCC supercell with a gold impurity 156
6.5 Transition levels for the gold substitutional defect 162
6.6 Schematic level diagram for the gold substitutional defect, following Watkins' model 165
6.7 Contour-surface plot of the MLWFs most strongly associated with the Au^0 defect centre (unrelaxed) 166
6.8 Projection of the three types of MLWFs on the valence+deep-level DOS 167
6.9 Contribution of all MLWFs in the supercell to the three defect levels in the band gap 168

7.1 Ideal Si–SiO2 interface supercells 176
7.2 Fitting of CRN parameters to DFT results. 182
7.3 Tetrahedral volume of the silicon ions in different oxidation states plotted along the z-direction 186
7.4 Normalised partial pair correlation functions for silica in the Si–SiO_2 system from the CRN-MC simulation 187
7.5 Segregation energy of the arsenic dopant at the ideal interface with respect to its bulk position 193
7.6 Calculated spread and displacement of the centre of the charge density distribution for the dopant donor level in the ideal interface system. 195

7.7 Segregation energy of the arsenic dopant at the disordered interface with respect to its bulk position. 196
7.8 Effect of lattice relaxation on the segregation energy 201

A.1 On-site force constant for bulk silicon calculated with the finite difference method for a range of ionic displacements. 214
A.2 Phonon dispersion and density of states for bulk silicon 215
A.3 Thermodynamic functions for bulk silicon 216
A.4 Effect of applying a real-space cut-off to the force constants matrix for bulk silicon 218

C.1 Contour-surface plots showing the two different Wannier representations of bulk silicon 226
C.2 Tight-binding matrix elements for the two Wannier representations 228
C.3 Silicon band structures calculated with the front- and back-bonded Wannier bases 231

D.1 Band structure for the BCC supercell calculated with the extended LCAO vacancy model 236
D.2 Effect of relaxation for the extended LCAO vacancy model . . . 237

List of Tables

5.1 Comparison of the potential alignment correction obtained from the MLWF and real-space Voronoi cell methods 125
5.2 Formation energies for the neutral unrelaxed vacancy in all supercells up to 256 atoms 126
5.3 Summary table of the results obtained for the 256-atom BCC supercell and the 1000-atom SC supercell with a vacancy 134
5.4 Transition levels for the 256-atom BCC supercell and the 1000-atom SC supercell with a vacancy 136

6.1 Formation energies for the neutral unrelaxed gold substitutional defect in all supercells up to 256 atoms 154
6.2 Summary table of the results obtained for the 256-atom BCC supercell with a gold substitutional defect 158
6.3 Transition levels for the 256-atom and 864-atom BCC supercells with a gold substitutional defect 160

7.1 List of parameters for the CRN model applied to the silicon–silica system 184
7.2 Summary table of the results obtained for the lattice relaxation of the doped system 200

Chapter 1

Introduction

1.1 Theory and simulation of materials

> If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generations of creatures, what statement would contain the most information in the fewest words? I believe it is the *atomic hypothesis* [. . .] that *all things are made of atoms—little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another*. In that one sentence, you will see, there is an *enormous* amount of information about the world, if just a little imagination and thinking are applied.
>
> R. P. Feynman, *The Feynman Lectures on Physics* [1]

It is perhaps surprising to think that the atomic theory of matter with which we are so familiar is barely a century old[1], dating back to the discovery of the electron by J. J. Thomson in 1897 and that of the nucleus by E. Rutherford in 1911. In this time, quantum mechanics has established itself as the fundamental physical theory in describing the behaviour of electrons, atomic nuclei

[1]We refer specifically to the modern theory of subatomic particles, which has its roots in the atomic theories of the 19th century. The earliest atomic theories, developed in ancient India and Greece, should be viewed as philosophical rather than scientific in nature.

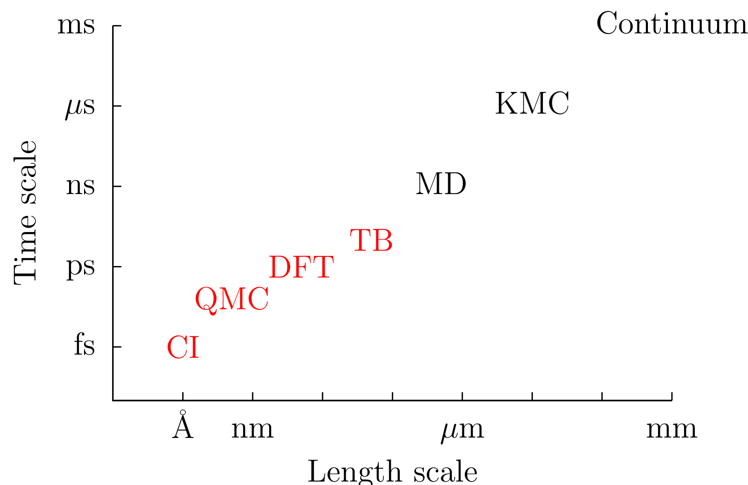


Figure 1.1: Hierarchy of materials modelling techniques. Explicitly quantum-mechanical methods are labelled in red. The positioning of the methods is not exact, but shows the typical scale at which each one can currently be applied. The methods referred to are: configuration interaction (CI), quantum Monte Carlo (QMC), density-functional theory (DFT), empirical tight-binding (TB), classical molecular dynamics (MD), kinetic Monte Carlo (KMC), and continuum modelling; this last method can extend significantly beyond the range shown.

and electromagnetic radiation. It is now clear that all of the intricate phenomena found in biology, chemistry and low-energy physics can in principle be traced back to the equations of quantum mechanics applied to large numbers of particles.

In practice, however, analytical solutions to these equations only exist for the very simplest systems; in order to be able to model realistic materials we need to use computational methods to find numerical solutions. Even so, problems of interest require a considerable computational effort and rely on a number of approximations. The field of materials modelling is a large and varied one, both in terms of the systems that are studied and the physical theories that are employed; these range from extremely ordered, periodic systems (crystalline condensed matter) to disordered or finite molecular ones (soft

matter such as colloids, polymers and other biological systems), and from conventional phases of matter (solids, liquids and gases) to more exotic ones (such as superconductors and Bose-Einstein condensates).

Different systems exhibit phenomena of interest on different length and time scales; accordingly, many different computational methods have been developed for materials simulation, incorporating aspects of physics, chemistry, materials science and engineering. Fig. 1.1 shows some of the most important techniques in materials modelling, and the typical scale at which each can be used. We focus mainly on small length and time scales, for which it is possible to employ first-principles or *ab initio* many-body quantum-mechanical methods; in particular, in this thesis we shall present results from calculations using density-functional theory [2–4] (DFT). DFT has become an extremely popular method due to its *ab initio* nature[2] and its balance of accuracy and computational speed, and has successfully been used to understand and predict a wide variety of properties for different classes of materials [5]. Furthermore, due to the availability of greater computing resources and ongoing advances in theoretical methods, the number of atoms that can be simulated with DFT has greatly increased in the last decade.

Fortunately, as the size of systems that can be modelled has increased, the corresponding size of systems that can be probed and controlled experimentally has decreased; this suggests a possible meeting point between first-principles simulations and experiment [6]. In fact, electronic structure calculations have already provided much insight into the properties of novel nanostructures such as quantum dots [7–10], semiconducting nanowires [11–13], carbon nanotubes [14, 15] and molecular electronics devices [16, 17]. Such studies show that the electronic properties of nanoscale materials depend strongly on the atomistic detail (e.g., surface reconstruction); therefore, *ab initio* simulations are often necessary for obtaining both qualitatively and quantitatively correct results.

[2]We note that the definition of what constitutes an *ab initio* method is somewhat ambiguous and context-dependent. In this thesis, we use the term simply to denote a method that does not rely on numerical parameters derived from experimental measurements (other than the fundamental physical constants), even if the choice of approximations employed in the underlying theory is guided by our experience of the method's performance in predicting experimental results.

As the codes used to perform such simulations become more stable, accurate and computationally inexpensive, it is increasingly the case that first-principles methods can be used routinely not only to explain experimental findings but also to predict the properties of as-yet unbuilt nanostructures, thus allowing the design of new devices.

1.2 First-principles defect calculations

In this thesis we present the results from our investigation of the properties of several different point defects in silicon-based systems from first-principles DFT calculations. The study of such systems is of significant technological interest, as the presence of point defects can have many consequences for the optical, electrical and mechanical properties of solids; in fact, the functioning of semiconducting devices depends crucially on the amount of impurities and lattice defects they contain.

The simulations themselves present a number of interesting theoretical challenges, as the disruption caused by the defect centre on the crystalline lattice can be extremely long-ranged; this requires the use of large system sizes in order to sufficiently approach the dilute limit to accurately calculate fundamental defect properties such as the formation energy, the lattice relaxation around the defect centre, and the relative stability of different defect configurations. However, the accuracy of the quantum-mechanical description of the electronic structure (dictated by the choice of the method used) can also have a large impact on these properties. It is often unclear whether the errors in the final results for a given study should be attributed to limitations in system size or in the accuracy of the underlying theory.

We first study the issue of finite size convergence in two deep-level defects in bulk crystalline silicon, the vacancy and gold substitutional. In the former case, we analyse the various contributions to the total finite size error, and investigate the effect of different k-point sampling schemes on the convergence with system size of two defect properties, the formation energy and the stable charge state transition levels. In the latter case, we use the insights obtained from the vacancy to obtain well-converged results for the gold impurity, show-

ing this to be crucial for recovering the correct behaviour of the isolated defect in DFT; the calculations provide a novel interpretation of the experimental results for this defect centre.

We then study a shallow-level dopant, the arsenic impurity, in a complex nanoscale environment, the interface between crystalline silicon and its amorphised oxide. We use a realistic large-scale model of the interface, and perform calculations of the impurity at many different lattice sites in the system, in order to build up a clear picture of the effect of a number of structural and environmental parameters (the proximity to the interface, and the local volume and stress of the substitutional site) on the segregation energy for this defect centre with respect to its position in bulk silicon.

Overall, we aim to show that modern large-scale DFT calculations can be employed to accurately simulate many properties of interest for point defects in semiconductors, thus providing both new insight into the electronic structure of bulk defects that have already been studied extensively for many decades (such as the gold impurity), and predictions for the behaviour of nanoscale systems that are becoming increasingly important in the field of nanotechnology (such as dopants near the semiconductor–oxide interface, which are expected to play a large role in future ultrasmall semiconducting devices).

1.3 Outline of thesis

We begin the thesis with a detailed review of the theoretical and computational quantum-mechanical methods that our simulations are based on; this is provided for the interested reader, and is recommended for those not already familiar with such methods. In Chapter 2, we describe the fundamental mathematical problem posed by many-body quantum mechanics for a system of nuclei and electrons, show how this can be simplified with the Born-Oppenheimer approximation for separating the nuclear and electronic degrees of freedom, and outline three methods for obtaining solutions to the problem; the first two of these are approximate (Hartree theory and Hartree-Fock theory), and the third, DFT, is in principle exact, but requires the approximation of an unknown functional of the electronic density for practical use. In Chapter 3,

we describe some important computational techniques for the application of DFT: the pseudopotential approximation, used to remove the chemically inert core electrons of the atoms and replace them with an effective non-local potential, and the use of basis sets to solve the eigenvalue problem, considering both plane waves and localised orbitals. Finally, we describe a reformulation of plane-wave DFT for obtaining a method that scales only linearly with the number of atoms in the system.

In Chapter 4, we introduce the subject of point defects: we describe the theory for calculating properties of interest such as the concentration at thermodynamic equilibrium, the formation energy and the defect transition levels, and discuss the important band gap problem in DFT and its implications for defect calculations, reviewing the proposed correction methods for it.

We then present and discuss the main results of the thesis, as introduced in the previous section. In Chapter 5, we describe our investigation of the finite size convergence properties for the silicon vacancy. In Chapter 6, we describe our investigation of the negative-U effect for the gold impurity in silicon. In Chapter 7, we describe our investigation of arsenic doping at the silicon–silica interface, first focusing on the generation of the interface between crystalline silicon and amorphous silica using a Monte Carlo method, and then on the calculation of segregation energies for arsenic at the interface using DFT. Finally, in Chapter 8, we summarise the results that we have presented, and discuss future work for these systems.

In the Appendices, we present further work that was carried out on a number of different topics: in Appendix A, we describe our implementation of phonon calculations in the linear-scaling DFT code ONETEP; in Appendix B, we describe a code that was written for interfacing between the CASTEP and WANNIER90 codes, in order to obtain the maximally-localised Wannier functions for our defect systems; in Appendix C, we describe our investigation of different Wannier representations of bulk silicon; and in Appendix D, we describe the extended LCAO model of the silicon vacancy that we developed for explaining the effect of different k-point sampling schemes on the symmetry of the relaxed defect centre.

Chapter 2

Many-body quantum mechanics

2.1 The Schrödinger equation

As noted in the previous chapter, the investigation of the microscopic properties of any material requires the solution of the equations of quantum mechanics for all of the electrons and nuclei in the system. In fact, one of the central postulates of quantum mechanics is that the behaviour of a system is described in its entirety by its wavefunction $|\Psi\rangle$, which is in turn governed in the non-relativistic limit by the time-dependent Schrödinger equation[1]:

$$\hat{H}\,|\Psi\rangle = \mathrm{i}\frac{\partial}{\partial t}\,|\Psi\rangle\,. \tag{2.1}$$

In this case we have used the Dirac ket notation for the wavefunction to accentuate the fact that the state of a system is representation invariant, although in practice we work in the position basis. $\hat{H}$, known as the Hamiltonian, is the Hermitian operator associated with the total energy of the system, which contains all of the interactions between particles.

In a one-particle system the wavefunction has three spatial coordinates $\mathbf{r}$ and one spin coordinate σ, and therefore can be written as $\Psi(\mathbf{r}, \sigma, t)$. For convenience, σ is defined such that the spin (i.e., the intrinsic angular momentum of the particle) measured along an arbitrary axis α is $S_\alpha = \sigma/2$ for a spin-$\frac{1}{2}$

[1]Dimensionless equations in Hartree atomic units are used throughout, for which $\hbar = e = m_e = 4\pi\epsilon_0 = a_0 = 1$.

particle (e.g., an electron); this is because a measurement on such a particle must always yield $S_\alpha = \pm\frac{1}{2}$, and so the two possible values of σ are ± 1 (often denoted by $\uparrow$ and $\downarrow$). In the more general case of an N-particle system, the formalism is unchanged, but the wavefunction must encode the spatial and spin information of all the particles. The complete many-body wavefunction is therefore written as $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \sigma_1, \sigma_2, \ldots, \sigma_N, t)$. We note that the many-body wavefunction has the same properties as the one-body one: in particular, $|\Psi|^2$ is the probability of finding the system in the state given by $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \sigma_1, \sigma_2, \ldots, \sigma_N\}$ at time t, and therefore integrating this function over all its spatial and spin dimensions must give 1 (the normalisation condition). Similarly, the expectation value of the operator $\hat{O}$ is given by $\langle\Psi|\,\hat{O}\,|\Psi\rangle$, with the corresponding integral being carried out over all the spatial and spin dimensions of the N particles.

If the Hamiltonian is time-independent, as is generally taken to be the case when calculating the properties of a crystal, we can separate the time variable in the wavefunction:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \sigma_1, \sigma_2, \ldots, \sigma_N, t) = \psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \sigma_1, \sigma_2, \ldots, \sigma_N)\,\zeta(t)\,. \tag{2.2}$$

Substituting into Eq. 2.1, we obtain the time-independent Schrödinger equation

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle\,, \tag{2.3}$$

which is an eigenvalue equation with the total energy of the system E as the eigenvalue and the time-independent part of the wavefunction $|\psi\rangle$ as the eigenvector. We note that, from Sturm-Liouville theory, the normalised eigenvector solutions of this equation form a complete orthonormal set, and so any valid state of the system can be expanded as a linear combination of them. The corresponding eigenvalues are real and ordered such that

$$E_0 < E_1 < E_2 < \ldots < E_n < \ldots \to \infty \tag{2.4}$$

(assuming non-degenerate solutions); the lowest eigenvalue, E_0, is the ground-state energy, while all higher eigenvalues are excited-state energies.

The time-dependence of the complete wavefunction $|\Psi\rangle$ is now given simply by

$$\zeta(t) = \mathrm{e}^{-\mathrm{i}Et}, \tag{2.5}$$

which does not affect the expectation values of the system. In the following discussion we can therefore neglect the time-dependent part of the wavefunction, and deal solely with the time-independent problem given by Eq. 2.3.

2.1.1 The variational principle

The quantity E in Eq. 2.3 is the expectation value of the Hamiltonian for an eigenstate of the system:

$$\langle\psi|\,\hat{H}\,|\psi\rangle = \langle\psi|\,E\,|\psi\rangle = E\,\langle\psi|\psi\rangle = E. \tag{2.6}$$

In general, we can consider any unnormalised wavefunction, which does not need to satify Eq. 2.3, and modify the definition of the expectation value to include a normalising prefactor[2]:

$$\overline{E}[\psi] = \frac{\langle\psi|\,\hat{H}\,|\psi\rangle}{\langle\psi|\psi\rangle}. \tag{2.7}$$

The notation emphasises the fact that $\overline{E}$ can be considered to be a functional of $|\psi\rangle$, since it maps any allowed complex wavefunction onto a real number. Let us now consider the change in energy $\delta\overline{E}[\psi]$ due to a small variation in

[2]The expectation value of the Hamiltonian is denoted $\overline{E}$ to distinguish it from the eigenvalue solutions E of the time-independent Schrödinger equation.

the wavefunction $|\delta\psi\rangle$; this is given by

$$
\begin{aligned}
\delta\overline{E}\left[\psi\right] =& \overline{E}\left[\psi+\delta\psi\right]-\overline{E}\left[\psi\right] \\
=& \frac{\langle\psi+\delta\psi|\,\hat{H}\,|\psi+\delta\psi\rangle}{\langle\psi+\delta\psi|\psi+\delta\psi\rangle}-\frac{\langle\psi|\,\hat{H}\,|\psi\rangle}{\langle\psi|\psi\rangle} \\
=& \frac{\langle\psi|\,\hat{H}\,|\psi\rangle}{\langle\psi|\psi\rangle+\langle\delta\psi|\psi\rangle+\langle\psi|\delta\psi\rangle}+\frac{\langle\delta\psi|\,\hat{H}\,|\psi\rangle+\langle\psi|\,\hat{H}\,|\delta\psi\rangle}{\langle\psi|\psi\rangle}-\frac{\langle\psi|\,\hat{H}\,|\psi\rangle}{\langle\psi|\psi\rangle} \\
&+\mathcal{O}\left(\delta\psi^2\right) \\
=& \frac{1}{\langle\psi|\psi\rangle}\Big\{\langle\delta\psi|\,\hat{H}\,|\psi\rangle+\langle\psi|\,\hat{H}\,|\delta\psi\rangle-\overline{E}\left[\psi\right]\left(\langle\delta\psi|\psi\rangle+\langle\psi|\delta\psi\rangle\right)\Big\} \\
&+\mathcal{O}\left(\delta\psi^2\right) \\
=& \frac{1}{\langle\psi|\psi\rangle}\Big\{\langle\delta\psi|\,\hat{H}-\overline{E}\left[\psi\right]|\psi\rangle+\langle\psi|\,\hat{H}-\overline{E}\left[\psi\right]|\delta\psi\rangle\Big\}+\mathcal{O}\left(\delta\psi^2\right),
\end{aligned}
\tag{2.8}
$$

which vanishes for all $|\delta\psi\rangle$ (thus making $\overline{E}\left[\psi\right]$ stationary) when $\hat{H}\,|\psi\rangle = \overline{E}\left[\psi\right]|\psi\rangle$. In other words, finding the stationary values of $\overline{E}\left[\psi\right]$ is equivalent to solving the eigenvalue problem given by Eq. 2.3. We note that it is possible to obtain the same result starting from the unnormalised definition of the expectation value of the Hamiltonian $\langle\psi|\,\hat{H}\,|\psi\rangle$ and imposing the normalisation constraint $\langle\psi|\psi\rangle = 1$ via an undetermined Lagrange multiplier, which would then be found to be equivalent to the eigenvalue E.

The formulation of the time-independent Schrödinger equation as a variational principle is particularly useful if we are interested in finding the lowest-lying eigenstates of a system, as it provides an alternative way of doing so that does not require the solution of the full eigenvalue problem. To illustrate this method, let us consider an arbitrary wavefunction $|\psi\rangle$, which we can expand using the complete set of eigenvalues of the Hamiltonian $\{\phi_n\}$ as a basis:

$$
|\psi\rangle = \sum_{n=0}^{\infty} c_n\,|\phi_n\rangle\,, \tag{2.9}
$$

where

$$
\hat{H}\,|\phi_n\rangle = E_n\,|\phi_n\rangle\,. \tag{2.10}
$$

Substituting Eq. 2.9 into Eq. 2.7, we find:

$$\begin{aligned}\overline{E}\left[\psi\right] &= \frac{\langle\psi|\,\hat{H}\,|\psi\rangle}{\langle\psi|\psi\rangle} \\ &= \frac{\sum_{n,m} c_n^* c_m \langle\phi_n|\,\hat{H}\,|\phi_m\rangle}{\sum_{n,m} c_n^* c_m \langle\phi_n|\phi_m\rangle} \\ &= \frac{\sum_{n,m} c_n^* c_m E_m \delta_{nm}}{\sum_{n,m} c_n^* c_m \delta_{nm}} \\ &= \frac{\sum_n |c_n|^2 E_n}{\sum_n |c_n|^2}.\end{aligned} \tag{2.11}$$

By definition, the ground-state energy $E_0 \leq E_n$. Therefore, it follows that

$$\sum_n |c_n|^2 E_0 \leq \sum_n |c_n|^2 E_n, \tag{2.12}$$

from which we arrive at the following variational inequality:

$$\overline{E}\left[\psi\right] = \frac{\sum_n |c_n|^2 E_n}{\sum_n |c_n|^2} \geq E_0. \tag{2.13}$$

Eq. 2.13 therefore shows that the expectation value of the total energy for any trial wavefunction gives an upper bound estimate on the ground-state energy of the system. Furthermore, if we express the trial wavefunction as the sum of the (unnormalised) ground-state wavefunction and a small error term, such that

$$|\psi\rangle = c_0\,|\phi_0\rangle + \sum_{n=1}^{\infty} c_n\,|\phi_n\rangle\,, \tag{2.14}$$

we find that

$$\overline{E}\left[\psi\right] = E_0 + \sum_{n\neq 0} \frac{|c_n|^2}{|c_0|^2}\left(E_n - E_0\right) + \mathcal{O}\left(|c_n|^4\right). \tag{2.15}$$

Therefore, it is clear that the error in $\overline{E}\left[\psi\right]$ with respect to E_0 is second-order in the corresponding error in the wavefunction (i.e., c_n), and that the energy functional will only equal the true ground-state energy for $|\psi\rangle = c_0\,|\phi_0\rangle$

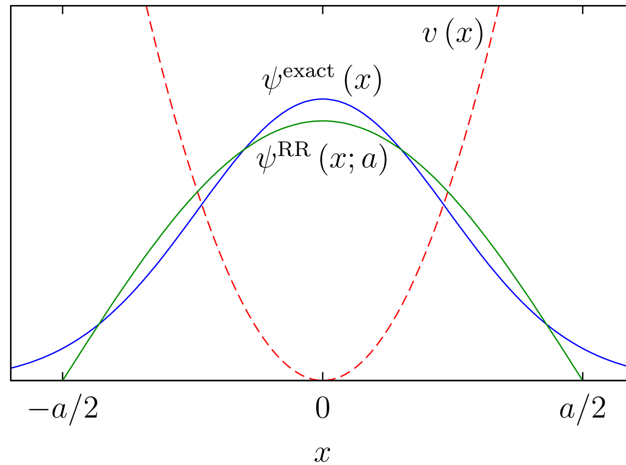


Figure 2.1: An illustration of a one-parameter Rayleigh-Ritz scheme applied to a 1D harmonic oscillator potential.

(assuming a non-degenerate ground state).

The variational inequality set out in Eq. 2.13 forms the basis of the Rayleigh-Ritz (RR) method for solving the time-independent Schrödinger equation. In this method, a trial wavefunction is constructed by choosing a functional form with a number of free parameters, and $\overline{E}[\psi]$ is then minimised by varying these parameters. By improving the accuracy of the trial wavefunction, either by using some physical intuition in choosing the functional form or increasing the number of parameters, a better estimate of E_0 can be obtained. We note that the use of the variational principle in this way is theoretically not restricted to the ground state, as excited states can also be found by constructing trial wavefunctions that are constrained to be orthogonal to all lower eigenstates.

A simple example of the RR method is shown in Fig. 2.1; here we are trying to find the ground-state energy and wavefunction for the 1D harmonic

oscillator potential $v(x) = \frac{1}{2}\omega^2 x^2$. We use the trial wavefunction

$$\psi^{\mathrm{RR}}(x;a) = \begin{cases} \sqrt{\frac{2}{a}}\cos\frac{\pi x}{a} & \text{if } |x| \leq \frac{a}{2} \\ 0 & \text{otherwise} \end{cases}, \tag{2.16}$$

which respects both the symmetry of the potential and the boundary conditions of the problem ($\psi(x) \to 0$ as $x \to \pm\infty$), and which depends parametrically on a. By minimising Eq. 2.7 with respect to this free parameter, we obtain the upper bound to the ground state energy $E_0^{\mathrm{RR}} \simeq 0.568\omega$; this is a good approximation of the true value ($E_0^{\mathrm{exact}} = \omega/2$), given the simplicity of the model.

2.1.2 Non-interacting particles

The many-body Hamiltonian for a system of N particles can be written as[3]

$$H = \sum_{i=1}^{N} t_i + V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N), \tag{2.17}$$

where

$$t_i = -\frac{1}{2m_i}\nabla^2_{\mathbf{r}_i} \tag{2.18}$$

is the single-particle kinetic energy operator for the i-th particle of mass m_i, and V is the potential energy of the classical system with the particles in positions $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$. Solving Eq. 2.3 exactly for such a system of interacting particles is an intractable problem for all but the simplest cases, and therefore a number of simplifying assumptions generally have to be employed in order to make it possible to arrive at an approximate solution for systems of interest; some of the methods used in this respect shall be discussed in this and the next chapter.

However, solving the many-body problem becomes simple for the case of perfectly non-interacting particles. This case can be viewed as an extreme simplification of a realistic system, and the solutions that result from it provide a

[3]In this definition we have neglected magnetic fields and relativistic effects.

useful starting point for understanding the more sophisticated approximations presented later.

For non-interacting particles, the potential energy is a sum of individual terms, each of which depends only on the coordinates of a single particle:

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = v_1(\mathbf{r}_1) + v_2(\mathbf{r}_2) + \ldots + v_N(\mathbf{r}_N); \tag{2.19}$$

consequently, the total Hamiltonian in Eq. 2.17 becomes a sum of single-particle Hamiltonians $\hat{h}_i$:

$$H = \sum_{i=1}^{N} \{t_i + v_i(\mathbf{r}_i)\} = \sum_{i=1}^{N} h_i. \tag{2.20}$$

This allows us to separate variables in the wavefunction, leading to a solution for Eq. 2.3 of the form

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \prod_{i=1}^{N} \phi_i(\mathbf{r}_i) \tag{2.21}$$

(we note that the spin variable has now been neglected, although it shall be reintroduced shortly). In other words, the many-body wavefunction has been reduced to a product of single-particle orbitals $\phi_i(\mathbf{r}_i)$. By substituting this wavefunction into the time-independent Schrödinger equation, we obtain:

$$\sum_{i=1}^{N} \frac{h_i \phi_i(\mathbf{r}_i)}{\phi_i(\mathbf{r}_i)} = E. \tag{2.22}$$

Since every term in the sum on the left-hand side of the equation must be constant, we can split the problem into a set of single-particle Schrödinger equations of the form

$$h_i \phi_i(\mathbf{r}_i) = \varepsilon_i \phi_i(\mathbf{r}_i), \tag{2.23}$$

where the single-particle energies ε_i sum to the total energy of the system E. In this picture, therefore, the many-body wavefunction can be constructed by choosing any eigenvector of the individual Schrödinger equation of each

particle:

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; j_1, j_2, \ldots, j_N) = \prod_{i=1}^{N} \phi_{i,j_i}(\mathbf{r}_i). \tag{2.24}$$

The notation has now been expanded to include the quantum number j_i that is selected from the set of solutions of Eq. 2.23 for the i-th particle. The corresponding total energy is

$$E(j_1, j_2, \ldots, j_N) = \sum_{i=1}^{N} \varepsilon_{i,j_i}. \tag{2.25}$$

The situation is simplified if we consider a system of identical particles, such as electrons. In fact, the single-particle Hamiltonians must also be identical, and so $\phi_i(\mathbf{r}_i)$ for every particle has to be chosen from the solutions of the same single-particle Schrödinger equation. Therefore, Eq. 2.23 only needs to be solved once, and the first subscript from $\phi_{i,j_i}(\mathbf{r}_i)$ and ε_{i,j_i} can be dropped.

It is tempting to assume that the ground-state wavefunction of such a system of electrons is therefore simply given by

$$\psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \prod_{i=1}^{N} \phi_0(\mathbf{r}_i), \tag{2.26}$$

with a total energy of $N\varepsilon_0$. However, this solution is only allowed for bosonic systems, for which the many-body wavefunction must be symmetric (i.e., unchanged under any permutation $\mathcal{P}$ of the particle labels $\{i\}$:

$$\psi^{\text{bos}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \psi^{\text{bos}}(\mathbf{r}_{\mathcal{P}_1}, \mathbf{r}_{\mathcal{P}_2}, \ldots, \mathbf{r}_{\mathcal{P}_N}). \tag{2.27}$$

Electrons, however, form a fermionic system, for which the many-body wavefunction must be antisymmetric:

$$\psi^{\text{ferm}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = (-1)^{\xi_{\mathcal{P}}} \psi^{\text{ferm}}(\mathbf{r}_{\mathcal{P}_1}, \mathbf{r}_{\mathcal{P}_2}, \ldots, \mathbf{r}_{\mathcal{P}_N}), \tag{2.28}$$

where $\xi_{\mathcal{P}} = \pm 1$ for an even $(+)$ or odd $(-)$ permutation.

Fortunately, the antisymmetry of the wavefunction can be restored in our

case by expressing the many-body wavefunction as a Slater determinant [18] of the same single-particle orbitals as before:

$$\psi\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; j_1, j_2, \ldots, j_N\right) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{j_1}\left(\mathbf{r}_1\right) & \phi_{j_1}\left(\mathbf{r}_2\right) & \ldots & \phi_{j_1}\left(\mathbf{r}_N\right) \\ \phi_{j_2}\left(\mathbf{r}_1\right) & \phi_{j_2}\left(\mathbf{r}_2\right) & \ldots & \phi_{j_2}\left(\mathbf{r}_N\right) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{j_N}\left(\mathbf{r}_1\right) & \phi_{j_N}\left(\mathbf{r}_2\right) & \ldots & \phi_{j_N}\left(\mathbf{r}_N\right) \end{vmatrix}. \tag{2.29}$$

Effectively, this wavefunction is the sum over all $N!$ permutations of particle labels of the simple symmetric product wavefunction given in Eq. 2.24, each multiplied by $(-1)^{\xi_{\mathcal{P}}}$ to ensure the overall antisymmetry. This wavefunction must still be an eigenfunction of the system, since it is a linear combination of individual solutions, all with the same energy (from Eq. 2.25). The only restriction on the set of single-particle orbitals is that they must all be orthogonal, and so all values of j_i must be different (otherwise, two rows of the matrix would be identical, and the determinant would vanish). Because of this, the lowest energy (and, hence, the ground state of the system) is constructed by selecting (or 'filling') the orbitals corresponding to the lowest N eigenvalues of the single-particle Schrödinger equation.

However, if we reintroduce the spin variable of the electrons we see that this is no longer the case: up to two electrons can occupy the same orbital, since their orthogonality can be maintained by the spin part of the wavefunction, thus far neglected. Therefore, the complete ground-state Slater determinant,

including spin, is[4]

$$\psi_0(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N,\sigma_1,\sigma_2,\ldots,\sigma_N)$$
$$=\frac{1}{\sqrt{N!}}\begin{vmatrix}\phi_1(\mathbf{r}_1)\chi_\uparrow(\sigma_1) & \phi_1(\mathbf{r}_2)\chi_\uparrow(\sigma_2) & \ldots & \phi_1(\mathbf{r}_N)\chi_\uparrow(\sigma_N)\\ \phi_1(\mathbf{r}_1)\chi_\downarrow(\sigma_1) & \phi_1(\mathbf{r}_2)\chi_\downarrow(\sigma_2) & \ldots & \phi_1(\mathbf{r}_N)\chi_\downarrow(\sigma_N)\\ \phi_2(\mathbf{r}_1)\chi_\uparrow(\sigma_1) & \phi_2(\mathbf{r}_2)\chi_\uparrow(\sigma_2) & \ldots & \phi_2(\mathbf{r}_N)\chi_\uparrow(\sigma_N)\\ \phi_2(\mathbf{r}_1)\chi_\downarrow(\sigma_1) & \phi_2(\mathbf{r}_2)\chi_\downarrow(\sigma_2) & \ldots & \phi_2(\mathbf{r}_N)\chi_\downarrow(\sigma_N)\\ \vdots & \vdots & \ddots & \vdots\\ \phi_{N/2}(\mathbf{r}_1)\chi_\uparrow(\sigma_1) & \phi_{N/2}(\mathbf{r}_2)\chi_\uparrow(\sigma_2) & \ldots & \phi_{N/2}(\mathbf{r}_N)\chi_\uparrow(\sigma_N)\\ \phi_{N/2}(\mathbf{r}_1)\chi_\downarrow(\sigma_1) & \phi_{N/2}(\mathbf{r}_2)\chi_\downarrow(\sigma_2) & \ldots & \phi_{N/2}(\mathbf{r}_N)\chi_\downarrow(\sigma_N)\end{vmatrix}. \quad (2.30)$$

where the single-particle orbitals are now labelled starting from 1, and

$$\chi_{\uparrow/\downarrow}(\sigma)=\begin{cases}1 & \text{if } \sigma=\uparrow/\downarrow\\ 0 & \text{otherwise}\end{cases}. \quad (2.31)$$

The corresponding ground-state energy of the system is

$$E_0=2\sum_{i=1}^{N/2}\varepsilon_i. \quad (2.32)$$

We note that, for any general single-particle operator of the form

$$O(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N)=\sum_{i=1}^{N}o_i(\mathbf{x}_i) \quad (2.33)$$

(where $\mathbf{x}_i$ may refer to any set of coordinates for particle i), it is easily shown that both a simple (unsymmmetrised) product wavefunction and the Slater determinant wavefunction which uses the same set of orthonormal single-particle

[4]For an open-shell system (one with an odd number of electrons, and, hence, an unpaired spin orbital), the last row of the determinant disappears, and $\phi_{N/2}\to\phi_{(N+1)/2}$. The total energy is similarly modified.

orbitals $\{\phi_i\}$, will give the same expectation value

$$\begin{aligned}\langle\psi|\,\hat{O}\,|\psi\rangle &= \sum_{i=1}^{N}\int \phi_i^*\left(\mathbf{x}\right)o_i\left(\mathbf{x}\right)\phi_i\left(\mathbf{x}\right)\,\mathrm{d}\mathbf{x}\\ &= \sum_{i=1}^{N}\langle\phi_i|\,\hat{o}_i\,|\phi_i\rangle\,;\end{aligned} \tag{2.34}$$

this is because, in the case of the Slater determinant, all cross terms will be cancelled by the presence of factors of the form

$$\int \phi_i^*\left(\mathbf{x}\right)\phi_{j\neq i}\left(\mathbf{x}\right)\,\mathrm{d}\mathbf{x} = 0. \tag{2.35}$$

It follows that the total density $n\left(\mathbf{x}\right)$ will also be the same in both cases, since it can be expressed as the expectation value of the density operator $\hat{n}\left(\mathbf{x}\right)$, which is of the form given in Eq. 2.33:

$$n\left(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N;\mathbf{x}\right) = \sum_{i=1}^{N}\delta\left(\mathbf{x}-\mathbf{x}_i\right). \tag{2.36}$$

The total density, therefore, is simply the sum of the single-particle orbital densities:

$$\begin{aligned}n\left(\mathbf{x}\right) = \langle\psi|\,\hat{n}\left(\mathbf{x}\right)|\psi\rangle &= \sum_{i=1}^{N}\int \phi_i^*\left(\mathbf{x}'\right)\delta\left(\mathbf{x}-\mathbf{x}'\right)\phi_i\left(\mathbf{x}'\right)\,\mathrm{d}\mathbf{x}'\\ &= \sum_{i=1}^{N}\left|\phi_i\left(\mathbf{x}\right)\right|^2.\end{aligned} \tag{2.37}$$

We emphasize the fact that Eq. 2.30 is only the exact ground state for a system of non-interacting fermions; once interactions between particles are considered, this is no longer the case. However, the set of all Slater determinants that can be constructed from the complete (infinite) set of single-particle orbitals themselves form a complete, orthonormal basis in terms of which the exact many-body wavefunction for an interacting system can be expanded. This idea forms the basis of several high-accuracy 'quantum chemistry' meth-

ods, such as configuration interaction [19] (CI), Møller-Plesset perturbation theory [20] (MP), and coupled-clusters [21] (CC).

2.2 The Born-Oppenheimer approximation

What does a realistic many-body Hamiltonian look like? If we consider a system of N electrons at positions $\{\mathbf{r}_i\}$ and M nuclei at positions $\{\mathbf{R}_\alpha\}$ with atomic numbers $\{Z_\alpha\}$ and nuclear masses $\{m_\alpha\}$, the Hamiltonian in Eq. 2.17 becomes

$$\begin{aligned} H^{\text{vib}} = &\sum_{i=1}^{N} t_i + \sum_{\alpha=1}^{M} t_\alpha - \sum_{i=1}^{N}\sum_{\alpha=1}^{M} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \\ &+ \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2}\sum_{\alpha=1}^{M}\sum_{\beta\neq\alpha}^{M} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}, \end{aligned} \tag{2.38}$$

where the first two terms contain the single-particle kinetic energy operators for the electrons and nuclei, and the following three terms are the electron-nucleus, electron-electron and nucleus-nucleus Coulomb interactions (in other words, the classical electrostatic potential energy $V(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\})$). The last two terms are multiplied by a factor of $\frac{1}{2}$ to correct for double counting.

The Hamiltonian $\hat{H}^{\text{vib}}$ and its corresponding eigenvalues E^{vib} and eigenvectors $\psi^{\text{vib}}(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\})$ all relate to the combined electronic-nuclear system, and are termed vibronic. Unfortunately, the vibronic Hamiltonian contains terms connecting every particle to every other particle, including between electrons and nuclei, and therefore does not allow for an exactly separable solution of the form described in Sec. 2.1.2. In fact, the problem set by Eq. 2.38 is completely intractable for most systems if one wishes to pursue an exact solution. It is necessary, therefore, to apply a number of approximations to reduce its complexity. The first approximation that is frequently applied is that proposed by Born and Oppenheimer to separate the electronic and nuclear degrees of freedom [22].

The physical reasoning behind the Born-Oppenheimer approximation is that the nuclei, having much larger masses than the electrons (by 3–5 orders

of magnitude), but a similar momentum, must therefore move much slower than them. This can be illustrated by considering a simple two-body model of an atom (known as the Lorentz oscillator), consisting of a nucleus of mass m_n at position x_n and an electron of mass m_e at position x_e connected by a spring with spring constant β. By transforming the problem into the coordinates $X = (x_n m_n + x_e m_e) / (m_n + m_e)$ (the centre of mass of the system) and $x = x_n - x_e$ (the distance between the two particles), we obtain a Hamiltonian of the separable form given by Eq. 2.20:

$$H = -\frac{1}{2M}\frac{\partial^2}{\partial X^2} - \frac{1}{2\mu}\frac{\partial^2}{\partial x^2} + \frac{1}{2}\beta x^2, \tag{2.39}$$

where $M = m_n + m_e$ is the total mass, and $\mu = m_n m_e / (m_n + m_e)$ is the reduced mass. The solutions in X are simply plane waves (arising from the continuous translational symmetry of the problem), while those in x are given by the quantum harmonic oscillator; the ground-state wavefunction for this coordinate is therefore

$$\phi_0(x) = \left(\frac{\mu\beta}{\pi^2}\right)^{\frac{1}{8}} \mathrm{e}^{-\frac{1}{2}\sqrt{\mu\beta}x^2}, \tag{2.40}$$

and the typical relative momentum of the particles p_x in this state is given by the RMS value of the momentum operator:

$$p_x = \sqrt{\langle\phi_0|\,\hat{p}_x^2\,|\phi_0\rangle} = \left(\frac{\mu\beta}{4}\right)^{\frac{1}{4}}. \tag{2.41}$$

The momentum of the two particles must be equal and opposite (with magnitude $p_x/2$), and, hence, the typical speed of each is simply inversely proportional to its mass. Given their relative masses, this result suggests that nuclei are typically 3–5 orders of magnitude slower than electrons.

The difference in the typical electronic and nuclear time scales means that it is reasonable to assume that the electrons relax almost instantaneously to their equilibrium configuration with respect to the nuclear positions; therefore, we can simply calculate the ground state electronic structure as a function of the

nuclear coordinates first and consider nuclear dynamics separately. In order to achieve this decoupling, we first assume a separable form for the vibronic wavefunction, in which the electronic part is parametrically dependent on the nuclear positions:

$$\psi^{\text{vib}}\left(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}\right) = \psi^{\text{el}}\left(\{\mathbf{r}_i\}; \{\mathbf{R}_\alpha\}\right) \psi^{\text{nuc}}\left(\{\mathbf{R}_\alpha\}\right) \tag{2.42}$$

By applying the vibronic Hamiltonian to this separable wavefunction, we obtain the following result (for legibility, we suppress the arguments of the wavefunction and potential terms):

$$\begin{aligned} H^{\text{vib}}\psi^{\text{vib}} =& \left\{\sum_{i=1}^{N} t_i + V^{\text{en}} + V^{\text{ee}}\right\} \psi^{\text{el}}\psi^{\text{nuc}} \\ &+ \left\{\sum_{\alpha=1}^{M} t_\alpha + V^{\text{nn}}\right\} \psi^{\text{el}}\psi^{\text{nuc}} \\ =& \left\{E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right) + \sum_{\alpha=1}^{M} t_\alpha + V^{\text{nn}}\right\} \psi^{\text{el}}\psi^{\text{nuc}} \\ =& \psi^{\text{el}} \left\{E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right) + \sum_{\alpha=1}^{M} t_\alpha + V^{\text{nn}}\right\} \psi^{\text{nuc}} \\ &+ \sum_{\alpha=1}^{M} \left\{\psi^{\text{nuc}} t_\alpha \psi^{\text{el}} - \frac{1}{m_\alpha}\left(\nabla_{\mathbf{R}_\alpha}\psi^{\text{nuc}}\right) \cdot \left(\nabla_{\mathbf{R}_\alpha}\psi^{\text{el}}\right)\right\}, \end{aligned} \tag{2.43}$$

where V^{en}, V^{ee} and V^{nn} are the three potential energy terms from Eq. 2.38, and we have defined the electronic energy $E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right)$ and corresponding electronic Hamiltonian $\hat{H}^{\text{el}}$, satisfying the eigenvalue equation

$$H^{\text{el}}\psi^{\text{el}} = \left\{\sum_{i=1}^{N} t_i + V^{\text{en}} + V^{\text{ee}}\right\} \psi^{\text{el}} = E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right) \psi^{\text{el}}. \tag{2.44}$$

In other words, the electronic part of the wavefunction is required to solve a time-independent Schrödinger equation for the electronic Hamiltonian, which is effectively that of a system of N electrons in an external potential V^{en} provided by M static, point-like nuclei in positions $\{\mathbf{R}_\alpha\}$. The eigenenergy

$E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right)$ is, in fact, the *adiabatic* electronic contribution to the total vibronic energy E^{vib}. However, there is also a non-adiabatic contribution given by the two terms in the second-to-last line of Eq. 2.43 (these do not vanish because of the parametric dependence of the electronic part of the wavefunction on the nuclear positions).

The non-adiabatic contributions to the energy are generally neglected, and so the two terms are set to zero; this is known as the adiabatic approximation. The most important effect that is lost by applying this approximation is that of electron-phonon interaction, which causes the vibronic wavefunction to accumulate contributions from electronic eigenstates other than the ground state as the nuclei move. In general, therefore, the adiabatic approximation will be most reasonable for cases where the gap between electronic eigenstates is larger than the typical energies for nuclear motion, thus making the transition probabilities small [23].

Once the non-adiabatic terms are neglected, the nuclear part of the wavefunction has to satisfy a time-independent Schrödinger equation of the form

$$H^{\text{nuc}}\psi^{\text{nuc}} = \left\{ \sum_{\alpha=1}^{M} t_\alpha + V^{\text{nn}} + E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right) \right\} \psi^{\text{el}} = E^{\text{vib,adiab}}\psi^{\text{nuc}}, \tag{2.45}$$

where we have defined the nuclear Hamiltonian $\hat{H}^{\text{nuc}}$; we note that, due to the adiabatic approximation, we are now solving only for the adiabatic contribution $E^{\text{vib,adiab}}$ to the total vibronic energy. As mentioned previously, $E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right)$ is found by solving Eq. 2.44. In practice, however, a further approximation is almost always employed: the nuclei are treated as point-particles obeying the laws of classical mechanics, and the forces on them are determined by the adiabatic electronic energy landscape given by $E^{\text{el}}\left(\{\mathbf{R}_\alpha\}\right)$ and the nucleus-nucleus Coulomb interaction term V^{nn}. Using this scheme, it is now possible either to minimise the total energy of the system with respect to the atomic positions, or to apply the methods of classical molecular dynamics [24] on the nuclei moving in the potential of the quantum-mechanical electrons; this is known as *ab initio* molecular dynamics [25, 26]. However, the lightest nuclei, most notably that of hydrogen, sometimes need to be treated quantum-mechanically; this is

usually done within a path-integral formalism [27, 28].

2.3 Hartree theory

Using the Bohr-Oppenheimer approximation, we can now focus solely on solving the electronic problem given in Eq. 2.44 for a fixed configuration of classical nuclei. Although simplified, the Schrödinger equation defined by the electronic Hamiltonian is nonetheless intractable, since it does not forgo the need to deal with the electronic many-body wavefunction directly. An intuitive solution is to approximate the behaviour of the interacting electrons with non-interacting particles in an effective potential. One of the first proposed methods to do so is the Hartree approach, in which the full wavefunction is assumed to be a product of one-electron orbitals, similarly to Eq. 2.21:

$$\psi^{\mathrm{H}}\left(\{\mathbf{r}_i\}\right) = \prod_{i=1}^{N} \phi_i\left(\mathbf{r}_i\right). \tag{2.46}$$

We impose the constraint that the orbitals must be orthonormal, following the insights into non-interacting systems discussed in Sec. 2.1.2, although, being a single-product wavefunction, $\left|\psi^{\mathrm{H}}\right\rangle$ is not properly antisymmetric.

The aim of the Hartree method is to optimise this trial wavefunction using the variational principle of Sec. 2.1.1, by minimising $\left\langle\psi^{\mathrm{H}}\right| \hat{H}^{\mathrm{el}} \left|\psi^{\mathrm{H}}\right\rangle$ with respect to variations in the one-electron orbitals. The expectation value of the total energy is given by three terms; two of them, the kinetic energy $\hat{T}$ and external potential due to the static nuclei $\hat{V}^{\mathrm{ext}} \equiv \hat{V}^{\mathrm{en}}$, are one-particle operators and, therefore, straightforward to evaluate, as discussed in Sec. 2.1.2:

$$\begin{cases} \left\langle\psi^{\mathrm{H}}\right| \hat{T} \left|\psi^{\mathrm{H}}\right\rangle & = \sum_{i=1}^{N} \left\langle\phi_i\right| \hat{t} \left|\phi_i\right\rangle \\ \left\langle\psi^{\mathrm{H}}\right| \hat{V}^{\mathrm{ext}} \left|\psi^{\mathrm{H}}\right\rangle & = \sum_{i=1}^{N} \left\langle\phi_i\right| \hat{v}^{\mathrm{ext}} \left|\phi_i\right\rangle \end{cases}, \tag{2.47}$$

where

$$v^{\mathrm{ext}}\left(\mathbf{r}; \{\mathbf{R}_\alpha\}\right) = -\sum_{\alpha=1}^{M} \frac{Z_\alpha}{\left|\mathbf{r} - \mathbf{R}_\alpha\right|}. \tag{2.48}$$

The last term, $\hat{V}^{\mathrm{ee}}$, however, is a two-particle operator; care must be taken in such cases, since the expectation value is not the same for a product wavefunction and a Slater determinant. In the case of $\left|\psi^{\mathrm{H}}\right\rangle$, we obtain:

$$\begin{aligned}\left\langle\psi^{\mathrm{H}}\right| \hat{V}^{\mathrm{ee}} \left|\psi^{\mathrm{H}}\right\rangle =& \frac{1}{2} \sum_{i=1}^{N} \left\{ \int \phi_1^* \left(\mathbf{r}_1\right) \phi_1 \left(\mathbf{r}_1\right) \, \mathrm{d}\mathbf{r}_1 \times \int \phi_2^* \left(\mathbf{r}_2\right) \phi_2 \left(\mathbf{r}_2\right) \, \mathrm{d}\mathbf{r}_2 \right. \\ & \times \ldots \times \int \phi_i^* \left(\mathbf{r}_i\right) \left[\sum_{j \neq i}^{N} \int \frac{\phi_j^* \left(\mathbf{r}_j\right) \phi_j \left(\mathbf{r}_j\right)}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} \, \mathrm{d}\mathbf{r}_j \right] \phi_i \left(\mathbf{r}_i\right) \, \mathrm{d}\mathbf{r}_i \\ & \left. \times \ldots \times \int \phi_N^* \left(\mathbf{r}_N\right) \phi_N \left(\mathbf{r}_N\right) \, \mathrm{d}\mathbf{r}_N \right\} \\ =& \frac{1}{2} \sum_{i=1}^{N} \left\langle\phi_i\right| \hat{v}_i^{\mathrm{ee}} \left|\phi_i\right\rangle ,\end{aligned} \tag{2.49}$$

where

$$v_i^{\mathrm{ee}} \left(\mathbf{r}\right) = \int \frac{\sum_{j \neq i}^{N} \left|\phi_j \left(\mathbf{r}'\right)\right|^2}{\left|\mathbf{r} - \mathbf{r}'\right|} \, \mathrm{d}\mathbf{r}'. \tag{2.50}$$

The operator $\hat{v}_i^{\mathrm{ee}}$ is the sum of the Coulomb potentials of the charge densities associated with the occupied one-electron orbitals, except for orbital i (since $\left\langle\phi_i\right| \hat{v}_i^{\mathrm{ee}} \left|\phi_i\right\rangle$ is the electrostatic energy for that electron in the potential created by all other electrons). In order to simplify the problem we neglect this necessary omission, and assume that each electron feels the mean Coulomb potential due to the total charge density of the system. This mean-field approximation gives us the Hartree potential[5],

$$v^{\mathrm{H}} \left(\mathbf{r}\right) = \int \frac{n \left(\mathbf{r}'\right)}{\left|\mathbf{r} - \mathbf{r}'\right|} \, \mathrm{d}\mathbf{r}', \tag{2.51}$$

which is a functional of the total electronic density n, instead of the individual orbital densities. Substituting $\hat{v}_i^{\mathrm{ee}}$ for $\hat{v}^{\mathrm{H}}$ in Eq. 2.49 introduces an unphysical self-interaction, since every electron now feels the Coulomb potential due to its own charge density as well as that of all other electrons; however, this will

[5]However, the first calculations on many-electron systems by Hartree did not make use of the approximation [29].

only introduce a small error if the one-electron orbitals are delocalised over the entire system, as is the case in bulk materials.

We now need to find the set of orthogonal orbitals that minimise the total energy, given by

$$E^{\mathrm{H}}\left[\{\phi_i\}\right] = \sum_{i=1}^{N} \langle\phi_i|\, \hat{t}\, |\phi_i\rangle + \sum_{i=1}^{N} \langle\phi_i|\, \hat{v}^{\mathrm{ext}}\, |\phi_i\rangle + \frac{1}{2}\sum_{i=1}^{N} \langle\phi_i|\, \hat{v}^{\mathrm{H}}\, |\phi_i\rangle . \tag{2.52}$$

To do so, we introduce a Lagrange multiplier λ_i for each of the N normalisation constraints:

$$F^{\mathrm{H}}\left[\{\phi_i\}\right] = E^{\mathrm{H}}\left[\{\phi_i\}\right] - \sum_{i=1}^{N} \lambda_i \langle\phi_i|\phi_i\rangle ; \tag{2.53}$$

there is no need to include Lagrange multipliers for the orthogonality constraints, as these will naturally be satisfied by our solution.

Let us consider the first-order change in our auxiliary function due to a small variation $\delta\phi_n$ in the orbital ϕ_n, which must vanish at stationary points of the functional:

$$\begin{aligned}
\delta F^{\mathrm{H}}\left[\phi_n\right] =& F^{\mathrm{H}}\left[\phi_n + \delta\phi_n\right] - F^{\mathrm{H}}\left[\phi_n\right] \\
=& F^{\mathrm{H,bra}}\left[\delta\phi_n\right] + F^{\mathrm{H,ket}}\left[\delta\phi_n\right] + \mathcal{O}\left(\delta\phi_n^2\right) \\
=& F^{\mathrm{H,bra}}\left[\Re(\delta\phi_n)\right] - \mathrm{i}F^{\mathrm{H,bra}}\left[\Im(\delta\phi_n)\right] + \\
& F^{\mathrm{H,ket}}\left[\Re(\delta\phi_n)\right] + \mathrm{i}F^{\mathrm{H,ket}}\left[\Im(\delta\phi_n)\right] + \mathcal{O}\left(\delta\phi_n^2\right) = 0,
\end{aligned} \tag{2.54}$$

where $F^{\mathrm{H,bra}}\left[\delta\phi_n\right]$ and $F^{\mathrm{H,ket}}\left[\delta\phi_n\right]$ are used to divide the first-order contributions into terms in which the variation is contained in a bra ($\langle\delta\phi_n|$) and ones in which it is contained in a ket ($|\delta\phi_n\rangle$). However, the auxiliary function must also be stationary with respect to a small variation $\delta\phi_n^*$; in this case, we obtain:

$$\begin{aligned}
\delta F^{\mathrm{H}}\left[\phi_n\right] =& F^{\mathrm{H}}\left[\phi_n + \delta\phi_n^*\right] - F^{\mathrm{H}}\left[\phi_n\right] \\
=& F^{\mathrm{H,bra}}\left[\delta\phi_n^*\right] + F^{\mathrm{H,ket}}\left[\delta\phi_n^*\right] + \mathcal{O}\left(\delta\phi_n^2\right) \\
=& F^{\mathrm{H,bra}}\left[\Re(\delta\phi_n)\right] + \mathrm{i}F^{\mathrm{H,bra}}\left[\Im(\delta\phi_n)\right] + \\
& F^{\mathrm{H,ket}}\left[\Re(\delta\phi_n)\right] - \mathrm{i}F^{\mathrm{H,ket}}\left[\Im(\delta\phi_n)\right] + \mathcal{O}\left(\delta\phi_n^2\right) = 0.
\end{aligned} \tag{2.55}$$

By adding and subtracting the last line of Eq. 2.54 with that of Eq. 2.55,

it is straightforward to show that both $F^{\mathrm{H,bra}}[\delta\phi_n]$ and $F^{\mathrm{H,ket}}[\delta\phi_n]$ must be separately equal to zero at all stationary points of the functional (in other words, the two conditions are equivalent); hence, it is only necessary to consider one of them to derive the correct solution. Choosing the first term, we obtain:

$$\begin{aligned} F^{\mathrm{H,bra}}[\delta\phi_n] = {} & \langle\delta\phi_n|\,\hat{t}\,|\phi_n\rangle + \langle\delta\phi_n|\,\hat{v}^{\mathrm{ext}}\,|\phi_n\rangle + \frac{1}{2}\langle\delta\phi_n|\,\hat{v}^{\mathrm{H}}\,|\phi_n\rangle \\ & + \frac{1}{2}\sum_{i=1}^{N}\left\langle\phi_i\left|\int\frac{\langle\delta\phi_n|\,\hat{n}(\mathbf{r}')\,|\phi_n\rangle}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}'\right|\phi_i\right\rangle \\ & - \lambda_n\langle\delta\phi_n|\phi_n\rangle \end{aligned} \tag{2.56}$$

However, the fourth term in this expression can be written as

$$\frac{1}{2}\sum_{i=1}^{N}\iint\frac{|\phi_i(\mathbf{r})|^2\,\delta\phi_n^*(\mathbf{r}')\,\phi_n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}'\,\mathrm{d}\mathbf{r} = \frac{1}{2}\langle\delta\phi_n|\,\hat{v}^{\mathrm{H}}\,|\phi_n\rangle\,, \tag{2.57}$$

and so is simply equivalent to the third term. Therefore, the factor of $\frac{1}{2}$ for the Hartree potential term is cancelled, giving us the final condition

$$\langle\delta\phi_n|\,\hat{t} + \hat{v}^{\mathrm{ext}} + \hat{v}^{\mathrm{H}} - \lambda_n\,|\phi_n\rangle = 0. \tag{2.58}$$

Eq. 2.58 is satisfied for an arbitrary variation $\delta\phi_n$ by solving an eigenvalue equation of the form

$$\hat{H}^{\mathrm{H}}\,|\phi\rangle = \varepsilon^{\mathrm{H}}\,|\phi\rangle\,, \tag{2.59}$$

where

$$\hat{H}^{\mathrm{H}} = \hat{t} + \hat{v}^{\mathrm{ext}} + \hat{v}^{\mathrm{H}} \tag{2.60}$$

is the effective mean-field Hamiltonian for a system of non-interacting particles, and the Lagrange multipliers become the eigenvalues $\{\varepsilon_i^{\mathrm{H}}\}$. The total energy is minimised, therefore, by selecting the orbitals corresponding to the lowest $N/2$ eigenvalues (due to spin, as explained in Sec. 2.1.2). In other words, the occupied set of orbitals can be used to construct a Slater determinant which is the exact antisymmetric ground-state solution for the approximate non-interacting Hamiltonian. We can now define a non-interacting energy,

corresponding to the sum of the eigenvalues for the occupied orbitals:

$$E_0^{\mathrm{H,non-int}} = 2\sum_{i=1}^{N/2} \varepsilon_i^{\mathrm{H}}. \tag{2.61}$$

The interacting energy, by comparison with Eq. 2.52, is

$$E_0^{\mathrm{H}} = E_0^{\mathrm{H,non-int}} - \frac{1}{2}\int n\left(\mathbf{r}\right) v^{\mathrm{H}}\left(\mathbf{r}\right)\,\mathrm{d}\mathbf{r}; \tag{2.62}$$

the last term restores the double-counting correction for the Hartree potential that is not present in the non-interacting system. It should be noted that, when implementing the Hartree method in practice, Eq. 2.59 needs to be solved self-consistently to ensure that the input and output densities agree, since the Hartree potential (and, hence, the Hamiltonian) depends on the electronic density obtained from the occupied orbitals that are the solutions to the equation. This is generally achieved by starting from a trial input density and using an iterative procedure to arrive at a self-consistent solution.

2.4 Hartree-Fock theory

The two major flaws with the Hartree method, as presented in the previous section, are the simple product form of the many-electron wavefunction used to minimise the energy of the interacting Hamiltonian, and the spurious Coulomb self-interaction of the electrons with their own charge density present in the Hartree potential. Both of these problems are resolved by the extension to the theory proposed by Fock [30], resulting in the popular Hartree-Fock method. The derivation of the method follows the same steps as those already outlined for the Hartree method, with the difference that the trial wavefunction is now taken to be a Slater determinant, thus satisfying the fermionic symmetry constraints:

$$\psi^{\mathrm{HF}}\left(\{\mathbf{r}_i\}\right) = \frac{1}{\sqrt{N!}}\sum_{\mathcal{P}=1}^{N!}\left(-1\right)^{\xi_{\mathcal{P}}}\prod_{i=1}^{N}\phi_{j_{i,\mathcal{P}}}\left(\mathbf{r}_i\right), \tag{2.63}$$

where the sum is over the $N!$ permutations of the particle indices $\{j_{i,\mathcal{P}}\}$. The expectation value of $\hat{H}^{\mathrm{el}}$ differs with respect to the product wavefunction only for the electron-electron Coulomb interaction term, $\hat{V}^{\mathrm{ee}}$, since it is a two-particle operator. For the Hartree-Fock trial wavefunction, this becomes:

$$\begin{aligned}
&\left\langle \psi^{\mathrm{HF}} \right| \hat{V}^{\mathrm{ee}} \left| \psi^{\mathrm{HF}} \right\rangle \\
=&\frac{1}{2\,(N!)} \sum_{\mathcal{P}=1}^{N!} \sum_{\mathcal{Q}=1}^{N!} (-1)^{\xi_{\mathcal{P}}} (-1)^{\xi_{\mathcal{Q}}} \\
&\times \sum_{i=1}^{N} \left\{ \int \phi^*_{j_{1,\mathcal{P}}}(\mathbf{r}_1)\, \phi_{j_{1,\mathcal{Q}}}(\mathbf{r}_1)\, \mathrm{d}\mathbf{r}_1 \times \int \phi^*_{j_{2,\mathcal{P}}}(\mathbf{r}_2)\, \phi_{j_{2,\mathcal{Q}}}(\mathbf{r}_2)\, \mathrm{d}\mathbf{r}_2 \right. \\
&\times \ldots \times \int \phi^*_{j_{i,\mathcal{P}}}(\mathbf{r}_i) \left[\sum_{k\neq i}^{N} \int \frac{\phi^*_{j_{k,\mathcal{P}}}(\mathbf{r}_k)\, \phi_{j_{k,\mathcal{Q}}}(\mathbf{r}_k)}{|\mathbf{r}_i - \mathbf{r}_k|} \mathrm{d}\mathbf{r}_k \right] \phi_{j_{i,\mathcal{Q}}}(\mathbf{r}_i)\, \mathrm{d}\mathbf{r}_i \\
&\left. \times \ldots \times \int \phi^*_{j_{N,\mathcal{P}}}(\mathbf{r}_N)\, \phi_{j_{N,\mathcal{Q}}}(\mathbf{r}_N)\, \mathrm{d}\mathbf{r}_N \right\}.
\end{aligned} \tag{2.64}$$

Due to the orthogonality of the orbitals, the product of integrals in the curly brackets will equal zero for all permutations except those for which $\{j_{l,\mathcal{P}} = j_{l,\mathcal{Q}}\}$ for all values of $l \neq i, k$. Therefore, there are only two types of permutations that contribute to the final result: those for which $\{j_{i,\mathcal{P}} = j_{i,\mathcal{Q}}, j_{k,\mathcal{P}} = j_{k,\mathcal{Q}}\}$, and those for which $\{j_{i,\mathcal{P}} = j_{k,\mathcal{Q}}, j_{k,\mathcal{P}} = j_{i,\mathcal{Q}}\}$; by diving these two cases into separate terms, we obtain:

$$\begin{aligned}
&\frac{1}{2\,(N!)} \sum_{\mathcal{P}=1}^{N!} \sum_{i=1}^{N} \int \phi^*_i(\mathbf{r}_i) \left[\sum_{k\neq i}^{N} \int \frac{\phi^*_k(\mathbf{r}_k)\, \phi_k(\mathbf{r}_k)}{|\mathbf{r}_i - \mathbf{r}_k|} \mathrm{d}\mathbf{r}_k \right] \phi_i(\mathbf{r}_i)\, \mathrm{d}\mathbf{r}_i \\
&- \frac{1}{2\,(N!)} \sum_{\mathcal{P}=1}^{N!} \sum_{i=1}^{N} \sum_{k\neq i}^{N} \int\!\!\int \frac{\phi^*_i(\mathbf{r}_i)\, \phi^*_k(\mathbf{r}_k)\, \phi_i(\mathbf{r}_k)\, \phi_k(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_k|} \mathrm{d}\mathbf{r}_k\, \mathrm{d}\mathbf{r}_i \\
=&\frac{1}{2} \sum_{i=1}^{N} \langle \phi_i | \hat{v}^{\mathrm{H}} | \phi_i \rangle - \frac{1}{2} \sum_{i=1}^{N} \langle \phi_i | \hat{k}^{\mathrm{x}} | \phi_i \rangle ,
\end{aligned} \tag{2.65}$$

where we have defined a non-local operator, known as the exchange operator:

$$\langle \mathbf{r} | \hat{k}^{\mathrm{x}} | \phi_i \rangle = \sum_{k=1}^{N} \int \frac{\phi^*_k(\mathbf{r}')\, \phi_i(\mathbf{r}')\, \phi_k(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}'. \tag{2.66}$$

We note that the self-interaction present in the Hartree potential is exactly cancelled by the equivalent terms in the exchange operator, due to the difference in sign. Furthermore, the expectation value of the exchange operator must be negative, and so the total energy will always be lower than that of the equivalent product wavefunction.

The minimisation of the total energy with respect to the set of one-electron orbitals can be derived in a similar fashion to that detailed in the previous section for the Hartree method; the final condition obtained by using the Hartree-Fock trial wavefunction is:

$$F^{\mathrm{HF,bra}}\left[\delta\phi_n\right] = \langle\delta\phi_n|\, \hat{t} + \hat{v}^{\mathrm{ext}} + \hat{v}^{\mathrm{H}} - \hat{k}^{\mathrm{x}} - \lambda_n \left|\phi_n\right\rangle = 0. \tag{2.67}$$

As before, Eq. 2.67 is satisfied for an arbitrary variation $\delta\phi_n$ by solving an eigenvalue equation of the form

$$\hat{H}^{\mathrm{HF}}\left|\phi\right\rangle = \varepsilon^{\mathrm{HF}}\left|\phi\right\rangle, \tag{2.68}$$

where

$$\hat{H}^{\mathrm{HF}} = \hat{H}^{\mathrm{H}} - \hat{k}^{\mathrm{x}}. \tag{2.69}$$

The non-interacting energy $E^{\mathrm{HF,non-int}}$ is defined analogously to $E^{\mathrm{H,non-int}}$, and the interacting energy now includes an additional double-counting correction, for the exchange operator:

$$\begin{aligned} E^{\mathrm{HF}} =& E^{\mathrm{HF,non-int}} - \frac{1}{2}\int n\left(\mathbf{r}\right) v^{\mathrm{H}}\left(\mathbf{r}\right)\,\mathrm{d}\mathbf{r} \\ &+ \sum_{i=1}^{N/2}\sum_{k=1}^{N/2} \iint \frac{\phi_i^*\left(\mathbf{r}\right)\phi_k^*\left(\mathbf{r}'\right)\phi_i\left(\mathbf{r}'\right)\phi_k\left(\mathbf{r}\right)}{\left|\mathbf{r}-\mathbf{r}'\right|}\,\mathrm{d}\mathbf{r}'\,\mathrm{d}\mathbf{r}. \end{aligned} \tag{2.70}$$

When taking into consideration the spin part of the one-electron orbitals, the exchange operator and the exchange double-counting term only contains half the number of terms that might be expected; this is because terms between orbitals of opposite spin are excluded, due the swapping of particle labels. For

example[6]:

$$\begin{aligned}\langle\phi_i\chi_\uparrow|\,\hat{k}^{\mathrm{x}}_{\phi_j\chi_\downarrow}\,|\phi_i\chi_\uparrow\rangle = \langle\phi_i|\,\hat{k}^{\mathrm{x}}_{\phi_j}\,|\phi_i\rangle\,\{&\chi_\uparrow(\uparrow)\,\chi_\downarrow(\uparrow)\,\chi_\uparrow(\uparrow)\,\chi_\downarrow(\uparrow)\\ &+\chi_\uparrow(\uparrow)\,\chi_\downarrow(\downarrow)\,\chi_\uparrow(\downarrow)\,\chi_\downarrow(\uparrow)\\ &+\chi_\uparrow(\downarrow)\,\chi_\downarrow(\uparrow)\,\chi_\uparrow(\uparrow)\,\chi_\downarrow(\downarrow)\\ &+\chi_\uparrow(\downarrow)\,\chi_\downarrow(\downarrow)\,\chi_\uparrow(\downarrow)\,\chi_\downarrow(\downarrow)\} = 0,\end{aligned} \tag{2.71}$$

since every term contains a factor $\chi_{\uparrow/\downarrow}(\downarrow/\uparrow) = 0$.

The determinantal form of the trial wavefunction ensures that the Hartree-Fock method respects the symmetry constraints for particle exchange of the Pauli exclusion principle; because of this, the difference between the Hartree and Hartree-Fock energies for a particular system is called the exchange energy. However, the accuracy of the Hartree-Fock solution is limited by the use of a single Slater determinant, which neglects all dynamical (non-mean-field) interactions between electrons other than exchange; the remaining discrepancy between the Hartree-Fock energy and the exact solution is called the correlation energy. Components of the correlation energy can be included in the calculation by using a multi-determinantal trial wavefunction; the exact solution, referred to as the full CI limit, requires an infinite number of determinants (for a complete set of single-particle orbitals). Alternatively, both the exchange and correlation components of the total energy can be approximated by a different method, that of density-functional theory, described in the next section.

Despite the complete neglect of correlation, Hartree-Fock has been shown to give accurate results in certain situations, most notably in calculations of bond lengths and angles for small molecules and atomic clusters; unfortunately, vibrational frequencies are typically overestimated, and binding energies are severely underestimated [31]. For solids, it provides a reasonable description of wide-gap insulators, although the band gap is overestimated [32], and is known to have several important problems in describing metals, mainly due to the fact that electronic screening effects are not well represented [33, 34].

---

[6] $\hat{k}^{\mathrm{x}}_{\phi_j}\,|\phi_i\rangle$ refers to the j-th term of the sum in Eq. 2.66.

2.5 Density-functional theory

2.5.1 The two theorems

Density-functional theory provides an alternative approach to solving the many-electron Schrödinger equation of Eq. 2.44. The central idea of DFT is to make use of an exact reformulation of the problem that features the three-dimensional electronic charge density as the fundamental variable instead of the $3N$-dimensional many-electron wavefunction, thus greatly reducing its complexity. This reformulation is based on two theorems by Hohenberg and Kohn [2]:

Theorem 1 *For a system of N interacting electrons in an external potential $v^{\mathrm{ext}}(\mathbf{r})$, this potential is uniquely determined (up to an additive constant) by the electronic ground-state density $n_0(\mathbf{r})$.*

Theorem 2 *It is possible to define an energy functional $E[n]$ of the electron density for a given $v^{\mathrm{ext}}(\mathbf{r})$ which is minimised and equal to the ground-state energy when $n(\mathbf{r}) = n_0(\mathbf{r})$.*

The first theorem can be proved by contradiction, by considering two external potentials $v^{\mathrm{ext}}(\mathbf{r})$ and $v^{\mathrm{ext}\prime}(\mathbf{r})$ (and corresponding Hamiltonians $\hat{H}$ and $\hat{H}'$) which differ by more than an additive constant and lead to non-degenerate ground states $|\psi_0\rangle$ and $|\psi_0'\rangle$ with ground-state energies E_0 and E_0'. By the variational principle[7],

$$E_0 = \langle\psi_0|\,\hat{H}\,|\psi_0\rangle < \langle\psi_0'|\,\hat{H}\,|\psi_0'\rangle\,. \tag{2.72}$$

We also note that

$$\langle\psi_0'|\,\hat{H}\,|\psi_0'\rangle = \langle\psi_0'|\,\hat{H}'\,|\psi_0'\rangle - \langle\psi_0'|\,\hat{H}' - \hat{H}\,|\psi_0'\rangle\,. \tag{2.73}$$

Since we can divide the Hamiltonian as $\hat{H} = \hat{F} + \hat{V}$, where $\hat{F}$ is the sum of the kinetic energy operator and the electron-electron Coulomb interaction and $\hat{V}$

[7] $|\psi_0\rangle$ and $|\psi_0'\rangle$ must be different, since otherwise we would find that $\left(\hat{v}^{\mathrm{ext}} - \hat{v}^{\mathrm{ext}\prime}\right)|\psi_0\rangle = (E_0 - E_0')\,|\psi_0\rangle$, and, hence, that the two potentials differ only by an additive constant, contrary to our initial assumption.

is the external potential, we obtain:

$$E_0 < E_0' - \langle\psi_0'|\,\hat{V}' - \hat{V}\,|\psi_0'\rangle = E_0' - \int n_0'(\mathbf{r})\left\{v^{\mathrm{ext}'}(\mathbf{r}) - v^{\mathrm{ext}}(\mathbf{r})\right\}\,\mathrm{d}\mathbf{r}; \quad (2.74)$$

similarly,

$$E_0' < E_0 - \langle\psi_0|\,\hat{V} - \hat{V}'\,|\psi_0\rangle = E_0 - \int n_0(\mathbf{r})\left\{v^{\mathrm{ext}}(\mathbf{r}) - v^{\mathrm{ext}'}(\mathbf{r})\right\}\,\mathrm{d}\mathbf{r}. \quad (2.75)$$

If we add these two inequalities together and assume that $n_0'(\mathbf{r}) = n_0(\mathbf{r})$, we arrive at the obviously contradictory statement

$$E_0 + E_0' < E_0 + E_0'; \quad (2.76)$$

thus, there is a one-to-one mapping between the external potential and the ground-state density of the system.

In order to prove the second theorem, we note that $n_0(\mathbf{r})$ determines both the number of electrons N and the external potential $v^{\mathrm{ext}}(\mathbf{r})$ of the Schrödinger equation of which it is a solution, and thereby all other ground-state properties of the system. Furthermore, there is also a one-to-one mapping between $n_0(\mathbf{r})$ and $|\psi_0\rangle$. Therefore, it is possible to define the following energy functional of the density, valid only for densities that are themselves ground-state solutions of the Schrödinger equation for a given external potential (known as v-representable):

$$E[n] = F[n] + V[n]\,. \quad (2.77)$$

This must be the case, since

$$F[n] = \langle\psi|\,\hat{F}\,|\psi\rangle \quad (2.78)$$

depends solely on N and $|\psi\rangle$ (the ground-state wavefunction associated with the v-representable density $n(\mathbf{r})$), and

$$V[n] = \langle\psi|\,\hat{V}\,|\psi\rangle = \int n(\mathbf{r})\,v^{\mathrm{ext}}(\mathbf{r})\,\mathrm{d}\mathbf{r} \quad (2.79)$$

depends solely on $n(\mathbf{r})$ and $v_{\text{ext}}(\mathbf{r})$. We can now see that the expectation value of the Hamiltonian for a wavefunction $|\psi\rangle$ is equal to the energy functional for the corresponding density $n(\mathbf{r})$, and so the variational principle ensures that, out of the set of all v-representable densities, the one mapped to the particular $v_{\text{ext}}(\mathbf{r})$ of interest (i.e., the ground-state density of the Schrödinger equation with that external potential) must minimise the functional and give the correct ground-state energy.

Unfortunately, this proof is not general enough to justify a practical variational scheme to find the correct ground-state energy by minimising the functional with respect to trial densities, as it is known that not all well-behaved densities are v-representable (in fact, various counter-examples have been found [35–38]). However, the constrained search approach of Levy [35,39] and Lieb [36] extends the proof to include any N-representable density, i.e., one that can be derived from a valid N-electron wavefunction, as well as applying also for systems with degenerate ground states. The constraint of N-representability is very weak, since it is known to be satisfied by any non-negative differentiable function $n(\mathbf{r})$ for which $\int n(\mathbf{r})\, \mathrm{d}\mathbf{r} = N$ [40]. The variational approach, therefore, can indeed be justified, and is considered in the next section.

2.5.2 The Kohn-Sham formalism

In order to minimise the Hohenberg-Kohn energy functional, and so find the ground-state density and energy, it is necessary to know the form of $F[n]$ (which is a *universal* functional, since it is the same for all N-electron systems). Although approximations to $F[n]$ based on the density have been proposed (including the precursor to DFT, Thomas-Fermi theory [41,42], and its extension by Dirac [43]), a more successful approach is that of Kohn and Sham [3] (KS). Their method is based on the realisation that the system of interacting electrons can be replaced by an auxiliary system of non-interacting fermions in a modified effective potential which, when solved using a Schrödinger-like equation, gives the same ground-state density.

This can be shown by considering the variational problem for the Hohenberg-

Kohn functional and applying the normalisation constraint on the density via a Lagrange multiplier:

$$\delta \left\{ F[n] + \int n(\mathbf{r}) v^{\mathrm{ext}}(\mathbf{r}) \, \mathrm{d}\mathbf{r} - \mu \left[\int n(\mathbf{r}) \, \mathrm{d}\mathbf{r} - N \right] \right\} = 0; \tag{2.80}$$

this gives us the Euler-Lagrange equation

$$\frac{\delta F[n]}{\delta n(\mathbf{r})} + v^{\mathrm{ext}}(\mathbf{r}) = \mu. \tag{2.81}$$

Following the Kohn-Sham formalism, we divide $F[n]$ into three terms:

$$F[n] = T^{\mathrm{non-int}}[n] + \frac{1}{2} \int n(\mathbf{r}) v^{\mathrm{H}}(\mathbf{r}) \, \mathrm{d}\mathbf{r} + E^{\mathrm{xc}}[n]; \tag{2.82}$$

the first term is the kinetic energy of a system of non-interacting fermions with the same density as that of the interacting electrons, the second term is the classical electrostatic energy (written in terms of the Hartree potential), and the third term, known as the exchange-correlation energy, accounts for the 'remaining' energy (i.e., the difference between the interacting and non-interacting kinetic energies, and the non-classical parts of the electron-electron interaction energy). Eq. 2.81 now becomes

$$\frac{\delta T^{\mathrm{non-int}}[n]}{\delta n(\mathbf{r})} + v^{\mathrm{KS}}(\mathbf{r}) = \mu; \tag{2.83}$$

the last term is the Kohn-Sham potential:

$$v^{\mathrm{KS}}(\mathbf{r}) = v^{\mathrm{ext}}(\mathbf{r}) + v^{\mathrm{H}}(\mathbf{r}) + v^{\mathrm{xc}}(\mathbf{r}), \tag{2.84}$$

where

$$v^{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E^{\mathrm{xc}}[n]}{\delta n(\mathbf{r})}. \tag{2.85}$$

Eq. 2.83 is equivalent to the Euler-Lagrange equation for a system of non-interacting fermions (for which $F[n] = T^{\mathrm{non-int}}[n]$) in an external potential $v^{\mathrm{KS}}(\mathbf{r})$. To find the true interacting ground-state density, therefore, we simply need to solve the corresponding non-interacting problem, as discussed in

Sec. 2.1.2. This leads to the Kohn-Sham eigenvalue equation

$$\hat{H}^{\mathrm{KS}} \left|\phi\right\rangle = \varepsilon^{\mathrm{KS}} \left|\phi\right\rangle , \tag{2.86}$$

where the Kohn-Sham Hamiltonian is given by

$$\hat{H}^{\mathrm{KS}} = \hat{t} + \hat{v}^{\mathrm{ext}} + \hat{v}^{\mathrm{H}} + \hat{v}^{\mathrm{xc}}. \tag{2.87}$$

The ground-state density is obtained by summing the single-particle densities of the lowest $N/2$ orbitals, counting each orbital twice to account for spin degeneracy, as in the Hartree and Hartree-Fock methods. The total energy of the non-interacting system is, therefore,

$$E_0^{\mathrm{KS}} = 2 \sum_{i=1}^{N/2} \varepsilon_i^{\mathrm{KS}}, \tag{2.88}$$

and that of the interacting system is

$$E_0 = E_0^{\mathrm{KS}} - \frac{1}{2} \int n\left(\mathbf{r}\right) v^{\mathrm{H}}\left(\mathbf{r}\right) \, \mathrm{d}\mathbf{r} - \int n\left(\mathbf{r}\right) v^{\mathrm{xc}}\left(\mathbf{r}\right) \, \mathrm{d}\mathbf{r} + E^{\mathrm{xc}}\left[n\right]. \tag{2.89}$$

The second term in this expression is the double-counting correction for the Hartree potential, and the last two terms make up the over-counting correction for the exchange-correlation potential.

2.5.3 Exchange and correlation

The Kohn-Sham formalism partitions the complicated many-body effects into a single unknown term, the exchange-correlation functional $E^{\mathrm{xc}}\left[n\right]$, thereby separating them from the simple known quantities associated with the auxiliary system of non-interacting fermions. The success of KS-DFT is based on the fact that the effect of this unknown functional can be approximated very accurately with simple models, especially for the case of bulk solids, even though the true functional itself is non-analytic and clearly hugely complex in its details.

The approximate exchange-correlation functional originally proposed by Kohn and Sham is the local density approximation (LDA); this is still one

of the most used today due to its simplicity and effectiveness. The LDA assumes that the contribution to the exchange-correlation energy from each point in space is equal to what it would be for a homogeneous electron gas with the same density throughout the whole system as is found at that point; therefore, the functional is local since it only depends on the density at each point independently of the rest of the system. The energy functional is defined as

$$E^{\mathrm{xc}}\left[n\right] = \int n\left(\mathbf{r}\right)\varepsilon^{\mathrm{xc,hom}}\left(n\left(\mathbf{r}\right)\right)\,\mathrm{d}\mathbf{r}, \tag{2.90}$$

where $\varepsilon^{\mathrm{xc,hom}}\left(n\left(\mathbf{r}\right)\right)$ is the exchange-correlation energy per electron in a homogeneous electron gas of density $n\left(\mathbf{r}\right)$; the corresponding exchange-correlation potential is

$$v^{\mathrm{xc}}\left(\mathbf{r}\right) = \frac{\delta E^{\mathrm{xc}}\left[n\right]}{\delta n\left(\mathbf{r}\right)} = \varepsilon^{\mathrm{xc,hom}}\left(n\left(\mathbf{r}\right)\right) + n\left(\mathbf{r}\right)\left.\frac{\mathrm{d}\varepsilon^{\mathrm{xc,hom}}\left(n\right)}{\mathrm{d}n}\right|_{n=n\left(\mathbf{r}\right)}. \tag{2.91}$$

$\varepsilon^{\mathrm{xc,hom}}\left(n\left(\mathbf{r}\right)\right)$ has been calculated analytically by in the high density limit by Gell-Mann and Brueckner [44], and numerically by Ceperley and Alder [45] using a stochastic quantum Monte Carlo method [46]; a parametrisation of this data by Perdew and Zunger [47] is usually used for DFT calculations.

It is also possible to consider the more general case of a spin-polarised system; this is done by separating the total density into the individual contributions from the two spin states, $n_{\uparrow}\left(\mathbf{r}\right)$ and $n_{\downarrow}\left(\mathbf{r}\right)$, and defining the exchange-correlation energy as a functional of both quantities[8]. In practice, it is usually more convenient to use the total electronic density

$$n\left(\mathbf{r}\right) = n_{\uparrow}\left(\mathbf{r}\right) + n_{\downarrow}\left(\mathbf{r}\right) \tag{2.92}$$

and the spin polarisation (or magnetisation) density

$$\zeta\left(\mathbf{r}\right) = n_{\uparrow}\left(\mathbf{r}\right) - n_{\downarrow}\left(\mathbf{r}\right). \tag{2.93}$$

[8]The most general case is that of non-collinear magnetism, in which the spin axis is allowed to vary in space; this requires the full 2×2 spinorial representation of the spin density matrix $n_{\alpha\beta}\left(\mathbf{r}\right)$, where $\alpha,\beta = \uparrow / \downarrow$. The collinear magnetism approximation, therefore, neglects the two off-diagonal terms of the matrix.

The extension of the LDA to spin-polarised systems is known as the local spin density approximation [48] (LSDA), for which $\varepsilon^{\mathrm{xc,hom}}\left(n\left(\mathbf{r}\right)\right)$ in Eq. 2.90 is replaced by $\varepsilon^{\mathrm{xc,hom}}\left(n\left(\mathbf{r}\right),\zeta\left(\mathbf{r}\right)\right)$. We note that if $n_{\uparrow}\left(\mathbf{r}\right)=n_{\downarrow}\left(\mathbf{r}\right)$, non-spin-polarised LDA is recovered.

It has been shown that the L(S)DA correctly reproduces certain known analytical properties of the exact exchange-correlation functional; this may partly account for its accuracy even in very inhomogeneous systems [49]. More advanced functionals based on the generalised gradient approximation [50, 51] (GGA), which depend on both the local density and density gradient, only show improvement on the L(S)DA when they are constructed to similarly enforce some of the physical properties of the exact functional [23, 34].

DFT with the L(S)DA has been shown to give accurate results in a variety of systems, most notably solids with strong covalent, ionic or metallic bonding; in fact, unlike Hartree-Fock theory, it describes both metals and insulators to a similar level of accuracy. Bond lengths are typically reproduced to within 1%, although they are usually underestimated; elastic constants and vibrational frequencies are less precise, and are also usually underestimated, while binding energies are usually overestimated [23, 49, 52]. Despite their successes, however, local and semi-local functionals completely fail to reproduce certain features that are crucial in correctly describing some systems, such as long-ranged dispersion interactions and strong on-site Hubbard-like correlations; these features can be partly corrected for by using more sophisticated DFT-based approaches [53, 54].

Chapter 3

Computational methods for DFT

In the previous chapter we described how the fully quantum-mechanical many-body problem for realistic materials (encapsulated in the Hamiltonian given in Eq. 2.38) can be greatly simplified by taking the following steps: (i) separating the electronic and nuclear degrees of freedom, and treating the nuclei as classical particles in a fixed configuration, (ii) reformulating (in principle exactly) the quantum-mechanical electronic problem in terms of the total density instead of the many-body wavefunction using DFT, and finally (iii) approximating the unknown DFT exchange-correlation functional with a simple local or semi-local functional. Taking these steps results in having to solve a single-particle Schrödinger equation with the effective Hamiltonian given in Eq. 2.87, which depends on the external potential from the fixed classical nuclei and the total electronic charge density; as noted previously for Hartree theory, the fact that the Hamiltonian is density-dependent means that the eigenvalue equation needs to be solved self-consistently.

In this chapter we outline some important methods used to find computational solutions to the Kohn-Sham equation in practical calculations; unfortunately, these will necessarily add either further physical approximations to the description of the system, as in the case of pseudopotentials, or numerical approximations to the calculations, as for the use of finite basis sets and discrete

reciprocal-space sampling in periodic systems.

3.1 The pseudopotential approximation

One of the main problems encountered when solving the Schrödinger equation numerically for a nuclear potential is that this potential, of the form $-\frac{1}{r}$, is not small, and so it is not possible to treat it as a perturbation on the free electron problem, as is done in the nearly-free electron model[1] [55]. This is reflected in the atomic orbitals, which decay exponentially as $r \to \infty$ and oscillate rapidly close to the nucleus in order to maintain orthogonality with each other.

The fundamental idea behind the pseudopotential approximation is to 'freeze' the tightly bound core electrons of each atom and replace them with an effective (or pseudo) potential which attenuates the strong Coulomb potential close to the nucleus. This is justified by the fact that the contribution of the core electrons to the total energy, although large, is only very weakly affected by the atom's chemical environment, and so does not play a significant role in its bonding properties; similarly, it is not unreasonable to assume that the core wavefunctions are unaffected when isolated atoms are brought together.

The main advantage of removing the core electrons is to relax the orthogonality constraint they pose on the wavefunctions of the remaining valence electrons; in fact, these are no longer required to be orthogonal to the core states (although they must still maintain orthogonality with each other), and so can reduce the number of nodes in their radial parts, with the lowest valence state for each angular momentum channel being nodeless. Therefore, the smooth pseudopotential results in the true valence wavefunctions being replaced by pseudised wavefunctions, which vary smoothly in the core region and only match the true ones outside the core radius r_c (see Fig. 3.1). The advantage of dealing with a smooth function is obviously a computational one, since this function will be less demanding to approximate accurately with either a truncated basis expansion or a real-space grid representation.

[1] It might seem strange, therefore, that the nearly-free electron model gives a reasonable description of many elemental metallic solids; pseudopotential theory, however, provides an indication as to why this is the case.

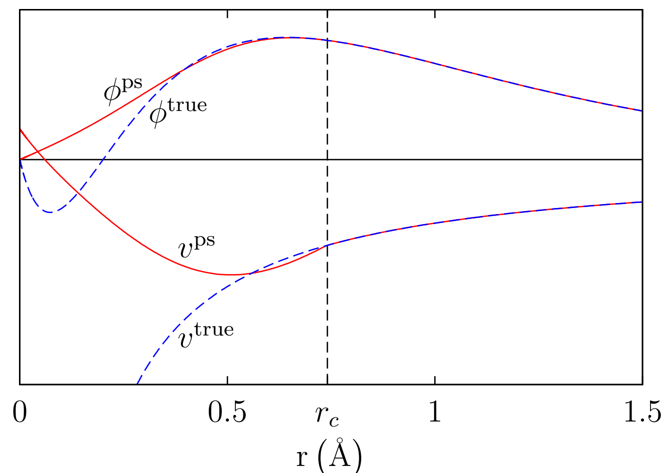


Figure 3.1: An example of the relationship between true and pseudised (ps) wavefunctions (ϕ) and potentials (v). The pseudised functions are smoother than their true counterparts for $r < r_c$, and match them exactly for $r \geq r_c$. The wavefunction shown is the 2s orbital of carbon; the pseudopotential was calculated with the OPIUM code [56] (version 3.3).

Modern pseudopotential methods are based on the approach taken by Philips and Kleinman [57] (inspired, in turn, by the earlier orthogonalised plane-wave method of Herring [58]), in which a valence state $|\phi_v\rangle$ is expressed in terms of a smooth pseudised wavefunction $|\phi_v^{\rm ps}\rangle$ and a linear combination of core wavefunctions:

$$|\phi_v^{\rm ps}\rangle = |\phi_v\rangle + \sum_c \alpha_{c,v} |\phi_c\rangle \, , \tag{3.1}$$

where c labels the core states and $\alpha_{c,v} = \langle \phi_c | \phi_v^{\rm ps} \rangle \neq 0$ (thus ensuring that the true valence wavefunction is orthogonal to all the core wavefunctions). The pseudised valence wavefunction can be shown to satisfy a Schrödinger equation with a suitably pseudised Hamiltonian:

$$\hat{H}^{\rm ps} |\phi_v^{\rm ps}\rangle = \varepsilon_v |\phi_v^{\rm ps}\rangle \, , \tag{3.2}$$

where

$$\begin{aligned}\hat{H}^{\mathrm{ps}} &= \hat{H} + \sum_{c} \left(\varepsilon_v - \varepsilon_c\right) \left|\phi_c\right\rangle \left\langle\phi_c\right| \\ &= \hat{T} + \hat{V} + \sum_{c} \left(\varepsilon_v - \varepsilon_c\right) \left|\phi_c\right\rangle \left\langle\phi_c\right| \\ &= \hat{T} + \hat{V}^{\mathrm{ps}}.\end{aligned} \tag{3.3}$$

In other words, we can substitute the true potential $\hat{V}$ with the pseudopotential $\hat{V}^{\mathrm{ps}}$ to obtain the same valence eigenvalue ε_v as before, but now with the smooth wavefunction $|\phi_v^{\mathrm{ps}}\rangle$ as the corresponding eigenvector.

The pseudopotential obtained in Eq. 3.3 is non-local and energy-dependent (on ε_v), since it acts differently depending on the valence wavefunction's angular momentum. The most general form of this type of l-dependent pseudopotential is, therefore,

$$\hat{V}^{\mathrm{ps}} = \sum_{l} \hat{v}_l^{\mathrm{ps}} \hat{P}_l, \tag{3.4}$$

where $\langle \mathbf{r} | \hat{v}_l^{\mathrm{ps}} \rangle = v_l^{\mathrm{ps}}\left(r\right)$ is a radial function, and we have defined the angular momentum projection operator $\hat{P}_l$, such that

$$\langle \mathbf{r} | \hat{P}_l | \phi \rangle = \sum_{m=-l}^{l} \int Y_{l,m}\left(\theta, \varphi\right) Y_{l,m}^{*}\left(\theta', \varphi'\right) \delta\left(r - r'\right) \phi\left(\mathbf{r}'\right) \, \mathrm{d}\mathbf{r}', \tag{3.5}$$

where $Y_{l,m}\left(\theta, \varphi\right)$ are spherical harmonics. This is known as a semi-local pseudopotential, since the potential $v_l^{\mathrm{ps}}\left(r\right)$ for each angular momentum channel is itself a local operator. It is generally convenient to separate semi-local pseudopotentials into a completely local part $v^{\mathrm{ps,loc}}\left(r\right)$ and a semi-local part using the difference $\Delta v_l^{\mathrm{ps}}\left(r\right)$ between the local part and the original radial potential for each channel:

$$\hat{V}^{\mathrm{ps}} = \hat{v}^{\mathrm{ps,loc}} + \sum_{l} \Delta \hat{v}_l^{\mathrm{ps}} \hat{P}_l. \tag{3.6}$$

The definition of $v^{\mathrm{ps,loc}}\left(r\right)$ is arbitrary; however, it is obviously most useful to define this as containing the long-ranged part of the bare $-\frac{1}{r}$ potential (which does not differ between angular momentum channels), therefore making $\Delta v_l^{\mathrm{ps}}\left(r\right)$ short-ranged and localised to the core region of the atom.

Unfortunately, the pseudopotential definition of Eq. 3.6 is non-separable and, hence, computationally expensive. This can be seen by considering the semi-local part of the matrix element V_{ij}^{ps} between two basis functions ϕ_i and ϕ_j; both for the case of a plane-wave basis or an atomic-like basis (in which the radial part is multiplied by a spherical harmonic), this results in integrals of the form

$$\int r^2 R_i^* (r) \, \Delta v_l^{\mathrm{ps}} (r) \, R_j (r) \; \mathrm{d}r, \tag{3.7}$$

where $R_i (r)$ is either a spherical Bessel function when working in a plane-wave basis, or the radial part of the orbital when working in an atomic-like basis. In both cases, this means that the number of integrals that have to be performed scales as the square of the number of basis functions. In order to achieve linear-scaling cost with basis set size, therefore, a different pseudopotential form, proposed by Kleinman and Bylander [59], is often used instead. The general form is

$$\hat{V}^{\mathrm{ps,KB}} = \hat{v}^{\mathrm{ps,loc}} + \sum_l \sum_{m=-l}^{l} \frac{\Delta \hat{v}_l^{\mathrm{ps}} \left| \phi_{l,m}^{\mathrm{at}} \right\rangle \left\langle \phi_{l,m}^{\mathrm{at}} \right| \Delta \hat{v}_l^{\mathrm{ps}}}{\left\langle \phi_{l,m}^{\mathrm{at}} \right| \Delta \hat{v}_l^{\mathrm{ps}} \left| \phi_{l,m}^{\mathrm{at}} \right\rangle}, \tag{3.8}$$

where

$$\phi_{l,m}^{\mathrm{at}} (\mathbf{r}) = R_l^{\mathrm{at}} (r) \, Y_{l,m} (\theta, \varphi) \tag{3.9}$$

are eigenfunctions of the pseudised Hamiltonian for the isolated atom. The Kleinman-Bylander pseudopotential acts on these reference eigenfunctions in an equivalent manner to the original semi-local pseudopotential, but is now separable, since it divides into two independent integrals for each of the basis functions.

Numerous schemes for generating pseudopotentials are available; most employ the constraint of norm-conservation first introduced by Hamann, Schlüter and Chiang [60], by which the pseudised wavefunction is required to maintain the same amount of charge within the core radius as the true wavefunction. Such 'norm-conserving' pseudopotentials can achieve a good degree of transferability between different chemical environments, provided that the energy range of the valence eigenvalues does not vary too drastically from the reference

energy (ε_v in Eq. 3.3) used in their construction.

More recently, a scheme of 'ultrasoft' pseudopotentials has been introduced by Vanderbilt [61], which relaxes the normalisation constraint while maintaining a good transferability. In this case, the missing valence charge density in the core region is represented by an auxiliary function, and the simple eigenvalue problem is substituted by a generalised problem; the result is that larger values of r_c can be safely employed, and sharp peaks of the pseudised wavefunctions in the core region can be avoided. Hence, the pseudised wavefunctions can generally be constructed so as to be smoother than in a norm-conserving scheme.

3.2 Plane waves and periodic systems

When numerically solving the Kohn-Sham eigenvalue equation given in Eq. 2.86 we can either use a real-space method such as finite difference [62, 63] to describe the single-particle orbitals and the total electronic density, or employ a basis set representation; the choice of basis set is crucial in determining the payoff between accuracy and speed for particular systems. A popular approach is to use localised atomic or atomic-like orbitals (such as Gaussians [64]) to capture most of the components of the solution with a small number of basis functions; however, it is often problematic to systematically improve the completeness of the basis and converge the results.

Alternatively, plane waves can be used as a basis; this approach is commonly used for performing calculations on crystalline solids, since the periodicity of the basis functions can be easily exploited, and their delocalised nature is not a drawback. Furthermore, the basis can be systematically converged, and many of the quantities required in a DFT calculation can be easily obtained by working in reciprocal space and using fast Fourier transforms [4,23,34] (FFTs).

The periodic boundary conditions imposed by the use of a plane-wave basis are naturally compatible with the case of an infinite periodic system, to which Bloch's theorem can be applied[2]. Bloch's theorem states that, given a one-

[2] A proof of the theorem can be found in, e.g., Ashcroft and Mermin [55].

electron Schrödinger equation of the form

$$H\phi_m(\mathbf{r}) = \left[-\frac{1}{2}\nabla^2_{\mathbf{r}} + v(\mathbf{r})\right]\phi_m(\mathbf{r}) = \varepsilon_m\phi_m(\mathbf{r}), \tag{3.10}$$

where the potential $v(\mathbf{r})$ is periodic such that

$$v(\mathbf{r}) = v(\mathbf{r}+\mathbf{T}) \tag{3.11}$$

for any Bravais lattice vector $\mathbf{T}$, the wavefunction may be written as

$$\phi_{m,\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}u_{m,\mathbf{k}}(\mathbf{r}), \tag{3.12}$$

where $\mathbf{k}$ is any real wavevector and $u_{m,\mathbf{k}}(\mathbf{r})$ is a function with the same periodicity as the potential:

$$u_{m,\mathbf{k}}(\mathbf{r}) = u_{m,\mathbf{k}}(\mathbf{r}+\mathbf{T}). \tag{3.13}$$

From this definition it follows that solutions differing by a reciprocal lattice vector $\mathbf{G}$ are equivalent; in fact, considering a wavevector $\mathbf{k}' = \mathbf{k}+\mathbf{G}$, we obtain:

$$\begin{aligned}\phi_{m,\mathbf{k}'}(\mathbf{r}) &= \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathbf{r}}u_{m,\mathbf{k}'}(\mathbf{r}) \\ &= \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}u_{n,\mathbf{k}}(\mathbf{r}) = \phi_{n,\mathbf{k}}(\mathbf{r}),\end{aligned} \tag{3.14}$$

where a band index n has now substituted the eigenvalue index m used previously[3]. The importance of this result lies in the fact that for an infinite periodic system we need only consider a finite number of occupied wavefunctions at each wavevector $\mathbf{k}$, also known as a k-point, inside the first Brillouin zone (the reciprocal-space unit cell defined by the reciprocal lattice vectors).

In order to show how we can exploit Bloch's theorem in a plane-wave basis, we first consider a 1D finite system with periodic boundary conditions, as shown in Fig. 3.2. The system consists of N repeating segments of an

[3] The wavevector $\mathbf{k}$ is the good quantum number associated with the translation operator $\hat{T}_{\mathbf{T}}$ (that acts on a function by shifting it by any Bravais lattice vector $\mathbf{T}$), that can be shown to commute with the Hamiltonian.

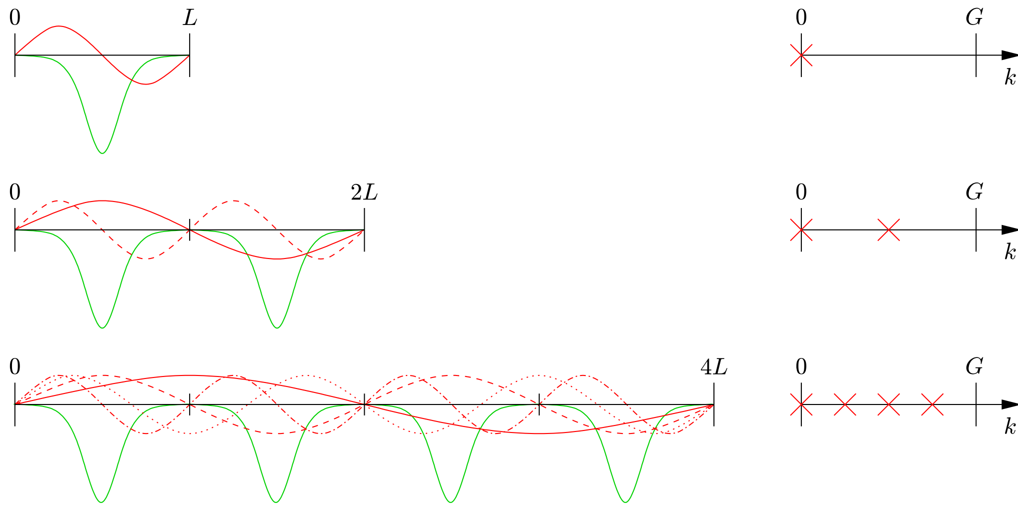


Figure 3.2: Schematic illustration of the use of a plane-wave basis in a finite 1D system. The external potential $v(x)$ of length L is shown in green; for each system size N the allowed plane waves with wavenumbers in the interval $(0, 2\pi/L]$ are shown in red. The wavenumbers are also plotted in reciprocal space, showing how the k-point density increases as $N \to \infty$ (note that $k = G = 2\pi/L$ has been shifted to $k = 0$).

arbitrary external potential $v(x)$ of length L; the total length of the system is therefore NL. In this case, we need only use a discrete basis of plane waves with wavenumbers $\{2\pi n/(NL), n \in \mathbb{Z}\}$. The Hamiltonian matrix elements are given by[4]

$$H_{nn'} = \int_0^{NL} \mathrm{e}^{-\mathrm{i}\frac{2\pi n}{NL}x} \left[-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + v(x) \right] \mathrm{e}^{\mathrm{i}\frac{2\pi n'}{NL}x}\,\mathrm{d}x. \tag{3.15}$$

The kinetic energy component becomes

$$T_{nn'} = \frac{1}{2}\left(\frac{2\pi n'}{NL}\right)^2 \int_0^{NL} \mathrm{e}^{\mathrm{i}\frac{2\pi}{NL}(n'-n)x}\,\mathrm{d}x, \tag{3.16}$$

which is only non-zero when $n' = n$ (i.e., the diagonal matrix elements). In-

[4]The normalisation of the basis functions has been neglected.

stead, the potential energy component becomes

$$\begin{aligned} V_{nn'} &= \int_0^{NL} v(x)\, \mathrm{e}^{\mathrm{i}\frac{2\pi}{NL}(n'-n)x}\, \mathrm{d}x \\ &= \int_0^{NL} \left[\sum_{m=-\infty}^{\infty} v_m \mathrm{e}^{\mathrm{i}\frac{2\pi m}{L}x} \right] \mathrm{e}^{\mathrm{i}\frac{2\pi}{NL}(n'-n)x}\, \mathrm{d}x \\ &= \sum_{m=-\infty}^{\infty} v_m \int_0^{NL} \mathrm{e}^{\mathrm{i}\frac{2\pi}{NL}(mN+n'-n)x}\, \mathrm{d}x, \end{aligned} \tag{3.17}$$

where v_m are the Fourier coefficients of $v(x)$. Therefore, $V_{nn'}$ will be non-zero if

$$mN + n' - n = 0 \tag{3.18}$$

for at least one value of m. For $N = 1$ this will always be true, and so none of the Hamiltonian matrix elements are zero. However, for $N = 2$ it is no longer the case; in fact, since mN is even, $V_{nn'}$ is zero when $(n' - n)$ is odd. This means that there is a decoupling of odd and even values of n, corresponding to two independent sets of wavenumbers:

$$k^{(1)} = \left\{ \frac{2\pi l}{L}, l \in \mathbb{Z} \right\}, \tag{3.19}$$

and

$$k^{(2)} = \left\{ \frac{\pi}{L}(2l+1), l \in \mathbb{Z} \right\}. \tag{3.20}$$

Therefore, the Hamiltonian becomes a block diagonal matrix, which can be separated into two smaller matrices, leading to two independent sets of eigenstates. It can easily be seen that for the general case there are N independent sets of wavenumbers, defined as

$$k^{(a)} = \left\{ \frac{2\pi}{L} \left[l + \frac{(a-1)}{N} \right], l \in \mathbb{Z} \right\}, \tag{3.21}$$

for $a = 1, 2, \ldots, N$; the sets are therefore evenly spaced within each Brillouin zone. Therefore, if we consider the limiting case of infinite N we ob-

tain a continuum of independent k-points in the interval $[0, 2\pi/L)$; we can find the eigenvalues at each k-point by diagonalising the submatrix containing the elements of the Hamiltonian between plane waves with wavenumbers $\{(2\pi l/L) + k, l \in \mathbb{Z}\}$, thus resulting in eigenfunctions of the form given by Eq. 3.12, in agreement with Bloch's theorem. Similarly, for the 3D infinite system the potential energy component of the Hamiltonian matrix element is

$$V_{\mathbf{k}\mathbf{k}'} = \sum_{\mathbf{G}} v_{\mathbf{G}} \int \mathrm{e}^{\mathrm{i}(\mathbf{G}+\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}} \, \mathrm{d}\mathbf{r}; \tag{3.22}$$

analogously to the 1D case, only wavevectors differing exactly by a reciprocal lattice vector $\mathbf{G}$ will produce non-zero elements, making them part of the same set. k-points inside the first Brillouin zone, therefore, are all independent of each other.

When performing plane-wave DFT calculations for periodic systems, there are two necessary numerical approximations that must be introduced. Firstly, the number of basis functions used to construct the Hamiltonian matrix must be made finite by neglecting all plane waves with wavevectors larger than a specified cut-off value; this is done in terms of a kinetic energy cut-off E^{cut}, such that only plane waves for which

$$\frac{1}{2} |\mathbf{k} + \mathbf{G}|^2 < E^{\mathrm{cut}} \tag{3.23}$$

are included in the calculation. Since higher-wavevector plane waves oscillate more rapidly, this cut-off effectively sets the resolution at which wavefunctions are described. As mentioned previously, the results obtained using the plane-wave basis can be systematically converged, and this is done simply by increasing the value of E^{cut}. Furthermore, the total energy of the system decreases monotonically as the kinetic energy cut-off of the basis is increased; this is guaranteed by the variational principle, since increasing E^{cut} will increase the variational freedom of the trial wavefunction without disallowing any solution obtainable with a lower value.

The second numerical approximation to be employed is that the continuous volume of independent k-points within the first Brillouin zone in reciprocal

space must be sampled discretely; any integral over its volume, as in the calculation of the total energy, is therefore replaced by a summation (this implies a certain interpolation between k-points). In practice, we rely on the smoothness of quantities of interest (such as the Kohn-Sham eigenvalues) as a function of $\mathbf{k}$ to obtain reasonable results even with a fairly coarse k-point sampling. One of the most common sampling techniques is that proposed by Monkhorst and Pack [65] (MP); this consists of a regular grid of $N_1 \times N_2 \times N_3$ points (in 3D), uniquely defined in terms of the reciprocal lattice vectors $\{\mathbf{G}_i\}$:

$$\mathbf{k}_{n_1,n_2,n_3} = \sum_{i=1}^{3} \frac{2n_i - N_i - 1}{2N_i} \mathbf{G}_i, \tag{3.24}$$

where $n_i = 1, 2, \ldots, N_i$. MP grids also allow for a systematic convergence of quantities of interest, by increasing the values $\{N_i\}$ to obtain a progressively finer representation of the Brillouin zone; this convergence, however, does not have to be monotonic. We note that the definition in Eq. 3.24 ensures that the grid is always centred around the high-symmetry point[5] $\mathbf{k} = \mathbf{0}$ (Γ), which will itself only be included if all values of N_i are odd; in fact, MP grids do not usually include high-symmetry k-points, as these generally present extremum values of $\mathbf{k}$-dependent quantities that can lead to a slow convergence of Brillouin zone summations.

Bloch's theorem, which allows us to restrict the independent eigenvalue solutions for periodic systems to a finite portion of reciprocal space, is a consequence of the translational symmetries of the Hamiltonian that follow from Eq. 3.11; additional symmetries of the Hamiltonian can be used to further reduce the volume of reciprocal space that needs to be considered (and, hence, the number of independent k-points that need to be sampled); in fact, for all crystals the Brillouin zone can be reduced to at least half the size due to the relationship between $\mathbf{k}$ and $-\mathbf{k}$. This can be shown by considering the time-reversal symmetry of the original time-dependent Schrödinger equation,

[5]The symmetry of a k-point $\mathbf{k}$ is determined by the subset of space group symmetries of the lattice for which the operation of the transformation matrix $\mathbf{R}$ results in the same k-point (or a translation by a reciprocal lattice vector $\mathbf{G}$): $\mathbf{R}\mathbf{k} = \mathbf{k} + \mathbf{G}$. See Appendix B for further discussion.

from which it follows that the Hamiltonian can always be chosen to be real[6]; therefore, the complex conjugate of any eigenfunction $\phi_m(\mathbf{r})$ is itself an eigenfunction with the same eigenvalue, since

$$(H\phi_m(\mathbf{r}))^* = (\varepsilon_m \phi_m(\mathbf{r}))^* \tag{3.25}$$

becomes

$$H\phi_m^*(\mathbf{r}) = \varepsilon_m \phi_m^*(\mathbf{r}). \tag{3.26}$$

For the case of a periodic system, we obtain:

$$\begin{aligned} \phi_{n,\mathbf{k}}^*(\mathbf{r}) &= \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}^*(\mathbf{r}) \\ &= \mathrm{e}^{\mathrm{i}(-\mathbf{k})\cdot\mathbf{r}} u_{n,-\mathbf{k}}(\mathbf{r}) = \phi_{n,-\mathbf{k}}(\mathbf{r}); \end{aligned} \tag{3.27}$$

therefore, the eigenvalue solutions at the two k-points $\mathbf{k}$ and $-\mathbf{k}$ are equivalent[7]. Additional space-group symmetries satisfied by the external potential (and, hence, the Hamiltonian) will further reduce the number of independent k-points in a similar fashion. The minimum portion of the Brillouin zone that contains all of the necessary information to describe it in its entirety is known as the irreducible wedge [23, 34].

3.3 Localised orbitals

As mentioned at the start of the previous section, it is possible to employ a basis set of localised functions to represent the Kohn-Sham eigenstates. In many ways, this approach of describing electronic structure is diametrically opposed to the plane-wave one, as it provides a physically intuitive, real-space view of the system that can be easily understood in terms of individual bonds and local chemistry, instead of the delocalised, reciprocal-space view offered by plane waves. Because of this, empirical models using localised functions

[6]This time-reversal symmetry is no longer valid in the presence of magnetic fields.

[7]Furthermore, the relationship between the Bloch functions at these k-points is $u_{n,-\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}^*(\mathbf{r})$, from which it follows that all Bloch functions at $\mathbf{k} = \mathbf{0}$ can always be chosen to be real. This is also the case for certain other k-points, for which $-\mathbf{k}$ differs from $\mathbf{k}$ by a reciprocal lattice vector.

can often provide a useful and accurate description of the main features of the electronic structure of a particular system even with a very small number of basis functions. This approach is also naturally well-suited to the study of molecules and clusters, in which it is desirable to concentrate most of the computational effort in describing the region of high electronic density, as opposed to the vacuum region; this is not possible in a plane-wave representation, as all of space is treated with an equal resolution. However, we emphasise that there can be no fundamental disagreement between the two representations, and, hence, that both methods must yield the same results when taken to convergence. In fact, the theory of Wannier functions can be viewed as a formal connection between the two, as discussed in Sec. 3.3.2.

3.3.1 The tight-binding approach

Although localised functions can be used to construct 'floating' basis sets [66] (i.e., ones in which the basis functions are fixed to definite points in space, independently of the positioning of the nuclei), it is generally most convenient to consider atom-centred basis sets, in which the basis functions take the form of atomic or atomic-like orbitals (AOs) which are tied to a particular nucleus. This idea forms the basis of the local combination of atomic orbitals (LCAO) theory, widely used for qualitative discussion of the properties of extended wavefunctions, also commonly referred to in this case as molecular orbitals (MOs).

The use of a localised, atom-centred basis set for electronic structure calculations is given the name of tight-binding [67], as this model is naturally better suited for cases in which the electrons are tightly bound to each atom. Tight-binding calculations range from fully self-consistent *ab initio* methods to empirical methods in which the orbitals are not explicitly specified, and the Hamiltonian matrix elements are fitted to reproduce results from experimental studies or other theoretical calculations.

We can consider a general tight-binding basis set for a periodically repeating system as being composed of localised orbitals of the form $\langle \mathbf{r} | \xi_{n,\alpha,\mathbf{T}} \rangle = \xi_{n,\alpha} \left(\mathbf{r} - \left(\mathbf{R}_\alpha + \mathbf{T} \right) \right)$, where this function is the n-th orbital associated with

atom α, centred at the nuclear coordinate $\mathbf{R}_\alpha$ in the unit cell defined by the Bravais lattice translation vector $\mathbf{T}$. By combining n and α into a single index m labelling all M orbitals belonging to a particular unit cell, this can be simplified to $\langle \mathbf{r} | \xi_{m,\mathbf{T}} \rangle = \xi_m \left(\mathbf{r} - (\mathbf{R}_m + \mathbf{T}) \right)$.

Due to the periodicity of the system, all eigenfunctions of the Hamiltonian must satisfy the Bloch condition (equivalent to that given by Eq. 3.12)

$$\phi_{\mathbf{k}} \left(\mathbf{r} + \mathbf{T} \right) = \mathrm{e}^{\mathrm{i}\mathbf{k} \cdot \mathbf{T}} \phi_{\mathbf{k}} \left(\mathbf{r} \right), \tag{3.28}$$

leading to the general form of the solution in the tight-binding basis

$$\phi_{\mathbf{k}} \left(\mathbf{r} \right) = \sum_{m=1}^{M} \sum_{\mathbf{T}} a_m \mathrm{e}^{\mathrm{i}\mathbf{k} \cdot \mathbf{T}} \xi_m \left(\mathbf{r} - (\mathbf{R}_m + \mathbf{T}) \right). \tag{3.29}$$

This can be easily shown to satisfy Eq. 3.28 for an arbitrary translation $\mathbf{T}'$ by making use of the property of the basis functions

$$\langle \mathbf{r} + \mathbf{T}' | \xi_{m,\mathbf{T}} \rangle = \langle \mathbf{r} | \xi_{m,\mathbf{T}-\mathbf{T}'} \rangle . \tag{3.30}$$

In order to find the eigenfunction for a particular Hamiltonian, therefore, we need to calculate the set of M coefficients $\{a_m\}$. Substituting Eq. 3.29 into the eigenvalue problem $\hat{H} \left| \phi_{\mathbf{k}} \right\rangle = \varepsilon_{\mathbf{k}} \left| \phi_{\mathbf{k}} \right\rangle$, we obtain:

$$\sum_{m=1}^{M} \sum_{\mathbf{T}} a_m \mathrm{e}^{\mathrm{i}\mathbf{k} \cdot \mathbf{T}} \hat{H} \left| \xi_{m,\mathbf{T}} \right\rangle = \varepsilon_{\mathbf{k}} \sum_{m=1}^{M} \sum_{\mathbf{T}} a_m \mathrm{e}^{\mathrm{i}\mathbf{k} \cdot \mathbf{T}} \left| \xi_{m,\mathbf{T}} \right\rangle . \tag{3.31}$$

If we now project onto the function $\langle \xi_{m',\mathbf{T}'} | \, \mathrm{e}^{-\mathrm{i}\mathbf{k} \cdot \mathbf{T}'}$, this becomes:

$$\sum_{m=1}^{M} \sum_{\mathbf{T}} a_m \mathrm{e}^{\mathrm{i}\mathbf{k} \cdot (\mathbf{T}-\mathbf{T}')} H_{(m',\mathbf{T}')(m,\mathbf{T})} = \varepsilon_{\mathbf{k}} \sum_{m=1}^{M} \sum_{\mathbf{T}} a_m \mathrm{e}^{\mathrm{i}\mathbf{k} \cdot (\mathbf{T}-\mathbf{T}')} S_{(m',\mathbf{T}')(m,\mathbf{T})}, \tag{3.32}$$

where

$$H_{(m',\mathbf{T}')(m,\mathbf{T})} = \langle \xi_{m',\mathbf{T}'} | \, \hat{H} \left| \xi_{m,\mathbf{T}} \right\rangle \tag{3.33}$$

and

$$S_{(m',\mathbf{T}')(m,\mathbf{T})} = \langle \xi_{m',\mathbf{T}'} | \xi_{m,\mathbf{T}} \rangle \tag{3.34}$$

are the customary Hamiltonian and overlap matrix elements for the basis, respectively. We can now see that Eq. 3.32 is equivalent for all translations $\mathbf{T}'$, since

$$\mathrm{e}^{\mathbf{ik}\cdot(\mathbf{T}-\mathbf{T}'')} \langle \xi_{m',\mathbf{T}''} | \hat{H} | \xi_{m,\mathbf{T}} \rangle = \mathrm{e}^{\mathbf{ik}\cdot(\mathbf{T}-\mathbf{T}'')} \langle \xi_{m',\mathbf{T}'} | \hat{H} | \xi_{m,\mathbf{T}+\mathbf{T}'-\mathbf{T}''} \rangle , \tag{3.35}$$

and similarly for the right-hand side of the equation (i.e., we have simply shifted the origin of the sum over $\mathbf{T}$, which has no effect for an infinite system). Therefore, we need only consider $\mathbf{T}' = \mathbf{0}$, leaving a set of M linear equations to be solved. These can be represented as a generalised eigenvalue problem of the form

$$\mathbf{H}^{\mathrm{TB}}\mathbf{a} = \varepsilon_{\mathbf{k}}\mathbf{S}^{\mathrm{TB}}\mathbf{a}, \tag{3.36}$$

where $\mathbf{a}$ is the vector of coefficients $\{a_m\}$, and we have defined a modified $M \times M$ Hamiltonian matrix $\mathbf{H}^{\mathrm{TB}}$ and overlap matrix $\mathbf{S}^{\mathrm{TB}}$, such that

$$H^{\mathrm{TB}}_{m'm} = \sum_{\mathbf{T}} \mathrm{e}^{\mathbf{ik}\cdot\mathbf{T}} H_{(m',\mathbf{0})(m,\mathbf{T})} \tag{3.37}$$

and

$$S^{\mathrm{TB}}_{m'm} = \sum_{\mathbf{T}} \mathrm{e}^{\mathbf{ik}\cdot\mathbf{T}} S_{(m',\mathbf{0})(m,\mathbf{T})}. \tag{3.38}$$

The advantage of working in a localised basis is now apparent, since the sum over lattice translations for each element of $\mathbf{H}^{\mathrm{TB}}$ and $\mathbf{S}^{\mathrm{TB}}$ can be drastically truncated; it is clear, in fact, that the magnitude of the Hamiltonian matrix element $H_{(m',\mathbf{0})(m,\mathbf{T})}$ will decay with the distance between the centres of the two orbitals $|\mathbf{R}_m + \mathbf{T} - \mathbf{R}_{m'}|$, with a similar form to the spatial decay of the orbitals themselves away from their centre. The truncation is usually done so as to only consider neighbouring unit cells. More generally, matrix elements between orbitals belonging to atoms whose separation is greater than a specified cut-off distance are discarded.

Tight-binding matrix elements can be classified according to the number of

atomic sites, or centres, that are involved in the integral. The overlap matrix elements and the kinetic energy term in the Hamiltonian matrix elements involve up to two centres, depending on whether the two orbitals m' and m belong to the same site or different sites; however, potential energy terms in the Hamiltonian also require three-centre integrals, in which the potential (approximated as a sum of atom-centred contributions) belongs to a different site from the two orbitals. Due to their computational cost, three-centre terms are usually neglected. Slater and Koster [68] showed how the calculation of the remaining two-centre terms can be greatly simplified by considering the symmetry of the orbitals, leading to a complete description of the matrix elements in terms of a small number of independent parameters.

3.3.2 Maximally-localised Wannier functions

Wannier functions [69–71] are a set of orthonormal localised basis functions used in periodic systems, defined in such a way as to span the same space as the delocalised eigenstates belonging to either a single band or a group of bands. The eigenstates, as shown in Sec. 3.2, are Bloch orbitals; however, there is an additional arbitrary phase $\theta_n(\mathbf{k})$ at each band n and k-point $\mathbf{k}$ that is periodic in reciprocal space and does not affect any physical quantities, and therefore is not fixed by the Schrödinger equation (this is an example of a gauge transformation):

$$\phi_{n,\mathbf{k}}(\mathbf{r}) \rightarrow \mathrm{e}^{\mathrm{i}\theta_n(\mathbf{k})}\phi_{n,\mathbf{k}}(\mathbf{r}). \tag{3.39}$$

The Wannier functions of a system are effectively the Fourier transforms of its Bloch orbitals, defined as

$$\begin{aligned} \langle \mathbf{r} | w_{n,\mathbf{T}} \rangle = w_n(\mathbf{r} - \mathbf{T}) &= \frac{V}{(2\pi)^3} \int \mathrm{e}^{\mathrm{i}\theta_n(\mathbf{k}) - \mathrm{i}\mathbf{k}\cdot\mathbf{T}} \phi_{n,\mathbf{k}}(\mathbf{r})\, \mathrm{d}\mathbf{k} \\ &= \frac{V}{(2\pi)^3} \int \mathrm{e}^{\mathrm{i}\theta_n(\mathbf{k}) - \mathrm{i}\mathbf{k}\cdot(\mathbf{T}-\mathbf{r})} u_{n,\mathbf{k}}(\mathbf{r})\, \mathrm{d}\mathbf{k}, \end{aligned} \tag{3.40}$$

where the integral is carried out over the entire Brillouin zone; due to the undefined phase shift, this representation is non-unique. The Wannier func-

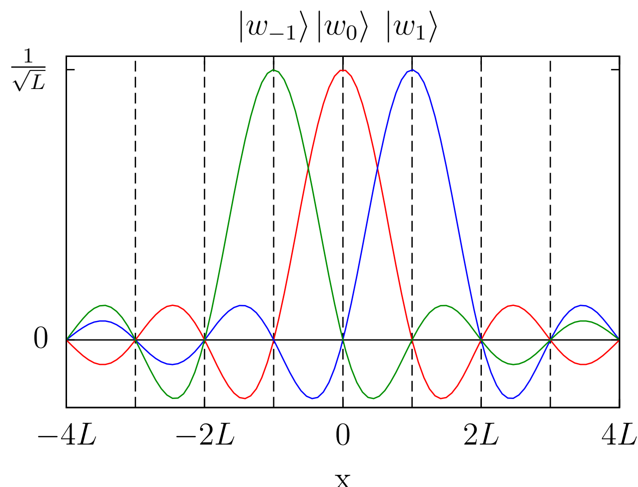


Figure 3.3: Three Wannier functions in neighbouring unit cells for the lowest band of the 1D free electron model, as described in the text.

tions are labelled by the band number n and the unit cell to which they are associated, defined by the translation vector $\mathbf{T}$, and functions in different cells transform into each other by a translation $\mathbf{T} - \mathbf{T}'$; in other words, they can be considered to be a tight-binding basis as defined in the previous section. Furthermore, they can be shown to form an orthonormal set, and, hence, $S_{(m',\mathbf{T}')(m,\mathbf{T})} = \delta_{(m',\mathbf{T}')(m,\mathbf{T})}$ and $\mathbf{S}^{\mathrm{TB}} = \mathbf{I}$.

Fig. 3.3 shows the simplest example of a single-band Wannier basis, for the lowest band of the 1D free electron model in a unit cell of length L. The normalised Bloch orbitals are given by

$$\phi_k(x) = \frac{1}{\sqrt{L}} \mathrm{e}^{\mathrm{i}kx}, \tag{3.41}$$

and the Wannier functions obtained from the 1D equivalent of Eq. 3.40 (setting

$\theta(k) = 0$) are, therefore,

$$\begin{aligned} \langle x | w_m \rangle = w(x - Lm) &= \frac{L}{2\pi\sqrt{L}} \int_{-\pi/L}^{\pi/L} \mathrm{e}^{\mathrm{i}k(x-Lm)}\,\mathrm{d}k \\ &= \frac{\sqrt{L}}{\pi} (-1)^m \frac{\sin\frac{\pi x}{L}}{(x - Lm)}. \end{aligned} \tag{3.42}$$

As can be seen from this expression, the Wannier function belonging to each unit cell for this system is given by a sinc function centred at $x = Lm$; when moving away from the unit cell they belong to, the functions decay with an oscillating tail, in order to satisfy the orthogonality requirements.

The definition of the Wannier basis given by Eq. 3.40 can be generalised to include linear combinations of Bloch orbitals from different bands; if we consider a group of N bands, we now obtain:

$$\langle \mathbf{r} | w_{n,\mathbf{T}} \rangle = w_n(\mathbf{r} - \mathbf{T}) = \frac{V}{(2\pi)^3} \int \sum_{m=1}^{N} U_{mn}^{(\mathbf{k})} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{T}-\mathbf{r})} u_{m,\mathbf{k}}(\mathbf{r})\,\mathrm{d}\mathbf{k}, \tag{3.43}$$

where $U_{mn}^{(\mathbf{k})}$ is an $N \times N$ unitary matrix that defines both the mixing between bands and the arbitrary phase factor at each k-point. A popular method of determining an 'optimal' $U_{mn}^{(\mathbf{k})}$ matrix, and, hence, of reducing the arbitrariness of the Wannier basis, is that of maximally-localised Wannier functions (MLWFs), first proposed in the formulation outlined below by Marzari and Vanderbilt [72]. Their approach consists in varying $U_{mn}^{(\mathbf{k})}$ so as to minimise the sum of the second central moments (or quadratic spreads) of all the Wannier functions, defined as

$$\begin{aligned} \Omega &= \sum_{n=1}^{N} \left\langle (\mathbf{r} - \langle \mathbf{r} \rangle_n)^2 \right\rangle_n \\ &= \sum_{n=1}^{N} \left\{ \left\langle r^2 \right\rangle_n - 2 \langle \mathbf{r} \cdot \langle \mathbf{r} \rangle_n \rangle_n + |\langle \mathbf{r} \rangle_n|^2 \right\} \\ &= \sum_{n=1}^{N} \left\{ \left\langle r^2 \right\rangle_n - |\langle \mathbf{r} \rangle_n|^2 \right\}, \end{aligned} \tag{3.44}$$

where $\langle \hat{o} \rangle_n = \langle w_{n,\mathbf{0}} | \, \hat{o} \, | w_{n,\mathbf{0}} \rangle$. By adding and subtracting the off-diagonal components of $|\langle \mathbf{r} \rangle_n|^2$ (i.e., those between two different Wannier functions $|w_{n,\mathbf{0}}\rangle$ and $|w_{m \neq n, \mathbf{T} \neq \mathbf{0}}\rangle$), Ω can be decomposed into the contributions

$$\Omega = \Omega_{\mathrm{I}} + \widetilde{\Omega}, \tag{3.45}$$

where

$$\Omega_{\mathrm{I}} = \sum_{n=1}^{N} \left\{ \langle w_{n,\mathbf{0}} | \, r^2 \, | w_{n,\mathbf{0}} \rangle - \sum_{m=1}^{N} \sum_{\mathbf{T}} |\langle w_{m,\mathbf{T}} | \, \mathbf{r} \, | w_{n,\mathbf{0}} \rangle|^2 \right\} \tag{3.46}$$

and

$$\widetilde{\Omega} = \sum_{n=1}^{N} \sum_{m \neq n}^{N} \sum_{\mathrm{T} \neq \mathbf{0}} |\langle w_{m,\mathbf{T}} | \, \mathbf{r} \, | w_{n,\mathbf{0}} \rangle|^2 . \tag{3.47}$$

It can be shown that both of these quantities are positive, and that Ω_{I} is gauge-invariant [72] and related to the electronic localisation length of the system [73]. The MLWFs, therefore, can be calculated by minimising $\widetilde{\Omega}$.

The 'wannierisation' procedure that follows from this definition of localisation is straightforward when considering a set of N isolated bands (i.e., ones that are completely separated by a finite gap from all other bands); however, difficulties arise in selecting N eigenstates at each k-point when the bands of interest are not isolated, but are instead entangled with other bands in certain portions of the Brillouin zone (as might be the case in metallic systems). For such cases, Souza *et al.* [74] have suggested a 'disentanglement' technique that extracts an optimally-connected subspace of eigenstates from the complete eigenspectrum. We note that the MLWFs describing the occupied manifold for an insulating system that satisfies time-reversal symmetry (or, more generally, any set of smooth bands in a system with such symmetry) have been shown to be both real (provided there is a unique minimum in the functional $\Omega[U]$) and exponentially localised[8] [75].

MLWFs can be viewed as the equivalent for periodic systems of Boys orbitals [76], that are used in a similar fashion for isolated molecules to construct

[8] The Wannier functions shown in Fig. 3.3 decay as x^{-1}, and so are not exponentially localised; this is inevitable, since the free electron model does not feature isolated bands, and the lowest band is discontinuous in its first derivative at $k = \frac{\pi}{L}(2m+1)$.

a localised representation of the space spanned by the extended eigenstates. The localised view of the electronic structure in real space given by MLWFs in periodic systems and Boys orbitals in molecular systems can provide an insightful picture of the chemical bonding, and can in fact be employed to distinguish the number and type of bonds present; such analysis has been used in the study of crystalline [74, 77] and disordered systems [78, 79], as well as defect centres [80]. Furthermore, MLWFs have also been used as an optimal tight-binding basis for k-space interpolation [81, 82], and for the study of the dielectric properties of materials within the modern theory of polarisation [83–85] (due to the formal connection in the latter case with MLWF theory).

3.3.3 The density matrix formalism and linear-scaling DFT

Conventional methods for solving the Kohn-Sham equation suffer from an unfavourable scaling: the computational cost increases asymptotically as the cube of the number of atoms in the system. This is because for N atoms, N eigenstates must be found, and each eigenstate extends over a volume generally proportional to N; this means that the amount of information in the system scales as N^2. However, the constraint that the eigenstates must be mutually orthogonal increases the computational cost by a further factor of N, thereby resulting in N^3 scaling. This fact severely limits the number of atoms that can routinely be treated in a DFT calculation, typically to a few hundred.

However, it has been noted that the amount of *useful* information contained in the Kohn-Sham eigenstates scales as N rather than N^2; this is related to the concept of quantum mechanical 'nearsightedness' [86] or electronic localisation, by which the observable properties of a region of space depend only weakly on other distant regions. When using the Kohn-Sham eigenstates this is not readily apparent, due to their delocalised nature. However, the problem can be recast in terms of the single-particle density matrix [87]

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i f_i \phi_i(\mathbf{r}) \phi_i^*(\mathbf{r}'), \tag{3.48}$$

where $\{\phi_i(\mathbf{r})\}$ are the Kohn-Sham eigenstates and $\{f_i\}$ their occupancies, which at zero temperature are restricted to 0 or 1. The corresponding operator, therefore, is defined such that

$$\hat{\rho}\,|\xi\rangle = \sum_i f_i\,|\phi_i\rangle\,\langle\phi_i|\xi\rangle\,. \tag{3.49}$$

The density matrix decays exponentially for insulators[9] [89, 90] and metals at finite electronic temperature [91], and therefore can be truncated to eliminate all the information that is not significant when calculating the observable properties of the system; this is the basis of several linear-scaling DFT methods (for a recent review, see Ref. [92]). We will focus on the approach used in the ONETEP code [93], which is closely related to the conventional cubic-scaling plane-wave basis set method.

Using the density matrix, the electronic charge density of the system can be calculated (assuming spin degeneracy) as

$$n(\mathbf{r}) = 2\rho(\mathbf{r},\mathbf{r})\,, \tag{3.50}$$

and the total energy of the non-interacting system as

$$E_0^{\mathrm{KS}} = 2\mathrm{Tr}\,(\boldsymbol{\rho}\mathbf{H})\,, \tag{3.51}$$

where the Hamiltonian and the density matrix are represented in an arbitrary complete orthonormal basis. The energy of the corresponding interacting system E_0 is given by Eq. 2.89 as before; the solution to the Kohn-Sham equation is obtained by minimising E_0 directly with respect to the density matrix, subject to the constraints of normalisation to N particles and idempotency:

$$\hat{\rho}^2 = \hat{\rho}. \tag{3.52}$$

The idempotency requirement is valid at zero temperature, when all eigenstates are either fully occupied or fully unoccupied.

[9] In fact, the localisation properties of the density matrix are connected to those of the Wannier functions for the occupied manifold [88].

In order to exploit the exponential decay of the density matrix we define it in terms of a basis of localised functions $\{\xi_\alpha(\mathbf{r})\}$, called non-orthogonal generalised Wannier functions (NGWFs), and write it in the separable form

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \xi_\alpha(\mathbf{r}) K^{\alpha\beta} \xi_\beta^*(\mathbf{r}'), \tag{3.53}$$

where we have defined the matrix $K^{\alpha\beta}$, known as the density kernel [87]. In fact, $K^{\alpha\beta}$ can be shown to be simply the density matrix in the set of duals of the NGWFs[10] $\{\xi^\alpha(\mathbf{r})\}$, defined such that

$$\langle \xi_\beta | \xi^\alpha \rangle = \delta_\beta^\alpha; \tag{3.54}$$

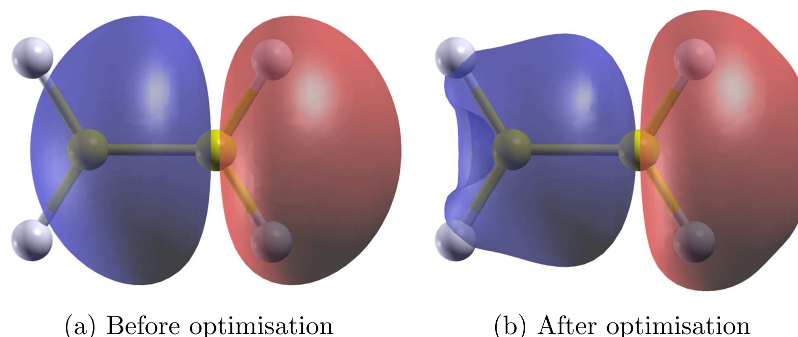
when working in a non-orthogonal basis, Eq. 3.51 becomes

$$E_0^{\mathrm{KS}} = 2\mathrm{Tr}(\mathbf{KH}). \tag{3.55}$$

The density kernel can be truncated by defining a cut-off radius r^{cut} and setting the matrix elements to zero when $|\mathbf{r}_\alpha - \mathbf{r}_\beta| > r^{\mathrm{cut}}$, where $\mathbf{r}_\alpha$ and $\mathbf{r}_\beta$ denote the centres of the localised NGWF orbitals. The energy minimisation for a given set of NGWFs is achieved by optimising the density kernel while enforcing the constraints on the density matrix; this is done through two different methods [94, 95]. We note that, in practice, it is not necessary to calculate the duals explicitly.

As discussed previously, it is difficult to improve the completeness of a localised basis set. In order to improve the accuracy of the calculation, therefore, it is generally necessary to significantly expand the size of the basis set, which is computationally very expensive; furthermore, it is usually not possible to do so in a systematic manner. ONETEP overcomes this problem by using a minimal number of NGWFs and optimising them *in situ* during the calculation with a conjugate gradients [4] (CG) minimisation scheme. This results in the NGWFs of each atom reflecting their molecular environment (an example of

[10] For an orthogonal basis the duals are equivalent to the original basis functions, and so the density kernel is equivalent to the density matrix.



(a) Before optimisation (b) After optimisation

Figure 3.4: Contour-surface plot showing the optimisation of an NGWF orbital during a ONETEP calculation. This example shows a p orbital centred on one of the carbon atoms of an ethene molecule; the optimisation distorts the orbital in response to the other atoms in the system.

this is shown in Fig. 3.4), and the NGWF basis set accurately spanning the occupied manifold. Consequently, the calculation of properties such as binding energies does not suffer from the well-known problem of basis set superposition error (BSSE) for finite atom-centred basis sets [96] (which is due to the fact that basis sets in systems with a different number or configuration of atoms do not span the same space), thus eliminating the need for correction techniques such as the commonly used counterpoise correction method [97, 98].

To achieve this optimisation, the NGWFs are expanded in terms of an underlying orthogonal basis of periodic sinc (psinc) functions [99]; these are localised functions that are related to plane-waves by discrete Fourier transform[11]. ONETEP uses a regular grid of points in real space with a psinc function centred at each point. The relationship between psincs and plane-waves allows for both an efficient calculation of the quantities of interest (using FFTs), and the possibility of systematically converging the accuracy of the NGWFs by varying a single parameter, the spacing between psinc functions; this convergence is variational with respect to the total energy, and is approximately equivalent to varying the kinetic energy cut-off employed in the plane-wave

[11]The periodicity of the psinc functions is a consequence of the discrete sampling of plane waves in reciprocal space; in fact, they can be obtained for a 1D system by discretising the integral in Eq. 3.42.

method [99, 100], since higher plane-wave energies are needed to describe a more closely spaced psinc grid.

ONETEP has been shown to scale linearly with the number of atoms in the system, and approximately as N_p^{-1} when parallelising the calculation over N_p cores [93, 95, 101, 102]; this has allowed calculations of up to tens of thousands of atoms on fewer than 100 cores [102]. The code has been used to study a wide range of systems, from crystalline solids and nanostructures to biological molecules [101].

We have implemented the calculation of phonon frequencies in ONETEP using a finite displacement method; our module is fully parallelised and integrated with the rest of the code, and allows for the automatic calculation of thermodynamic quantities (the free energy, entropy, internal energy, and specific heat as a function of temperature). A full description of our implementation is given in Appendix A; we also present our results for a 1000-atom supercell of bulk silicon, showing the phonon dispersion curves and specific heat calculated with ONETEP to be in good agreement with experimental measurements. Finally, we discuss the possibility of using the information from finite displacement calculations in large supercells with ONETEP to build model dynamical matrices for calculating the phonon frequencies of extremely large systems that cannot be simulated directly.

Chapter 4

Point defects in crystalline materials

4.1 Classification of defects

The defining feature of a crystalline solid is its symmetry: by assuming an infinite crystal of perfectly repeating units we can use this symmetry to simplify our mathematical analysis of its properties, as shown in Sec. 3.2. It is obvious that this assumption cannot be entirely true, since a real crystal has a finite macroscopic size. Of more importance, however, are the naturally occurring imperfections within the crystal on the atomic and mesoscopic scales. These are referred to as crystalline defects; the importance of their effects in real systems has long been recognised, and a discussion on the subject is a standard feature of most textbooks on solid state physics [55, 103, 104].

Crystalline defects can be classified according to their dimension as point defects (0D), line defects (1D), plane defects (2D), or bulk defects (3D). Point defects are not extended in space in any dimension, and so are considered to involve a local irregularity in the crystal lattice, generally involving only a few atomic sites. The basic types of point defects are:

- vacancies, given by the absence of an ion on a site that is occupied in the perfect crystal;
- interstitials, given by the presence of an ion on a site that is unoccupied

in the perfect crystal;

- substitutionals, given by the replacement of an ion in the lattice with one of a different species.

Furthermore, we can distinguish between native (or intrinsic) defects, that only involve atoms of the pure crystal, and impurity-related (or extrinsic) defects, that introduce foreign atoms into the lattice. Point defects can also combine into a complex, such as an interstitial defect with a nearby vacancy. However, the formation of a large cluster of defects with a regular structure should be considered a bulk defect (e.g., a collection of vacancies in neighbouring lattice sites produces a void).

The standard notation for describing point defects is that of Kröger and Vink [105]; in this system, defects are defined as M_S^C, where M is the symbol of the defect species, S is the lattice site occupied by the defect, and C is the charge state of the defect centre. M can be an atomic symbol or v for a vacancy. Similarly, S can be an atomic symbol (for a vacancy or substitutional defect) or i for an interstitial defect. C can be a cross ($^\times$) for a neutral defect or a number of dots ($^\cdot$) or dashes ($'$) to indicate the amount of net positive or negative charge at the defect site (in units of e), respectively. For example, $\mathrm{v}_{\mathrm{Si}}^{\cdot\cdot}$ refers to a doubly positive vacancy in silicon, while $\mathrm{Au}_{\mathrm{Si}}^{\times}$ refers to a neutral gold substitutional defect in silicon. An equally valid notation for C, which shall be used henceforth, is to write the charge state numerically ($\mathrm{v}_{\mathrm{Si}}^{2+}$ and $\mathrm{Au}_{\mathrm{Si}}^{0}$ for the two previous examples).

Defect complexes can also be treated with this notation, by enclosing the individual defects in brackets and denoting the overall charge of the complex outside the brackets; for example, a singly negative complex made from a vacancy and a self-interstitial in silicon is written $(\mathrm{v}_{\mathrm{Si}}, \mathrm{Si}_{\mathrm{i}})^{1-}$. Each individual defect within the complex may additionally be labelled with its contribution to the total charge, similarly to an ionic compound.

Finally, we note that point defects can be classified as stoichiometric or non-stoichiometric, depending on whether they preserve the stoichiometry of the perfect crystal. Stoichiometric defects must necessarily be native, and can be subdivided into three categories:

- Frenkel defects (also known as Frenkel pairs), given by a complex made from a vacancy and an interstitial of the same species carrying opposite charges, thus ensuring the overall charge neutrality of the system (e.g., $\left(v_{Ag}^{1-}, Ag_{i}^{1+}\right)^{0}$ in AgCl);

- Schottky defects, given by a complex made from multiple vacancies of all the species present in the crystal (in their stoichiometric proportions) with an overall neutral charge (e.g., $\left(v_{Ti}^{4-}, 2v_{O}^{2+}\right)^{0}$ in TiO_2);

- antisite complexes[1], made from multiple substitutional defects resulting from the exchange of positions of different species in the crystal lattice (e.g., $\left(As_{Ga}^{3-}, Ga_{As}^{3+}\right)^{0}$ in GaAs).

Frenkel defects can be thought of as being created by moving an ion from its lattice site into an interstitial position, while for Schottky defects this happens by moving a group of ions from their original lattice sites to new sites at the surface of the crystal, thereby increasing its volume. The latter, therefore, decrease the density of the crystal, while the former have little effect on it. In the case of a monatomic crystal such as silicon, the situation is somewhat simpler: Schottky defects can be considered to be single neutral vacancies, and antisites do not exist; furthermore, neutral self-interstitials are not required to form a Frenkel pair to maintain stoichiometry.

4.2 Defect concentration at equilibrium

All crystals at finite temperature contain a finite concentration of defects; in fact, even extremely pure crystals with a negligible percentage of impurities will feature some or all of the native defects presented in the previous section. Any defect introduced into a crystalline lattice will increase its energy, and so there is an activation energy associated with its formation. Therefore, we can expect more defects to form at higher temperatures, and the corresponding equilibrium defect concentration to be greater. Furthermore, once formed the

[1] A single antisite defect is, however, not stoichiometric.

defect will have a characteristic diffusion speed, also dependent on temperature, allowing it to either migrate to the surface of the crystal or recombine with its 'complementary' defect in order to reform the pristine lattice. A rapid cooling of the crystal will result in the defect being frozen into position.

The temperature-dependent concentration of different types of defects can be estimated by considering the crystal to be in thermodynamic equilibrium; in this case, the state of the system is found by minimising the appropriate thermodynamic potential. Most experiments are done at constant temperature and pressure, and so this will be the Gibbs free energy,

$$G = U - TS + PV, \tag{4.1}$$

where the symbols have their usual meaning. Taking into consideration only the simplest defect, the single vacancy, and a monatomic lattice, we can think of the crystal as being composed of a total of N_0 lattice sites, of which N are occupied and n are vacant. The total volume, therefore, can be approximated as $N_0 v_0 = (N + n)\, v_0$, where v_0 is the volume per ion in the pristine lattice.

If n is much smaller than N, it is reasonable to assume that the internal energy depends only on the number of vacancies but not their arrangement, since it will be unlikely for two defects to be found close together; this is known as the dilute limit, since each defect is considered to be effectively isolated in an otherwise perfect lattice. We note that n is nevertheless an extensive thermodynamic variable, since the ratio $\frac{n}{N_0}$ (the defect concentration) is finite.

The entropy is composed of two contributions, vibrational (S^{vib}) and configurational (S^{conf}). The configurational term is simply $S^{\mathrm{conf}} = k_{\mathrm{B}} \ln W$, where k_{B} is Boltzmann's constant and W is the number of distinguishable configurations of the system. For $n = 0$ (the perfect crystal), $W = 1$ and so $S^{\mathrm{conf}} = 0$; in general, W is equal to the number of ways of arranging the host ions and vacancies in the $N + n$ available lattice sites:

$$W = \frac{(N + n)!}{N!n!}. \tag{4.2}$$

The total change in the free energy of the system when creating n defects

is then

$$\Delta G = n\Delta u - nT\Delta s^{\mathrm{vib}} - k_{\mathrm{B}}T\ln\frac{(N+n)!}{N!n!} + nPv_0, \tag{4.3}$$

where Δu is the difference in internal energy between a pristine system and one with an isolated defect, and Δs^{vib} is the corresponding difference in vibrational entropy. The number of defects at thermodynamic equilibrium is

$$\frac{\mathrm{d}\Delta G}{\mathrm{d}n} = 0; \tag{4.4}$$

this gives

$$\Delta u - T\Delta s^{\mathrm{vib}} - k_{\mathrm{B}}T\frac{\mathrm{d}}{\mathrm{d}n}\ln\frac{(N+n)!}{N!n!} + Pv_0 = 0. \tag{4.5}$$

Using Stirling's approximation ($\ln N! \simeq N\ln N - N$ for large N), we obtain:

$$\Delta u - T\Delta s^{\mathrm{vib}} - k_{\mathrm{B}}T\ln\frac{(N+n)}{n} + Pv_0 \simeq 0, \tag{4.6}$$

and, hence, the temperature-dependent equilibrium defect concentration $c\,(T)$ is given by

$$c\,(T) = \frac{n}{N_0} \simeq \mathrm{e}^{(-\Delta u + Pv_0)/k_{\mathrm{B}}T}\mathrm{e}^{\Delta s^{\mathrm{vib}}/k_{\mathrm{B}}}. \tag{4.7}$$

It should be noted that, similarly to the entropy, the internal energy can also be divided into two contributions: $\Delta u = E_f + \Delta u^{\mathrm{vib}}$, where E_f is known as the defect formation energy (i.e., the difference in structural energy), while Δu^{vib} is the difference in vibrational energy. The contributions from both the vibrational terms and Pv_0 are usually small compared to that from E_f [55], and so these terms are often neglected, leaving only

$$c\,(T) \simeq \mathrm{e}^{-E_f/k_{\mathrm{B}}T}. \tag{4.8}$$

The calculation of the formation energy for different defects shall be discussed in detail in Sec. 4.4.1.

Due to the assumption of non-interacting defects, Eq. 4.8 can be trivially generalised to consider any number of different types of defects in the same crystal. Thus, if we consider each type of defect i to have n_i occurrences, the

corresponding equilibrium concentration is[2]

$$c_i\left(T\right) = \frac{n_i}{N_0} \simeq N_i \mathrm{e}^{-E_{f,i}/k_\mathrm{B}T}, \tag{4.9}$$

where N_i is the number of inequivalent configurations in which the defect can be incorporated on or around the same lattice site (since now $W_i = N_i \frac{(N+n)!}{N!n!}$). For simple defects, such as vacancies or substitutionals in which no symmetry breaking occurs, $N_i = 1$.

It is important to consider the validity of the assumption of thermodynamic equilibrium, as the defects are created during the growth of the crystal, which is clearly a non-equilibrium process. However, it is not necessary for all the dynamical processes involved to be in equilibrium, but only for those relating to the formation and migration of defects to be close to equilibrium; this will be achieved if the mobility of defects is high enough to allow their concentration to equilibrate [106].

Defect concentrations are typically very low, although, due to the exponential relationship given by Eq. 4.8, can vary over many orders of magnitude. Defect formation energies are on the order of a few electronvolts (the same as the dissociation energy of a chemical bond), and typical growth temperatures for semiconducting crystals range between 800–1100°C; therefore, we can estimate a lower limit for the typical defect concentration in such crystals as $c \sim \mathrm{e}^{-1\mathrm{eV}/\left(9\times10^{-5}\mathrm{eVK}^{-1}\right)\left(10^3\mathrm{K}\right)}$, giving a value on the order of 10^{-5} (ten defects per million atoms of the crystal), thus justifying the assumption of working in the dilute limit. We note that experimental studies of the self-interstitial in silicon have estimated the defect formation enthalpy ($E_f + \Delta u^\mathrm{vib} - Pv_0$ in our analysis) for this particular defect to be 3.8 eV, and the vibrational entropy contribution Δs^vib to be 9.9 k_B, resulting in a concentration orders of magnitude lower than our simple estimate [107]. Similar values have been reported for the silicon vacancy, although the experimental evidence is less clear [108, 109]. The large differences in concentration caused by small differ-

[2] We note that this analysis is valid only if there are no constraints between the number of defects of each type; the presence of any constraint, most notably the overall charge neutrality of the system, will reduce the concentration of the most abundant defect type and raise that of the most abundant 'complementary' type in order to satisfy it.

ences in formation energy generally lead to one type of defect being dominant in a particular crystal; silicon is a notable exception, since vacancies and self-interstitials both play an important role [109, 110].

4.3 Doping in semiconductors

The behaviour of semiconducting materials doped with impurities has been the subject of thorough and ongoing research since the discovery of the transistor in the late 1940s [111]. This is because point defects strongly influence the electrical conductivity of semiconductors by adding states in the forbidden band gap region between the occupied valence band and unoccupied conduction band, thus changing the number of charge carriers available (either electrons in the conduction band or hole quasiparticles in the valence band). Therefore, it is possible to accurately control the electrical properties of semiconducting materials such as silicon and germanium by adjusting the type and concentration of impurities added to the pure (intrinsic) crystal to make a doped (extrinsic) crystal.

The number of charge carriers is accurately described by the application of Fermi-Dirac statistics to the single-particle energy levels of the crystal's band structure. In an intrinsic semiconductor, the number of electrons per unit volume in the conduction band n and the that of holes in the valence band p, must be equal; using this relationship, it is straightfoward to derive an expression for the position of the electronic chemical potential μ_e in the band gap:

$$\mu_e = \varepsilon_v + \frac{1}{2}E_g + \frac{3}{4}k_{\mathrm{B}}T \ln \frac{m_h^*}{m_e^*}, \tag{4.10}$$

where ε_v is the valence band edge, E_g is the size of the band gap, and m_h^* and m_e^* are the effective masses for the hole and conduction-band electron, respectively (these depend on the density of states close to the band edges). Since $\frac{m_h^*}{m_e^*} \simeq 1$, the last term in the expression is small, and so the chemical potential remains very close to the centre of the band gap. Following from this result, typical charge carrier concentrations at room temperature can be estimated to be on the order of 10^{15} cm^{-3} (one free electron or hole per ten

million atoms of the crystal). Similarly to the analysis of defect concentrations in the previous section, the number of charge carriers increases exponentially with temperature.

The problem with intrinsic semiconductors is the inherent symmetry between positive and negative charge carriers, since most practical applications require an imbalance between the two[3]; this is achieved if a significant number of impurities (or dopants) are present. Typically, dopants can be classified either as electron donors or acceptors. Donors can become positively charged by donating electrons to the system, resulting in n-type doping (since negative charge carriers are created). Conversely, acceptors can become negatively charged by accepting electrons from the system, resulting in p-type doping (since positive charge carriers are created). In the case of silicon, which is a Group IV element, the most common n-type dopants are Group V elements such as phosphorus and arsenic, while the most common p-type dopants are Group III elements such as boron and aluminium.

The ionised dopant will act as either an attractive or repulsive centre with a charge Q opposite to that of the charge carriers it has contributed to the system. Therefore, it gives rise to an effective hydrogen-like potential of the form

$$V\left(r\right) = -\frac{Q}{\epsilon_r r}, \tag{4.11}$$

where ϵ_r is the relative dielectric constant of the material it is embedded within. ϵ_r can be quite large for semiconductors (≈ 12 for silicon) due to their small band gap and, therefore, the potential is substantially attenuated from that of the impurity atom in free space. Furthermore, an electron responding to this potential should be treated as a free quasiparticle with an effective mass m_e^*, rather than the free electron mass. The radius of its ground state orbit is, therefore,

$$r_0 = \frac{\epsilon_r}{Q m_e^*}, \tag{4.12}$$

[3]The conductivity of the semiconductor, however, is not zero even when the negative and positive charge carrier concentrations are equal, since electrons and holes have a different mobility in the lattice.

and the corresponding binding energy is

$$E_b = \frac{Q^2 m_e^*}{2\epsilon_r^2}. \tag{4.13}$$

Using the above value of ϵ_r for silicon, $Q = 1$ and an effective mass of 0.2, we see that the radius of this orbit is 60 $a_0 \simeq 32$ Å (much larger than the interatomic spacing, thereby justifying the use of the dielectric constant of the medium and the quasiparticle effective mass), and the binding energy is 7×10^{-4} Ha $\simeq 0.02$ eV, which is indeed on the same order as experimentally observed values. Therefore, $E_b \ll E_g$ in most semiconductors, and so the effect of the impurity ion on the band structure is to add localised defect levels within the band gap, close to one of the band edges. A donor impurity will add an occupied state below the conduction band minimum, from which electrons can be excited into the conduction band, creating negative charge carriers. Instead, an acceptor impurity will add an empty state above the valence band minimum into which electrons can be excited, leaving behind positive charge carriers. In other words, although the defect levels themselves are localised and, hence, not conducting[4], the introduction of the dopant shifts the position of the electronic chemical potential from its intrinsic position as given by Eq. 4.10 so as to increase the concentration of one type of carrier (the majority carrier), and decrease that of the other type (the minority carrier). In practice, it is common for both donor and acceptor impurities to be present in different concentrations; this general situation is illustrated in the schematic density of states diagram in Fig. 4.1.

Not all impurity-related defects in semiconductors can be considered as dopants; in fact, species with significantly different valence chemistry from that of the host can provide a more severe disruption to the lattice (both in terms of the electronic structure and the ionic configuration around the defect site), resulting in a number of defect levels in the band gap positioned far from the band edges, in certain cases even close to the middle of the gap.

[4]In fact, due to the small overlap between the wavefunctions of defect levels at different impurity sites, there is the possibility of conduction even at zero temperature due entirely to the tunnelling of electrons between defects. This phenomenon is known as impurity band conduction.

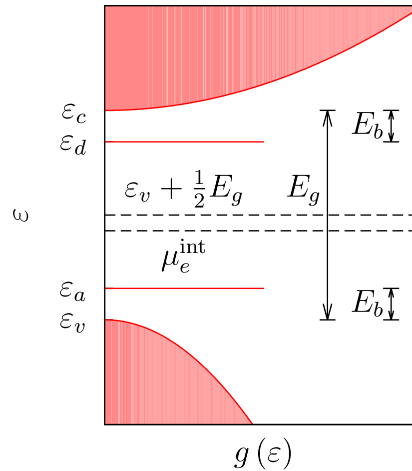


Figure 4.1: Density of states for a doped semiconductor, featuring a donor level ε_d and an acceptor level ε_a close to the band edges. E_g is the band gap and E_b the binding energy for the impurities, as estimated in Eq. 4.13. The position of the middle of the band gap and the intrinsic electronic chemical potential μ_e^{int} are also shown. The energy differences E_b and $\varepsilon_v + \frac{1}{2}E_g - \mu_e^{\mathrm{int}}$ have been exaggerated.

These are known as deep-level defects or traps; such defects interfere with the useful doping processes in extrinsic semiconductors by trapping charge carriers (since the ionisation energy for a deep level is generally $\gg k_{\mathrm{B}}T$), and can consequently lead to a significant decrease in the conductivity of the system [111, 112]. In general, transition metal impurities are a source of deep levels in semiconducting materials, as well as lattice vacancies and interstitials.

4.4 Simulation of defect properties

The study of point defects in crystalline solids using first-principles electronic structure simulations has received a considerable amount of attention in recent years (for a review, see Ref. [106]). From the point of view of electronic structure, the introduction of a point defect destroys the translational symmetry of

the lattice and the concept of a band structure, taking us back to the infinite problem in real space and a Γ-point only calculation in reciprocal space.

Consequently, the two most common approaches are to embed the point defect either in a finite cluster or a periodically repeating supercell of the host material. Both methods have advantages and disadvantages. Cluster calculations avoid interactions between periodic images of the defect, which, as shall be discussed in detail and demonstrated in our calculations, are the most significant drawback of the supercell method. However, a similar problem is introduced by the interaction between the defect and the surfaces of the cluster; in general, the finite size of the system will produce quantum confinement effects that are known to significantly affect its band structure and electronic properties [113,114] (in fact, studies of doped nanocrystals have shown noticeable deviations from bulk values, as well as completely novel effects [115,116]). Furthermore, in order to study the eigenvalue levels originating from the defect states it is necessary to eliminate defect-like surface states resulting from broken bonds, either by surface reconstruction or hydrogenation; this means that a large number of atoms will be needed for the surface layers, which will add to the computational expense while not contributing to the bulk-like environment that we are trying to create for the defect.

The supercell method, instead, correctly describes the band structure of the host crystal and its bulk properties independently of the supercell size. Indeed, this approach is ideal when the electronic structure associated with the defect is highly localised, since k-point sampling techniques can be employed to accurately describe the extended bulk-like states (and, hence, various important properties of the crystal in relation to its response to point defects, such as its permittivity), but will not affect the description of the defect levels themselves, since a perfectly localised level will produce an energy band which is flat throughout the entire Brillouin zone. Unfortunately, this is never the case: the defect band will always feature some dispersion, caused by the interaction between defect centres. In practice, we are dealing exactly (within the accuracy of the underlying electronic structure method) with an unrealistic situation, that of a material with an extremely high concentration of defects, typically 3–10 orders of magnitude larger than experimentally observed val-

ues, distributed periodically instead of randomly. We encounter an additional problem when simulating charged defect centres: there is no absolute reference for the electrostatic potential in periodic structures, which is needed in this case to calculate the electronic chemical potential (see Eq. 4.15); this is not a problem for cluster calculations, since it possible to define an unambiguous zero of potential.

We also note an alternative approach that is sometimes employed, that of Green's function scattering methods [117–121]. The Green's function is first calculated for the host material and then used to determine perturbative changes caused by the defect centre; therefore, the band structure of the pristine lattice is well described, and interactions with images of the defect are avoided. However, the method is challenging to apply to large systems, and the results are harder to interpret than for supercell calculations. The simplicity and relative accuracy of the supercell method, combined with the ongoing development and increasing popularity of many efficient plane-wave DFT codes that use periodic boundary conditions, have therefore resulted in this method becoming the most commonly used approach for defect calculations.

Another important consideration is the level of theory used to describe the electronic structure of the system. As described in Chapters 1–3, this can range from empirical tight-binding methods to post-Hartree-Fock (quantum chemistry) methods. Empirical tight-binding calculations suffer from the problem of parametrising the matrix elements for orbitals close to the defect site, whose value may deviate substantially from bulk regions. An incorrect parametrisation can lead to qualitatively wrong conclusions on the position and nature of the defect levels, as demonstrated by tight-binding and DFT studies of native defects in gallium nitride [122–124]. More advanced, first-principles tight-binding methods, in which the matrix elements are calculated explicitly, overcome the parametrisation problem, although the localised orbitals must be chosen carefully in order to provide a good description of the defect environment; for the most accurate results, a self-consistent field (SCF) procedure can be used [125].

As already discussed, plane-wave pseudopotential DFT has many desirable properties: it is unbiased and systematically convergeable, and is known to

perform well in solids while remaining reasonably computationally inexpensive; however, the disruption of the uniform bulk electronic structure at the defect site provides a challenging test of the transferability of the pseudopotentials and accuracy of the exchange-correlation functional. Studies of point defects have generally found local and semi-local functionals to be in good qualitative agreement [126, 127]; however, both types of functional suffer from the 'DFT band gap problem' [128–130], whereby the fundamental energy gap of insulators and semiconductors is severely underestimated (by about 50%) by the Kohn-Sham band structure, which is clearly problematic for assessing the position of defect levels in the gap. Many approaches have been proposed to overcome this problem, ranging from simple post-processing corrections to the use of more exotic functionals and the rigorous theoretical treatment of excited states going beyond standard ground-state DFT (see Secs. 4.4.4 and 4.4.5 for further discussion on the subject).

The application of electronic structure methods going beyond the accuracy of DFT is not yet common for defect systems, although some studies using the diffusion Monte Carlo [46] (DMC) method have been presented, e.g., for the silicon self-interstitial [110, 131], the vacancy in diamond [132], and Schottky defects in magnesium oxide [133]. Although all of these studies report substantial differences between the DMC and LDA/GGA-DFT defect formation energy at the same system size (∼0.5–1 eV), the DMC calculations are unfortunately restricted to very small supercells (16–54 atoms) due to the increased computational cost of the method. For such system sizes, spurious defect–defect interaction effects are known to play a large role; therefore, the finite system size error is estimated from DFT calculations on larger supercells, and the value obtained is used to correct the QMC result (this is the case for all the cited studies). Furthermore, it is currently difficult to obtain ionic forces from QMC that allow for a structural optimisation of the system (although progress is being made in this repect [134]); instead, DFT must be relied on to provide an optimised structure as input for the QMC calculation, the accuracy of which is of course not guaranteed.

In the rest of this section we will describe the theoretical framework used in first-principles simulations to calculate the most important properties of point

defects in solids. We note that the discussion in Secs. 4.4.1–4.4.3, regarding the rigorous definition of defect formation energies, chemical potentials, and transition levels, is equally valid for all electronic structure methods; however, Secs. 4.4.4 and 4.4.5 will focus exclusively on an important theoretical issue in DFT in relation to defect calculations, that of band gap underestimation.

4.4.1 Formation energies

The key quantity of interest for a point defect is its formation energy E_f, as discussed in Sec. 4.2; however, it is not straightforward to define this rigorously. The defect formation energy can be loosely thought of as the energy required to 'embed' the defect in a perfect crystal, through a combination of adding, removing and rearranging atoms. In this manner we can define E_f (following Zhang and Northrup [135]) by considering the defect in a large volume of crystal that entirely contains the disruptions to the lattice. E_f is then the energy difference between the volume containing the defect and the same volume filled with bulk crystal, after accounting for the adding and removing operations needed to create the defect; for an uncharged defect, therefore, we obtain:

$$E_f = E^{\mathrm{def}} - E^{\mathrm{bulk}} - \sum_i \Delta n_i \mu_i, \tag{4.14}$$

where i labels the atomic species, Δn_i is the difference in the number of atoms between the defect and bulk configurations in the volume of crystal being considered, and μ_i is the chemical potential of the species.

For example, in the case of a vacancy we need to subtract from the bulk energy the contribution of the atom that is removed to create the unoccupied site. For a monatomic crystal, the chemical potential of the species is well-defined: it is simply equal to the energy per atom in the bulk crystal[5]. However, noticeable complications arise when considering a single vacancy in a crystal containing more than one species, since it is necessary to find a rigorous way

[5] This implies, in effect, that the missing atom is actually contributing to the creation of more crystal; indeed, this is the tacit assumption that was made in Eq. 4.3 with the addition of the term nPv_0. Furthermore, it is important to note that we are also assuming that all unit cells in the crystal are bulk-like in nature, since surface terms have not been considered.

of partitioning the energy of the unit cell between its constituent species. In fact, stoichiometric defects are the only class of defects for which the definition of chemical potentials is trivial: this is because either $\Delta n = 0$ for all species (since the defect is created simply by moving ions within the supercell, as in the case of Frenkel defects and antisite complexes), or all species have been removed in their stoichiometric proportions, thereby allowing a well-defined collective chemical potential (equal to the energy per unit cell) to replace the individual ones of Eq. 4.14

In the case of a substitutional impurity defect we need to both subtract the contribution of the bulk atom and add the contribution of the impurity atom that is introduced in its place. To do so we need to assign a chemical potential to the impurity species α. We can imagine a reservoir for this species from which we extract the atom; μ_α is then the energy per atom of the species in its lowest-energy bulk crystal structure. However, this approach contains no information of the formation conditions of the system; some knowledge of the process by which the defect is formed can generally be used to obtain a better estimate of μ_α. The issue of defining chemical potentials is discussed in more detail in Sec. 4.4.2.

If the defect is charged the situation is more complicated, since E_f also depends on the bulk electronic chemical potential μ_e. In this case, Eq. 4.14 becomes

$$E_f^q = E^{\mathrm{def},q} - E^{\mathrm{bulk}} - \sum_i \Delta n_i \mu_i + q\mu_e, \tag{4.15}$$

where q is the charge of the system. μ_e is, by definition, the Fermi level; however, for an insulator or semiconductor this is at an undefined position within the band gap (which will depend on the other defects in the crystal, as discussed in Sec. 4.3, and any external potentials applied to the system). The formation energy is therefore a function of the position of μ_e between the valence band maximum (VBM) and the conduction band minimum (CBM). Consequently, it is convenient to define the chemical potential in terms of its relative position to the VBM ε_v:

$$\mu_e = \varepsilon_v + \Delta\mu_e \tag{4.16}$$

As shall be discussed in Sec. 5.4, a correction term is usually applied to Eq. 4.16 to align the energy levels of the defect supercell with those of the bulk supercell.

4.4.2 Chemical potentials

As noted in the previous section, it is not straightforward to define the chemical potentials to use in Eq. 4.14 for a multi-component system; in fact, it is not possible to do so uniquely, as they depend on the experimental growth conditions. However, it is possible to place firm upper and lower bounds on their values independently of these conditions.

Since the chemical potential for a species represents the energy needed to remove or add one of its atoms during the formation of a defect, in order to define it we need to answer the following two questions: (i) what reservoir is the atom being removed to or taken from, and (ii) how much of the total energy of that reservoir can be attributed to this species alone. When considering one of the atomic species that constitute the host material, it is reasonable to assume that the reservoir of atoms is the material itself. Therefore, the chemical potential of a monatomic crystal is exactly equal to the energy per atom of the crystal. However, if more than one species is present, all that can be calculated is the energy per unit cell; this needs to be divided between the species. We can define a range within which the chemical potential of each species must lie; as for the case of charged defects, therefore, all formation energies should strictly be considered to be functions of the atomic chemical potentials as well as the electronic one. For example, in the case of gallium nitride [106], we have:

$$\mu_{\mathrm{GaN}} = \mu_{\mathrm{Ga}} + \mu_{\mathrm{N}}; \tag{4.17}$$

we can place upper limits on μ_{Ga} and μ_{N} by noting that for GaN to form, neither of these chemical potentials can be higher that the energy per atom of its species' elemental configuration. This means that

$$\mu_{\mathrm{Ga}} \leq \mu_{\mathrm{Ga}}^{\mathrm{bulk}} \tag{4.18}$$

and

$$\mu_{\mathrm{N}} \leq \mu_{\mathrm{N}}^{\mathrm{N}_2}, \tag{4.19}$$

where $\mu_{\mathrm{Ga}}^{\mathrm{bulk}}$ is the energy per atom of bulk Ga and $\mu_{\mathrm{N}}^{\mathrm{N}_2}$ that of an N_2 gas molecule. By combining these two inequalities with Eq. 4.17, we can also place lower limits on the chemical potentials:

$$\mu_{\mathrm{Ga}} \geq \mu_{\mathrm{GaN}} - \mu_{\mathrm{N}}^{\mathrm{N}_2} \tag{4.20}$$

and

$$\mu_{\mathrm{N}} \geq \mu_{\mathrm{GaN}} - \mu_{\mathrm{Ga}}^{\mathrm{bulk}}. \tag{4.21}$$

These upper and lower bounds correspond experimentally to extreme Ga-rich or N-rich growth conditions. Therefore, it might sometimes be appropriate to simply set one of the two chemical potentials to its upper bound, depending on the experimental setup that is being considered; otherwise, the defect formation energy can be plotted as $E_f(\xi)$ for $0 \leq \xi \leq 1$ and

$$\begin{cases} \mu_{\mathrm{Ga}}(\xi) = & \mu_{\mathrm{Ga}}^{\mathrm{bulk}} + \xi\left(\mu_{\mathrm{GaN}} - \mu_{\mathrm{N}}^{\mathrm{N}_2} - \mu_{\mathrm{Ga}}^{\mathrm{bulk}}\right) \\ \mu_{\mathrm{N}}(\xi) = & \mu_{\mathrm{GaN}} - \mu_{\mathrm{Ga}}^{\mathrm{bulk}} + \xi\left(\mu_{\mathrm{N}}^{\mathrm{N}_2} - \mu_{\mathrm{GaN}} + \mu_{\mathrm{Ga}}^{\mathrm{bulk}}\right) \end{cases}. \tag{4.22}$$

We should also note that these expressions are temperature and pressure dependent; while this might only have a negligible effect for bulk solids, it will certainly have a strong effect on gases such as N_2.

The 'richness' of one or the other component is described by the degree of non-stoichiometry of the compound (e.g., $\mathrm{Ga}_x\mathrm{N}_{1-x}$ in the example above, where x is close to the exact stoichiometric ratio $x_s = \frac{1}{2}$). For a given small deviation of x from $x_s = \frac{m}{m+n}$ in a general ordered $\mathrm{A}_m\mathrm{B}_n$ binary alloy with non-stoichiometric composition $\mathrm{A}_x\mathrm{B}_{1-x}$, Hagen and Finnis [136, 137] have shown that it is possible to derive from thermodynamic considerations analytical formulae that can be solved to find the chemical potentials of the two species and the point defect concentrations (as discussed in Sec. 4.2) of the two native vacancies and antisites as a function of x and temperature[6]. The derivation

[6]Interstitials are not taken into consideration in this analysis.

depends on identifying the defect type that accommodates the deviation from stoichiometry for A-rich or B-rich conditions (and, hence, the only defect type that persists at zero temperature); this is known as the constitutional defect. For A-rich (B-rich) conditions, the constitutional defect can be either v_B (v_A) or A_B (B_A); this can be determined by comparing the sum of formation energies of several defects forming (non-interacting) stoichiometric complexes, that can therefore be calculated unambiguously by atomistic simulation [137].

For impurities it is not as simple to define the reservoir; this depends on the physical provenance of the impurity atoms in the system of interest. A hard upper bound is given as before by the energy per atom of the species' elemental bulk phase. This upper bound can usually be reduced by considering the energy of other compounds that might form between the impurity and host species (e.g., the possible formation of Si_3N_4 when incorporating silicon into gallium nitride [138]). However, the most direct way of estimating the chemical potential is simply to consider the experimental setup on a case-by-case basis, and therefore determine the nature of the reservoir of impurity atoms; the energy per atom can then be calculated for this particular phase of the element.

4.4.3 Transition levels

The calculation of the positioning of the levels introduced by point defects in the band gap is fundamental for comparison with experimental results, as these are usually what is used to identify the defect. However, the experimentally measured levels do not correspond to the one-electron eigenvalues given by the Kohn-Sham band structure.

Following the careful description of Baraff *et al.* [139], we define the stable charge state transition level (also referred to as the thermodynamic transition level or occupancy level) $E\,(m/n)$ to be the value of the Fermi level with respect to the perfect crystal valence band edge at which there is a crossing of the defect formation energies of the two charge states m and n, leading to a change in the most stable state from m to n as the Fermi level is raised. We can calculate the stable charge state transition levels by equating the expression

in Eq. 4.15 for two different charge states and solving to find the value of $\Delta\mu_e$ at which their defect formation energy is equal. In general, this is given by

$$\begin{aligned} E\left(m/n\right) =& \frac{E_f^n\left(\Delta\mu_e = 0\right) - E_f^m\left(\Delta\mu_e = 0\right)}{m-n} \\ =& \frac{E^{\mathrm{def},n} - E^{\mathrm{def},m} - \left(m-n\right)\varepsilon_v}{m-n}; \end{aligned} \tag{4.23}$$

therefore, $E\left(m/n\right)$ is simply the energy difference between the two states of a system composed of a point defect and an electron reservoir, where the defect exchanges $(m-n)$ electrons with the reservoir during the transition[7]. In the case of a shallow defect level, the $E\left(1+/0\right)$ transition for a donor or the $E\left(0/1-\right)$ transition for an acceptor is usually called the thermal ionisation energy or the donor/acceptor ionisation energy.

The stable charge state transition levels can be represented graphically by plotting the defect formation energy of each charge state as a function of the electronic chemical potential. Each charge state will produce a linear function with a slope equal to its charge. The transition levels are given by the various crossings present; the overall stable state is the one with the lowest formation energy at a given Fermi level position. In general the stable state will change across the band gap, as shown in Fig. 4.2.

In the above definition, is it implied that the ionic configuration around the defect centre is allowed to relax independently for the initial and final state of the transition; in fact, the equilibrium configuration will change, sometimes substantially, between charge states. Therefore, these levels correspond to those measured in quasi-equilibrium experiments such as deep-level transient spectroscopy [141] (DLTS) and temperature-dependent measurements of resistivity and Hall data[8].

Optical transition levels can also be defined using Eq. 4.23, although the energy of the final state $E^{\mathrm{def},n}$ is calculated using the relaxed configuration of

[7]The energy difference is divided by the number of electrons exchanged, to give the cost per electron of the transition.

[8]More precisely, the experimentally measured quantity is the activation energy of the transition; this might be greater than the stable charge state transition level due to the presence of an energy barrier between states.

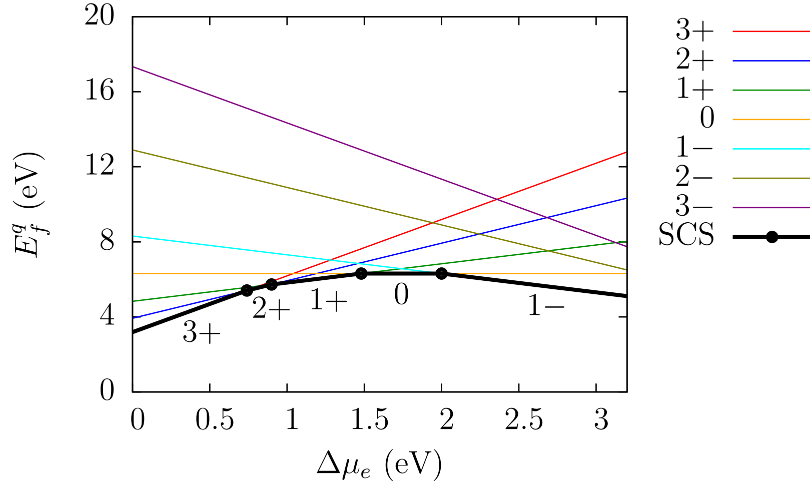


Figure 4.2: An example of how the stable charge state for a given defect changes depending on the position of the Fermi level in the band gap. E_f is the defect formation energy and $\Delta\mu_e$ is the electronic chemical potential (i.e. the Fermi level) relative to the VBM. The size of the band gap is 3.2 eV. The thick black line indicates the stable charge state (SCS), and the circles mark the transition levels between different stable states. The data was taken from plane-wave DFT calculations on the N_i defect in wurtzite GaN by Limpijumnong and Van de Walle [140].

the initial state; this is valid for experiments in which the system cannot relax during the transition process, such as photoluminescence (PL). Therefore, the optical absorption energy $E^{\text{opt}}(q/q+n)$ due to the excitation of an electron from the defect level into the CBM ($n=+1$), or the optical emission due to the inverse recombination process ($n=-1$), is given by

$$\begin{aligned} E^{\text{opt}}(q/q+n) =& E_f^{q+n}(\Delta\mu_e = E_g) - E_f^q(\Delta\mu_e = E_g) \\ =& E^{\text{def},q+n} - E^{\text{def},q} + n(\varepsilon_v + E_g), \end{aligned} \tag{4.24}$$

where both formation energies are calculated using the equilibrium structure for charge q. Analogously, the absorption energy $E^{\text{opt}}(q/q-n)$ due to the excitation of an electron from the VBM into the defect level ($n=+1$), or the

corresponding optical emission ($n = -1$), is given by

$$\begin{aligned} E^{\text{opt}}\left(q/q-n\right) =& E_f^{q-n}\left(\Delta\mu_e = 0\right) - E_f^{q}\left(\Delta\mu_e = 0\right) \\ =& E^{\text{def},q-n} - E^{\text{def},q} - n\varepsilon_v. \end{aligned} \tag{4.25}$$

Hence, the optical absorption energy is greater than the stable charge state transition level for the same transition by an amount E^{rel}, equal to the difference in energy for the final charge state in the ionic configuration of the initial state and its fully relaxed configuration; this is known as the Franck-Condon shift [142–144].

4.4.4 The band gap problem in DFT

The calculation of defect formation energies using Eq. 4.15 and stable charge state transition levels using Eq. 4.23 is subject to a number of diverse finite size errors; these arise from the use of the supercell approach, and are therefore present regardless of the level of theory employed in the electronic structure calculations. We discuss such finite size effects in Chapters 5 and 6.

However, the use of DFT presents an additional problem for the calculation of defect formation energies, and, hence, transition levels; this is the well-known band gap problem for insulators and semiconductors, i.e., the fact that the single-particle gap E_g^{KS}, given by the energy difference between the Kohn-Sham eigenvalues at the valence and conduction band edges, is observed to be much smaller than the experimentally measured (fundamental) gap. However, it is not possible in general to interpret the Kohn-Sham single-particle energies in a physically meaningful way; specifically, they cannot be equated with the quasiparticle excitation energies that define the fundamental band gap E_g^{QP}. Indeed, it is known that the single-particle band gap underestimation is an inherent feature of DFT [128–130], and is not a spurious effect caused by the commonly-used approximations to the exchange-correlation functional; in fact, the 'true' Kohn-Sham band structure resulting from the exact functional also features such an underestimation. This is due to the existence of a derivative discontinuity in the exact exchange-correlation potential when an electron is

added to the extended system, defined as

$$\Delta^{\mathrm{xc}} = v^{\mathrm{xc},N+1}\left(\mathbf{r}\right) - v^{\mathrm{xc},N}\left(\mathbf{r}\right), \tag{4.26}$$

where $v^{\mathrm{xc},M}\left(\mathbf{r}\right)$ is the exact exchange-correlation potential for an M-electron system, and the difference Δ^{xc} is independent of $\mathbf{r}$. The effect of the discontinuity is to apply a rigid shift to the single-particle band structure upon addition of an electron; the fundamental gap is, therefore,

$$\begin{aligned} E_g^{\mathrm{QP}} =& \varepsilon_{N+1}^{\mathrm{KS},N+1} - \varepsilon_N^{\mathrm{KS},N} \\ =& \varepsilon_{N+1}^{\mathrm{KS},N} + \Delta^{\mathrm{xc}} - \varepsilon_N^{\mathrm{KS},N} = E_g^{\mathrm{KS}} + \Delta^{\mathrm{xc}}, \end{aligned} \tag{4.27}$$

where $\varepsilon_i^{\mathrm{KS},M}$ is the i-th Kohn-Sham eigenvalue of the M-electron system. Approximate analytic functionals such as the LDA or GGA do not include the effect of the derivative discontinuity; however, calculations have shown the discontinuity to be a substantial component of the quasiparticle band gap [145]. For example, $\Delta^{\mathrm{xc}} = 0.58$ eV in silicon [146], accounting for approximately half of the experimental gap of 1.16 eV; the exact Kohn-Sham single-particle gap, therefore, is $E_g^{\mathrm{KS}} = 1.16 - 0.58 = 0.58$ eV, which is in reasonable agreement with that calculated with the LDA (0.52 eV).

The fundamental gap can also be defined as the difference between the ionisation potential I and the electron affinity A:

$$E_g^{\mathrm{QP}} = I - A = \left(E_0^{N-1} - E_0^N\right) - \left(E_0^N - E_0^{N+1}\right), \tag{4.28}$$

where E_0^M is the ground-state energy of the M-electron system. The connection between Eqs. 4.27 and 4.28 is given by Janak's theorem [147], which states that the variation of the DFT total energy with respect to an orbital occupation f_i is equal to the Kohn-Sham eigenvalue of the orbital:

$$\frac{\partial E_0}{\partial f_i} = \varepsilon_i^{\mathrm{KS}}. \tag{4.29}$$

For the true exchange-correlation functional, it has been shown that the total energy in the grand canonical ensemble (allowing for a non-integer number of

electrons) is a linear piecewise function, connecting the value of the energy at integer electron numbers with straight lines [148, 149]. It follows, therefore, that

$$\left.\frac{\partial E_0^M}{\partial M}\right|_{N-\delta} = \varepsilon_N^{\mathrm{KS},N} = -I \tag{4.30}$$

and

$$\left.\frac{\partial E_0^M}{\partial M}\right|_{N+\delta} = \varepsilon_{N+1}^{\mathrm{KS},N+1} = -A \tag{4.31}$$

(where $N \pm \delta$ is the total number of electrons slightly above and below N, respectively), and, hence, that the two definitions of the fundamental gap are consistent.

It is important to note that Eqs. 4.30 and 4.31 are not valid for finite system sizes when using approximate functionals, since they no longer obey the correct piecewise behaviour. Instead, functionals that are explicitly dependent on the density, such as the LDA and GGA, exhibit a convex behaviour as a function of M between integer electron numbers, and so the magnitude of the derivative $\left.\frac{\partial E_0^M}{\partial M}\right|_{N-\delta}$ (and, consequently, that of the Kohn-Sham eigenvalue $\varepsilon_N^{\mathrm{KS},N}$, since Janak's theorem is still valid) underestimates the energy difference $E_0^{N-1} - E_0^N$. Similarly, the magnitude of the derivative $\left.\frac{\partial E_0^M}{\partial M}\right|_{N+\delta}$ (equal, in the case of convex functionals, to the Kohn-Sham eigenvalue $\varepsilon_{N+1}^{\mathrm{KS},N}$ [150]) overestimates the energy difference $E_0^N - E_0^{N+1}$. Therefore,

$$I - A = \varepsilon_{N+1}^{\mathrm{KS},N} - \varepsilon_N^{\mathrm{KS},N} + \Delta^{\mathrm{convex}} = E_g^{\mathrm{KS}} + \Delta^{\mathrm{convex}}, \tag{4.32}$$

where $\Delta^{\mathrm{convex}} > 0$ is the correction due to the functional's convex behaviour for fractional charges [151, 152]. This suggests that $I - A$ should give a more accurate value of the fundamental gap than the difference between the highest occupied eigenvalue and lowest unoccupied eigenvalue in the N-electron system, which has indeed been shown to be the case for atoms and small molecules [127, 153].

Furthermore, Eqs. 4.28 and 4.32 suggest that, since the fundamental gap can be calculated as a difference of total energies, and is not equal to the single-particle gap, it might be possible to calculate it accurately for the case

of semiconductors even with a local or semi-local treatment of the exchange-correlation functional; unfortunately, this is not the case. In fact, it has been argued that $\Delta^{\mathrm{convex}} = 0$ for all functionals in the limit of infinite system size [128, 129]; it follows that for local and semi-local functionals the energy difference $I - A$ must converge to the single-particle gap E_g^{KS} as the system size is increased (since $\Delta^{\mathrm{xc}} = 0$), as confirmed by calculations on various semiconductors in LDA [154] and GGA [127]. This phenomenon can be viewed as a consequence of the poor description of the added electron or hole, which is incorrectly delocalised over the entire system by local and semi-local functionals due to their convex behaviour (caused, from a physical point of view, by the self-interaction error present in such functionals) [127, 151, 152]; this, in turn, spuriously lowers the total energies E_0^{N-1} and E_0^{N+1} with respect to E_0^N, thereby affecting both I and A.

4.4.5 Band gap correction schemes

The band gap problem, as discussed in the previous section, is a fundamental limitation of LDA/GGA-DFT. It is clear that the calculation of point defects in different charge states will be equally problematic, since the formation energies and transition levels will be affected by the charge delocalisation error. This realisation has led to many proposals for correction methods (for reviews, see Refs. [127, 155]), that can be broadly divided into two categories: post-LDA/GGA methods, and *ex post facto* corrections to the LDA/GGA results.

In the former category, the aim is to use a different level of theory for the calculation itself, in order to obtain a single-particle band structure that reproduces the fundamental gap and correctly places the defect levels within it; examples include the use of non-local (hybrid) exchange-correlation functionals [131, 156–163], adapted pseudopotentials [164, 165], the DFT+U method [127, 166–168], and the GW method [162, 169, 170]. In particular, numerous studies have made use of hybrid functionals, that aim to reduce the self-interaction error made by local and semi-local descriptions of exchange and correlation, and, hence, generally significantly improve the single-particle band gap with respect to experimental measurements. This is achieved by substituting a

portion of the exchange term from an LDA/GGA functional with either the exact exchange from Hartree-Fock theory (as for the B3LYP [171, 172] and PBE0 [173] functionals), or a short-ranged screened exchange (sX) term (as for the sX-LDA [174] and HSE [175] functionals). The accuracy of such functionals for the treatment of the electronic structure of point defects is still unclear, as is their relative performance (that could dictate the preference for a specific functional over others for particular types of defect). We note, however, that the HSE functional has shown good agreement both with DMC (in a study of the silicon self-interstitial [131]) and with the GW method (in a study of the Mg_{Ga} defect in gallium nitride [162]). The oxygen vacancy in hafnia provides a good test case for the use of hybrid functionals for defect studies, as calculations on this system have been carried out with sx-LDA [158], B3LYP [159], and PBE0 [160, 161]: in all cases, the band gap is significantly improved compared to the one calculated with the LDA or PBE (GGA), increasing it from 4.3 eV to 5.8–6.1 eV (the experimental gap is estimated at $\sim$5.9 eV); however, the positions of the defect levels for the charge states ranging from $+2$ to -2 vary between studies by $\sim$5–30% within the band gap (0.3–2.0 eV), and lead to qualitatively different descriptions of the stable charge state transition levels. The level of uncertainty between different hybrid functionals is comparable to the difference between PBE and PBE0, for which the level positions differ by $\sim$1.5 eV, mainly due to the band edge shift [161].

Similarly to the use of hybrid functionals, the DFT+U method can be employed to widen the single-particle band gap, by augmenting the exchange-correlation functional with screened Coulomb interactions on atomic sites, provided by 'Hubbard projectors' for specific angular momentum channels [54]. This is generally only applied to d and f orbitals; however, it has also been used for s and d orbitals in the study of the oxygen vacancy in zinc oxide [127], in order to raise the level of the unoccupied conduction band edge (by acting on the Zn s states), and lower that of the occupied valence band edge (by acting on the Zn d states). The strength of the potentials U_s and U_d were fixed empirically so as to recover the experimental band gap. The calculations, however, reveal a number of problems with this approach: as noted by the authors of the study, although the method arguably provides some insight of how the

defect levels should be placed within the quasiparticle gap, spurious errors in the description of the oxide introduced by the empirical correction to the band gap significantly reduce the predictive power of the method.

The GW method stands out amongst the post-LDA/GGA correction methods as the only approach that provides a formal connection to the electronic excitations of the system, and, hence, a rigorous theoretical framework for calculating the quasiparticle energies [176–178]. Encouragingly, GW calculations of the formation energy for the silicon self-interstitial in various configurations correct ∼70% of the discrepancy between LDA-DFT and DMC calculations [110, 131, 169]. Unfortunately, the GW method suffers from the same drawback as DMC: both are currently prohibitively expensive for large supercells.

In the latter category of band gap correction methods, instead, the aim is to apply a post-processing correction to the results obtained from a standard LDA/GGA calculation, in order to improve the defect levels positions. A common approach is to perform a 'scissors' operation on the band structure, effectively shifting the empty states so as to align the CBM with experiment. The effect this operation should have on the defect levels inside the gap is usually unclear. The defect levels can be thought of as being tied either to the valence band or the conduction band, depending on their character; by examining the wavefunctions of these states it should therefore be possible to decide whether to shift a particular level together with the CBM or keep it pinned to the VBM. A more rigorous approach is to expand the defect levels in terms of the defect-free host wavefunctions (which form a complete basis), and apply a perturbation to the Hamiltonian that correctly opens the band gap to the experimental value; however, this is process suffers from large and uncontrolled errors due to the difficulty of defining a physically reasonable perturbation [127]. Another approach is to apply an empirical correction by using a reference (or marker) defect [179, 180]. The idea is to calculate the transition levels in the same way for both the defect of interest and a standard defect for which the experimental values are well known; then, the difference between the calculated and experimental levels for the reference defect is used to correct the levels for the defect of interest. This scheme works best when

the two defects are very similar, as it depends crucially on the cancellation of errors between the two calculations.

In conclusion, it is interesting to note, from a pragmatic point of view, the inverse relationship that exists for any computational investigation of point defect between errors caused by the cluster/supercell size (finite size errors) and those caused by the limitations of the electronic structure method employed (theory errors, e.g., matrix element parametrisation errors in the tight-binding method, band gap and other exchange-correlation functional errors in DFT, fixed-node errors in DMC [46]). This can be understood by considering the following question: for a fixed computational cost, how can the accuracy of the calculated defect properties be maximised? Clearly, we might opt either for a large system size and an inexpensive (but imprecise) level of theory, or a small system size and a more accurate level of theory. In the former case, finite size errors will be small and theory errors will be large, while in the latter case, the reverse will be true. We could imagine, therefore, that the optimum approach depends on the relative magnitude of these two sources of error; furthermore, the accuracy and computational expense of possible correction methods can also be taken into account. It is clear, however, that the relative importance of different types of error is sensitively dependent on the nature of the specific defect under investigation, and so is hard to predict: for example, although they are both native deep-level defects in the same material, the silicon self-interstitial shows significant differences between different levels of theory [110,131,169], but exhibits a simple convergence with increasing system size [181], while the silicon vacancy exhibits non-trivial finite size effects, but is arguably in good agreement with available experimental results once these are properly treated (as shown in the next chapter).

Chapter 5

The silicon vacancy

In this chapter we present the results from our investigation of the silicon vacancy in bulk silicon (denoted v^q, where q is the charge state of the defect centre)[1] using DFT. The main quantities of interest that we calculate are the defect formation energy and stable charge state transition levels, as defined in the previous chapter.

The silicon vacancy is one of the most well-studied defects in semiconductors, and is often treated as a test case for theoretical and computational methods; nevertheless, it has been noted that even for the neutral defect centre, different studies have shown a large scatter in results for basic quantities such as the defect formation energy; this is generally attributed to finite size effects that lead to a particularly slow convergence of the quantities of interest with respect to system size.

For a point defect in an infinite crystal lattice the relaxation of atomic positions around the defect can extend to many successive shells of atoms. In simulations with period boundary conditions this gives rise to elastic interactions between the defect and its images in neighbouring supercells; the relaxation must therefore be contained within one supercell. Additionally, there is a spurious electrostatic interaction between a defect and its periodic replicas that depends on the size of the supercell.

Despite previous studies of the convergence of the defect formation energy,

[1] In the following chapters we neglect the subscript Si from the Kröger-Vink notation of defect centres.

there is still not a clear understanding of the relative importance of different sources of finite size error, and of the best approach to take in order to minimise them. Our aim, therefore, is to investigate systematically and quantify the main contributions to the total finite size error in the supercell approach, focusing on spurious electrostatic interactions, elastic interactions and wavefunction orthogonality constraints between periodic images of the defect centre. Our calculations demonstrate that the defect formation energy and stable charge state transition levels exhibit different convergence behaviour with respect to supercell size, depending on the Brillouin zone sampling used: the former benefits from the use of a dense k-point grid, while the latter from sampling at the Γ-point only. The reasons for this difference will be discussed in the text. In addition, we present the first calculations of maximally-localised Wannier functions for this defect centre, and show how they can be used for two different purposes: (i) as an accurate probe of localised properties of the system, in order to calculate the potential alignment correction to the valence band maximum of charged defect supercells, and (ii) as a way of analysing the chemical bonding around the defect centre, by finding an optimal visualisation of the electronic structure in real space. We relate the MLWFs associated with the defect centre for each charge state to the canonical LCAO model of the vacancy by Watkins, showing that the qualitative description given by this simple model is in full agreement with parameter-free DFT calculations.

The rest of this chapter is organised as follows: in Sec. 5.1, we review the previous experimental and theoretical results for the vacancy. In Sec. 5.2, we present Watkins' LCAO model of the system. In Sec. 5.3, we describe the computational techniques that we employ and give the technical details of our simulations. In Sec. 5.4, we illustrate the two methods we use to perform the potential alignment correction and show preliminary results comparing them. In Sec. 5.5, we present our main results; first we describe the convergence properties of the unrelaxed (Sec. 5.5.1) and relaxed (Sec. 5.5.2) neutral vacancy, and then we describe the results obtained for different charge states of the defect (Sec. 5.5.3) in terms of the transition levels (Sec. 5.5.4) and the MLWFs of the system (Sec. 5.5.5). In Sec. 5.6, we give a brief summary of our main conclusions.

5.1 Previous studies

The isolated single vacancy in bulk silicon is one of the two main native defects of interest for the material (the other being the self-interstitial). The vacancy is known to play an important role in self-diffusion and impurity diffusion [109, 182], and is readily produced both thermally [108, 109] and as a result of irradiation [183].

The vacancy and related defect centres in silicon have been studied experimentally using several different techniques (for a review, see Ref. [184]). Electron paramagnetic resonance (EPR) measurements [185] and, later, electron-nuclear double resonance (ENDOR) measurements [186, 187] on the charged states of the vacancy have given information on the symmetry and spatial distribution of the highest unoccupied electronic state localised at the defect site, leading to the adoption of the Watkins model (described in the following section) as the canonical description of the defect centre [183].

Of particular interest in the case of the silicon vacancy is the ordering of the stable charge state transition levels. As first predicted by Baraff *et al.* [139] and then verified by Watkins and Troxell [184, 188] on the basis of experimental measurements of the stable charge transition levels (given in Table 5.4) obtained by EPR [189,190] and DLTS [190–192], the isolated silicon vacancy exhibits a 'negative-effective-U' effect, by which the stable charge state changes directly from v^{2+} to v^{0} as the electronic chemical potential is raised; the v^{1+} state is therefore only metastable. This effect has been explained as a consequence of the lattice relaxation for different charge states, which lowers the formation energy of the neutral vacancy much more than the singly positive vacancy. It has also been suggested that the same effect might be observed in the sequence v^{0}, v^{1-}, v^{2-}.

Later studies of the vacancy and substitutional phosphorus complex in silicon (known as an E-centre), both by DLTS [193–196] and positron annihilation lifetime spectroscopy [197] (PALS), have given information on the change in volume of this defect centre between charge states, showing an outwards relaxation on electron emission for the $(\mathrm{P}, \mathrm{v})^{1-} \rightarrow (\mathrm{P}, \mathrm{v})^{0}$ transition. These studies have argued that, since phosphorus is a shallow electron donor in silicon, and

the two elements neighbour on the periodic table, the P substitutional has little effect on the lattice relaxation, which can then be attributed to the vacancy alone. Furthermore, assuming the donor electron from the phosphorus is trapped by the vacancy, the measured transition can be equated to $v^{2-} \rightarrow v^{1-}$ (i.e., resulting in the complete transition $(P^{1+}, v^{2-})^{1-} \rightarrow (P^{1+}, v^{1-})^{0}$). This conclusion is supported by a different PALS study [198], comparing the relative lattice relaxation of the single v^{1-} centre and the $(P, v)^{1-}$ complex. However, it has been noted that there are significant differences in the electronic structure of the two centres [199]. These experiments also provide no indication of the volume of the neutral defect centre, and so it is still unclear whether the vacancy causes an overall outwards or inwards relaxation of the host lattice around it.

The accurate determination of the vacancy formation enthalpy has also proven challenging, due to the difficulty of unambiguously interpreting the experimental data. In an early study of the E-centre, Watkins and Corbett estimated the formation enthalpy at 3.6 eV from EPR and ENDOR measurements [199]; one of the assumptions used to derive this value (the temperature-independence of the vacancy migration enthalpy) was later criticised by Van Vechten, who estimated it instead at 2.4 eV [200]. However, PALS measurements have also suggested a value of 3.6 eV, with a vibrational entropy contribution to the total defect concentration (from Eq. 4.9) on the order of 10 k_{B} [201].

Alongside the experimental investigations, many computational studies of the silicon vacancy have also been undertaken, using a variety of theoretical techniques (for reviews, see Refs. [202, 203]); these have yielded a large scatter in quantitative results, as well as qualitative disagreements between different methods and even between different studies using the same level of theory. Early calculations using interatomic force fields suggested an inwards relaxation around the defect centre [204, 205], while Green's function calculations have shown an outwards relaxation [206–208]. Later tight-binding [209, 210] and *ab initio* DFT [202, 203, 211–214] studies have both shown an inwards relaxation, but a recent study using the sx-LDA screened exchange functional has instead suggested an outwards relaxation [156].

These discrepancies can be partially attributed to the different approximations contained in the various levels of theory; however, restricting ourselves only to the DFT studies, we find a very broad range of calculated formation energies ($\sim$2 eV spread in the values, see Fig. 5.6 for a comparison of selected studies with our results), as well as qualitative disagreements in the symmetry of the relaxed defect centre. In particular, Probert and Payne [203], Puska *et al.* [202], and Wright [213] have performed studies of the finite size convergence properties of the defect, showing a noticeable dependence of calculated quantities on system size. We shall refer to these studies in the discussion of our results in Secs. 5.5.1 and 5.5.3.

5.2 Watkins' vacancy model

Fig. 5.1 illustrates the model deduced by Watkins from EPR studies of the silicon vacancy, and later confirmed by the ENDOR measurements. The model is based on that proposed previously by Coulson and Kearsley in their study of the equivalent defect centre in carbon [215]. Only four orbitals need to be taken into account, creating a 'defect molecule' that is embedded within the perfect crystal lattice, and interacts only weakly with it; the orbitals originate from the four valence bonds that are broken by removing the central atom. We can assume that they still point towards the defect centre in the undistorted lattice to maintain the local tetrahedral symmetry of the atoms surrounding the vacancy. The four dangling bonds are labelled $\mathrm{a}_1, \mathrm{a}_2, \mathrm{b}_1, \mathrm{b}_2$ (these are all equivalent when the point group symmetry of the defect centre is $\mathrm{T_d}$).

We can now construct four linear combinations of these orbitals, resulting in one completely symmetric s-type nodeless combination, and three equivalent p-type combinations with a nodal plane (these are classified as a_1 and t_2, respectively, in Mulliken's symmetry notation [216]). The s state is lowest in energy and lies within the silicon bulk valence band, while the p states form a triply degenerate deep defect level in the band gap. For the neutral vacancy, each of the four neighbouring atoms will contribute one electron to the these levels, thereby resulting in partially occupied defect level and a degenerate ground state.

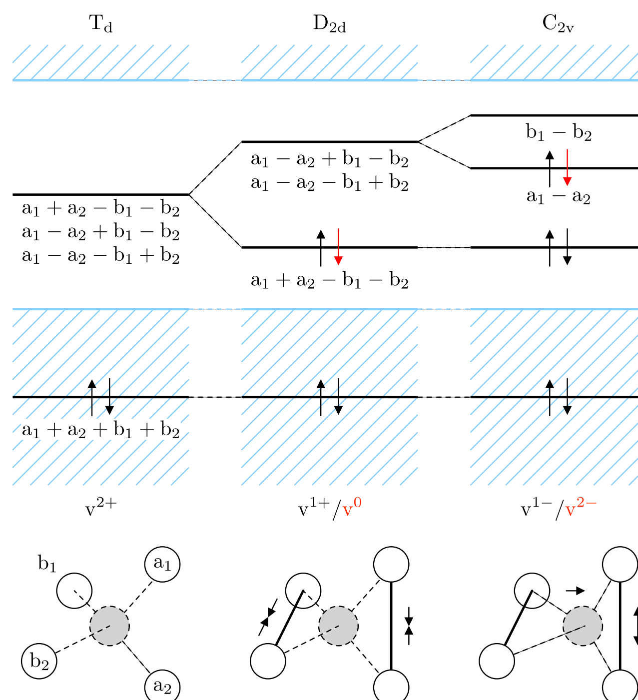


Figure 5.1: LCAO model for the vacancy in silicon. The four orbitals associated with the neighbouring atoms of the defect site are the dangling bonds resulting from the removal of the central silicon atom. The figure shows the predicted ionic configuration and point group symmetry for different charge states of the vacancy due to Jahn-Teller distortion. For the D_{2d} and C_{2v} configurations, the first charge state (in black) refers to the spin-polarised case with only one electron in the highest occupied orbital, and the second charge state (in red) refers to the non-spin-polarised case with two electrons.

This configuration, however, is forbidden by the Jahn-Teller theorem [217], which states that a non-linear molecule with a degenerate ground state must be unstable with respect to a symmetry-lowering distortion that lifts the degeneracy. Therefore, restricting ourselves to the charge states $q = -2, \ldots, +2$, it is clear that the vacancy will only remain in its undistorted T_d symmetry for v^{2+}, since the defect levels are completely unoccupied. Instead, for v^{1+} and v^0 only one of the defect levels must be occupied; the system is lowered in energy by forming pairs between the four neighbouring atoms (hence the a, b labels), thus creating two new bonds from the four original orbitals and lowering the

symmetry of the defect centre to D_{2d}. The p state which is lowered is the one that can be combined with the s state to produce the two bonding states $a_1 + a_2$ and $b_1 + b_2$. The remaining two p states are still degenerate, and can now be combined to produce the corresponding antibonding states $a_1 - a_2$ and $b_1 - b_2$.

Finally, for v^{1-} and v^{2-} (in which an additional defect level is occupied) the degeneracy of these p states is lifted by a distortion which differentiates between the atom pairs, such as the stretching of one of the new bonds. This further lowers the symmetry of the system to C_{2v}. We assume here that the state that is lowered in energy is the combination forming the antibond of the *longer* bond ($a_1 - a_2$ in the figure), since this will now rehybridise with the bonding orbital $a_1 + a_2$ to produce two independent, fully occupied dangling orbitals a_1 and a_2.

Despite its success in explaining the symmetry information given by experiment, this empirical model cannot provide more than a qualitative understanding of the defect centre; for quantitative predictions of the various defect properties, it is instead necessary to employ a more accurate electronic structure method. To this end, we investigate the vacancy from a first-principles perspective with DFT; nevertheless, we shall make extensive use of Watkins' model, both in this and the next chapter, as a reference point for analysing our DFT results.

5.3 Computational methodology

5.3.1 Supercell calculations within DFT

Following our general definition of the defect formation energy from Eqs. 4.15 and 4.16, we can write a specific definition for a vacancy with charge q in a supercell of bulk silicon having N atoms in its pristine (i.e., defect-free) configuration:

$$E_f^q = E^{\mathrm{vac},q} - \left(\frac{N-1}{N} \right) E^{\mathrm{bulk}} + q \left(\varepsilon_v + \Delta\mu_e \right) . \tag{5.1}$$

The determination of the valence band maximum ε_v is discussed in more detail in Sec. 5.4.

We study supercells of up to 1000 atoms; these are constructed from the three unit cell shapes belonging to the cubic crystal system: FCC (the primitive cell for the diamond lattice, with 2 atoms), simple cubic (SC) (with 8 atoms), and BCC (with 32 atoms). The supercells are then made from $n \times n \times n$ unit cells.

For supercell sizes up to and including 256 atoms, the calculations are performed using the CASTEP code [218] (version 5.0); this employs the plane-wave pseudopotential DFT method discussed in previous chapters. The LDA is used to describe exchange and correlation. For charged supercell calculations, a compensating uniform background charge is added to satisfy the condition of overall charge neutrality for the periodic unit cell.

We use two pseudopotentials for silicon: CASTEP's 'on-the-fly' Vanderbilt ultrasoft pseudopotential (USPP), and a norm-conserving pseudopotential (NCPP); both have four valence electrons, corresponding to the electronic configuration $3s^2 3p^2$. The pseudopotentials give accurate lattice constants for bulk silicon (5.39 Å and 5.38 Å, respectively, compared with an experimental value of 5.43 Å); this slight underestimation of less than 1% is typical for LDA-DFT. The bulk moduli obtained are also very accurate (100 GPa and 98 GPa, respectively, compared with an experimental value of 100 GPa). The DFT-optimised lattice parameter is used in all calculations, in order to avoid spurious stresses on the supercell.

We follow the methodology outlined by Probert and Payne in their study of the neutral silicon vacancy [203] for the convergence of quantities of interest with respect to the numerical parameters of the simulation (most importantly, the plane-wave energy cut-off and the Brillouin zone sampling); however, the parameters are converged *independently* for all supercell sizes. Our convergence tolerance for the defect formation energy is 10 meV. We use a regular Monkhorst-Pack mesh of k-points for the Brillouin zone integration; in the text, the convergence parameter k^{MP} refers to a $k^{\mathrm{MP}} \times k^{\mathrm{MP}} \times k^{\mathrm{MP}}$ grid centred on the Γ-point. The issue of Brillouin zone sampling shall be discussed both in this and the next chapter; in general, we make use of two sampling

schemes: Γ-point only sampling (i.e., $k^{\mathrm{MP}} = 1$), and what we call a 'dense' sampling. This last term we shall use to indicate that the defect formation energy has been converged to within our stated tolerance with respect to k^{MP} for the supercell in question.

We perform calculations on both unrelaxed and relaxed geometries. For the relaxation, a quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) scheme is used. All the atoms in the supercell are allowed to move, and their positions are slightly randomised at the start of the procedure to allow for symmetry breaking. The supercell lattice vectors, however, are held constant at their bulk value. Our convergence tolerance parameters for the geometry optimisation procedure are 5×10^{-3} eV/Å for the RMS force, 1×10^{-4} Å for the RMS ionic displacement, and 1×10^{-5} eV for the energy per atom difference between BFGS iterations. We note that forces generally converge slower than energies; therefore, the convergence of the basis set cut-off energy and k-point sampling grid has to be checked explicitly with respect to the RMS force for the unrelaxed system.

Smearing of the Fermi surface is required for calculations on unrelaxed geometries in particular, when partially occupied degenerate defect levels are present in the band gap. This smearing is approximately equivalent to performing the calculation at a finite electronic temperature, for which the occupancies of the Kohn-Sham states vary smoothly according to the Fermi-Dirac distribution (in practice, a Gaussian distribution is applied [219]). This occupancy smearing, however, is used principally to improve the numerical stability of the code's energy minimisation algorithm when dealing with metals (or, in our case, defects states at the Fermi level). We note that the smearing width can have a noticeable effect on the ionic relaxation and formation energy for large supercells. We therefore reduce it to less than 1 meV in all calculations; this does not affect the numerical stability of the minimisation algorithm, and is small enough to have a negligible effect on our results.

5.3.2 Linear-scaling DFT calculations

For our largest calculations, on a 1000-atom SC supercell, we use the linear-scaling DFT code ONETEP (version 2.4), described previously in Sec. 3.3.3. The 'cross-over', namely, the system-size at which it becomes computationally more efficient to use ONETEP as opposed to conventional cubic-scaling DFT, is highly system dependent and we estimate it to lie at around 500 atoms for bulk silicon. For this reason we use conventional plane-wave DFT for supercells smaller than this cross-over, and linear-scaling DFT for the largest supercell. We use the same norm-conserving pseudopotential as for CASTEP, and nine NGWFs on each silicon atom with a truncation radius of 3.97 Å. The NGWFs are described by an underlying basis set of psinc functions with a spacing of 0.27 Å in each direction (approximately equivalent to a plane-wave energy cut-off of 800 eV); this spacing is chosen to be commensurate with that of the pristine crystal lattice, so that each atom is positioned on a psinc grid point. We do not truncate the density kernel.

The plane-wave accuracy of ONETEP has already been demonstrated for crystalline silicon [220]; however, we note that defect formation energy calculations are particularly challenging, since we are interested in a small energy difference between two systems with a large total energy.

5.3.3 Maximally-localised Wannier functions

In Sec. 3.3.1, we described how a group of delocalised Bloch states can be transformed into a set of localised Wannier functions that span the same space; the transformation can be made unique by finding the most localised representation of the electronic structure, defined in terms of the spreads of the Wannier functions. Computationally, this gives rise to a minimisation problem that can be tackled by using well-known algorithms such as conjugate gradients schemes.

To perform this wannierisation procedure and compute the MLWFs for the defect systems, we make use of the WANNIER90 code [221] (version 1.2). We employ the extended Bloch states of the defect supercells calculated with CASTEP and the norm-conserving pseudopotentials. We have written a stan-

dalone CASTEP-TO-WANNIER90 interface code to calculate the required information from the Bloch states for performing the wannierisation and subsequent visualisation of the MLWFs; this includes the projection of a set of user-defined trial orbitals onto the subspace, that is used both for the disentanglement of non-isolated bands, and to compute an initial guess for the $U_{mn}^{(\mathbf{k})}$ matrix (which is varied during the wannierisation procedure to minimise the total spread $\Omega(U)$). Full details of the CASTEP-TO-WANNIER90 code are given in Appendix B.

Since we use a large supercell of 256 atoms for our MLWF calculations, we only include the Γ-point wavefunction in the wannierisation routine, even though the defect formation energy is found to require an extended k-point grid for convergence. This is necessary to reduce the computational cost, since WANNIER90 is a serial code. However, we expect a negligible difference in the properties of the MLWFs.

5.4 Determining the electronic chemical potential

As described in Sec. 4.4.1, the formation energy of charged defects depends on the electronic chemical potential, which is given relative to the valence band maximum eigenvalue position ε_v. ε_v can equivalently be defined as the energy difference between the pure host material and the host with a single electron hole. In the limit of a dilute hole gas, this energy difference is equivalent to the VBM eigenvalue of the pure host supercell. For a finite supercell calculation, however, the energy difference between the bulk system with and without the hole only slowly converges to the infinite limit as the system size is increased; this is due to the convex behaviour of the local exchange-correlation functional for fractional charges at finite system sizes, that gives rise to a correction term Δ^{convex}, defined in Eq. 4.32. Instead, the VBM will always remain at the same position (given a sufficient Brillouin zone sampling), which is the limiting value (see Fig. 10 in Ref. [154]). As we have argued in Sec. 4.4, this should be considered a 'theory error', as it is inextricably linked to the approximations

contained in the exchange-correlation functional; even in the limit of infinite system size, in fact, the Kohn-Sham eigenvalues are not correctly placed.

However, there is an additional problem in defining the position of the VBM in the supercell approach, that is purely a finite size error (i.e., independent of the level of theory employed): ε_v is defined only up to a constant, as there is no absolute reference for the electrostatic potential in the supercell. For consistency between the total energy of the charged defect supercell and the energy of the VBM, therefore, a finite size correction is usually employed to align the average potential of the supercell with and without the defect [127, 154, 222]. This uncertainty vanishes in the limit of very large supercells, as the perturbation to the host potential becomes negligible.

We can therefore take two approaches to determine ε_v: either calculate the energy difference of the bulk supercell with $q = 0$ and $q = +1$, or use the bulk VBM eigenvalue and employ a potential alignment correction to achieve consistency with the charged defect supercell. It is important to note that these two approaches are quite different. The former explicitly corrects for the convex behaviour of the functional (expected however to be small for moderately large supercells of hundreds of atoms), but shifts the potential reference level in an uncontrolled manner based on the difference of the average potential of the bulk supercell with and without the electron hole. The latter, instead, does not correct the convexity error, but explicitly adjusts the potential reference level. Unfortunately, the two approaches are not compatible, due to the unknown shift in potential in the former. In our calculations, therefore, we compare both methods (see Sec. 5.5.3).

We now focus on the potential alignment correction, since it is not straightforward to evaluate in practice. In the rest of this section we describe the two methods that we employ for calculating it: a partitioning of the real-space potential using Voronoi cells, and an MLWF-based approach. Both methods are tested on the silicon vacancy in a 256-atom supercell. To the best of our knowledge, this is the first time these methods have been used to calculate the correction.

5.4.1 The real-space Voronoi cell method

Our first method is based on considering the electrostatic potential directly. In such approaches, the correction is typically determined by plotting the electrostatic potential obtained from the DFT calculation, and averaging it either in the x-y plane, within atomic spheres or over primitive cells of the host material. The important point is that the averaging must be done in some localised manner, such that it is possible to measure the value of the bulk-like potential in the charged defect supercell by considering a region far away from the defect centre. The correction to the defect formation energy is then $q\left(V^{\mathrm{vac}} - V^{\mathrm{bulk}}\right)$, where V^{bulk} and V^{vac} are the average potentials in the bulk supercell and in a 'bulk-like' region of the defect supercell.

In order for the correction to be unambiguous and accurate, the averaging volume should be as small as possible while still covering a region of space that is completely representative of the bulk material (i.e., all regions in the bulk supercell should give the macroscopic average for the material). Our approach is to use Voronoi cells: each region corresponds to the real-space volume of points closest to one particular atom in the supercell. There are, therefore, as many regions as there are atoms. The electrostatic potential, however, is given on a discrete real-space grid whose spacing depends on the basis set cut-off energy and which, in general, is not commensurate with the atomic spacing, leading to inaccuracies in the averaging procedure that may swamp the difference between the bulk and defect supercells. In order to overcome this drawback, we Fourier-interpolate the potential onto a finer grid that is commensurate with the atomic spacing, which results in each Voronoi cell containing exactly the same number of points. We note that care must be taken to give the correct fractional weighting to points which are directly between two or more atoms.

The potential alignment correction is then determined by considering Voronoi cells belonging to the atoms which are furthest from the defect centre, as shown in Fig. 5.2.

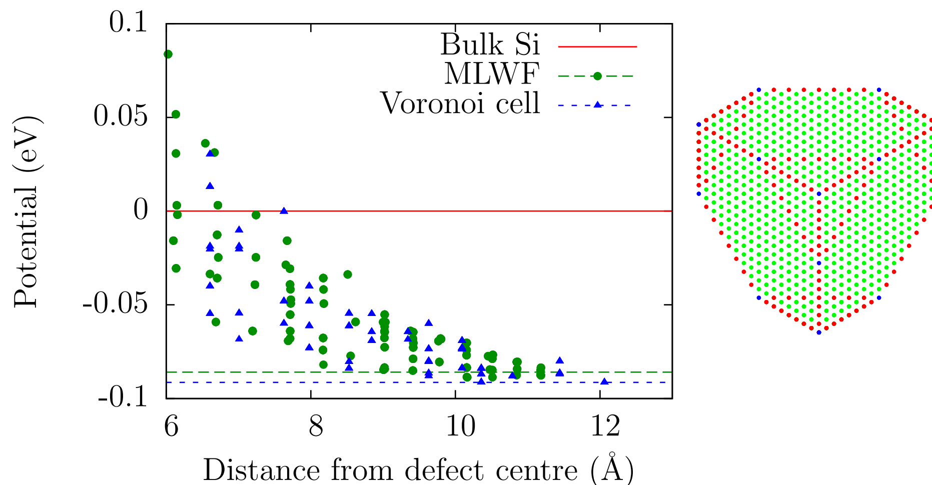


Figure 5.2: Potential alignment correction for v^{2-} using the MLWF and real-space Voronoi cell methods. The values are adjusted with respect to the bulk supercell. The dashed horizontal lines show the value for the furthest point from the defect centre. The Voronoi cell used is shown on the right, viewed in the [111] direction; green dots are surfaces, red dots are edges and blue dots are corner points. Each Voronoi cell contains 8192 points of the Fourier-interpolated potential.

5.4.2 The MLWF method

Our second approach to calculating the potential alignment correction is based on matrix elements of the Hamiltonian in the basis of MLWFs. The shift in the reference of potential that is sought is reflected in the position of the energy eigenvalues; however, comparing the eigenvalues of bulk-like states in the defect supercell with those of the bulk supercell is problematic since the eigenstates are delocalised. MLWFs, on the other hand, provide a probe of the localised properties of the system. Once the Hamiltonian is transformed into a basis of MLWFs, the potential alignment correction can be determined by considering on-site matrix elements $H_{nn} = \langle \omega_n | \hat{H} | \omega_n \rangle$ (where ω_n is a MLWF belonging to the cell at the origin) for Wannier functions whose centres, defined by their first moments, are far away from the defect centre. The shift in potential is then

Charge	MLWF	Voronoi cell	Diff.
2+	−19.1	−8.7	−10.4
1+	−32.9	−27.0	−5.9
1−	−67.0	−69.2	2.2
2−	−85.9	−91.4	5.5

Table 5.1: Comparison of the potential alignment correction obtained from the MLWF and real-space Voronoi cell methods for all charge states of the vacancy. The discrepancy between the two methods is listed in the last column. All values are given in meV.

simply $H_{nn}^{\mathrm{vac}} - H_{nn}^{\mathrm{bulk}}$. In practice, as shown in Fig. 5.2, we plot this difference for all n as a function of the distance of the Wannier function centre from the defect centre, and the value of the potential alignment correction is taken as the mean of the values which are furthest from the defect centre.

This method is feasible since the wannierisation procedure results in the same partitioning of the electronic density in the bulk and defect supercells everywhere except in the direct neighbourhood of the defect; hence it is possible to identify Wannier functions that are associated with the defect centre and Wannier functions that are not. In the latter case, therefore, it is straightforward to unambiguously match equivalent Wannier functions in the bulk and defect supercells.

From Fig. 5.2 it can be seen that the MLWF correction method is in excellent agreement with the Voronoi cell method. Among all our calculations for the various charge states (listed in Table 5.1), the maximum discrepancy is 0.01 eV. The MLWF method gives a finer representation of the system than averaging over atomic sites (as in the Voronoi cell approach), since each silicon atom has four bonding Wannier functions connecting it to its neighbours. Therefore, there are twice as many Wannier functions as atoms, and hence twice as much information to consider.[2]

MLWFs are now a standard tool for the analysis of electronic structure calculations; therefore, this approach provides a simple and accurate method

[2]Only the valence bands have been taken into account for the wannierisation.

			k-point vol.	E_f^0 (eV)	
N	Symm.	k^{MP}	$(10^{-3}\mathrm{\AA}^{-3})$	Total	Kinetic
2	FCC	8	12.35	2.65	-10.19
8	SC	7	4.61	3.13	-11.49
16	FCC	6	3.66	3.10	-11.68
32	BCC	6	1.83	3.83	-12.08
54	FCC	4	3.66	3.51	-12.36
64	SC	6	0.91	3.81	-12.45
128	FCC	4	1.54	3.79	-12.53
216	SC	3	2.17	4.05	-12.51
250	FCC	3	1.87	3.95	-12.60
256	BCC	3	1.83	4.13	-12.47

Table 5.2: List of all supercells up to 256 atoms with their respective symmetries. Also listed are the value of k^{MP} used (converged with respect to the formation energy for each supercell), the corresponding k-point volume in reciprocal space, the unrelaxed defect formation energy and the kinetic energy contribution to this value.

for calculating the potential alignment correction, since all the necessary information is readily available once the standard wannierisation procedure has been performed.

5.5 Results

5.5.1 The neutral unrelaxed vacancy

In order to study the finite size convergence properties of the system we first simulate the neutral vacancy in its unrelaxed state, with all the atoms in their perfect crystalline positions. As explained by Watkins' model, this results in a triply degenerate deep defect level which is partially filled with two electrons in the neutral charge state.

As stated previously, each supercell is converged independently. We now use the ultrasoft pseudopotential, resulting in a converged cut-off energy of

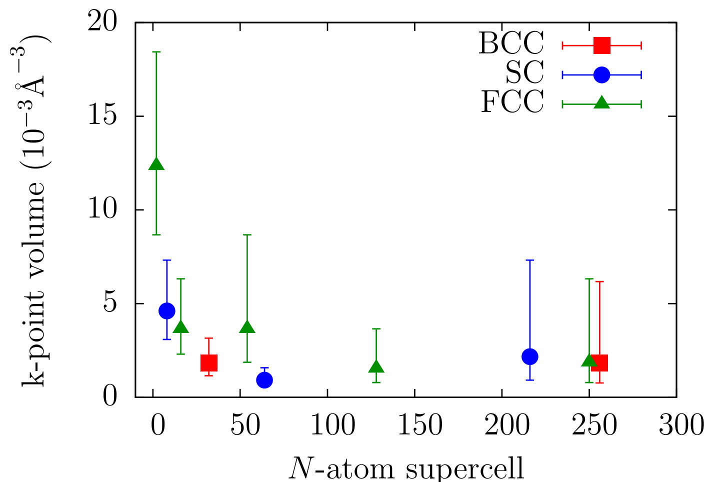


Figure 5.3: Convergence of k-point volume. The vertical bars show the volumes corresponding to $k^{\mathrm{MP}} \pm 1$ in each case.

400 eV for all system sizes. The results are reported in Table 5.2. The pattern of the convergence for the three symmetries agrees very well with that calculated by Probert and Payne using the GGA [203], although our results vary by -0.25 eV to 0.10 eV with respect to theirs, generally being higher for FCC supercells and lower for SC and BCC. The convergence is strongly dependent on supercell geometry: BCC supercells show the fastest convergence, followed by SC and then FCC. This has been explained by considering the direction along which there appears a spurious movement of charge caused by the finite system size for each of the three geometries [203].

It is noteworthy that even for the largest supercells it is necessary to include more points than Γ for proper convergence of Brillouin zone integrals. The resulting k-point volume in reciprocal space for each supercell is shown in Fig. 5.3. It is clear that convergence to a stable value does not occur until at least $N = 64$; therefore, caution should be taken when extrapolating convergence parameters from small to large supercells, as is often done.

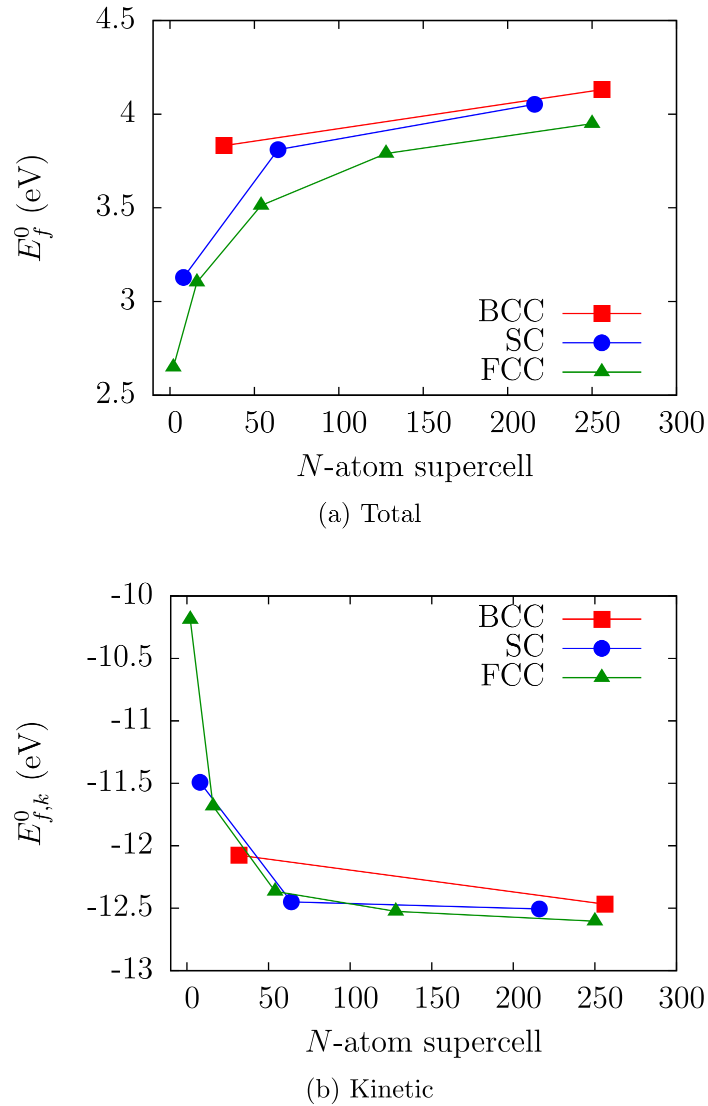
These results represent fictitious systems with an extremely high vacancy concentration and a long-range ordering of the defect centres; however, what we are really interested in is how well they describe the properties of a truly

isolated vacancy. As can be seen from the values in Table 5.2 and Fig. 5.4a, the defect formation energy increases monotonically for each individual supercell geometry. We therefore take our highest value of 4.13 eV (for the 256-atom BCC supercell) as a lower bound to the unrelaxed isolated vacancy formation energy.

The slow convergence of the formation energy as a function of system size is due to the spurious interaction between periodic images of the defect centre. As the defect is both unrelaxed and uncharged, this is neither caused by elastic interactions nor monopole-monopole electrostatic interactions; rather, it is the result of higher multipole interactions and overlap between the wavefunctions of the periodically repeated defect centres.

The importance of wavefunction overlap may be gauged by considering the kinetic energy contribution to the defect formation energy, which we define analogously to Eq. 5.1, except using only the non-interacting kinetic energy contribution to the total energy instead of the total energy itself. The results are listed in the right-hand column of Table 5.2 and shown in Fig. 5.4b. It is immediately apparent that the kinetic energy component of the defect formation energy varies on a scale *larger* than the total formation energy as the supercell size is increased, and that, even at our largest supercell sizes, it is not converged to better than 0.1 eV. This slow convergence is caused by subtle but non-negligible changes in the defect wavefunctions of neighbouring defect centres, that occur in order to maintain orthogonality as the periodic images of the defect move apart. While the electronic density is rather insensitive to them and, therefore, the electrostatic interactions are not affected, the kinetic energy is not. As a result, any correction scheme that accounts solely for the classical electrostatic interaction will not be sufficient in this case to predict the correct formation energy for an infinite system beyond this level of accuracy.

We can also understand this effect by considering the electronic structure in terms of MLWFs. In this picture, the concept of a defect level is replaced by that of a defect Wannier function, which describes the local environment at the defect centre. For a system to be perfectly converged with respect to supercell size the tails of the Wannier functions at one defect centre must have a negligible overlap with those at neighbouring defects, and so remain



(a) Total

(b) Kinetic

Figure 5.4: Convergence of the defect formation energy with system size (both the total and the kinetic energy contribution) for different supercell geometries.

unchanged as the system size is further increased.

We consider one last question before moving on to ionic relaxations: is it desirable to use a dense k-point grid when simulating point defects? Although this is obviously necessary for describing the delocalised bulk states, it seems reasonable that Γ-point sampling might give a better approximation of the

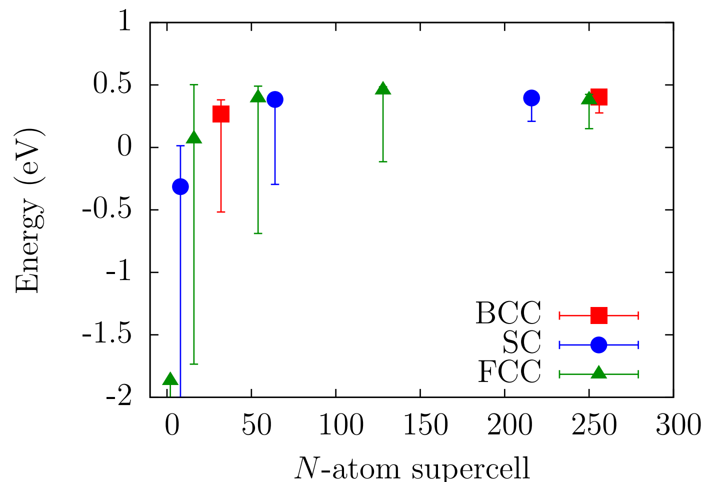


Figure 5.5: Dispersion of the unrelaxed defect level. The symbols show the position of the level at Γ, and the bars show the extent of the dispersion obtained from our dense k-point sampling. The positions are given with respect to the bulk VBM.

localised defect levels in the dilute limit, as a way of deliberately eliminating unwanted dispersion, and by occupying the state which most closely resembles the bonding defect state for the infinite system (see Sec. 6.4.3 and Appendix D for further discussion). Indeed, Fig. 5.5 shows that the position of the defect level at Γ rapidly converges to a fixed value, even if the total dispersion of the level throughout the Brillouin zone is large.

As far as the defect formation energy is concerned, at least, our results show that Γ-point calculations do not give a better estimate of the defect formation energy: for the 256-atom supercell E_f^0 is underestimated (as compared to our lower bound estimate) both for the unrelaxed and relaxed cases, by 0.2 eV and 0.4 eV respectively. This conclusion is in agreement with two other studies on the relaxed silicon vacancy: Puska *et al.* [202], who report that a 2^3 MP sampling shows a faster convergence than Γ-point sampling for the neutral charge state; and Wright [213], whose results for the 216-atom supercell show a greater average error across all charge states for Γ-point sampling than any finer MP grid sampling. We note that, conversely, quantities of interest other than the defect formation energy might benefit from Γ-point only sampling;

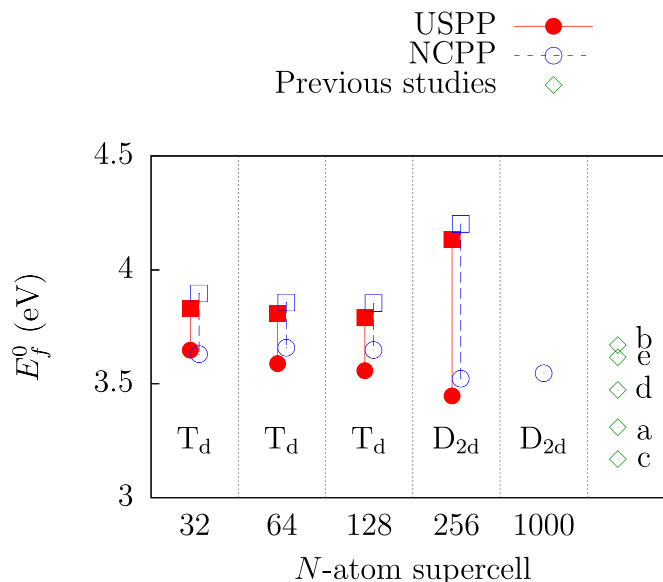
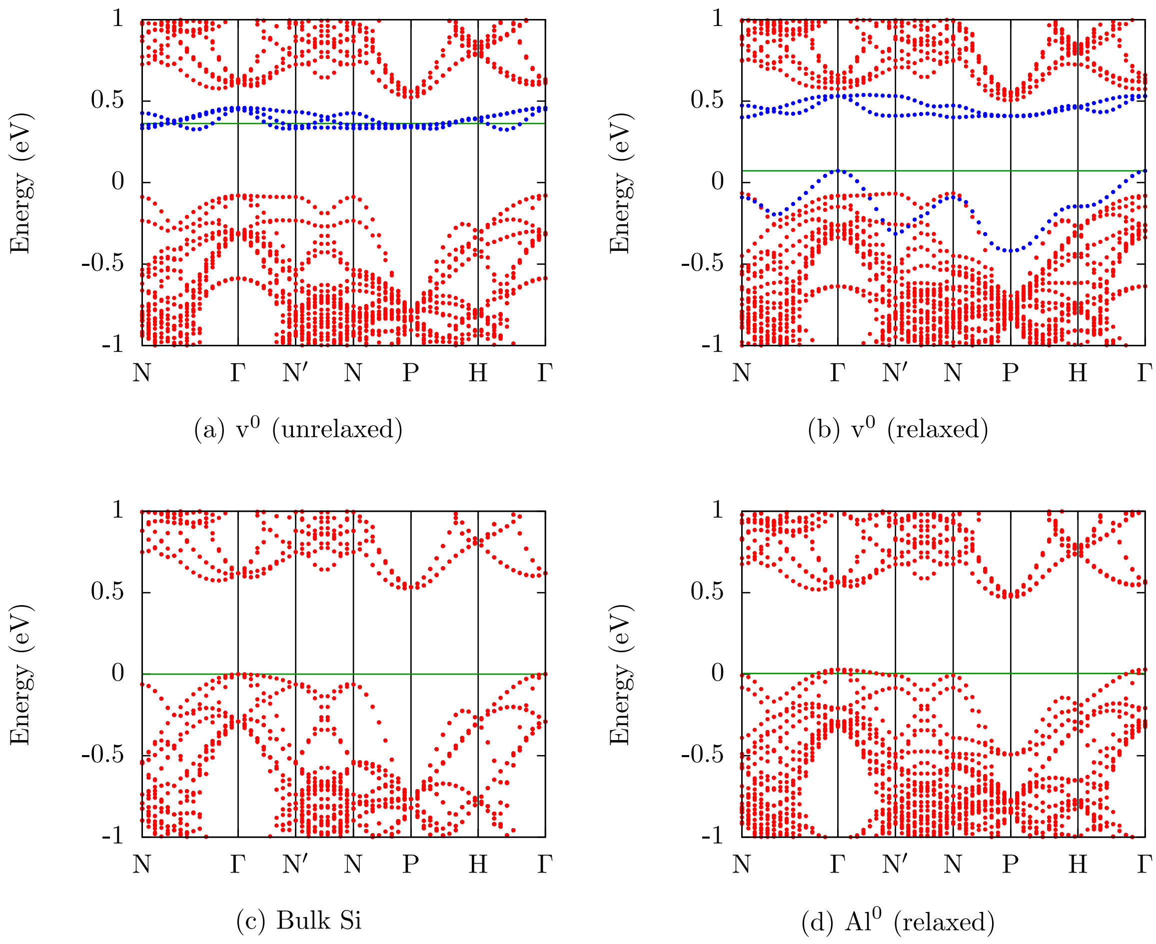


Figure 5.6: Relaxation effects for the neutral vacancy. Squares show the defect formation energy for the unrelaxed lattice and circles for the relaxed one. Labelled below is the point group symmetry of the defect centre after relaxation. The last column shows the relaxed formation energy from previous studies: a) Ref. [202] ($N = 216$, $k_{\mathrm{MP}} = 2$, LDA); b) Ref. [214] ($N = 216$, $k_{\mathrm{MP}} = 2$, GGA); c) Ref. [203] ($N = 256$, $k_{\mathrm{MP}} = 2$, GGA); d) Ref. [213] ($N = 1000$, $k_{\mathrm{MP}} = 3$, LDA); e) Ref. [213] ($N = 1000$, $k_{\mathrm{MP}} = 3$, GGA). Details of our calculations are given in the text.

as we shall see in Sec. 5.5.4, the stable charge state transition levels are an example of this.

5.5.2 The neutral relaxed vacancy

We now discuss the effect of relaxing the ionic positions in the supercell containing the neutral vacancy. In particular, we are interested in the possible change in point group symmetry at the defect centre, and the change in defect volume, corresponding either to an inwards or outwards relaxation of the atoms surrounding the vacancy. The symmetry of the relaxed defect centre is deduced by monitoring the change in bond lengths between the four nearest neighbours of the vacancy site. We perform the calculations with ultrasoft



(a) v^0 (unrelaxed)

(b) v^0 (relaxed)

(c) Bulk Si

(d) Al^0 (relaxed)

Figure 5.7: Band structures for the 256-atom BCC supercell. The full horizontal line (in green) indicates the level of the highest occupied orbital. For the vacancy supercells, the three defect levels are shown in blue, and all other bands are shown in red; the entanglement with the top of the valence band in the relaxed case has been estimated by calculating the wavefunction overlap between neighbouring k-points.

(norm-conserving) pseudopotentials, using the converged value of the numerical parameters: a Brillouin zone sampling of $k^{\mathrm{MP}} = 3$ ($k^{\mathrm{MP}} = 4$) for the 256-atom supercell, and an energy cut-off of 400 eV (800 eV).

On relaxation of the ionic positions, the predicted symmetry for v^0 from Watkins' model ($\mathrm{D}_{2\mathrm{d}}$) is not seen in any supercell smaller than 256 atoms, as shown in Fig. 5.6. Calculations on the 32-atom BCC, 64-atom SC and 128-atom FCC supercells do not exhibit a change in symmetry, only an inwards relaxation. Consequently, the defect formation energy is only lowered by ∼0.2 eV in these cases. We note that the ionic positions at the end of the geometry optimisation will contain some numerical noise, on the order of 1–10 mÅ; this will swamp very weak Jahn-Teller distortion effects, which might nevertheless be present in the small supercells.

The 256-atom BCC supercell undergoes the predicted change in symmetry and a reduction in defect volume by more than 40%. This results in the defect formation energy being lowered by 0.7 eV (full results are given in Table 5.3). Therefore, even though the unrelaxed defect formation energy is highest for the largest supercell, after relaxation it becomes the lowest, demonstrating the well-known importance of Jahn-Teller distortion and the long-ranged nature of the elastic interactions which can lead to both qualitatively incorrect relaxation patterns and quantitatively inaccurate formation energies in small supercells.

These results confirm several previous DFT studies on this system [202, 203, 213]. In particular, Puska *et al.* [202] report that small supercells can show a range of different symmetries depending on the k-point sampling, and there is some evidence from their results that Γ-point only sampling favours $\mathrm{D}_{2\mathrm{d}}$. It is clear from our results that a dense k-point sampling favours the lattice's unrelaxed tetrahedral point group symmetry T_d for small supercells. This qualitative difference between sampling schemes is explained within a tight-binding model of the system in Appendix D.

In the figure we also show the results obtained with ONETEP for the 1000-atom SC supercell. The calculation has been fully relaxed, starting from a random displacement of the atomic positions[3]. This calculation is in excellent

[3]The unrelaxed defect formation energy has not been calculated with ONETEP, since the density matrix method does not allow for a degenerate ground state without an explicit

						Bond lengths (Å)			Defect vol.	E_f^q (eV)	
Charge	N	k^{MP}	Rel.	Pspot.	Symm.	a–a	b–b	a–b	(Å^3)	Hole	VBM
2+	256	1	Yes	USPP	T_d	3.22	3.22	3.22	3.93	2.86	2.85
	256	3	Yes	USPP	T_d	3.63	3.63	3.63	5.61	3.62	3.66
1+	256	1	Yes	USPP	D_{2d}	2.93	2.93	3.27	3.62	2.99	3.00
	256	3	Yes	USPP	D_{2d}	2.86	2.86	3.40	3.73	3.50	3.55
	1000	1	Yes[a]	NCPP	D_{2d}	2.87	2.87	3.37	3.70	-	3.50
0	256	1	No	USPP	T_d	3.81	3.81	3.81	6.54	3.90	
	256	3	No	USPP	T_d	3.81	3.81	3.81	6.54	4.13	
	256	1	Yes	USPP	D_{2d}	2.83	2.83	3.31	3.53	3.06	
	256	3	Yes	USPP	D_{2d}	2.86	2.86	3.40	3.73	3.45	
	256	4	No	NCPP	T_d	3.81	3.81	3.81	6.50	4.20	
	256	4	Yes	NCPP	D_{2d}	2.86	2.86	3.41	3.74	3.52	
	1000	1	Yes	NCPP	D_{2d}	2.87	2.87	3.37	3.70	3.55	
1−	256	1	Yes	USPP	$\sim C_{2v}$	3.10	2.70	3.24 (av.)	3.50	3.48	3.44
	256	3	Yes	USPP	$\sim C_{2v}$	2.94	2.87	3.26 (av.)	3.55	3.80	3.71
						a–a	a–c	$\Delta\alpha$			
2−	256	1	Yes	USPP	D_{3d} (split)	3.45	2.60	6.63°	2.86	3.70	3.57
	256	3	Yes	USPP	D_{3d} (split)	3.46	2.60	6.52°	2.88	3.92	3.72

Table 5.3: Summary table of the results obtained for the 256-atom BCC supercell and the 1000-atom SC supercell. Bond lengths are given between the atoms surrounding the vacancy, as labelled in Fig. 5.9a for v^{2+}, v^{1+}, v^0, v^{1-}, and Fig. 5.10a for v^{2-}. Atoms labelled with the same letter are equivalent by symmetry. $\Delta\alpha = \alpha - 90°$ is the distortion of the regular octahedron in the split vacancy configuration. The defect volume is calculated from the tetrahedron formed by the four neighbours of the vacancy site. The defect formation energy is given for $\Delta\mu_e = 0$. Symmetry labels are given in Schönflies notation.

[a] For this system we use the relaxed atomic positions obtained for the 1000-atom v^0 supercell, without further relaxation.

agreement with the 256-atom system calculated with CASTEP with the same pseudopotential, both in terms of the relaxed configuration (33 mÅ RMS difference in the lengths of the bonds at the defect centre) and the relaxed defect formation energy (23 meV difference).

Fig. 5.7 shows band structure calculations for the 256-atom BCC supercell. At this system size, the three defect levels are not very dispersed, and clearly distinguishable for the unrelaxed lattice in the top half of the DFT band gap (Fig. 5.7a). The effect of relaxation (Fig. 5.7b) can be directly compared with the schematic level diagram in Fig. 5.1; one of the three levels is lowered by almost 0.5 eV, while the other two are unchanged and only slightly raised to the top of the band gap. For comparison, we also include the band structure for the relaxed neutral aluminium substitutional defect in the same supercell (Al^0, in Fig. 5.7d); this is a typical shallow-level acceptor in silicon. While the deep defect levels of the vacancy are clearly separated from the bulk-like band structure, the shallow acceptor level of the aluminium impurity is visible only as a small perturbation on the valence band edge host states.

5.5.3 Charged vacancies

The relaxed symmetry and defect formation energy for the charged states of the vacancy are calculated using the 256-atom supercell in CASTEP with the ultrasoft pseudopotential. We repeat the calculations with both Γ-point sampling and a $k^{\mathrm{MP}} = 3$ (dense) grid. The results for all charge states are given in Table 5.3.

As explained in Sec. 5.4, there are two possible approaches for estimating the position of ε_v in the system: we can either use a calculation of the bulk supercell with an electron hole (denoted 'Hole' in Tables 5.3 and 5.4), or the bulk VBM value with a potential alignment correction[4] ('VBM'). The agreement between the two approaches is better for the positively charged vacancies

(non-linear-scaling) diagonalisation.

[4]The results given use the Voronoi cell method for determining the potential alignment; this is to ensure consistency among all results, since for the 1000-atom supercell calculations with ONETEP only this method is appropriate. As stated previously, all comparisons between the Voronoi cell and MLWFs methods for 256-atom supercell calculations with CASTEP are in agreement to within 0.01 eV.

		$N = 256$				$N = 1000$	Ref. [202] $N = 216$	Ref. [214][a] $N = 216$	Ref. [213][a] $N = 1000$	
		$k^{\mathrm{MP}} = 1$		$k^{\mathrm{MP}} = 3$		$k^{\mathrm{MP}} = 1$	$k^{\mathrm{MP}} = 1$	$k^{\mathrm{MP}} = 2$	$k^{\mathrm{MP}} = 3$	
	Expt. [184]	Hole	VBM	Hole	VBM	VBM	LDA	GGA	LDA	GGA
$E\,(2+/1+)$	0.13	0.13	0.16	$\times$	$\times$	-	0.19	0.13	0.27*	0.27
$E\,(2+/0)$	0.09*	0.10*	0.11*	$\times$	$\times$	-	0.15*	0.10*	0.28	0.19*
$E\,(1+/0)$	0.05	0.07	0.06	$\times$	$\times$	0.04	0.11	0.06	0.28*	0.12
$E\,(0/1-)$	-	0.43	0.38	0.35	0.26	-	0.57	0.37	0.76	0.63
$E\,(0/2-)$	-	0.32*	0.25*	0.24*	0.13*	-	0.49*	0.37*	0.66*	0.53*
$E\,(1-/2-)$	-	0.22	0.13	0.13	0.00	-	0.40	0.36	0.56	0.42

Table 5.4: Transition levels for the 256-atom BCC supercell and the 1000-atom SC supercell. Results from previous studies are also shown. Our calculations use the LDA functional. Crosses ($\times$) are used when there is no transition in the band gap. Asterisks (*) denote thermodynamically stable transitions. All values are given in eV.

[a] The transition levels have been calculated from the quoted values of the formation energy of the various charge states using Eq. 4.23.

(−0.05 eV to 0.01 eV) than the negatively charged ones (0.05 eV to 0.21 eV); however, these uncertainties are small enough not to affect the stable charge state transition level ordering, discussed in the next section.

We defer detailed discussion of the electronic structure of the different configurations to Sec. 5.5.5; for now, we simply state that all the charge states apart from v^{2-} are in agreement with Watkins' model. For v^{2-}, a completely different 'split vacancy' configuration is found, with a point group symmetry of D_{3d} (in agreement with previous studies by Nieminen *et al.* [202, 214] and Wright [213]). The symmetry predicted by the model (C_{2v}) was found as a metastable state. For all charge states there are no qualitative differences between the Γ-point only and multiple k-point calculations, although the former consistently showed a greater inwards relaxation and thus a smaller defect volume.

We note that the relaxation of v^{1-} is the most problematic; in fact, if the initial random displacement of the atomic positions is not large enough, the system consistently converges to a metastable T_d configuration with a completely spin-polarised triply degenerate defect level (i.e., filled with three spin-aligned electrons). Furthermore, our final lowest configuration is only approximately C_{2v}, as there is a small but noticeable splitting of ∼0.1 Å of the bond lengths between the two pairs of neighbours of the vacancy. Nieminen and Wright both report that the LDA does not give the expected symmetry for this charge state, although they find a D_{3d} configuration. As noted by Ganchenkova *et al.* [214], the vacancy exhibits several metastable configurations with different symmetries but comparable formation energies; therefore, it is likely that the ordering of these configurations is strongly dependent both on the numerical parameters of the simulation and the details of the pseudopotential.

5.5.4 Transition levels

We can calculate the stable charge state transition levels by using the expression in Eq. 4.23. Our transition levels are given in Table 5.4; as can be seen, the Γ-point calculations predict the negative-U behaviour both for the negatively and positively charged levels as expected, and are in good agreement

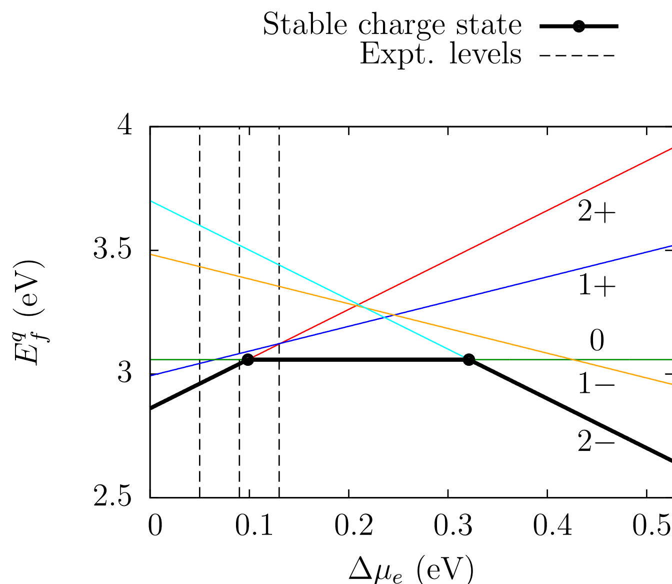


Figure 5.8: Formation energy of the different charge states of the vacancy as a function of the electronic chemical potential (plotted relative to the VBM). The energy range shown covers the DFT band gap for silicon. The thermodynamically stable charge state at each point is highlighted in bold, and the circles indicate the level position for the stable transitions. The dashed vertical lines show the available experimental values for the transition levels. The results shown are for the 256-atom BCC supercell with Γ-point sampling.

with the available experimental results (also shown in Fig. 5.8). Surprisingly, the calculations using a dense k-point sampling ($k_{\mathrm{MP}} = 3$) give the opposite ordering for the sequence v^{2+}, v^{1+}, v^{0}; this results in no transitions in the band gap between these levels, and only v^{0} being a stable charge state. The negative-U behaviour is however still present for the negatively charged levels, although their positioning is lower than for the Γ-point calculations.

The Γ-point, therefore, gives better estimates of the transition levels as compared to experiment. However, as stated previously, the absolute value of the defect formation energy is not as well converged with respect to the dilute limit when using this Brillouin zone sampling. However, calculations of v^{0} and v^{1+} in the 1000-atom supercell at Γ (also included in Tables 5.3 and 5.4) are converged with respect to system size both in terms of the transition

level $E\,(1+/0) = 0.04$ eV and the absolute value of the formation energy $E_f^0 = 3.55$ eV. Therefore, this suggests that at this system size Γ-point sampling can be employed for simultaneous convergence of the both quantities of interest.

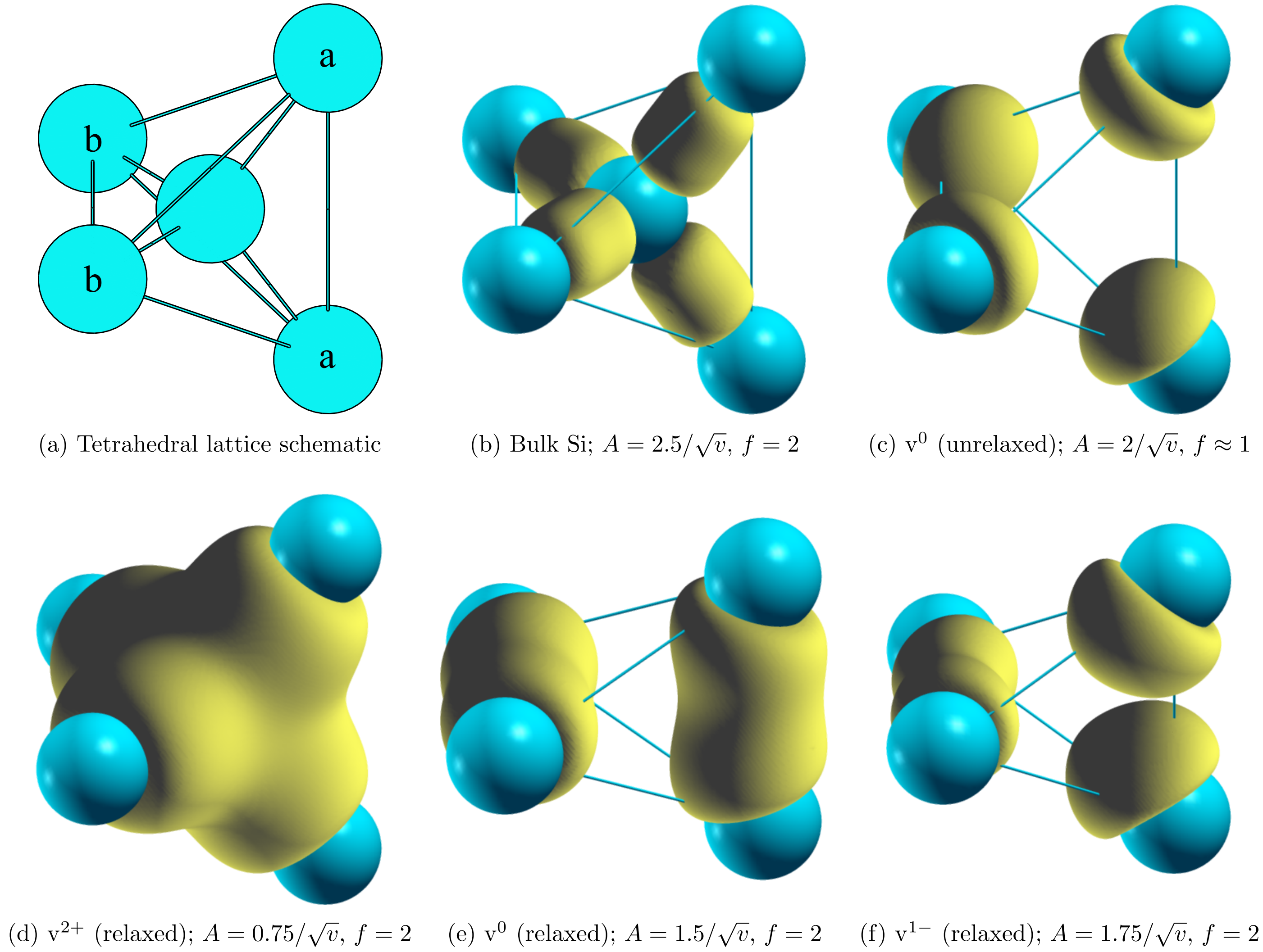
5.5.5 Visualisation using MLWFs

The localised view of the electronic structure provided by Wannier functions is particularly useful in the study of point defects, since the Bloch states associated with the defect levels are not completely localised due to the periodicity of the system; in fact, in small supercells it is not possible to disentangle them from the bulk band structure.

In the case of bulk silicon, previous work has shown that it is possible to recover the typical σ bond orbitals between silicon ions by wannierisation of the valence band; alternatively, the bottom conduction bands (once disentangled from higher bands) give the corresponding antibonding orbitals, while treating both these sets of bands together as a single manifold produces four sp^3 orbitals on each ion [72, 74]. The bulk silicon σ orbitals obtained from our calculations are shown in Fig. 5.9b. Further work that we have carried out on the wannierisation of the valence+conduction manifold of bulk silicon is given in Appendix C, showing that the conventional forward-facing sp^3 orbitals are not the most localised representation of the subspace, as is generally assumed.

For the wannierisation of the defect systems we only use the occupied manifold, except in the case of the unrelaxed vacancy where we include all the defect levels. We can loosely define the concept of a defect Wannier function to be any MLWF in the defect supercell which differs qualitatively from the σ bond MLWFs of the bulk system. For all charge states, such defect Wannier functions are only present within the first ionic shell around the vacancy; between the first and second shell we already recover σ-like bonding orbitals.

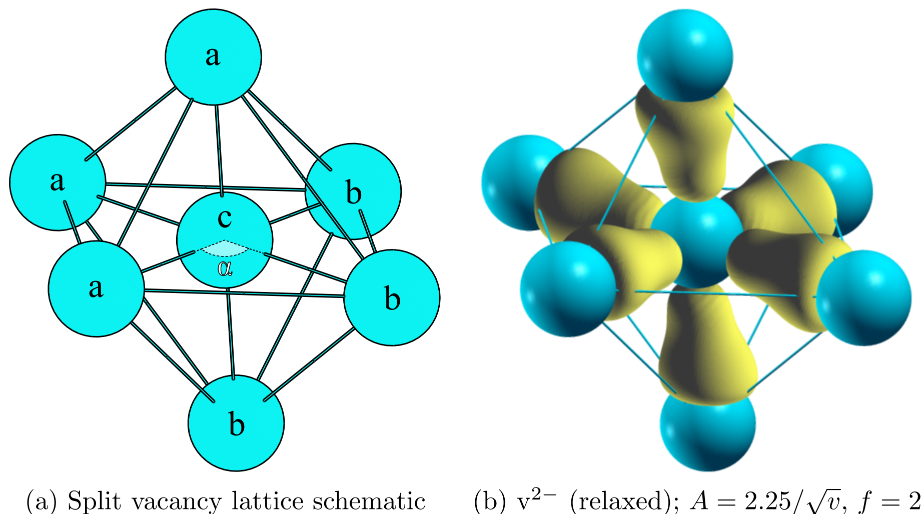
Fig. 5.9c shows the defect Wannier functions for the neutral unrelaxed system; as expected, these are sp^3 orbitals pointing towards the vacancy. This confirms the validity of the central assumption in Watkins' vacancy model, namely that of taking into account only four dangling bond orbitals. We stress that these orbitals cannot be obtained by wannierising the three visible



(a) Tetrahedral lattice schematic (b) Bulk Si; $A = 2.5/\sqrt{v}$, $f = 2$ (c) v^0 (unrelaxed); $A = 2/\sqrt{v}$, $f \approx 1$

(d) v^{2+} (relaxed); $A = 0.75/\sqrt{v}$, $f = 2$ (e) v^0 (relaxed); $A = 1.5/\sqrt{v}$, $f = 2$ (f) v^{1-} (relaxed); $A = 1.75/\sqrt{v}$, $f = 2$

Figure 5.9: Contour-surface plots of the MLWFs most strongly associated with the defect centre for different charge states of the vacancy. Each separate closed surface is an individual Wannier function. The amplitude A of the contour shown in each figure is given in the caption in terms of the primitive cell volume v. In general the Wannier functions also have small negative components on the backbonds (not pictured). f is the electronic occupancy of a single orbital.



(a) Split vacancy lattice schematic (b) v^{2-} (relaxed); $A = 2.25/\sqrt{v}$, $f = 2$

Figure 5.10: Split vacancy configuration for the doubly negative charge state of the defect centre; details of the contour-surface plot of the MLWFs are given in the caption of Fig. 5.9.

defect levels in the gap alone, as we also need the nodeless combination which lies within the valence band; therefore, the entire valence manifold plus the defect levels in the gap must be wannierised as a whole.

The defect Wannier functions obtained for v^{2+}, v^0, and v^{1-} in their relaxed configurations (Figs. 5.9d–5.9f) also support Watkins' model. For the doubly positive vacancy, the defect levels in the gap are empty and therefore not included; the result is a single MLWF corresponding to the symmetric s-type nodeless combination of the four sp^3 orbitals. The neutral vacancy includes one lowered defect level only, and produces the two bonding orbitals between the pairs of nearest neighbours of the defect centre. The singly negative vacancy includes two defect levels in the gap (for the majority spin), producing one bonding orbital (between the pair of ions with the shortest bond length) and two sp^3 orbitals.

Finally, we show the split vacancy configuration obtained for v^{2-} in Fig. 5.10. In this case, one of the neighbours (labelled c) moves halfway along the line connecting it with the vacancy site, thus placing itself at the centre of an octahedron made up of two pairs of second-nearest neighbours (labelled a and

b). This octahedron is approximately regular (having as faces equilateral triangles); however, there is a small distortion due to the difference in distance between a–a/b–b pairs and a–b pairs. This distortion is quantified by the angle α between any three ions a, c, b; for a regular octahedron, α is a right angle, and so the distances a–a, b–b, a–b are identical. In general, the split configuration can be specified with only two parameters: the distortion angle α and the distance a–c between the central ion and any of the ions forming the octahedral cage.

The wannierisation of the occupied manifold produces six defect Wannier functions, each one a bond between the central ion and one of its neighbours. This suggests that the c ion is forming six sp^3d^2 orbitals which then bond to the dangling sp^3 orbitals of the a, b ions. The inwards relaxation of the ionic positions also shortens the bonds and reduces the distortion of the octahedron. This configuration is favoured since these bonds are shorter than the ones obtained by dimerisation in the C_{2v} arrangement predicted by Watkins' model, and so are closer in length to the bulk silicon bond. However, only in the case of the doubly negative vacancy are there enough electrons to fully occupy the six orbitals.

5.6 Conclusions

In this chapter we have studied the silicon vacancy in all its charge states using the supercell approach and plane-wave pseudopotential DFT. Our calculations confirm the slow finite size convergence of defect formation energies and transition levels, due to electrostatic interactions and wavefunction overlap between periodic images of the defect, and long-ranged ionic relaxations. The impact of each of these has been quantified, and it has been found that all three provide non-negligible contributions to the total error. In particular, we note that the importance of wavefunction overlap, which is a purely quantum-mechanical effect, is generally overlooked in studies of point defects, and could therefore account for the well-known unreliability of classical finite size correction schemes. In fact, we find that such schemes can actually further slow down the convergence, due to the simplified view of the electronic structure

that is employed; we defer further discussion on this topic to the next chapter (see Sec. 6.3).

In addition, due to the hybridisation of the defect levels with the perfect crystal band structure, we find that the choice of k-points has a noticeable impact on the results. In particular, Γ-point calculations converge faster than calculations with uniform, multiple k-point sampling when considering stable charge state transition levels (given by differences in formation energies of different charge states), and vice-versa when considering the absolute value of formation energies.

Due to the overall slow convergence, the calculations may still benefit from the use of even larger supercells than the ones currently employed. This, however, will present additional numerical challenges: the defect formation energy is an *intensive* quantity that is obtained by taking the difference of two total energies that are *extensive* with the system size. As a result, to achieve a given accuracy for E_f as the supercell size is increased, it is necessary to proportionately increase the precision per atom in the total energy calculations in order to avoid the result being swamped by numerical noise. Encouragingly, however, we have shown that linear-scaling DFT methods can be used to obtain results with a comparable level of accuracy to conventional plane-wave methods.

Finally, we have also introduced two methods for correcting the alignment of the valence band maximum for charged defects: one is based on averaging the electrostatic potential using a Voronoi cell construction, and the other on Hamiltonian matrix elements in a basis of MLWFs. The two methods give excellent mutual agreement and constitute simple and robust ways to calculate potential alignment corrections. However, the determination of ε_v remains a somewhat uncontrolled source of error, as demonstrated by the discrepancy between the potential alignment correction ('VBM') and total energy difference ('Hole') approaches, which can be as small as 0.01 eV (for positively charged defects), or as large as 0.2 eV (for negatively charged defects).

Chapter 6

Gold in silicon

In this chapter we present the results from our investigation of the gold substitutional defect in bulk silicon (denoted Au^q) using DFT. The insights obtained from the silicon vacancy regarding the finite size convergence of defect properties in the supercell approach are applied to the study of this impurity-related defect, which has not yet been well-characterised by modern first-principles methods.

Our calculations of the gold centre represent a significant improvement on previous theoretical studies of this defect, for the following reasons: (i) a broader range of charge states has been investigated (from 1+ to 3−), (ii) large supercells are used (up to 864 atoms), and (iii) careful consideration has been given to finite size effects, in particular, the sampling of electronic states in the Brillouin zone. We argue that it is important for the sampled states to possess the full point group symmetry of the isolated defect in an infinite crystal, and investigate the efficiency of various high-symmetry k-point sampling schemes for the 256-atom BCC supercell; the Γ-point only sampling is shown to offer the fastest convergence of the stable charge state transition levels with system size.

From our converged results we show that, similar to the vacancy, the gold substitutional defect becomes a non-spin-polarised negative-U centre due to the effect of Jahn-Teller distortion. The resulting stable charge state transition levels are in excellent agreement with the donor and acceptor levels measured

experimentally, and therefore allow for a clear identification of these levels; our results suggest that the assumption used in previous studies on the nature of the two experimental levels is incorrect.

The rest of this chapter is organised as follows: in Sec. 6.1, we review the previous experimental and theoretical results for the gold impurity. In Sec. 6.2, we give the technical details of our simulations. In Sec. 6.3, we discuss the application of corrections for the electrostatic interaction of charged point defects, and show preliminary results for the case of the gold centre. In Sec. 6.4, we present our main results; first we describe the convergence properties of the neutral gold impurity (Sec. 6.4.1), then we describe the results obtained for different charge states of the defect (Sec. 6.4.2), in terms of the transition levels (Sec. 6.4.3) and Watkins' vacancy model (Sec. 6.4.4), and finally we describe the electronic structure of system by examining the MLWFs that are obtained from the valence+deep-level manifold (Sec. 6.4.5). In Sec. 5.6, we give a brief summary of our main conclusions.

6.1 Previous studies

The role of gold as an impurity in silicon has been extensively studied, due to its technological importance in the semiconductor industry (notably, through the use of gold contacts to semiconducting devices [223]). Gold doping introduces deep defect levels into the band structure of silicon; as previously noted, such deep-level impurities are extremely effective in reducing minority carrier lifetimes, and can cause a strong reduction in the conductivity of the sample [112]. This system has also received renewed interest due to the use of gold as a catalyst in the growth of semiconducting nanowires using the vapor-liquid-solid (VLS) method [224], which can result in the presence of gold impurities in excess of the bulk solubility [225, 226].

Since 1957, two defect levels have been identified as the main contributions from gold impurity centres: a donor level at $\varepsilon_v + 0.35 \pm 0.02$ eV, and an acceptor level at $\varepsilon_v + 0.62 \pm 0.02$ eV [227]. These measurements have been confirmed by several different experimental techniques (for a review, see Ref. [228]), including a relatively recent study [229] using DLTS measurements, which has

clearly identified the defect centre as a single amphotheric gold substitutional. However, the electronic structure of this defect is still a matter of debate; this is mainly due to the lack of an EPR signal, in contrast to the isoelectronic Pt^- defect, which has been well characterised using this technique [230]. A possible explanation for this situation has been given by Anderson [231], who proposed that rapid tunnelling of the defect between two trigonal C_{2v} configurations results in an average value of $g_\perp \simeq 0$. This would make the microwave transition probability between states small and hence any EPR signal difficult to detect. This explanation is supported by Watkins *et al.* [232] in Zeeman studies of the substitutional gold excitation spectra; on the other hand, Son *et al.* [233] have attributed an EPR spectrum to this defect centre, concluding that it is paramagnetic with $S = 1/2$, but with substantially different g values to those found in Ref. [232], which makes it unlikely that both studies are examining the same centre. In each case there is a problem with unambiguously identifying the defect centre, due to the presence of other impurities and the large number of Au-related complexes that are known to exist [234–236].

Computationally, self-consistent Green's function calculations of the gold substitutional defect in an undistorted silicon lattice [228] have shown the possibility of both the donor and acceptor levels arising from the same centre, in contrast to the results from a previous cluster model [237, 238] using the multiple scattering Xα [239] (MS-Xα) method; however, the predicted level positions are 0.21 eV and 0.26 eV higher than the experimental measurements for the donor and acceptor levels, respectively. More recently, a DFT cluster simulation [179] using a relaxed ionic configuration also calculated these two levels, although their positions were found to be 0.14 eV and 0.12 eV lower than the experimental donor and acceptor levels, respectively, in spite of the application of an empirical correction (this is the 'reference defect' method discussed in Sec. 4.4.5; for this study, the C_i defect in silicon was taken as the reference).

The LCAO model of the silicon vacancy, introduced in the previous chapter, has also been adapted for the case of transition metal substitutional impurities (belonging to the Ni- and Cu-headed groups of the period table) in silicon [240, 241]. The defect centre is now treated as a perturbed vacancy; the

impurity atom's valence d orbital is taken to be fully occupied, with the remaining valence electrons forming vacancy-like defect levels. This model explains both the paramagnetic behaviour and the lattice distortion observed experimentally for the negatively charged palladium and platinum defect centres in silicon [242], as well as the negatively charged nickel centre in germanium [243]. Furthermore, the early computational studies of the gold centre cited above have confirmed the symmetry of the defect states in the gap to be those expected from the vacancy model [228, 237, 238]. We shall refer to this model in the discussion of our DFT results in Secs. 6.4.4 and 6.4.5.

6.2 Computational methodology

As previously noted, we apply the results from our investigation of the finite size convergence properties of the vacancy in the supercell approach to the study of the gold defect centre. Consequently, the general methods already described for the plane-wave DFT simulations (Sec. 5.3.1) and the calculation of MLWFs (Sec. 5.3.3) are also applied in the simulations presented in this chapter. Therefore, we now describe only those technical details that are specific to the simulation of the gold substitutional defect.

The definition of the defect formation energy is almost identical to that given in Eq. 5.1 for the vacancy; however, we must now include the chemical potential of the gold species μ_{Au}:

$$E_f^q = E^{\mathrm{def},q} - \left(\frac{N-1}{N}\right) E^{\mathrm{bulk}} - \mu_{\mathrm{Au}} + q\left(\varepsilon_v + \Delta\mu_e\right). \tag{6.1}$$

μ_{Au} is taken to be the energy per atom of ideal FCC gold calculated with the DFT-optimised lattice parameter. We note that the chemical potential of the impurity species does not affect the stable charge state transition levels (the main quantity of interest for the present study), as this term in the defect formation energy cancels between different charge states in Eq. 4.23. The valence band maximum reference ε_v for the electronic chemical potential is calculated by the total energy difference approach (between the neutral bulk supercell and the same supercell with a single electron hole), as it is computationally

simpler than the potential alignment method (the formation energy in this approach is entirely determined by total energy calculations), and has proven as precise with respect to experiment for calculating transition levels.

We study supercells of up to 864 atoms using the CASTEP code; these correspond to those employed for the vacancy up to 256 atoms (Tables 5.2 and 6.1), and a large 864-atom BCC supercell (constructed from $3 \times 3 \times 3$ BCC unit cells of 32 atoms). For the Brillouin zone sampling, we use the MP k-point sampling scheme employed previously; additionally, for large supercells, we perform calculations on specific high-symmetry k-points (these are detailed in the text). The LDA is used in all calculations to describe exchange and correlation. The BFGS scheme is used for calculating relaxed geometries; our convergence tolerance parameters for this procedure are 6×10^{-3} eV/Å for the RMS force, 5×10^{-4} Å for the RMS ionic displacement, and 1×10^{-5} eV for the energy per atom difference between BFGS iterations[1].

We use the same ultrasoft and norm-conserving pseudopotentials described previously for silicon, and two similar pseudopotentials for gold (we do not mix ultrasoft and norm-conserving pseudopotentials in the same calculation). The gold USPP has 19 valence electrons (corresponding to the electronic configuration $5s^2 5p^6 5d^{10} 6s^1$ in the free atom), and the NCPP has 11 (the $5s^2 5p^6$ orbitals are treated as core states in this case). The USPP gives the optimised lattice constant for bulk FCC gold to within 1% of experiment (4.04 Å, compared with an experimental value of 4.08 Å), and the bulk modulus to within 10% (190 GPa, compared with an experimental value of 173 GPa). The elastic constants for the FCC lattice C_{11}, C_{12}, and C_{44} are also found to be in reasonable agreement with experiment (with errors within 5%, 10%, and 20%, respectively). The NCPP, instead, gives the optimised lattice constant to within 2%

[1]We note that the determination of the point group symmetry of the relaxed defect centre is now complicated by the presence of the gold ion, that might shift with respect to the centre of mass of the four neighbouring silicon ions. Therefore, we have written a code for automatically finding the permissible symmetry operations for a set of ionic coordinates, from which the molecular point group symmetry can be easily deduced; this is done by using a Nelder-Mead downhill simplex algorithm [244] to find (for each type of symmetry element) the axes which minimise an error function describing the accuracy of the symmetry operation in terms of the RMS difference between the original and transformed sets of ionic coordinates.

of experiment (4.01 Å), and the bulk modulus to within 20% (207 GPa).

6.3 Electrostatic image charge corrections

Before presenting the main results for the gold centre, we shall briefly discuss the merits of supercell image charge correction schemes for charged point defects in relation to our particular defect system. The idea was first introduced by Leslie and Gillan [245], who proposed the first-order (monopole-monopole) correction term for the defect formation energy

$$\Delta E_f^q = \frac{q^2\alpha}{2L\varepsilon} + \mathcal{O}\left(L^{-3}\right), \tag{6.2}$$

where q is the charge of the defect centre, L is the cube root of the supercell volume, ε is the dielectric constant of the host crystal, and α is the Madelung constant of the cell (related to its shape). This correction, therefore, approximates the system as an infinite periodic lattice of interacting point charges. Makov and Payne [246] extended the correction to include the third-order term, taking into account the monopole-quadrupole interaction:

$$\Delta E_f^q = \frac{q^2\alpha}{2L\varepsilon} - \frac{2\pi qQ}{3L^3} + \mathcal{O}\left(L^{-5}\right), \tag{6.3}$$

where Q is the second radial moment of the aperiodic charge associated with the defect centre. As noted by Lento *et al.* [181], this is not a well-defined quantity, as the treatment of the Coulomb interaction in dielectric materials requires the separation of the external charges, from which Q should be calculated, and the induced charges due to screening. By taking the difference in the ground-state charge density of the supercell with and without the defect, however, we calculate the sum of these two quantities. It is therefore unclear how to proceed; for the case of the silicon self-interstitial, Lento *et al.* use the charge density of an isolated silicon atom in free space. This approach is clearly not appropriate for substitutional defects, especially if the valence electrons become strongly bonded with other lattice sites.

Leaving aside the uncertainties in calculating Q, the accuracy of this ap-

proach has been questioned by several studies [247–249], that suggest that the defect charge is too delocalised to be well-represented in practice by a multipole expansion. Nevertheless, it has also been empirically observed [155, 250, 251] that the finite size scaling of the formation energy for charged defects calculated from a number of different supercell sizes can sometimes be accurately fitted to a function of the same form as that given by Eq. 6.3:

$$E_f^q(L) = E_f^q(L = \infty) + \frac{\gamma_1}{L} + \frac{\gamma_3}{L^3}, \tag{6.4}$$

where $E_f^q(L = \infty)$ is now taken to be the converged formation energy in the dilute limit. However, the sign of the first-order term in some cases is found to be opposite to that suggested by Eqs. 6.2 and 6.3, thereby seriously calling into question the validity of the extrapolation to the dilute limit. Lany and Zunger [127] have attributed this to additional errors from potential alignment effects.

We have tested the effectiveness and reliability of two possible electrostatic correction schemes on the unrelaxed gold substitutional defect in its most highly charged state, Au^{3-}, for which the correction is expected to have the most noticeable effect. The two schemes we use are: (i) the Leslie-Gillan first-order correction given by Eq. 6.2, and (ii) the empirical finite size scaling fitting given by Eq. 6.4. We avoid the Makov-Payne correction due to the difficulty in defining Q; however, the effectiveness of the Leslie-Gillan correction should give an indication of the general reliability of the multipole expansion approach.

Fig. 6.1 shows the results for both schemes. As can be seen, the Leslie-Gillan correction clearly does not speed up the convergence; indeed, its application *increases* the overall range of formation energies between the smallest and largest supercells by 38%. The correction increases the formation energy, as must always be the case for the first-order term in the multipole expansion.

The reliability of the finite size scaling fitting procedure, instead, is tested by comparing fits to different data sets, given by the three supercell geometries. Unfortunately, there is a large discrepancy (on the order of 1 eV) between the values of $E_f^q(L = \infty)$ obtained in each case. We have further tested the robustness of the method by discarding the smallest supercell from the fit of

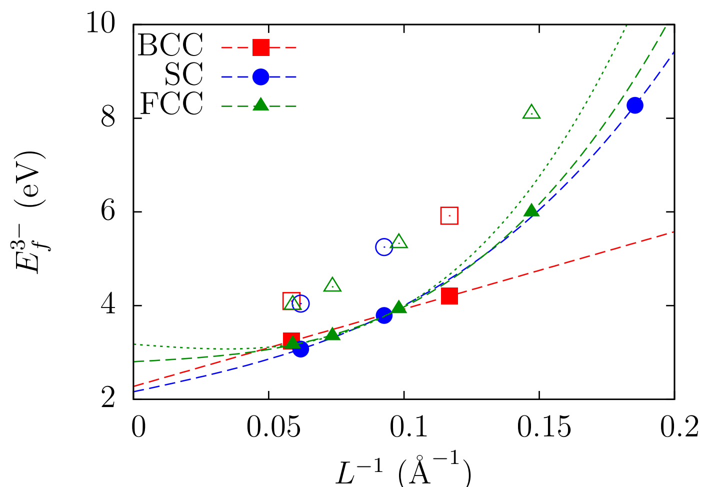


Figure 6.1: Electrostatic corrections of the defect formation energy for the Au^{3-} centre. Filled symbols show the uncorrected formation energy, and empty symbols show the same quantities with the addition of the Leslie-Gillan correction term from Eq. 6.2. The dashed lines show the best fit to the data of a first-order finite size scaling term (for BCC), and first- and third-order terms (for SC and FCC); in addition, the fine dashed green line shows the best fit obtained for FCC after discarding the smallest supercell ($L^{-1} \simeq 0.15$ A^{-1}).

the FCC data points (as this geometry has the most available points); this results in a non-negligible shift of $E_f^q (L = \infty)$ of 0.5 eV. Finally, we note that the fitting procedure, despite displaying a large uncertainty in the extrapolated result, suggests that $E_f^q (L = \infty) < E_f^q (L)$ for the calculated range of L, and is therefore in qualitative disagreement with the Leslie-Gillan correction.

In conclusion, therefore, we have shown that neither the direct application of an electrostatic correction based on multipole expansion, nor the fitting of the available data to an empirical finite scaling function gives a clear improvement on the uncorrected results for the defect formation energy. We suggest, in fact, that the multipole expansion greatly overestimates the magnitude of the necessary correction, due to the noticeably delocalised and non-point-like nature of the defect wavefunction (as demonstrated both by our calculation of the kinetic energy contribution to the total defect formation energy in Sec. 5.5.1, and the MLWF analysis of the defect levels in Sec. 6.4.5); we note that if we

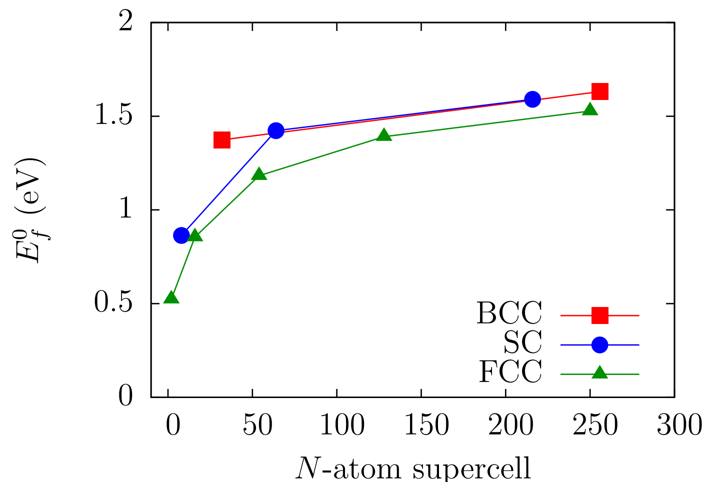


Figure 6.2: Convergence of the defect formation energy with system size for different supercell geometries.

consider the opposite limit of a completely delocalised defect charge (which might arguably be a better approximation of the defect in small and medium size supercells), the electrostatic correction now tends to zero. We therefore do not apply an electrostatic correction in the results presented in the rest of this chapter, and, as for the vacancy, use a large supercell to check the convergence of the main quantity of interest in our investigation.

6.4 Results

6.4.1 The neutral defect centre

The analysis of the gold centre offered by Watkins' vacancy model suggests definite similarities between the two point defects, notably the bonding of the valence electrons at the defect centre and the resulting symmetry of the defect states. This, in turn, also suggests the possibility of the two defects exhibiting similar finite size convergence properties in electronic structure calculations; to this end, we first simulate the neutral gold substitutional defect centre, investigating the converge properties of the unrelaxed and relaxed lattices.

N	Symm.	k^{MP}	k-point vol. (10^{-3}Å^{-3})	E_f^0 (eV)
2	FCC	8	12.35	0.52
8	SC	7	4.61	0.86
16	FCC	6	3.66	0.86
32	BCC	5	3.16	1.37
54	FCC	6	1.08	1.18
64	SC	5	1.58	1.42
128	FCC	4	1.54	1.39
216	SC	3	2.17	1.59
250	FCC	4	0.79	1.53
256	BCC	4	0.77	1.63

Table 6.1: List of all supercells up to 256 atoms with their respective symmetries. Also listed are the value of k^{MP} used (converged with respect to the formation energy for each supercell), the corresponding k-point volume in reciprocal space, and the unrelaxed defect formation energy.

Table 6.1 lists the defect formation energy obtained for the different supercells with an undistorted lattice and ultrasoft pseudopotentials. We use a converged value of 400 eV for the cut-off energy (for a convergence tolerance of 10 meV in the formation energy). The pattern of convergence observed for the defect formation energy (shown in Fig. 6.2) is extremely similar to that found for the vacancy (Fig. 5.4a): there is a monotonically-increasing convergence for individual supercell geometries, with BCC and SC supercells showing a faster convergence than FCC. However, the overall range of values between all system sizes is 25% smaller than for the vacancy, and the difference in formation energy between the three largest supercells (216-atom SC, 250-atom FCC, and 256-atom BCC) is only 40 meV, 50% smaller than the equivalent value for the vacancy formation energies. Table 6.1 also shows the converged (dense) k-point grids used for the calculations; these are almost identical to those employed for the vacancy supercells, again suggesting a noticeable similarity between the electronic structure of the two defect centres.

The converged value of the unrelaxed gold defect formation energy itself

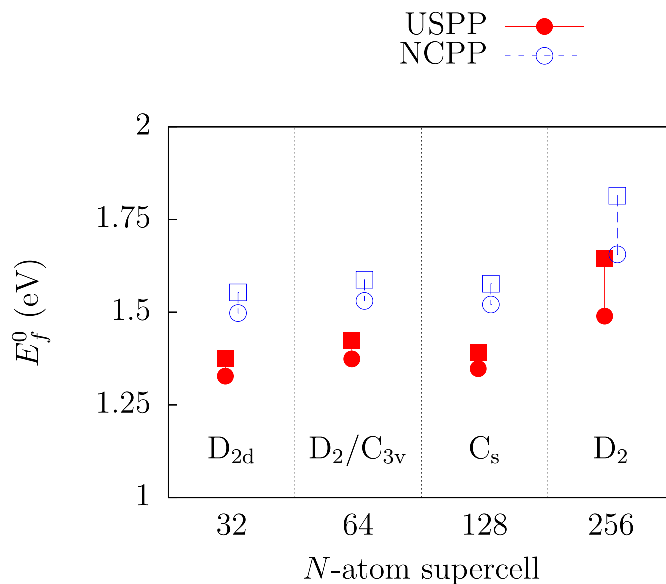


Figure 6.3: Relaxation effects for the neutral gold substitutional defect. Squares show the defect formation energy for the unrelaxed lattice and circles for the relaxed one. Labelled below is the point group symmetry of the defect centre after relaxation (USPP/NCPP).

(∼1.6 eV) is much smaller than the corresponding vacancy formation energy; however, we note that the gold chemical potential should be considered as an upper bound to the possible range of experimental values (see Sec. 4.4.2), and, hence, the calculated formation energy represents a lower bound estimate.

As predicted by Watkins' model, the simulations with an undistorted lattice result in three defect levels within the silicon band gap, degenerate at Γ. When the ionic positions are allowed to relax, we therefore expect the degeneracy of these levels to be lifted by Jahn-Teller distortion. For Au^0, this is assumed to result in the same symmetry as the relaxed v^{1-} centre (C_{2v}). Our simulations on the relaxed geometry for the 256-atom supercell, however, predict a lower symmetry of D_2 (we shall discuss symmetry considerations in more detail in Sec. 6.4.4). The calculations are performed with ultrasoft (norm-conserving) pseudopotentials, using a converged energy cut-off of 400 eV (800 eV), and the Brillouin zone samplings listed in Table 6.1.

Fig. 6.3 shows the finite size dependence of the relaxed symmetry of the

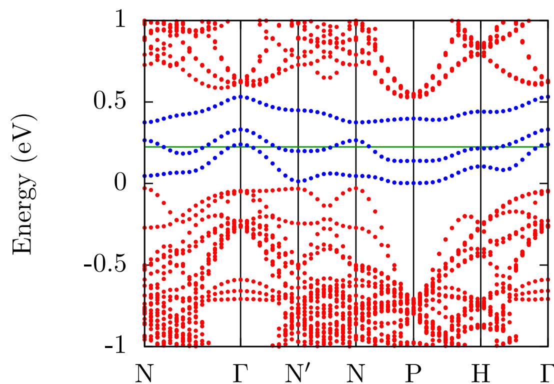


Figure 6.4: Band structure for the 256-atom BCC supercell with a relaxed neutral gold substitutional defect. The full horizontal line (in green) indicates the level of the highest occupied orbital. The three defect levels are shown in blue, and all other bands are shown in red. The corresponding band structures for bulk silicon and the relaxed vacancy are shown in Figs. 5.7b and 5.7c, respectively.

gold defect centre; as with the vacancy (Fig. 5.6), only the largest supercell of 256 atoms shows a substantial lowering of the defect formation energy from the relaxation (by 0.3 eV, compared with 0.7 eV for the vacancy), revealing the spurious effect of elastic interactions between images of the defect centre for smaller supercells[2]. However, even for the largest supercell there is only a modest change in volume at the defect site (an increase of less than 5%, compared with a reduction of more than 40% for the vacancy). We also note that the symmetry of the smaller supercells is not consistent (as is seen for the vacancy, which exhibits T_d symmetry for all supercells smaller than 256 atoms), and is only higher than D_2 for the smallest system size.

Fig. 6.4 shows the band structure calculated for the 256-atom supercell with the neutral gold defect centre after relaxation. The three defect levels are split by Jahn-Teller distortion, and result evenly spread out, covering the entire DFT band gap. Additionally, at this system size there is still a noticeable

[2]We note an almost constant offset of ∼0.17 eV in the defect formation energy between the two pseudopotentials; this can be attributed to the difference in the ideal structure obtained for FCC gold, from which the impurity chemical potential is calculated.

dispersion on the order of 0.1 eV; nonetheless, the defect levels are clearly distinguishable in the band gap, and are not entangled with either the valence or conduction band-edge states, as is the case for the relaxed neutral vacancy. This band structure can be directly compared with the schematic level diagram in Fig. 6.6.

6.4.2 Charged defect centres

We now consider the gold substitutional defect in all its charge states from +1 to −3; these correspond to the three defect levels in the band gap being filled with between two and six electrons. We have calculated the relaxed symmetry and defect formation energy for each charge state using the 256-atom supercell and USPP. We repeat the calculations with both Γ-point sampling and a dense $k_{\mathrm{MP}} = 3$ grid. The results are summarised in Table 6.2.

All charge states show a small outwards relaxation, which decreases as more electrons are added to the defect centre. In the majority of cases there is no qualitative difference in relaxation between the Γ-point only and multiple k-point calculations, although the latter results in a higher symmetry for Au^{3-} (this cannot, however, be considered a Jahn-Teller distortion, as the unrelaxed lattice does not feature a partially filled group of degenerate levels). In general, however, the difference between symmetries is found to be small, and the relaxation procedure shows evidence of an extremely flat energy landscape. This suggests that the system should be considered as a multi-symmetry defect, since the small energy barriers between different symmetries will allow it to explore several metastable configurations even at low temperatures. An analogous situation has been described for the vacancy in a previous DFT study [214] (see also Sec. 5.5.3). However, comparing the results presented in this and the previous chapter, it is clear that this effect is even more noticeable for the gold centre; this should be expected, as the relaxation of the lattice that causes the change in symmetry is much less pronounced in this case. Unfortunately, due to the numerical noise in the ionic positions at the end of the relaxation procedure, this also makes the point group symmetry of the defect centre in our simulations less well-defined than for the vacancy, as

					Bond lengths (Å)					Defect vol.	
Charge	k^{MP}	Rel.	Pspot.	Symm.	a–a	b–b	a–b	a–c	b–c	(Å^3)	E_f^q (eV)
1+	1	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	1.07
	3	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	1.46
	1	Yes	USPP	D_{2d}	3.71	3.71	3.95	2.37	2.37	6.78	0.85
	3	Yes	USPP	D_{2d}	3.72	3.72	3.99	2.39	2.39	6.90	1.26
0	1	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	1.39
	3	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	1.64
	1	Yes	USPP	$\sim D_2$	4.00	3.98	3.86/3.73	2.36	2.37	6.72	1.22
	3	Yes	USPP	$\sim D_2$	4.02	4.01	3.85/3.75†	2.37	2.37	6.79	1.49
	1	No	NCPP	T_d	3.81	3.81	3.81	2.33	2.33	6.50	1.55
	3	No	NCPP	T_d	3.81	3.81	3.81	2.33	2.33	6.50	1.81
	1	Yes	NCPP	D_2	4.00	3.99	3.86/3.73	2.37	2.37	6.74	1.42
	3	Yes	NCPP	$\sim D_2$	4.03	4.01	3.85†/3.75	2.37	2.38	6.80	1.66
1−	1	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	1.87
	3	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	1.98
	1	Yes	USPP	$\sim D_{2d}$	4.02	4.02	3.75†	2.35	2.35	6.60	1.59
	3	Yes	USPP	D_{2d}	4.04	4.04	3.76	2.36	2.36	6.64	1.76
2−	1	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	2.39
	3	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	2.45
	1	Yes	USPP	$\sim D_{2d}$	4.02	4.01	3.75†	2.35	2.35	6.60	2.18
	3	Yes	USPP	$\sim D_{2d}$	3.98	3.98	3.77†	2.35	2.35	6.64	2.30
3−	1	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	2.95
	3	No	USPP	T_d	3.81	3.81	3.81	2.34	2.34	6.54	3.01
	1	Yes	USPP	$\sim D_{2d}$	4.03	4.00	3.76†	2.35	2.36	6.61	2.77
	3	Yes	USPP	$\sim T_d$	3.82	3.80	3.83	2.34	2.34	6.60	2.88

Table 6.2: Summary table of the results obtained for the 256-atom BCC supercell. Bond lengths are given between pairs of Si ions a,b surrounding the defect centre and the Au impurity ion c (as labelled in Fig. 6.6). Equivalent bonds in each system agree to within 0.01 Å, except for ones labelled with a dagger (†), which agree to within 0.03 Å.

the differences in bond lengths between different point groups is smaller.

We shall make use of the information in Table 6.2 in discussing the Jahn-Teller distortion and electronic structure of the system in Secs. 6.4.4 and 6.4.5. However, before doing so, we first turn to the central question in our study of gold in silicon: how well do the *converged* DFT transition levels compare with experiment, and what can they tell us about the nature of this defect centre?

6.4.3 Transition levels

The stable charge state transition levels obtained from our calculations are given in Table 6.3 for a number of different relaxed and unrelaxed geometries and k-point sampling schemes. We also list the experimentally observed donor and acceptor levels, and the transition levels reported from two previous computational studies. Both of these studies restricted their investigation to the $E\,(1+/0)$ and $E\,(0/1-)$ transitions, assuming them to correspond to the two experimental levels; the discrepancies between the calculated and experimental values, on the order of 0.1 eV, were hypothesised to be due either to lattice relaxation [228] (not included in the model) or finite size effects [179]. While it is clear from our calculations that both of these factors are crucial for correctly placing the transition levels, we shall argue that the underlying assumption of the assignment of the experimental levels is incorrect; instead, we propose that the two transitions of interest are $E\,(1+/1-)$ and $E\,(1-/3-)$.

We have already shown for the silicon vacancy that Γ-point sampling is preferable for calculating the transition levels, as the calculated levels for the 256-atom supercell with this sampling are in good agreement (within 0.03 eV) both with experiment and with the results obtained for a large 1000-atom supercell; a dense k-point sampling at this system size, instead, results in qualitatively wrong behaviour (the transition levels of interest are not in the band gap). Due to the noticeable similarities in the finite size convergence properties of the gold centre and the vacancy, shown in the previous sections, we expect the Γ-point sampling to give the most accurate results for the gold impurity transition levels; we consider this case next.

The levels calculated with Γ-point sampling are shown in Fig. 6.5. As can

		$N = 256$						$N = 864$		
		Γ ($k^{\mathrm{MP}} = 1$)		$k^{\mathrm{MP}} = 3$		$\Gamma + \mathrm{H}$	P	Γ ($k^{\mathrm{MP}} = 1$)		
	Expt. [227]	Rel.	Unrel.	Rel.	Unrel.	Rel.	Rel.	Rel.[a]	Ref. [228]	Ref. [179]
$E\,(1{+}/0)$	-	0.38	0.32*	0.23*	0.18*	0.27	0.54	-	0.56	0.21
$E\,(1{+}/1{-})$	0.35	0.37*	0.40	0.25	0.26	0.28*	0.55*	0.34	-	-
$E\,(0/1{-})$	-	0.37	0.48*	0.27*	0.34*	0.28	0.55	-	0.88	0.50
$E\,(0/2{-})$	-	0.48	0.50	0.41	0.40	-	-	-	-	-
$E\,(1{-}/2{-})$	-	0.59	0.52*	0.54*	0.46*	-	-	-	-	-
$E\,(1{-}/3{-})$	0.62	0.59*	0.54	0.56	0.51	-	-	-	-	-
$E\,(2{-}/3{-})$	-	0.59	0.56*	0.58*	0.56*	-	-	-	-	-

Table 6.3: Transition levels for the 256-atom and 864-atom BCC supercells. Results from previous studies are also shown. Asterisks (*) denote thermodynamically stable transitions. Experimental values for the donor and acceptor levels are placed in line with our proposed corresponding transition levels, as explained in the text. All values are given in eV.

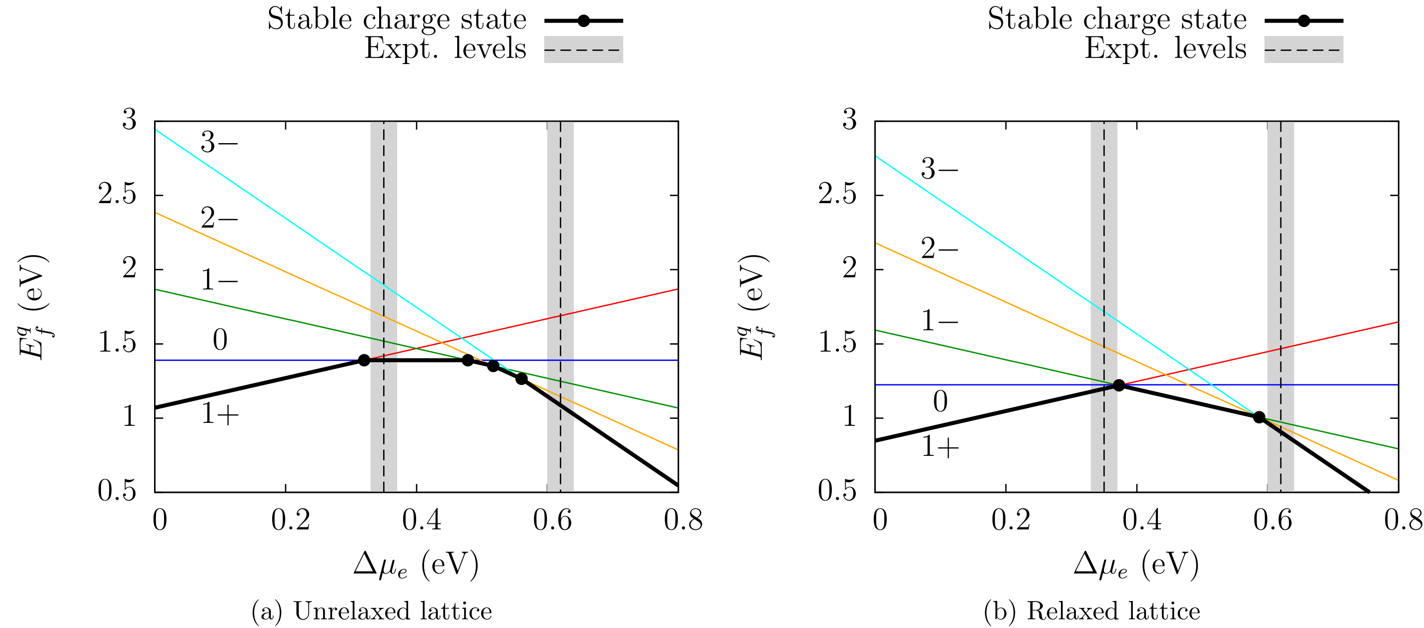
[a] For this system we use the relaxed atomic positions obtained for the 256-atom Au^{1+} and Au^{1-} supercells with $k^{\mathrm{MP}} = 1$, embedded in the larger supercell.

be seen, the ionic relaxation has a large effect on the level ordering and positions. For the unrelaxed lattice, all the charge states from 1+ to 3− exhibit a thermodynamically stable region, with the spin-polarised neutral charge state having the lowest defect formation energy for a range of $\Delta\mu_e$ of width 0.16 eV. The lattice relaxation, however, significantly lowers the formation energy of Au^{1-} (between 0.07 eV and 0.16 eV more than the other charge states), resulting in both the neutral and doubly negative charge states being cut off, and two thermodynamically stable double electron transitions appearing, from Au^{1+} to Au^{1-}, and from Au^{1-} to Au^{3-}. In other words, the defect centre is a negative-U system as a direct consequence of the Jahn-Teller lattice distortion, analogously to the well-known negative-U behaviour of the vacancy.

The position of the two thermodynamically stable transition levels for the relaxed system $E\,(1+/1-)$ and $E\,(1-/3-)$ are in good agreement with the donor and acceptor levels measured experimentally, with a discrepancy of +0.02 eV and −0.03 eV, respectively. We note that experimental measurements also vary on the order of 0.01 eV [227–229].

Table 6.3 also reports the transition levels calculated using $k_{\mathrm{MP}} = 3$; in this case, the levels are shifted closer to the valence band edge and the negative-U behaviour is only approximate, with small regions of stability of width 0.04 eV for both Au^0 and Au^{2-}. A similar effect is observed for the vacancy, as the transitions for the sequence v^0, v^{1-}, v^{2-} with $k_{\mathrm{MP}} = 3$ are also shifted towards the valence band edge with respect to those calculated with Γ-point sampling, and the transitions for the sequence v^{2+}, v^{1+}, v^0 with $k_{\mathrm{MP}} = 3$ are located within the valence band instead of the band gap (Table 5.4).

Our calculations on the two defect centres clearly show the results using Γ-point sampling to be better converged with respect to the large supercell limit (see Sec. 5.5.4 and discussion below), as well as in better agreement with the available experimental values, than those from a dense k-point sampling. Why is this? Due to the importance of relaxation on the transition level positioning, we argue that a key consideration is the correct description of the symmetry-breaking Jahn-Teller distortion of the truly isolated defect centre within the constraints of a periodically repeated supercell. This may be understood as follows:



(a) Unrelaxed lattice

(b) Relaxed lattice

Figure 6.5: Formation energy of the different charge states of the defect as a function of the electronic chemical potential (plotted relative to the VBM). The thermodynamically stable charge state at each point is highlighted in bold, and the circles indicate the level position for the stable transitions. The dashed vertical lines show the experimental values for the transition levels (the uncertainty range is shown in grey). The results shown are obtained with Γ-point sampling.

Any substitutional impurity in an unrelaxed diamond lattice has T_d point group symmetry; a repeating supercell lattice of such impurities will also have T_d symmetry, provided that the supercell is in the cubic crystal system (as is the case for all those used in this study). The Γ-point is somewhat special, in that the Bloch states at Γ form an irreducible representation of the full point group symmetry of the supercell lattice (i.e., the symmetry of the k-point is the same as that of the lattice, as described in Fn. 5 in Sec. 3.2). Hence, when sampling the Brillouin zone of the supercell at the Γ-point only, the Kohn-Sham energy level degeneracies will be correctly described, and the levels correctly occupied (with reference to the isolated system). This is not true, in general, for other k-points, for which the degeneracy of the levels is lifted even without lattice distortion. In fact, amongst the high-symmetry points of the three cubic Bravais lattices, there are only three k-points (other than Γ) that possess the required degree of symmetry: $R = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ for the SC lattice, and $H = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$ and $P = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ for the BCC lattice[3]; there are none for the FCC lattice. R and H have O_h symmetry, and P has T_d symmetry. The impact of k-point symmetry on the description of the electronic structure and ionic relaxation of the defect centre is further discussed in Appendix D.

Previous defect studies have suggested the use of Γ + R sampling in a SC supercell for silicon [202, 252], and Γ + H sampling in a BCC supercell for cubic metals [253]. For such sampling schemes is it necessary to impose a fixed occupation of the defect levels across the Brillouin zone, to avoid a non-uniform filling between k-points. We have calculated the transitions for the sequence Au^{1-}, Au^{0}, Au^{1+} with Γ + H sampling and P-point only sampling (listed in Table 6.3); in both cases, the negative-U effect is maintained, such that $E\,(1{+}/1{-})$ is the only thermodynamically stable transition. The positioning of this level, however, changes significantly with respect to the Γ-point only sampling case (by −0.09 eV and +0.18 eV, respectively). In order to check which of these sampling schemes gives the most accurate level positioning, we have performed the calculation using a large 864-atom BCC supercell, by embedding the relaxed ionic positions around the defect site ob-

[3]The high-symmetry points are given as fractional coordinates of the reciprocal lattice vectors for each Bravais lattice.

tained from the 256-atom calculation into the larger system. The transition level obtained is at E_v + 0.34 eV; amongst the different sampling schemes tested in the 256-atom system, Γ-point only has the smallest discrepancy with this value (+0.03 eV). Therefore, this gives an indication that the Γ-point only sampling scheme provides a faster finite size convergence with respect to the stable charge transition levels than either a dense k-point mesh or other high-symmetry sampling schemes, and, hence, the most accurate results from the ones listed in Table 6.3.

Finally, we have also tested for additional effects due to spin-orbit coupling. Calculations were performed with the ABINIT code [254] (version 6.6); we use the LDA exchange-correlation functional, Hartwigsen-Goedecker-Hutter (HGH) norm-conserving pseudopotentials [255], Γ-point sampling, and a plane-wave cut-off energy of 800 eV. The addition of spin-orbit coupling gives rise to a shift of approximately −0.1 eV in the formation energies that is almost constant across the various charge states. Therefore, when considering transition levels, this shift almost entirely cancels between different charge states, resulting in negligible changes to the transition levels ($E\,(1+/1-)$ and $E\,(1-/3-)$ being lowered by only 15 meV and 17 meV, respectively) and the relaxed ionic configurations, and no qualitative change in the negative-U nature of the defect centre.

6.4.4 Watkins' model

We now analyse the relaxation patterns obtained for the different charge states (Table 6.2) using Watkins' model, as illustrated in Fig. 6.6. The lattice will only remain in its undistorted T_d symmetry for Au^{3-}, since the defect levels are completely occupied in this charge state. However, D_{2d} symmetry is obtained via either the removal or addition of an electron from the neutral defect centre; this is achieved by forming pairs between the four neighbouring atoms of the defect centre, although in the former case the distance between pairs decreases by 0.1 mÅ, whilst in the latter it increases by 0.2 mÅ (hence, the level ordering of the states in the band gap is reversed in these two cases). We note that Au^{1+} exhibits the same behaviour as the isoelectronic v^0 centre, but Au^{1-}

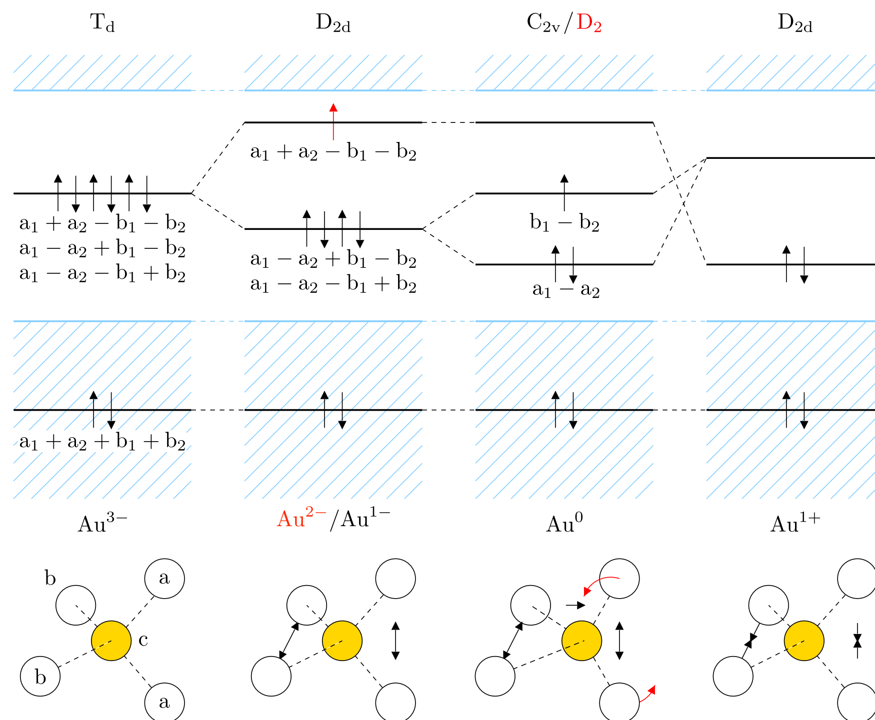


Figure 6.6: Schematic level diagram showing the effect of Jahn-Teller distortion on the defect levels for different charge states, following Watkins' LCAO model.

differs from v^{2-}, which instead features C_{2v} symmetry.

For Au^0, the degeneracy is lifted on all the defect levels; this is conventionally assumed to be caused by a second distortion which differentiates between the pairs, resulting in C_{2v} symmetry. Our calculations show some evidence of such behaviour, as the two stretched bonds between ion pairs differ in length for this charge state, and the Au ion consistently moves along [100] towards the centre of mass of the longer of the two (similarly to what was found in a previous study [179]). However, we also find a much more prominent distortion: a rotation of the ion pairs about the [100] axis which breaks the symmetry of the four a–b bond lengths. This distortion is sufficient in itself to reduce the symmetry of the defect centre to D_2 and lift the degeneracy of the three defect levels. The distortion of the defect centre from our DFT calculations is therefore fully consistent with a Watkins-like model, although two of the

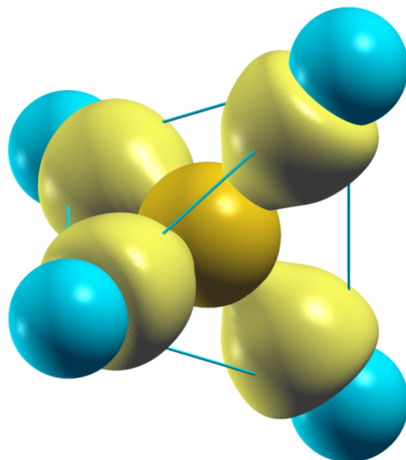


Figure 6.7: Contour-surface plot of the MLWFs most strongly associated with the Au^0 defect centre (unrelaxed); $A = 2/\sqrt{v}$, $f \approx 5/4$ (details are given in the caption of Fig. 5.9).

charge states (Au^0 and Au^{1-}) feature alternative symmetries to the ones that are generally assumed.

6.4.5 MLWF analysis

The calculation of the defect Wannier functions for the gold centre is more challenging than for the vacancy, as the distortion patterns obtained for the various charge states, although qualitatively similar, are much less pronounced, and the energy gap between the highest occupied Bloch state and the lowest unoccupied one is small. Consequently, the wannierisation of the occupied manifold does not result in chemically-intuitive orbitals; this is because the lattice distortion is not large enough to result in a clear hybridisation of the underlying orbitals from Watkins' model.

Therefore, we show only the MLWFs obtained from a wannierisation of the entire valence+deep-level manifold (Fig. 6.7); in this case, there are four Wannier functions of note associated with the defect, corresponding to the four orbitals used in the LCAO model (a_1, a_2, b_1, and b_2 in Fig. 6.6). For a substitutional defect centre, these orbitals are themselves hybrids of the sp^3-

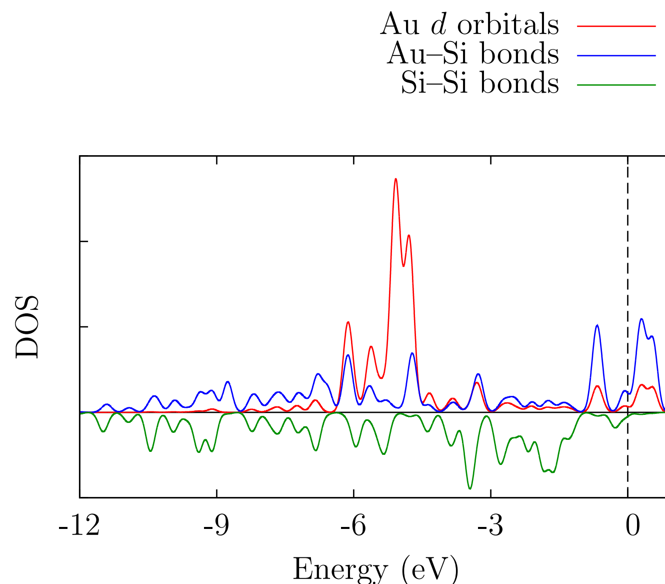


Figure 6.8: Projection of the three types of MLWFs (averaged over all Wannier functions of that type) on the valence+deep-level DOS. The dashed vertical line shows the position of the valence band edge. For clarity, the Si–Si bonds projection is plotted on the negative axis.

like dangling bonds surrounding the defect centre and the central impurity orbitals [228] (in this case, the sp^3 orbitals created from the 6s and 6p atomic subshells of gold); the result, therefore, is four σ-like bonding orbitals between the gold ion and its silicon neighbours. Unlike the typical Si–Si bonding orbitals, they do not posses a symmetry plane at the centre of the bond and parallel with the bond axis, although they are symmetrical with respect to rotation about that axis. There is no noticeable difference in this picture between the unrelaxed and relaxed geometries, as well as between different charge states of the defect.

In addition, the wannierisation procedure recovers the five orbitals of the closed 5d subshell of gold. By plotting the projected density of states onto the three different types of Wannier functions (Au–Si σ bonds, Si–Si σ bonds, and Au d orbitals), shown in Fig. 6.8, we see that the 5d orbitals' main contribution is a localised peak deep within the valence band of silicon, confirming that they are only weakly coupled to the rest of the system, and play little part in

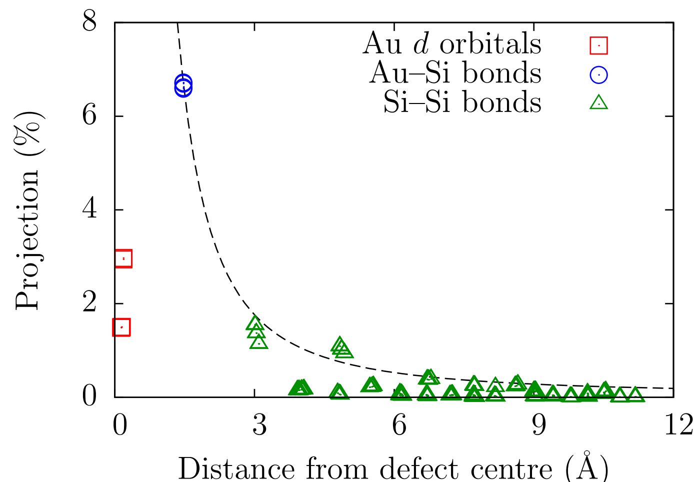


Figure 6.9: Contribution of all MLWFs in the supercell to the three defect levels in the band gap. The dashed line is a second-order polynomial decay fitted to the values of the Wannier functions with the largest contributions at each distance range.

the defect levels in the gap. The position of the main peak ($\sim\varepsilon_v - 5.1$ eV) is in reasonable agreement with that measured by photoemission for gold in silicon [256] ($\sim\varepsilon_v - 4.5$ eV), although our simulation does not include the large spin-orbit splitting that is also observed.

Fig. 6.9 shows the contribution from the projected DOS of every Wannier function in the 256-atom supercell to the three defect levels in the band gap. The strongest contribution comes from the four Au–Si bonds, and there is an algebraically-decaying contribution from Si–Si bonds in subsequent shells around the defect centre, despite the exponential localisation of the Wannier functions themselves. This decay is strongly directional, as indicated by the presence of Wannier functions with very small contributions to the defect levels even at short distances from the gold ion. This is in qualitative agreement with Probert and Payne's analysis of the defect charge density for the vacancy [203].

6.5 Conclusions

In this chapter we have studied the substitutional gold centre in supercells of bulk silicon, making use of the methodology developed for the silicon vacancy. We have found the defect centre to exhibit many similarities with the vacancy, both in terms of the finite size convergence properties and the analysis of the valence electron bonding at the defect centre in its various charge states; this is attributed to the fact that the gold impurity also forms a 'Watkins defect', as demonstrated by the relaxed ionic configuration of the different charge states in the 256-atom supercell, as well as the analysis of the defect supercell using MLWFs, showing the gold's $5d$ orbitals to interact only weakly with the defect states of interest. We note that our calculations have shown the relaxed neutral defect centre to have D_{2d}, rather than the conventionally assumed C_{2v}, symmetry; however, this is not incompatible with a description of the defect based on Jahn-Teller distortion within Watkins' LCAO vacancy model.

Using the results from the largest supercells employed we have argued that the defect forms a negative-U centre which depends on the Jahn-Teller distortion, resulting in a double level system with the stable charge state transitions (1+/1) and (1/3), and, hence, no stable spin-polarised charge state. The effect of spin-orbit coupling has also been tested, and found to have a negligible effect on our results. Our investigation, therefore, suggests that the defect is not paramagnetic, thus proving a simple explanation for the well-known missing EPR signal for the system. However, we are not aware of any direct experimental evidence to support this conclusion, and other explanations have also been proposed (as discussed in Sec. 6.4). Therefore, we believe that further investigation is needed, both experimentally and computationally; in the latter case, calculations using more accurate levels of theory should be undertaken to verify the robustness of the DFT ordering of the state charge transition levels.

Chapter 7

Doping at the Si–SiO_2 interface

In this chapter we present the results from our investigation of substitutional arsenic defects in silicon close to and at the interface between crystalline silicon and both crystalline and amorphous silica. In contrast to the silicon vacancy and the gold impurity discussed in previous chapters, arsenic is a shallow-level impurity in silicon, widely used for n-type doping; Watkins' LCAO model, that we have thus far employed extensively, is therefore not applicable in this case. Instead, the simple model that we must adopt is that of a charged silicon-like defect site with a weakly bound electron creating a shallow donor level below the conduction band edge, as described briefly in Sec. 4.3.

The Si–SiO_2 interface is a common feature in modern silicon-based complementary metal-oxide-semiconductor (CMOS) technology for the fabrication of integrated circuits; in devices such as MOS field-effect transistors (MOSFETs), the insulating oxide is placed next to the doped semiconductor, thereby potentially exposing the impurity atoms to the interface. Due to the ongoing miniaturisation drive for such devices, understanding the effects of the interface on the dopant distribution and properties is becoming an increasingly important concern. Indeed, in some modern integrated circuits, MOSFET channel lengths have been reduced to a few tens of nanometres, with the device properties being determined by only about 100 dopant atoms; it can be argued, therefore, that semiconductor technology is approaching the remit of nanothechnology.

The generation of candidate interface structures is in itself an active area of research in the field of materials modelling, since obtaining a realistic atomic configuration for the system is clearly important for analysing its properties. Therefore, in the first half of this chapter (Sec. 7.1), we describe our preliminary investigation of the undoped interface. We adopt a multiscale approach, making use of a Monte Carlo (MC) method in order to access the required time scales for amorphising the oxide and generating a disordered interface; we parametrise our MC model from DFT by using an energy-matching technique, resulting therefore in a computationally inexpensive *ab initio* method for simulating such structures. Our results show that this method is capable of correctly reproducing the qualitative features of the interface, athough the amorphous silica has a greater density than that measured experimentally; the possible causes for this are discussed.

In the second half of this chapter (Sec. 7.2), we investigate the segregation of arsenic dopants in silicon at the interface using DFT. We focus on the role of electrically active defects in bulk-like four-fold coordinated configurations, due to their importance for nanoscale devices. We employ two candidate structures obtained from our study of the interface: an ordered system with α-cristobalite, and a disordered one with amorphous silica.

Previous computational investigations of donor impurities at the interface have been restricted to small clusters or supercells, typically studying only a few selected defect sites in the first monolayer of silicon below the oxygen-bonded layer. The emphasis of these studies, therefore, has generally been on identifying novel defect configurations at the interface (featuring a qualitatively different bonding topology from the undoped network) that passivate the defect and favour segregation. In the present study, instead, we employ a large supercell and simulate both ordered and disordered interfaces. The size of the system allows us to accurately characterise the long-range quantum confinement effect due to the interface, showing a small energy barrier for segregation. We also investigate the impact of the local stress at the defect site on its segregation energy, and show that a simple 'particle in a box' model can be used to explain the calculated segregation energies at all substitutional silicon sites, independently of the oxidation state.

The rest of this chapter is organised as follows: in Sec. 7.1.1, we review previous experimental and theoretical studies of the Si–SiO_2 interface. In Sec. 7.1.2, we describe the simulation of the interface with two crystalline polymorphs of the oxide using hand-built models relaxed with DFT, and present our results for them. In Sec. 7.1.3, we describe our *ab initio* MC simulations of the interface with the amorphous oxide, and analyse its properties. In Sec. 7.2.1, we review previous studies relating to dopant segregation at the interface. In Sec. 7.2.2, we give the technical details of our simulations of the doped system, and in Sec. 7.2.3 we present and discuss the results of our calculations. Finally, in Sec. 7.2.4, we give a brief summary of our main conclusions.

7.1 Simulating the interface

7.1.1 Previous studies

Due to its crucial role in semiconductor technology, the Si–SiO_2 interface has been the subject of many experimental and theoretical studies. Several different experimental techniques have provided information on the atomic-scale structure of the interface (for a review, see Ref. [257]); however, due to the complexity of the system, a precise interpretation of the available data has proven challenging, and sometimes controversial. To this end, there is an ongoing effort to develop accurate models based on first-principles simulations for characterising the atomic arrangement at the interface.

Both scanning transmission electron microscopy (STEM) and x-ray scattering measurements have confirmed the oxide to be in its amorphous state [257]; however, the density of the oxide close to the interface is higher than that of silica glass by up to $\sim$10% [258, 259]. Indeed, investigations using high-resolution transmission electron microscopy [260, 261] (HRTEM) and x-ray diffraction [262] have detected a cristobalite-like oxide structure close to the interface; this polymorph of silica is denser than the amorphous structure by $\sim$5%.

Early studies of the width and roughness of the interface layer between the

crystalline silicon and the amorphous oxide, by a variety of techniques, did not yield a consistent picture, with different results ranging from an abrupt interface to an extended one of width ∼7 Å [263]. However, a number of more recent studies [263–266] using synchrotron radiation photoemission spectroscopy (PES) and x-ray photoelectron spectroscopy (XPS) have provided compelling evidence for the presence of all three intermediate oxidation states of silicon Si^{1+}, Si^{2+}, Si^{3+} at the interface, suggesting a gradual change from Si^{0} (bulk silicon) to Si^{4+} (bulk silica); these were determined by measuring the $2p$ core-level chemical shifts for the silicon due to its environment. The various oxidation states are therefore taken to correspond to four-fold coordinated silicon atoms with an increasing number of Si–O bonds replacing the Si–Si bonds of bulk silicon, ranging from four Si–Si bonds for Si^{0} to four Si–O bonds for Si^{4+}. The three intermediate oxidation states, belonging exclusively to the interface structure, are referred to as suboxide states[1]. The spatial distribution of the suboxide states suggests an extended interface of a few oxide monolayers, with the Si^{1+} and Si^{2+} states localised within a region of width 6–10 Å [264], and the Si^{3+} state extending further into the oxide [263–266].

Both electrical [272] and electron spin resonance [273] (ESR) measurements show a very low concentration of interfacial defects such as dangling bonds (less than one per 10^4 atoms in the interfacial layers), thus indicating an almost perfect bonding network with only a negligible number of over- or under-coordinated atoms.

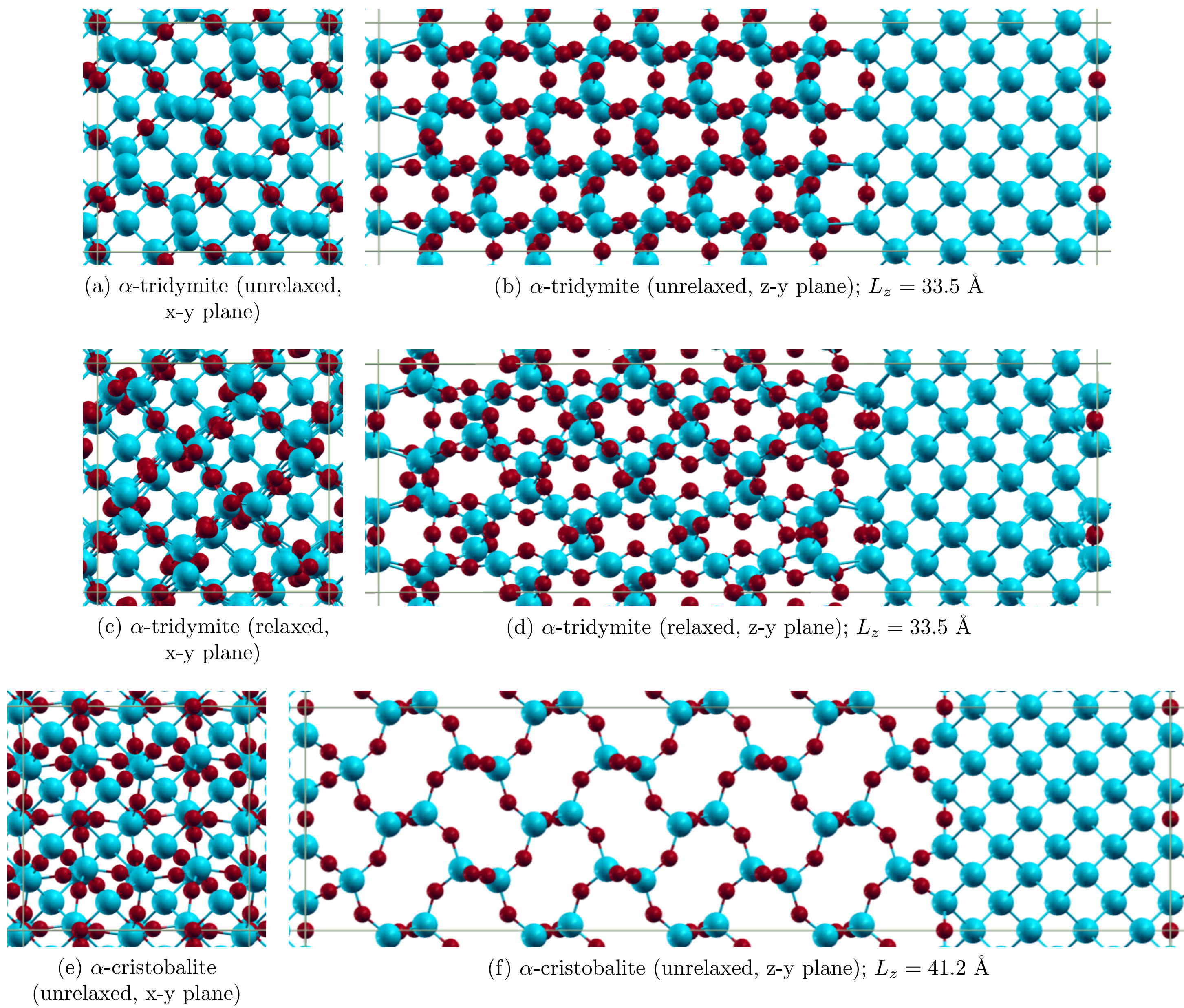
Atomistic and quantum-mechanical simulations have been employed both to find and test candidate structures for the interface. Several studies have employed hand-built models, simulated using either an empirical potential [274] or DFT [268–270, 275]. In particular, Pasquarello *et al.* have constructed model interfaces by attaching a crystalline silica polymorph (either tridymite [268–270] or β-cristobalite [270]) to the bulk terminated Si(001) surface and fixing the bond density mismatch either by dimerisation of silicon atoms at the

[1]This simple first-nearest neighbour picture was challenged by Banaszak Holl and McFeely [267], who suggested that second-nearest neighbour interactions might also play an important role; however, later theoretical studies [268–271] found this not to be the case, with the simpler picture providing the correct interpretation of the experimental measurements.

surface or by adding oxygen bridges between them; in this way, the various suboxide states are created. The candidate structures are then relaxed with DFT, and used to obtain information on the distribution of bond lengths and angles, the width of the interfacial region, and the $2p$ core-level shifts for the different oxidation states of silicon (by calculating the expectation value of the core orbital on the local self-consistent potential). In a later study by the same authors [276], the regular structure (with crystalline silica) is used as the initial configuration for an *an initio* molecular dynamics simulation at high temperature lasting 24 ps; the resulting structure is free from coordination defects, features all suboxide states in roughly equal proportions, and shows an excess of silicon at the interface.

A different approach is taken by Ng and Vanderbilt [277], and Tu and Tersoff [278, 279], who employ a MC method based on the continuous random network [280] (CRN) model introduced by Wooten, Winter and Weaire [281] (also known as the WWW method). In this approach, a defect-free bonding network is enforced, and the simulation proceeds by stochastic 'bond-switching' moves; the method is described in more detail in Sec. 7.1.3. This simple model allows for a long-time exploration of the relevant configuration space, from which optimal low-energy interface structures are found [278, 279], and the kinetics of the oxidation process is investigated [277]. However, we note that the highly ordered oxygen-bridge structure found by this method does not agree with the interpretation of the data from ion-scattering experiments [282].

Finally, a recent computational investigation of the interface [283] was conducted using classical molecular dynamics with newly developed Tersoff-like interatomic potentials [284]; a large system containing $\sim$2000 atoms was simulated in 128 runs of $\sim$2.8 ns each (using the replica-exchange method for sampling configurational space [285]). The simulation resulted in a suboxide layer of 6 Å and the Si^{3+} state showing the greatest extension into the oxide, in good agreement with experimental measurements. Furthermore, a cristobalite-like configuration of the oxide at the interface was observed, thus providing the first theoretical confirmation of this experimentally observed phenomenon.



(a) α-tridymite (unrelaxed, x-y plane)

(b) α-tridymite (unrelaxed, z-y plane); $L_z = 33.5$ Å

(c) α-tridymite (relaxed, x-y plane)

(d) α-tridymite (relaxed, z-y plane); $L_z = 33.5$ Å

(e) α-cristobalite (unrelaxed, x-y plane)

(f) α-cristobalite (unrelaxed, z-y plane); $L_z = 41.2$ Å

Figure 7.1: Ideal Si–SiO2 interface supercells. Silicon atoms are shown in blue, and oxygen atoms in red. The systems are periodic in all directions; the thin solid lines denote the cell boundary. L_z is the length of the supercell in the z-direction.

7.1.2 Ideal relaxed interfaces

For the creation of an ideal, highly ordered interface between silicon and silica, both in crystalline form, we follow the method of Pasquarello *et al.* [268–270] for constructing model interfaces with fully coordinated atoms. We then perform a structural relaxation of the system using DFT in order to eliminate spurious forces on the atoms and examine the resulting electronic structure.

We consider two different interfaces, both obtained by attaching a polymorph of silica to a Si(001) surface (made from 2×2 cubic cells of bulk silicon); these are: (i) α-tridymite, following Pasquarello *et al.* [269], and (ii) α-cristobalite, following Tu and Tersoff [279]. In each case the silica unit cell is stretched so as to be commensurate with the silicon lattice. α-tridymite is orthorhombic, with 8 silicon atoms in the conventional cell; by compressing the two longest sides of the cell by 8% and 13%, they match the silicon cubic cell diagonal length (7.6 Å). Instead, α-cristobalite is tetragonal, with 4 silicon atoms in the conventional cell; by expanding the two equal sides by 8%, they match the lattice parameter of the silicon cubic cell (5.4 Å). In both cases the silica's c-axis is perpendicular to the interface plane; the interface supercell is therefore tetragonal, and is made from 8 monolayers of silica and 9 layers of bulk silicon along the z-direction, parallel to the c-axis. We note that both polymorphs become significantly strained, with an increase in density from 2.20 g/cm^3 to 2.73 g/cm^3 in the former case, and a decrease from 2.32 g/cm^3 to 1.98 g/cm^3 in the latter. By comparison, the density of silica glass is 2.20 g/cm^3, and that of quartz is 2.65 g/cm^3.

Finally, we correct the bond density mismatch at the interface with the addition of oxygen bridges between atoms in the 'surface' layer of silicon, thus saturating the remaining dangling bonds after attaching the silica (these exist because the atoms have interlayer bonding in silica but not in silicon). The interfaces with α-tridymite and α-cristobalite require one and two oxygen bridges per silicon cubic cell, respectively. This results in all the suboxide states of silicon being present at the interface with α-tridymite (see Fig. 1a in Ref. [269] for a schematic illustration), but only the Si^{2+} state being present at that with α-cristobalite. The two interface systems are shown in Fig. 7.1;

the α-tridymite system has 272 atoms, and the α-cristobalite system has 280 atoms.

The ideal interface supercells are relaxed within DFT using the CASTEP code; we employ norm-conserving pseudopotentials (the silicon pseudopotential described previously, and an oxygen pseudopotential with 6 valence electrons, corresponding to the electronic configuration $2s^2 2p^4$ in the free atom), a cut-off energy of 800 eV, Γ-point sampling of the Brillouin zone, and the LDA exchange-correlation functional. The shape of the supercell is allowed to vary during the geometry optimisation; our convergence tolerance parameters for the relaxation procedure are 1×10^{-1} eV/Å for the RMS force, 1×10^{-3} Å for the RMS ionic displacement, 1×10^{-2} GPa for the maximum component of the supercell's stress tensor, and 1×10^{-5} eV for the energy per atom difference between iterations.

The relaxed structure for α-tridymite is shown in Figs. 7.1c and 7.1d; there is a negligible change in the shape of the supercell, but a large distortion of the silica configuration; the bonding network, however, remains unchanged. Of particular interest is the difference between the two interfaces present in the supercell (these can be seen in the figure at the edge and in the middle of the cell). Despite a qualitative agreement with respect to the previous description of the construction of the interface, it is not possible to achieve an equivalent bonding topology at both interfaces, as the orthorhombic cell of silica does not posses the necessary space group symmetry. This inequality can be seen in Fig. 7.1d, as the first two layers of silicon at each interface exhibit noticeably different relaxation patterns; in particular, there is a greater disruption to the perfect bulk silicon arrangement for the interface at the edge of the cell. The difference between the two interfaces leads to a spurious long-range interaction over the entire system: analysing the charge density obtained from the DFT simulation, we observe a transfer of charge of $\sim$0.029 e/Å^2 between the interfaces (i.e., almost one electron per silicon cubic cell). The resulting dipole creates an electric field across the system, which can be expected to have a large effect when calculating the properties a point defect near one of the two interfaces.

For α-cristobalite, instead, the relaxation produces no noticeable change in

the atomic configuration, and only a small decrease of 0.6 Å of the supercell's length in the z-direction, resulting in a negligible increase in density for the crystalline silica. In this case, however, the two interfaces in the supercell are equivalent by symmetry; hence, no charge transfer can occur, and the interaction between interfaces can be expected to be small. Therefore, we make use of this system as our 'ideal' interface in the main results presented in Sec. 7.2; furthermore, we use it as the initial configuration in our Monte Carlo simulations for creating a crystalline silicon–amorphous silica interface with the continuous random network model (discussed in the next section).

7.1.3 Disordered interfaces with the CRN model

As previously discussed, a more realistic model of the Si–SiO_2 interface in semiconductor devices must take into account the amorphous nature of the oxide. To obtain a representative 'snapshot' of such an interface, it is necessary to sample the configuration space of the systems, as is done in, e.g., *ab initio* and classical molecular dynamics simulations. A more coarse-level approach, that allows for an efficient and quick configuration sampling, is provided by Monte Carlo methods [286]: instead of simulating the dynamics of the system, these generate a sequence of states using a Makov chain procedure and a predefined set of possible trial moves, that are either accepted or rejected stochastically according to a Boltzmann probability distribution. The resulting trajectories can then be used to measure the thermodynamic properties of the system.

The success of MC simulations in describing a particular system is usually dependent on the appropriate selection of the state variables of interest (to limit the region of configuration space that can be sampled), and definition of the trial move for exploring the possible configurations. To do so, we employ the canonical continuous random network model for network glasses, first used in conjunction with MC techniques to simulate silicon [281, 287] and germanium [281], and later for the silicon–silica system [277–279]. Additionally, we parametrise the model by fitting to data obtained from DFT total energy calculations of 100 snapshots of the interface; thus, we present the first entirely

ab initio application of the CRN-MC method to this system.

We now briefly summarise the theory and algorithms that we have implemented in a new CRN-MC code for the purpose of obtaining snapshots of the interface with the amorphised oxide[2]. The main idea is to characterise the system solely in terms of its bonding topology $\mathcal{B}$ (i.e., a neighbour list specifying all pairs of atoms connected by bonds). The total energy of the system E^{tot}, however, is given by a Keating-like potential [288], based on bond lengths and angles (and defined, therefore, also by the set of atomic positions $\{\mathbf{r}_i\}$):

$$E^{\mathrm{tot}}\left(\mathcal{B},\{\mathbf{r}_i\}\right)=\sum_i^{N_b} k_b^{(i)}\left(b_i-b_0^{(i)}\right)^2+\sum_i^{N_\theta} k_\theta^{(i)}\left(\cos\theta_i-\cos\theta_0^{(i)}\right)^2, \qquad (7.1)$$

where N_b (N_θ) is the total number of bonds between two atoms (angles between three atoms connected by two bonds), b_i (θ_i) is the value of the i-th bond length (angle), $b_0^{(i)}$ ($\theta_0^{(i)}$) is the equilibrium value of the bond length (angle) of type i, and $k_b^{(i)}$ ($k_\theta^{(i)}$) is the weight associated with the bond (angle) of type i. The determination of these parameters is discussed below. As the system contains silicon and oxygen atoms, there are two types of bond (Si–Si, and Si–O), and four types of angles (Si–Si–Si, Si–Si–O, Si–O–Si, and O–Si–O); O–O bonds are forbidden, so as to enforce the same bond coordination as the crystal structure.

In order to achieve a unique representation of the system as a function of the bonding topology (which is a discrete quantity), we define a constrained energy E, obtained by minimising the total energy in Eq. 7.1 with respect to $\{\mathbf{r}_i\}$)[3]:

$$E\left(\mathcal{B}\right)=\min_{\{\mathbf{r}_i\}} E^{\mathrm{tot}}\left(\{\mathcal{B}\},\mathbf{r}_i\right). \qquad (7.2)$$

The MC simulation generates trial moves by introducing a change in $\mathcal{B}$, and uses the resulting change in magnitude of the constrained energy before and

[2] The code is written in Fortran90 and fully parallelised using MPI.

[3] It is important for the bonding topology to be consistent with the atomic positions; therefore, in order to avoid the occurrence of 'false' neighbours, we introduce an additional cubic repulsion term between atoms that are not bonded, but whose distance from each other is smaller than the equilibrium bond length (the repulsion goes to zero outside of this radius, and therefore does not contribute to the energy for any reasonable atomic arrangement) [287]. A Verlet neighbour list is used to compute this term.

after the move ΔE in the stochastic calculation of the move's outcome (using the conventional Metropolis algorithm [289] with an acceptance probability given by the Boltzmann factor $e^{-\Delta E/k_B T}$). Therefore, the minimisation of the total energy with respect to the atomic positions must be performed after each trial move in order to determine ΔE, and, hence, the move's outcome; this minimisation is done with a conjugate gradients scheme.

The trial move that evolves the system is the single bond-switch defined by Wooten, Winter and Weaire [281] and adapted for the case of silica by Tu *et al.* [287]: in its simplest form, this is done by selecting four atoms connected by three bonds a–b–c–d, and switching the pairing between the first and last bond to create the new topology a–c–b–d (the middle bond is unaffected). It is important to note that this bond-switch is designed so as to leave all coordination numbers unchanged, ensuring that no atom in the system can become over- or under-coordinated during the course of the simulation. For the case of the silicon–silica system, the trial move operates in a similar manner, although it is complicated by additional constraints: silicon and oxygen atoms must always have four and two bonds, respectively, and oxygen, as previously stated, must only bond to silicon. Silicon atoms, instead, are free to have between zero and four bonds to oxygen, thereby allowing the formation of the various suboxide states at the interface.

We have tested our code by using it to calculate the glass transition temperature T_m between crystalline and amorphous silicon for the CRN model; this is in agreement with that calculated by Tu *et al.* [287] (0.50 eV $< k_B T_m <$ 0.55 eV). In this case, we make use of the parameters for the model from the cited study; these are chosen empirically so as to reproduce known properties of amorphous silicon. However, for the final parameters used in our simulations of the silicon–silica interface, our aim is determine their values from first-principles simulation.

The simplest parameters to fit from first-principles are those relating exclusively to silicon: the equilibrium Si–Si bond length $b_0^{(\text{Si–Si})}$ and Si–Si–Si angle $\theta_0^{(\text{Si–Si–Si})}$, and their respective weights $k_b^{(\text{Si–Si})}$ and $k_\theta^{(\text{Si–Si–Si})}$; the first two are given by the equilibrium crystalline silicon structure calculated with DFT, and the last two can be derived, assuming a simple analytic form for the equa-

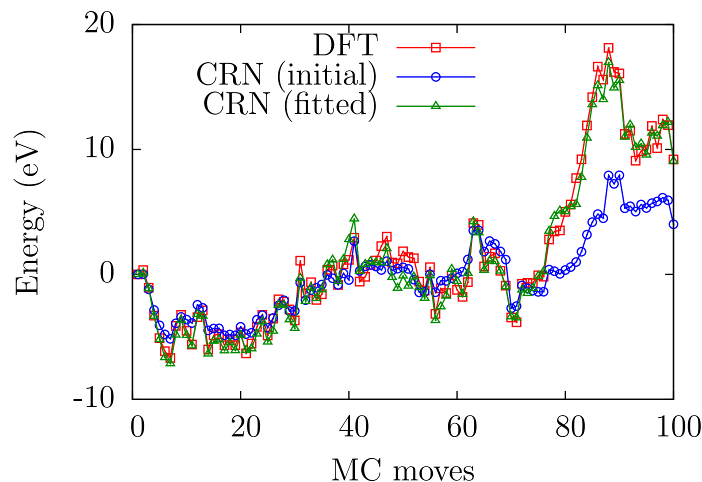


Figure 7.2: Fitting of CRN parameters to DFT results.

tion of state, from the material's bulk modulus (giving information on bond stretching) and Young's modulus (giving information on bond stretching and bending), both of which are straightforward quantities to calculate from simulation. Table 7.1 lists the values calculated by this method.

However, it is not clear how to fit the parameters relating to silica ($b_0^{(\text{Si–O})}$, $\theta_0^{(\text{Si–O–Si})}$, and $\theta_0^{(\text{O–Si–O})}$, and their respective weights) using the same approach. There are two reasons for this difficulty: (i) different polymorphs of silica give incompatible data for this simple parametrisation (as shown in Table 7.1), and (ii) elastic constants cannot be equated in a simple manner with bond stretching and bending, as the fractional positions of the ions in the unit cell change noticeably as the cell is distorted[4]. Finally, the parameters relating to the Si–Si–O angle cannot be derived from silica alone, as they refer specifically to the suboxide states at the silicon–silica interface. In previous work, these have therefore previously been assigned by averaging between the values relating to the Si–Si–Si and O–Si–O angles [277, 278].

In order to perform the fitting, therefore, we use a more powerful and

[4]Taking this into account is important for calculating, e.g., the bulk moduli of different polymorphs of silica, as these result much too high if the ions are not allowed to relax in the distorted unit cell.

general method, that of energy-matching the force field to DFT calculations for a number of snapshots of the entire interface system [290]. The snapshots are obtained from a MC run at high temperature (∼5800 K) using the values of the parameters given in previous studies. We use $N_s = 100$ snapshots of the 280-atom system, and fit the parameters by minimising the normalised RMS error in the energy difference between all pairs of snapshots calculated with Eq. 7.1 with respect to the DFT total energies:

$$S^{\text{fit}} = \frac{\sqrt{\sum_{i,j}^{N_s} \left(\left(E_i^{\text{CRN}} - E_j^{\text{CRN}} \right) - \left(E_i^{\text{DFT}} - E_j^{\text{DFT}} \right) \right)^2}}{\sqrt{\sum_{i,j}^{N_s} \left(E_i^{\text{DFT}} - E_j^{\text{DFT}} \right)^2}}. \tag{7.3}$$

Fig. 7.2 shows the results of the fitting procedure. As can be seen, the values of the total energy from the initial MC run do not agree well with the DFT results for each snapshot; this is especially clear for the last 24 snapshots, which deviate substantially from the DFT trajectory. Instead, the energies from the CRN model agree very closely with DFT after the energy-matching method is used; 74% of the snapshots are in better agreement with DFT after the fitting procedure, and the overall error S^{fit} decreases from 0.53 to 0.13. This also confirms that the model provides an accurate representation of the system for the region of configuration space that is explored by the bond-switching method. We note that the fitting procedure is robust, i.e., we recover similar values for the parameters by only fitting to half the data set (the snapshots labelled 1 to 50 in Fig. 7.2); in this case, we still find a decrease in S^{fit} for the data points that have not been used in the fitting, indicating that the large correlation between snapshots (as each one differs from the previous one by a single MC move) does not have a large impact on the fitted parameters.

The parameters calculated by the fitting procedure are listed in Table 7.1. The values are generally qualitatively similar to those used in previous studies (i.e., the starting values for the minimisation of S^{fit}); the only value which is significantly changed by the fitting is the bending weight for the Si–O–Si angle, which increases by 2.59 eV. This can be expected to have an impact on the properties of silica, as the increased stiffness of the angle will strengthen the short-range crystal-like atomic ordering.

	b_0 (Å)		θ_0 (°)			
	Si–Si	Si–O	Si–Si–Si	Si–Si–O	Si–O–Si	O–Si–O
Ref. [278]	2.35	1.60	109.5	109.5	180.0	109.5
Fitted	2.33	1.58	109.5	127.1	146.4	117.0
Expt.	2.35	1.60/1.61	109.5	-	146.5/143.7	109.5/109.5
DFT	2.33	1.60/1.60	109.5	-	146.9/143.5	109.5/109.5

	k_b (eV/Å^2)		k_θ (eV)			
	Si–Si	Si–O	Si–Si–Si	Si–Si–O	Si–O–Si	O–Si–O
Ref. [278]	4.54	13.50	1.79	1.97	0.38	2.16
Fitted	4.56	14.12	0.83	2.01	2.97	2.92
Expt.	5.08[a]	-	0.79[b]	-	0.51/1.51[c]	-
DFT	4.92[a]	-	1.89[b]	-	0.48/1.26[c]	-

Table 7.1: List of parameters for the CRN model applied to the silicon–silica system, comparing the values from a previous theoretical study and those calculated by our fitting procedure. Also listed are the partial sets of values derived either from experimental measurements or DFT simulations of equilibrium structures, bulk moduli and Young's moduli; for silica-related parameters, these are listed for α-cristobalite/α-quartz (experimental measurements are taken from Refs. [292] and [293], respectively).

[a]Calculated as $2\sqrt{3}B_0^{(\mathrm{Si})}b_0^{(\mathrm{Si-Si})}$, where $B_0^{(\mathrm{Si})}$ is the bulk modulus of silicon.

[b]Calculated as $\frac{27}{512}\left(Y_0^{(\mathrm{Si})}a_0^3/2 - k_b^{(\mathrm{Si-Si})}a_0^2/3\right)$, where $Y_0^{(\mathrm{Si})}$ is the [100] Young's modulus of silicon [291], and a_0 is the equilibrium cubic lattice parameter.

[c]Calculated as $\frac{9}{8}\left(B_0^{(\mathrm{SiO_2})}V_0/N_\theta\right)/\left(1-\cos\theta_0^{(\mathrm{Si-O-Si})}\right)^2$, where $B_0^{(\mathrm{SiO_2})}$ is the bulk modulus of silica, V_0 is the equilibrium cell volume, and N_θ is the number of Si–O–Si angles in the cell (this expression is derived by assuming fixed fractional positions for Si ions and fixed Si–O bond lengths upon cell distortion).

For generating the final interface structure between crystalline silicon and its amorphous oxide, we make use of the interface with α-cristobalite described in the previous section as our initial configuration, for two different system sizes: a 304-atom (small) supercell, with a cross-sectional area of 115.9 Å^2 (i.e., a 2×2 Si(001) surface, as shown in Fig. 7.1), and a 1600-atom (large) supercell, with a cross-sectional area of 463.6 Å^2 (i.e., a 4×4 Si(001) surface). The length of the system is allowed to change in the z-direction, while the cross-sectional area is fixed. The first two layers of bulk silicon are free to take part in bond-switching moves during the MC simulation, but the layers below are fixed in their crystalline bonding configuration.

We first perform an equilibration of the system at high temperature ($k_{\mathrm{B}}T =$ 0.5 eV, causing a rapid amorphisation of the oxide), and then perform a slow annealing to a lower temperature ($k_{\mathrm{B}}T = 0.1$ eV); this is observed to result in the system finding a low-energy local minimum, with a disordered interface and the oxide in its amorphous state. For the small system, we equilibrate for 4000 accepted moves, and anneal for 1000 accepted moves; for the large system, instead, we use 500 and 1500 accepted moves for these two stages.

Although the initial configuration only includes a single suboxide state (Si^{2+}), the disorder introduced at the interface by the bond-switching mechanism results in all the suboxide states being present at the end of the simulation. For the large system, the Si^{2+} state maintains the highest concentration, with 56 occurrences (compared with only 5 each for Si^{1+} and Si^{3+}); this is due to the persistence of most of the oxygen bridges present in the initial ideal interface. Previous MC studies have argued that these bridge structures play a key role in the interface [278, 279]; however, our results for the small system (which was equilibrated for a much larger number of moves) show no remaining bridge structures at the end of the simulation, and a similar concentration of all suboxide states (13, 8, and 9 occurrences for Si^{1+}, Si^{2+}, and Si^{3+}, respectively). This suggests that the results for the large system are not sufficiently equilibrated, and that the observed high concentration of oxygen bridges is spurious.

Fig. 7.3 shows the distribution along the interface and volume of the different oxidation states of silicon. The volume is calculated from the tetrahedron

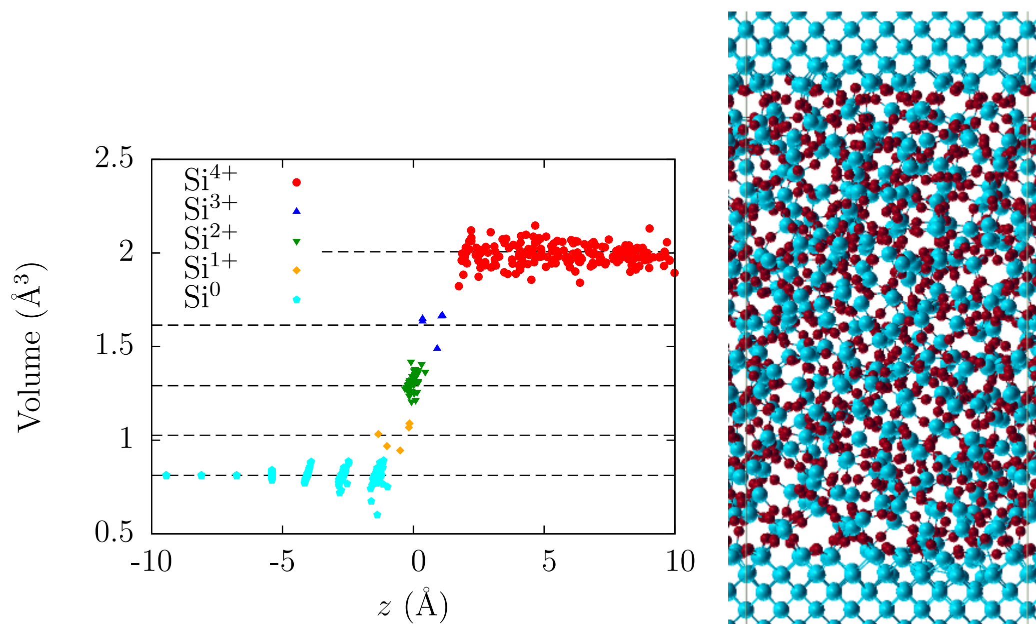


Figure 7.3: Tetrahedral volume of the silicon ions in different oxidation states plotted along the z-direction (perpendicular to the interface plane). The dashed lines indicate the ideal volume for each oxidation state. The large supercell is shown on the right in the z-y plane; the width of the oxide region in the z-direction is of ∼36 Å.

formed by the four neighbours of the ion (we use either the middle of the bond for a Si–Si bond, or the oxygen position for a Si–O bond). We also show the ideal volume for each state, calculated from the DFT equilibrium bond lengths, giving an indication of the local strain for each ion; we note a significant strain for the first few layers of silicon in the bulk oxidation state Si^0. Taking these layers into account, the width of the interface can be seen to be between 5 Å and 10 Å, in agreement with experimental observations.

The large supercell allows us to investigate the properties of bulk-like silica at the centre of the oxide layer. The average density for the amorphous silica is ∼2.5 g/cm^3, with no noticeable increase at the interface; this is a 25% increase from the density of the initial structure (strained α-cristobalite), and is 14% higher than the experimental value for silica glass. We suggest two possible causes for this discrepancy. Firstly, the Si–Si distance is underestimated in

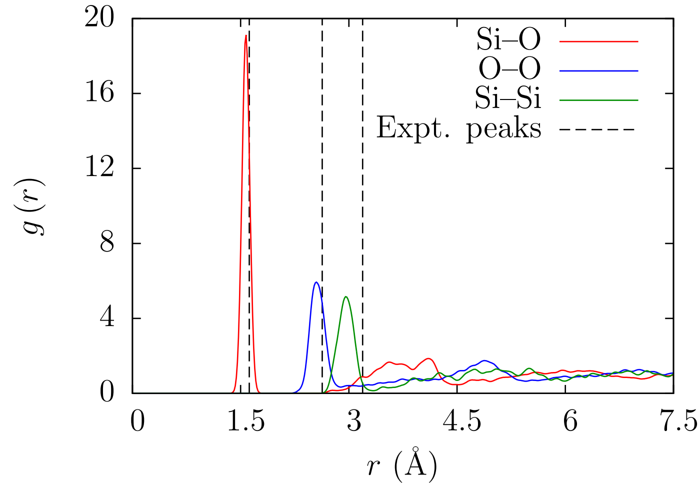


Figure 7.4: Normalised partial pair correlation functions for silica in the Si–SiO_2 system from the CRN-MC simulation (averaged over all the ions in a region of width 4 Å at the centre of the oxide layer). The dashed vertical lines show the positions of the first three peaks measured experimentally [294], corresponding to the smallest Si–O, O–O, and Si–Si distances, respectively.

the CRN model; this can be seen from the pair correlation function, shown in Fig. 7.4. Although the positions of the first and second peaks are in good agreement with experiment, the third peak (corresponding to the distance between silicon ions connected by an oxygen) is at a smaller distance for the model; the position of the peak is close to the equilibrium Si–Si distance calculated from the parameters $b_0^{(\text{Si–O})}$ and $\theta_0^{(\text{Si–O–Si})}$, suggesting that the strength of the short-range ordering is overestimated by the model. Secondly, it has been suggested that the low density of silica might be due to the presence of voids in the structure (with no dangling bonds), with sizes on the order of 10 Å [295]. These formations are not present in the simulations described here; interestingly, however, we have found that such voids *do* spontaneously appear during the course of the MC run if the supercell size is fixed so as to impose the experimental density for silica. The fact that the voids are not formed in the unconstrained cell might be a consequence of the finite size of the oxide layer between the interfaces; this suggests the possibility that the increased density of silica close to a Si–SiO_2 interface is due to an absence of voids in

this region.

In conclusion, we have studied the formation of the interface between crystalline silicon and amorphous silica, starting from a hand-built model of an ideal ordered interface with crystalline silica, and performing a MC simulation on the continuous random network model of the system to amorphise the oxide and obtain a disordered interface. The simulation is an example of *ab initio* multiscale modelling, as the CRN model is parametrised from first-principles DFT calculations. Our results show the formation of all the suboxide states at the interface, although their relative proportions are difficult to determine quantitatively, as these are found to be sensitive to the simulation history even for large numbers of MC moves; qualitatively, however, the interface is in good agreement with the features suggested by previous experimental and theoretical studies. The final structure obtained from the MC run for the small supercell has been additionally relaxed with DFT to a tight convergence tolerance of 5×10^{-3} eV/Å for the RMS force on all the ions; this resulted only in small adjustments of the ionic positions, and no change in the bonding network. Furthermore, the analysis of the charge distribution in the relaxed DFT structure shows a negligible dipole between the two non-equivalent interfaces in the system, in contrast to what is observed for the ideal α-tridymite interface structure (see discussion in previous section). Therefore, we make use of this amorphised structure (after relaxation with DFT) as our ‘disordered’ interface in the study of the arsenic substitutional defect in silicon close to the interface.

7.2 Arsenic segregation at the interface

7.2.1 Previous studies

The behaviour of dopant species at semiconductor interfaces has been characterised by a number of different studies in the last few decades; as previously noted, as the junction depth of modern transistors decreases, the influence of the interface becomes an increasingly important consideration with regards to the performance of the device.

Several studies have reported the uphill thermal diffusion and segregation of dopant atoms to the Si–SiO_2 interface region during the high-temperature anneal following implantation, both for silicon donors (phosphorus [296, 297], arsenic [298–307] and antimony [298, 308]) and acceptors (boron [301, 302, 309, 310]). Up to a monolayer of dopant atoms can be collected in this region [296, 298] (as is assumed by simple segregation models for simulating CMOS technology processes [311]). The dopants were initially believed to be located on the oxide side of the interface, as their pileup was observed to disappear upon etching the oxide [312, 313]; however, later studies using XPS [299], medium ion scattering [301] (MEIS) and STEM [303] measurements have shown them instead to be on the silicon side. Most importantly, the majority of the atoms trapped at the interface are deactivated [298], i.e., they no longer behave as shallow dopants; this is confirmed by Auger electron spectroscopy (AES) measurements, showing that the dopants stay neutral, and so do not release any electrons into the channel [314]. This segregation and deactivation phenomenon, therefore, leads to significant dose loss in the device, that can affect the transistor's threshold voltage by up to 50% [315].

Although the focus of previous theoretical studies (described below) has so far been on the processes leading to donor deactivation, it is also important to note that the amount of electrically active segregated dopants has been estimated to be 10–20% of the total by recent secondary ion mass spectrometry (SIMS) measurements [316]. These active defect centres have not yet been characterised in models of the doped interface system; however, an understanding of their properties is crucial for the development of novel electronic nanoscale devices, as the reduced size will make the interaction between the active dopants and the interface one of the most important effects for determining the properties of the system. Indeed, a very recent study has observed single arsenic dopant signatures at low temperatures in a nanoscale CMOS channel [317]; the ionisation energy of the arsenic was found to be much larger than in bulk, which was postulated to be due to dielectric and quantum confinement effects caused by the proximity of the donor to the interface (as supported by calculations on a very similar system, that of donor impurities in semiconducting nanowires embedded in a material with a low

dielectric constant [318]).

Several *ab initio* computational studies (all using DFT) have been undertaken with the aim of investigating the issue of phosphorus and arsenic segregation to the Si–SiO$_2$ interface [319–324]. All the cited studies made use of highly regular, hand-built models of the interface, generally similar to those described in Sec. 7.1.2; only the most recent one by Kong *et al.* [324] additionally employed a disordered interface between crystalline silicon and amorphous silica in a 160-atom supercell.

Dabrowski *et al.* [319, 320] (and, later, Kim and Chang [322]) showed the bonding of donor ions to oxygen to be energetically unfavourable, and proposed two different segregation and deactivation mechanisms based on trapping at fully saturated three-fold coordinated sites at the interface (either as a single donor substituting an under-coordinated silicon with a dangling bond, or as a donor pair at defect-free interfaces). Following these results, Ravichandran and Windl [321] showed that the presence of hydrogen at the interface can strongly enhance the energy gain from segregation, and argued that this might be a more realistic mechanism for explaining the high pileup rates observed experimentally. Zhou *et al.* [323] investigated the donor at four-fold coordinated substitutional sites at the interface, reporting a severe distortion upon relaxation for the neutral defect centre, resulting in three short bonds and one long one (thus creating an approximately three-fold coordinated site); this distortion disappears for the positively charged defect, with a large calculated energy barrier for the release of the electron. Finally, Kong *et al.* [324] recently identified three novel defect complexes at the interface, which were found to be energetically more favourable than complexes in bulk as a result of the reduced strain for the structural distortion due to the flexible Si–O–Si angles; these complexes were also found to be electrically inactive.

7.2.2 Computational methodology

As noted in previous sections, we make use of two different supercells for our defect calculations: (i) an ordered ideal interface between crystalline silicon and α-cristobalite (featuring oxygen bridges at the silicon surface to compen-

sate for the bond density mismatch between the two materials, as described in Sec. 7.1.2), and (ii) a disordered interface between crystalline silicon and amorphous silica (generated with the CRN-method, as described in Sec. 7.1.3, and featuring all the suboxide states of silicon). Both supercells contain 472 atoms in total, with 33 monolayers of silicon along the z-direction[5]. The width of the silicon layer is therefore ∼42.5 Å, and that of the silica layer is ∼29.6 Å in the ideal system and ∼23.4 Å in the disordered one. As explained previously, the supercells contain two interfaces due to the periodic boundary conditions; in the ideal system these are identical, but in the disordered one they are not, and, hence, a large interfacial total area (231.8 Å^2) is available to us for our investigation.

Our DFT simulations are performed with the CASTEP code, using the numerical parameters detailed in Sec. 7.1.2. Before the introduction of the dopant atom into the system, the two undoped supercells are relaxed to their equilibrium geometry (to a tolerance of 1×10^{-2} eV/Å in the RMS force). We then perform simulations of the system containing a single arsenic substitutional defect at a silicon site; as shown in the next section, we investigate many different sites in the supercell (including all the suboxide states of silicon). The segregation energy E_s for a particular site is defined simply as

$$E_s = E^{\text{def}} - E^{\text{def,ref}}, \tag{7.4}$$

where E^{def} is the total energy of the supercell with the dopant placed at the site of interest (typically close to or at the interface), and $E^{\text{def,ref}}$ is the same quantity for the dopant placed at a reference site, in this case taken to be at the centre of the silicon layer (as it has the most bulk-like character). Only neutral defects are considered; unless otherwise stated, the doped supercells are unrelaxed. We note that the defect formation energy is not well-defined by the Zhang-Northrup approach used in bulk, due to the ambiguity in defining the host chemical potential at inequivalent lattice sites [325]; however, this is not a problem for defining the segregation energy, as the number of atoms for

[5]For the disordered system, we use the small supercell from the MC simulations and insert additional layers of perfect bulk silicon.

each species is constant between the two configurations.

We employ a norm-conserving pseudopotential for the arsenic species with 15 valence electrons, corresponding to the electronic conguration $4s^2 3d^{10} 4p^3$ in the free atom. Preliminary calculations of the substitutional arsenic defect in a 256-atom BCC supercell of bulk Si are in good agreement with experimental measurements of the defect properties: the ionisation energy (calculated as the position of the stable charge transition level $E\,(1+/0)$ with respect to the conduction band edge[6]) is 47 meV, compared with an experimental value of 49 meV [326], and the relaxed As–Si bond length is 2.43 Å, compared with experimental values of 2.41 ± 0.02 Å [327] and 2.43 Å [328].

7.2.3 Results

The ideal interface

We first investigate the energetics of arsenic segregation at the ideal interface, shown in Fig. 7.5. Due to the ordered atomic arrangement, there are only a small number of inequivalent defect sites, generally one per monolayer; the only important exception is the silicon monolayer immediately beneath the first oxygen-bonded layer, which features two inequivalent sites due to the positioning of the oxygen ions in the first layer of silica (excluding the oxygen bridges, that do not break the symmetry)[7]. Consequently, the ionic relaxation results in a difference in the position of these two sites of 0.30 Å in the z-direction. The local strain (deduced from the volume of the tetrahedron defined by the four nearest neighbours of the lattice site) is also different, both in sign and magnitude: for the site closer to the interface it is slightly positive, but with a volume almost identical to the equilibrium one found in bulk (6.52 Å^3, compared with an equilibrium volume of 6.50 Å^3), while for that further from the interface it is negative, with a significantly smaller volume (6.15 Å^3).

[6]By using ε_c as the reference for calculating the transition level, we implicity employ a scissor operator correction (see Sec. 4.4.5), making the assumption that the shallow donor level from arsenic shifts together with the CBM.

[7]This symmetry breaking is present also in other silicon monolayers further from the interface; however, as the ionic and electronic structure depends primarily on the local chemical environment, the differences between inequivalent sites are negligible for all layers expect the one closest to the interface.

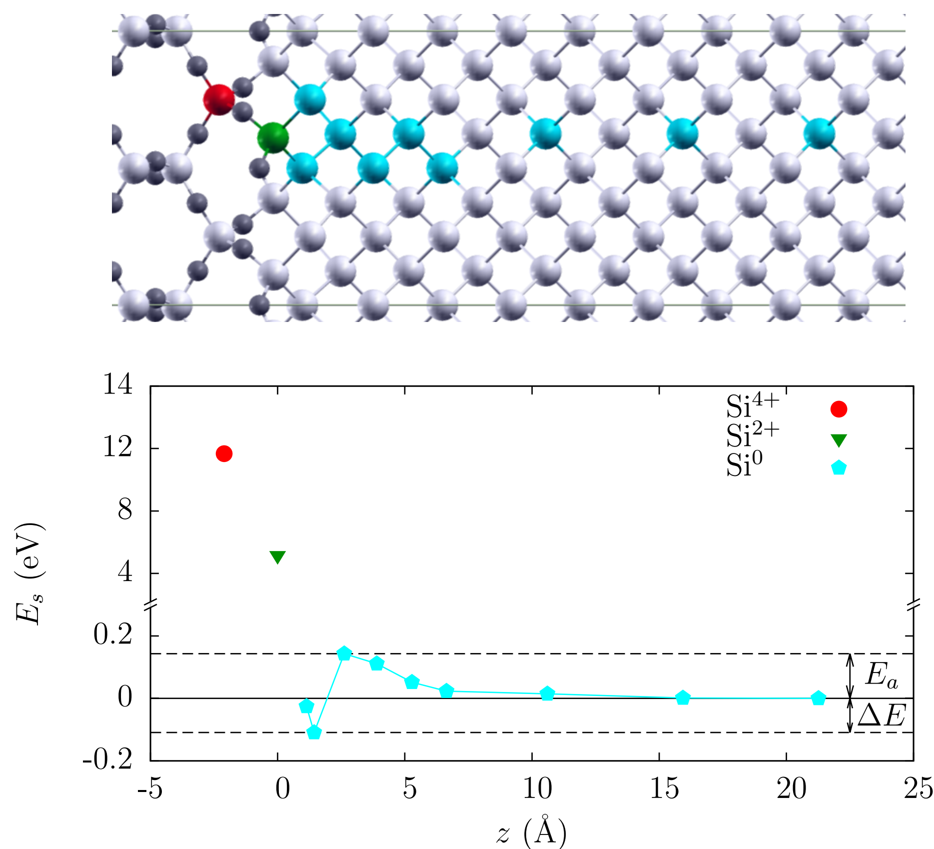


Figure 7.5: Segregation energy of the arsenic dopant at the ideal interface with respect to its bulk position. The different symbols indicate the various oxidation states of the silicon sites that are occupied by the substitutional arsenic. E_a is the activation energy and ΔE the maximum segregation energy. The supercell is shown above; silicon atoms are shown in white, and oxygen atoms in black. The substitutional sites for the arsenic atom are shown in colour (with the specific colour denoting its oxidation state, as labelled in the key of the graph).

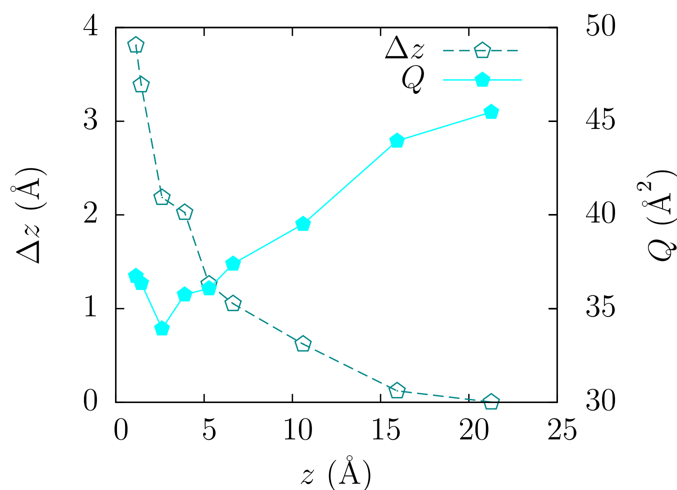
Fig. 7.5 shows the arsenic segregation energy as a function of the distance to the interface (the zero is taken as the position of the Si^{2+} suboxide layer); we include the energy of the dopant at the Si^{2+} site and the first Si^{4+} site in silica. As can be seen from these calculations, the direct bonding of the dopant atom to oxygen (As–O) is highly unfavourable, with an energy penalty of 5.17 eV for the doubly oxygen-coordinated defect site and 11.67 eV for the

fully oxygen-coordinated one. This is in agreement with previous theoretical studies [319–324], as well as the experimental evidence that the segregated atoms are found on the silicon side of the interface [299, 301, 303]. However, the cause of this unfavourable bonding has not previously been investigated, and has been assumed to be a chemical effect [319]; from the results for the disordered interface (discussed in the next part of this section), we shall argue that it is due to a confinement effect caused by the proximity of the oxygen charge density to the defect site.

Restricting ourselves to the fully silicon-coordinated defect sites, we see two separate regimes: firstly, a long-ranged steady increase in the segregation energy as the As ion approaches the interface from bulk Si; this reaches its maximum value in the second Si^0 monolayer from the interface, with a small energy penalty of 0.14 eV. Secondly, a drop in the segregation energy at the first Si^0 monolayer (closest to the interface) for both the inequivalent sites previously described; these dopant sites result energetically favourable respect to the bulk position, and, hence, can drive segregation to the interface. Assuming a simple migration of the As ion between lattice sites, we can consider the energy barrier at the second monolayer as a migration or activation energy for the segregation process, as shown in the figure[8]. We note, however, that the energy gain from segregation is modest: 0.11 eV for the site with a large negative strain, and only 0.03 eV for that with a small positive strain.

All the defect sites belonging to the first regime have a negligible strain, and are almost perfectly bulk-like in their local bonding environment both in terms of bond lengths and angles. We therefore assume that the increase in segregation energy is a quantum confinement effect due to the silica, affecting the long-range decay of the defect wavefunction. This effect is analogous to that shown in DFT studies of semiconducting nanocrystals [116] and nanowires [332], in which the defect formation energy for a substitutional dopant atom increases as the system size decreases. Furthermore, we can quantify the confinement

[8]As is known to diffuse in Si via a dual vacancy-interstitial mechanism, with the possible addition of some concerted (direct) exchange between lattice sites [329]; the interface segregation activation energy should be considered as an addition to the energy barrier for normal dopant diffusion in bulk (estimated at 1.07 eV from experimental measurements [330], and 1.19 eV from theoretical investigations [331]).



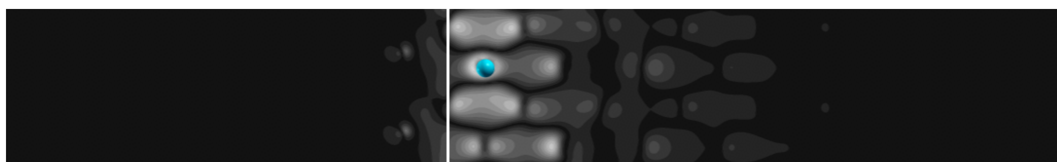
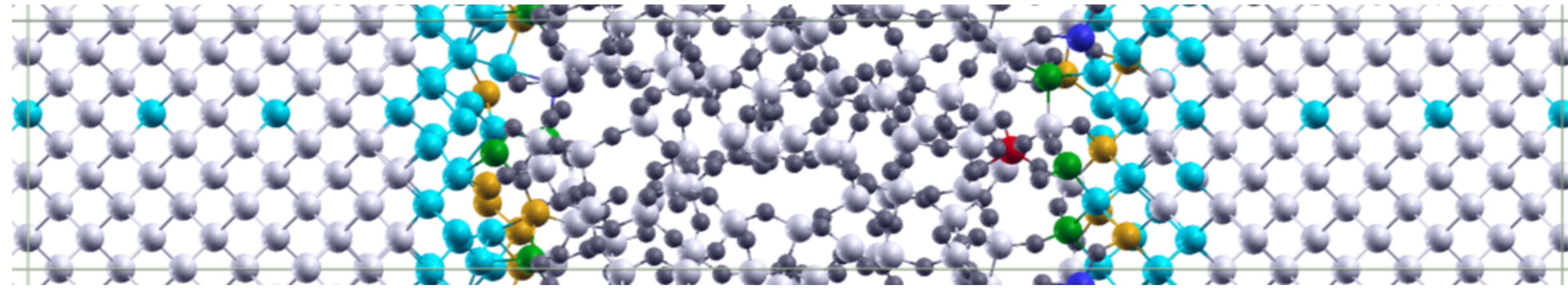
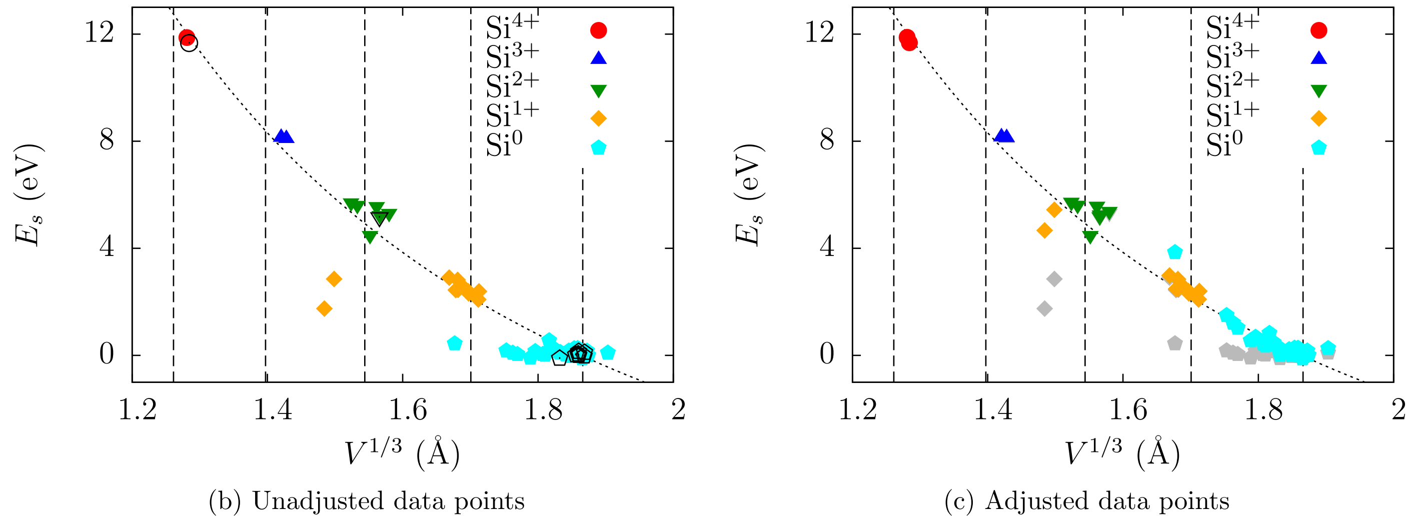


Figure 7.6: Calculated spread Q and displacement of the centre of the charge density distribution from the arsenic position in the z-direction Δz for the dopant donor level at different sites in the ideal interface system. The values of Δz are positive, as the charge distribution is shifted away from the silica region ($z < 0$). An example charge density distribution contour map for a defect site close to the interface is shown below in the z-y plane; the vertical white line indicates the position of the interface, with the silica region to the left. A square root scale is used, with lighter regions indicating a higher density.

by calculating the quadratic spread of the charge density of the donor level eigenstate; this is shown in Fig. 7.6 for all the Si^0 defect sites. As can be seen, the spread monotonically decreases (up to the second Si^0 monolayer) as the As ion approaches the interface; the example density plot shows that this is due to the charge distribution being confined by the interface, resulting in a long-range decay into the bulk silicon region and an abrupt decay into the silica. This asymmetric decay also results in the centre of the charge density being shifted from the position of the defect site and away from the interface. As expected, the calculated magnitude of the shift in the z-direction increases



(a) The disordered interface supercell used for the calculations (details are given in the caption of Fig. 7.5).



(b) Unadjusted data points

(c) Adjusted data points

Figure 7.7: Segregation energy of the arsenic dopant at the disordered interface with respect to its bulk position, as a function of the cube root of the defect volume V. The vertical dashed lines indicate the value of $V^{1/3}$ corresponding to the ideal volume for each oxidation state. Filled coloured symbols indicate sites in the disordered system, and empty black symbols ones in the ideal system. The fine dashed line is a function of the form $V^{-2/3}$ fitted to the data points for the ideal system only. In (b), the unadjusted data points are shown; in (c), the same data points are shown with the addition of a strain energy term (as explained in the text), with the unadjusted values now appearing in light grey.

as the As ion approaches the interface (shown in the same figure).

The first regime, therefore, illustrates the large-scale confinement of the defect wavefunction at the interface between materials with different dielectric constants, which has a relatively small effect on the segregation energy (the ionisation energy, instead, is expected to be greatly influenced by this, as discussed in previous studies [317,318]). This effect depends only on the distance of the dopant to the interface, instead of on the detailed ionic configuration around the defect site. However, it is the local configuration that has the greatest effect on the segregation energy: this is evident from the low energy of the sites at the first Si^0 monolayer, and the high energy of the Si^{2+} site, despite both of them residing in the same large-scale interface environment (i.e., having an approximately equal position in the z-direction).

In order to explain the effect of the local environment on the segregation energy, therefore, we propose a simple 'particle in a box' model, for which the energy increases as the defect volume decreases due to the confinement of the defect charge at the arsenic site; the defect volume can be approximated as before by the tetrahedron defined by its four nearest neighbours in the lattice. The model is applicable for all suboxide states, and is consistent with the calculated segregation energies, since the defect volume decreases as the oxidation of the site increases. However, the local strain at the undoped defect site must also be considered: irrespective of its sign, such a strain will decrease the segregation energy, as it will make it favourable for the Si ion to be placed in bulk (and, consequently, for the As ion to be placed at the site of interest); this is consistent with the calculated negative segregation energy for the highly strained site at the first Si^0 monolayer. In the next part of this section, we test this model quantitatively for the case of the disordered interface.

The disordered interface

In the disordered interface system (shown in Fig. 7.7a) there are a large number of inequivalent sites for all suboxide states, allowing us to investigate in detail the effect of defect volume and local strain on the arsenic segregation energy; we perform calculations on 67 sites in total.

Fig. 7.7b shows the As segregation energy for all the simulated defect sites as a function of the cube root of the defect volume $V^{\mathrm{As}} \equiv V$, both for the ideal and disordered interfaces. We fit a function of the form $V^{-2/3}$ (from the 3D 'particle in a box' model) to the data points obtained for the ideal interface only (Si^{0}, Si^{2+} and Si^{4+} sites). As can be seen, most of the data points for the disordered interface are in excellent agreement with this curve; in particular, the Si^{1+} sites follow it closely for a small range of volumes.

The ideal volume for each oxidation state is also indicated, showing that the majority of defect sites have little strain. However, there are a small number of outlying data points, mainly for the Si^{0} and Si^{1+} sites. These points all correspond to highly strained configurations, and have noticeably lower energies than the prediction obtained from the fitted curve. As already explained, this can be understood by considering the energy gain of the doped system with respect to the reference configuration, as the Si ion is effectively moved from the strained site to an unstrained bulk site with lower energy. The total segregation energy for a particular site, therefore, can be divided into two main contributions: (i) the energy of the As ion from the 'particle in a box' model, and (ii) a 'strain energy', due to the difference between the defect volume of the site in question occupied by an Si ion (i.e., the undoped site) and the ideal volume of that site based on its oxidation state; this latter term always lowers the total segregation energy, as the strain on the Si ion disappears when it is moved to the bulk. We note that the same long-range confinement regime as for the ideal interface is observed for defect sites in bulk-like Si far from the interface; however, for the first three Si^{0} monolayers this is swamped by the effect of the local environment.

Fig. 7.7c shows the same data points for the disordered system, now adjusted by the addition of an approximate strain energy for each defect site. For this adjustment, we use a function of the form $\left(V^{\mathrm{Si}} - V_0^{(i)}\right)^2$, where V^{Si} is the volume of the undoped site, and $V_0^{(i)}$ is the ideal volume for oxidation state i. For unrelaxed defect configurations, $V^{\mathrm{As}} = V^{\mathrm{Si}}$, and so the superscript can be neglected; however, upon lattice relaxation (as considered in the next part of this section) this is no longer the case. The function is fitted so as

to give the best agreement of all the points with the $V^{-2/3}$ curve. This adjustment successfully shifts the outlying data points upwards so as to be in good agreement with the curve, and has a negligible effect on all other points. Furthermore, different oxidations states now appear together in mixed clusters along the curve, thus supporting the use of a single model for all states. We stress that the true segregation energy is the unadjusted value obtained from the DFT total energy calculations; the adjustment is only used to obtain an approximation of the contribution to the segregation energy from the local quantum confinement effect on the dopant, and, hence, to illustrate the accuracy of the local volume+strain model in describing the total segregation energy. The model only uses three free parameters in total (the two coefficients for the $V^{-2/3}$ and $\left(V^{\text{Si}} - V_0^{(i)}\right)^2$ terms, and a constant shift); consequently, the ratio of the number of data points to the number of fitting parameters is very high.

It is clear, therefore, that obtaining a negative segregation energy for any suboxide state other than Si^0 is extremely unlikely, as a large positive strain is needed to compensate for the small defect volume; no such sites are found in our MC simulations of the amorphous interface (both for the small supercell considered here and the large supercell described in Sec. 7.1.3). For the Si^0 sites close to the interface, however, it is possible to find favourable lattice sites for segregation even with a negative local strain (as is the case for almost all sites in our interface systems), if the decrease in energy due to the strain is large enough to overcome the increase due to the reduced defect volume. In the disordered system, we find four sites with a negative value of E_s out of a total of 45 sites in the first few Si^0 monolayers at the interface. The maximum energy gain from segregation is of 0.11 eV.

Lattice relaxation

So far, we have only considered unrelaxed defect configurations; however, the As substitutional defect in bulk Si features a small outwards relaxation (as described in Sec. 7.2.2), that might differ in magnitude for sites at the interface due to the additional strain field. We perform structural relaxation calculations

System	Ox. state	Pos.	Rel.	Bond lengths (Å) a	b	c	d	Defect vol. (Å^3)	Strain	E_s (eV)
Ideal	Si^0	Bulk Si	No	2.32	2.32	2.32	2.32	6.43	−0.01	0.00[a]
			Yes	2.42	2.42	2.42	2.42	7.25	0.16	−0.23
		Int.	No	2.31	2.31	2.30	2.30	6.15	−0.05	−0.11
			Yes	2.39	2.38	2.38	2.38	6.81	0.05	−0.39
	Si^{2+}	Int.	No	2.35	2.31	1.65*	1.63*	3.84	0.04	5.17
			Yes	2.46	2.41	1.87*	1.85*	4.68	0.27	3.51
Disordered	Si^0	Bulk Si	No	2.32	2.32	2.32	2.32	6.43	−0.01	−0.01
			Yes	2.42	2.42	2.42	2.41	7.24	0.11	−0.24
		Int.	No	2.49	2.34	2.32	2.31	6.49	−0.00	−0.12
			Yes	2.75	2.38	2.38	2.37	7.37	0.13	−0.50
	Si^{1+}	Int.	No	2.66	2.47	2.38	1.66*	3.27	−0.34	1.74
			Yes	2.82	2.43	2.39	1.94*	3.68	−0.25	0.75

Table 7.2: Summary table of the results obtained for the lattice relaxation of the doped system. The oxidation state and position in the supercell (either in the bulk-like region in the middle of the Si layer or in the interface region) are listed for each defect site considered. Also listed are the bond lengths between the As ion and its four neighbours a,b,c,d; asterisks (*) denote an As–O bond. The strain is defined as $\left(V^{\mathrm{As}} - V_0^{(i)}\right)/V_0^{(i)}$, where V^{As} is the volume of the doped site[b], and $V_0^{(i)}$ is the ideal volume for the undoped site of oxidation state i.

[a] This configuration is taken as the energy reference for both systems.

[b] For unrelaxed defect configurations, $V^{\mathrm{As}} = V^{\mathrm{Si}}$, and so the value of the strain listed in the table is effectively the strain on the undoped site; however, for relaxed configurations this is no longer the case, and so the value listed does not correspond to a 'true' strain, and should instead be regarded simply as an indication of the change in defect volume upon relaxation.

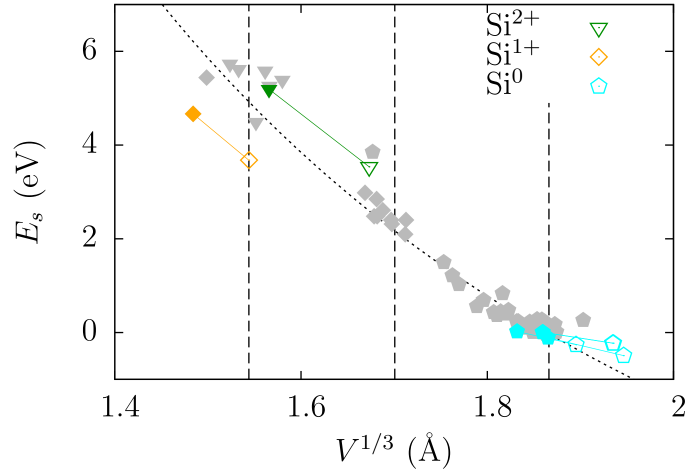


Figure 7.8: Effect of lattice relaxation on the segregation energy. Data points for both the ideal and disordered systems are shown. Filled coloured symbols indicate the unrelaxed energy, and empty coloured symbols the relaxed energy. The energy is adjusted by the addition of a strain energy term (as described in Sec. 7.2.3), applied using the unrelaxed defect volume both before and after relaxation. The unrelaxed values for all other sites appear in light grey.

for six defect sites in total, three each for the ideal and disordered interfaces. In both systems we relax the bulk-like site at the centre of the Si layer, the Si^0 site at the interface with the greatest energy gain from segregation, and a partially oxidised site at the interface (Si^{2+} for the ideal system and Si^{1+} for the disordered one). The relaxed configurations and energies are given in Table 7.2.

For all sites, the calculations show an outwards relaxation and a decrease in energy of the system on the order of 0.1–1 eV. The energy gain is greater for the Si^0 sites at the interfaces that those in bulk, resulting in an increase in the magnitude of the segregation energy from 0.11 eV to 0.16 eV in the ideal system, and from 0.11 eV to 0.26 eV in the disordered one. We note that the relaxed bond lengths and decrease in total energy for the defect site at the centre of the Si layer are in good agreement with those calculated in the 256-atom BCC supercell of bulk Si. As observed previously for the P substitutional defect at the Si–SiO_2 interface [323], the optimised structure for the neutral

As substitutional defect at the interface exhibits one As–Si bond that is longer than the other three. This effect is most pronounced in the disordered system, with a difference of 0.37 Å between the long bond and the short ones; however, the dopant still appears to be active after relaxation, as there is a negligible shift in the position of the donor level in the Kohn-Sham band structure.

Finally, Fig. 7.8 shows the effect of the outwards ionic relaxation on the segregation energy in terms of our local quantum confinement model. The increase in the defect volume due to the relaxation is expected to simply shift the segregation energy down the $V^{-2/3}$ curve fitted previously. As can be seen, the calculated relaxed segregation energies are in reasonable agreement with the model; in particular, the relaxation of the Si^{2+} site in the ideal system results in the energy shifting from the Si^{2+} to the Si^{1+} cluster of data points along the curve, thus giving further support to the underlying assumption of the model, i.e., that the most important factors in determining the segregation energy are the defect volume and the local strain, as opposed to the nature of the chemical bonds.

7.2.4 Conclusions

In this chapter we have studied the energetics of neutral substitutional arsenic defects at silicon sites close to and at the Si–SiO_2 interface. We use supercells with perfect bonding topologies, and, hence, no dangling bonds at the interface; therefore, all defect sites are four-fold coordinated, although the oxidation of the undoped site can vary between Si^0 and Si^{4+}. It is found that defect configurations with As–O bonds are highly energetically unfavourable respect to bulk-Si-like defect sites with four As–Si bonds. Using the information from 78 defect sites in total (11 in the ideal system and 67 in the disordered one), we have shown that the segregation energy for all possible oxidation states is well described by a $V^{-2/3}$ relationship, where V is the defect volume estimated by the position of its four neighbours, corresponding therefore to a simple ‘particle in a box’ or quantum confinement picture; as the Si–O bond is significantly shorter than the Si–Si bond, partially and fully oxidised defect sites have a smaller volume and, hence, a higher energy.

The strain of the undoped defect site is also shown to have a crucial effect on the segregation, as highly strained sites exhibit a lower segregation energy than that predicted by the quantum confinement effect alone, as they favour the reference unstrained bulk position for the Si atom that is removed by the substitutional impurity. Therefore, although the confinement effect is generally unfavourable for segregation, we find a number of strained Si^0 sites within the first three Si monolayers at the interface in the disordered system that are energetically favourable with respect to the bulk site (corresponding to a density of $\sim$0.02 $Å^{-2}$), thereby driving the segregation process; the largest energy gain from segregation is calculated to be 0.26 eV after relaxation of the ionic positions. In conclusion, therefore, our calculations confirm the presence of electrically active segregated defects at the interface that has been suggested from experimental measurements.

Chapter 8

Conclusions

In this thesis, we have studied the properties of several point defect systems in silicon using DFT. We have both tackled theoretical issues relating to the accuracy of the calculations due to finite size effects, and presented novel results for defect centres of technological relevance. Furthermore, we have demonstrated the use of a variety of computational methods for the study of such systems: (i) MLWFs, for an *ab initio* (i.e., parameter-free) interpretation of the chemical bonding around the defect site, and for the accurate calculation of valence band offsets in charged supercells; (ii) linear-scaling DFT based on a density matrix formalism, for the simulation of large supercells with plane-wave accuracy, and (iii) CRN-MC parametrised from *ab initio* data, for the generation of large-scale disordered perfect networks at finite temperature, as a starting point for more accurate DFT calculations.

For the vacancy in bulk silicon, we have investigated the well-known issue of the slow convergence of defect properties with system size in the supercell approach. We separate the total error into three main contributions: the electrostatic interaction between defect charge densities, the wavefunction overlap between defect centres, and the elastic interaction upon relaxation of the ionic positions; furthermore, the determination of the VBM position can be considered an additional finite size error for charged defect supercells. Unfortunately, our results indicate that there isn't a clear dominant interaction that accounts for most of the error; therefore, it appears unlikely that any single correction

scheme can provide a reliable method of extrapolating to the limit of infinite system size. We note that even for the case of highly charged defects (as described for the gold substitutional), the monopole-monopole interaction does not dominate for supercells up to 256 atoms, and, hence, the Leslie-Gillan correction [245] does not improve the convergence, contrary to what might reasonably be expected.

We have also investigated the effect of k-point sampling on the convergence; in particular, we find the choice of sampling to play an important role in defect systems featuring a large Jahn-Teller distortion and a change of point group symmetry upon ionic relaxation, with a dense sampling scheme favouring the undistorted lattice, and a Γ-point (or other k-points with sufficiently high symmetry) scheme favouring the correct symmetry-lowering distortion. However, the quantitative result is also affected by the description of the host material, for which the dense sampling is clearly more accurate. Overall, we find that the dense sampling offers a faster convergence for defect formation energies, while the Γ-point sampling is preferable for the stable charge state transition levels [333].

By carefully converging the defect properties of interest with respect to the system size, we have shown that LDA-DFT is in good agreement with the known experimental results for both the vacancy and the gold impurity; this is true both for qualitative properties of the system (the symmetry of different charge states and the outward relaxation on electron emission), and quantitative ones (the position of the transition levels in the band gap, as well as the formation energy of the vacancy compared with the available experimental estimates). In particular, despite the band gap problem in DFT, the transition levels (including the ionisation energy of arsenic in bulk silicon) are predicted to within a few hundredths of an electronvolt; this level of accuracy has allowed us to unambiguously assign the two known gold levels to specific double electron transitions, leading to our prediction of the negative-U behaviour for this centre.

It is important to note that our results for the finite size convergence properties of these defect systems are not 'universal', and we do not recommend them indiscriminately as a code of best practice for *any* point defect calcula-

tion. Instead, it is clear from our results and previous studies that different types of defects can exhibit markedly different convergence patterns (e.g., it is unlikely that a defect centre with a trivial relaxation pattern will benefit from a Γ-point sampling). Nevertheless, due to the noticeable similarity in the properties of the vacancy and gold substitutional (the almost identical converge of the formation energy for the unrelaxed defect, and the negative-U effect due to Jahn-Teller distortion), we suggest that most defects encompassed by Watkins' model (vacancies or impurity substitutionals in tetrahedrally-bonded semiconductors) should also follow the same convergence patterns; this could also be extended to the case of binary semiconductors in the zinc blende structure.

We have also studied the segregation properties of a shallow-level defect, the arsenic substitutional in silicon, at the interface between crystalline silicon and its amorphous oxide. For this disordered system we employ a multiscale approach, making use of a computationally inexpensive CRN model combined with a simple classical potential based on the bond lengths and angles between atoms to run MC simulations at finite temperature, and, hence, generate a realistic interface structure to use in the DFT simulations of the system with the additional inclusion of a dopant atom. We have investigated both the long-range effect of the interface on the arsenic segregation energy (showing a small activation barrier due to the quantum confinement of the defect wavefunction tail), and the local effect of the volume and strain of the substitutional site (showing that segregation is favoured for sites with a larger volume than in bulk silicon, and for highly strained sites independently of the sign of the strain). Our results suggest that the largest contribution to the segregation energy is given by the local environment of the defect site, and that this can be reasonably characterised with a minimal three-parameter model.

8.1 Future work

An obvious general solution to the problem of finite size errors in point defect calculations is to explicitly converge the results up to very large supercell sizes. Although this is indeed feasible (as we have shown for our defects), such calculations have not yet become routine, and the continuing development of

accurate and robust linear-scaling DFT codes is very important in this respect. It is also interesting to note that the study of point defects is an area in which linear-scaling DFT does not encounter the serious problem of combinatorial explosion of configuration space with increasing system size, as only the first few shells of atoms around the defect are significantly perturbed from their equilibrium lattice position, irrespective of supercell size.

In the case of the ONETEP code, quasi-plane-wave accuracy is achieved by using an underlying psinc basis, and allowing full variational freedom to the localised orbitals within a cut-off radius. Inevitably, however, this leads to additional numerical parameters that need to be monitored (i.e., the real-space cut-off for both the NGWFs and the density kernel); furthermore, the corresponding physical quantities are not exponentially localised for systems with a vanishing band gap, such as partially occupied deep defect levels, or even shallow levels close to a band edge.

A more serious, and yet unsolved, problem that we encountered in our simulations is that of converging the NGWF optimisation to a high enough tolerance for accurate calculations of the total energy and ionic forces. The conjugate gradients minimisation algorithm that is employed, in fact, is found to exhibit very poor convergence properties for certain systems, for which the quantity that is minimised (the RMS NGWF gradient) reaches a plateau beyond which it is hard to obtain further progress towards the energy minimum. This is observed to occur when there is a significant build up of charge at the boundary of the NGWF sphere, since in such cases the NGWF gradient features non-negligible components outside of the sphere that are 'shaved' in order to maintain the strict localisation of the orbital. We have found this problem to be particularly serious in the description of silica, due to the large extent of the orbitals centred on the oxygen ions.

We have proposed a projected gradient method to address this issue, by which the psinc components of the NGWF gradient are modulated by the projection onto a smaller subspace defined by a set of atom-centred orbitals that decay smoothly to zero at the boundary imposed by the cut-off radius. This atom-centred basis set should ideally give the NGWFs as much variational freedom as possible in the central region of the NGWF sphere, but constrain

them to decay smoothly at the boundary. Consequently, the minimisation algorithm will converge to a constrained energy minimum that is higher than the true minimum, but is also expected to avoid the gradient stagnation problems of the unconstrained algorithm.

Due to the NGWF optimisation problem, we did not make use of linear-scaling DFT in our study of the doped Si–SiO_2 interface system. Instead, we employed conventional cubic-scaling plane-wave DFT, thus limiting the supercell size to ∼500 atoms (nevertheless, as previously noted, this is much larger than the system sizes used in previous studies of the doped interface). The finite size error in the segregation energy of the neutral dopant in the interface system is expected to be negligible in the z-direction (perpendicular to the interface), and on the order of 0.01 eV in the x-y plane[1]. We find a much larger finite size error, on the order of 0.1 eV, for the arsenic ionisation energy (i.e., 200% of the value itself), that can be attributed almost entirely to the total energy of the $q = +1$ defect supercell. Interestingly, however, this large error is caused by the tetragonal cell shape; in fact, for a cubic supercell with a much smaller volume (we use the 64-atom SC one for comparison), the corresponding error is found to be an order of magnitude smaller. Because of this, it was not possible to calculate accurate arsenic ionisation energies at the interface in the present study. We expect to find an increase in the ionisation energy of defects close to the interface due to dielectric confinement effects, as suggested by recent experimental results. To this end, we postulate that linear-scaling DFT codes such as ONETEP could be employed in future studies to simulate extremely large interface supercells, in which the finite size error introduced by the tetragonal cell shape is negligible.

[1] This is estimated by comparing the calculated $E\,(1{+}/0)$ level of the defect in the 64-atom SC supercell and the 256-atom BCC supercell; the error in the segregation energy is assumed to be similar to that in the transition levels, as both quantities are calculated from the difference in total energy of two defect supercells.

Appendix A

Phonon calculations in ONETEP

The calculation of the phonon frequencies of nanostructures has many important applications; they can be used for understanding and predicting spectra obtained from vibrational spectroscopy experiments (such as infrared and Raman spectroscopy, and inelastic neutron scattering), as well as the thermodynamic properties of materials (such as specific heat, lattice expansion, and phase stability). Furthermore, in relation to point defects, the change of vibrational energy and entropy in the lattice when creating a defect site affects the equilibrium concentration of that defect (see Sec. 4.2). In general, the first-principles calculation of the vibrational modes of a material, and, from these, the free energy and other thermodynamic functions, is known as *ab initio* thermodynamics [334].

For stable solids and other harmonic systems we make use of the harmonic approximation, in which the total energy of the system E^{tot} is expanded to quadratic order in the displacement of the ions about their equilibrium positions:

$$E^{\mathrm{tot}} = E^{\mathrm{eq}} + \frac{1}{2} \sum_{a,\alpha,\kappa,a',\alpha',\kappa'} u_{a,\alpha,\kappa} \phi_{\kappa,\kappa'}^{\alpha,\alpha'} \left(a, a'\right) u_{a',\alpha',\kappa'}, \qquad \text{(A.1)}$$

where E^{eq} is the equilibrium energy and $u_{a,\alpha,\kappa}$ denotes a small displacement of ion α belonging to unit cell a in the Cartesian coordinate direction κ from its equilibrium position; $\boldsymbol{\phi}\left(a, a'\right)$ is known as the force constants matrix, defined

as

$$\phi_{\kappa,\kappa'}^{\alpha,\alpha'}(a,a') = \frac{\partial^2 E}{\partial u_{a,\alpha,\kappa}\partial u_{a',\alpha',\kappa'}}. \tag{A.2}$$

It can be shown that the phonon frequencies $\omega_{\mathbf{q},n}$ at wavevector $\mathbf{q}$ are the eigenvalues of the dynamical matrix $\mathbf{D}(\mathbf{q})$, which can be calculated from the Fourier transform of the force constants matrix:

$$D_{\kappa,\kappa'}^{\alpha,\alpha'}(\mathbf{q}) = \frac{1}{\sqrt{M_\alpha M_{\alpha'}}}\sum_a \phi_{\kappa,\kappa'}^{\alpha,\alpha'}(a,0)\,\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{R}_a}, \tag{A.3}$$

where M_α is the mass of ion α and $\mathbf{R}_a$ is the lattice vector displacement for unit cell a. The vibrational free energy for the unit cell is then given by

$$F(T) = \frac{1}{2}\sum_{\mathbf{q},n}\omega_{\mathbf{q},n} + k_\mathrm{B}T\sum_{\mathbf{q},n}\ln\left(1-\mathrm{e}^{-\omega_{\mathbf{q},n}/k_\mathrm{B}T}\right), \tag{A.4}$$

where the first term is the zero-point energy of the system, and the second term is the temperature-dependent part of the free energy. We note that in the limit of an infinite periodic system the sum over $\mathbf{q}$ should be replaced by an integral of the phonon dispersion curves over the first Brillouin zone.

There are two broad approaches for the *ab initio* calculation of the force constants matrix: perturbation methods based on linear response theory [335,336], and finite difference methods based on the calculation of ionic forces[1] for small ionic displacements introduced in the structure [339, 340]. Despite requiring a more complicated formalism, the linear response method is generally preferred for calculations on crystalline materials, since it allows all phonon wavevectors to be computed using the primitive unit cell; for the finite difference method, instead, a supercell of material large enough to capture the real-space decay of the elements of the force constants matrix is needed.

We have implemented the calculation of phonon frequencies in ONETEP using the finite difference method; this is because we are naturally interested in taking advantage of the linear-scaling nature of the code, which makes the

[1]The force on ion α is given by $\mathbf{F}_\alpha = -\partial E_0/\partial\mathbf{R}_\alpha$, where E_0 is the total ground-state energy in the Born-Oppenheimer approximation, and $\mathbf{R}_\alpha$ is the position of the classical ion. This expression is evaluated by making use of the Hellmann-Feynman theorem [337, 338].

routine calculation of large supercells feasible (indeed, the 'cross-over' point above which linear-scaling DFT methods becomes more computationally efficient than conventional cubic-scaling methods is estimated at several hundred atoms for solids, since there is no empty space in the system). The calculation of the phonon frequencies of a regular crystal is therefore straightforward, as the number of ionic displacements required is proportional to the number of atoms in the primitive unit cell (by symmetry), and, hence, is independent of the supercell size used.

We calculate the elements of the force constants matrix by using the central-difference formula

$$\phi_{\kappa,\kappa'}^{\alpha,\alpha'} \approx \frac{F_{\alpha,\kappa}^{+} - F_{\alpha,\kappa}^{-}}{2d}, \tag{A.5}$$

where $F_{\alpha,\kappa}^{\pm}$ is the force on ion α in direction κ caused by a displacement $\pm d$ of ion α' in direction κ'. Therefore, $6N$ calculations are needed in total, where N is the number of atoms in the unit cell of the crystal. However, each of these calculations is simply a small perturbation on the equilibrium configuration. Therefore, the converged set of NGWFs $\{\xi_{\beta}(\mathbf{r})\}$ and density kernel $\mathbf{K}$ that are obtained from the equilibrium ground-state calculation (and that contain all the information about its electronic structure, as discussed in Sec. 3.3.3) are used as the starting guess for each of the displacement calculations. The standard minimisation algorithms employed by the code therefore converge in a greatly reduced number of iterations, as this starting guess already represents to a good approximation the ground state of the new (perturbed) system.

We have tested the method using a 1000-atom simple cubic supercell of bulk silicon. We use the same norm-conserving pseudopotential as described in Sec. 5.3.1, a plane-wave cut-off energy of ∼800 eV, and nine NGWFs on each silicon atom with a truncation radius of 4.0 Å. We do not truncate the density kernel. For bulk silicon, the calculation of the force constants matrix can be reduced to a single displacement calculation, as the two atoms in the primitive cell are symmetrically equivalent, as well as the six displacements on each atom. The phonon frequencies can then be calculated at an arbitrary q-point by applying an effective truncation of the force constants (and, hence, of the sum in Eq. A.3) far away from the displaced atom.

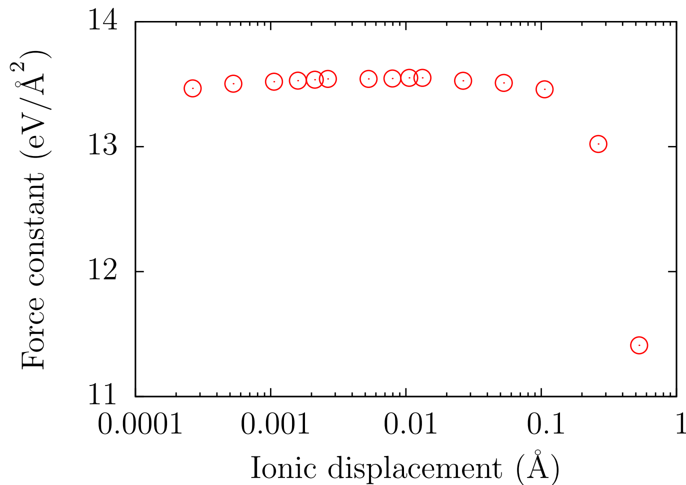


Figure A.1: On-site force constant for bulk silicon calculated with the finite difference method for a range of ionic displacements.

Fig. A.1 shows the on-site force constant $\phi^{\alpha;\alpha}_{\kappa,\kappa}$ calculated by finite difference for a range of ionic displacements d. The value of $\phi^{\alpha;\alpha}_{\kappa,\kappa}$ is approximately constant for a wide range of displacements, and is observed to deviate for $d > 0.1$ Å. We note here that the accuracy of the calculated ionic forces has a large impact on the phonon frequencies; ONETEP has been shown to be in good agreement with conventional plane-wave DFT in this respect, with forces agreeing to within 0.1 eV/Å for a plane-wave cut-off energy of 800 eV [341] (the discrepancy is mostly due to the fact that the basis sets employed are not entirely equivalent in their high-frequency components; in fact, as the cut-off is increased, the forces converge to an almost perfect agreement). Therefore, in order to minimise the impact of this discrepancy, we fix the ionic displacement to ∼53 mÅ (0.1 a_0), which is close to the large-displacement limit of the linear range (producing a force on the ion of 0.71 eV/Å).

The calculated phonon dispersion curves along high-symmetry directions are shown in Fig A.2; these are in good agreement with the experimental data. In order to recover the correct linear behaviour of the long-wavelength

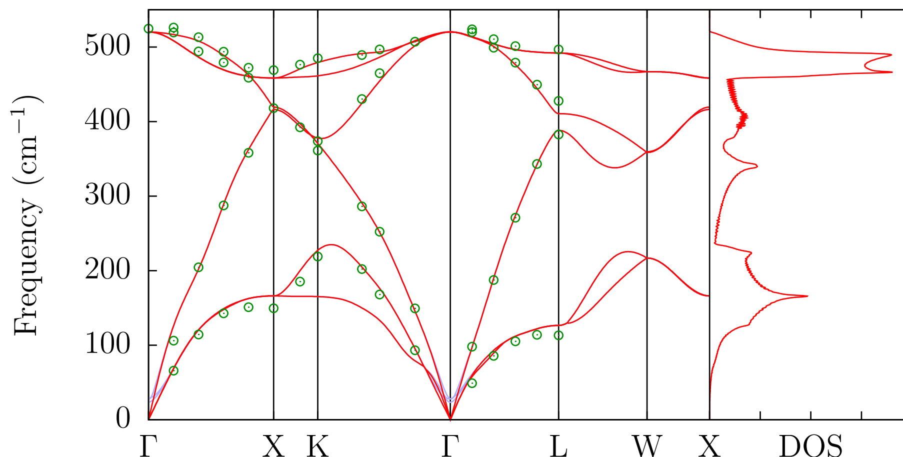


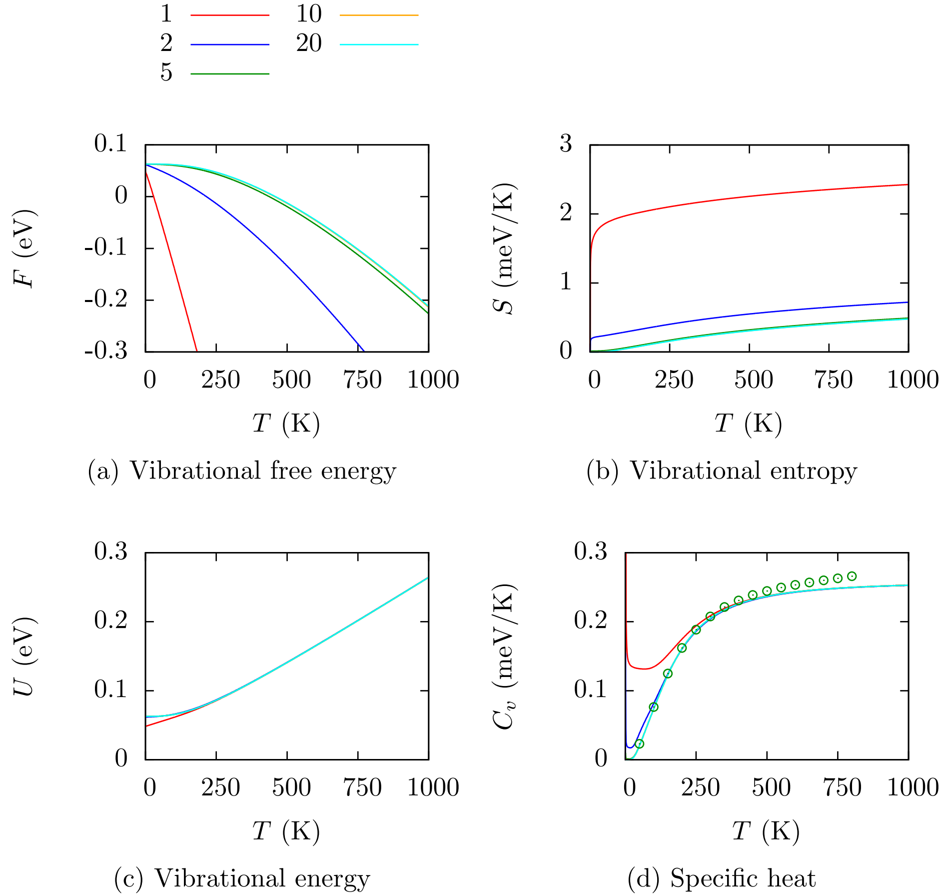
Figure A.2: Phonon dispersion and density of states for bulk silicon. The dispersion curves calculated with the acoustic sum rule correction are shown in red, and the uncorrected ones are shown in light blue (these are only visible for the acoustic modes around Γ). Experimental phonon frequencies from Ref. [342] are denoted by open circles. The density of states is calculated from a 100 × 100 × 100 grid of q-points.

acoustic modes around the Γ-point, we enforce the acoustic sum rule by a post-processing correction to the forces. The acoustic sum rule is given by

$$\sum_{\alpha,a} \phi^{\alpha,\alpha'}_{\kappa,\kappa'}(a,0) = 0 \tag{A.6}$$

(i.e., for any displacement, the sum of the forces on all the ions in a given direction must be equal to zero). The correction, therefore, subtracts from each individual force component the spurious non-zero total of the sum on the left hand side of this equation divided by the total number of atoms in the system. As can be seen from the figure, the sum rule correction only affects the three acoustic vibrational modes close to the Γ-point, and has a negligible effect on all other phonon modes.

Using Eq. A.4 we calculate the vibrational free energy of the crystal as



(a) Vibrational free energy

(b) Vibrational entropy

(c) Vibrational energy

(d) Specific heat

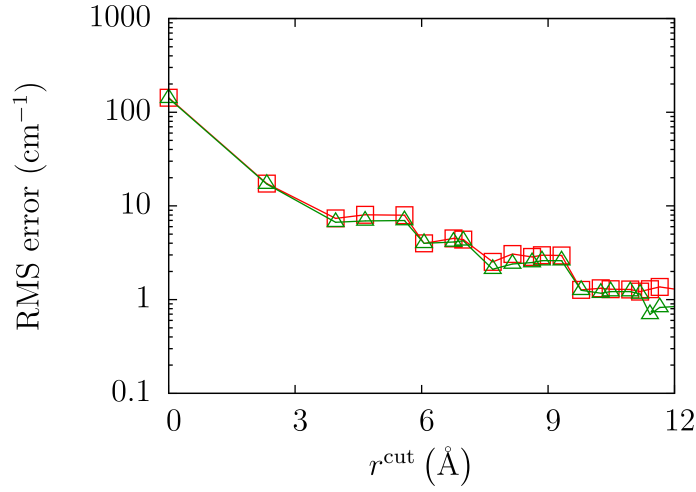
Figure A.3: Thermodynamic functions for bulk silicon obtained by summing over different q-point meshes (the values in the key refer to an $n \times n \times n$ grid of q-points); $n = 10$ and $n = 20$ are indistinguishable in all plots. Experimental measurements of the specific heat from Ref. [343] are denoted by open circles. All quantities are given per atom.

a function of temperature; from this, all other thermodynamic functions can be calculated using the well-known thermodynamic relations. In Fig. A.3, we show the free energy, entropy, internal energy and specific heat obtained from our calculations. The thermodynamic functions need to be converged with respect to the fineness of the q-point grid used to approximate the integral over the dispersion curves in q-space; the figures show a $10 \times 10 \times 10$ grid to be sufficient in this respect. The specific heat per atom is in excellent agreement

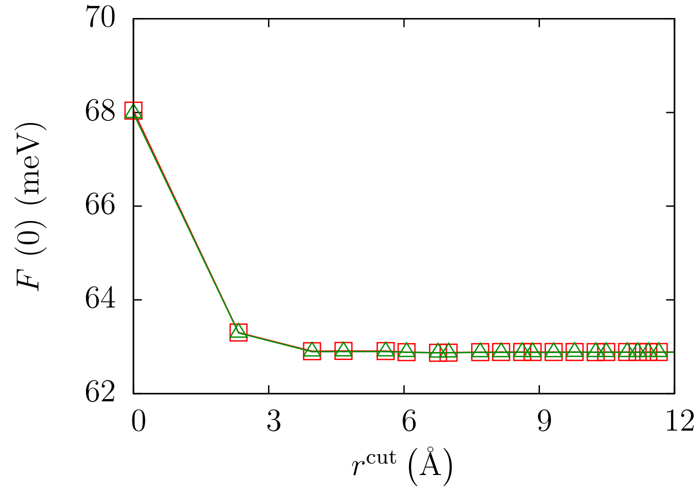
with experiment at low temperatures, but is increasingly underestimated at higher temperatures as anharmonic effects (neglected in our model) become more important. At 800 K, there is a 6% discrepancy between the calculated and experimental values, similar to previous results reported for LDA-DFT [344].

It is interesting to note the similarity between the finite difference supercell method for phonon calculations and the tight-binding band structure interpolation method using maximally-localised Wannier functions discussed in Appendix C. In both cases, a real-space truncation is used to construct either the dynamical matrix (in the former case) or the Hamiltonian matrix (in the latter) at an arbitrary wavevector in the Brillouin zone. The maximum range of the MLWF-derived tight-binding basis is dependent both on the size of the supercell and the fineness of the k-point grid in the calculation of the Bloch functions used for the wannierisation procedure; therefore, the most computationally efficient approach for increasing this range is to fix the system size to the primitive unit cell and increase the number of k-points. Instead, the maximum range for the force constants matrix is dependent only on the supercell size, and so large systems are needed for an accurate calculation of the phonon frequencies.

We have investigated how variations in the range of the dynamical matrix affect the accuracy of the calculated phonon frequencies, by imposing a real-space cut-off radius r^{cut}: the force constant $\phi_{\kappa,\kappa'}^{\alpha,\alpha'}(a,0)$ is set to zero if the distance between ion α in unit cell a and ion α' in the reference unit cell is greater than this cut-off. Fig. A.4 shows the effect of the cut-off on the RMS error in the phonon frequencies throughout the Brillouin zone (with respect to the maximum cut-off, given by half the lattice parameter of the 1000-atom supercell), and the zero-point energy $F(0)$. As can be seen, the real-space decay of the force constants is extremely rapid: the error in the zero-point energy is reduced to 0.02% by considering only on-site, first- and second-nearest neighbour interactions in the lattice ($r^{\mathrm{cut}} = 3.81$ Å). Furthermore, the RMS error in the phonon frequencies decays exponentially with cut-off. This behaviour suggests, therefore, the possibility of constructing model dynamical matrices of very large systems by combining short-ranged force constants matrices of



(a) RMS error in the phonon frequencies



(b) Zero-point energy

Figure A.4: Effect of applying a real-space cut-off to the force constants matrix. Open green triangles show the data with the acoustic sum rule correction, and open red squares show the uncorrected data.

smaller systems, in complete analogy with a method that has been recently demonstrated for Hamiltonian matrices in a MLWF basis [345].

Appendix B

The CASTEP-TO-WANNIER90 code

In this appendix we describe the algorithms and functionality of the CASTEP-TO-WANNIER90 interface code, that was written in order to use the Kohn-Sham Bloch states obtained from CASTEP as input for WANNIER90 to find the MLWFs of the system. Examples of the Wannier functions obtained in this way can seen in Figs. 5.9, 5.10 and 6.7. The interface code is written in Fortran90, and parallelised using MPI; external routines are provided by the FFTW library.

The main information required by WANNIER90 from the underlying electronic structure code is contained in the two matrices $M_{mn}^{(\mathbf{k},\mathbf{b})}$ and $A_{mn}^{(\mathbf{k})}$; the definition and a brief explanation for each is given below. Additionally, the Bloch states at all k-points need to be provided on a real-space grid in order to visualise the Wannier functions (however, this is not required for the wannierisation procedure itself), and the eigenvalues corresponding to the Bloch states need to be provided in the case of a disentanglement procedure, as well as for the calculation of the interpolated band structure, density of states and Fermi surface of the system.

As explained in Sec. 3.3.2, the MLWF representation of the subspace of interest is found by minimising the non-gauge-invariant contribution $\widetilde{\Omega}$ to the sum of the spreads of the Wannier functions, given by Eq. 3.47. $\widetilde{\Omega}$ requires the calculation of the matrix elements of the position operator between Wannier functions; as shown by Blount [346], these can be expressed in reciprocal space

in terms of the matrix elements of $\nabla_{\mathbf{k}}$ between the Bloch functions $|u_{n,\mathbf{k}}\rangle$:

$$\langle w_{m,\mathbf{T}}| \, \mathbf{r} \, |w_{n,\mathbf{0}}\rangle = \mathrm{i}\frac{V}{(2\pi)^3} \int \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{T}} \langle u_{m,\mathbf{k}}| \, \nabla_{\mathbf{k}} \, |u_{n,\mathbf{k}}\rangle \, \mathrm{d}\mathbf{k}, \tag{B.1}$$

where $|w_{m,\mathbf{T}}\rangle$ is the Wannier function in cell $\mathbf{T}$ associated with band m.

WANNIER90 assumes that the Brillouin zone is discretised on a uniform Monkhorst-Pack grid of k-points, and approximates the reciprocal-space integral in Eq. B.1 with a first-order finite difference scheme. As detailed in Ref. [221], $\widetilde{\Omega}$ can then be expressed solely in terms of the overlap matrix

$$M_{mn}^{(\mathbf{k},\mathbf{b})} = \langle u_{m,\mathbf{k}}|u_{n,\mathbf{k}+\mathbf{b}}\rangle \, , \tag{B.2}$$

where $\{\mathbf{b}_i\}$ is the set of vectors connecting k-point $\mathbf{k}$ with its nearest neighbours on the grid. The wannierisation algorithm uses this discrete representation to evolve $U_{mn}^{(\mathbf{k})}$ so as to minimise the total spread functional.

An initial guess for $U_{mn}^{(\mathbf{k})}$ must also be provided to WANNIER90 as the starting point of the minimisation routine. Although it is possible to use the Bloch functions themselves for this initial guess, it is generally preferable for the user to define a set of localised trial orbitals $\{g_n(\mathbf{r})\}$ (as many as the resulting MLWFs), that are projected onto the (non-periodic) eigenfunctions $|\phi_{n,\mathbf{k}}\rangle$ in order to provide the necessary information to generate a trial $U_{mn}^{(\mathbf{k})}$ matrix. The projections are given by the matrix

$$A_{mn}^{(\mathbf{k})} = \langle \phi_{m,\mathbf{k}}|g_n\rangle \, . \tag{B.3}$$

$A_{mn}^{(\mathbf{k})}$ is similarly used in WANNIER90's disentanglement routine (for the case of non-isolated bands), to generate an initial guess for the optimal subspace [74].

Our interface code calculates $M_{mn}^{(\mathbf{k},\mathbf{b})}$ and $A_{mn}^{(\mathbf{k})}$ from Eqs. B.2 and B.3, respectively; the input data it requires is provided in two files: `<seedname>.nnkp` (generated by WANNIER90 with the pre-processing command line option `-pp`), and either `<seedname>.check` (the standard wavefunction output file generated by CASTEP for self-consistent calculations) or `<seedname>.orbitals` (the corresponding file for non-self-consistent band structure calculations). From

these, the code extracts the following information:

- the unit cell lattice vectors and the number of k-points in the MP grid (from WANNIER90);
- the set of vectors $\{\mathbf{b}_i\}$ for each k-point (from WANNIER90);
- the definition of the trial orbitals $\{g_n(\mathbf{r})\}$ (from WANNIER90);
- the Fourier coefficients $\left\{c_{\mathbf{G}}^{(n,\mathbf{k})}\right\}$ in the plane-wave expansion of the Bloch functions for the irreducible set of k-points (from CASTEP);
- the space group symmetries of the system (from CASTEP);
- the Kohn-Sham eigenvalues $\{\varepsilon_{n,\mathbf{k}}\}$ (from CASTEP).

The Fourier coefficients for a periodic Bloch function are defined as

$$u_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{G}}^{(n,\mathbf{k})} \mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathrm{r}}; \tag{B.4}$$

the elements of the two matrices are therefore calculated as

$$M_{mn}^{(\mathbf{k},\mathbf{b})} = \sum_{\mathbf{G}} \left(c_{\mathbf{G}}^{(m,\mathbf{k})}\right)^* c_{\mathbf{G}}^{(n,\mathbf{k}+\mathbf{b})} \tag{B.5}$$

and

$$A_{mn}^{(\mathbf{k})} = \sum_{\mathbf{G}} \left(c_{\mathbf{G}}^{(m,\mathbf{k})}\right)^* \widetilde{g}_n(\mathbf{k}+\mathbf{G}), \tag{B.6}$$

respectively, where $\widetilde{g}_n(\mathbf{k})$ is the Fourier transform of the n-th trial orbital. CASTEP-TO-WANNIER90 is restricted to the use of $1s$-like trial orbitals, defined as

$$g_n(\mathbf{r}) = \frac{\alpha_n^{3/2}}{\sqrt{\pi}} \mathrm{e}^{-\alpha_n|\mathbf{r}-\mathbf{r}_n|}, \tag{B.7}$$

where $\mathbf{r}_n$ is the centre of the n-th orbital, and α_n its spread. The Fourier transform is given by

$$\widetilde{g}_n(\mathbf{k}) = \frac{8\alpha_n^{5/2}\sqrt{\pi}}{\left(k^2+\alpha_n^2\right)^2} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_n}. \tag{B.8}$$

The evaluation of the Bloch states at all k-points in the MP grid is computationally inefficient for the underlying electronic structure code, as the number of independent k-points can be reduced to those within the irreducible wedge of the Brillouin zone by making use of the symmetries of the Hamiltonian (see Sec. 3.2). The interface code, therefore, performs the inverse operation, 'filling in' the complete MP grid from the irreducible wedge using the list of space group symmetries provided by CASTEP and time-reversal symmetry.

The simplest case is that of k-points related by time-reversal symmetry, for which we have shown in Eq. 3.27 that $\phi^*_{n,\mathbf{k}}(\mathbf{r}) = \phi_{n,-\mathbf{k}}(\mathbf{r})$. Hence, the relationship between the Fourier coefficients of the Bloch states for band n at these two k-points is given by

$$c_{\mathbf{G}}^{(n,\mathbf{k})} = \left(c_{-\mathbf{G}}^{(n,-\mathbf{k})}\right)^* . \tag{B.9}$$

Space group symmetries are defined by a unitary transformation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$, for which it can be shown that

$$\phi_{n,\mathbf{k}}(\mathbf{R}\mathbf{r} + \mathbf{t}) = \phi_{n,\mathbf{k}'}(\mathbf{r}) , \tag{B.10}$$

where

$$\mathbf{R}\mathbf{k}' = \mathbf{k} \tag{B.11}$$

(the translation does not affect reciprocal space). Expanding Eq. B.10 in terms of Fourier coefficients, we obtain:

$$\mathrm{e}^{\mathrm{i}\mathbf{R}\mathbf{k}'\cdot\mathbf{R}\mathbf{r}}\mathrm{e}^{\mathrm{i}\mathbf{R}\mathbf{k}'\cdot\mathbf{t}}\sum_{\mathbf{G}}\left[c_{\mathbf{G}}^{(n,\mathbf{k})}\mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathbf{t}}\right]\mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathbf{R}\mathbf{r}} = \mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{r}}\sum_{\mathbf{G}} c_{\mathbf{G}}^{(n,\mathbf{k}')}\mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathbf{r}}. \tag{B.12}$$

Hence, since $\mathbf{R}\mathbf{k}'\cdot\mathbf{R}\mathbf{r} = \mathbf{k}'\cdot\mathbf{r}$ and $\mathbf{G}\cdot\mathbf{R}\mathbf{r} = \mathbf{R}^{-1}\mathbf{G}\cdot\mathbf{r}$, we arrive at the following relationship for the Fourier coefficients:

$$c_{\mathbf{G}}^{(n,\mathbf{k})} = c_{\mathbf{G}'}^{(n,\mathbf{k}')}\mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{t}}, \tag{B.13}$$

where $\mathbf{G}' = \mathbf{R}^{-1}\mathbf{G}$, and we neglect the constant phase shift $\mathrm{e}^{\mathrm{i}\mathbf{R}\mathbf{k}'\cdot\mathbf{t}}$. We note that $|\mathbf{k}' + \mathbf{G}'| = |\mathbf{R}^{-1}(\mathbf{k} + \mathbf{G})| = |\mathbf{k} + \mathbf{G}|$; therefore, since the plane-wave

basis set used to expand the Block function $|u_{n,\mathbf{k}}\rangle$ is defined in CASTEP by a spherically symmetric cut-off in reciprocal space centred at $-\mathbf{k}$, the transformed Fourier component $\mathbf{G}'$ will necessarily always be contained in the truncated basis set for k-point $\mathbf{k}'$ (this is not the case in, e.g., ONETEP, as the use of a psinc basis is equivalent to a *cube* of k-points centred on the origin in reciprocal space).

Finally, we note that is it sometimes necessary to translate a Bloch function by a reciprocal lattice vector $\mathbf{G}'$ (e.g., when considering the neighbours of a k-point at the Brillouin zone edge); in this case, from Bloch's theorem, $\phi_{n,\mathbf{k}}(\mathbf{r}) = \phi_{n,\mathbf{k}+\mathbf{G}'}(\mathbf{r})$, leading to a shift in the Fourier coefficients:

$$c_{\mathbf{G}}^{(n,\mathbf{k})} = c_{\mathbf{G}-\mathbf{G}'}^{(n,\mathbf{k}+\mathbf{G}')}. \tag{B.14}$$

Using the above relationships, the interface code calculates the Fourier coefficients for the Bloch functions at all k-points in the Brillouin zone; to reduce the memory cost, each Bloch function is written to a temporary scratch file instead of being held in memory. The scratch files are then read back in pairs for the calculation of $M_{mn}^{(\mathbf{k},\mathbf{b})}$, or singly for that of $A_{mn}^{(\mathbf{k})}$. This also allows for a simple parallelisation strategy for shared-disk architectures: each matrix element is calculated independently by a single process, and the necessary data for the operation is retrieved from the scratch files (that are accessible to all processes).

The data outputted by the interface code is in the required format for use in a WANNIER90 calculation; it is divided in the following files:

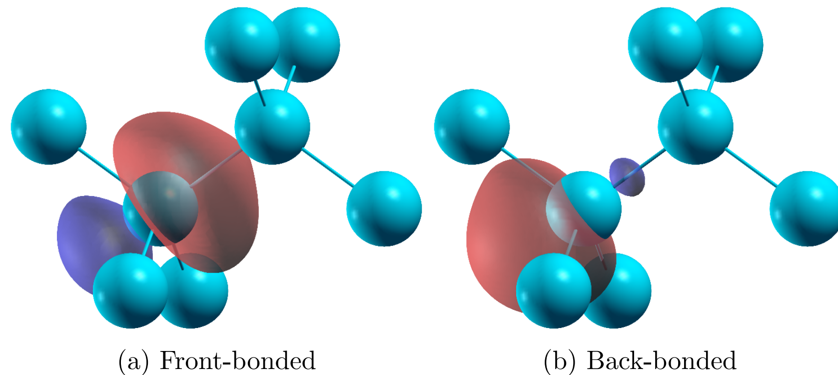
- `<seedname>.mmn`–the $M_{mn}^{(\mathbf{k},\mathbf{b})}$ matrix;
- `<seedname>.amn`–the $A_{mn}^{(\mathbf{k})}$ matrix;
- `<seedname>.eig`–the Kohn-Sham eigenvalues $\{\varepsilon_{n,\mathbf{k}}\}$ at all k-points in the Brillouin zone;
- `UNKp.s`–the Bloch functions at k-point `p` and spin `s` on a real-space grid (these are obtained by fast Fourier transform of the Fourier coefficients).

Appendix C

Wannier representations of bulk silicon

Several previous studies have considered the construction of Wannier functions from the Bloch eigenstates of bulk silicon and similar tetrahedral semiconductors, some focusing on the valence manifold [72, 347–351] (four bands in the primitive cell) and others also on the possibility of including the lowest four bands in the conduction manifold [74,352,353]. In particular, Souza *et al.* [74] have shown how their disentanglement procedure allows us to extract these low-lying bands from the conduction manifold in the absence of a band gap between them and higher states. The resulting four valence bands and four conduction bands can either be treated separately or combined; in the former case, the maximally-localised Wannier functions obtained are bonding and antibonding states centred on the Si–Si bond, while in the latter case they are sp^3-like orbitals, tetrahedrally arranged around each atom and with their major lobe pointing along the bond towards the neighbouring atom (this shall be referred to as the front-bonded case). This corresponds to the chemically-intuitive picture that is employed in most tight-binding models of silicon, in which the valence $3s$ and $3p$ states hybridise to form the characteristic sp^3 orbitals that are a feature of tetrahedral systems.

However, our calculations show that these conventional front-bonded orbitals are not the most localised Wannier representation of the subspace, and,



(a) Front-bonded (b) Back-bonded

Figure C.1: Contour-surface plots showing the two different Wannier representations of bulk silicon. The major lobe (taken to be positive) is shown in red, and the minor lobe (taken to be negative) in blue.

hence, are not strictly the MLWFs of the system. Indeed, we find an alternative Wannier representation that is more localised, corresponding to a set of orbitals also with tetrahedral sp^3 symmetry, but with their *minor* lobe pointing along the bond and their major lobe pointing in the opposite direction; hence, this shall be referred to as the back-bonded case. The front- and back-bonded Wannier functions are illustrated in Fig. C.1.

Both representations can be recovered from a wannierisation of the same bands, as they correspond to different local minima of the spread functional $\Omega\,[U]$ defined in Eq. 3.44. The algorithm employed to minimise $\Omega\,[U]$ is generally found to be robust and insensitive to the set of trial orbitals that are provided at the start of the procedure to generate an initial guess for the $U_{mn}^{(\mathbf{k})}$ matrix[1]; however, in some systems local minima can be found, depending on the initial trial orbitals provided. As discussed in Sec. 3.3.2, the Wannier functions at the global minimum (i.e, the MLWFs) should be real, and so any minimum that features significant imaginary components to the Wannier functions can be identified as being local. However, it is sometimes possible for local minima to produce Wannier functions that are real, making it difficult to identify the true global minimum. This is the case for the valence+conduction

[1]For the case of non-isolated bands, the trial orbitals are also used as an initial guess for the disentanglement routine that extract the optimally-connected subspace.

manifold of bulk silicon, since both the front- and back-bonded orbitals are real.

For our calculations, we employ the QUANTUM ESPRESSO code [354] (version 4.1) for the initial calculation of the band structure from LDA-DFT. We perform the calculations on a 64-atom cell of silicon with a $2 \times 2 \times 2$ MP sampling of the Brillouin zone (shifted so as to include Γ). We use an ultrasoft pseudopotential for silicon, and a plane-wave energy cut-off of 340 eV. The disentanglement and wannierisation of the Bloch states is then performed with the WANNIER90 code (version 1.2).

The initial trial orbitals used for these procedures are sp^3-like Gaussian orbitals; in order to obtain the front- and back-bonded Wannier functions, these trial orbitals are oriented with their major lobe pointing either along the bond or in the opposite direction, respectively. We note that it is also possible to obtain a mixed representation, in which some atoms feature front-bonded Wannier functions and others back-bonded ones, depending solely on the orientation of the initial trial orbitals placed on each atom. However, due to the orthogonality constraints, the front- and back-bonded Wannier functions in the mixed representations are slightly modified depending on the orientation of those on neighbouring atoms. In the following discussion, therefore, we restrict ourselves to the case of a purely front- or back-bonded representation of the system, in which all the Wannier functions are identical.

Our calculations show the front-bonded Wannier functions to have a spread of 2.54 Å^2 each, and the back-bonded ones to have a spread of 2.10 Å^2 each; therefore, the back-bonded representation, although counter-intuitive from a chemical perspective, is indeed the more localised of the two. It seems unlikely that other, even more localised representations exist, since the use of randomised initial trial orbitals for the wannierisation routine is observed to always result in either the front- or back-bonded representation, or a mixture of the two. Therefore, it is probable that the back-bonded Wannier functions correspond to the global minimum of the spread functional, and so are the maximally-localised representation of the subspace in question.

We have investigated the use of front- or back-bonded Wannier functions as a minimal tight-binding basis for the system. It is important to note that,

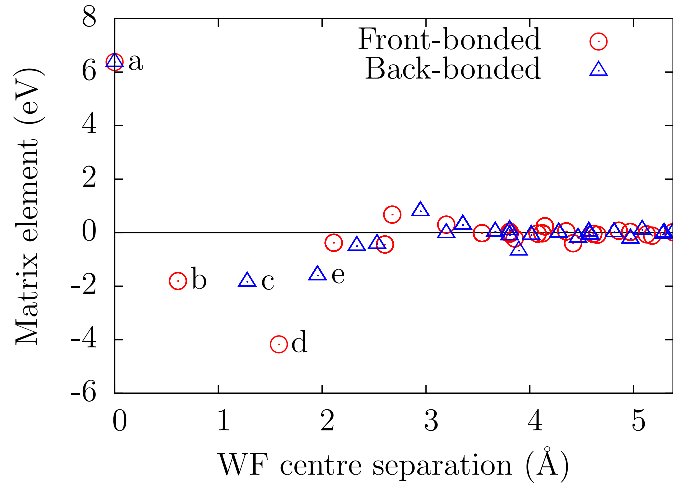


Figure C.2: Tight-binding matrix elements for the two Wannier representations. The labels are explained in the text.

since both sets of functions possess the same symmetry as sp^3 orbitals, they are identical in terms of the symmetries of the resulting tight-binding Hamiltonian; however, the magnitude of corresponding matrix elements (MEs) is very different. Fig. C.2 shows the value of the Hamiltonian MEs for both representations as a function of the separation between Wannier function centres (since all Wannier functions in a given representation are identical, only one reference function needs to be taken into consideration). As can be seen, both the values of the MEs and the separation distances are different, since the Wannier function centres are positioned differently in the two representations (as might be expected, they are situated along the bond for the front-bonded case and behind the bond for the back-bonded case). The most significant contributions have been labelled a–e, and correspond to the following MEs:

- a–the diagonal MEs H_{nn} (these have an identical value in both representations, since $\mathrm{Tr}\,(\mathbf{H})$ must be the same in both cases as they describe the same subspace);

- b–MEs between Wannier functions belonging to the same atom (front-

bonded representation);

- c–MEs between Wannier functions belonging to the same atom (back-bonded representation);
- d–MEs between front-bonded Wannier functions on the same bond (i.e., with overlapping major lobes);
- e–MEs between a back-bonded Wannier function and *six* other such Wannier functions, two each on the three neighbouring atoms that are *not* situated along the bond of the reference Wannier function (i.e., the bond on which the minor lobe is found).

As can be seen from the above description, b and c are the equivalent MEs in the two representations, but d and e are not; their equivalent elements are less significant in the other representation, and the centres of the Wannier functions involved are further apart.

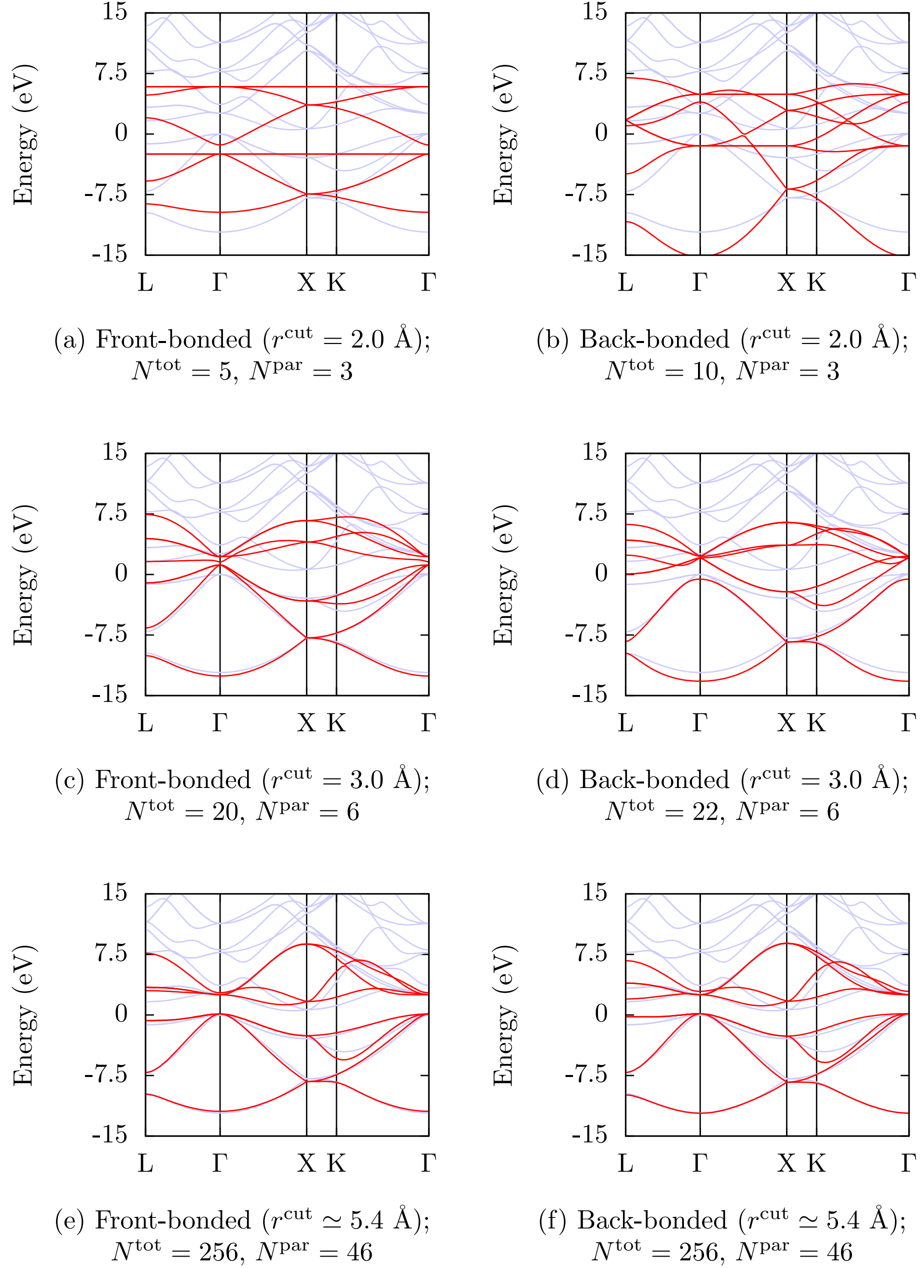
These matrix elements, calculated for the 64-atom supercell (in order to achieve a sufficiently small overlap between Wannier functions in neighbouring cells), have been used to construct a tight-binding Hamiltonian for the primitive 2-atom unit cell. The silicon band structure obtained from the tight-binding Hamiltonian in the front- and back-Wannier bases can then be compared to the full *ab initio* band structure for the primitive cell, as shown in Fig. C.3. In particular, we investigate the effect of neglecting MEs between Wannier functions whose separation is greater than a cut-off distance r^{cut}. As expected, the front- and back-Wannier bases converge to approximately equivalent band structures as r^{cut} is increased to its maximum value (given by half the lattice parameter of the 64-atom supercell), which, in turn, agree closely with the *ab initio* band structure.

However, it is interesting to note that the front-bonded Wannier functions provide a much more reasonable description of the system for small values of r^{cut}. For $r^{\mathrm{cut}} = 2.0$ Å, only the MEs labelled a, b and d are used for the front-bonded case, and a, c and e for the back-bonded case. The Hamiltonian for the front-bonded representation is now equivalent to the simple Weaire-Thorpe model for tetrahedrally-bonded solids [355, 356], and so, as expected,

provides a qualitatively accurate description of the valence bands and the band gap. In contrast, the Hamiltonian for the back-bonded representation shows no band gap and no clear distinction between the valence and conduction bands, despite including more non-zero elements. Increasing r^{cut} to 3.0 Å, there is a noticeable improvement in the shape of the bands for both representations; however, for the back-bonded case the highest two valence bands are separated from the lowest two and entangled with the conduction bands, and so the band gap is still absent.

We conclude, therefore, that the back-bonded Wannier representation, although more localised in terms of the spread functional, is clearly less well-suited as a tight-binding basis for silicon than the chemically-intuitive front-bonded representation. The reason for this can be understood in terms of the increased isolation of the back-bonded Wannier functions: in fact, Fig. C.2 shows that their centres are further apart than those of the front-bonded Wannier functions; hence, they interact only relatively weakly with a large number of other functions, with no single interaction being dominant. Instead, the front-bonded Wannier functions have a clear single dominant interaction, that between the two functions on the same bond, accounting for most of the band energy[2]. We note that a similar study [357] of GaAs, GaN, and InN, in which five bands were wannierised to obtain an sp^3s^*-like basis, also reports that both chemically-intuitive and non-chemically-intuitive Wannier representations can be obtained; the most efficient representation for designing a tight-binding basis was found to be system-dependent.

[2]Indeed, an even simpler model of the system than the Weaire-Thorpe model is the molecular (or bond order) model, in which only this interaction is considered, resulting in $2N$ bonding states and $2N$ antibonding states for an N-atom system, and no interaction between bonds.



(a) Front-bonded ($r^{\text{cut}} = 2.0$ Å); $N^{\text{tot}} = 5$, $N^{\text{par}} = 3$

(b) Back-bonded ($r^{\text{cut}} = 2.0$ Å); $N^{\text{tot}} = 10$, $N^{\text{par}} = 3$

(c) Front-bonded ($r^{\text{cut}} = 3.0$ Å); $N^{\text{tot}} = 20$, $N^{\text{par}} = 6$

(d) Back-bonded ($r^{\text{cut}} = 3.0$ Å); $N^{\text{tot}} = 22$, $N^{\text{par}} = 6$

(e) Front-bonded ($r^{\text{cut}} \simeq 5.4$ Å); $N^{\text{tot}} = 256$, $N^{\text{par}} = 46$

(f) Back-bonded ($r^{\text{cut}} \simeq 5.4$ Å); $N^{\text{tot}} = 256$, $N^{\text{par}} = 46$

Figure C.3: Silicon band structures calculated with the front- and back-bonded Wannier bases. MEs between Wannier functions whose separation is greater than r^{cut} are set to zero. The band structure obtained from the Wannier basis is shown in red, and the *ab initio* band structure calculated directly from the DFT code is shown in light blue. N^{tot} is the total number of MEs per atom in the tight-binding Hamiltonian, and N^{par} is the number of independent MEs.

Appendix D

The extended LCAO vacancy model

We have shown, both for the silicon vacancy (Chapter 5) and the gold substitutional centre (Chapter 6), qualitative differences in the finite size convergence behaviour of the system using Γ-point only and dense k-point sampling. In particular, when using a dense sampling, the correct relaxation of the defect centre is only recovered for large supercells containing at least 256 atoms; smaller supercells exhibit either no symmetry change upon relaxation (for the vacancy), or a variety of spurious symmetries (for the gold centre). However, there is evidence that Γ-point only calculations recover the correct symmetry even for small supercells [202].

In addition, our results for the gold centre in the 256-atom BCC supercell show a greater decrease in the formation energy due to relaxation ΔE_f when using a Γ-point sampling instead of a $k_{\mathrm{MP}} = 3$ grid for all charge states apart from the neutral one. This difference ranges from -0.02 eV for Au^0 to $+0.07$ eV for Au^{1-}. This is consistent with the idea that the dense sampling 'resists' the relaxation, leading to smaller changes in bond lengths, a smaller energy gain, and no change of symmetry for small supercells.

We use a simple LCAO model to explain these finite size effects; this we have developed as an extension to Watkins' model of the vacancy (described in Sec. 5.2). For a perfectly isolated impurity which does not interact with

the host lattice, the four dangling bonds combine with each other to form the tight-binding Hamiltonian

$$H = \begin{pmatrix} -a & -b & -b & -b \\ -b & -a & -b & -b \\ -b & -b & -a & -b \\ -b & -b & -b & -a \end{pmatrix}, \tag{D.1}$$

where a and b are the absolute values of the on-site and two-centre matrix elements. The eigenvalues are: $-a-3b$, corresponding to the nodeless combination buried within the silicon valence band, and the triply degenerate level $-a+b$, corresponding to the three combinations with a nodal plane which lie in the band gap.

There are two additions that we make to this Hamiltonian. Firstly, we add a term $\Delta = \alpha Q$ to four off-diagonal elements, which breaks the tetrahedral symmetry and is proportional to a hypothetical tetragonal distortion Q of the system. Q relates to the bond length between a pair of dimerised neighbours of the defect centre, where $Q = 0$ corresponds to the undistorted bond length, and $Q = 1$ to the limit of zero bond length. The Hamiltonian now becomes:

$$H^{(\Delta)} = \begin{pmatrix} -a & -b-\Delta & -b & -b \\ -b-\Delta & -a & -b & -b \\ -b & -b & -a & -b-\Delta \\ -b & -b & -b-\Delta & -a \end{pmatrix}. \tag{D.2}$$

This arrangement pairs up the first and second orbitals, and the third and fourth ones. The eigenvalues are now: $-a-3b-\Delta$ for the buried level, and $-a+b\pm\Delta$ for the defect levels in the gap, where the energy of one level is lowered by Δ and the other two are raised by Δ. This distortion is appropriate to describe v^0 and Au^{1+}, since only one of the three levels is occupied (and the resulting symmetry of the defect centre is D_{2d}); therefore, the total electronic energy of the system is $-2\left(a+b+\Delta\right)$ (neglecting spin). However, the dimerised ions will also feel a repulsion $f/\left(1-Q\right)$. The total energy is minimised at the equilibrium distortion $Q^{\mathrm{eq}} = 1 - \sqrt{f/\left(2\alpha\right)}$; the

corresponding decrease in formation energy from the unrelaxed to the relaxed lattice is $\Delta E_f = \left(\sqrt{f} - \sqrt{2\alpha}\right)^2$.

We can estimate Q^{eq} from the change of bond lengths for Au^{1+} (v^0 is not reliable for this purpose, as there is a significant additional inwards breathing relaxation). From our calculations, we find that $Q^{\mathrm{eq}} \simeq 0.026$, and, hence, that $f \simeq 1.9\alpha$.

This model is valid for the isolated defect, and therefore gives the 'correct' behaviour of the system with no finite size effects. The second addition to the basic Hamiltonian takes into account the interaction of the four orbitals between different images of the defect centre in a supercell calculation. We ensure that each orbital interacts only with the nearest orbitals from the defect images, which will depend on their relative orientations, as it is important to preserve the overall tetrahedral symmetry of the underlying lattice and the cubic symmetry of the supercell. In the simplest approximation each orbital interacts with six extra orbitals in total, two each from three neighbouring defect images, with a small additional term c in the Hamiltonian.

This new Hamiltonian $H^{(\Delta,c)}$, which approximates the finite size effects of the supercell method, can be solved numerically, or analytically for certain high-symmetry k-points. It correctly predicts that at Γ and P the three defect levels are degenerate for the undistorted lattice, as is seen in the DFT calculations. The eigenvalues are the same as for the standard Watkins model Hamiltonian H, but with a shift of $+2c$ at Γ and $-2c$ at P for the triply degenerate level, and $-6c$ at Γ and $+6c$ at P for the buried level. We note that this does not agree with our observed effect for the undistorted neutral vacancy that the Γ-point quickly converges with system size to the position of the isolated defect level; however, it correctly identifies Γ and P as corresponding (approximately, for the DFT results) to the upper and lower limits of the extent of the defect levels' dispersion.

Fig. D.1 shows the band structure for a BCC supercell calculated with $H^{(\Delta,c)}$; this can be compared with the DFT band structure for the unrelaxed vacancy in Fig. 5.7a. As can be seen, the degeneracy of the three defect levels is correct at all high-symmetry points in the Brillouin zone.

We can now calculate the equilibrium distortion Q^{eq} and corresponding

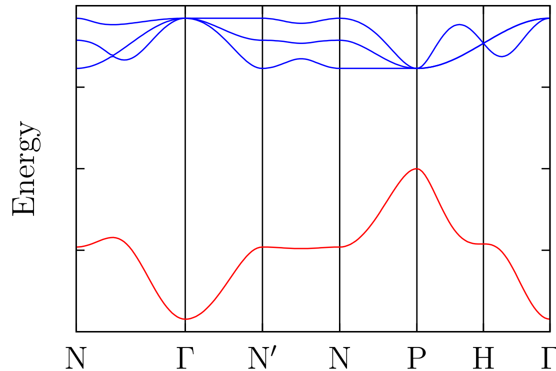


Figure D.1: Band structure for the BCC supercell calculated with the extended LCAO vacancy model. The three defect levels in the band gap are shown in blue, and the buried level is shown in red.

decrease in formation energy from the undistorted to the distorted lattice ΔE_f for different k-points. For the case of the Γ-point the finite size effects cancel, except for a constant shift in the total energy of $-4c$; therefore, both Q^{eq} and ΔE_f are identical to the isolated defect centre. The model, therefore, provides a simple explanation as to why Γ-point calculations exhibit the correct relaxation even for small supercells.

However, this cancellation does not occur for non-high-symmetry k-points. As a comparison, therefore, we numerically investigate the behaviour of the system using a $k^{\mathrm{MP}} = 3$ mesh and setting $f = 1.9\alpha$ (as estimated from the Au^{1+} DFT calculation). This leaves only two independent parameters, the ratios c/b and α/b. c is expected to be significantly smaller than b, and clearly $c \to 0$ as the supercell lattice parameter $L \to \infty$.

It is found that for this extended k-point sampling scheme the model features a discontinuity in Q^{eq}: below a critical value of c, the correct distortion is obtained, but above it the system remains completely undistorted. This is consistent with our observation of an abrupt change in symmetry for v^0, which remains $\mathrm{T_d}$ for supercells smaller than 256 atoms and switches to $\mathrm{D_{2d}}$

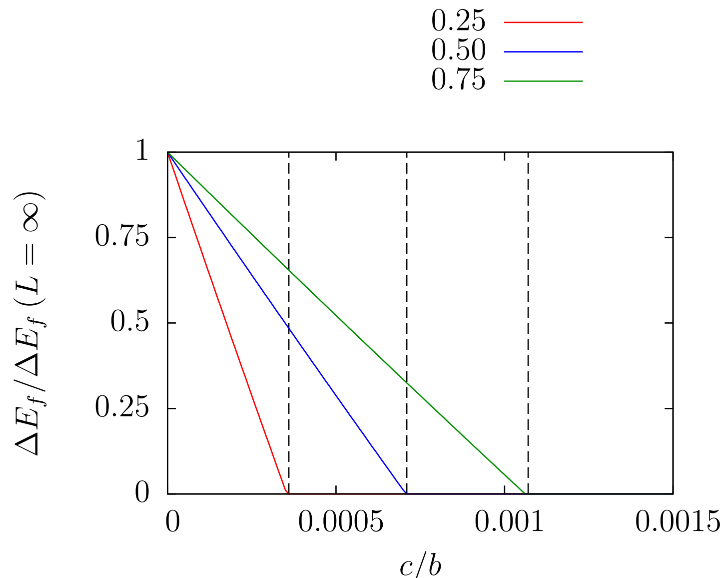


Figure D.2: Effect of relaxation for the extended LCAO vacancy model. The graph shows the decrease in formation energy upon relaxation ΔE_f as a fraction of the correct value (at $L = \infty$). The key gives the different values of α/b used in the calculation. The dashed vertical lines show the critical value of c for each calculation below which the correct distortion is recovered.

for $\geq$256 atoms. Furthermore, ΔE_f is zero in the undistorted region, and increases approximately linearly in the distorted region to the correct value at $c = 0$ ($L = \infty$). This suggests that small supercells with dense k-point sampling will significantly underestimate the gain in energy from Jahn-Teller distortion, even if they show the correct change in symmetry. Fig. D.2 shows ΔE_f as a function of c for various values of α.

In conclusion, we have shown that an extended Watkins LCAO model can qualitatively explain the dispersion of the defect levels in the band gap for Watkins defects, as well as the differences in relaxation for defect supercells when using Γ-point only and dense k-point sampling schemes. Specifically, it confirms that Γ-point only sampling will obtain the correct relaxation for all supercell sizes, while dense k-point sampling will result in no distortion for

small supercells, and an underestimation of the energy gain from relaxation even for larger supercells which exhibit the correct distortion, as observed in our DFT calculations.

Unfortunately, this simple model does not explain the fast finite size convergence of the stable charge state transition levels when using a Γ-point sampling (observed in our DFT calculations); instead, it predicts that the defect levels should converge towards the middle of the band as the dispersion due to the interaction between supercells (given by c) decreases, as is generally the case in such models. We therefore suggest that this effect is caused by a modification of the matrix elements between orbitals at the same defect site (given by a and b) as the system size is increased; in particular, a decrease in the magnitude of the on-site matrix elements with increasing L would rigidly shift the eigenvalues upwards, and, hence, could lead to the apparent convergence of the defect levels towards the upper portion of the band, close to the Γ-point level. Hence, the correct finite size convergence properties cannot emerge naturally from the tight-binding Hamiltonian; rather, the modification of the matrix elements must either be estimated from a more sophisticated model, or calculated directly from the data obtained in the DFT simulations. In order to verify this hypothesis, therefore, it would be necessary to examine the change in magnitude of the on-site matrix elements for the defect MLWFs obtained from a range of supercell sizes.

Bibliography

[1] R. P. Feynman, R. B. Leighton, and M. Sands, in *The Feynman Lectures on Physics* (Addison-Wesley, Reading, Mass., 1963), Vol. 1, pp. 1–2.

[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[3] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[4] M. C. Payne *et al.*, Rev. Mod. Phys. **64**, 1045 (1992).

[5] J. Hafner, C. Wolverton, and G. Ceder, MRS Bull. **31**, 659 (2006).

[6] N. Marzari, MRS Bull. **31**, 681 (2006).

[7] I. Vasiliev, S. Öğüt, and J. R. Chelikowsky, Phys. Rev. Lett. **86**, 1813 (2001).

[8] A. J. Williamson *et al.*, Phys. Rev. Lett. **89**, 196803 (2002).

[9] L. D. Negro *et al.*, Appl. Phys. Lett. **88**, 183103 (2006).

[10] A. Franceschetti, L. W. Wang, G. Bester, and A. Zunger, Nano Lett. **6**, 1069 (2006).

[11] B. Delley and E. F. Steigmeier, Appl. Phys. Lett. **67**, 2370 (1995).

[12] X. Zhao, C. M. Wei, L. Yang, and M. Y. Chou, Phys. Rev. Lett. **92**, 236805 (2004).

[13] Z. Wu, J. B. Neaton, and J. C. Grossman, Phys. Rev. Lett. **100**, 246804 (2008).

[14] C. D. Spataru, S. Ismail-Beigi1, L. X. Benedict, and S. G. Louie, Phys. Rev. Lett. **92**, 077402 (2004).

[15] M. Lazzeri *et al.*, Phys. Rev. Lett. **95**, 236802 (2005).

[16] M. D. Ventra, S. T. Pantelides, and N. D. Lang, Phys. Rev. Lett. **84**, 979 (2000).

[17] D. A. Scherlis and N. Marzari, J. Phys. Chem. B **108**, 17791 (2004).

[18] J. C. Slater, Phys. Rev. **34**, 1293 (1929).

[19] C. D. Sherrill and H. F. S. III, Adv. Quantum Chem. **34**, 143 (1999).

[20] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[21] R. J. Bartlett, J. Phys. Chem. **93**, 1697 (1989).

[22] M. Born and J. R. Oppenheimer, Ann. d. Physik **84**, 457 (1927).

[23] J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules* (Cambridge University Press, Cambridge, 2006).

[24] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, San Diego, 1996).

[25] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[26] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, Phys. Rev. B **45**, 1538 (1992).

[27] M. J. Gillan, Phys. Rev. Lett. **58**, 563 (1987).

[28] X.-Z. Li, M. I. J. Probert, A. Alavi, and A. Michaelides, Phys. Rev. Lett. **104**, 066102 (2010).

[29] D. R. Hartree, Proc. Cambridge Phil. Soc. **24**, 89, 111, 426 (1928).

[30] V. Fock, Z. Phys. **61**, 126 (1930).

[31] C. W. Bauschlicher Jr., Chem. Phys. Lett. **246**, 40 (1995).

[32] A. Svane, Phys. Rev. B **35**, 5496 (1987).

[33] G. F. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, 2005).

[34] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2004).

[35] M. Levy, Phys. Rev. A **26**, 1200 (1982).

[36] E. Lieb, Int. J. Quant. Chem. **24**, 243 (1983).

[37] H. Englisch and R. Englisch, Phys. Status Solidi (b) **123**, 711 (1984).

[38] H. Englisch and R. Englisch, Phys. Status Solidi (b) **124**, 373 (1984).

[39] M. Levy, Proc. Natl. Acad. Sci. **76**, 6062 (1979).

[40] T. L. Gilbert, Phys. Rev. B **12**, 2111 (1975).

[41] L. H. Thomas, Proc. Cambridge Phil. Roy. Soc. **23**, 542 (1927).

[42] E. Fermi, Rend. Accad. Naz. Lincei **6**, 602 (1927).

[43] P. A. M. Dirac, Proc. Cambridge Phil. Roy. Soc. **26**, 376 (1930).

[44] M. Gell-Mann and K. A. Brueckner, Phys. Rev. **106**, 364 (1957).

[45] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[46] W. M. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

[47] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[48] U. von Barth and L. Hedin, J. Phys. C **5**, 1629 (1972).

[49] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

[50] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[51] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[52] N. Argaman and G. Makov, Am. J. Phys. **68**, 69 (2000).

[53] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[54] V. I. Anisimov *et al.*, Phys. Rev. B **48**, 16929 (1993).

[55] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College Publishing, Philadelphia, 1976).

[56] `http://opium.sourceforge.net` (2009).

[57] J. C. Phillips and L. Kleinman, Phys. Rev. **116**, 287 (1959).

[58] C. Herring, Phys. Rev. **57**, 1169 (1940).

[59] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[60] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[61] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[62] T. L. Beck, Rev. Mod. Phys. **72**, 1041 (2000).

[63] J. R. Chelikowsky, N. Troullier, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994).

[64] S. F. Boys, Proc. R. Soc. Lond. A **200**, 542 (1950).

[65] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[66] M. Sprik and M. L. Klein, J. Chem. Phys. **91**, 5665 (1989).

[67] C. M. Goringe, D. R. Bowler, and E. Hernández, Rep. Prog. Phys. **60**, 1447 (1997).

[68] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

[69] G. H. Wannier, Phys. Rev. **52**, 191 (1937).

[70] W. Kohn, Phys. Rev. **115**, 809 (1959).

[71] J. D. Cloizeaux, Phys. Rev. **129**, 554 (1963).

[72] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[73] I. Souza, T. Wilkens, and R. M. Martin, Phys. Rev. B **62**, 1666 (2000).

[74] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).

[75] C. Brouder *et al.*, Phys. Rev. Lett. **98**, 046402 (2007).

[76] S. F. Boys, Rev. Mod. Phys. **32**, 296 (1960).

[77] N. Marzari and D. Vanderbilt, AIP Conf. Proc. **436**, 146 (1998).

[78] M. Fornari, N. Marzari, M. Peressi, and A. Baldereschi, Comput. Mater. Sci. **20**, 337 (2001).

[79] R. Martonak, C. Molteni, and M. Parrinello, Comput. Mater. Sci. **20**, 293 (2001).

[80] S. Goedecker, T. Deutsch, and L. Billard, Phys. Rev. Lett. **88**, 235501 (2002).

[81] X. Wang, J. R. Yates, I. Souza, and D. Vanderbilt, Phys. Rev. B **74**, 195118 (2006).

[82] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Phys. Rev. B **75**, 195121 (2007).

[83] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[84] D. Vanderbilt and R. D. King-Smith, Phys. Rev. B **48**, 4442 (1993).

[85] R. Resta, Rev. Mod. Phys. **66**, 899 (1994).

[86] W. Kohn, Phys. Rev. Lett. **76**, 3168 (2006).

[87] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).

[88] W. Kohn, Int. J. Quantum Chem. **56**, 229 (1995).

[89] L. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001).

[90] S. Ismail-Beigi and T. A. Arias, Phys. Rev. Lett. **82**, 2127 (1999).

[91] S. Goedecker, Phys. Rev. B **58**, 3501 (1998).

[92] D. R. Bowler and T. Miyazaki, Rep. Prog. Phys. **75**, 036503 (2012).

[93] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005).

[94] C. K. Gan, P. D. Haynes, and M. C. Payne, Comput. Phys. Commun. **134**, 33 (2001).

[95] P. D. Haynes, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, Phys. Status Solidi (b) **243**, 2489 (2006).

[96] P. D. Haynes, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, Chem. Phys. Lett. **422**, 345 (2006).

[97] H. B. Jansen and P. Ros, Chem. Phys. Lett. **3**, 140 (1969).

[98] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).

[99] A. A. Mostofi, P. D. Haynes, C.-K. Skylaris, and M. C. Payne, J. Chem. Phys. **119**, 8842 (2003).

[100] A. A. Mostofi, C.-K. Skylaris, P. D. Haynes, and M. C. Payne, Comput. Phys. Commun. **147**, 788 (2002).

[101] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Phys.: Condens. Matter **20**, 064209 (2008).

[102] N. D. M. Hine *et al.*, Comput. Phys. Commun. **180**, 1041 (2009).

[103] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1953).

[104] J. S. Blakemore, *Solid State Physics*, 2nd ed. ed. (Cambridge University Press, Cambridge, 1985).

[105] F. A. Kröger and H. J. Vink, *Solid State Physics* (Academic Press, New York, 1956), Vol. 3.

[106] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[107] F. Morehead, in *Defects in Electronic Materials*, edited by M. Stavola, S. J. Pearton, and G. Davies (Materials Research Society, Pittsburgh, 1987), p. 99.

[108] S. Dannefaer, P. Mascher, and D. Kerr, Phys. Rev. Lett. **56**, 2195 (1986).

[109] P. M. Fahey, P. B. Griffin, and J. D. Plummer, Rev. Mod. Phys. **61**, 289 (1989).

[110] W.-K. Leung *et al.*, Phys. Rev. Lett. **83**, 2351 (1999).

[111] H. J. Queisser and E. E. Haller, Science **281**, 945 (1998).

[112] W. M. Bullis, Solid State Electron. **9**, 143 (1966).

[113] X. Zhao, C. M. Wei, L. Yang, and M. Y. Chou, Phys. Rev. Lett. **92**, 236805 (2004).

[114] S. Öğüt, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. Lett. **79**, 1770 (1997).

[115] S. C. Erwin *et al.*, Nature **436**, 91 (2005).

[116] G. M. Dalpian and J. R. Chelikowsky, Phys. Rev. Lett. **96**, 226802 (2006).

[117] A. R. Williams, P. J. Feibelman, and N. D. Lang, Phys. Rev. B **26**, 5433 (1982).

[118] G. A. Baraff and M. Schluter, Phys. Rev. B **28**, 2296 (1983).

[119] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. B **21**, 3545 (1980).

[120] R. Car, P. J. Kelly, A. Oshiyama, and S. T. Pantelides, Phys. Rev. Lett. **52**, 1814 (1984).

[121] I. Gorczyca, A. Svane, and N. E. Christensen, Solid State Commun. **101**, 747 (1997).

[122] D. W. Jenkins and J. D. Dow, Phys. Rev. B **39**, 3317 (1989).

[123] D. W. Jenkins, J. D. Dow, and M. H. Tsai, J. Appl. Phys. **72**, 4130 (1992).

[124] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).

[125] M. Elstner *et al.*, Phys. Rev. B **58**, 7260 (1998).

[126] C. Stampfl and C. G. Van de Walle, Phys. Rev. B **65**, 155212 (2002).

[127] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008).

[128] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[129] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

[130] L. J. Sham and M. Schlüter, Phys. Rev. B **32**, 3883 (1985).

[131] E. R. Batista *et al.*, Phys. Rev. B **74**, 121102 (2006).

[132] R. Q. Hood, P. R. C. Kent, R. J. Needs, and P. R. Briddon, Phys. Rev. Lett. **91**, 076403 (2003).

[133] D. Alfè and M. J. Gillan, Phys. Rev. B **71**, 220101 (2005).

[134] A. Badinski, P. D. Haynes, J. R. Trail, and R. J. Needs, J. Phys.: Condens. Matter **22**, 074202 (2010).

[135] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).

[136] M. Hagen and M. W. Finnis, Mater. Sci. Forum **245**, 207 (1996).

[137] M. Hagen and M. W. Finnis, Phil. Mag. A **77**, 447 (1998).

[138] J. Neugebauer and C. G. Van de Walle, in *Festkörperprobleme/Advances in Solid State Physics*, edited by R. Helbig (Vieweg, Braunschweig/Wiesbaden, 1996), Vol. 35, p. 25.

[139] G. A. Baraff, E. O. Kane, and M. Schlüter, Phys. Rev. B **21**, 5662 (1980).

[140] S. Limpijumnong and C. G. Van de Walle, Phys. Rev. B **69**, 035207 (2004).

[141] D. V. Lang, J. Appl. Phys. **45**, 3023 (1974).

[142] J. Franck and E. G. Dymond, Trans. Faraday Soc. **21**, 536 (1925).

[143] E. U. Condon, Phys. Rev. **28**, 1182 (1926).

[144] E. U. Condon, Phys. Rev. **32**, 858 (1928).

[145] M. Lannoo, M. Schlüter, and L. J. Sham, Phys. Rev. B **32**, 3890 (1985).

[146] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[147] J. F. Janak, Phys. Rev. B **18**, 7165 (1978).

[148] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).

[149] W. Yang, Y. Zhang, and P. W. Ayers, Phys. Rev. Lett. **84**, 5172 (2000).

[150] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Phys. Rev. B **77**, 115123 (2008).

[151] P. Mori-Sánchez, A. J. Cohen, and W. Yang, Phys. Rev. Lett. **100**, 146401 (2008).

[152] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Science **321**, 792 (2008).

[153] P. Mori-Sánchez, A. J. Cohen, and W. Yang, J. Chem. Phys. **125**, 201102 (2006).

[154] C. Persson, Y.-J. Zhao, S. Lany, and A. Zunger, Phys. Rev. B **72**, 035211 (2005).

[155] C. W. M. Castleton, A. Höglund, and S. Mirbt, Phys. Rev. B **73**, 035215 (2006).

[156] J. Lento and R. M. Nieminen, J. Phys.: Condens. Matter **15**, 4387 (2003).

[157] J. M. Knaup *et al.*, Phys. Rev. B **72**, 115323 (2005).

[158] K. Xiong, J. Robertson, M. C. Gibson, and S. J. Clark, Appl. Phys. Lett. **87**, 183505 (2005).

[159] J. L. Gavartin *et al.*, Appl. Phys. Lett. **89**, 082908 (2006).

[160] P. Broqvist and A. Pasquarello, Appl. Phys. Lett. **89**, 262904 (2006).

[161] A. Alkauskas, P. Broqvist, and A. Pasquarello, Phys. Rev. Lett. **101**, 046405 (2008).

[162] A. Stroppa and G. Kresse, Phys. Rev. B **79**, 201201 (2009).

[163] P. Broqvist, A. Alkauskas, and A. Pasquarello, Phys. Rev. B **78**, 075203 (2008).

[164] C. Stampfl *et al.*, Phys. Rev. B **61**, R7846 (2000).

[165] J. Li and S.-H. Wei, Phys. Rev. B **73**, 041201 (2006).

[166] A. Janotti and C. G. Van de Walle, Phys. Rev. B **76**, 165202 (2007).

[167] A. Janotti and C. G. Van de Walle, Appl. Phys. Lett. **92**, 032104 (2008).

[168] A. K. Singh, A. Janotti, M. Scheffler, and C. G. Van de Walle, Phys. Rev. Lett. **101**, 055502 (2008).

[169] P. Rinke, A. Janotti, M. Scheffler, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 026402 (2009).

[170] F. Bruneval and G. Roma, Phys. Rev. B **83**, 144116 (2011).

[171] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[172] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[173] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).

[174] A. Seidl *et al.*, Phys. Rev. B **53**, 3764 (1996).

[175] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).

[176] L. Hedin, Phys. Rev. **139**, A796 (1965).

[177] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (1999).

[178] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[179] A. Resende, R. Jones, S. Öberg, and P. R. Briddon, Phys. Rev. Lett. **82**, 2111 (1999).

[180] A. Carvalho *et al.*, Phys. Rev. B **75**, 115206 (2007).

[181] J. Lento, J.-L. Mozos, and R. M. Nieminen, J. Phys.: Condens. Matter **14**, 2637 (2002).

[182] U. Gösele, A. Plössl, and T. Y. Tan, in *Process Physics and Modelling in Semiconductor Technology*, edited by G. R. Srinivasan, C. S. Murphy, and S. T. Dunham (Electrochemical Society, Pennington, N. J., 1996), p. 309.

[183] G. D. Watkins, in *Radiation Damage in Semiconductors* (Dunod, Paris, 1964), p. 97.

[184] G. D. Watkins, in *Deep Centres in Semiconductors*, edited by S. T. Pantelides (Gordon and Breach, New York, 1986), p. 147.

[185] G. D. Watkins, in *Defects and Their Structure in Non-metallic Solids*, edited by B. Henderson and A. E. Hughes (Plenum, New York, 1976), p. 203.

[186] M. Sprenger, S. H. Muller, and C. A. J. Ammerlaan, Physica B+C **116B**, 224 (1983).

[187] M. Sprenger, S. H. Muller, E. G. Sieverts, and C. A. J. Ammerlaan, Phys. Rev. B **35**, 1566 (1987).

[188] G. D. Watkins and J. R. Troxell, Phys. Rev. Lett. **44**, 593 (1980).

[189] G. D. Watkins, in *Lattice Defects in Semiconductors, The Institute of Physics Conference Series*, edited by F. A. Huntley (The Institute of Physics, Bristol/London, 1974), Vol. 23, p. 1.

[190] G. D. Watkins, J. R. Troxell, and A. P. Chatterjee, in *Defects and Radiation Effects in Semiconductors, The Institute of Physics Conference Series*, edited by J. H. Albany (The Institute of Physics, Bristol/London, 1978), Vol. 46, p. 16.

[191] L. C. Kimerling, in *Radiation Effects in Semiconductors, The Institute of Physics Conference Series*, edited by N. B. Urli (The Institute of Physics, Bristol/London, 1976), Vol. 31, p. 221.

[192] J. C. Brabant, M. Pugnet, and T. I. o. P. C. S. J. Barbolla, in *Radiation Effects in Semiconductors*, edited by N. B. Urli (The Institute of Physics, Bristol/London, 1976), Vol. 31, p. 200.

[193] C. E. Barnes and G. A. Samara, Mater. Res. Soc. Symp. Proc. **46**, 471 (1985).

[194] G. A. Samara, Phys. Rev. B **37**, 8523 (1988).

[195] G. A. Samara, Phys. Rev. B **39**, 11001 (1989).

[196] G. A. Samara, Phys. Rev. B **39**, 12764 (1989).

[197] J. Makinen, P. Hautojarvi, and C. Corbel, J. Phys.: Condens. Matter **4**, 5137 (1992).

[198] J. Mäkinen *et al.*, Phys. Rev. B **39**, 10162 (1989).

[199] G. D. Watkins and J. W. Corbett, Phys. Rev. **134**, A1359 (1964).

[200] J. A. Van Vechten, Phys. Rev. B **33**, 2674 (1986).

[201] S. Dannefaer, P. Mascher, and D. Kerr, Phys. Rev. Lett. **56**, 2195 (1986).

[202] M. J. Puska, S. Pöykkö, M. Pesola, and R. M. Nieminen, Phys. Rev. B **58**, 1318 (1998).

[203] M. I. J. Probert and M. C. Payne, Phys. Rev. B **67**, 075204 (2003).

[204] F. P. Larkins and A. M. Stoneham, J. Phys. C **4**, 143 (1971).

[205] F. P. Larkins and A. M. Stoneham, J. Phys. C **4**, 154 (1971).

[206] O. Gunnarsson, O. Jepsen, and O. K. Andersen, Phys. Rev. B **27**, 7144 (1983).

[207] U. Lindefelt, Phys. Rev. B **28**, 4510 (1983).

[208] M. Scheffler, J. P. Vigneron, and G. B. Bachelet, Phys. Rev. B **31**, 6541 (1985).

[209] C. Z. Wang, C. T. Chan, and K. M. Ho, Phys. Rev. Lett. **66**, 189 (1991).

[210] L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999).

[211] A. Antonelli and J. Bernholc, Phys. Rev. B **40**, 10643 (1989).

[212] P. E. Blöchl *et al.*, Phys. Rev. Lett. **70**, 2435 (1993).

[213] A. F. Wright, Phys. Rev. B **74**, 165116 (2006).

[214] M. G. Ganchenkova *et al.*, Mater. Sci. Eng. B **159-160**, 107 (2009).

[215] C. A. Coulson and M. J. Kearsley, Proc. R. Soc. Lond. A **241**, 433 (1957).

[216] R. S. Mulliken, J. Chem. Phys. **23**, 1997 (1955).

[217] H. A. Jahn and E. Teller, Proc. R. Soc. Lond. A **161**, 220 (1937).

[218] S. J. Clark *et al.*, Z. Kristallogr. **220**, 567 (2005).

[219] C. L. Fu and K. M. Ho, Phys. Rev. B **28**, 5480 (1983).

[220] C.-K. Skylaris and P. D. Haynes, J. Chem. Phys. **127**, 164712 (2007).

[221] A. A. Mostofi *et al.*, Comput. Phys. Commun. **178**, 685 (2008).

[222] S. Pöykkö, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **53**, 3813 (1996).

[223] J. W. Mayer, Gold Bull. **17**, 18 (1984).

[224] R. S. Wagner and W. C. Ellis, Appl. Phys. Lett. **4**, 89 (1964).

[225] J. E. Allen *et al.*, Nature Nanotech. **3**, 168 (2008).

[226] S. H. Oh *et al.*, Nano Lett. **8**, 1016 (2008).

[227] C. B. Collins, R. O. Carlson, and C. J. Gallagher, Phys. Rev. **105**, 1168 (1957).

[228] A. Fazzio, M. J. Caldas, and A. Zunger, Phys. Rev. B **32**, 934 (1985).

[229] J. W. Petersen and J. Nielsen, Appl. Phys. Lett. **56**, 1122 (1990).

[230] F. G. Anderson, F. S. Ham, and G. D. Watkins, Phys. Rev. B **45**, 3287 (1992).

[231] F. G. Anderson, J. Phys.: Condens. Matter **3**, 4421 (1991).

[232] G. D. Watkins, M. Kleverman, A. Thilderkvist, and H. G. Grimmeiss, Phys. Rev. Lett. **67**, 1149 (1991).

[233] N. T. Son, T. Gregorkiewicz, and C. A. J. Ammerlaan, Phys. Rev. Lett. **69**, 3185 (1992).

[234] S. D. Brotherton, P. Bradley, A. Gill, and E. R. Weber, J. Appl. Phys. **55**, 952 (1984).

[235] M. Höhne, Phys. Status Solidi (b) **99**, 651 (1980).

[236] M. Höhne, Phys. Status Solidi (b) **109**, 525 (1982).

[237] J. A. Alves *et al.*, J. Phys. C: Solid State Phys. **17**, 771 (1984).

[238] J. L. A. Alves and J. R. Leite, Phys. Rev. B **30**, 7284 (1984).

[239] K. H. Johnson, Annu. Rev. Phys. Chem. **26**, 39 (1975).

[240] G. D. Watkins, Physica B+C **117B-118B**, 9 (1983).

[241] G. D. Watkins and P. M. Williams, Phys. Rev. B **52**, 16575 (1995).

[242] H. H. Woodbury and G. W. Ludwig, Phys. Rev. **126**, 466 (1962).

[243] G. W. Ludwig and H. H. Woodbury, Phys. Rev. **113**, 1014 (1959).

[244] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).

[245] M. Leslie and M. J. Gillan, J. Phys. C: Solid State Phys. **18**, 973 (1985).

[246] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[247] T. R. Paudel and W. R. L. Lambrecht, Phys. Rev. B **77**, 205202 (2008).

[248] D. Segev and S.-H. Wei, Phys. Rev. Lett. **91**, 126406 (2003).

[249] A. Gali *et al.*, Phys. Rev. B **75**, 045211 (2007).

[250] C. W. M. Castleton and S. Mirbt, Phys. Rev. B **70**, 195202 (2004).

[251] A. F. Wright and N. A. Modine, Phys. Rev. B **74**, 235209 (2006).

[252] D. J. Chadi *et al.*, Phys. Rev. Lett. **79**, 4834 (1997).

[253] T. Korhonen, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **54**, 15016 (1996).

[254] X. Gonze *et al.*, Comput. Phys. Commun. **180**, 2582 (2009).

[255] C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).

[256] I. Abbati *et al.*, J. Vac. Sci. Technol. **17**, 930 (1980).

[257] A. C. Diebold *et al.*, Mater. Sci. Semicond. Process. **2**, 103 (1999).

[258] N. Awaji *et al.*, Jpn. J. Appl. Phys. **35**, L67 (1996).

[259] S. D. Kosowsky *et al.*, Appl. Phys. Lett. **70**, 3119 (1997).

[260] N. Ikarashi, K. Watanabe, and Y. Miyamoto, Phys. Rev. B **62**, 15989 (2000).

[261] N. Ikarashi, K. Watanabe, and Y. Miyamoto, J. Vac. Sci. Technol. A **21**, 495 (2003).

[262] I. Takahashi *et al.*, Jpn. J. Appl. Phys. **42**, 7493 (2003).

[263] F. J. Himpsel *et al.*, Phys. Rev. B **38**, 6084 (1988).

[264] P. J. Grunthaner, M. H. Hecht, F. J. Grunthaner, and N. M. Johnson, J. Appl. Phys. **61**, 629 (1987).

[265] F. Rochet *et al.*, J. Non-Cryst. Solids **216**, 148 (1997).

[266] J. H. Oh *et al.*, Phys. Rev. B **63**, 205310 (2001).

[267] M. M. Banaszak Holl and F. R. McFeely, Phys. Rev. Lett. **71**, 2441 (1993).

[268] A. Pasquarello, M. S. Hybertsen, and R. Car, Phys. Rev. Lett. **74**, 1024 (1995).

[269] A. Pasquarello, M. S. Hybertsen, and R. Car, Appl. Phys. Lett. **68**, 625 (1996).

[270] A. Pasquarello, M. S. Hybertsen, and R. Car, Phys. Rev. B **53**, 10942 (1996).

[271] D. R. Hamann, Phys. Rev. B **61**, 9899 (2000).

[272] S. C. Witczak, J. S. Suehle, and M. Gaitan, Solid-State Electron. **35**, 345 (1992).

[273] A. Stesmans and V. V. Afanas'ev, J. Phys.: Condens. Matter **10**, L19 (1998).

[274] I. Ohdomaria, H. Akatsua, Y. Yamakoshia, and K. Kishimotoa, J. Non-Cryst. Solids **89**, 239 (1987).

[275] R. Buczko, S. J. Pennycook, and S. T. Pantelides, Phys. Rev. Lett. **84**, 943 (2000).

[276] A. Pasquarello, M. S. Hybertsen, and R. Car, Nature **396**, 58 (1998).

[277] K.-O. Ng and D. Vanderbilt, Phys. Rev. B **59**, 10132 (1999).

[278] Y. Tu and J. Tersoff, Phys. Rev. Lett. **84**, 4393 (2000).

[279] Y. Tu and J. Tersoff, Thin Solid Films **400**, 95 (2001).

[280] W. H. Zachariasen, J. Am. Chem. Soc. **54**, 3841 (1932).

[281] F. Wooten, K. Winer, and D. Weaire, Phys. Rev. Lett. **54**, 1392 (1985).

[282] A. Bongiorno, A. Pasquarello, M. S. Hybertsen, and L. C. Feldman, Phys. Rev. Lett. **90**, 186101 (2003).

[283] D. Fischer, A. Curioni, S. Billeter, and W. Andreoni, Appl. Phys. Lett. **88**, 012101 (2006).

[284] S. R. Billeter, A. Curioni, D. Fischer, and W. Andreoni, Phys. Rev. B **73**, 155329 (2006).

[285] Y. Sugitaa and Y. Okamotoa, Chem. Phys. Lett. **314**, 141 (1999).

[286] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2005).

[287] Y. Tu, J. Tersoff, G. Grinstein, and D. Vanderbilt, Phys. Rev. Lett. **81**, 4899 (1998).

[288] P. N. Keating, Phys. Rev. **145**, 637 (1966).

[289] N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953).

[290] P. Tangney and S. Scandolo, J. Chem. Phys. **117**, 8898 (2002).

[291] J. J. Wortman and R. A. Evans, J. Appl. Phys. **36**, 153 (1965).

[292] R. T. Downs and D. C. Palmer, Am. Mineral. **79**, 9 (1994).

[293] L. Levien, C. T. Prewitt, and D. J. Weidner, Am. Mineral. **65**, 920 (1980).

[294] S. Susman, K. J. Volin, D. G. Montague, and D. L. Price, Phys. Rev. B **43**, 11076 (1991).

[295] C. R. Helms and E. H. Poindexter, Rep. Prog. Phys. **57**, 791 (1994).

[296] P. B. Griffin, S. W. Crowder, and J. M. Knight, Appl. Phys. Lett. **67**, 482 (1995).

[297] R. Duffy *et al.*, Appl. Phys. Lett. **86**, 081917 (2005).

[298] G. A. Sai-Halasz, K. T. Short, and J. S. Williams, IEEE Electron Device Lett. **6**, 285 (1985).

[299] R. Kasnavi *et al.*, J. Appl. Phys. **87**, 2255 (2000).

[300] S. Whelan *et al.*, J. Appl. Phys. **90**, 3873 (2001).

[301] J. A. Van den Berg *et al.*, J. Vac. Sci. Technol. B **20**, 974 (2002).

[302] M. J. P. Hopstaken *et al.*, Appl. Surf. Sci. **231**, 688 (2004).

[303] M. Ferri *et al.*, J. Appl. Phys. **99**, 113508 (2006).

[304] A. Parisini *et al.*, Appl. Phys. Lett. **92**, 261907 (2008).

[305] C. Steen *et al.*, J. Appl. Phys. **104**, 023518 (2008).

[306] L. Pei *et al.*, J. Appl. Phys. **104**, 043507 (2008).

[307] F. Ruffino *et al.*, Appl. Phys. Lett. **96**, 093116 (2010).

[308] C. Steena, P. Pichlera, and H. Ryssela, Mater. Sci. Eng. B **154**, 264 (2008).

[309] H. C.-H. Wang *et al.*, IEEE Electron Device Lett. **22**, 65 (2001).

[310] R. Duffy *et al.*, Appl. Phys. Lett. **82**, 3647 (2003).

[311] M. Orlowski, Appl. Phys. Lett. **55**, 1762 (1989).

[312] Y. Sato, M. Watanabe, and K. Imai, J. Electrochem. Soc. **140**, 2670 (1993).

[313] Y. Sato, K. Imai, H. Yomezawa, and T. Shigematsu, J. Electrochem. Soc. **140**, L177 (1993).

[314] J. Dabrowski *et al.*, J. Vac. Sci. Technol. B **18**, 2160 (2000).

[315] H.-H. Vuong *et al.*, IEEE Trans. Electron. Devices **43**, 1144 (1996).

[316] J. Frühauf, Ph.D. thesis, Technische Universität München, 2005.

[317] M. Pierre *et al.*, Nature Nanotech. **5**, 133 (2010).

[318] M. Diarra, Y.-M. Niquet, C. Delerue, and G. Allan, Phys. Rev. B **75**, 045301 (2007).

[319] J. Dabrowski *et al.*, J. Vac. Sci. Technol. B **18**, 2160 (2000).

[320] J. Dabrowski *et al.*, Phys. Rev. B **65**, 245305 (2002).

[321] K. Ravichandran and W. Windl, Appl. Phys. Lett. **86**, 152106 (2005).

[322] Y.-S. Kim and K. J. Chang, Appl. Phys. Lett. **87**, 041903 (2005).

[323] Z. Zhou *et al.*, J. Appl. Phys. **98**, 076105 (2005).

[324] N. Kong, T. A. Kirichenko, G. S. Hwang, and S. K. Banerjee, Phys. Rev. B **80**, 205328 (2009).

[325] R. Rurali and X. Cartoixà, Nano Lett. **9**, 975 (2009).

[326] F. J. Morin and J. P. Maita, Phys. Rev. **96**, 28 (1954).

[327] A. Erbil, W. Weber, G. S. Cargill, and R. F. Boehme, Phys. Rev. B **34**, 1392 (1986).

[328] V. Koteski *et al.*, Nucl. Instrum. Meth. B **200**, 60 (2003).

[329] A. Ural, P. B. Griffin, and J. D. Plummer, J. Appl. Phys. **85**, 6440 (1999).

[330] E. L. Elkin and G. D. Watkins, Phys. Rev. **174**, 881 (1968).

[331] J. Xie and S. P. Chen, J. Phys.: Condens. Matter **11**, 7219 (1999).

[332] J. Han, T.-L. Chan, and J. R. Chelikowsky, Phys. Rev. B **82**, 153413 (2010).

[333] F. Corsetti and A. A. Mostofi, Phys. Rev. B **84**, 035209 (2011).

[334] D. Alfè, G. A. D. Wijs, G. Kresse, and M. J. Gillan, Int. J. Quantum Chem. **77**, 871 (2000).

[335] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni, Phys. Rev. B **43**, 7231 (1991).

[336] A. Dal Corso, A. Pasquarello, and A. Baldereschi, Phys. Rev. B **56**, R11369 (1997).

[337] H. Hellmann, *Einführung in die Quantenchemie* (Franz Deuticke, Leipzig, 1937).

[338] R. P. Feynman, Phys. Rev. **56**, 340 (1939).

[339] K. Kunc and R. M. Martin, Phys. Rev. Lett. **48**, 406 (1982).

[340] G. Kresse, J. Furthmüller, and J. Hafner, Europhys. Lett. **32**, 729 (1995).

[341] N. D. M. Hine *et al.*, Phys. Rev. B **83**, 195102 (2011).

[342] G. Dolling, Proc. Symp. Inelastic Scattering Neutrons in Solids and Liquids, Chalk River, IAEA, Vienna **2**, 37 (1963).

[343] P. D. Desai, J. Phys. Chem. Ref. Data **15**, 967 (1986).

[344] S. Wei, C. Li, and M. Y. Chou, Phys. Rev. B **50**, 14587 (1994).

[345] M. Shelley, N. Poilvert, A. A. Mostofi, and N. Marzari, Comput. Phys. Commun. **182**, 2174 (2011).

[346] E. I. Blount, Solid State Phys. **13**, 305 (1962).

[347] H. Teichler, Phys. Status Solidi (b) **43**, 307 (1971).

[348] E. O. Kane and A. B. Kane, Phys. Rev. B **17**, 2691 (1978).

[349] C. Tejedor and J. A. Vergés, Phys. Rev. B **19**, 2283 (1979).

[350] S. Satpathy and Z. Pawlowska, Phys. Status Solidi (b) **145**, 555 (1988).

[351] P. Fernández, A. Dal Corso, A. Baldereschi, and F. Mauri, Phys. Rev. B **55**, R1909 (1997).

[352] W. Kohn, Phys. Rev. B **7**, 4388 (1973).

[353] B. Sporkmann and H. Bross, J. Phys.: Condens. Matter **9**, 5593 (1997).

[354] P. Giannozzi *et al.*, J. Phys.: Condens. Matter **21**, 395502 (2009).

[355] D. Weaire and M. F. Thorpe, Phys. Rev. B **4**, 2508 (1971).

[356] M. F. Thorpe and D. Weaire, Phys. Rev. B **4**, 3518 (1971).

[357] M. Wahn and J. Neugebauer, Phys. Status Solidi (b) **243**, 1583 (2006).