A theoretical investigation of gas source growth of the Si(001) surface

D.R.Bowler
Wolfson College
D.Phil Thesis
Trinity 1997

Preface

This thesis is an account of the work carried out by the author during the period October 1994 to June 1997, under the supervision of Dr. G.A.D.Briggs and Professor D.G.Pettifor. It has not been previously submitted for a degree at this or any other university. The work of others, which has been drawn upon, is acknowledged as such where used. Parts of this thesis have been published or are in press, and a list of such parts is given below.

Owen J.H.G., Bowler D.R., Goringe C.M., Miki K. and Briggs G.A.D. (1995) *Surf. Sci. Lett.* **341**, L1042.
Bowler D.R. and Goringe C.M. (1996) *Surf. Sci. Lett.* **360**, 489.
Owen J.H.G., Bowler D.R., Goringe C.M., Miki K. and Briggs G.A.D. (1996) *Phys. Rev. B* **54**, 14153.
Owen J.H.G., Bowler D.R., Goringe C.M., Miki K. and Briggs G.A.D. (1997) *Surf. Sci. Lett.* **382**, L678.
Bowler D.R., Aoki M., Goringe C.M., Horsfield A.P. and Pettifor D.G. (1997) *Modelling Simul. Mat. Sci. Eng.* **5**, 199.
Owen J.H.G., Miki K., Bowler D.R., Goringe C.M., Goldfarb I. and Briggs G.A.D. (1997a) *Surf. Sci.* In press.
Owen J.H.G., Miki K., Bowler D.R., Goringe C.M., Goldfarb I. and Briggs G.A.D. (1997b) *Surf. Sci.* In press.
Goringe C.M., Bowler D.R. and Hernández E.H. (1997) *Rep. Prog. Phys.* In press.
Goringe C.M. and Bowler D.R. (1997) *Phys. Rev. B*, In press.

A theoretical study of gas source growth of the Si(001) surface

D.R.Bowler, Wolfson College

Thesis submitted for the degree of Doctor of Philosophy

Trinity 1997

Abstract

The growth of the Si(001) surface from gas sources such as disilane is technologically important, as well as scientifically interesting. The aspects of growth covered are: the clean surface, its defects and steps; the action of bismuth, a surfactant; the diffusion behaviour of hydrogen in different environments; and the entire pathway for formation of a new layer of silicon from adsorption of fragments of disilane to nucleation of dimer strings. The theoretical methods used, density functional theory and tight binding, are described. Four linear scaling tight binding methods are compared. The construction of the tight binding parameterisations used is also explained.

The structure of the most common defect on the Si(001) surface is identified by comparison of the electronic structure with scanning tunneling microscopy (STM) images. The energy and structure of steps is calculated, and their kinking behaviour is modelled, achieving good agreement with experimental results.

Two unusual features which form when bismuth is placed on the surface and annealed are investigated. The first has possible applications as a quantum wire, and its structure and growth are described. The second relates to a controversial area in the field; a structure is proposed which fits all available experimental evidence.

The behaviour of hydrogen is vital to understanding growth, as large amounts are deposited during disilane growth. After validating the tight binding parameterisation against DFT and experiment for the system of a single hydrogen diffusing on the clean Si(001) surface, the barriers for diffusion on the saturated surface, down a step and away from a defect are found, and prove to be in good agreement with available experimental data.

The pathway for the formation of a new layer of silicon from disilane is described step by step, giving barriers and structures for all events. The interaction with experiment is highlighted, and demonstrates that great benefit accrues from such close work, and that the atomistic modelling techniques used in the thesis produce results in close agreement with reality.

Contents

List of Figures

Acknowledgements

One of the nicest things about writing this thesis is being able to acknowledge publicly those people who have contributed to the work over the last three years, in many different ways.

Firstly, my supervisors **Andrew Briggs** and **David Pettifor**, who have allowed me the freedom to pursue what I found interesting, while being happy to provide advice, guidance and help wherever needed. I have greatly enjoyed my time working with them. I would also like to thank my new employer, **Mike Gillan**, for allowing me to work on my thesis as well as on CONQUEST, and for his insights into the subject.

It is often the case that while the supervisors provide overall guidance and ideas, post docs give a more detailed level of help. I have been extremely fortunate in working closely with two post docs in particular: **Chris Goringe** and **Andrew Horsfield**. They have both helped me enormously over the last three years, while being very patient with my incessant questions. Chris has helped me more with the systems I have modelled, while Andrew has taught me more about the methods I have used, but there has also been a great deal of overlap.

As well as theoretical work, this thesis revolves around experiments, and those experimentalists I have worked with deserve thanks for their patience and enthusiasm. First and foremost of these is **James Owen**, whose thesis should be consulted for details of any of the experiments mentioned herein. Our collaboration has proved fruitful, and our friendship has enriched my time in Oxford. Other people in the lab with whom I have worked closely include **Ilan Goldfarb**, **Kazushi Miki** and **Holger Nörenberg**.

The Materials Modelling Laboratory is a superb environment in which to work, and there are many people there who have helped me at one time or another. **Neil Long** performs wonders in maintaining the computers, and has been very helpful and supportive during the writing of this thesis. Other people who have helped me include **Alex Bratkovsky, Mike Fearn, Hideaki Fujitani** and **Tony Paxton**.

Many friends have contributed to keeping me and Erica sane and happy during this time, and I should like to thank **Mike and Hester, Steve and Heather, John and Claire, Chris and Su**; as well as these friends around the country, there have been many friends in the lab and around Oxford who have contributed to the profits of the Royal Oak with us: **James, Catherine, Adam, Adrian, Chris, Paul, Rob, James and Joanna**.

Finally I would like to acknowledge my family: my grandmother, for her love and friendship over many years, for teaching me the piano and sharing the Blüthner; my father, for showing me how to do physics *properly*, and for willing help during the writing of this thesis; and Erica, whose love and support has made all this possible.

In Memory of P.M.Bowler
1939-1993

Chapter 1

Introduction: Gas Source Growth

The life so short, the craft so long to learn

Hippocrates, *Aphorisms* I, i

One of my favourite possessions as a boy was a crystal of copper sulphate which I had grown. I had used a common kit, which was started by forming a seed crystal through evaporation of a supersaturated solution, and then growing the larger crystal over time by suspending it in a concentrated solution. The growth of silicon from gas source hydrides, the key theme of this thesis, is rather different, and requires more expensive equipment, but preserves a common theme: the fascination of watching crystals grow.

Gas-source growth (so called because the silicon source is a gas rather than silicon atoms evaporated from a Knudsen cell, as is the case for solid source growth) involves the family of silane gases (SiH_4, Si_2H_6 and so on) which are the silicon analogue of the alkanes. Traditionally, growth of silicon and silicon/germanium multilayers has been performed using solid source growth; recently, gas source growth has been proposed as an alternative, mainly for cleanliness and doping reasons. Solid source molecular beam epitaxy (SSMBE) uses cells at up to 2000°C, which introduces significant contamination problems, and the necessary substrate temperatures can cause dopant and germanium segregation - these problems are avoided by gas source MBE (GSMBE), which uses room temperature gas, and lower substrate temperatures. The passivating and surfactant effects of the hydrogen on the substrate convey additional benefits. While industrial processes use silane itself (due mainly to the lower cost of production) academic investigations tend to use disilane (Si_2H_6), which breaks up more easily, allowing growth at lower temperatures[1], and without the use of an activator such as chlorine. However, this difference should result in little change to the results found in this work.

The study of growth from these gaseous hydride precursors has a number of interesting and important facets. The differences between it and solid source growth are extremely interesting, and are due both to the hydrogen present, and the different pathways to formation of clean silicon dimers, as explained in Chapter 7. The extension of GSMBE to SiGe alloys is extremely exciting for possible future applications in the electronics industry. At the most basic level, the study is interesting because the key atomic reactions involved in the growth are not understood.

The work which is presented in this thesis involves the computational modelling of the (001) surface of silicon (the face most commonly used for industrial growth and device fabrication), the behaviour of two relevant elements on the surface (hydrogen, which comes from the gas itself, and bismuth, which is another possible surfactant though is more likely to be used in SSMBE than GSMBE), and the reactions which occur in the growth of a new layer of silicon from disilane. This allows the entire pathway to be understood, which will have implications for larger-scale models of growth, and hence actual industrial processes. There is also an investigation of the best tight binding method for modelling these processes, and a description of the development of appropriate parameterisations to allow this.

One of the key ingredients for the success of the theoretical study was a very close interaction with experiments. This interaction will be expounded upon later in the introduction, and it has been my aim, where possible, throughout the thesis to highlight those areas which would have been impossible without interaction - both from the theoretical side, and the experimental side. James Owen has already been acknowledged as a very important co-worker, and his thesis (Owen 1996) forms an important companion volume to this, detailing the experiments more fully than is possible here.

[1] Indeed, lower temperatures are demanded, since the sticking coefficient of disilane drops over 700K (Gates 1988).

1.1 Materials modelling

At the recent inauguration of the Convex Exemplar in the Materials Modelling Laboratory (MML), I was talking to a director of a medium sized industrial firm, which specialises in alloys. He was telling me that, in the last ten years, materials modelling has gone from being "worse than useless" in the prediction of the properties of new materials (specifically, heats of formation were a factor of ten out) to actually producing useful predictions today. It is the ability to model materials at the level of a few atoms sufficiently accurately that macroscopic properties can be predicted that is the eventual goal of all materials scientists. The motto of the MML is "linking the length scales" - that is, understanding how to relate results using one technique to those from another. One of the best examples of this is looking at how a material cracks - at the tip of the crack, individual bonds are breaking, and the quantum mechanical description of these bonds must be correct; a little further away, two new surfaces are forming, and the behaviour of the atoms must be accounted for; as the crack broadens, the microstructure of the material becomes important; and at the largest scales, continuum elasticity theory can predict strain and heat dissipation. At each of these levels, the modelling method described is sufficiently accurate for the results required, and allows simulations to take place in a reasonable amount of time.

The length scales covered range from Ångströms to metres, and the time scales from femtoseconds to minutes, or even hours. The work here covers only scales from Ångströms up to a few tens of nanometres, and time scales from femtoseconds to a few tenths of a picosecond. However, there is a very high possibility that the reaction barriers which have been found herein will be used in a kinetic Monte Carlo model, which should enable simulations to cover up to a micron of growth on a cell 100 $\times$ 100 Ångströms, covering an hour of simulated time. The technique to be used has already been proven by Battaile, Srolovitz and Butler (1997) for diamond growth. The eventual aim of the whole project is to produce predictions of layers grown which will be sufficiently accurate to enable planning *before* growth commences in a reactor.

1.2 Experiments and theory

While I am not an experimentalist *per se*, I have had to learn a fair amount about various different experimental techniques, their limitations, and (most importantly) what they can and cannot measure. The work which I have performed over the last three years has revolved around the real space data which Scanning Tunneling Microscopy (STM) can give, though other techniques have been relevant, and certainly results from the literature must be born in mind. I shall therefore give a brief description of the principles behind the various relevant surface science techniques and their areas of applicability.

Since the 1960s, surface science has burgeoned as a discipline in its own right, as the development of ultra-high vacuum (UHV) systems, and computers and codes, have enabled rigorous understanding of the relevant processes (Zangwill 1988). Since the invention of the STM (Binnig *et al.* 1982), an atomic-scale, real space understanding has become possible. The importance of UHV as a means of preparing a "clean" surface, as opposed to a "real" surface (Many, Goldstein and Grover 1965) cannot be underestimated; this ability to image atomic processes has brought about the work in this thesis.

The scanning tunneling microscope works in a very simple way: a sharp, conducting tip is brought to within a few Ångströms of a conducting surface, and a voltage is applied between them. When the tip and surface are brought together, electrons can tunnel through the barrier (caused by the insulating vacuum gap) in a well-known quantum mechanical effect. By applying a voltage between the two, the tunneling in one direction is enhanced over the tunneling in the other, giving rise to a current. The magnitude of the current is strongly related to the size of the gap (from a simple treatment, it appears to be exponential) and also to the electronic properties of the surface. If the tip is scanned over the surface, a map of the topography can be built up. The most common mode of imaging, one used for all pictures in this thesis, is known as *constant current*; a simple feedback mechanism is used to adjust the height of the tip, so that the current flowing remains constant, and the height of the tip as it scans is displayed. This is suitable for high magnification work at reasonable temperatures. If images at high temperature (over 700K) or images of very rough surfaces are desired, then the tip must be scanned at constant height, and a display based on the logarithm of the current is produced.

The STM cannot give any chemical information, except in as much as the electronic behaviour of different species may give rise to a variation in contrast. One valuable tool for obtaining broadly averaged chemical information is Auger Electron Spectroscopy (AES). In this technique, a beam of high energy electons (over 1 keV) is directed at the sample, and the back-scattered electron spectrum, $N(E)$, is collected. Taking the

derivative of this spectrum, $\mathrm{d}N(E)/\mathrm{d}E$, magnifies small oscillations in the signal due to *secondary* electrons. If an incoming electron collides with an atom in the solid, and ionises a 1s electron, then a hole in the 1s shell will result. A 2s electron can drop down to fill this, and the transition energy can eject a *second*, Auger electron from the 2p shell, which has an energy characteristic of the transitions involved and the atom in which it occurred. These have been tabulated for all of the atoms in the periodic table except for hydrogen and helium; by measuring the relative strengths of the different peaks in an Auger spectrum, a measure of the proportions of different elements can be built up. A related procedure, X-ray photoemission spectroscopy (XPS), again takes advantage of the core-hole binding energies. The sample is illuminated with a source of monochromatic X-rays, and the spectrum of emitted electrons displays peaks at kinetic energies characteristic of elements in particular environments. Both of these techniques are sensitive to about 1% of surface coverage.

An alternative, though destructive, way of obtaining chemical composition is to remove the top few layers and perform mass spectroscopy on them. Secondary Ion Mass Spectroscopy (SIMS) does just this, by bombarding the surface with a beam of ions or atoms with high energies (greater than 1 keV). Atoms, or clusters of atoms, which are sputtered off are then sent into a mass spectrometer, yielding information at the level of 10^{-6} of a monolayer - far more sensitive than AES or XPS.

Crystal structure can be very important, particularly during growth, and there are two common ways of monitoring this. Low energy electron diffraction (LEED) fires electrons with energies in the range of 20-500 eV at the surface, and collects the *elastically* back-scattered electrons. These then form a Fraunhofer diffraction pattern which is the Fourier transform of the surface atomic arrangement. A sharp pattern (shown by crisp spots) implies a well-ordered substrate, whereas a diffuse one will indicate a basic structure, but some disorder; it is worth noting that the symmetry can be lower than that implied on the LEED screen. A related technique, that of reflection high energy electron diffraction (RHEED), is used most often to monitor crystal growth. A beam of high energy electrons is fired at the surface at an extreme grazing incidence angle. Those electrons which are scattered through only a few degrees will sample the first 1-2 atomic layers, and no more. If the specular beam is monitored during a growth experiment, then regular oscillations are seen, corresponding to smooth and rough surfaces, which allows layer-by-layer growth to be monitored.

There are many experimental techniques which are relevant to this area, and I have given a brief introduction to the most important. However, there is more to the experimental-theoretical interaction than simply taking experimental results and modelling them, or *vice versa*. If the interaction proceeds on a day-to-day basis, as was possible in the department for myself and James Owen, then a mutual trust and respect is built up, and the level of interaction deepens. Ideas can form over coffee (or beer !) which require joint expertise; a deeper understanding results on both sides when explanations must be made; a synergy occurs which results in new understandings, and a feedback mechanism comes into play: as results are discussed, new experiments or calculations come to light, which then feed back into the discussion. In this way, new pathways have been discovered - for instance, the square feature which will be described in Chapter 7 would never have been found without the interaction, and the formation of dimer strings from this would have remained a mystery.

1.3 Interpreting STM images

When STM first appeared, there appeared to be little problem interpreting the images. Tersoff and Hamann (1983) produced a theory of STM current flow which worked well for metals though not for semiconductors; however, as time has gone on, the question of how easy it is to obtain reliable, quantitative interpretations of STM images has become far more vexed. Even in the (apparently) simple system of a metal surface interacting with a metal tip, there are many areas of interest, in which recent progress has been made (Clarke *et al.* 1996). As well as the question of tip-sample interactions, there is the question of simulated images, and the effect of the electric field on the images. Issues related to some of these areas have been addressed by Ness and Fisher (1997) and Fisher, Blöchl and Briggs (1997).

However, tip-sample interaction and simulating images, while fascinating, are not the concern of this thesis. The main drive of the thesis is surface structure and surface reactions: how they can be modelled, and how they can be related to the available experimental data. As the most readily available, and most appropriate, data is from the STM in Oxford, a means of interpreting the images and relating them to calculations has been developed. The guiding principles used are listed below.

1. Initially assume that all contrast is due to geometrical effects (J.B.Pethica, Private Communication).

2. Where appropriate (e.g. when the image changes with bias voltage) assume that there is an electronic effect.

3. When electronic effects are important, make the simplest possible assumptions: the energy-resolved charge density or local densities of states from modelling are *in some way* related to the image behaviour.

These principles have performed well in general, allowing a fruitful interaction to develop. For those interested in the questions discussed above, it could be interesting to examine where, and why, these principles break down; however, this must be left to others.

1.4 Kinetics

Much of this thesis is concerned with the calculation of reaction barriers and their relationship to the experimental temperatures at which certain process are seen to occur. The formal theory of kinetics states that a reaction *rate* will occur as (Christian 1965):

$$k_f = \left(\frac{kT}{hQ_\nu}\right) \exp\left(\frac{-\Delta G}{kT}\right), \tag{1.1}$$

where Q_ν is the vibrational partition function of the reactants and ΔG is the free energy difference between the start point and the transition state. There are two key assumptions which are made in the derivation of this formula (see Christian 1965 and Glasstone, Laidler and Eyring 1941 for further discussion), which are that:

1. The reaction has a well-defined start and end, with the one proceeding to the other via a continuous change of coordinates. These initial and final states are assumed to be local minima.

2. There is an energy barrier between the initial and final states for any choice of reaction path, and the most-favoured reaction path is that path with the lowest energy barrier (which is therefore a saddle-point in the energy).

Equation 1.1 can be rewritten, using the fact that $\Delta G = \Delta H - T\Delta S$ (where ΔH is the internal energy and ΔS is the entropy of the system), and factorising the first term into a characteristic frequency of vibration, ν,

$$\begin{aligned} k_f &= \nu \exp\left(\frac{-\Delta G}{kT}\right), \\ &= \nu \exp\left(\frac{\Delta S}{k}\right) \exp\left(\frac{-\Delta H}{kT}\right). \end{aligned} \tag{1.2}$$

This theory was further developed and applied to solid-state reactions by Vineyard (1957), who showed that Eq.(1.2) could be further simplified to give:

$$k_f = \nu^* \exp(-\Delta H/kT), \tag{1.3}$$

with $\nu^* = \nu \exp(\Delta S/k)$, and that the entropy could be written in terms of the frequencies of the normal modes of the system. Vineyard made the further assumption that the motions of the atoms in the system both at the start point and at the transition state can be treated as small, harmonic oscillations when deriving this theory.

The formal theory as presented above has one drawback, as far as the calculation of reaction rates is concerned: it is extremely difficult to calculate a frequency, ν^*, for use in the formula. It is also important to realise that these frequencies are temperature dependent.

When reaction rates have been calculated for comparison with experiment, or equivalently, when a barrier has been inferred from a reaction rate which has been measured, various assumptions have been made:

1. The barrier found is an internal energy, not a free energy (this places all temperature dependence into the prefactor, ν^*, and uses Eq.(1.3) for comparisons).

2. Over the temperature range concerned (which is generally not more than 50K) the prefactor remains temperature independent.

3. The best frequency prefactor available is used; where possible, this is taken from a fit to experimental data, or measurements of frequency modes made using various types of spectroscopy.

The reader should be aware that the inaccuracies inherent in the approach given above will be quoted as an error; for instance, at about 600K, the use of a frequency of 10^{13} rather than 10^{14} will change the barrier calculated by about 0.1 eV.

1.5 Overview

The reader may find it helpful at this point to have an overview of the thesis, so that he or she can gain a broad view of the material involved. One of the more pleasurable aspects of writing this thesis has been the drawing together of three years' work into a coherent whole, and I would like to pass this feeling on to the reader. This thesis has two main themes: the application of tight binding and density functional theory to the problem of growth of silicon from disilane (and related topics); and the importance of experimental-theoretical interaction. Without the experimental work which has been progressing during my time (and in particular without James Owen) very little of what is in this thesis would have happened.

The thesis starts with a description of the theory which underlies the techniques I have used. Density functional theory is described, along with several of the areas which I feel that it is important to understand before DFT results can be relied upon. The basis for tight binding is then discussed, and four of the recent linear scaling methods are presented in some detail; this area is a particular strength in Oxford, and that expertise has been exploited. Chapter 3 is rather more involved in the nitty gritty details of modelling, and starts with a comparison of the four linear scaling methods available in the MML. The systems examined are biased towards those relevant for this thesis (i.e. semiconductors and their surfaces). The chapter ends by describing three tight binding parameterisations which I have developed for this work.

Before any problems of reactions on surfaces can be considered, the clean surface and its defects must be understood. Chapter 4 presents the silicon (001) surface and the structure of the most common defect. A discussion of steps and their kinking behaviour rounds it off. Chapter 5 considers a new surfactant, bismuth (a horrible element to model), and presents results for some fascinating structures which have been seen in STM after Bi adsorption. Chapter 6 deals with hydrogen, both as a prototypical system for studying adsorbate diffusion, and as a key player in the growth of silicon from disilane. The key motivating subject for this thesis, growth of silicon from disilane, is addressed in Chapter 7. This covers the whole process, from the basic adsorption products, to the nuclei for small islands, and anti-phase boundaries (features which only occur when a nearly complete layer of silicon has been grown) at all times remaining in close contact with experimental results, which are mainly those generated in Oxford. I should explain that the references are sorted alphabetically by chapter, and where a reference occurs in two chapters, it will be found in the references twice.

Chapter 2

Theoretical Techniques

It is the behavior and distribution of the electrons around the nucleus that gives the fundamental character of an atom: it must be the same for molecules.

C. A. Coulson, 1951

As atoms of different sorts go to make up everything around us, understanding how they bond to each other and react is extremely important; it is this understanding that atomistic modelling seeks to achieve. The essential aim of all such techniques is to obtain the total energy and force for a given arrangement of atoms, so $E = E(\mathbf{R}_i)$, where E is energy, and $\mathbf{R}_i$ indicates the positions of the atoms, and $\mathbf{F}(\mathbf{R}_i) = -\nabla_{\mathbf{R}_i} E$; once this energy has been obtained, a whole world of simulation opens up: the minimum energy structure for a set of atoms can be found, molecular dynamics can be performed, phonon modes calculated, etc. There are many ways to find the energy, ranging from the so-called *ab initio* techniques (literally "from the beginning" - meaning that the atomic numbers and their coordinates are the only input), through semi-empirical techniques such as tight binding (TB), where quantum mechanics is retained, but chemical intuition is used to simplify the problem enormously, to the level of empirical potentials, where bonds are represented by empirically fitted laws such as a harmonic potential, and quantum mechanics is sacrificed for computational efficiency. The work in this thesis has been performed using two quantum mechanical techniques: the accurate, *ab initio* method of Density Functional Theory (DFT), solved using the Local Density Approximation (LDA) or one of the Generalized Gradient Approximations (GGA); and the less accurate (though highly effective) semi-empirical technique of tight binding, specifically using a method known as the Density Matrix Method (DMM).

The main aim of this chapter is to leave the reader with an understanding of the techniques which have been used throughout the rest of this work. To this end, I shall describe the key theoretical points behind each method, and also try to address some of the problems facing each of them. The theory behind four of the new linear scaling TB schemes (where computation effort scales with N, the number of atoms in the system, rather than N^3, as is true for traditional methods) is described here, while the results of an investigation into which method is best for semiconductor surfaces are presented in Chapter 3, so that it is clear why I have chosen to use the DMM.

2.1 General ideas

In this section, I will briefly discuss a useful approximation and an important theorem which apply to all the theoretical methods which will be presented, a method of representing the (almost) infinite solids in an efficient manner and different ways of finding reaction barriers.

2.1.1 The Born-Oppenheimer approximation

The difference in mass between the electrons and the ions in the systems to be calculated is so great (or, equivalently, the electrons have a relaxation time so much shorter than the ions) that the electrons respond almost instantaneously to changes in the position of the ions. This leads to the Born–Oppenheimer approximation (Born and Oppenheimer 1927), which states that the degrees of freedom of the electrons and the ions can be uncoupled. Another way of looking at this is to say that for each position of the nuclei,

the electrons can be assumed to be in the ground state. This is tremendously important for almost all calculations in materials science and condensed matter physics, as it means that for each atomic configuration, the electronic degrees of freedom can be relaxed to the ground state, and the energy for that configuration calculated. This removes the extremely difficult problem of calculating the interactions between moving electrons and ions.

However, the assumption that the electrons and ions can be treated separately is not always a good one, particularly when an electron is excited into a long-lived state, or performs a transition non-adiabatically. This kind of reaction has been studied by Coker and Xiao (1995), and has been successfully applied in the tight binding formalism (see section 2.3) to studies of H_2 in water (Xiao and Coker 1995) and photodissociation of I_2 in Ar clusters (Batista and Coker 1996). HeadGordon and Tully (1995) have also studied non-adiabatic molecular dynamics for other systems, but have not yet published results of this formalism, presenting it at meetings only (Abstracts of the American Chemical Society 1995-7).

2.1.2 The Hellmann-Feynman theorem

The Hellmann-Feynman theorem (Hellmann 1937, Feynman 1939) relates to a derivative of the total energy of a system, and is usually applied to finding forces on an atom. It shows that the force on an ion is the same as the expectation value of the derivative of the Hamiltonian with respect to that ion position; if a calculation scheme is to have forces and energies in agreement, then the *numerical* derivative of the energy must match the analytically calculated force - the method is said to obey the Hellmann-Feynman theorem if this is true. I have found two demonstrations of this theorem which are particularly helpful. The first involves the density matrix for a system, which will be dealt with in section 2.5. For this theorem to be generally true, the density matrix in question must be the two-particle density matrix (which is required in density functional theory to obtain the Hartree energy correctly). Consider,

$$\begin{aligned} E &= Tr[\rho H] = \sum_{ij} \rho_{ij} H_{ji}, \\ \frac{\partial E}{\partial \lambda} &= \sum_{ij} \frac{\partial E}{\partial \rho_{ij}} \frac{\partial \rho_{ij}}{\partial \lambda} + \sum_{ij} \frac{\partial E}{\partial H_{ij}} \frac{\partial H_{ij}}{\partial \lambda}, \end{aligned} \tag{2.1}$$

where λ is the variable with respect to which the derivative is taken (typically an atomic coordinate), and ρ is the density matrix of the system. If, as assumed above, the electrons are in their ground state, then the first derivative ($\partial E/\partial \rho_{ij}$) will be necessarily zero, and the third derivative ($\partial E/\partial H_{ij}$) will be ρ_{ji} (a simple result from the definition of a trace), so we are left with:

$$\begin{aligned} \frac{\partial E}{\partial \lambda} &= \sum_{ij} \rho_{ji} \frac{\partial H_{ij}}{\partial \lambda}, \\ &= Tr\left[\frac{\partial H}{\partial \lambda}\rho\right] \\ &= Tr\left[\rho\frac{\partial H}{\partial \lambda}\right], \end{aligned} \tag{2.2}$$

where the last line comes from the symmetry of the matrices.

The other way of looking at the theorem involves considering the derivative with respect to position, and noting that as the ion moves, the wave functions will change, so that the full derivative can be expanded out in terms of the wave functions (Payne *et al.* 1992),

$$\frac{\mathrm{d}E}{\mathrm{d}\mathbf{R}_I} = \frac{\partial E}{\partial \mathbf{R}_I} + \sum_i \frac{\partial E}{\partial \psi_i} \frac{\partial \psi_i}{\partial \mathbf{R}_I} + \sum_i \frac{\partial E}{\partial \psi_i^*} \frac{\partial \psi_i^*}{\partial \mathbf{R}_I}, \tag{2.3}$$

where ψ_i are the eigenfunctions of the system. However, it is easy to show that the last two terms (the sum terms) are zero. Given that $\partial E/\partial \psi_i^*$ is just $H\psi_i$, these terms can be rewritten,

$$\sum_i \langle \frac{\partial \psi_i}{\partial \mathbf{R}_I} \mid H\psi_i \rangle + \sum_i \langle \psi_i H \mid \frac{\partial \psi_i}{\partial \mathbf{R}_I} \rangle. \tag{2.4}$$

However, for the eigenstates, $H\psi_i = \lambda_i\psi_i$, so

$$\begin{aligned}\sum_i \langle \frac{\partial\psi_i}{\partial \mathbf{R}_I} \mid \lambda_i\psi_i\rangle \quad &+ \quad \sum_i \langle \psi_i\lambda_i \mid \frac{\partial\psi_i}{\partial \mathbf{R}_I}\rangle \\ &= 2\sum_i \lambda_i \frac{\partial}{\partial \mathbf{R}_I}\langle\psi_i \mid \psi_i\rangle \\ &= 0, \end{aligned} \tag{2.5}$$

as $\langle\psi_i \mid \psi_i\rangle$ is a constant. Note that the differential can be taken outside the integral as the variables are different. This means that the electronic degrees of freedom **do not** contribute to the derivative of the energy, so that

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}\mathbf{R}_I} &= \frac{\partial E}{\partial \mathbf{R}_I} \\ &= \frac{\partial\langle\psi_i \mid H \mid \psi_i\rangle}{\partial \mathbf{R}_I} \\ &= \langle\psi_i \mid \frac{\partial H}{\partial \mathbf{R}_I} \mid \psi_i\rangle. \end{aligned} \tag{2.6}$$

This theorem then allows the force, as well as the energy, to be calculated for a given atomic configuration *without* recalculating the electronic states, or finding their derivatives. This makes force calculations simple and more manageable in terms of time.

2.1.3 Representing an infinite solid

It is unfortunate that the systems which physicists, chemists and materials scientists model are, in effect, infinite, while the computational resources available are all too finite. There are two common ways of achieving a manageable system size: clusters and the supercell method. The cluster method (which I have not used) assumes that, by taking a finite cluster of atoms, and terminating the dangling bonds with hydrogen, the area of interest which is at the centre of the cluster will be adequately described. The problem, of course, is that edge effects will play a large role unless the cluster is very large. This method is generally used for quantum chemical calculations, where only a very few atoms can be treated.

The supercell method defines a unit cell, and then (conceptually) repeats that cell throughout space in an appropriate lattice. This is illustrated schematically in Figure 2.1. The unit cell then must be boxed - i.e. if an atom goes beyond the right-hand boundary, it re-appears just inside the left-hand boundary. This is commonly used in physical calculations, where the electronic structure is solved in reciprocal space - lending itself nicely to an infinitely repeated real space problem.

2.1.4 Modelling Transitions and Reactions

How a molecule passes from one state into another is a fundamental problem for chemists and materials scientists; different methods for calculating such processes are examined here. The essence of the calculation is how to get from the initial state to the final state in such a manner that the lowest energy path (i.e. that which the reaction would itself follow) is found, if indeed the final state is known. All of the calculations presented in this thesis fall into the category of problems where the end points of the reaction are known, and the barrier is desired.

In a reaction where the start and end points are known, there is often a clear choice for a "reaction coordinate" - that is, a coordinate which evolves smoothly from start to finish. A good example of this is the position of a hydrogen atom along the dimer row as it diffuses along the dimer row. If this single coordinate is constrained to different values (generally evenly spaced along the path) and the minimum energy for the system found for those values, then a good approximation to the barrier can be taken to have been found. This is illustrated schematically in Figure 2.2. In many cases, the reaction will proceed smoothly from start to finish, indicating that a smooth path has been found. Where there are two paths which intersect to provide a single total pathway, the barrier for crossing from one to the other is unknown, and a better method is be required for refinement of the calculation.

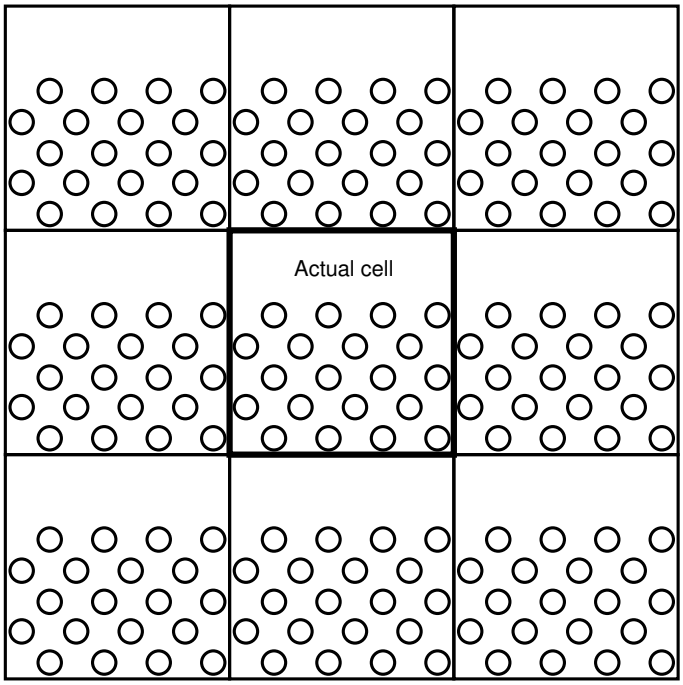

Figure 2.1: A schematic picture of the supercell method of representing an infinite solid for modelling. Only nine cells are shown but, theoretically, an infinite number exist both in the plane of the diagram and out of it.

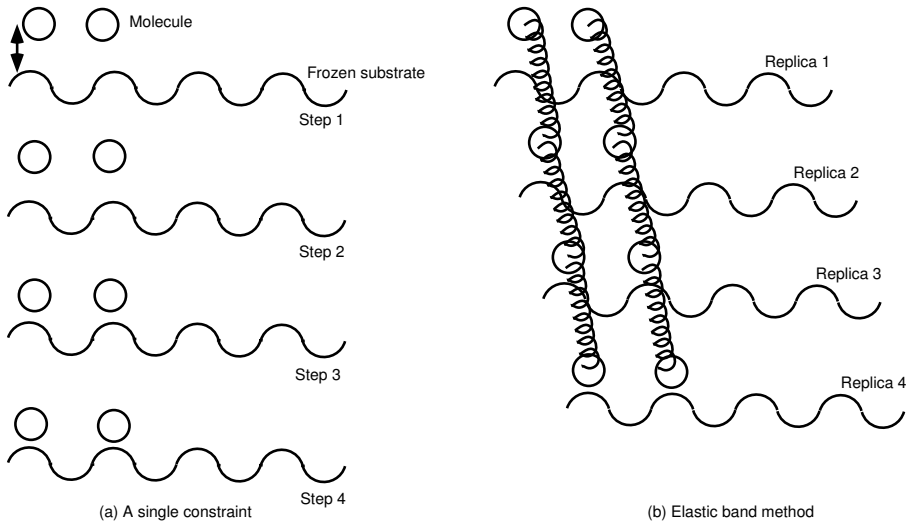

Figure 2.2: A schematic diagram of (a) the single constraint method for finding a reaction path and (b) the "elastic band method". The substrate is assumed to be frozen, and a diatomic molecule is adsorbing. The variable constrained in (a) is the height of the centre of mass of the molecule, shown by an arrow.

An extension of the single constraint method, where there are several different pathways available, is to calculate an energy surface. This is most commonly performed for an adsorption or diffusion reaction (see, for instance, White, Bird and Payne 1996; Bird and Gravil 1997), and can yield valuable information about steering processes, as well as possible mechanisms (Kay *et al.* 1995). For an adsorbing or diffusing molecule, the in-plane positions of one of the atoms is fixed over some point in the surface, and the minimum energy for the system is found. This is then repeated for other values of these positions, yielding an energy surface. The lowest energy pathway for the reaction can then be found rather easily, and further analysis performed on the surface if desired.

A recent, exciting method which has been proposed by Jónsson and Mills (Jónsson and Mills 1997 and Mills, Jónsson and Schenter 1995) involves an approximation to the actual minimum energy path. The start and end points are required, and the path is approximated with a series of replicas of the system, which traverse the reaction. The analogues of each atom in each replica are connected together with springs, and the total energy of the **entire** system is minimised with respect to the positions of the atoms. This process is illustrated in Fig. 2.2. In the limit of an infinite number of replicas, this method finds the minimum energy path; as the number is reduced, the path found becomes steadily more of an approximation. The problem with this method as described is that the springs will tend to pull the replicas off the exact path, leading to "corner cutting"; by introducing "nudging" - that is, once an approximation to the path can be made, resolving spring forces along the path and all other forces perpendicular to the path (the state which the system would be in given a real pathway) the path found becomes much closer to the minimum energy path, and the dependence on the spring constants disappears.

There are, however, drawbacks even to the nudged elastic band method, particularly when the number of replicas is relatively low (for instance eleven, which is a compromise between speed and accuracy when using a quantum mechanical method such as tight binding). In essence there are two constraints being applied to *every* atom in the system (though in practice this will only apply to the reactant species), where in the method described above only one constraint is applied to the entire system. The result of these constraints is the raising of the energy barrier found (as the system has less degrees of freedom than otherwise). There are also problems of implementation: using quenched molecular dynamics proves extremely slow, as the system can be ill-conditioned; however, using a better minimisation scheme (such as conjugate gradients) results in extreme difficulties implementing the nudging. Recently, Kresse and Bates have implemented the method (Kresse and Bates, Private Communication), and used it to find the barrier for proton transfer from an adsorbed methanol group to the TiO_2 substrate (Bates, Kresse and Gillan 1997). They found that implementing the method on the Cray T3D (which is highly parallel) enabled relaxation in reasonable timescales.

Given the relative simplicity of the reactions being modelled in this thesis, and the available computing power, it was felt that the simpler method of single constraints for finding the energy path would be appropriate, and has been used throughout, except where the extension to an energy surface was appropriate.

2.2 Density functional theory

Density functional theory is an accurate, *ab initio* theory which has proved extremely powerful since its inception in 1964. In this section, I will explain the basic ideas underlying the theory, how it works, and also the approximations used to solve the many-body problem of electron-electron interactions. This is not meant to be an in depth analysis of the subject; for such a treatment, the reader is referred to, for instance, Parr and Yang (1989) or Jones and Gunnarsson (1989).

In the first of two papers which defined the methods, Hohenberg and Kohn (1964) proved two relatively simple theorems which radically simplified the problem of calculating electronic structure. For a system of N electrons, the Schrödinger equation and density operators can be written:

$$\hat{H}\Psi = E\Psi, \tag{2.7}$$

$$\hat{H} = \sum_{i=1}^{N}(-\frac{1}{2}\nabla_i^2) + \sum_{i=1}^{N} v(\mathbf{r}_i) + \sum_{i<j}^{N} \frac{1}{r_{ij}}, \tag{2.8}$$

$$\rho(\mathbf{r}_1) = N \int \dots \int \mid \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \mid^2 \mathrm{d}\mathbf{r}_2 \dots \mathrm{d}\mathbf{r}_N, \tag{2.9}$$

where $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ is the many-electron wavefunction, $\hat{H}$ is the Hamiltonian and $\rho(\mathbf{r}_1)$ is the electron density at $\mathbf{r}_1$. The ground state energy and wavefunction can be found by minimising the energy with respect to Ψ; for a given system, the ground state wavefunction (and hence charge density) is wholly determined by N, the number of electrons, and $v(\mathbf{r})$, the external potential (generally due to the atoms).

The first Hohenberg-Kohn theorem vastly simplifies the problem by proving that, to within an additive constant, $v(\mathbf{r})$ and N are determined *solely* by $\rho(\mathbf{r})$[1]. This is effectively reversing the previous statement that

[1]This original theorem is actually not quite correct. The infinite spaces of $v(\mathbf{r})$s and $\rho(\mathbf{r})$s are not equivalent, and there are $\rho(\mathbf{r})$s which *cannot* lead to a $v(\mathbf{r})$. This problem is raised in Levy (1982) and Lieb (1983), and discussed extensively in Parr and Yang (1989).

$v(\mathbf{r})$ determines $\rho(\mathbf{r})$. We can then write:

$$E_v[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho], \tag{2.10}$$

$$V_{ne}[\rho] = \int \rho(\mathbf{r})v(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{2.11}$$

where the square brackets denote a *functional*, E_v represents the energy as a functional of ρ for a given $v(\mathbf{r})$ and T, V_{ne} and V_{ee} are the kinetic, electron-ion and electron-electron energies respectively.

The second Hohenberg-Kohn theorem makes this transformation useful; it shows that for any trial density, $\tilde{\rho}(\mathbf{r})$, which satisfies the physically reasonable conditions that $\tilde{\rho}(\mathbf{r}) \geq 0$ and $\int \tilde{\rho}(\mathbf{r})\mathrm{d}\mathbf{r} = N$,

$$E_{v,GS} \leq E_v[\tilde{\rho}], \tag{2.12}$$

so that the minimum energy is given by the ground state charge density only, and that density is formed from the ground state wavefunction. This means that the charge density can be treated as the variational parameter in a minimisation.

This reformulation was extremely significant, and valuable, but had not yet placed the subject in a state where it was actually usable. Kohn and Sham (1965) continued the development to a point where the theory became practically useful. They performed another reformulation, whereby they constructed one electron wavefunctions, which did not interact, and then lumped all of the many-body interactions into one functional - which is known as the exchange and correlation functional. I will discuss this functional more later, but first I will treat the one electron wavefunctions.

Recall that the energy can be written as a functional of the charge density,

$$E[\rho] = \int \rho(\mathbf{r})v(\mathbf{r})\mathrm{d}\mathbf{r} + T[\rho] + V_{ee}[\rho]. \tag{2.13}$$

Kohn and Sham defined a set of non-interacting wavefunctions, ψ_i, which move in an *effective* potential:

$$v_{\text{eff}}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid}\mathrm{d}\mathbf{r}' + v_{XC}(\mathbf{r}), \tag{2.14}$$

where $v_{XC}(\mathbf{r})$, the exchange and correlation potential, is given by $\delta E_{XC}[\rho]/\delta\rho(\mathbf{r})$, and the ψ_i's satisfy:

$$[-\frac{1}{2}\nabla^2 + v_{\text{eff}}(\mathbf{r})]\psi_i = \epsilon_i\psi_i,$$

$$\rho(\mathbf{r}) = \sum_{i=0}^{N} \mid \psi_i(\mathbf{r}) \mid^2, \tag{2.15}$$

the Kohn-Sham equations. This reformulation has performed two tricks: firstly, it has made the kinetic energy which is calculated the kinetic energy for a *non-interacting* electron gas (which is easy to find), and secondly all the non-classical part of the electron-electron interactions has been put into the exchange and correlation functional (along with the correction associated with the kinetic energy simplification). So, the exchange and correlation energy can now be written:

$$E_{XC}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho], \tag{2.16}$$

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{\mid \mathbf{r}' - \mathbf{r} \mid}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}', \tag{2.17}$$

where $T_s[\rho]$ is the kinetic energy for a non-interaction electron gas, $T[\rho]$ is the kinetic energy for the full electron gas, and $J[\rho]$ is the Coulomb integral for the electrons. Now all the difficult calculations have been placed into a single term: $E_{XC}[\rho]$. It is the calculation of this term which introduces the approximations generally associated with DFT, a subject which will be dealt with in the next section; until this point, DFT is an exact reformulation of the many-body problem.

2.2.1 Approximating the exchange and correlation energy

The whole area of electron-electron interactions, and many-body theory, is enormously complex; all that I am aiming to present in this section is an overview of those areas which impinge on DFT. I also do not claim to be an expert in the field; to quote Parr and Yang (1989), "The problem remains, and this problem is of surpassing difficulty, of how to calculate T[ρ], and how to calculate the nonclassical part of $V_{ee}[\rho]$."

Conceptually, the exchange energy of a system of electrons is associated with the Pauli principle: because two electrons of the same spin must have a spatial separation, the electron-electron repulsion energy is reduced, and this is known as the *exchange* energy. The correlation energy is harder; it is actually defined as the difference between the correct many-body energy and the Hartree-Fock energy (which puts in the exchange integral between two electrons, as given above in Eq. 2.17 but assumes that there is *no* correlation).

The simplest approach taken in DFT is the Local Density Approximation (LDA). Here, the assumption is made that for each point in space with a density $\rho(\mathbf{r})$, the exchange and correlation energy *at that point* is the same as for a uniform electron gas with that density, $\rho(\mathbf{r})$. That is,

$$E_{XC}^{LDA}[\rho] = \int \rho(\mathbf{r})\epsilon_{XC}(\rho)\mathrm{d}\mathbf{r}, \tag{2.18}$$

$$V_{XC}^{LDA}(\mathbf{r}) = \frac{\delta E_{XC}^{LDA}}{\delta\rho(\mathbf{r})} \tag{2.19}$$

$$= \epsilon_{XC}(\rho) + \rho(\mathbf{r})\frac{\delta\epsilon_{XC}(\rho)}{\delta\rho}, \tag{2.20}$$

where $\epsilon_{XC}(\rho)$ is the exchange and correlation energy for a uniform electron gas of density ρ. Various parameterisations exist for the LDA, which give extremely similar results for total energy. These are all based on calculations for the exchange and correlation energy of a homogeneous electron gas, which can be split into the exchange energy (which can be written analytically) and the correlation energy (which can't). The Dirac exchange formula (1930) for a uniform electron gas of density ρ is:

$$E_x^{\mathrm{LDA}}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int \rho(\mathbf{r})^{4/3}\mathrm{d}\mathbf{r}. \tag{2.21}$$

The correlation energy is generally interpolated analytically, based on a series of quantum Monte Carlo calculations made by Ceperly and Alder (1980) for the exchange and correlation of the electron gas.

Many researchers have found that the major source of error in the LDA is in the exchange energy (see Section 8.7 of Parr and Yang (1989) for references and a more detailed discussion of the rest of this section). By considering the gradient of the density, as well as the value, at a given point, it is possible to make the exchange energy which is calculated more accurate. There have been a number of so-called Generalised Gradient Approximations (GGAs) proposed (Perdew 1985; Perdew and Yue 1986; Lee, Yang and Parr 1988; and Perdew and Wang 1991), which use an entirely empirical form to include the gradient of the electron gas as well as its value in the correlation calculation. In those calculations which use GGA in this thesis, the Perdew and Wang (1991) functional has been used. To give a flavour of the added complication which these gradient functionals introduce, the Perdew and Yue (1986) functional (which is relatively simple) is given below:

$$E_x^{\mathrm{GGA}}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int \rho(\mathbf{r})^{4/3}F(s)\mathrm{d}\mathbf{r}, \tag{2.22}$$

$$s = \frac{\mid\nabla\rho(\mathbf{r})\mid}{(2k_F\rho(\mathbf{r}))}, \tag{2.23}$$

$$k_F = (3\pi^2\rho(\mathbf{r}))^{1/3}, \tag{2.24}$$

$$F(s) = (1 + 1.296s^2 + 14s^4 + 0.2s^6)^{1/15}. \tag{2.25}$$

While LDA is acknowledged generally to overbind (in other words, the adsorption energies found with LDA are too high, and the diffusion barriers generally too low), and all of the GGAs reduce this problem (see, for instance, Hammer, Jakobsen and Nørskov 1993), they are still not an entirely adequate solution. A

recent study (Nachtigall *et al.* 1996) has found that the Perdew-Wang (1991) functional can underestimate the barrier for H_2 elimination from Si_2H_6 when compared to accurate, quantum chemical (configuration interaction) calculations. Even so, it improves the calculation of reaction barriers, and is a reasonable GGA functional to use - the Lee, Yang and Parr (1988) is found to do slightly better, though Nachtigall *et al.* found that the best fit was obtained by using a 3-centre exchange functional proposed by Becke (1988). Another question which arises is that of self-consistency: is it better to achieve a self-consistent gradient-corrected charge density, or to correct *post hoc* ? It seems that the *post hoc* correction (i.e. calculating the GGA energy and the LDA energy for the charge density *after* the relaxation is complete, and applying that difference as a correction) is entirely adequate for the calculation of reaction barriers (Hammer, Jakobsen and Nørskov 1993 and White, Bird, Payne and Stich 1994) and this approach has been taken in this thesis. The development of new schemes for the exchange and correlation energy, particularly if the basis set for DFT calculations shifts from plane waves to localised orbitals, will be extremely interesting, and very important, for the materials modelling community.

2.2.2 Pseudopotentials

To calculate the electronic wavefunctions for a relatively large slab including all the electrons is an extremely hard task; in an element such as silicon, where the outer, or valence, electrons are in the 3s and 3p orbitals, their wavefunctions must be orthogonal to those of the inner, core states due to the Pauli exclusion principle. This entails introducing nodes into these wavefunctions, which in turn gives them a high kinetic energy in the core region, requiring a high plane wave cutoff (a concept which will be explained in the next section) and increasing computational time enormously. What is done in practice to make the calculation easier is to replace both the inner electrons and potential with another potential, because the orthogonality requirement, due to the inner, core electrons, means that the nuclear charge has less effect on the outer electrons - they are *screened*. There is a close cancellation between the attractive, Coulomb potential from the nucleus and the repulsive orthogonality requirement from the core electrons. This can be taken advantage of, by replacing the strong, nuclear Coulomb potential inside a certain radius with a weaker *pseudopotential*, and modelling only the valence electrons (which should now be referred to more properly as *pseudoelectrons*). This is illustrated schematically in Fig. 2.3. There is a great deal more to pseudopotentials, which is discussed in many places, such as Cohen and Heine (1970) and Yin and Cohen (1982); two of the more important aspects will be touched on briefly below.

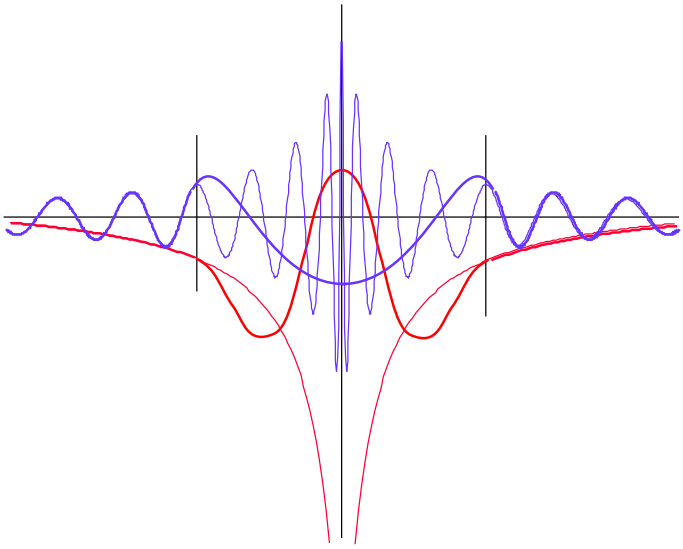

Figure 2.3: A schematic representation of the potentials (red lines) and wavefunctions (blue lines) for an atom. The real potential and wavefunction are shown with thin lines, while the pseudopotential and wavefunction are shown in thick lines. Outside the cutoff region (vertical black lines) the two are identical. *Picture courtesy of Chris Goringe.*

Electrons with different angular momentum components (e.g. s or p electrons) will scatter differently from

an atomic potential (or a pseudopotential). The best way to correctly represent this different scattering with a pseudopotential is to use different projectors and potentials for different angular momentum components:

$$V_{NL} = \sum_{lm} \mid lm\rangle V_l \mid lm \mid, \tag{2.26}$$

where V_l is the pseudopotential for angular momentum component l. This form of pseudopotential is known as a *non-local* pseudopotential, and it allows the maintenance of "softness" of a pseudopotential (i.e. allowing the plane wave cutoff to be kept low) while correctly reproducing the scattering and phase shift of the electrons.

Another key feature of a pseudopotential is that the wavefunctions and charge density outside the core region should be identical to those produced by the genuine ionic potential. If the pseudo-charge density within the core is made to be equal to the real charge density in the core, then this condition, known as *norm conservation*, is fulfilled. Norm-conserving pseudopotentials were first introduced by Starkloff and Joannopoulos (Joannopoulos *et al.* 1977, Starkloff and Joannopoulos 1977) for local pseudopotentials only, and accurately described the valence energies and wave functions of many heavy atoms accurately. The method has been extended to non-local pseudopotentials by several groups (Redondo *et al.* 1977, Hamann *et al.* 1979, Zunger and Cohen 1979, Kerker 1980, Bachelet *et al.* 1982). As Hamann *et al.* (1979) observed, the matching of the real and pseudo wavefunctions outside the core guarantees that the first order energy dependence of the scattering is correct, so that the scattering is described correctly for a wide range of energies. The most recent development in pseudopotentials involves relaxing the norm conservation criterion, in order to reduce further the plane wave cutoff required (in particular for extremely hard elements such as oxygen)(Vanderbilt 1990). This in turn requires an energy-dependent potential to reproduce the energy scattering properties for the same wide range of energies. The use of Vanderbilt pseudopotentials (also known as ultrasoft pseudopotentials) can speed up calculations by a factor of between 2 and 10 (Georg Kresse, Private Communication).

2.2.3 A plane wave code: CASTEP

The question of how best to represent the electronic wavefunctions in the calculation is a vexed one, which has had many solutions: Plane Waves, Orthogonalised Plane Waves, Linearised Muffin-Tin Orbitals, Gaussian Orbitals, Projector Augmented Waves and B-splines, to mention just a few. The LDA and *post hoc* GGA calculations performed in the course of this thesis used the plane wave, DFT code CASTEP (Payne *et al.* 1992). The wavefunctions are expanded out to a maximum **G** vector, which is equivalent to an energy cutoff, as plane waves:

$$\psi_j = \sum_{\mathbf{G}<\mathbf{G}_{max}} c^j_{\mathbf{k}} e^{i(\mathbf{k}+\mathbf{G}\cdot\mathbf{r})}, \tag{2.27}$$

where **k** is the k-point being considered in the Brillouin zone. In this thesis, a plane wave cutoff of 115 eV is used for Si; this is entirely adequate for energy difference convergence. When hydrogen is being modelled, a cutoff of 200 eV is used, which is also adequate for energy difference convergence (as the memory and time requirements for calculations scale exponentially with cutoff, the lowest cutoff possible is generally used). The single k-point of (0,0.25,0) was used for all calculations. This code uses non-local pseudopotentials of the Kerker type (Kerker 1980) in the Kleinman-Bylander form (1982). In all simulations performed herein, structures were considered relaxed when the force was less than 0.01 Å/eV, and the energy was changing by less than 1 meV.

CASTEP and its sister code CETEP (a parallel implementation) have been used successfully to model many systems, including dislocations in Si (Bigger *et al.* 1992), the Si(113) and (111) surfaces (Bird *et al.* 1992; Stich *et al.* 1992) and chlorine dissociation on Si(111) (De Vita *et al.* 1993, Stich *et al.* 1993); the GaAs(001) surface and adsorption of trimethyl gallium (Goringe *et al.* 1997); lithium and magnesium oxide defects (De Vita *et al.* 1992) and magnesium oxide surface reactions (Pugh and Gillan 1994); and dissociation of H_2 at the Cu(100) and Mg(0001) surfaces (White and Bird 1993; Bird *et al.* 1993).

Almost all of the calculations in the thesis have used a standard unit cell: two dimer rows wide, with two dimers in each row; and five layers of silicon deep, with the bottom layer frozen in bulk-like positions, and terminated in hydrogen. The pseudopotential used for silicon was a standard non-local pseudopotential, with the *s*-component local, while the hydrogen potential was simply a 1/r local dependence.

2.3 Tight binding

Tight binding is a semi-empirical method for electronic structure calculations. While it retains the quantum mechanics of the electrons, the Hamiltonian is parameterised and simplified before the calculation, rather than constructing it from first principles. The method was initially due to Slater and Koster (1954), in which the ground work was laid. A recent review summarises the different approaches to tight binding, and its applications (Goringe, Bowler and Hernández 1997).

The calculation can be split into two parts: the electronic part (which is addressed by solving the equation $H\psi = E\psi$) and the repulsive terms, which are added to obtain the correct total energy and forces. A clear justification of the method has been provided by Sutton *et al.* (1988), where it is shown, from density functional theory, that the total energy can be split up into a band structure term (the electronic part of the calculation) and a sum of pair-like terms (which account for electron-electron repulsion, etc.). In general, the matrix elements and repulsive potentials are further split into an equilibrium value (which is easily fitted to the bulk crystal properties) and a scaling term (which is fitted to a variety of properties).

Conceptually, tight binding works by postulating a basis set which consists of atomic-like orbitals (i.e. they share the angular momentum components of the atomic orbitals, and are easily split into radial and angular parts) for each atom in the system, and the Hamiltonian is then parameterised in terms of various high symmetry interactions between these orbitals. For the work in this thesis, which considers tetrahedral semiconductors, a conceptual basis set of 1 s-like orbital and 3 p-like orbitals has been used. In the most common form of tight binding (nearest neighbour, orthogonal TB) the orbitals are *assumed* to be orthogonal[2] and interactions between different orbitals are only allowed to be non-zero within a certain distance, which is placed somewhere between the first and second nearest neighbours in the crystal structure. A further simplification which is made is to neglect three-centre integrals (i.e. an interaction between orbitals on atoms A and B which is mediated by the potential on atom C), meaning that each interaction is a function of the distance between the atoms only. Pettifor (1977) showed that these interactions were about 10% of the total interaction for d-band metals, which, while not justifying the neglect, gives a quantitative picture of the error incurred.

So, the Schrödinger equation $H\Psi^{(n)} = E\Psi^{(n)}$ needs to be rewritten. First of all, the eigenvector $\Psi^{(n)}$ is expanded out in terms of the basis functions:

$$\Psi^{(n)} = \sum_{i\alpha} c_{i\alpha}^{(n)} \phi_{i\alpha}, \tag{2.28}$$

where i covers the atoms and α the orbitals on each atom, and $\phi_{i\alpha}$ is an orbital on atom i. The Hamiltonian can then be written as a matrix which must be fitted to experimental or *ab initio* data, as

$$H_{i\alpha,j\beta} = \langle i\alpha \mid \hat{H} \mid j\beta \rangle, \tag{2.29}$$

where $\mid i\alpha\rangle$ represents the state $\phi_{i\alpha}$. Typically the elements $H_{i\alpha,j\beta}$ are split into electronic and scaling parts, which enables the electronic part to be fitted to the crystal at equilibrium, and the scaling terms to be fitted separately (typically to elastic constants).

There are two main types of electronic interaction which need to be parameterised: on-site and inter-site. The on-site integrals are relatively easy to understand, and represent the energies of the orbitals:

$$\begin{aligned} \langle \phi_s \mid \hat{H} \mid \phi_s \rangle &= E_s, \\ \langle \phi_p \mid \hat{H} \mid \phi_p \rangle &= E_p. \end{aligned} \tag{2.30}$$

The inter-site interactions are easiest understood if the bond is assumed to be along the x-axis (the procedure for constructing any inter-site interaction from the different integrals given below, due to Slater and Koster (1954), is explained in full in Chapter 3):

$$\begin{aligned} \langle \phi_{i,s} \mid \hat{H} \mid \phi_{j,s} \rangle &= h_{ss\sigma}, \\ \langle \phi_{i,s} \mid \hat{H} \mid \phi_{j,p_x} \rangle &= h_{sp\sigma}, \\ \langle \phi_{i,p_x} \mid \hat{H} \mid \phi_{j,p_x} \rangle &= h_{pp\sigma}, \end{aligned}$$

[2]This assumption is justified elegantly in Slater and Koster (1954). Essentially, they argue that new orbitals, which are orthogonal and share the symmetry properties of the original orbitals, can be made by linear combinations of non-orthogonal atomic orbitals.

$$\langle \phi_{i,p_y} \mid \hat{H} \mid \phi_{j,p_y} \rangle = h_{pp\pi},$$

$$\langle \phi_{i,p_z} \mid \hat{H} \mid \phi_{j,p_z} \rangle = h_{pp\pi}, \tag{2.31}$$

and all other interactions are zero. The parameterisation of the Hamiltonian is vital to the success of a calculation, and will be dealt with fully in Chapter 3.

The details given above apply only to the electronic energy, which allows the band energy to be found from $E_{band} = 2\sum_i \epsilon_i$, where ϵ_i are the eigenvalues of the occupied states of the Hamiltonian. To obtain a total energy for the system being modelled, the effects of ion-ion repulsion, and those parts of the electron-electron interaction neglected above, need to be accounted for. This is most often done via a pair potential, which is again fitted to *ab initio* data or experiment.

The most common way of solving the Schrödinger equation, certainly until recently, was by diagonalisation of the Hamiltonian, using Bloch functions and a sum over k-points in the Brillouin zone to correctly find the energy. A method for choosing special k-points for good sampling has been given by Monkhorst and Pack (1976). The drawback with direct diagonalisation is that it scales with the cube of the number of atoms in the system, known as $O(N^3)$ scaling, leading to prohibitive time and memory requirements for large unit cells. There are advantages to diagonalisation in k-space, which include the automatic generation of the wavefunctions. In this thesis, I have used direct diagonalisation only to generate local densities of states. All other work has been performed using a tight binding method known as the density matrix method. This is one of a class of methods which have $O(N)$ scaling, and will be described in the next section.

The justification for the tight binding method, and particularly for the pair potential form of the repulsion which must be added, has been given elegantly by Sutton *et al.* (1988). Starting from DFT, they use a non-self-consistent approximation known as the Harris-Foulkes approximation (Harris 1985; Foulkes 1987), which argues that because of the variational nature of DFT, any error in an input charge density causes a second order error in the output charge density, to produce a reformulation of DFT which corresponds to the different terms in the TB total energy, and show that those terms which do not correspond to the band energy can be written (approximately) as pair potentials. This important paper put the common forms of tight binding on a rigorous theoretical footing.

The tight-binding calculations performed in this thesis have used a standard unit cell, in terms of depth and often length: one dimer row with six dimers; and ten layers of silicon deep, with the bottom five layers frozen in bulk-like positions. The tight-binding parameterisations described in Chapter 3 were used for all interactions. Unless stated, the density matrix method (which is described in Section 2.5) has been used, with a cutoff defined by 3 hops.

2.4 Linear scaling tight binding methods

As described above, TB has one major disadvantage: the computational time scales with the cube of the number of atoms. Recently, many methods have been proposed which improve on this (Li, Nunes and Vanderbilt 1993; Daw 1993; Galli and Parrinello 1992; Mauri, Galli and Car 1993; Ordejón *et al.* 1993; Pettifor 1989; Aoki 1993; Goedecker and Colombo 1994; Kress and Voter 1995; Horsfield 1996; Stechel, Williams and Feibelman 1994), and break down into two basic categories: variational methods and moments methods. Variational methods, as their name suggests, seek to minimise the total energy with respect to some electronic degrees of freedom, while moments methods use an elegant theorem which relates the electronic structure of an atom to the local environment. Computer codes which implemented four of these methods were readily available in Oxford: the Density Matrix Method (Li, Nunes and Vanderbilt 1993; Daw 1993), the Global Density of States method (Horsfield 1996), the Fermi Operator Expansion method (Goedecker and Colombo 1994;Voter, Kress and Silver 1996) and the Bond Order Potential method (Pettifor 1989, Aoki 1993). The key question to be addressed was which method was most suitable for the systems to be studied. To this end, and to understand more generally the applicability of each method, a number of tests were applied: cohesive energies for bulk diamond, silicon and titanium; the *unrelaxed* vacancy formation energy for diamond, silicon and titanium; the forces for the unreconstructed (001) surface of silicon; and the time required to obtain a given level of accuracy for the vacancy in diamond and titanium. Further tests have been performed, and the interested reader is referred elsewhere (Bowler *et al.* 1997) for the full investigation. In this section, a brief justification of the idea underlying all linear scaling methods (including more recent density functional methods) will be given, followed by a description of each of the four methods to be compared. The comparison is given in Chapter 3.

All of these methods rely on a piece of chemical intuition, which is that covalent bonding is local. There is no rigorous theoretical proof for this, but the decay of Wannier functions (functions which resemble the postulated TB orbitals closely) which is exponential in semiconductors and a power law in metals in one dimension (Kohn 1959) is sometimes quoted. What is offered here is a demonstration of localisation, shown by the band energy for an atom at the centre of a cluster of different sizes, solved by exact diagonalisation. The clusters are made up of atoms which all lie within a certain number of hops of the centre, an idea which relates to the moments methods which will be described in Section 2.6. The error in band energy relative to the perfect bulk value is shown in Fig. 2.4 for three materials: titanium, silicon and carbon. As can be

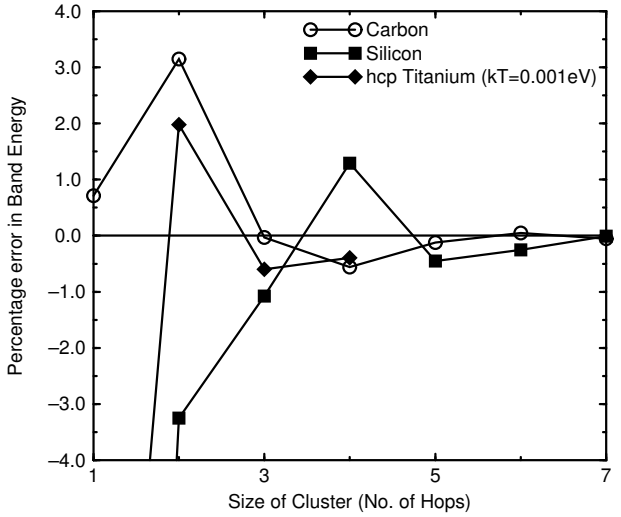

Figure 2.4: The error in band energy for the atom at the centre of a cluster for different cluster sizes.

seen, and as would be expected for an insulator, the energy for carbon is converged well at three shells, and almost completely converged by 5 shells. Silicon, which is a semiconductor, and would therefore be expected to have a longer range interaction, is showing good convergence at 5 hops, and full at 7 hops. The titanium plot, which stops at 4 hops because of computer memory restrictions, shows some convergence, but is likely to take some time to converge, due to the delocalised nature of metallic bonding.

A more quantitative way of showing the localisation of bonding is to examine the density matrix between the central atom and all other atoms in a cluster. The density matrix element between two atoms can be defined as:

$$\rho_{ij} = \sum_n \sum_{\alpha,\beta} c^n_{i\alpha} c^n_{j\beta}, \tag{2.32}$$

where n indexes the eigenvalue number. The more basic definition of the density matrix element between two orbitals is $\rho_{i\alpha,j\beta} = \sum_n c^n_{i\alpha} c^n_{j\beta}$; the above definition between atoms is then just the sum over α and β, and the trace of ρ_{ij} gives the total number of electrons in the system. The density matrix element between atoms in a cluster is shown in Fig. 2.5 for carbon, silicon and titanium.

Figure 2.5a shows the results for carbon and silicon, which are expected to be similar, as they both have band gaps. Clearly, the bonding in carbon is extremely localised, with almost no electrons between atoms which are more than 3 hops apart. Silicon is less extremely localised, but it should be noted that the variations beyond 3 bond lengths are of the order of 0.01 electrons. Beyond about 5 hops, these values will be affected by edge effects from the cluster, and will therefore be a little unreliable. Figure 2.5b shows the same plot for titanium at high and low electronic temperature. As is expected for a metal at low electronic temperature, there are still large fluctuations of the density matrix at relatively long range, reflecting the delocalised band structure of the metal. However, going to high electronic temperature localises the bonding, and makes the assumptions behind the order N methods valid. The danger of this approach is that there is a large electronic entropy contribution to the energy of the system. This can be overcome by using the Gillan functional (Gillan 1989) to extrapolate back to low temperatures for the energy.

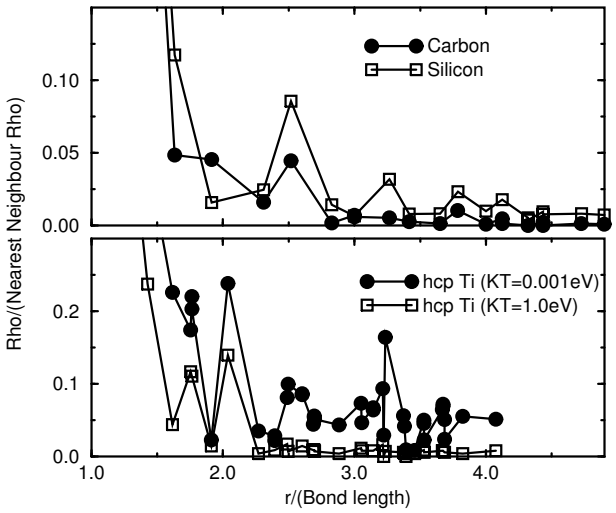

Figure 2.5: The density matrix between the atom at the centre of a cluster and all the other atoms in the cluster for the largest cluster available, for C, Si and Ti. (a) C and Si. The nearest neighbour elements are 0.217 and 0.194 electrons respectively. (b) Ti. Note that the plots are at different electronic temperature.

The assertion that covalent bonding is localised has been illustrated in two different ways, and seems to be reasonable. Metallic bonding can be induced to be local by increasing electronic temperature, though this can lead to problems in total energy. With this understanding, the methods which exploit it can be presented.

2.5 The density matrix method

The density matrix method (DMM) was simultaneously proposed by Li, Nunes and Vanderbilt (1993) and by Daw (1993), though from different arguments. It revolves around the density matrix operator,

$$\hat{\rho} = \sum_i |\psi_i\rangle f(E_i) \langle\psi_i|, \tag{2.33}$$

which is the operator notation for the density matrix defined in Eq.2.32 in Section 2.4. This acts as a projector onto the occupied subspace of eigenvalues; it has eigenvalues of zero (unoccupied) or one (occupied). In the matrix notation where the eigenvectors are expanded in terms of atomic basis functions, the elements represent a number of electrons between basis functions, which can be summed to give the number of electrons in a bond. The number of electrons in the system, the band energy for kT=0 (U) and the corresponding contribution to the forces from the band energy can be written in terms of $\hat{\rho}$:

$$N_e = 2Tr[\hat{\rho}] \tag{2.34}$$

$$U = 2Tr[\hat{\rho}\hat{H}] \tag{2.35}$$

$$\mathbf{F}_i = -2Tr\left[\hat{\rho}\frac{\partial \hat{H}}{\partial \mathbf{R}_i}\right], \tag{2.36}$$

where Tr indicates taking the trace of a matrix. It is important to note that the trace of a matrix is independent of the basis set used to evaluate it. The above results follow from the fact that in the basis set which diagonalises the Hamiltonian, the density matrix only has non-zero elements on the diagonal where states are occupied - as follows from the discussion above. The ground state energy for kT=0 can be found by minimising U with respect to the elements of some initial ρ subject to two constraints: idempotency of the density matrix ($\hat{\rho}^2 = \hat{\rho}$) and constant number of particles (N_e = constant). In the density matrix

method, the elements of a trial density matrix are used as variational degrees of freedom with respect to which the energy is minimised. To impose idempotency on the density matrix, which is equivalent to it having eigenvalues of 0 or 1, it is replaced with the result of the McWeeney transformation (McWeeney 1960) of this trial density matrix,

$$\tilde{\rho} = 3\sigma^2 - 2\sigma^3, \tag{2.37}$$

where σ is identified as the *trial* density matrix and $\tilde{\rho}$ is identified as the *physical* density matrix, such that the expectation value of an operator $\hat{A}$ is given by $Tr[\tilde{\rho}\hat{A}]$. At the minimum energy, the McWeeney transformation forces $\tilde{\rho}$ to have eigenvalues of 0 or 1, and the band energy has a single local minimum where $\tilde{\rho}$ is equal to the true density matrix, $\hat{\rho}$, as given by Eq. 2.33. There are also runaway solutions, but these can be avoided easily by explicit construction.

The second constraint is easily achieved by varying the chemical potential, μ, at each step. In the implementation used in this work, the chemical potential for electrons (μ) is introduced into the minimisation by working with the grand potential, $\Omega = U - \mu N_e = 2Tr[\tilde{\rho}(H - \mu)]$. Minimising the energy, and fixing the number of electrons, are then performed in a concerted fashion (Goringe 1995), with respect to the elements of the trial density matrix, σ.

As described so far, this method is exact, and not linear scaling. In order to achieve linear scaling, the density matrix method takes advantage of the fact that the elements of the density matrix between two atoms tend to zero as the distance between them tends to infinity, as described in the previous section. A cutoff radius (R_c) is defined beyond which all elements of the trial density matrix are set to zero. This leads to a sparse density matrix, which gives linear scaling because the number of elements in the variational minimisation is proportional to L^2N, where N is the number of elements in the system and L is the number of atoms in the sphere.

However, imposing this cutoff leads to a density matrix which is no longer exactly idempotent, though the McWeeney transformation reduces the error. If the trial density matrix has idempotency errors to first order, then the physical density matrix will have idempotency errors to second order (this point is discussed in more detail in Goringe (1995)). The transformation forces the eigenvalues of the density matrix which emerges to be clustered about 0 and 1, rather than exactly equal to 0 and 1, as would be the case for an idempotent operator.

The method is variational which means that all non-exact density matrices give an energy above the exact energy, so $\Omega \geq \Omega_{exact}$; also, as more degrees of freedom are added (i.e. as the cutoff is extended), the value of Ω will converge to the exact value. More importantly, the Hellmann-Feynman theorem is obeyed by the DMM, which means that the forces are exactly equal to the numerical derivative of the energy; this is vital when trying to obtain the minimum energy for a given configuration of atoms. It should be noted that in the implementation of this method used for this work, the cutoff is defined not by a radius, but by the number of hops away from an atom. This enables easy comparisons with moments methods (see Chapter 3) and is also much faster than the methods described in Goringe (1995).

2.6 Recursion methods

Recursion methods, also known as moments methods, take advantage of the fact that the local density of states of an atom depends upon the local environment. The *total* band energy for a system of atoms can be obtained from the *total* density of states (DOS) for the system of atoms, $n(E)$:

$$E_{band} = \int^{E_f} En(E)\mathrm{d}E. \tag{2.38}$$

However, this is a *global* property of the system; the *global* density of states can be decomposed into a set of *local* densities of states (LDOS) (Friedel 1954), thus:

$$n(E) = \sum_{i\alpha} n_{i\alpha}(E), \tag{2.39}$$

where $n_{i\alpha} = \sum_k c_{i\alpha}^{(k)} \delta(E - E_k) c_{i\alpha}^{(k)}$. For a finite basis set, the LDOS has a finite width, a definite mean, and a shape. These properties, and hence the entire LDOS, can be fully defined by its *moments* (Heine 1980). The p^{th} moment of the local density of states, $n_{i\alpha}$, is $\mu_{i\alpha}^{(p)}$, where

$$\mu_{i\alpha}^{(p)} = \int E^p n_{i\alpha}(E) dE. \tag{2.40}$$

There is a useful identity (Ducastelle and Cyrot-Lackmann 1970) which enables a correspondence between the moment and powers of the Hamiltonian to be drawn:

$$\mu_{i\alpha}^{(p)} = \int E^p n_{i\alpha}(E) dE = \langle i\alpha \mid \hat{H}^p \mid i\alpha \rangle, \tag{2.41}$$

$$\mu_{i\alpha}^{(p)} = \sum_{j_1\beta_1,\ldots,j_{p-1}\beta_{p-1}} H_{i\alpha,j_1\beta_1} H_{j_1\beta_1,j_2\beta_2} \ldots H_{j_{p-1}\beta_{p-1},i\alpha}, \tag{2.42}$$

which corresponds to hopping around a lattice along closed loops of length p, if $H_{i\alpha,j\beta}$ is the hopping parameter for orbital $i\alpha$ to $j\beta$. This identity shows a simple connection between the local bonding of an atom and its electronic structure.

While the moments of a function seem to offer a promising way of reconstructing a function, most methods are unstable. However, the recursion method (Haydock 1980) is an important exception. This is a Green's function method for building the LDOS from a set of moments.

The LDOS is given in terms of the Green's function $G_{i\alpha,i\alpha}(Z)$ by:

$$n_{i\alpha}(E) = -\frac{1}{\pi} \lim_{\eta \to 0} Im\{G_{i\alpha,i\alpha}(E + i\eta)\}. \tag{2.43}$$

This form for the LDOS is useful, because the Green's function can be written in terms of a continued fraction, whose components are related to the elements of the tridiagonalised Hamiltonian of the system (Haydock 1980) [3]; the element $G_{00}(Z)$ (where $G_{nm}(Z) = \langle U_n \mid \hat{G}(Z) \mid U_m \rangle$) can be found from:

$$G_{00}(Z) = \cfrac{1}{Z - a_0 - \cfrac{b_1^2}{Z - a_1 - \cfrac{b_2^2}{Z - a_2 - \cfrac{b_3^2}{\ddots}}}}. \tag{2.44}$$

The Lanczos recursion algorithm (Lanczos 1950) is an efficient scheme for tridiagonalising a matrix. Consider a matrix $\mathbf{H}$, which corresponds to the operator $\hat{H}$. Let the tridiagonal matrix be $\mathbf{H'}$, whose diagonal elements are given by a_n and whose off-diagonal elements are given by b_n. If the states which tridiagonalise the matrix are $\mid U_n \rangle$, then:

$$H'_{mn} = \langle U_m \mid \hat{H} \mid U_n \rangle = \begin{cases} a_n, & \text{if } m = n; \\ b_n, & \text{if } m = n-1; \\ b_{n+1}, & \text{if } m = n+1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.45}$$

The Lanczos recursion algorithm for tridiagonalising a matrix is based on the following recursion relation:

$$\hat{H} \mid U_n \rangle = a_n \mid U_n \rangle + b_n \mid U_{n-1} \rangle + b_{n+1} \mid U_{n+1} \rangle \tag{2.46}$$

and the condition that the tridiagonalising states are orthonormal ($\langle U_n \mid U_m \rangle = \delta_{n,m}$). To apply this relation to a matrix, a starting state (which will generally be a TB basis function) is chosen. This gives a_0 immediately from Eq. 2.45 ($a_0 = \langle U_0 \mid \hat{H} \mid U_0 \rangle$). The construction requires that b_0 be zero, so the product $b_1 \mid U_1 \rangle$ can be found from Eq. 2.46. From the orthonormality imposed on $\mid U_n \rangle$, the value of b_1 can be obtained, and $\mid U_1 \rangle$ can be found. This then can be applied to the process again, giving a_1, b_2 and $\mid U_2 \rangle$ and so on. If $\mid U_0 \rangle$ is set equal to a basis function, $\mid \psi_{i\alpha} \rangle$, then $G_{i\alpha i\alpha}$ can be obtained from the continued fraction given in equation 2.44, and from $G_{i\alpha i\alpha}$ the LDOS can be obtained from Eq. 2.43.

For an infinite system, there are an infinite number of levels in the continued fraction, so the exact values are replaced by estimated ones after a certain level. See Beer and Pettifor (1984) for details of the simplest of these terminations.

Since the above may not bear any obvious relation to a moments method, it is useful to make the following observation:

$$\mu_{i\alpha}^{(n)} = \langle i\alpha \mid \hat{H}^n \mid i\alpha \rangle$$

[3] A tridiagonal matrix has non-zero elements only on the diagonal and one column either side.

$$
\begin{aligned}
&= \langle U_0 \mid \hat{H}^n \mid U_0 \rangle \\
&= \sum_{m_1 \ldots m_{n-1}} \langle U_0 \mid \hat{H} \mid U_{m_1} \rangle \langle U_{m_1} \mid \hat{H} \mid U_{m_2} \rangle \ldots \langle U_{m_{n-1}} \mid \hat{H} \mid U_0 \rangle. \qquad (2.47)
\end{aligned}
$$

The first few moments defined by the above are:

$$
\begin{aligned}
\mu_{i\alpha}^{(0)} &= 1, \\
\mu_{i\alpha}^{(1)} &= a_0, \\
\mu_{i\alpha}^{(2)} &= a_0^2 + b_1^2, \\
\mu_{i\alpha}^{(3)} &= a_0^3 + 2a_0 b_1^2 + a_1 b_1^2, \\
\mu_{i\alpha}^{(4)} &= a_0^4 + 3a_0^2 b_1^2 + 2a_0 a_1 b_1^2 + a_1^2 b_1^2 + b_1^2 b_2^2 + b_1^4. \qquad (2.48)
\end{aligned}
$$

These equations can be inverted to give the recursion coefficients in terms of the moments. Every extra moment that is calculated allows another recursion coefficient to be found. However, such an inversion becomes numerically unstable beyond about 20 moments. It is often desirable to go higher than this in numbers of moments; in such a case, Eq. 2.46 is used, as it is stable for many levels.

The recursion method is therefore an efficient Green's function method for the density of states. The band energy can be written in terms of a Green's function as:

$$
E_{band} = -\frac{1}{\pi} \sum_{i\alpha} \lim_{\eta \to 0} \mathrm{Im} \int^{E_f} \mathrm{d}E G_{i\alpha,i\alpha}(E + \mathrm{i}\eta) E, \qquad (2.49)
$$

with $G_{i\alpha,i\alpha}$ obtained from Eqs. 2.44-2.46.

Recursion and forces The derivation of forces from the recursion method is extremely helpful in understanding the formalism of the Bond Order Potential method (BOP), and also shows several of the problems with the recursion method. The obvious manner of obtaining the contribution to the force on an atom from the band energy is to differentiate the band energy with respect to the atomic coordinate.

However, this is extremely computer intensive for more than a few levels of recursion (see Section 2.6.1 for more details). An alternative is to use the density matrix form of the force, Eq. 2.36. This is possible because the density matrix element $\rho_{i\alpha,j\beta}$ can be written in terms of a Green's function element, $G_{i\alpha,j\beta}$:

$$
\rho_{i\alpha,j\beta} = -\frac{1}{\pi} \lim_{\eta \to 0} \mathrm{Im} \int^{E_f} \mathrm{d}E G_{i\alpha,j\beta}(E + i\eta). \qquad (2.50)
$$

The problem of how to find this Green's function element is very important. The approach taken in the recursion method starts from bonding and anti-bonding states:

$$
\begin{aligned}
\mid + \rangle &= \frac{1}{\sqrt{2}} \left(\mid i\alpha \rangle + \mid j\beta \rangle \right) \\
\mid - \rangle &= \frac{1}{\sqrt{2}} \left(\mid i\alpha \rangle - \mid j\beta \rangle \right) \qquad (2.51)
\end{aligned}
$$

From these two states, the Green's function elements G_{++} and G_{--} can be found using the recursion technique. If the states $\mid + \rangle$ and $\mid - \rangle$ are expanded out, and G_{--} is subtracted from G_{++}, then the following expression for $G_{i\alpha,j\beta}$ is arrived at:

$$
G_{i\alpha,j\beta} = \frac{1}{2} \left[G_{++} - G_{--} \right]. \qquad (2.52)
$$

The result just derived is known as the inter-site (IS) method, while that derived in the previous section is the site diagonal (SD) method, as it works with $G_{i\alpha,i\alpha}$. Both of these methods have severe problems -

the problem with forces from the SD method has been alluded to above, while the IS method has appalling convergence for calculation of the bond energy (Glanville *et al.* 1988), and a symmetry problem, which means that the energy and forces are not invariant for rotation about crystal axes (Inoue and Ohta 1987). These problems have led to a variety of different solutions, the first of which was the matrix recursion method (MRM) (Inoue and Ohta 1987; Paxton, Sutton and Nex 1987; Paxton and Sutton 1989; Jones and Lewis 1984). In the MRM, the recursion coefficients are matrices rather than scalars (typically each matrix represents the orbitals on one atom), which removes the problem of rotational invariance; it does not, however, solve the problem of the IS convergence (Aoki 1993). For further information on the MRM, see the references given above. It is extremely important, but very poorly convergent. The improvement of the convergence, and remedy for all the problems mentioned at the end of the previous section, led to the development of BOP. Before looking at BOP, two alternative moments approaches will be examined.

2.6.1 The global density of states method

It has been shown in Eq. 2.49 that the band energy can be written in terms of a one particle Green's function, $G_{i\alpha,i\alpha}$. However, for molecular dynamics simulations a force is also required. For constant chemical potential, the contribution to the force on an atom k from the band energy is:

$$\begin{aligned}\mathbf{F}^{(k)}_{band} &= -\frac{\partial E_{band}}{\partial \mathbf{r}_k} \\ &= -\frac{1}{\pi}\sum_{i\alpha}\lim_{\eta\to 0}\mathrm{Im}\int^{E_f}\mathrm{d}E\frac{\partial G_{i\alpha,i\alpha}(E+\mathrm{i}\eta)}{\partial \mathbf{r}_k}E. \end{aligned} \tag{2.53}$$

The force depends on the derivative of the Green's function; if the expression for G_{00} from Eq. 2.44 is taken, setting $\mid U_0\rangle = \mid i\alpha\rangle$ and using the fact that the recursion coefficients depend on the moments (Eq.(2.48)), then the chain rule for partial differentiation gives:

$$\frac{\partial G_{00}(Z)}{\partial \mathbf{r}_k} = \sum_n\sum_{p=1}^{2n+1}\frac{\partial G_{00}(Z)}{\partial a_n}\frac{\partial a_n}{\partial \mu^{(p)}_{i\alpha}}\frac{\partial \mu^{(p)}_{i\alpha}}{\partial \mathbf{r}_k} + \sum_n\sum_{p=1}^{2n}\frac{\partial G_{00}(Z)}{\partial b_n}\frac{\partial b_n}{\partial \mu^{(p)}_{i\alpha}}\frac{\partial \mu^{(p)}_{i\alpha}}{\partial \mathbf{r}_k}. \tag{2.54}$$

The derivatives of G_{00} with respect to the recursion coefficients can be evaluated without difficulty ($\partial G_{00}(Z)/\partial a_n = G_{0n}(Z)G_{n0}(Z)$ and $\partial G_{00}(Z)/\partial b_n = 2G_{0(n-1)}(Z)G_{n0}(Z)$), as can the derivatives of the recursion coefficients with respect to the moments (Horsfield 1996). The problem with direct differentiation of the Green's function, alluded to above, arises when the derivatives of the moments with respect to atomic positions are considered. Formally, the derivative of the p^{th} moment is given by:

$$\begin{aligned}\frac{\partial \mu^{(p)}_{i\alpha}}{\partial \mathbf{r}_k} &= \frac{\partial}{\partial \mathbf{r}_k}\sum_{j_1\beta_1\ldots j_{p-1}\beta_{p-1}} H_{i\alpha,j_1\beta_1}H_{j_1\beta_1,j_2\beta_2}\ldots H_{j_{p-1}\beta_{p-1},i\alpha} \\ &= \sum_{j_1\beta_1\ldots j_{p-1}\beta_{p-1}}\left\{\frac{\partial H_{i\alpha,j_1\beta_1}}{\partial \mathbf{r}_k}H_{j_1\beta_1,j_2\beta_2}\ldots H_{j_{p-1}\beta_{p-1},i\alpha}\right. \\ &\quad \left. + H_{i\alpha,j_1\beta_1}\frac{\partial H_{j_1\beta_1,j_2\beta_2}}{\partial \mathbf{r}_k}\ldots H_{j_{p-1}\beta_{p-1},i\alpha} + \ldots\right\}. \end{aligned} \tag{2.55}$$

This expression is, generally, very slow to evaluate on a computer (though it can be done for a low order moment expansion(Foiles 1993)). For a moment of order p, because of the sum inherent in the moments, p contributions to the force (in effect p moments) must be evaluated, each of which has 3N components (where N is the number of atoms in the system).

The method for working around this is to use *global* moments, rather than local moments, leading to the Global Density of States method(GDOS) (Horsfield 1996). The global moments are given by:

$$\begin{aligned}\bar{\mu}^{(p)} &= \int^{E_f}\mathrm{d}E n(E)E^p \\ &= \sum_{i\alpha,j_1\beta_1\ldots j_{p-1}\beta_{p-1}} H_{i\alpha,j_1\beta_1}H_{j_1\beta_1,j_2\beta_2}\ldots H_{j_{p-1}\beta_{p-1},i\alpha}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i\alpha} \langle i\alpha \mid \hat{H}^p \mid i\alpha \rangle \\
&= \sum_{i\alpha} \mu_{i\alpha}^{(p)} \\
&= \mathrm{Tr}\left\{ \hat{H}^p \right\}. \qquad (2.56)
\end{aligned}
$$

A key point of GDOS, in terms of obtaining forces, is to note that matrices within a trace can be permuted without affecting the value of the trace. This leads to the derivative of the moments with respect to the atomic positions being written as:

$$\frac{\partial \bar{\mu}^{(p)}}{\partial \vec{r}_k} = p \sum_{i\alpha, j_1\beta_1 \ldots j_{p-1}\beta_{p-1}} \frac{\partial H_{i\alpha, j_1\beta_1}}{\partial \vec{r}_k} H_{j_1\beta_1, j_2\beta_2} \ldots H_{j_{p-1}\beta_{p-1}, i\alpha}, \qquad (2.57)$$

which is extremely easy to calculate on a computer.

The GDOS method is used with direct inversion of *global* moments, following Eq. 2.48. This means that it is unstable beyond about 20 moments, and can be memory intensive as all moments must be stored. There is also a reduced rate of convergence of energy for inhomogeneous systems, when compared with a local density of states method (as the inhomogeneity is averaged over the whole system). It does, however, have forces which are exact derivatives of energy, which is not true for FOE and BOP (see below). The linear scaling comes from working with moments, which, even though global, can be calculated locally and summed. Truncation of the sum is then possible, and linear scaling results.

2.6.2 The Fermi operator expansion method

The Fermi Operator Expansion (Goedecker and Colombo 1994; Goedecker and Teter 1995; Voter, Kress and Silver 1996) (FOE) is an alternative, and conceptually simple, method of obtaining the density matrix, albeit at the expense of introducing a finite (and often significant) electronic temperature. The method starts by defining the Fermi matrix,

$$\mathbf{F}_{\mu,\mathrm{T}} = f\left(\frac{\mathbf{H} - \mu}{\mathrm{kT}} \right), \qquad (2.58)$$

where $f(x) = 1/(1 + \exp(x))$, the Fermi function. The Fermi matrix plays the same role as the real-space density matrix in Eqs. 2.35 and 2.36.

Now, the Fermi function only has to cover a finite width, that is the width of the density of states for the system in question (or the difference between the minimum and maximum eigenvalues of the Hamiltonian). Within this range, it can be represented by a polynomial in the energy,

$$f(x) = \sum_{p=0}^{n_{pl}} C_p E^p, \qquad (2.59)$$

which means that the Fermi matrix $F_{\mu,T}$ can be represented as a polynomial in the Hamiltonian,

$$\mathbf{F}_{\mu,T} = \sum_{p=0}^{p=n_{pl}} C_p \hat{H}^p. \qquad (2.60)$$

This then gives the expression for one element of the Fermi matrix as:

$$\langle i\alpha \mid F_{\mu,T} \mid j\beta \rangle = \sum_{p=0}^{p=n_{pl}} C_p \langle i\alpha \mid \hat{H}^p \mid j\beta \rangle, \qquad (2.61)$$

which is clearly a sum of moments of the Hamiltonian. Conceptually, then, the method works by fitting a polynomial to the fermi function over the range of the eigenvalues. Then, using the coefficients of this polynomial and the moments of the Hamiltonian, elements of the finite temperature density matrix, or the Fermi matrix, are constructed.

The order of the polynomial is given as $n_{pl} \approx W/\mathrm{kT}$ (Goedecker and Colombo 1994). As described, the method is $O(N^2)$, as each element of the Fermi matrix will require n_{pl} operations, and there are N^2

elements. However, as discussed above, bonding in semiconductors and insulators (and, at high electronic temperature, metals) is local, and so the Fermi matrix can be truncated beyond a certain cutoff radius, leading to $O(N)$ scaling.

If the simple moments of the Hamiltonian are used, then the method becomes unstable rather quickly. In practice, a Chebyshev polynomial is used (Goedecker and Teter 1995), which leads to a recursion relation for the coefficients:

$$p_{\mu,T}(H) = \frac{c_0}{2} + \sum_{j=1}^{n_{pl}} c_j T_j(H), \text{and} \tag{2.62}$$

$$T_0(H) = I$$

$$T_1(H) = H$$

$$T_{j+1} = 2HT_j(H) - T_{j-1}(H) \tag{2.63}$$

The Gillan scheme (1989) for extrapolating the T=0 energy from a high temperature energy is available to extrapolate back to zero temperature.

Once the Fermi matrix has been truncated, the forces calculated using Eq. 2.36 are not exactly equal to the derivative of energy (Voter, Kress and Silver 1996); the formalism derived by Voter, Kress and Silver (1996) does, however, give a force which is the exact derivative of the energy.

2.6.3 The bond order potential method

The Bond Order Potential method (Pettifor 1989, Aoki 1993, Horsfield *et al.* 1996) is an efficient method for obtaining the off-diagonal elements of the density matrix, $\rho_{i\alpha,j\beta}$ in terms of a single-site Green's function, $G_{i\alpha,i\alpha}$. This enables both energy and force to be found in an efficient linear scaling manner. It derives its name from the expression for the bond energy,

$$U_{bond} = \sum_{i \neq j} h(R_{ij})\Theta_{ij}, \tag{2.64}$$

where $h(R_{ij})$ are the hopping elements, and Θ_{ij} is the *bond order*, $\Theta_{i\alpha,j\beta} = 2\rho_{i\alpha,j\beta}$. This expression resembles a pair potential, but Θ_{ij} is environment dependent (Pettifor 1989). To assist in understanding the mathematical formalism used in BOP, an earlier solution will be considered first.

The two-site BOP expansion The two-site BOP expansion (Pettifor 1989,Pettifor and Aoki 1991, Aoki and Pettifor 1993) extends the simple result of Eq. 2.52 to a more general form. Consider the state:

$$\mid U_0^\lambda\rangle = \frac{1}{\sqrt{2}}\left[\mid i\alpha\rangle + \mathrm{e}^{i\theta} \mid j\beta\rangle\right], \tag{2.65}$$

where $\theta = \cos^{-1}(\lambda)$. Here, λ defines a mixing between the two states; the previous states ($\mid +\rangle$ and $\mid -\rangle$) are given by $\lambda = -1$ and $\lambda = 1$. Now consider the Green's function element G_{00}, and expand out $\mid U_0^\lambda\rangle$,

$$\begin{aligned} G_{00}^\lambda(Z) &= \langle U_0^\lambda \mid \hat{G}(Z) \mid U_0^\lambda\rangle \\ &= \frac{1}{2}\left[\langle i\alpha \mid \hat{G}(Z) \mid i\alpha\rangle + \langle j\beta \mid \hat{G}(Z) \mid j\beta\rangle\right] \\ &+ \lambda\langle i\alpha \mid \hat{G}(Z) \mid j\beta\rangle \end{aligned} \tag{2.66}$$

A value for $G_{i\alpha,j\beta}$ can be found by differentiating Eq. 2.66 with respect to λ, which yields

$$G_{i\alpha,j\beta}(Z) = \frac{\partial G_{00}^\lambda(Z)}{\partial\lambda} \tag{2.67}$$

and

$$\begin{aligned} \Theta_{i\alpha,j\beta} &= \frac{\partial N^\lambda}{\partial\lambda} \\ N^\lambda &= -\frac{2}{\pi}\lim_{\eta\to 0}\mathrm{Im}\int^{E_f} \mathrm{d}E G_{00}^\lambda(E + i\eta), \end{aligned} \tag{2.68}$$

where the physical interpretation of N^λ is the number of electrons in the state $\mid U_0^\lambda\rangle$. The expression for $G_{i\alpha,j\beta}$ can be further expanded, by using the chain rule and the dependence of G_{00}^λ on the recursion coefficients, a_n^λ, b_n^λ,

$$G_{i\alpha,j\beta} = \sum_{n=0}^{\infty} \frac{\partial G_{00}^\lambda}{\partial a_n^\lambda} \delta a_n^\lambda + \sum_{n=1}^{\infty} \frac{\partial G_{00}^\lambda}{\partial b_n^\lambda} \delta b_n^\lambda, \tag{2.69}$$

where $\delta a_n^\lambda = \partial a_n^\lambda / \partial\lambda$ and $\delta b_n^\lambda = \partial b_n^\lambda / \partial\lambda$. This gives an expression for the bond order,

$$\begin{aligned} \Theta_{i\alpha,j\beta} &= \sum_{n=0}^{\infty} \frac{\partial N^\lambda}{\partial a_n^\lambda} \delta a_n^\lambda + \sum_{n=1}^{\infty} \frac{\partial N^\lambda}{\partial b_n^\lambda} \delta b_n^\lambda \\ &= -2 \left[\sum_{n=0}^{\infty} \chi_{0n,n0}(E_f) \delta a_n^\lambda + \sum_{n=1}^{\infty} \chi_{0(n-1),n0} 2\delta b_n^\lambda \right]. \end{aligned} \tag{2.70}$$

The functions $\chi_{0m,n0}(E_f)$ are called *response functions* and are calculated from:

$$\chi_{0m,n0}(E_f) = \frac{1}{\pi} \lim_{\eta\to 0} \mathrm{Im} \int^{E_f} \mathrm{d}E G_{0m}^\lambda(E + i\eta) G_{n0}^\lambda(E + i\eta), \tag{2.71}$$

and the Green's functions are derived from the recursion relation

$$(Z - a_n^\lambda) G_{nm}^\lambda(Z) - b_n^\lambda G_{n-1,m}^\lambda(Z) - b_{n+1}^\lambda G_{n+1,m}^\lambda(Z) = \delta_{n,m} \tag{2.72}$$

using the fact that $G_{0n} = G_{n0}$, and finding the starting value of G_{00}^λ from the continued fraction in Eq. 2.44. By expanding out the derivatives of the recursion coefficients (δa_n and δb_n) with respect to the moments, the link back to moments methods can be seen:

$$\delta a_n^\lambda = \frac{\partial a_n^\lambda}{\partial\lambda} = \sum_{r=1}^{2n-1} \frac{\partial a_n^\lambda}{\partial \mu_\lambda^{(r)}} \frac{\partial \mu_\lambda^{(r)}}{\partial\lambda} = \sum_{r=1}^{2n-1} \frac{\partial a_n^\lambda}{\partial \mu_\lambda^{(r)}} \zeta_{i\alpha j\beta}^{r-1} \tag{2.73}$$

$$\delta b_n^\lambda = \frac{\partial b_n^\lambda}{\partial\lambda} = \sum_{r=1}^{2n-1} \frac{\partial b_n^\lambda}{\partial \mu_\lambda^{(r)}} \frac{\partial \mu_\lambda^{(r)}}{\partial\lambda} = \sum_{r=1}^{2n-1} \frac{\partial b_n^\lambda}{\partial \mu_\lambda^{(r)}} \zeta_{i\alpha j\beta}^{r-1}, \tag{2.74}$$

where $\mu_\lambda^{(r)} = \langle U_0^\lambda \mid \hat{H}^r \mid U_0^\lambda \rangle$ and $\zeta_{i\alpha j\beta}^{r-1} = \langle i\alpha \mid \hat{H}^r \mid j\beta \rangle$, an *interference term.*

The physical interpretation for Eq. 2.70 is as follows. The dependence of $\Theta_{i\alpha,j\beta}$ on the number of electrons in the bond is expressed in the functions $\chi_{0m,n0}$. These functions have a weak dependence on atomic coordination, which enters through the derivatives of the recursion coefficients, δa_n and δb_n.

The two-site BOP expansion, while it provides better convergence than Eq. 2.52, is still problematical. It is more slowly convergent for the bond energy than single site recursion, which leads to a breakdown in equivalence between the single site (or site diagonal) and intersite expansions, and it is still affected by the symmetry problem referred to above. The solution to these problems was found in a reformulation of BOP.

The single site BOP expansion The single site BOP expansion requires a reasonably high level of formal mathematics for rigorous proof and derivation, which is not the purpose here. For rigorous definitions and derivations of this subject, the interested reader is referred to Aoki (1993) and Horsfield *et al.* (1996). The discussion below is intended to describe enough of the formalism to enable a physical understanding of the method.

In the previous section, it was shown that the off-diagonal Green's function element, $G_{i\alpha,j\beta}$ could be obtained from the composite state $\mid U_0^\lambda\rangle = \frac{1}{\sqrt{2}}\left[\mid i\alpha\rangle + \mathrm{e}^{i\theta} \mid j\beta\rangle\right]$, where $\theta = \cos^{-1}(\lambda)$. From this Green's function element the bond order between the orbitals $i\alpha$ and $j\beta$, $\Theta_{i\alpha,j\beta}$, can be obtained. Now consider a label which is different for every bond in the system, $\Lambda_{i\alpha,j\beta}$. This can be thought of as an element of a matrix Λ; the element can be given by the inner product between two vectors,

$$\Lambda_{i\alpha,j\beta} = (e_{i\alpha}^\Lambda \mid e_{j\beta}^\Lambda). \tag{2.75}$$

It is useful to introduce an *auxiliary space* here. The vectors which span it will be denoted as $\mid e_\nu^0)$, where the round bracket is used to show the difference between the auxiliary space and the space associated with the atomic orbitals.

The vectors $\mid e^{\Lambda}_{i\alpha})$ which were used above to construct the elements of the matrix Λ can be defined so as to have various useful properties: they exist in the space which is spanned by the orthonormal vectors $\mid e^{0}_{\nu})$; they are each associated with one orbital in the space spanned by the Hamiltonian, hence the index $i\alpha$; and they are only ever used in conjunction with an atomic orbital.

The generalisation of the vector $\mid U^{\lambda}_{0}\rangle$ can be written:

$$\mid W^{\Lambda}_{0}\} = \sum_{i\alpha} \mid i\alpha\rangle \mid e^{\Lambda}_{i\alpha}), \tag{2.76}$$

where, again, the curly bracket is used to denote the difference between this vector and the others in the system. $\mid W^{\Lambda}_{0}\}$ is a generalisation of the previous vector as it now links *all* the basis functions in the system, but will still be used as the starting vector for the recursion.

If it is used as the first vector for the recursion relation for G_{00}, then

$$\begin{aligned} G^{\Lambda}_{00} &= \{W^{\Lambda}_{0} \mid \hat{G}(Z) \mid W^{\Lambda}_{0}\} \\ &= \sum_{i\alpha,j\beta} (e^{\Lambda}_{i\alpha} \mid \langle i\alpha \mid \hat{G}(Z) \mid j\beta\rangle \mid e^{\Lambda}_{j\beta}), \\ &= \sum_{i\alpha,j\beta} G_{i\alpha,j\beta}(Z)\Lambda_{i\alpha,j\beta} \end{aligned} \tag{2.77}$$

This allows the central part of the one site BOP expansion to be written:

$$G_{i\alpha,j\beta}(Z) = \frac{\partial G^{\Lambda}_{00}}{\partial \Lambda_{i\alpha,j\beta}}. \tag{2.78}$$

As the values of $\Lambda_{i\alpha,j\beta}$ and the vectors $\mid e^{\Lambda}_{i\alpha})$ have been left completely general, they can now be specified. If the value $\Lambda_{j\beta,k\gamma} = \delta_{i,j}\delta_{i,k}\delta_{\alpha,\beta}\delta_{\alpha,\gamma}$ is chosen (the notation relates to a piece of mathematical formalism described in Horsfield *et al.* (1996)), then the following simple and central result can be derived:

$$G^{\Lambda}_{00} = G_{i\alpha,i\alpha} \tag{2.79}$$

As G^{Λ}_{00} can clearly be written as a continued fraction using the recursion method as described above, then much of the formalism developed for the recursion method can be used here. Using the equivalent of Eq. 2.69, the expression for an off-site Green's function element can be rewritten as

$$\begin{aligned} G_{i\alpha,j\beta} = \quad & \textstyle\sum_{n=0}^{\infty} G^{\Lambda}_{0n}(Z)G^{\Lambda}_{n0}(Z)\frac{\partial a^{\Lambda}_{n}}{\partial \Lambda_{i\alpha,j\beta}} + \\ & 2\textstyle\sum_{n=1}^{\infty} G^{\Lambda}_{0,(n-1)}(Z)G^{\Lambda}_{n,0}(Z)\frac{\partial b^{\Lambda}_{n}}{\partial \Lambda_{i\alpha,j\beta}}, \end{aligned} \tag{2.80}$$

where G^{Λ}_{0n} is given by the recursion relation shown in Eq. 2.72. By integrating this as before (c.f. Eq. 2.70), the following expression for the bond order can be found:

$$\Theta_{i\alpha,j\beta} = -2\left[\sum_{n=0}^{\infty} \chi^{\Lambda}_{0n,n0}\frac{\partial a^{\Lambda}_{n}}{\partial \Lambda_{i\alpha,j\beta}} + 2\sum_{n=1}^{\infty} \chi^{\Lambda}_{0(n-1),n0}\frac{\partial b^{\Lambda}_{n}}{\partial \Lambda_{i\alpha,j\beta}}\right], \tag{2.81}$$

with $\chi_{0m,n0}$ found from Eq. 2.71.

The derivatives of the recursion coefficients, $\partial a^{\Lambda}_{n}/\partial\Lambda_{i\alpha,j\beta}$ and $\partial b^{\Lambda}_{n}/\partial\Lambda_{i\alpha,j\beta}$ can be evaluated simply and stably using a recurrence relation. There are a number of other issues which are too complicated for treatment here: the proof that within the new formalism the intersite and site diagonal expressions for the energy are equivalent; the rotational invariance of the new formalism (obtained by working with moments averaged over the magnetic quantum number); and the truncation of the recursion series for the derivatives of the recursion coefficients. For a discussion of all of the above issues, the reader is again directed to Aoki (1993) and Horsfield *et al.* (1996).

Conclusions The theoretical bases for all of the computational methods used in this thesis have been explained, and some of the problems with each of them highlighted.

Density functional theory is the most accurate, and different approximations to the exchange and correlation energy have been mentioned. The commonly used LDA is sufficient for static energy calculations, while the added accuracy of the GGA is needed for energy barriers.

Tight binding is a fast quantum mechanical method which is based in DFT. Four linear scaling methods have been examined in some detail. The density matrix method uses the elements of a trial density matrix as variational degrees of freedom with respect to which the energy is minimised. This is similar to obtaining the charge density in the system, and yields an energy and forces in exact correspondance.

All of the moments methods use the rather elegant mapping between the bonding local to an atom and the moments of its density of states to find the band energy efficiently. The divergence between the methods occurs when forces are required, and a variety of different techniques are used to obtain these quickly.

Chapter 3

Technical Issues: Linear Scaling and Parameterisations

For the nature of the chemical bond is the problem at the heart of all chemistry.

Bryce Crawford, Jr., 1953

While the previous chapter has covered the theories which underlie all of the calculations in my thesis, this chapter addresses more detailed issues, specifically where I have developed a new understanding or parameterisation. The new field of linear scaling theories in tight binding has been touched on in Chapter 2, and the mathematical basis for several of these theories has been explained. In the first part of this chapter, the results of an investigation into the suitability of these different methods for calculating energies and forces for different systems are presented, using tests which identify the best method to use for the calculations in my thesis. After that I describe the concept of a tight binding parameterisation, and detail three parameterisations which I have fitted (for bonds between Si atoms, between Si and H atoms and between Si and Bi atoms - all of which are extremely relevant to silicon growth: the first two to the growth from hydride gas sources, and the last to surface properties).

3.1 A comparison of linear scaling tight binding methods

In this section, the results generated from the methods described above in sections 2.4- 2.6 are compared for different systems. The tests have been performed for energy convergence, force convergence and efficiency (i.e. time per step for a given level of accuracy) for an insulator (carbon), a semiconductor (silicon), a metal (titanium) and a molecule (benzene), though several of the tests will not be presented here for reasons of space. The full set of comparisons can be seen in Bowler *et al.* (1997). All of the graphs, except for the efficiency plots, have a similar form which might be considered confusing, and so I will spend a little time explaining how the results have been laid out.

When calculating a given moment, there are many contributions from the different possible paths, as explained in section 2.6. For instance, the tenth moment of a system has contributions from the path which takes five hops away from atom i, and then five hops back again along the same path, as well as the path which consists of hopping to a given nearest neighbour, j, and back again five times, as well as many others. It is possible to evaluate *only* those moments which fit within a cluster of a given size, so that the order of the moment calculated can be increased without increasing the cluster size; this fits well into the linear scaling justification presented in section 2.4. (So-called exact moments correspond to evaluating the moments up to that moment which defines the cluster size - so, for instance, up to the eleventh moment in a five hop cluster). On the graphs, there are five boxes containing plots, which correspond to clusters made by hopping away from the central atom between three and seven times. For the density matrix method, the cluster size defines the cutoff radius of the density matrix; the x-axis has no significance. For the moments methods, the x-axis denotes the number of moments evaluated used within that cluster size.

3.1.1 Total energies and vacancy formation energies

All of the results in this section are presented as errors relative to the value calculated using an exact diagonalisation of the Hamiltonian in k-space, which was fully converged with respect to k-point sampling, and not relative to an *ab initio* or experimental result. One of the main aims here is to investigate energy *difference* convergence, rather than absolute energy convergence, as the bulk of the work in this thesis is in comparing the energetics of different structures.

Figure 3.1 shows the error in the binding energy for carbon. As it is an insulator, with extremely localised bonding, all of the methods would be expected to converge well with it, and this is indeed the case. By five hops, and twenty moments where applicable, all methods are converged, except for FOE. As will be seen, FOE requires more moments to converge, but can achieve them extremely quickly.

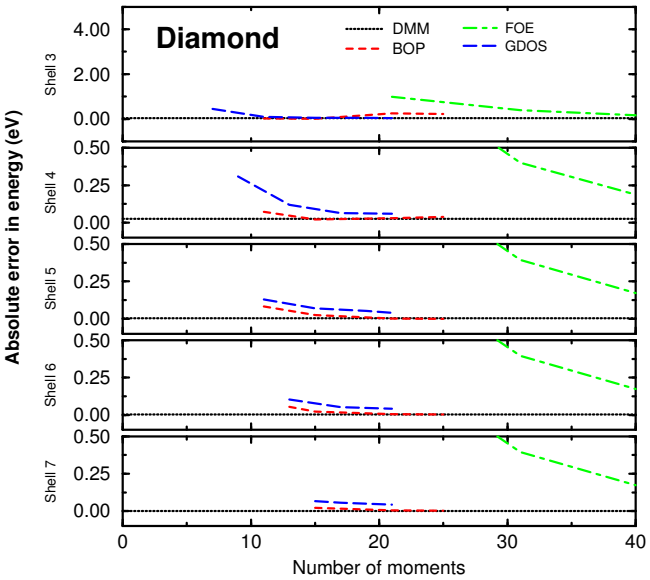

Figure 3.1: The error in binding energy for diamond relative to a k-space result (7.25 eV per atom).

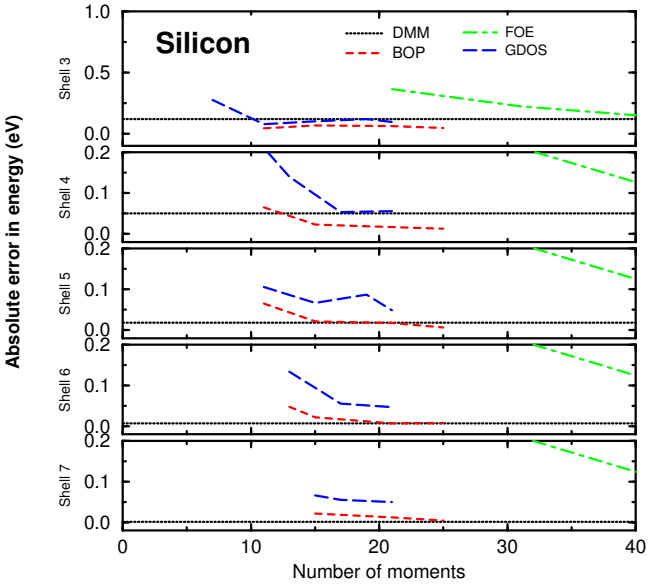

Figure 3.2: The error in binding energy for Si in the diamond structure relative to a k-space result (5.47 eV per atom).

Figure 3.2 shows the absolute error in binding energy for silicon in the diamond structure. As it is a semiconductor, it should take larger clusters to converge the energy than for diamond, and this is indeed what is seen. The convergence is good at five levels, and excellent by seven, with 20 moments entirely adequate for the moments methods.

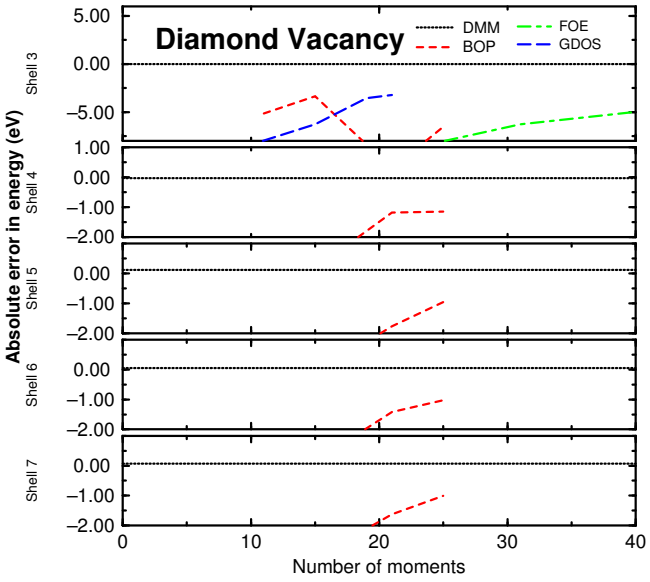

Figure 3.3: The error in unrelaxed vacancy formation energy for diamond relative to a k-space result (10.41 eV).

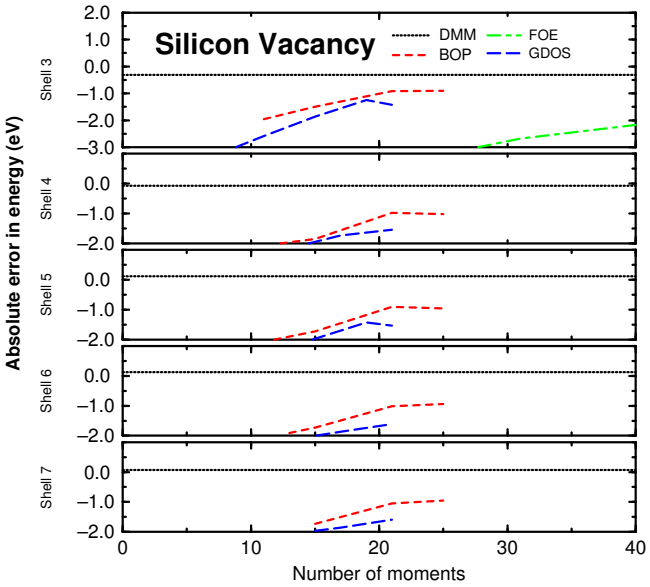

Figure 3.4: The error in unrelaxed vacancy formation energy for Si in the diamond structure relative to a k-space result (5.48 eV).

The vacancy formation energies shown in Figures 3.3 and 3.4 are for unrelaxed vacancies, and are calculated for 63 atom unit cells (i.e. 64 atom perfect bulk cells with one atom removed). The energy is found by subtracting 63/64×(energy for the bulk unit cell) from the energy for the vacancy cell. As can be seen from both figures, the DMM is the only method which is even close at any cluster level, and is well converged at three hops, and extremely well converged at five. It is this result (that energy differences

are well converged) which justifies use of three hops for the calculations to be performed throughout this thesis. The explanation for the difference between the methods lies in the fact that moments methods have difficulties resolving narrow features in the density of states (Kress and Voter 1995), while the DMM does not. The density matrix found in the DMM is exact within the limits of the approximation; this means that it will resolve sharp features well, where moments methods will have to include extremely high moments to achieve such resolution. It is important to note that the sign and magnitude of the error for the DMM varies because the value plotted is a *difference* in energies, and is no longer variational.

3.1.2 Forces and efficiency

There are two key questions to be addressed when considering the forces which a method produces: whether the Hellmann-Feynman theorem is obeyed (i.e. whether the forces calculated analytically match the numerical derivatives of the energy as discussed in Chapter 2); and whether the forces match the k-space well enough to reproduce MD results. The graphs shown in Figure 3.5 answer this question for the z-force on an atom on the unreconstructed Si(001) surface, though in a slightly confusing manner. On the graph, the solid black line is the k-space result, while for the three O(N) methods, there are two lines each: the analytic force and the numerical force. If the two lines for a given method do not match, the Hellmann-Feynman theorem is not satisfied and there will be difficulty in stating the energy achieved in a relaxation calculation, and great difficulty in achieving energy conservation during MD runs. Fig. 3.5 shows that the force from the DMM is almost indistinguishable from the force in k-space, whereas the moments methods have grave difficulty on both fronts, though FOE achieves good consistency.

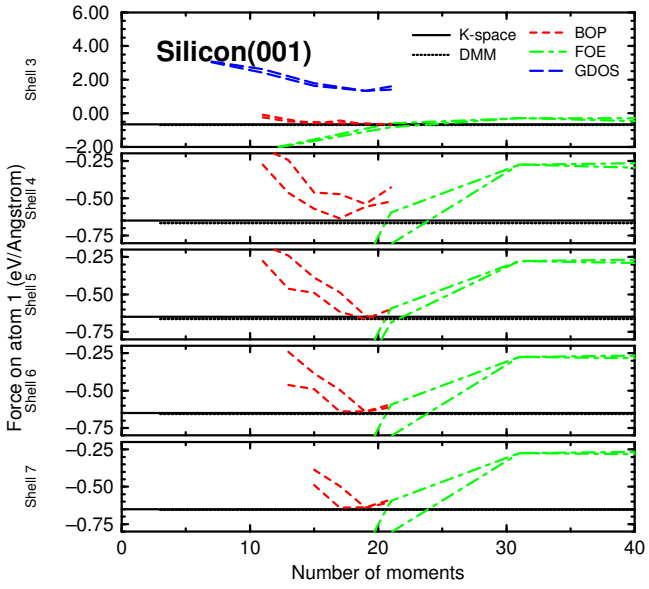

Figure 3.5: The z force on an atom in the unreconstructed Si(001) surface, compared to a k-space result. The two lines for each method represent numerical and analytical forces.

The final question to consider when deciding which of the linear scaling methods to use concerns efficiency: how long a given method takes to reach a given level of accuracy. For the purposes of the work in this thesis, the diamond vacancy formation energy has been considered, and is shown in a logarithmic plot for time in Figure 3.6. From this, it is clear that the only method of choice for insulators and semiconductors, particularly when there are gap states, is the DMM, and this has been used throughout.

3.1.3 Titanium: moments' revenge !

The above sections may have left the reader with the impression that the DMM is the only linear scaling tight binding method to use. This is hardly a fair impression to leave; the examples shown above were designed to test the systems to be considered in this thesis, and not all of the systems which could be modelled. In Figure 3.7, the vacancy formation energy for hcp titanium is shown; while DMM is performing as well as any

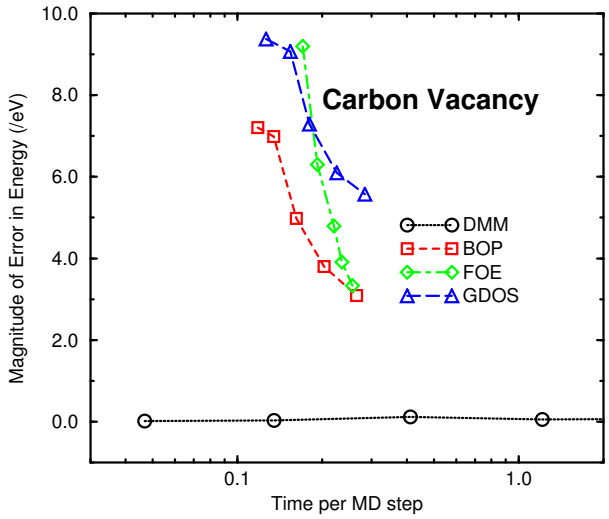

Figure 3.6: Time taken for a given accuracy for the different O(N) methods for the diamond vacancy.

of the other methods for this system, there are two drawbacks. Firstly, as can be seen in the figure, beyond five shells the results stop; for clusters larger than this, the DMM requires above 256 MB of memory, which is not available in the MML, whereas the other methods had memory spare even at 7 levels.

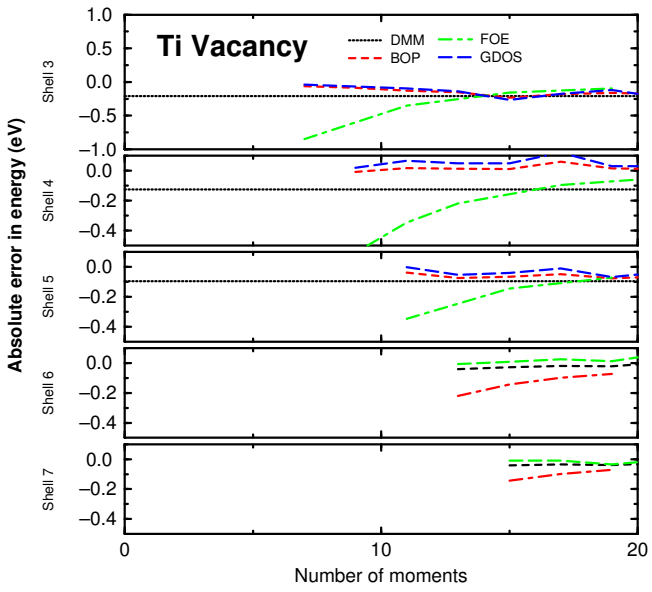

Figure 3.7: The error in unrelaxed vacancy formation energy for hcp titanium relative to a k-space result (0.88 eV).

The second drawback of the DMM can be seen in Figure 3.8, where the efficiency for the titanium vacancy formation energy is plotted. The graph makes its own point, though it is worth noting that of the moments methods, BOP or FOE are probably the methods of choice for this kind of calculation; if simplicity of programming and understanding is the aim, then FOE is the method to use, whereas for greater flexibility and potential accuracy, BOP has the edge.

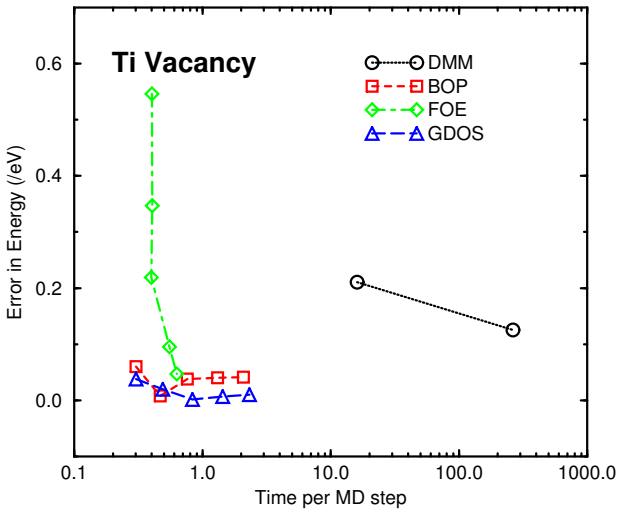

Figure 3.8: Time taken for a given accuracy for the different methods for a vacancy in hcp titanium.

3.2 Tight binding parameterisations

As well as the choice of method, as discussed above, the parameterisation used for a tight binding calculation strongly affects the success of the calculation. There are various issues which need to be addressed: what the parameterisation has been fitted to; how transferable the parameterisation is; what form is used for the scaling of the parameterisation; and how applicable the parameterisation is to the situation being modelled. It is possible that in several years time, the whole question of fitting a parameterisation may be academic with the development of so called *ab initio* tight binding methods, such as Sankey and Niklewski (1989). This class of methods explicitly constructs basis functions on each atom in the system, and then tabulates their integrals for a large range of separations. The Hamiltonian is generally set up using a Harris-Foulkes type approach (in which the superposed atomic charge densities are used as the input to the Kohn-Sham equations, and self-consistency is neglected), before being diagonalised using standard methods. Thus the Hamiltonian is still parameterised, but based on *ab initio* methods.

In general, when constructing a parameterisation, there are three areas which must be addressed: the electronic parameters (which are generally fitted to bulk properties, such as band structure), the scaling parameters (which represent how the electronic parameters change with distance) and the repulsive potential (which represents everything which has been neglected until now, and is generally taken to be a pairwise function). The electronic parameters include the on-site energies (e.g. E_s and E_p) and the hopping parameters (e.g. $\mathrm{h}_{ss\sigma}$, which describe electronic transitions). The scaling parameters multiply the equilibrium eletronic parameters and the repulsive potential; a variety of forms for scaling will be discussed below.

In the original tight binding paper by Slater and Koster (1954), only the electronic part of the parameterisation was covered. They provided a table which gave all possible two centre integrals between s, p and d-like orbitals, based on particular symmetry relations (for instance, a p_x orbital will have σ and π components to its overlap with another p_x orbital). For two atoms at points $\mathbf{R}_i$ and $\mathbf{R}_j$, with direction cosines in the direction $\mathbf{R}_j$-$\mathbf{R}_i$ given by l, m and n, then the Hamiltonian elements between different orbitals on i and j are shown in Table 3.1. It is these relationships which were referred to in section 2.3, when considering orbital overlap.

Slater and Koster fitted the Hamiltonian matrix elements to *ab initio* calculations of band structure, at certain high symmetry points in the Brillouin zone. The paper considers the band structure of various different materials, and looks forward to the kind of calculations being performed today. It is a marvelous example of how a scientific paper should be written.

Chadi and co-workers (Chadi and Cohen 1975; Chadi and Martin 1976;Chadi 1977; Chadi 1979a,b) developed tight binding parameters for silicon, germanium, gallium arsenide and other III-V semiconductors. The method assumes a scaling of the hopping parameters as d^{-2} for two atoms (i and j) separated by a

$H_{s,s}$	$h_{ss\sigma}$
$H_{s,x}$	$l\cdot h_{sp\sigma}$
$H_{x,x}$	$l^2\cdot h_{pp\sigma}$ - $(1-l^2)\cdot h_{pp\pi}$
$H_{x,y}$	$lm\cdot h_{pp\sigma}$ - $lm\cdot h_{pp\pi}$
$H_{x,z}$	$ln\cdot h_{pp\sigma}$ - $ln\cdot h_{pp\pi}$

Table 3.1: The Hamiltonian elements between two atoms (i and j) for s and p-like orbitals. Note that x,y and z refer to p_x, p_y and p_z orbitals, and that l, m and n are the direction cosines between atoms i and j. After Slater and Koster (1954).

distance d (this is similar to the method of Harrison (1980), which is discussed in the next paragraph) and a repulsive term which had no equilibrium value, U_0, but was given by:

$$\Delta U = \sum_{i>j}(U_1\epsilon_{ij} + U_2\epsilon_{ij}^2), \tag{3.1}$$

where ϵ_{ij} is the fractional change in bond length. U_1 was fitted so that the total energy was at a minimum for the equilibrium volume, and U_2 was fitted to give accurate harmonic forces, so as to give good results for elastic constants and transverse phonons. The final version of the silicon parameterisation (Chadi 1979c) gave good results for the (001) surface, but the bulk modulus was still wrong. Chadi and his co-workers fitted to data from x-ray and photoemission band spectra, and *ab initio* band structure results, though they acknowledged that they could only fit the valence band and not the conduction band with a nearest neighbour, sp^3 model.

Froyen and Harrison (1979) compared the band structures produced by linear combination of atomic orbital (LCAO) theory and free electron theory, and using certain selected differences in energy from these two methods found a universal formula for hopping elements in terms of a dimensionless coefficient $\eta_{ll'm}$,

$$h_{ll'm} = \eta_{ll'm}\frac{\hbar^2}{md^2}, \tag{3.2}$$

where d is the distance between ions and m is the mass of the electron. This method works extremely well for semiconductors in the tetrahedral structure, but not for close-packed structures (Paxton, Sutton and Nex 1987). The whole bond orbital model of Harrison is well described in his book on the subject (Harrison 1980), where he derives the d^{-2} scaling of hopping parameters, and also the d^{-4} scaling of repulsive terms, which form an alternative to the repulsive model used by Chadi and described above.

One of the most successful tight binding parameterisations is that of Vogl, Hjalmarson and Dow (1983). (A useful, and increasingly common, measure of the success of a scientific paper is the citation count - Vogl, Hjalmarson and Dow has been cited over 600 times). Two things were shown in their paper: that the hopping parameters could be fitted to high symmetry points of band structures; and that adding a single, excited state (which they call the s^* state) can greatly improve the bottom of the conduction band, and thus the shape of the gap. In their paper they give parameters for sixteen tetrahedral semiconductors (from carbon, silicon and germanium through to zinc telluride) all of which correctly fit the value and direction (in k-space) of the gap.

A major concern with all tight binding parameterisations is that of transferability: how well a parameterisation will perform in an environment which is different to that used for fitting. This problem was addressed by Goodwin, Skinner and Pettifor (1989), who produced a nearest neighbour, orthogonal tight binding parameterisation for silicon which fitted well to the diamond phase and reproduced the energy and volume differences for the beta-tin, simple cubic, fcc and bcc phases reasonably well. They achieved this by increasing the E_s to E_p separation beyond that expected physically (Chadi had 6.45 eV, Harrison 7.03 eV and GSP 8.3 eV) and introducing a rescaling of the hopping parameters, shown in Eq.3.3.

$$f(r) = (\frac{r_0}{r})^n exp\{n(-(\frac{r}{r_c})^{n_c} + (\frac{r_0}{r_c})^{n_c})\}, \tag{3.3}$$

The next three sections cover three parameterisations which have been developed in the course of this thesis, for interactions between Si atoms, between Si and H atoms, and between Si and Bi atoms. There are already several parameterisations for both Si-Si and Si-H bonds which are used in the literature (e.g. Chadi (1979c), Goodwin, Skinner and Pettifor (1989), Kwon *et al.* (1994), Min *et al.* (1992) and Panzarini and Colombo (1994)); it might be reasonable to ask why another parameterisation is necessary. The answer

which emerges is that the parameterisations which are available are all unsuitable in one way or another for the systems which are to be studied in this paper. There was, however, no parameterisation available for the Si-Bi work planned, and the work presented is alone in the field.

3.3 Si-Si bonding

The parameterisation of Goodwin, Skinner and Pettifor (1989) (GSP) was designed to fit to various phases of silicon at equilibrium bond lengths - notably including the metallic phases. This was achieved in the simple sp^3, orthogonal tight binding model by artificially increasing the separation between E_s and E_p beyond that appropriate for the tetrahedral structure; unfortunately, the increased separation weakens the π-bonding of silicon dimers on Si(001) and renders it unsuitable for modelling the chemistry of the Si(001) surface. The weakening can be seen quite simply by modelling two unit cells: a clean Si(001) surface, and a Si(001) surface with one hydrogen adsorbed on one end of a dimer. The dimer which has had hydrogen adsorbed on one end should be longer, as the π-bond in the dimer has been broken (and, indeed, in the LDA modelling described in Chapter 6, the bond length for a clean dimer is 2.28 Å, and for a partially hydrogenated dimer 2.43 Å). Unfortunately, the GSP parameterisation predicts that the bond *shortens* when a hydrogen adsorbs; this suggests that the next available orbital after the σ-bond is the σ^* anti-bonding orbital, which is emptied when the hydrogen adsorbs, leading to the shortening.

The early parameterisation of Chadi (1979c) described the band structure rather well, and had the correct separation of E_s and E_p, but gave poor elastic constants, due to the assumption of $1/d^2$ scaling (see Goringe (1995) for further discussion of this point). The parameterisation of Kwon *et al.* (1994) is rather more complicated, and uses a cutoff which is more than nearest-neighbour, which degrades the performance of linear scaling methods enormously. This parameterisation and the GSP parameterisation, in attempting to be transferable to all phases of silicon, are less suited to a detailed study of one aspect of silicon chemistry. For this reason, a new parameterisation was created specifically for the Si(001) surface.

The aim of the parameterisation presented here was to reproduce well the chemistry of the Si(001) surface, as well as bulk Si. This led to a number of requirements: a good fit to the bulk band structure; good elastic constants; the correct separation of E_s and E_p; and accurate reproduction of small distortions of the structure. As Chadi's parameterisation has the correct separation of E_s and E_p, and reproduces the band structure well, the values he chose for the energy levels and hopping parameters were adopted. The scaling rules must now be fitted; the GSP form for scaling (Eq.3.3) was used, and the following criteria fitted to:

1. The diamond and β-tin structure absolute volumes
2. The diamond-β-tin energy difference
3. The bulk modulus, $C^{'}$ and C_{44}^0

E_s	E_p	$h_{ss\sigma}$	$h_{sp\sigma}$	$h_{pp\sigma}$	$h_{pp\pi}$	ϕ_0
-12.2	-5.75	-1.938	1.745	3.050	-1.075	3.44566

r_0	r_c	d_c	n	n_c	m	m_c
2.35	3.8661	3.8521	1.9771	6.8702	4.7104	7.0531

Table 3.2: Parameters for Si-Si interactions. All energies are in eV.

This fitting yielded the parameters shown in Table 3.2. The quality of the fit is shown in Table 3.3. Various tests of the model, which involved both Si-Si interactions and Si-H interactions, are given in the next section. It is worth noting here that the fit is extremely good, except for the C' constant, though this is nothing unusual - it is normally at least this far off in a minimal basis set. With this exception, the fit is

Source	Diamond		β-tin		Elastic constants		
	a_0	Energy	a_0	Energy	B	$C^{'}$	C_{44}^o
TB	5.43	0.00	4.842	0.266	0.998	0.363	1.099
Expt/LDA	5.43	0.00	4.822	0.266	0.990	0.510	1.110

Table 3.3: Fitting results for the silicon-silicon interactions. Energies are in eV, lattice constants in Ångströms and elastic constants in Mbar.

almost perfect.

3.4 Si-H bonding

As has been explained above, it has been necessary to create new parameterisations for Si-Si and Si-H, although such sets already existed. The reasons for the rejection of the Si-H parameterisations now follow. The parameterisation of Panzarini and Colombo (1994) for H-Si interactions uses the GSP silicon parameterisation, and thus will not work well at surfaces for two reasons: firstly, because the Si-Si bonds are poorly described, and secondly because the gap between the silicon E_s (or, equivalently, the silicon E_p) and the hydrogen E_s will be incorrect. Also, they did not fit the scaling terms to any H-Si stretched bonds, which is appropriate for the bulk diffusion they were investigating, but is inappropriate for surface diffusion. The parameterisation of Min *et al.* (1992) used poor Si-Si parameters, and was not fitted to extended bonds. Again, a new parameterisation was created to fit to the following situations: extended bonds, and interactions at the Si(001) surface.

In the SiH_4 molecule, the only bonding involved is between Si and H. As it is also tetrahedral, the molecule is ideal for modelling Si-H interactions. The electronic eigenstates are available from experiments; theoretically, the Hamiltonian for the silane molecule can be solved analytically; the bonding (s_b, t_b) and anti-bonding (t_a, s_a) eigenvalues are given by[1]:

$$s_b = \frac{1}{2}\left[\left(E_s^{Si} + E_s^H\right) - \sqrt{\left(E_s^{Si} - E_s^H\right)^2 + 16h_{ss\sigma}^2}\right] \tag{3.4}$$

$$t_b = \frac{1}{2}\left[\left(E_p^{Si} + E_s^H\right) - \sqrt{\left(E_p^{Si} - E_s^H\right)^2 + \frac{16}{3}h_{sp\sigma}^2}\right] \tag{3.5}$$

$$t_a = \frac{1}{2}\left[\left(E_p^{Si} + E_s^H\right) + \sqrt{\left(E_p^{Si} - E_s^H\right)^2 + \frac{16}{3}h_{sp\sigma}^2}\right] \tag{3.6}$$

$$s_a = \frac{1}{2}\left[\left(E_s^{Si} + E_s^H\right) + \sqrt{\left(E_s^{Si} - E_s^H\right)^2 + 16h_{ss\sigma}^2}\right] \tag{3.7}$$

These can be inverted, to give the hopping parameters in terms of the energy levels of hydrogen and silicon. The hopping parameters will be written in terms of the bonding eigenvalues, and the onsite energies for silicon and hydrogen:

$$h_{ss\sigma} = \frac{1}{16}\sqrt{\left[\left(E_s^{Si} + E_s^H\right) - 2s_b\right]^2 - \left(E_s^{Si} - E_s^H\right)^2} \tag{3.8}$$

$$h_{sp\sigma} = \frac{3}{16}\sqrt{\left[\left(E_p^{Si} + E_s^H\right) - 2t_b\right]^2 - \left(E_p^{Si} - E_s^H\right)^2} \tag{3.9}$$

If no energy zero is defined, then there are three energy differences available for fitting (e.g. s_a - s_b, and the other levels relative to s_b), and three variables to fit (E_s^H, $h_{ss\sigma}$ and $h_{sp\sigma}$). A perfect fit to these requires the silicon E_s and E_p splitting to be reduced to 5.45 eV; as the Si-Si parameters have already been fitted, with the splitting between E_s and E_p fixed at 6.45 eV, the requirement for reducing E_{sp} will be disregarded. As the simple tight binding model which is being used rarely reproduces excited states well, only the occupied (bonding) eigenvalues will be fitted to, as shown in Eqns 3.8 and 3.9, which should not greatly affect the quality of the modelling. The other electronic degrees of freedom are the separation of the hydrogen E_s and silicon E_s levels, and the absolute level of the silicon E_s level relative to the experimental eigenspectrum (all other energy level are taken relative to the silicon E_s). The scaling of the Si-H bond for small displacements was fitted to the binding energy curve for the silane molecule. The final degree of electronic freedom (the absolute level of the silicon E_s level) can be understood more easily when it is noted that the hopping terms, $h_{ss\sigma}$ and $h_{sp\sigma}$, depend on E_s^{Si}.

The experimental data being fitted to are the eigenvalues of the molecule and the frequency of the symmetric bending mode. The other data come from *ab initio* calculations: the binding energy curve about equilibrium from a fully self-consistent LDA calculation There are two electronic degrees of freedom (the

[1] I am indebted to Andrew Horsfield for pointing this out to me; I have confirmed these results

absolute position of the silicon E_s level, and the relative position of the hydrogen E_s level with respect to the silicon E_s level), which are fitted to four pieces of data, while there are seven scaling degrees of freedom (using GSP scaling as shown in Eqn. 3.3, there are n, n_c and r_c for hopping and repulsive terms and ϕ_0) which are fitted to a 26 point binding energy curve.

The parameters from the fitting are shown in Table 3.4. In this situation, the GSP-style rescaling is used to obtain a clean cutoff at 3.4 Å (a requirement found from fitting to extended bonds; the simple d^{-n} scaling for the hopping and repulsive parameters reproduced the binding energy curve extremely well, making the GSP terms redundant). The tests of the parameterisation are described in the next section.

E_s	E_p	$h_{ss\sigma}$	$h_{sp\sigma}$	$h_{pp\sigma}$	$h_{pp\pi}$	ϕ_0
-8.4	N/A	-3.834	4.734	N/A	N/A	7.4399

r_0	r_c	d_c	n	n_c	m	m_c
1.474	3.4	3.4	-2.6752	20.0	-4.2302	20.0

Table 3.4: The parameters for Si-H interactions. All energies are in eV.

The fitting to extended bonds is somewhat harder; in the end, I have chosen to fit to a fictitious reaction. I have used the process of SiH_4 and SiH_3 (separated by 5 Å) transferring a hydrogen from the SiH_4 to the SiH_3 so that the end product is SiH_3 plus SiH_4. The entire system remains frozen during this transfer, with the hydrogen placed at different points between the two Si atoms. The Si-H interaction is completely negligible once the distance is more than 3.4 Å (i.e. the charge density between the two is zero, and the energy does not change once this distance is increased, hence the cutoff mentioned above). An additional constraint upon the fitting was that the equilibrium curvature of the binding energy curve should remain the same, so as not to disturb the vibrational modes already fitted. As it turned out, the parameters as already fitted gave a good fit to the *ab initio* energies, as shown in Figure 3.9. The forces from LDA and the tight binding fit are also shown in this figure, and are in remarkably good agreement.

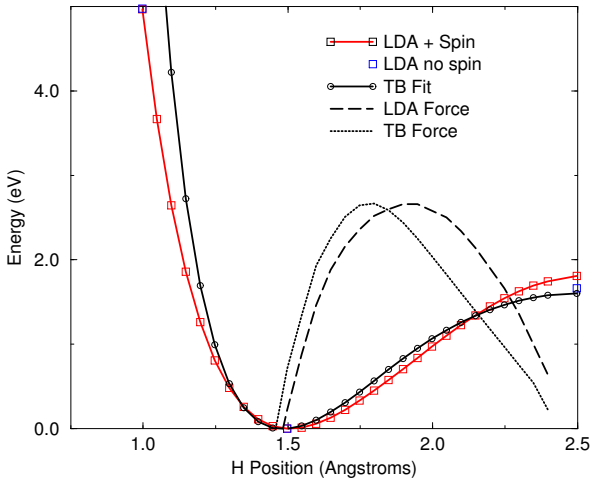

Figure 3.9: The energy for a fictitious reaction, $SiH_4 + SiH_3 \rightarrow SiH_3 + SiH_4$, with the coordinates of all atoms except for the central hydrogen kept fixed. The results from LDA and tight binding are shown, along with three points to indicate the effect of spin, and the force (by numerical differentiation of the energy curves).

As will be discussed in detail in Chapter 6, there is one remaining problem with the modelling of diffusion on the surface. When there is a long, weak bond, there is little charge density between the two atoms (e.g. between a hydrogen and the atom towards which it is diffusing), which means that physically, there is little interaction. Unfortunately, in tight binding, the repulsive term is sufficiently large that, rather than

Eigenvalue	TB	Expt
One	-16.6	-17.3
Two	-13.1	-13.3
Three	-12.1	-12.1
Four	-10.5	-10.7
Vibrational mode	TB	Expt
Symm Stretch	286	265
Symm Bend	114	112
Asymm Bend	107	105
Wag	74	78
Twist	55	62
Si-Si	56	54

Table 3.5: Eigenvalues and stretching modes for the disilane molecule, from tight binding and experiment. Eigenvalues are in eV, stretching modes in meV.

Molecule	Si-H	Si-Si	H-Si-H	H-Si-Si
Disilane	1.48 (1.49)	2.32 (2.33)	109.7 (108.6)	109.2
Disilene	1.48 (1.49)	2.25 (2.22)	110.7 (110)	116.3

Table 3.6: The bond lengths and angles for disilane (Si_2H_6) and disilene (Si_2H_4). Experimental values are given in brackets. Lengths are in Ångströms, angles in degrees.

a weak attraction between the atoms, there is a weak repulsion. This means that the barriers calculated by tight binding are rather high; however, they should all suffer from about the same effect, meaning that comparisons between different barriers in different environments should be accurate. An estimate of the effect of the repulsion is mentioned in Chapter 6.

3.4.1 Tests of the parameterisations

The parameterisations have been tested on a number of different systems, and perform extremely well. The asymmetric bending mode of the silane molecule is modelled as 108 meV, where experimentally it should be 113 meV. The four lowest eigenvalues and stretching modes for disilane are given in table 3.5, and the bond lengths and angles for disilane and disilene are given in table 3.6. These results indicate that the parameterisation is extremely good for Si-H bonds in a range of environments; along with the parameterisation for C-H bonds by Horsfield *et al.* (1996), it indicates that tight binding can achieve remarkable results given a relatively low level of computational effort.

The above calculations of molecules show remarkable agreement with experiment, but do not tackle a large system with a surface. Three of these are offered as tests of the parameterisation. First, the single missing dimer defect of Chapter 4. The LDA results give the rebonded defect to be 0.42 eV more stable than the unbonded; this parameterisation gives it as 0.45 eV more stable. Secondly, the SiH_2 fragments of Chapter 7 are correctly ordered in stability by this parameterisation, except for the on-dimer fragment (which, with a 60^o bond angle, is a difficult system for tight binding to model). Thirdly, the stability of Si ad-dimers in different adsorption sites on the Si(001) surface is correctly ordered by this parameterisation. There are two positions in which the dimer can sit (on the dimer row or over the trench; these are explained in more detail in Chapter 7) and two orientations for each: epitaxial (perpendicular to the substrate dimers) and non-epitaxial (parallel to the substrate dimers). The trench dimers are less stable than the row dimers, and the non-epitaxial orientation is more stable than the epitaxial one. The parameterisation correctly predicts these facts, though the trench dimers are more stable than they should be according to LDA; this is again a feature of the 60° bond angles formed in these structures, which are too far from what was fitted to be modelled accurately. So with the exception of structures with approximately 60° bond angles, the parameterisation is extremely good for the Si(001) surface, and its interactions with Si and H.

3.5 Si-Bi and Bi-Bi bonding

The creation of a silicon-bismuth and bismuth-bismuth parameterisation presented a number of challenges: Bi is sufficiently large that the valence electrons achieve speeds through the core which make relativistic corrections necessary; there is no AB Si-Bi compound, making the fitting of electronic parameters to an experimental bandstructure impossible; and Bi in its native state is a semimetal in an arsenic-like structure, with puckered graphitic sheets which are stacked sufficiently close together that the coordination is almost 6-fold (which means that fitting to a tetrahedral bonding environment is hard).

Accordingly, the parameterisation has been put together from a number of different sources. The effect on the E_s and E_p levels of relativistic corrections have been taken from a full, relativistic solution of the Dirac equation for the Bi atom, using Hideaki Fujitani's ATMDAT program. The offset of the Bi E_s level from the Si E_s level was taken from the same program. The pre-relativistic levels were taken from Harrison (1980), as were the hopping parameters.

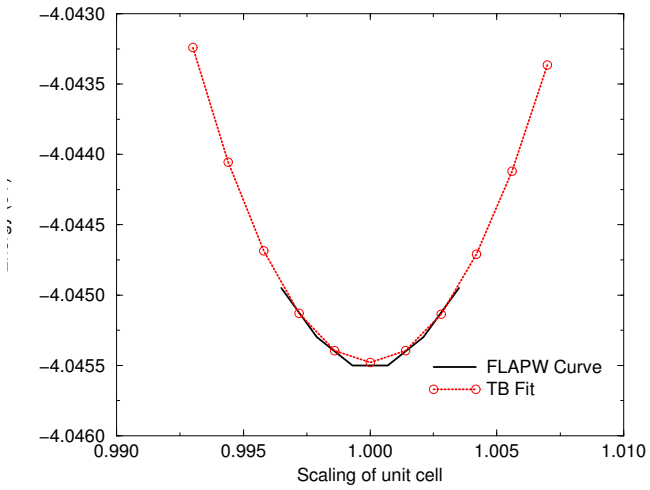

Figure 3.10: The scaling for a fictitious tetrahedral Bi cell, calculated using FLAPW and the TB parameterisation. a_0 is 7.135 Å.

E_S	E_P	$h_{ss\sigma}$	$h_{sp\sigma}$	$h_{pp\sigma}$	$h_{pp\pi}$	ϕ_0
-15.3	-5.87	-1.684	2.213	3.897	-0.974	5.15855

r_0	r_c	d_c	n	n_c	m	m_c
3.09	0.0000	0.0000	3.7847	0.0000	5.5413	0.000

Table 3.7: The parameters for Bi-Bi interactions.

The scaling for the Bi-Bi bonds and the Si-Bi bonds is again hard to fit. An initial estimate of the Bi-Bi bond length in tetrahedral form was made by assuming that the ratio of bond lengths for Bi in fcc and diamond was the same as for silicon; this ratio was then applied to the Bi bulk bond length, and taken as the diamond bond length. The Si-Bi bond length was taken as the geometric mean of Si-Si and Bi-Bi.

However, this is extremely crude, and a better fit was constructed. Binding energy curves for Bi in tetrahedral form and zincblende SiBi were calculated using Hideaki Fujitani's FLAPW code, and the tight binding scaling parameters and repulsive potentials were fitted to these. The fits are shown in Figures 3.10 and 3.12. The most interesting feature of this fitting is that the bond lengths derived from the FLAPW calculations were very much larger than those predicted from the simple prescription given above (3.09 Å rather than 2.52 Å). The reason for this can be seen in Figure 3.11, where the s-band is clearly about 10 eV below the Fermi level, and is thus forming a filled, stable lone pair state, and making Bi fundamentally

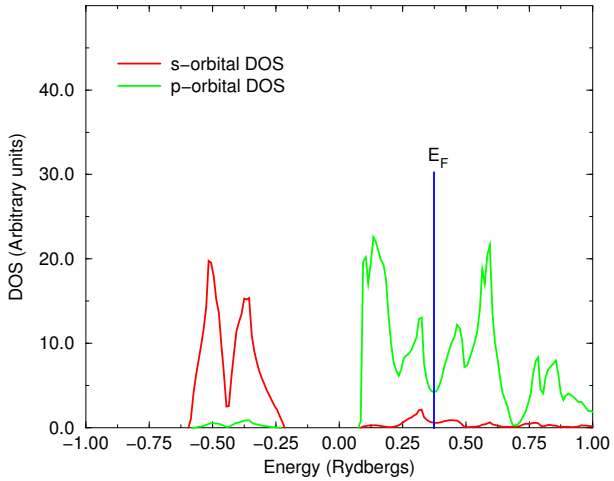

Figure 3.11: The density of states for a fictitious diamond Bi cell, calculated using FLAPW.

p-valent; thus the tetrahedral bonding environment will be much much worse, relatively speaking, for Bi than fcc would be for Si, making the bond lengths longer than anticipated. A similar effect, though far smaller, can be seen in Figure 3.13.

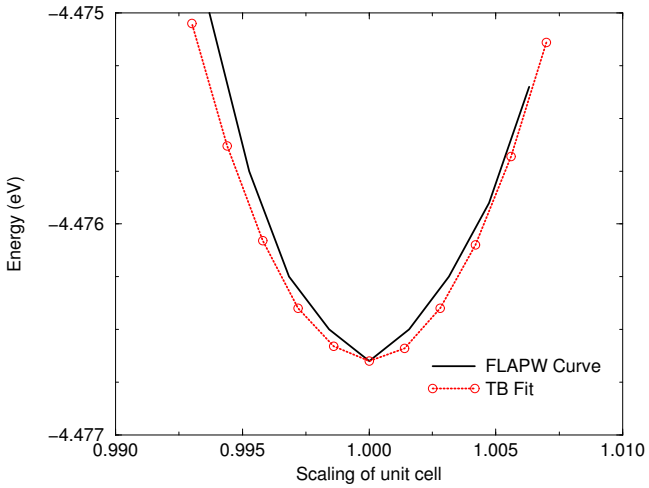

Figure 3.12: The scaling for a fictitious tetrahedral zincblende SiBi cell, calculated using FLAPW and the TB parameterisation. a_0 is 6.34 Å.

$h_{ss\sigma}$	$h_{sp\sigma}$	$h_{pp\sigma}$	$h_{pp\pi}$
-1.802	2.367	4.169	-1.042

r_0	r_c	d_c	n	n_c	m	m_c	ϕ_0
2.75	0.0000	0.0000	2.7673	0.0000	4.2876	0.0000	6.0138

Table 3.8: The parameters for the Si-Bi interaction.

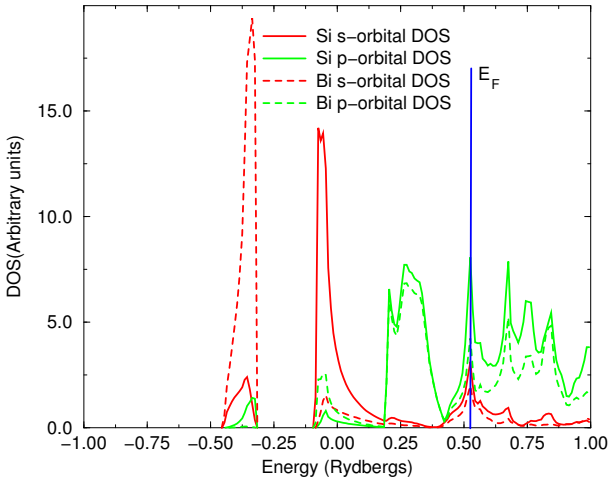

Figure 3.13: The density of states for a fictitious zincblende SiBi cell, calculated using FLAPW.

The Bi-Bi bonding in the tetrahedral structure cannot really be tested, except through comparison of energies with experiment; however, the robustness and transferability can be investigated by modelling the native bismuth structure, or at least a puckered sheet of bismuth, which it extremely close to the native structure. The simulation yielded a cell with nearest neighbour spacing of 3.09 (compared to 3.10 in reality) and puckering angles of 100°, compared to 90° in reality. It seems that the Bi-Bi bonding is remarkably well described.

The Si-Bi bonding is not so easy. There is no AB alloy, and we will have to trust that the fitting performed is adequate. There are good reasons to assume this: the fitting procedure followed is very similar to others followed in this thesis and elsewhere, which have produced good results; the Harrison scheme, which was used for the E_s and E_p levels of bismuth, is extremely successful for semiconductors; and the results (presented in Chapter 5) are very sensible, with Bi taking up tetrahedral bonding positions, and allow a clear understanding of a physical problem to be developed.

Conclusions In this chapter, the issues of which linear scaling method to use for modelling semiconductors and how to fit parameterisations for tight binding have been addressed. The Density Matrix method has been found conclusively to be the best method for modelling semiconductors, particularly when there are narrow features in the density of states. Parameterisations for Si-Si, Si-H and Si-Bi bonds have been described, along with the fitting procedures. Those tests available (which are far more numerous for Si-Si and Si-H than for Si-Bi) suggest that the parameterisations are extremely good.

Chapter 4

The Clean Silicon(001) Surface: Defects and Steps

...Those sciences are vain and full of errors which are not born from experiment, the mother of certainty...

Leonardo Da Vinci, 1452-1519

In this chapter, the simulations of real systems begin. Before any of the more complex reactions can be understood, the structure of the surface must be investigated. The clean Si(001) surface has been well studied for many years, although the detailed structure has only recently been confirmed. I shall describe this structure, as it is the substrate for all the calculations which follow. Then I shall move on to a study of strained bonding, which has led to the confirmation of the structure of the most common defect in the Si(001) surface, and one of the step terminations. This also illustrates very elegantly the importance of interaction between experiment and theory. The structure of steps, and their kinking, will round off the chapter.

4.1 The perfect surface

The silicon surface was first observed in 1959 in a LEED (Low Energy Electron Diffraction) study by Schlier and Farnsworth (1959). They found evidence for a (2x1) periodicity, which they attributed to a reconstruction based on dimer formation. Further investigations showed evidence of higher periodicities (p(2x2) and c(4x2) for example) which depended on sample preparation and treatment. A variety of models were proposed, which featured dimer models, and various other more drastic reconstructions (e.g. the dimer plus chain model of Northrup (1985) or the vacancy model of Poppendieck, Ngoc and Webb (1978)).

With the advent of STM, a real space image of the silicon surface could be obtained, an example of which is shown in Fig. 4.1. Tromp, Hamers and Demuth did just this (1985) and followed it up with a more comprehensive paper (Hamers, Tromp and Demuth 1986) in which they compared the various models to their pictures. This effectively established the dimer as the basic unit of the surface reconstruction[1]. However, further controversy ensued. There was significant doubt about whether the dimers were symmetric or buckled (as first suggested by Chadi (1979)), and also about the high level of defects in the surface. A schematic diagram of the different reconstructions is shown in figure 4.2.

Various different modellers predicted that symmetric (Pandey 1985, Ihara *et al.* 1990) or buckled (Chadi 1979,Roberts and Needs 1990) dimers were more stable - the essential point being that the energy difference between the two was small. Symmetric dimers seemed unlikely, as the surface would become metallic due to the degeneracy of surface states. The question as to why dimers might appear symmetric in the STM if they were really buckled was answered with the idea that they were flipping very quickly between the two possible buckled states (Dabrowski and Scheffler 1992). The breakthrough in understanding came when Wolkow (1992) imaged the Si(001)-(2x1) surface as he cooled it to 120 K. He saw a significant increase in the number of buckled dimers as he cooled, finding a maximum of 80% buckling at 120 K. He also observed

[1]It is interesting to note that Harrison (1980), p. 240, said of the dimer reconstruction, "...such a reconstruction seems unlikely, since all low-energy dangling hybrids are occupied and all high-energy dangling hybrids are empty and bonding cannot occur without promoting electrons to high-energy hybrids". This interpretation has been proved wrong by experiment - another example of why interaction is vital !

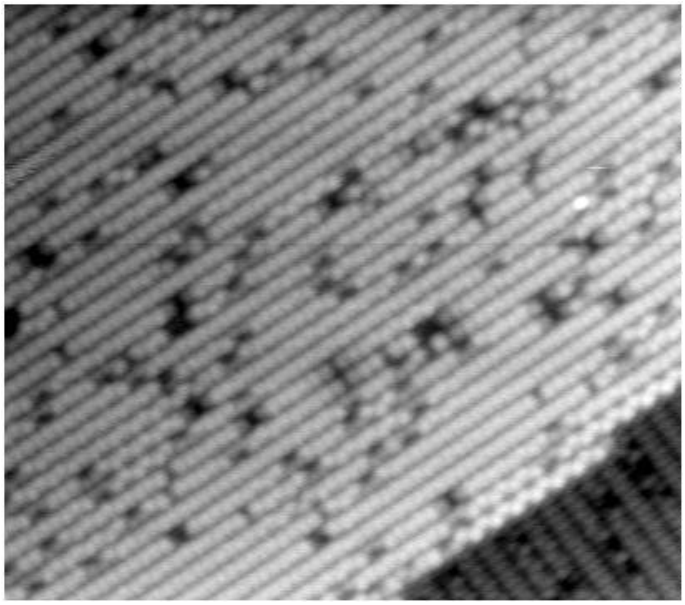

Figure 4.1: An STM image of the Si(001) surface. The lines running diagonally across the image from bottom left to top right are the rows of dimers (as shown in Fig. 4.2). There is a step in the bottom right corner, where the dimer row direction rotates by 90°. Some pinned, buckled dimers can be seen at the step edge. *Image courtesy of Holger Nörenberg.*

(as have others) that buckling at room temperature occurs next to step edges and defects, where pinning will raise the barrier for flipping.

The conclusion which can be drawn from all of this work is that the silicon (001) surface consists of alternately buckled dimers which are bistable, and oscillate backwards and forwards between their two states. Previous calculations which found that the symmetric orientation was more stable may well have been insufficiently converged (Dabrowski and Scheffler 1992), and a recent study (Ramstad, Brocks and Kelly 1995) has found that, with insufficient convergence with respect to plane waves, symmetric dimers are more stable, and that the difference between p(2x2) and c(4x2) is too small to resolve accurately in LDA.

As a demonstration of the tight binding parameterisation which will be used throughout this work, I have performed calculations of the energetics of the three different reconstructions described above, using the standard unit cell. Relative to the c(4x2), which was found to be (just) the most stable, the p(2x2) has an excess surface energy of 0.1 meV per dimer (which is well below the limit of resolution of LDA, let alone tight binding; essentially the conclusion to draw here is that they are equally stable) and the (2x1) has an excess surface energy of 77 meV per dimer, which is in good agreement with LDA calculations. In general, I have used the p(2x2) reconstruction in this work, as the unit cells are a factor of two smaller, and, when working with steps, less complex.

4.2 Electronic enhancement and strained bonding

It was recently observed (Owen *et al.* 1995) that, at low imaging voltages (i.e. relatively near to the Fermi level), the single missing dimer defect on the Si(001) surface appears to have been enhanced - the dimers either side of it brighten markedly. In this section, the behaviour is presented, and modelling solutions from both LDA and TB are given, along with a simple physical interpretation of the phenomenon.

In STM images of the Si(001) surface, for instance Fig. 4.1, there are always a certain number of dark features, which have been attributed variously to dimer vacancies (DVs) (Tromp, Hamers and Demuth 1985; Hamers, Tromp and Demuth 1986), adsorbed water molecules (Anderson and Köhler 1993) and interstitial dimers (Ihara *et al.* 1990). In the experiments which were performed (details of which are given in Owen *et al.* (1995)), there was almost no change in the appearance of the surface between room temperature and 400^oC, by which time water should have desorbed. The interstitial dimer has been found to be unstable in the extensive modelling study of defects by Wang, Arias and Joannopoulos (1993), and so the majority

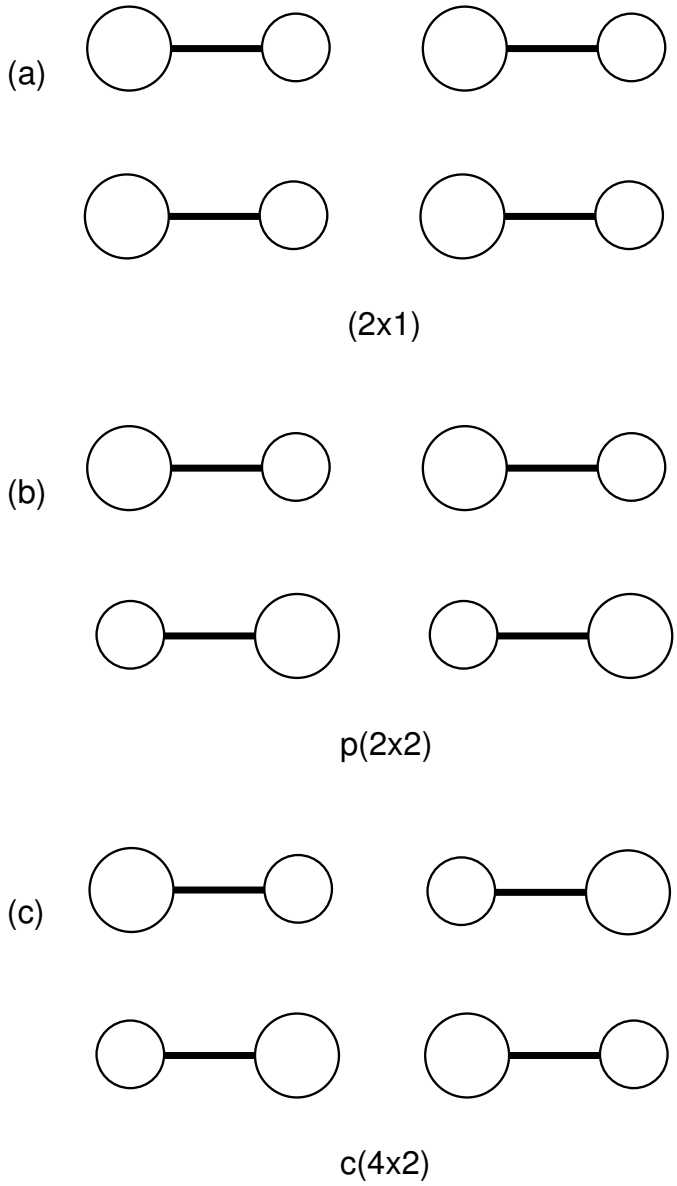

Figure 4.2: A schematic diagram of the Si(001) surface, showing the three major reconstructions: (a) (2x1) (b) p(2x2) and (c) c(4x2). The dimers are all buckled, with the larger atoms higher than the smaller ones.

of the dark features seen on the surface are taken to be missing dimer defects[2], known as 1DVs (or, if two dimers are missing, 2DVs, etc).

These defects have a local effect upon the surface, which may be important in adsorption or desorption (Wu, Ionova and Carter 1993) or processes such as diffusion (Roland and Gilmer 1993). Recently, Carter and Carter have found evidence that the key factor in surface reactivity is the number of dangling bonds on the surface (Carter and Carter 1995) which will be affected by the structure of a defect. The structure of defects is therefore important when studying reactions on the surface.

When a single dimer is removed, the exposed second-layer atoms can either move apart and relax without forming bonds across the gap, or move together to form bonds across the gap, as had been first suggested by Pandey (1985). Wang, Arias and Joannopoulos (1993), in a careful study of defect structures and energetics using LDA, calculated the latter structure to be lower in energy by 0.42 eV. Since the two proposed structures have different numbers of dangling bonds, they might be expected to have different electronic structures and

[2]The possibility of these defects being due to the segregation to the surface of bulk impurities such as carbon or nickel has also been ruled out. See Owen *et al.* (1995) for a further discussion of this point.

reactivities. Hamers and Köhler (1990) found using Scanning Tunneling Spectroscopy (STS) that the surface near 1-DV and 2-DV defects has a reduced band gap compared to the perfect surface. It was not known, however, which of the two proposed structures would lead to a reduced band gap.

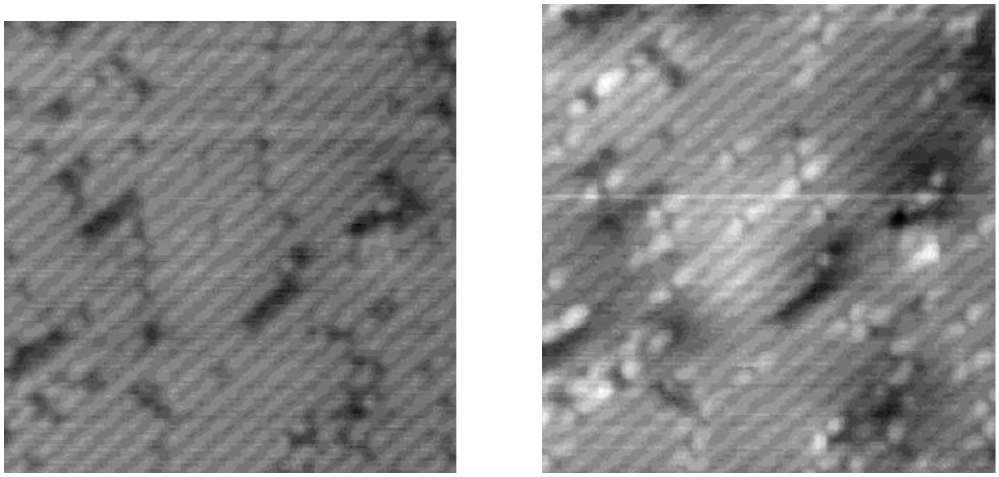

Figure 4.3: STM images of the Si(001) surface at (a) -1V and (b) -0.5V. The dimer rows run from bottom left to top right, and defects are visible as dark features in the dimer rows. Single missing dimer defects show an enhancement in (b).*Images courtesy of James Owen*

Images of the same area of the Si(001) surface at two different bias voltages (-1V and -0.6V) are shown in Figure 4.3(a) and (b). While the resolution is not good enough to make out individual dimers, defects can be seen as dark features of various sizes in the dimer rows. The image at lower bias voltage, i.e. nearer to the Fermi level (Fig. 4.3(b)), shows a marked change from the image at higher bias voltage ((Fig. 4.3(a)). The most common feature, the 1DV, is now highlighted on either side with a white patch. It is clear that this must be an electronic effect, as the only change between images is of bias voltage. This is indicative of a localised state further into the band gap than for normal dimers, which is the cause of the reduced band gap found by Hamers and Köhler. It is interesting to note that around some other defects, typically larger ones, a darkening is visible, which is indicative of a lack of charge or a surface state which is further out of the band gap than for normal dimers. For both of these features, it is the dimers either side of the defect which are affected, rather than the defect itself.

Three possible structures for the defect were modelled with TB (illustrated schematically in Figure 4.4): with the second layer atoms forming a bond across the gap ('bonded'); with the second layer atoms relaxing away from the gap ('broken'); and with one pair bonding across the gap and the other pair relaxing away from it ('twisted'). A unit cell consisting of two standard unit cells side-by-side was used, making a cell six dimers long, and two dimer rows wide. A single defect was created in one of the rows. This found that the bonded defect was 0.45 eV lower in energy than the broken defect - a value remarkably close to that of Wang, Arias and Joannopoulos (1993). The twisted structure was the highest in energy.

The charge density for the system was found by putting a reduced version of the tight binding unit cell into CASTEP , without re-relaxing the atomic coordinates. The dimer rows were truncated to four dimers in length, and the slab was reduced to 5 layers of silicon. The top 25 Kohn–Sham eigenstates were then projected out, with their energies given by the corresponding Kohn–Sham eigenvalues. STM images were simulated in a crude, but in this case effective, manner. The density of states in the tip was taken to be uniform, with the occupation given by a Fermi function with kT=0.025eV. The imaging bias was represented by an offset between the Fermi level in the tip and the sample. The contribution of each eigenstate in the sample to the tunneling current was taken to be equal to the available density of states in the tip. The tip was taken to be point-like, which results in a much improved resolution over the real STM images.

The results of this modelling for the bonded and broken structure are shown in Figure 4.5, along with the clean surface for comparison. At high bias voltage (left hand column), the two structures look extremely similar, with the dimers either side of the defect being virtually indistinguishable. At low bias voltage (right hand column), the situation has changed greatly. The bonded structure shows an enhancement of the dimers either side of the defect, whereas the broken structure shows a clear darkening of the dimers either side of the defect.

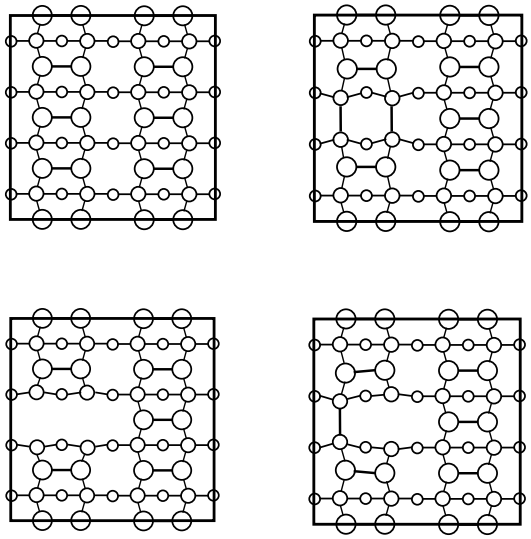

Figure 4.4: A schematic picture of the clean surface and the three defect structures considered in the modelling. From left to right, in the top row the clean surface and bonded structure, and in the bottom row, the broken and twisted structures.

One interesting method of comparison between STM and theoretical calculations is via LDOS: the STM tunneling current (and hence the contrast seen) is *in some way* related to the LDOS for each atom. The result of a tight binding calculation in k-space (which used the same cell and parameters as before) giving the LDOS for the up atoms of a dimer on a perfect Si surface, and a dimer next to a rebonded 1DV are shown in Figure 4.6.

Fig. 4.6 clearly shows (for filled states only, as discussed in Chapter 3) that there is a change in the surface states in the band gap for the dimer next to the 1DV, yielding a large state close to the Fermi level. This state is what is responsible for the enhancement. A simple LCAO argument demonstrating how this works is given in Figure 4.7. As the electronic states progress from atomic orbital levels to hydrids and then bonds and bands, the structure becomes more complex. At the top of Fig. 4.7, the situation is shown for bulk Si, which forms perfect tetrahedral bonds at 109.3^o. This bonding and structure give rise to the band gap which makes Si an insulator. Below this is a picture of what happens at the surface. The atoms in the dimers are pulled away from the perfect tetrahedral angle, which reduces the bonding-antibonding splitting at the bond stage, and results in bands in the gap - also known as surface states. The picture would change even further if more constraint were put on the dimers - for instance by second layer bonding which pulls them away from the equilibrium dimer position. This is clearly what is happening in the STM images shown in Fig. 4.3(b), resulting in the enhanced behaviour either side of the defect. As an extension to this, consider what would happen if the second layer atoms were allowed to relax. In this case, the dimers would be *closer* to their equilibrium structures (i.e. the bulk structure) and the surface states pushed *away* from the Fermi level. This explains the darkened appearance of the broken defect, and suggests that any darker areas on the surface (as seen in Fig. 4.3(b)) are associated with relaxed bonds and structures similar to the broken defect.

Thus the structure of the 1DV has been identified conclusively, and it is actually *less* reactive than the normal surface, as one dimer is missing and the second layer is rebonded. Defects need not enter into calculations of reactions, unless specifically desired.

4.3 Step energies and structures

Si(001) wafers, as obtained by cleaving from a boule grown commercially, are not perfectly flat. Indeed, there will be a large step density on the surface, though this can vary from area to area. There are two kinds of steps, because of the (2x1) periodicity: one where the dimer rows in the upper terrace are parallel to the step edge (A-type) and the other where they are perpendicular to the step edge (B-type); these are illustrated in Figure 4.8. The B-type step edge has two possible terminations (the unbonded and the rebonded) depending

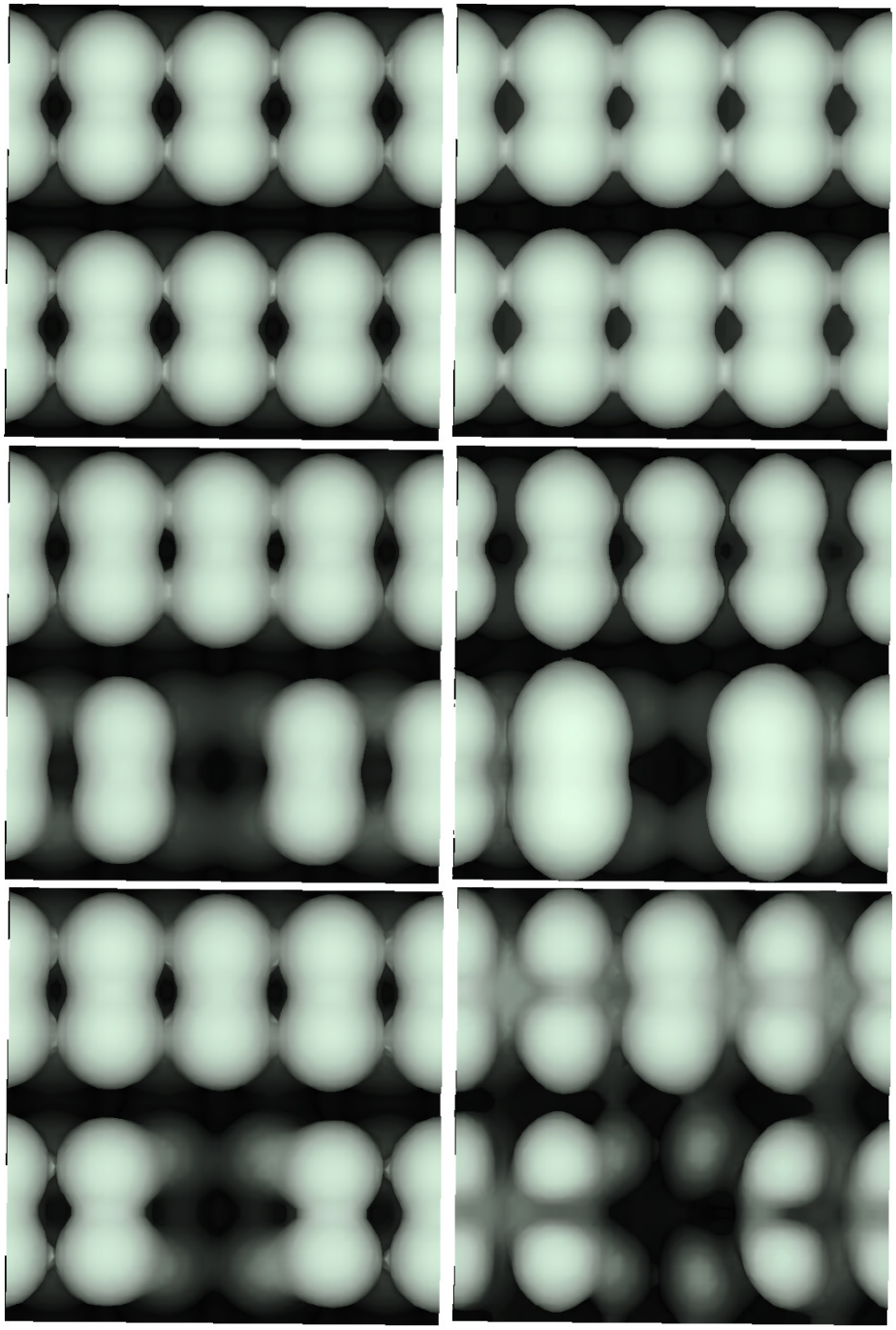

Figure 4.5: The simulated images for the clean surface (top row), the bonded defect (middle row) and the broken defect (bottom row). The left hand column represents imaging far from the Fermi level, while the right hand column represents imaging near the Fermi level.

on where the step ends relative to the dimer rows in the lower terrace - the rebonded is shown in Fig. 4.8, and the unbonded would have another dimer in the upper terrace, removing the lower terrace dimers just above the dotted line.

The A-type step is generally very long and straight, with an occasional kink seen at room temperature, whereas the B-type step is seen to be kinked often at room temperature. In this section, we shall examine the size of unit cell required to model steps successfully, and the kinking energy for A-type and rebonded B-type steps.

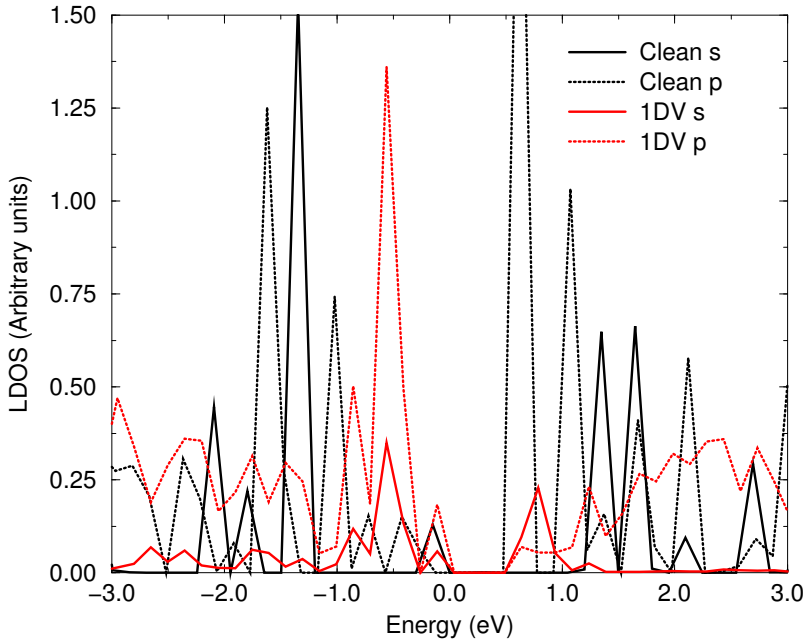

Figure 4.6: The local densities of states calculated using tight binding for the up atoms of a dimer in a clean surface (black) and next to a 1DV (red). Note that the empty states are not accurately modelled by tight binding, and are only shown to indicate the rough band structure. All energies are relative to the Fermi level.

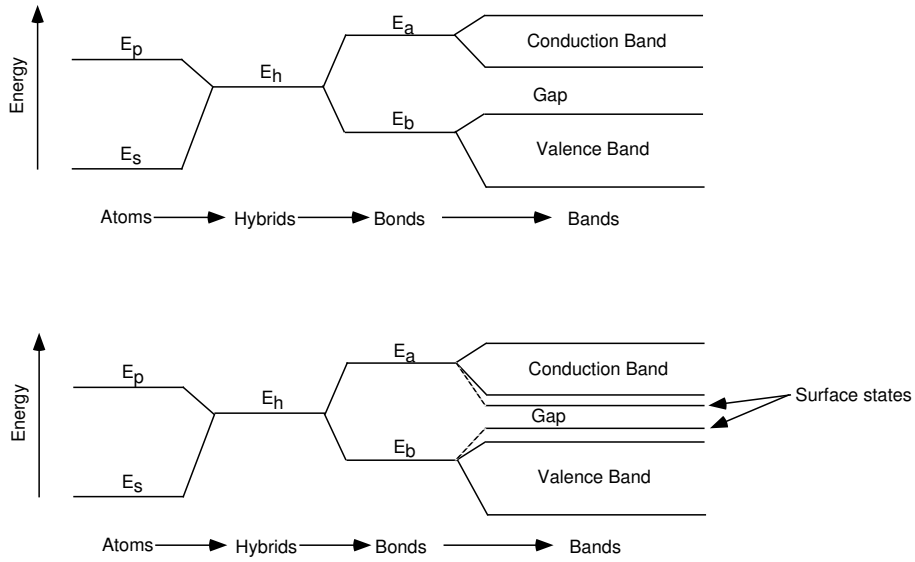

Figure 4.7: A simple LCAO picture of bonding, showing how atomic orbitals hydbrise, form bonding and anti-bonding orbitals, which combine to form bands in the solid. The top diagram shows conceptually the structure for bulk silicon, while the bottom demonstrates the effect of distorting bonds - bands in the gap.

4.3.1 Terrace sizes for rebonded and unbonded steps

In the repeated supercell method of simulation (as described in Chapter 2), care must be taken to isolate features in one cell from their images in other cells. This is generally done by making the cell large enough

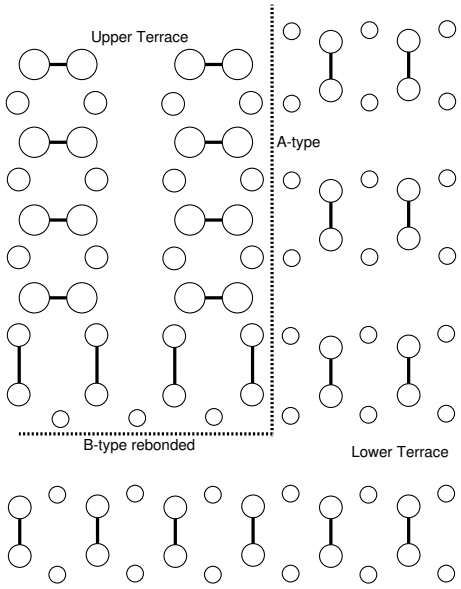

Figure 4.8: A schematic picture of the two types of step on the Si(001) surface: the A-type and the B-type (here shown in its rebonded configuration).

that any interaction is negligible. With steps, the isolation must be achieved across two terraces, one at a lower level and one at an upper level. The excess surface energy for rebonded and unbonded B-type steps and for the A-type step have been calculated using tight binding. The standard unit cell was used, but lengthened (to lengths given in the figures, up to 18 dimers long) and with some of the top layer removed to give a pair of steps. These are shown in Figures 4.9 and 4.10.

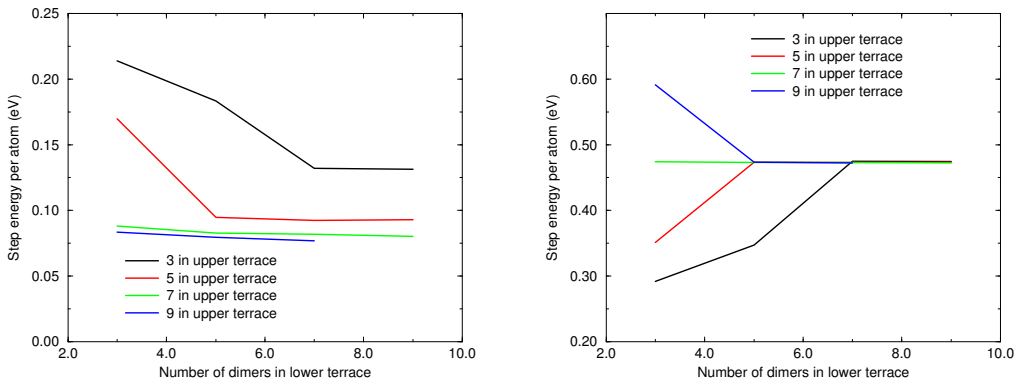

Figure 4.9: The step energies for different terrace sizes for (a) the rebonded B-type step and (b) the unbonded B-type step on Si(001).

To calculate the excess step energy for these cells, I found the excess surface energy (i.e. $E_{cell} - n \times E_{bulk}$, where there are n atoms in the unit cell and the energy for a bulk silicon atom is E_{bulk}) for the stepped cell and a clean, unstepped surface of the same surface area. The excess step energy is then given by the difference between these two (though divided by two, as there are two steps in the unit cell). There was no need to compensate for the hydrogen termination in these calculations as the termination energy was the same for each cell, and cancelled.

The calculation was not quite that simple, however. The lower terrace has only four layers of silicon atoms before the bulk-like silicon, whereas the upper terrace has five layers. The difference in excess surface energy between the four layer and five layer cases is only 0.02 eV per dimer, but without this correction

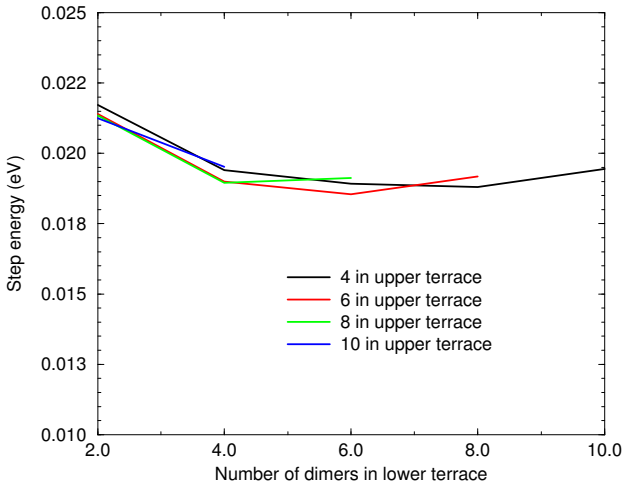

Figure 4.10: The step energies for different terrace sizes for the A-type step on Si(001).

the step energy shown in the figures has a linear increase, which would lead to an infinite energy at infinite separation.

The final values show that for the B-type step, the energy is converged with respect to the lower terrace within 5 dimers, and with respect to the upper terrace within 5 or 7 dimers. The values per atom shown in Figure 4.9 (a) and (b) are 0.08 eV/a for the rebonded step and 0.473 eV/a for the unbonded step. The A-type step is less sensitive to cell size, and is converged for all values of upper terrace width, and beyond four dimers in the lower terrace. The value of energy per surface lattice constant is 0.019 eV. It should be noted that the scale in Figure 4.10 is ten times finer that in Figure 4.9(b) leading to the appearance of poorer convergence.

There are several values to compare these energies against. Chadi (1987) calculated the structures and energies of single and double height steps on the Si(001) surface using tight binding. He found a value of 0.01 eV/atom for the A type step, 0.075 eV/atom for the rebonded B type step and that the unbonded step is "energetically less favourable by at least 0.16 eV/atom". The only other quantum mechanical study (Bogusławski *et al.* 1994) used DFT, but only had 3 dimers in the upper terrace and did not quote step edge energies, so is not very helpful. The experimental evidence will be discussed fully in the section on kinks.

4.3.2 Step edge enhancement

In the same way that certain defects show an enhancement at low imaging voltage, so also certain step edges show this enhancement. An example of this is shown in Figure 4.11.

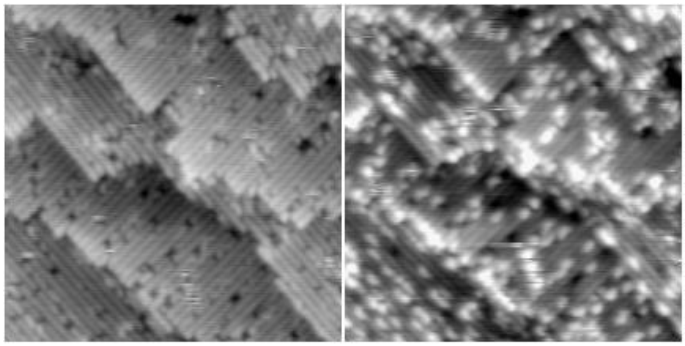

Figure 4.11: The Si(001) surface at -1V and -0.5V, showing B-type step edge enhancement. *Images courtesy of James Owen.*

The significance of this enhancement extends to two other chapters; both the stability and diffusion of hydrogen near step edges depends on the structure of the step in question (Chapter 6) and the growth of silicon occurs at step edges and islands, but may be inhibited by hydrogen (Chapter 7). The rebonded B-type step (which is the lower energy termination) bears a great deal of similarity to one half of the rebonded defect discussed earlier, and might be expected to have the same enhancement behaviour seen before for that system. Accordingly, a similar calculation was performed for the unbonded and rebonded B-type steps as was performed for the defect, using LDA.

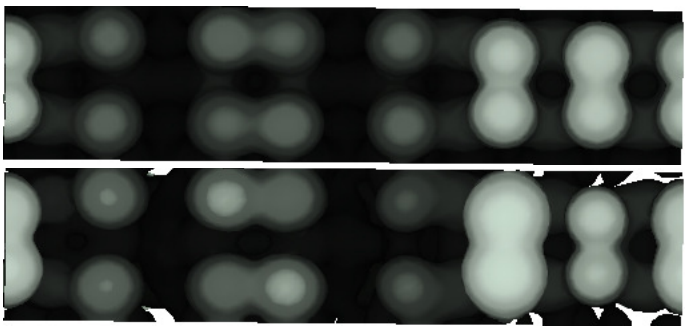

Figure 4.12: The rebonded step, showing imaging far from the Fermi level (above) and near to the Fermi level (below).

The unit cell used was eight dimers long, and five layers deep, with the bottom layer constrained to lie in bulk-like positions, and terminated with hydrogen. Five of the top layer dimers were removed, to create a double step in the unit cell, which was only three dimers long. While this may seem contradictory to the conclusions drawn above about length of step required, the cell was only being used for charge densities and not energetics, and the positions, and hence strains, for the atoms are acceptably similar to a larger cell. This choice was a compromise between isolation and computational possibility, as a larger cell would have been prohibitively slow. To obtain symmetrical images, as would be seen in STM, two calculations had to be performed with the dimers in the second calculation flipped the other way to the dimers in the first calculation. The charge densities were then added to show a symmetrical image. The simulated STM images for the rebonded step edge are shown in Fig. 4.12, and for the unbonded step edge in Fig. 4.13. Clearly, the same effect is at work here as in the defect case, and an enhanced step edge can be taken to be a B-type rebonded step edge (the A-type or B-type identification is made from the direction of the dimers in the upper terrace).

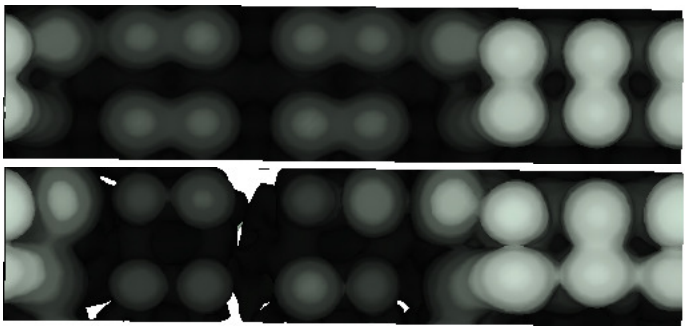

Figure 4.13: The unbonded step, showing imaging far from the Fermi level (above) and near to the Fermi level (below).

There is a question about the saturation of step edges with hydrogen during disilane growth which will be addressed fully in Chapters 6 and 7. The essence of the question is whether or not saturating the *bottom* of the step will promote or decrease the enhancement at the top. This can be easily answered by using the same technique as above, and the simulated images are shown in Figure 4.14.

It is clear that the hydrogenation of the bottom part of the step does not stop the enhancement of the step, and that this should not be taken as a counter to the argument in Chapters 6 and 7.

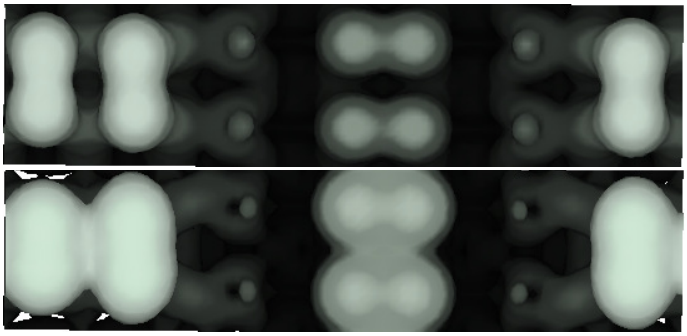

Figure 4.14: The hydrogenated, rebonded step, showing imaging far from the Fermi level (above) and near to the Fermi level (below).

4.3.3 Kink structures and energies for A- and B-type steps

At growth temperatures, the step edges on the Si(001) surface become mobile, and kinks are seen to form and move around. Kinks[3] in the A-type and B-type steps are illustrated schematically in Figure 4.15. At room temperature, the A-type step is extremely smooth, while the B-type step kinks frequently, and appears rough. The step edges can play a large part in the surface dynamics of growth, and understanding the energetics of kinking is important. There has been a large amount of experimental work on step kinking (Pearson *et al.* 1995a; Pearson *et al.* 1995b; Swartzentruber *et al.* 1990; Swartzentruber *et al.* 1993; Bartelt, Tromp and Williams 1994; Zandvliet *et al.* 1992) which has produced various results: firstly, the step edges energies mentioned above, along with a "corner" energy (that is, the energy generated by the corners of a kink); secondly, the energetic behaviour of the separation and depth of kinks; finally, the dynamics of kink formation, diffusion and extinction. As the majority of results are for the B-type step kinks, these will be presented first.

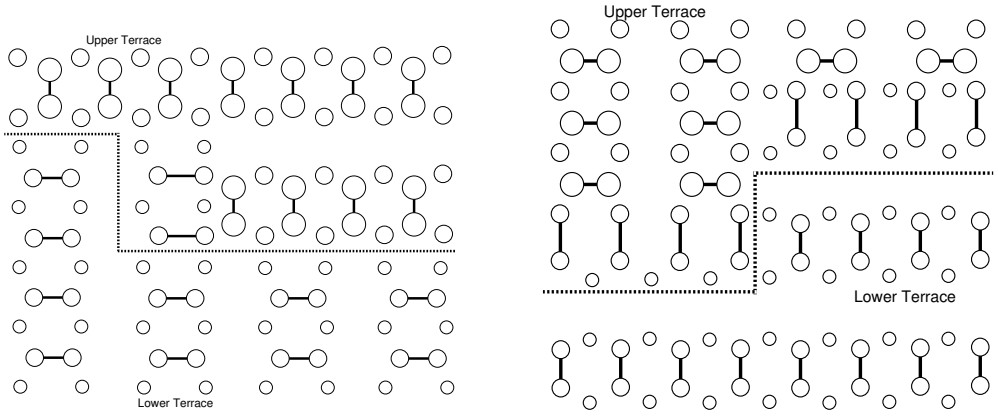

Figure 4.15: The structure of the A-type step and the rebonded B-type step after the introduction of a kink (with two dimers in the kink) as used in the calculation of kinking energy. The small circles represent the rebonding of the B-type step edge.

As was the case with the step energies discussed above, it is important to understand how big the supercell must be to prevent kinks from interacting. Two cells were investigated using TB, both of which had two kinks (one up and one down) of depth two atoms (this terminology may be confusing; the kink depth is exactly as illustrated in Fig. 4.15 for the B-type step, where the depth is two atoms). The unit cell was the equivalent of 14 dimers long, with four dimer rows, and the usual depth and termination. In the first cell, the separation of the kinks in both directions was increased simultaneously, whereas in the second, only one segment of step edge was lengthened. These cells are illustrated schematically in the insets of Figure 4.16 and the energies for the kinks are given for the first case in Fig. 4.16(a), and in the second case in

[3]Note that in the entire discussion, "a kink" is taken to mean a 90° bend, a length of the other kind of step and *another* 90° bend. Thus, after a kink in a B-type step, the B-type step will continue.

Fig. 4.16(b). From these it can be seen that the energy associated with a kink is converging to a limit with separation, and that the kink-kink interaction is small even with two dimer rows separation (less than 0.01 eV) and negligible at three. The kinking energy of the B-type step for a cell with a kink of depth one atom was also investigated, and proved to be 1.04 eV - mainly because of the unfavourable unbonded termination forced by the depth of one dimer. Kinks of depth one are almost never seen in experiment, and therefore will be discounted.

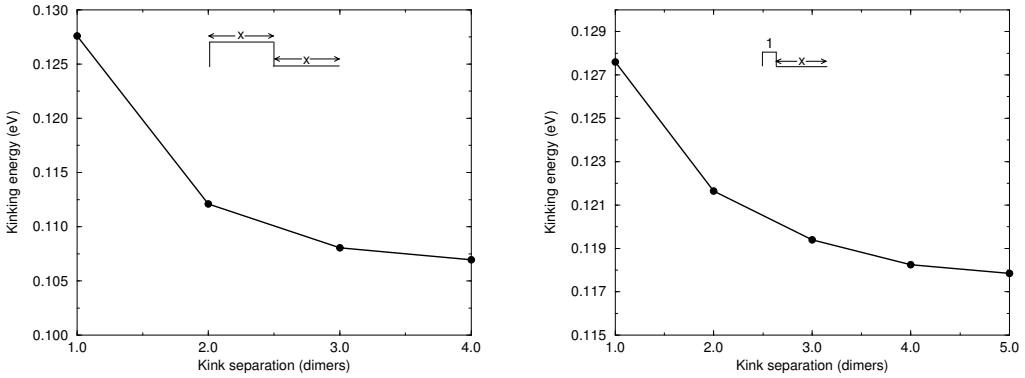

Figure 4.16: Convergence of kinking energy. (a) for isolated kinks and (b) for two kinks with separation 1 dimer. The insets illustrate how the x-axes relate to the unit cells used.

Now that the isolation issue has been addressed, the first question to be answered is that of how best to compare with experimental results. All subsequent experimental papers rely on the analysis of Swartzentruber *et al.* (1990), whose formalism will be examined in detail, and a useful extension derived. To summarise their assertions, they said that when examining the kinks on a B-type step, there are two distributions: firstly, the separation of kinks, s, which behaves as $P(s) = P_k(1-P_k)^{s/2-1}$, where P_k is the probability of any kink at the site under consideration; secondly the depth of a given kink, n, which has a probability proportional to the expontential of its energy, $P(n) \propto \exp\left[(-n\varepsilon_A + E_C)/\mathrm{kT}\right]$, where ε_A is the energy of the step exposed when kinking (the A-type step).

To understand these better, consider an arbitrary length of B-type step with a step energy of ε_B. If we introduce a kink into this system, the energy will be raised by $n\varepsilon_A + E_C$, where n is the depth of the kink, ε_A is the energy of the step edge exposed and E_C is the "corner energy". The probability of this kink occuring will be, as stated above, proportional to the exponential of its energy. The probability of *any* kink occurring at a given site is then simply the sum over the probabilities of all kinks, divided by the partition function:

$$P(n) = \exp\left[-\left(n\varepsilon_A + E_C\right)/\mathrm{kT}\right]/Z, \tag{4.1}$$

$$Z = 1 + 2\sum_{n=1}^{\infty} \exp\left[-\left(n\varepsilon_A + E_C\right)/\mathrm{kT}\right], \tag{4.2}$$

$$P_k = \frac{2\sum_{n=1}^{\infty} P(n)}{Z},$$

$$= \frac{2\exp\left[\left(-\varepsilon_A - E_C\right)/\mathrm{kT}\right]}{1 - \exp\left(-\varepsilon_A/\mathrm{kT}\right) + 2\exp\left[\left(-\varepsilon_A - E_C\right)/\mathrm{kT}\right]}, \tag{4.3}$$

where some simple algebra has been performed to obtain the final result. The form for $P(s)$ arises quite simply: to obtain a separation of $s/2$ dimers, there must be $s/2-1$ segments where kinking did not occur (hence the factor of $(1-P_k)^{s/2-1}$) and then a single occurence of a kink. Values for P_k are available directly from experiment as the number of kinks observed divided by the total number of sites. Swartzentruber *et al.* (1990) provided a plot of $P(s)$, and also derived values for ε_A and E_C from plots of $P(n)$. Thus the values available for comparison with experiment are $P(s)$, ε_A and E_C. All of these absolute quantities depend on the temperature at which the step edge is assumed to have equilibrated, which is not known; the ratio ε_A/E_C is a more independent measure of the accuracy of the calculation, and can be used for comparisons

with the theoretical results. It is possible to obtain a good estimate of the temperature by comparison with the theoretical results, once validated.

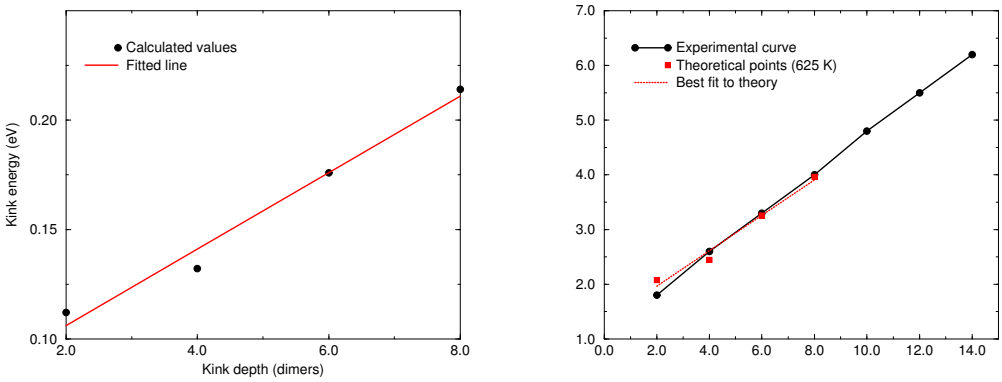

Figure 4.17: (a) Energies for a kink in a B-type step of different depths, calculated using TB. The fitted line gives an energy for the A-type step of 0.018 eV and a corner energy of 0.071 eV. (b) Comparison between experimental points for E(n) and theoretical points from (a) with kT taken equivalent to 625 K. Experimental points from Swartzentruber *et al.* (1990), by kind permission.

Figure 4.17(a) shows the variation of kink energy with a depth ranging from 2 to 8 atoms for a rebonded B-type step, calculated using TB. The unit cells used ensured that there was always a separation of five dimers between steps adjacent to kinks, so that the length varied between 14 and 26 dimers. The configuration is as shown in Fig. 4.15 for a depth of two atoms, and consists of two dimer rows, a kink up, two dimer rows and a kink down, repeating infinitely. The separation between kinks of two dimer rows was chosen as a compromise between accuracy and feasibility - the kink of depth eight dimers had 2,500 atoms in the unit cell. The line shown is a best fit, and gives a step edge energy for the A-type step exposed of 0.018 eV (which should be compared to 0.019 found in Section 4.3.1 and differs by only 1 meV) and a corner energy of 0.071 eV. These have a ratio of 0.246. Fig. 4.17(b) shows the experimental line for E(n)/kT from Swartzentruber *et al.* (1990), along with the same theoretical points in (a), but scaled with a kT appropriate for 625 K. The close correspondence suggests that the appropriate temperature to use for these experimental results is 625 K. The issue of temperature is discussed more fully below.

All the experimental papers to be described measured the kink distributions with an STM at room temperature (except where stated). The sample was annealed at a high temperature, and radiation quenched to room temperature to perform observations. This procedure leads to a large uncertainty in the temperature used to obtain absolute values for energies. Swartzentruber *et al.* (1990) found values of ε_A/kT=0.37±0.03 eV and E_C/kT=1.1±0.3 eV; the errors on ε_A/kT and E_C/kT are correlated, so that if ε_A/kT goes up, E_C/kT goes down (this arises from fitting a straight line to the experimental data). The ratio for the values given is 0.336, but going one standard deviation away to 0.34 and 1.4 gives 0.243 - so the values are well within experimental error of the theoretical results. A temperature of 875K was assumed, which gives absolute values of ε_A=0.028 eV and E_C=0.09 eV; in a later paper (Swartzentruber *et al.* 1993), a value of 625K is quoted, which gives values of ε_A=0.02±0.01 eV and E_C=0.06±0.01 eV, which is in much better agreement. Given later studies (Swartzentruber and Schacht 1995; Kitamura *et al.* 1993; Zandvliet, Elswijk and van Loenen 1992; Pearson *et al.* 1995a,b) which suggest that step edges start to move on a timescale of tens of seconds at between 550 and 600 K, this value of 625K seems much more reasonable, and fits well with the value derived from theoretical results. A temperature which agrees with ε_A, E_C and P_k will be given below. Other measurements of step kink distributions give results of: ε_A=0.026 eV and E_C=0.06 eV, with a ratio of 0.43 (Zandvliet *et al.* 1992), though these results include kinks forced by the miscut of the crystal which will give a low value of E_C and a correspondingly high value of the ratio; and ε_A=0.02 eV and E_C=0.07 eV, with a value of ε_B of 0.075 eV (Bartelt, Tromp and Williams 1994). This last study used a relatively new technique known as LEEM (low energy electron microscopy) which is a large area technique, resulting in relatively large errors in the values found. Overall, the experimental results are in good agreement with the theoretical values found, though the errors are sufficiently large that the best conclusion that can be drawn is that there is no contradiction.

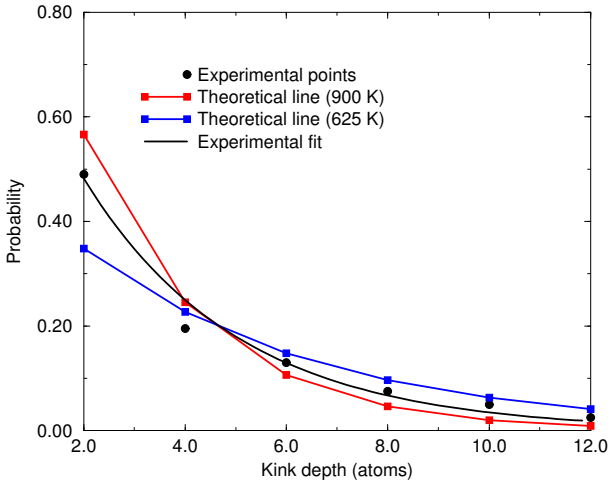

Figure 4.18: Probability of separation of kinks. Experimental points from Swartzentruber *et al.* (1990), by kind permission. The best fit line gives a P_k of 0.48. The theoretical curves have P_ks of 0.57 and 0.36, based on ε_A=0.018 eV and E_C=0.07 eV, and values of kT as indicated.

The final test of the results is to calculate a value of P_k from the theoretical values, using an appropriate value of kT. This is shown in Figure 4.18 for kT equivalent to 625 K and 900 K, along with the experimental results from Swartzentruber *et al.* (1990). The value of P_k is clearly highly dependent on the value of kT used, and the exact ratio of ε_A to E_C; however, the theoretical curve for 625 K is a good fit except for the first point. The experimental data as given in Swartzentruber *et al.* (1990) have no error bars, and cannot be made to fit the form of P(s) for any value of P_k. A temperature of 625 K gives values of ε_A and E_C of 0.018 eV and 0.076 eV for the set of experimental numbers which match the theoretical ratio; this, along with the good fit obtained in Fig. 4.17(b), may be an indication of the effective temperature at which the steps were quenched.

The kinking of an A-type step is rarer, but values of ε_B have been found nevertheless, though assuming that the corner energy is not different to that for a B-type step. To test this, a single calculation for the kinking of an A-type step was performed with TB, using a standard unit cell which was ten dimers long, and five dimer rows wide. This yielded a value for the corner energy of 0.078 eV, which is in almost exact agreement with the value from B-type steps. This is reassuring, given that the corners are almost exactly identical, and means that the energy found for the B-type step edge energy can be trusted. Swartzentruber *et al.* (1990) found that $\varepsilon_B/kT = 1.2\pm0.1$, but rather than using the E(n) plot above, they used a probability. This gives a ratio of 1.09 or 0.86 for E_C=1.1 or 1.4, and a value for ε_B of 0.065±0.005 eV, which is a little lower than the value of 0.08 eV found in Section 4.3. As they give almost no details of the calculation, it is hard to know where the differences might occur. The value obtained by Bartelt, Tromp and Williams (1994) of 0.075 eV is in much better agreement with the calculated results.

Conclusions In this chapter, the Si(001) surface has been introduced. After examining the reconstructions which characterise it, the effects of strain on bonding have been investigated, enabling the identification of a defect structure, and the confirmation of step structures. The structures and energetics of kinked steps have been studied, yielding good agreement with experiment. Now that the surface on its own has been examined, adsorbates can be added with confidence.

Chapter 5

Bismuth on Si(001)

Every valley shall be exalted, and every mountain and hill shall be made low: and the crooked shall be made straight, and the rough places plain:

Isaiah 40:4a (King James Version)

The growth of heteroepitaxial films, and heterostructure devices, is becoming increasingly important to the electronics industry. However, this form of growth often presents problems, as lattice mismatch between layers can cause segregation and islanding, which are potentially detrimental to device performance. One of the most interesting ways of solving this problem involves the use of a surfactant - a substance which stays on the surface during growth, and has some beneficial effect on the surface growth processes.

Ge/Si heteroepitaxial films are ideal for semiconductor manufacturers as they involve little change to existing fabrication technology, and yet allow the tuning of electronic structure and the extension of silicon devices to frequency regions where GaAs (or other expensive III-V materials) have been the only materials available. However, Ge on Si has a 4% lattice mismatch, and forms 3D islands when grown on a Si(001) substrate (an effect which is sometimes used deliberately - see Chapter 8). Ge also segregates to the surface when Si is grown over a Ge layer, leading to a diffuse interface (Fukatsu, Usami and Shiraki 1993). The use of gas-source MBE, where disilane and germane are used as the group IV source, rather than solid source MBE has been shown to reduce Ge segregation during growth of Si/Si_xGe_{1-x} heterostructures (Ohtani *et al.* 1994). With these sources, hydrogen acts as a surfactant, improving surface quality. More common, however, is to use a large atom which will not easily incorporate in the growing substrate. Group V atoms such as As and Sb have been very popular for such applications (Copel *et al.* 1989; Fujita *et al.* 1990).

Sakamoto *et al.* (1994) showed that a surfactant layer of bismuth has two effects on SSMBE of Ge/Si heterostructures[1], one beneficial and the other detrimental. A sandwiched structure was grown, alternating a thick layer of Si with a thin layer (4 ML) of Ge, four times and ended with a cap of Si. After the third layer of Ge, however, they deposited Bi, so that the growth of the final pair of layers before the cap (i.e. thick Si and thin Ge) were influenced by Bi. The structure was then investigated with high resolution TEM. The results of the first Ge layer were as expected - the Ge/Si interface was flat (suggesting that the Si surface on which the Ge grew was flat) but the Ge layer formed large islands, and the interface between the Ge and the next layer of Si was highly diffuse. Bi had a marked effect on this behaviour. It was deposited after the third layer of Ge, and had the immediate effect of making the new Si/Ge interface sharp, reducing the diffuseness seen before. It also prevented islanding of the Ge in the fourth layer; it had a detrimental effect, however, producing a far rougher Si layer (the fourth Si layer) and Ge/Si interface than had been seen before. Katayama *et al.* (1996) studied the effect of Sb and Bi on SSMBE of Ge on Si(001). An ion scattering technique was used to measure the mixing of Ge and Si, and the crystal perfection of the resulting layers. It was found that for increasing levels of surfactant, the intermixing was suppressed and the crystal quality improved. Sakamoto *et al.* (1993) used secondary ion mass spectroscopy (SIMS) to show that when Bi desorbs (above about 550° C), the amount of Bi left in the silicon is less than 5×10^{16} atoms cm^{-3} (the resolution limit of SIMS).

It is reasonably clear that Bi can act as a good surfactant for Ge/Si heterostructure growth. However, the behaviour of Bi on the Si(001) surface is still poorly understood, and so its surfactant action cannot be clearly understood. There have been several experimental studies, which have found variously (1x2)

[1]It is conventional to write the growing layer followed by the substrate, so Ge/Si indicates growing germanium on silicon.

reconstructions (Noh *et al.* 1994; Franklin *et al.* 1995; Lyman *et al.* 1996; Kim *et al.* 1996; Qian *et al.* 1996), (2x2) reconstructions (Noh *et al.* 1994; Kim *et al.* 1996; Qian *et al.* 1996) and a strain-induced (2xn) reconstruction (Hanada and Kawai 1991; Park *et al.* 1994). The STM studies of Noh *et al.* (1994) and Park *et al.* (1994) show that at room temperature, Bi initially adsorbs as adatoms, sited between Si dimers on top of the rows, which then pair up to form dimers, perpendicular to the underlying Si dimers, and still on top of the dimer rows. These dimers would give rise to a (1x2) periodicity. The Bi ad-dimers align from row to row to give a (2x2) periodicity. There have been some calculations for Bi on Si(001) (Tang and Freeman 1994, Franklin *et al.* 1995, Qian *et al.* 1996) which used cluster calculations in LDA, with the DMOL code. These calculations do not use relativistic corrections, and (from the information available) do not perform relaxations on the cell, which makes their results a little dubious. It was found that the lowest energy state for Bi on the Si(001) surface is as a dimer perpendicular to the underlying dimers, but do not consider the two possible sites for it (i.e. over the trench between Si dimer rows as well as on top of the Si dimer rows). The (2x2) structure was found to be more stable than the (1x2) structure, in opposition to the experiments. A recent experiment using high resolution electron energy loss spectroscopy (HREELS) and ultraviolet photospectroscopy (UPS) (Gavioli, Betti and Mariani 1997) has found that the Bi dimers on the surface have a larger band gap than the Si dimers and a large density of states above the Fermi level, a point which will be important later.

The work in this thesis is part of a study of Bi on Si(001) with Kazushi Miki and James Owen. The work in Oxford involved STM, while SIMS and RHEED results have been obtained in Japan. The specific areas being studied were the minimum surface energy state for a Bi dimer on (or in) the Si(001) surface, and then two of the more unusual features generated during the STM study: large numbers of straight lines, running for hundreds of nanometers without kinking, which formed during annealing at the desorption temperature[2], and a c(4x4) structure which formed if the surface was held close to the desorption temperature while adsorbing Bi. An interesting point which has emerged from the calculations (which is not surprising in view of its size) is that the Bi atoms are always either adsorbed on the surface or substituted for Si atoms in the surface: they are never in the second layer.

5.1 Stable structures for bismuth

The first task in investigating the behaviour of bismuth on Si(001), and particularly in investigating the complex structures seen in STM, is to identify the lowest energy sites on the surface for Bi adatoms and dimers. This is a hard problem, as the usual procedure (to use a chemical potential of some sort for the Bi atoms) is not possible here, for there is no reservoir (for example, the bulk energy for silicon atoms; there is no SiBi alloy to use for this purpose). Accordingly, I have chosen to compare the excess surface energy (compared to a clean, perfect silicon surface) **plus** the Bi chemical potential per Bi atom. This is achieved by subtracting off the Si chemical potential multiplied by the number of silicon atoms from the energy, followed by the excess surface energy for a perfect silicon surface of the same area, and dividing by the number of bismuth atoms. The excess surface energy plus bismuth adsorption energy per Bi atom for a variety of structures (illustrated in Figure 5.1) is given in Table 5.1. All calculations were performed using TB and the parameterisation developed in Chapter 3, with two standard unit cells side-by-side, to create a unit cell six dimers long and two dimer rows wide.

The notation in the table may be a little confusing, so each category will be described. For the adatoms, the most stable position is the intra-row position - that is, bonded to two dimers in the same row ((a) in the ad-atoms section). There is another position which is almost as stable,(b), where the Bi atom is in line with the dimer, bonded to one end only, and sticks out over the trench. The two other possibilities, where it bonds to two dimers either side of the trench (inter-row (c)), or where it inserts into a dimer bond (in-dimer (d)), are much less stable. It is clear that the in-dimer has broken the silicon dimer sigma bond, rendering it less stable; the cause of the lack of stablility for the inter-row is not clear. Given that the bismuth ad-atom is never more than two coordinate, it is not surprising that tight binding produces results at variance with STM images; it is not expected to produce accurate predictions for adatoms.

With the dimers, an additional factor enters the question of stability: the Bi-Bi bond length. The most stable position is as a substitutional dimer ((a) in the substitutional section), followed by the epitaxial trench dimer ((a) in the ad-dimer section). It is perhaps not surprising that the substitutional dimer is more stable than any of the others, as it has the lowest excess surface energy. A comparison between the stabilities of the ad-dimers formed by bismuth and those formed by silicon is also interesting. The ordering for silicon

[2]A very recent publication (Naitoh *et al.* 1997) shows evidence of similar structures, though not in such profusion or detail

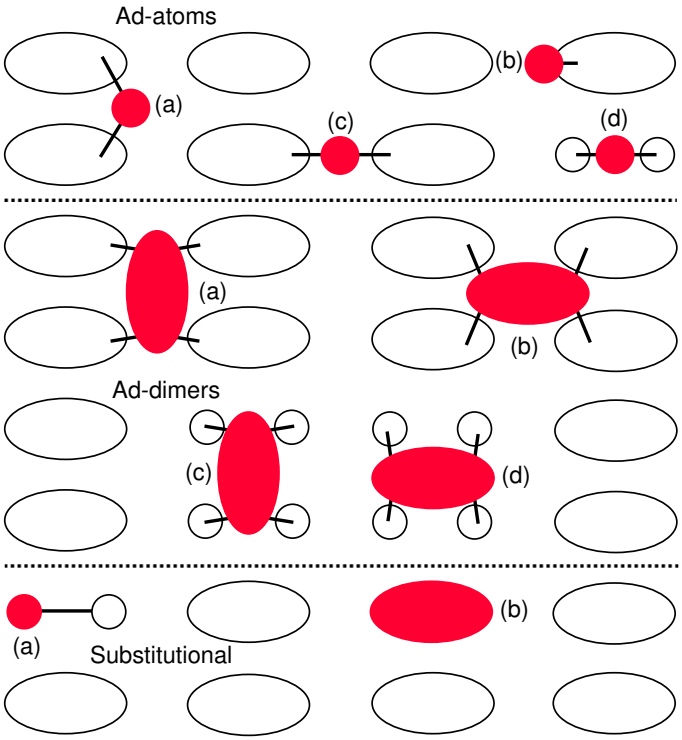

Figure 5.1: A schematic diagram, showing the possible adsorption sites for Bi ad-atoms, ad-dimers and substitutional features. Letters refer to the appropriate area in Table 5.1. Note that the dimer with one substitutional ad-atom ((a) in the bottom section) is shown as a pair of atoms, not an oval, for clarity.

Structure	No. of atoms	Energy (eV/Bi atom)
Ad-atoms		
Intra-row (a)	1	-5.87
Off dimer (b)	1	-5.86
Across trench (c)	1	-5.54
In-dimer (d)	1	-5.53
Ad-dimers		
Trench(Epi) (a)	2	-6.58
Trench(Non) (b)	2	-5.91
Row(Epi) (c)	2	-5.73
Row(Non) (d)	2	-4.32
Substitutional		
Atom(Subs) (a)	1	-6.33
Dimer(Subs) (b)	2	-7.14

Table 5.1: The excess surface energy plus bismuth adsorption energy per pair of atoms for Bi atoms in various positions on the Si(001) surface. The positions are shown in Fig. 5.1.

has the two structures on top of the rows (i.e. in the same positions as (c) and (d) in Fig. 5.1) as the lowest in energy (and almost indistinguishable) and then the non-epitaxial trench dimer, with the epitaxial trench dimer being the worst. For bismuth, which is a much larger atom, the situation is almost completely reversed: the epitaxial dimer on top of the dimer row has broken the underlying bonds, leading to a much worse energy than the trench dimer, which is the only position where there is enough room for the Si-Bi bonds to relax fully.

The normally observed configuration for Bi (in STM) is as a dimer; it is a useful confirmation of the Si-Bi parameterisation developed that the dimer is found to be the lowest energy conformation; however, the most commonly observed location for the dimer, over the Si dimer rows, is calculated to be much higher in

energy than the dimer over the trench between the Si dimer rows. In view of the accurate predictions later in the chapter, it may well be that this disparity is due to kinetics. As the adatoms are observed to adsorb on top of the rows, the pathway to formation of ad-dimers on top of the rows may well be enhanced. The low-energy positions of these dimers (substitutional and in the epitaxial position over the trench between dimer rows) will be important in the coming sections.

5.2 The straightest lines in the world ?

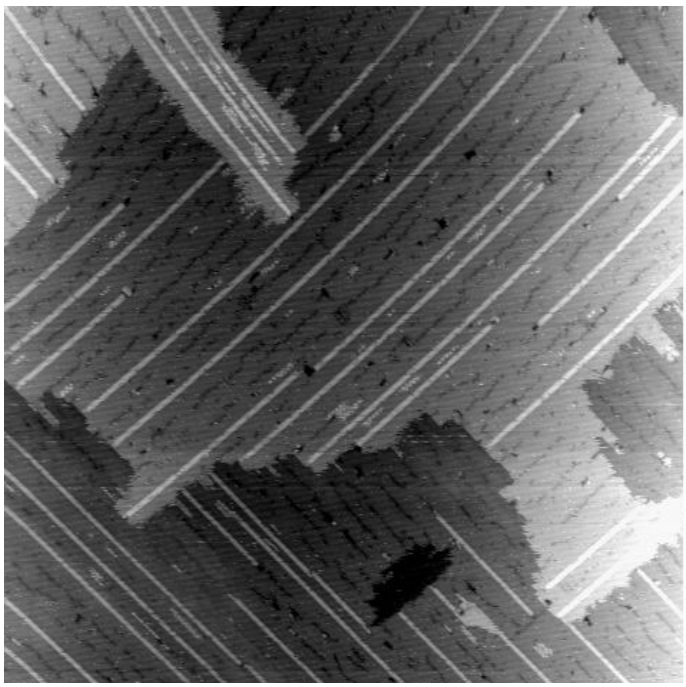

Figure 5.2: An STM image showing the lines formed on the Si(001) surface after annealing at 545° C, which appear white against the darker Si background. Several of the lines in the centre of the image are terminated in defects. A peninsula is visible in the top left of the image. Image bias of -2.0 V. The slight curvature visible on the picture is due to thermal drift. *Image courtesy of Kazushi Miki.*

The first of the interesting features which will be investigated is perhaps the most startling, and is potentially useful as a basis for quantum wires. If bismuth is adsorbed onto silicon well below the desorption temperature, the standard (2xn) structure is formed. When this sample is heated to around the desorption temperature (545°C), and annealed there, the surface undergoes a radical change. Large, flat terraces of silicon are formed (indicating the efficacy of Bi as a surfactant) which have lines running across them, perpendicular to the dimer rows. These lines are perfectly straight and run for hundreds of nanometres, probably making them the straightest lines in the world. An example is shown in Figure 5.2

These lines take up the equivalent of three dimers in the Si(001) surface, and have an internal structure which shows two features, in phase with the dimer rows, and brighter than them at high bias, though darker than them at low voltage. In line with the principles of interpretation discussed in Chapter 1, this suggests that they are at the same height as the silicon dimers, but with a larger band gap. A close up of the line is shown in Figure 5.3.

There are two questions to ask about these lines, to which atomistic modelling should be able to provide pointers, if not answers. Firstly, why are they so straight ? Secondly, how do they form ? The answers to both of these questions are intimately concerned with the structure of the lines, which therefore becomes of paramount importance.

There are various structures which would match the description given above, the most plausible of which are shown in Figure 5.4. The energies for these structures (as in the first section, an excess surface energy plus a Bi adsorption energy, though this time per dimer rather than per atom) are shown in Table 5.2, calculated using TB and a unit cell ten dimers long and one dimer wide with the standard termination. The

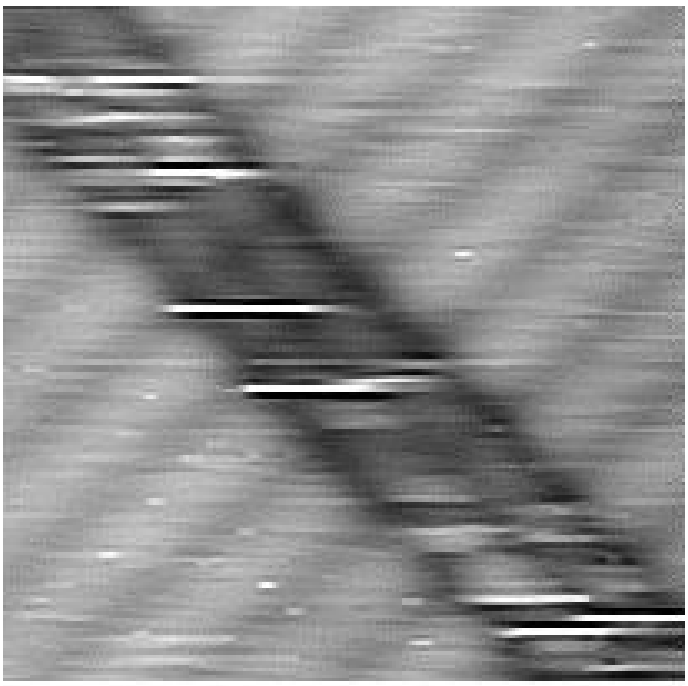

Figure 5.3: A close up of the Bi line seen in Fig. 5.2. The line (which is darker than the surrounding silicon and runs from top left to bottom right) has two features in line with the dimer rows. Image bias of -0.6V. *Image courtesy of Kazushi Miki*

link between all of these structures is the substitution of two dimers and a missing dimer defect for three silicon dimers; with Bi involved, the energetics of the 1DV are more favourable (due to the longer bond lengths) and the dimers will be closer together than with pure silicon. The two most stable structures ((a) and (c)) have bismuth in the top layer, not in the second; the out-of-phase structure, (c), is unlikely to be a candidate, as it would appear out-of-phase in STM images, and this leaves the first structure, (a). LDOS calculations for the Bi dimer, compared to a clean Si dimer, in structure (a), are shown in Figure 5.5 and indicate a wider band gap in keeping with STM and UPS observations. The energetic stability and close correspondence with the STM images suggest that this structure is indeed what is seen as the lines. These lines should probably be referred to as quantum anti-wires, due to their band gap. It is possible, however, that they could be used as the substrate for quantum wires. More experiments will be needed to explore this exciting possibility.

Structure	Energy (eV/dimer)
(a) Bi1DV	-14.544
(b) Bi1DV(lower)	-8.999
(c) Bi1DV(Out of Phase)	-14.429
(d) Bi1DV(AllBi)	-11.502

Table 5.2: The excess surface energy plus bismuth adsorption energy per pair of atoms for various possible structures for the long lines seen on the Si(001) surface after annealing at the Bi desorption temperature. The different structures are in the same order as in Fig. 5.4.

Having identified the structure of the line, the questions of kinking and growth can be addressed. The most obvious cause for the straightness of the lines, particularly at the high temperatures used, would be thermodynamic - the kinked structure may have an energy which is so much higher than the unkinked structure that the population of the kinked state is nearly zero. The simplest test of this is to take a unit cell, and calculate the increase in energy after forcing it to kink. This kind of approach was extremely successful in identifying the GaAs(001)-β2(2x4) reconstruction (Avery *et al.* 1996), where similarly straight dimer rows are seen. However, when it is done for the line structure identified (using TB and a unit cell ten

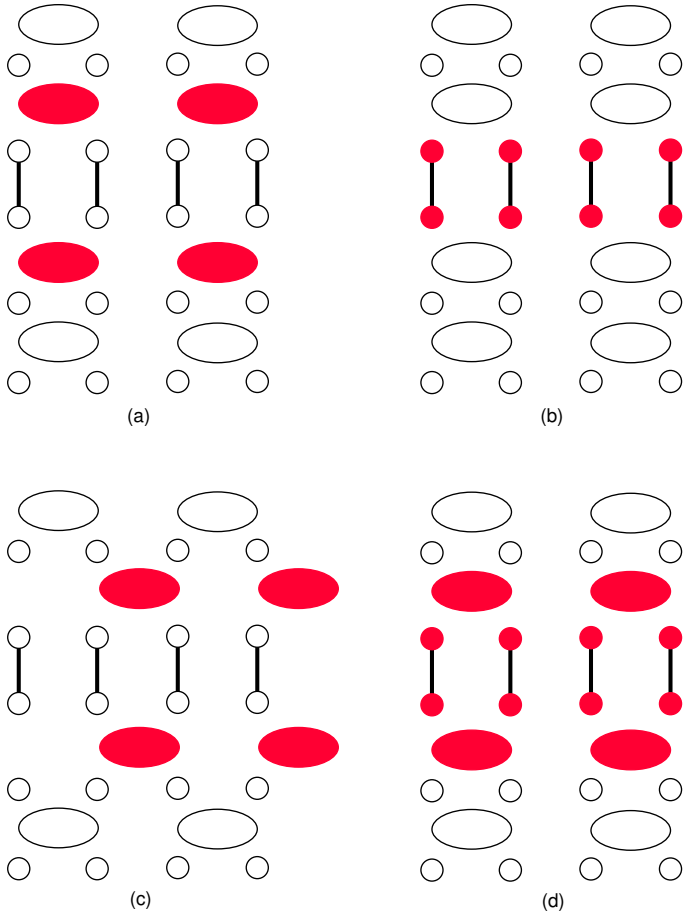

Figure 5.4: Various proposed structures for the long line structure formed by Bi at high temperature on Si(001). Red atoms are Bi. (a) A single missing dimer defect with Bi dimers substitutionally either side. (b) As for (a), but with the Bi atoms in the second layer. (c) As for (a) but with the Bi dimers out of phase with the Si dimer rows. (d) As for (a) but with the second layer atoms as well as the first layer atoms Bi.

dimers long and six dimer rows wide with the usual termination), the increase in energy due to the kink is far too small to account for the straightness of the lines (less than 0.1 eV per kink). The reason for this seems to be the almost complete isolation of the dimer rows - there is only a third layer atom connecting the two, and this is hardly strained by the kink[3]. The possibility that there is a long, conjugated bond along the line (which would give an interaction between rows) is ruled out on two counts: firstly, the STM images, which show a clear gap between dimer rows; secondly, modelling of this structure (where the Bi atoms in the dimers would be much more evenly spaced) which proves to be unstable. Clearly, some other mechanism is at work. The first clue can be found it Table 5.2, where it can be seen that the structure with the Bi atoms in the second layer is disastrously poor in energy - at least 5 eV per dimer worse than the structures where Bi is in the upper layer only. This suggests (not surprisingly, given its size) that the energetic cost of placing a Bi atom in the second layer is prohibitively high, and provides a clue to the straightness of the lines. The normal mode for diffusion of defects (Watanabe and Ichikawa 1997; Yamasaki, Uda and Terakura 1996), where one dimer is displaced into the second layer, while two second layer atoms rise to form a new dimer (a process which has a barrier of 1.6 eV (Kitamura, Lagally and Webb 1993; Yamasaki, Uda and Terakura 1996)), is barred, as the Bi dimers cannot go into the second layer. So, if the lines were to *form* straight, they would *remain* straight. Therefore the formation mechanism, which must account for the straightness of the lines, must be elucidated. The experimental evidence is that there are different areas where these lines form: in the terraces; as islands on the terraces or jutting out from step edges (peninsulas); and in the terraces, but a layer lower than the surrounding silicon (inlets). The first two of these are shown in Fig. 5.2,

[3]The GaAs(001) case mentioned above had a larger trench, with a dimer forming out of phase in the third layer, which gave a connection between unit cells.

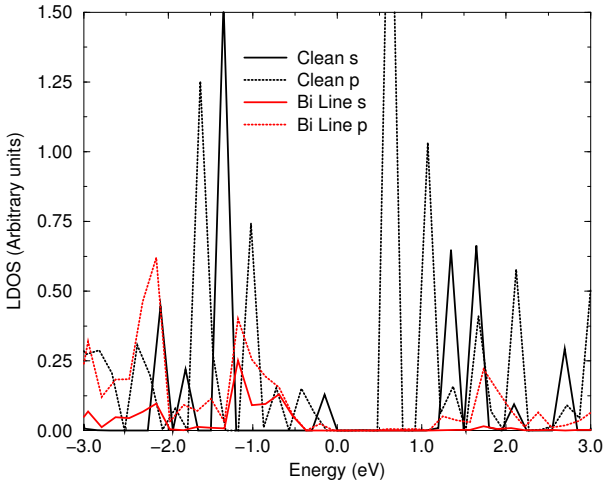

Figure 5.5: LDOS for the lowest energy structure proposed for the Bi line. The band gap is clearly wider than for clean Si dimers.

while the third is sufficiently rare that it will be discounted, except as further evidence that Bi cannot enter the second layer of the growing crystal.

Two possible modes of formation have been modelled, which are shown by two kinds of line (equivalent in structure, though not in location): the first kind grows *in* the terrace, while the second kind either grows out from a step (continuing a line in the terrace) or forms *on* the terrace, surrounding itself with a small, ragged island of silicon. While both of these growth modes are illustrated in Figure 5.6, they will be dealt with separately.

Firstly, the growth in the terrace will be examined. All calculations were performed using TB and a unit cell ten dimers long, and between six and eight dimer rows wide (depending on the termination of the line described below). In Figure 5.2, it is reasonably clear that many lines end in a (roughly) square defect. This is (by extension of the notation in Chapter 4) a pair of 3DVs, and it is likely that these are involved in the growth. What must be calculated to elucidate the mechanism is the energetics of the termination of a short segment of line. The unit cell used for the different calculations had 3 dimer rows containing Bi (as shown in the upper part of Fig. 5.6), possibly some termination, and 3 dimer rows of Si to isolate the termination. These show that, while a single 3DV is less stable than no termination (by 0.65 eV), two 3DVs are more stable than no termination (by 0.2 eV); this explains the STM images, and gives a strong indication of the growth mode. A line is most stable if it is terminated in a pair of 3DVs rather than in perfect Si dimer rows or a single 3DV, and at the temperatures used in the experiments (about 800 K) the Si will be sufficiently mobile for these to form. It is energetically more favourable for Bi dimers to be in the surface than on it (as described in the previous section), and so any Bi dimers on the surface will move to fill in the 3DV, creating a new segment of line, terminated in a single 3DV. More silicon will move out of the surface (relocating to step edges or islands on the surface), creating a new stable pair of 3DVs. The mechanism repeats while Bi is available on the surface, and is shown schematically in the upper half of Fig. 5.6. While growth of these lines was observed in STM movies, it proceeded too fast to allow confirmation of this growth mode. The only drawback to this mechanism is that a pair of 3DVs offset from the Bi line by one dimer is calculated to be as stable as a pair of 3DVs in line. However, this configuration is never seen in experiment. It may be that the relative stability of the offset cell was due to the unit cell being too small; the size used was nearly at the limit of available computing power and memory. The experimental evidence is a good start, and indicates that this mechanism is certainly plausible.

The growth of peninsulas from step edges, and line segments on top of terraces (as seen in Fig. 5.2), also demands an explanation. The proposed process is shown schematically in the lower half of Fig. 5.6. (The calculations used TB, and a unit cell made up of two standard unit cells side-by-side, making a unit cell two dimer rows wide and six dimers long.) It has already been established that the lowest energy position for a Bi dimer is over the trench between dimer rows in the epitaxial position (Fig. 5.6(a)). This line can extend

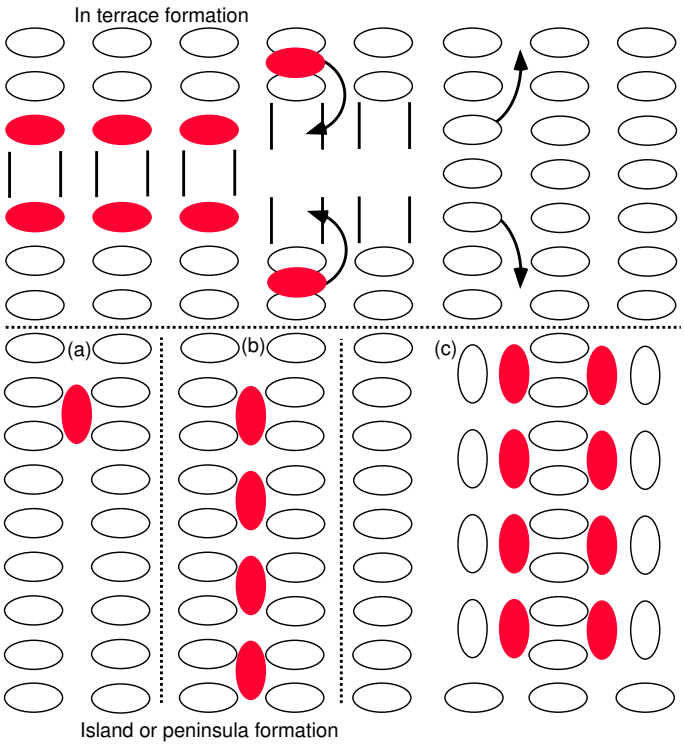

Figure 5.6: Schematic diagram showing the two modes of formation of the Bi lines. Top: the growth mode in the terrace. As two Bi dimers fill in the first 3DV, a new 3DV forms at the end. The line grows towards the right of the diagram. Second layer rebonding is indicated with double lines. Bottom: Growth of peninsulas or islands, shown for an island. (a) A Bi dimer rests over the trench between Si dimer rows. (b) This extends to become an unkinked line of Bi dimers. (c) A second line of Bi dimers forms, and is the whole is surrounded by silicon ad-dimers.

along the trench (and is 0.2 eV more stable than an isolated Bi dimer) and does not kink (the energy for two dimers, one over the trench and one on the dimer row, is 0.55 eV worse than two dimers over the trench), as shown in Fig. 5.6(b). Once a line is established, it is 0.2 eV better for it to have two silicon dimers next to it, building a small island; it is, however, then 0.2 eV better to add a *second* Bi dimer, epitaxially oriented, over the neighbouring trench than to continue the silicon building. This then has silicon added next to it, building a small island which contains a line (Fig. 5.6(c)). The only difference which would be seen at a step edge would be that Bi dimers arriving initially would be lining up with the dimers in the line in the upper terrace. The exact details of formation after (b) are unclear, and might vary from island to island. Unfortunately, only the presence, and not the growth, of these features has been observed in STM. It is also conceivable that they could form as a result of etching during annealing.

5.3 A mysterious reconstruction

It has recently been found that a c(4x4) reconstruction can be induced on the Si(001) surface; this has been observed after the adsorption of boron (Wang, Hamers and Kaxiras 1995), trimethyl gallium (Nörenberg, Bowler and Briggs 1997), bismuth (Miki *et al.* 1997), hydrogen (Uhrberg *et al.* 1992) and silicon (Zhang, Kulakov and Bullemer 1996). At first glance, these all have certain similar features in STM: in filled states, bright blobs every four dimers, with an offset between dimer rows; in empty states, the bright blobs are every other dimer, and there is no offset between dimer rows, and at higher bias, there is more detail visible - between alternate bright blobs (i.e. every four dimers), there are split features (similar to the empty states Si dimer in appearance). These features are illustrated schematically in Figure 5.7. The apparent size and contrast of these features vary between different experiments, though there is some consistency: the features labelled A in Fig. 5.7 are smaller in filled states and larger in empty states than those labelled B for the group III and V cases, whereas in the clean Si case A and B are more uniform.

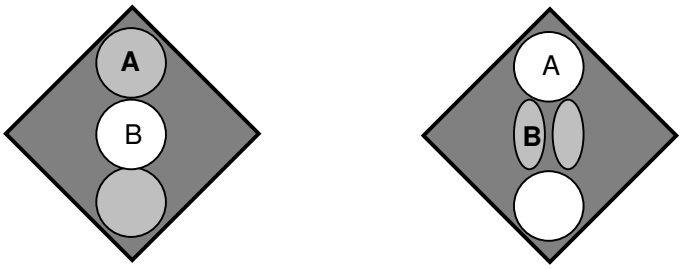

Figure 5.7: The STM appearance in (a) filled states and (b) empty states for the c(4x4) reconstruction seen on Si(001) after adsorption of several different compounds. The image is schematic. In STM, the apparent size and contrast of features A and B vary from adsorbate to adsorbate. The diamond surrounding the features delineates the unit cell. See text for details, and compare with insets in Fig. 5.8.

There is something peculiar occurring in these experiments, for all the different preparation modes give relatively similar STM images. There has been a recent suggestion (Kazushi Miki, Private Communication) that all of these reconstructions are the same, being induced by small amounts of an adsorbate which strains the surface and stabilises a clean Si c(4x4) reconstruction, such as that suggested by Uhrberg *et al.* (1992). This theory is based on SIMS experiments which show the presence of about 1% of carbon in the top few monolayers. With the clean Si c(4x4) results, this explanation seems not unlikely; however, with the group III and V compounds there is good evidence that the adsorbate is involved in large amounts in the reconstruction. Wang, Hamers and Kaxiras (1995) used Auger measurements to confirm that they had up to half a monolayer of boron on the surface in their images, which indicates strongly that boron is involved in the reconstruction; they also found an increase in the amount of the reconstruction with increased dose. Nörenberg, Bowler and Briggs (1997) also found an increase in the amount of c(4x4) with Ga exposure, but were unable to measure the Ga coverage in Auger. The shape of the patches which form in these and in the Bi results (Miki *et al.* 1997) are highly ragged, and often form at step edges, which suggest a reconstruction based on embedding rather than small amounts of strain. With these results in mind, an identification based on the presence, rather than the absence, of Bi will be made. However, there are some intriguing unexplained results still in the literature, mainly concerned with clean Si, which cannot be solved here. It may well be that the explanation of Miki is correct for these results.

The c(4x4) structure forms when bismuth is deposited while the clean Si(001) surface is being annealed at about 570°C. An example is shown in Figure 5.8, which shows the reconstruction at different bias voltages, with magnifications and schematics clarifying the structures. The essential features to be noted are: in filled states, the most prominent feature, which occurs every fourth dimer, forms squares at 45° to the dimer rows (this feature is shown with a white dot between two grey dots in the inset in Fig. 5.8(a)) and is surrounded by two grey dots; in empty states, at low bias voltage (i.e. near the Fermi level), a square pattern of prominent features is visible which is smaller and not rotated with respect to the dimer rows; at higher bias voltage, a faint double feature becomes visible between every fourth brighter blob, which corresponds to the prominent features in filled states (this is illustrated schematically in the inset of Fig. 5.8(d)).

There are two important observations which help to identify the structure of this phase: firstly, that a silicon dimer appears as a dot in filled states and as a double feature in empty states; secondly that HREELS of Bi dimers (Gavioli, Betti and Mariani 1997) shows that on the Si(001) surface, they have a large density of states above the Fermi level - significantly larger than the clean Si(001) surface. These two facts suggest a structure, schematically illustrated in Figure 5.9(a): the prominent feature in filled states, which splits into two blobs in empty states, is a silicon dimer, while the prominent feature in empty states, which is darker in filled states, is a bismuth dimer. Another possible unit cell, where the Si dimer is missing (which is seen occasionally in large patches of the reconstruction) is shown in Fig. 5.9(b).

The so-called alpha structure, which forms the majority of the reconstruction, shows all the right features to match with the STM images, and has a very reasonable energy - 6.5 eV per Bi dimer (found using TB and a unit cell four dimers long and two dimer rows wide with the usual termination). The beta structure is slightly less stable, which might account for the lower proportion seen. As seen in the case of the lines, the possibility of the Bi being in the second layer is ruled out by the energetic cost - 5 eV per Bi dimer. There is no experimental evidence for the growth of this mode, and there are few theoretical clues. This will have

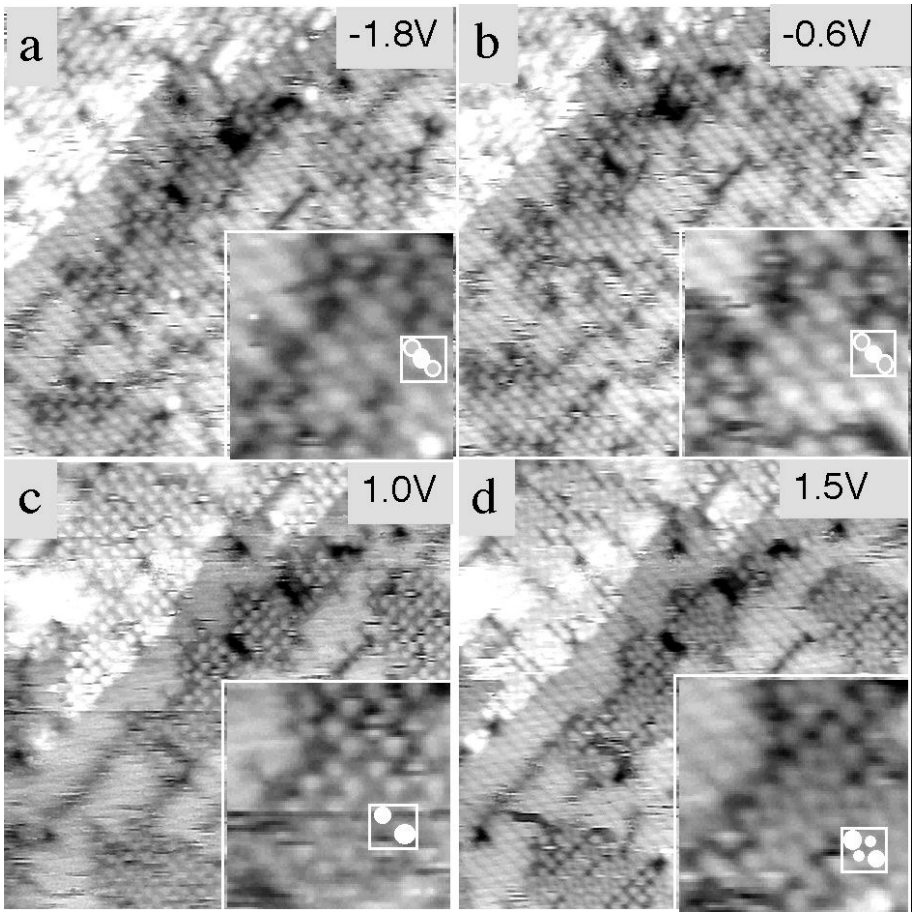

Figure 5.8: The c(4x4) structure formed when Bi is deposited on Si(001) near the Bi desorption temperature. Different bias voltages are shown: (a) -1.8V, (b) -0.6V, (c) 1.0V and (d) 1.5V. Images are 29nm square, and the insets are 6.3 nm square. *Image courtesy of Kazushi Miki.*

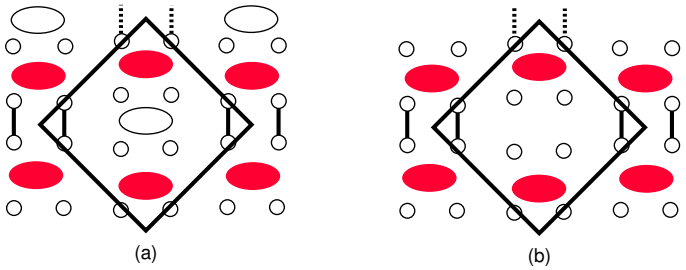

Figure 5.9: Two proposed structures for the Bi c(4x4) structure. (a) The alpha structure, which has the double feature in empty states. (b) The beta structure, which has no double feature in empty states. Red atoms are Bi. The dashed lines at the top of the central dimer row represent rebonding in the second layer which goes out of the visible area.

to be left to future work.

Conclusions This chapter has covered some of the most exciting areas in silicon surface studies - adsorption, surfactants, controversial reconstructions and the possibility of quantum wires. The parameterisation developed in Chapter 3 has generated results which are in reasonable agreement with experimental observations. Models for the Bi line structure and growth and the Bi c(4x4) reconstruction have been proposed which fit all available experimental data.

Chapter 6

Hydrogen on Si(001)

... A molecular system ... (passes) ... from one state of equilibrium to another ... by means of all possible intermediate paths, but the path most economical of energy will be the more often traveled.

Henry Eyring, 1945

Hydrogen is an interesting system to study for two main reasons: firstly, it is a prototypical system for diffusion, which is easy to image in STM, and is therefore good to model, as there is comparative data available; secondly, it plays an important role in the growth of silicon from disilane, as described in Chapter 7. At low temperatures, the silicon growth rate is limited by the hydrogen desorption rate (Mokler *et al.* 1993 and Wu and Nix 1994) since the adsorbed hydrogen can block incoming disilane molecules or fragments, although growth can proceed slowly on a hydrogen saturated surface (Boland 1991a). It can also influence the rate of diffusion of Si atoms (Eaglesham *et al.* 1993, Vasek *et al.* 1995), which then affects the morphology and the rate of growth.

6.1 Hydrogen diffusing on Si(001)

Hydrogen adsorbs on the surface of silicon by breaking the π-bond of the silicon dimer (Wu and Carter 1991) and bonding to one of the dangling bonds thus created. The remaining half-filled dangling bond appears as a bright blob in the STM, while the hydrogen at the other end is relatively dark (Boland 1991b), due to the bonding energy gain. Hydrogen adsorbs randomly at coverages of less than 0.1ML (Boland 1992,1993), but after that it pairs up onto a single dimer (Widdra *et al.* 1995). This arrangement, with one hydrogen on each end of a silicon dimer, is the most stable phase for adsorption, as the silicon atoms are both four coordinate, with no dangling bonds. Widdra *et al.* found that this pairing occurs at adsorption temperatures between 150K and 600K, and that the amount of pairing was independent of temperature; they proposed a mobile precursor mechanism (similar to that proposed by Sinniah *et al.* (1989) for desorption) to explain this. In this mechanism, the incoming hydrogen atom is in an excited, band state with a finite lifetime, which will sample several different adsorption sites. If the lifetime allows it to sample, say, 10 sites, then for coverages below 0.1 ML, the hydrogen distribution will appear random, and above that it will be almost entirely paired. Confirming that the diffusion barrier is sufficiently high to prevent diffusion on adsorption over a good part of the temperature range used (150-600 K) would lend weight to this argument. These two experiments suggest that the diffusion barrier is sufficiently high to prevent diffusion upon adsorption; this is the only experimental evidence available on diffusion apart from the work connected with this thesis (Owen *et al.* 1996).

There have been many theoretical investigations of hydrogen diffusion on Si(001), though none of them have been conclusive. Wu and Carter have performed many cluster calculations using CI techniques and initially (Wu and Carter 1992) predicted diffusion barriers of 2.0 eV along the dimer rows and 2.7 eV across the dimer rows. However, they used a small cluster consisting of 9 silicon atoms terminated with 12 "pseudo-hydrogen atoms", and did not allow the substrate to relax during diffusion, which suggests that their barrier will give a poor answer to the actual barrier. The CI method finds a very accurate answer, but the degree of relevance of the answer found to the physical system is rather small, due to the use of clusters. When they later performed different calculations (Wu, Ionova and Carter 1994) using empirical potentials (based on the Stillinger-Weber potential) which they had generated from SCF-LCAO calculations, and allowed the

slab substrate to relax, they found a significantly lowered barrier of 1.5 eV for diffusion along the dimer row and a slightly lowered value of 2.5 eV for diffusion across dimer rows.

This lower barrier fits with calculations by Vittadini, Selloni and Casarin (1993) who found a barrier of 1.3 eV, though I would argue that using empirical potentials to model a process where the bond breaking is vital, as Wu, Ionova and Carter did, will only get results by luck. Vittadini, Selloni and Casarin used LDA, and a repeated slab for their calculations, and allowed the top four layers of the slab to relax. They claim that the major difference between their calculations and those of Wu and Carter (1992) is the lack of surface relaxation in Wu and Carter. They attempt to estimate the magnitude of this effect by performing their diffusion calculations *without* allowing the substrate to relax, and conclude that the effect can be as large as 1 eV.

All of these sources of evidence indicate that the diffusion of hydrogen on Si(001) has an activation barrier of at least 1.3 eV, and that hydrogen is unlikely to hop across the dimer rows (as hopping across dimer rows is likely to occur in a similar temperature regime as desorption).

It is worth discussing briefly here the effect of the zero-point energy of hydrogen on diffusion barriers. If we assume that the dominant mode which is contributing to this is the Si-H bond stretch, which has a value of about 2100 cm^{-1}, then the zero-point energy (using the standard form of $\frac{1}{2}\hbar\omega$) is about 0.04 eV for each of the three directions. This might double at the mid-point (but is unlikely to increase any more), leading to a change of up to 0.04 eV. Such a value fits well with the expectation found from EELS data that the change in zero-point energy for hydrogen even for heavy structural modifications varies by less than 3 % (Stucki *et al.* 1983). Thus it will be a relatively minor change.

6.2 STM results and phenomenological modelling

While it seems a little unusual to be describing the experimental work in detail in a thesis which is primarily theoretical, the modelling does interact with the experiment, and the results are so elegant that they deserve repetition. Images obtained at up to 500 K show small white dots (which are the result of hydrogen adsorption, and are the dangling bond at the unhydrogenated end of the dimer) which do not move over periods of up to 15 minutes (a pair of images 15 minutes apart are shown in Figure 6.1). The means that the *maximum* hopping rate of the hydrogen is once every fifteen minutes.

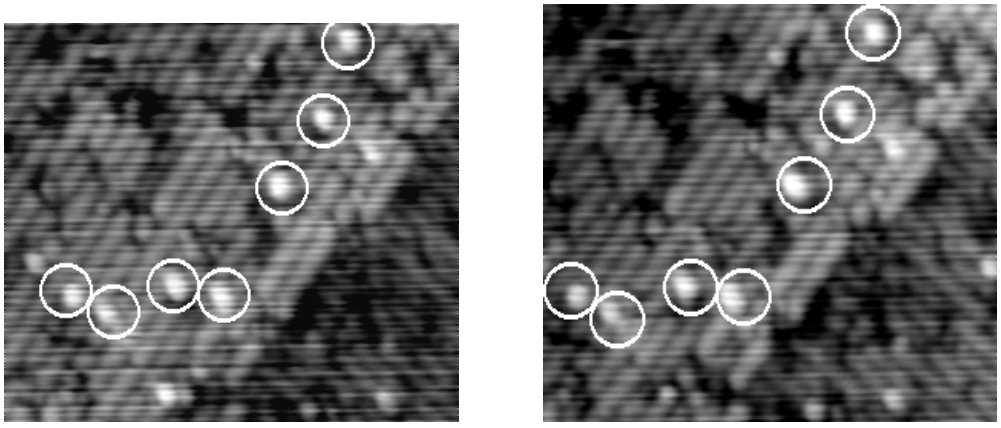

Figure 6.1: STM images taken at 500 K, with imaging conditions of -1.3V, 0.08 nA. The images are 14 nm wide, and separated by 15 minutes. The grey lines are the Si dimer rows, while the bright white dots are the result of hydrogen adsorption. Note that there is no change between the two pictures. *Images courtesy of James Owen.*

Once the sample is warmed to about 600 K, the hydrogen starts to move between pictures. Figure 6.2 shows two pictures taken at 600 K, separated by 10 seconds. The dots are moving, and appear to be slightly fuzzy, when compared to the previous, lower temperature pictures. This change in appearance will be discussed below. This temperature gives two limits on the hopping rate: the minimum hopping rate is once every ten seconds, and the maximum is the scan rate of the tip, which is once every 1/50th of a second.

If the sample is warmed further, the hydrogen starts to move on the timescale of the tip scan. An example of the behaviour at 640 K is shown in Figure 6.3. The STM tip scans down the image in a series of 512 line

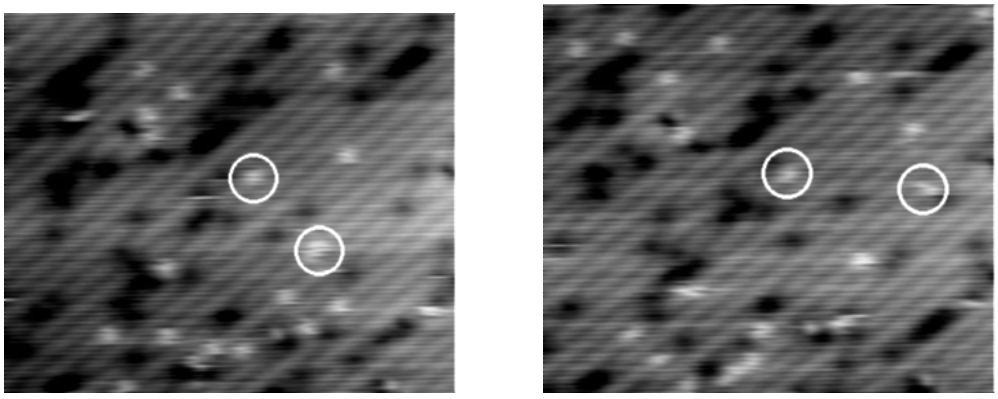

Figure 6.2: STM images taken at 600 K, with imaging conditions of -1.0V, 0.08 nA. The images are 20 nm wide, and separated by 10 seconds. While the two dots on both pictures cannot be categorically identified as the same, they probably are, in which case both have moved considerably. Note also the slightly fuzzy appearance, as discussed in the text. *Images courtesy of James Owen.*

scans. On reasonably high resolution, each dimer will be broken down into between 3 and 5 lines - i.e., the STM tip samples each dimer 3-5 times. Once the hydrogen is hopping once every 1/50th of a second, it can move off any given dimer onto another in the time that the tip makes a single line scan. This means that on any given dimer, out of the 4 linescans in which it is sampled by the tip, it may be occupied, unoccupied, occupied and finally unoccupied, or any other combination; more to the point, it will appear on several dimers in this fashion, elongating its normal appearance. This will lead to the smudged appearance which has been circled in Fig. 6.3, and gives a *minimum* hopping rate, as we know that it is hopping faster than the tip can scan.

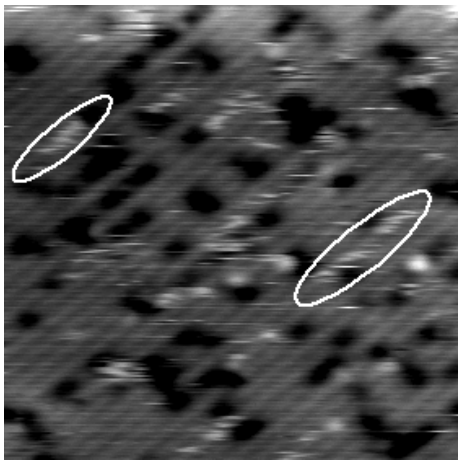

Figure 6.3: STM image taken at 640 K, with imaging conditions of -1.0V, 0.08 nA. The image is 30 nm wide. The fuzzy dots in Fig. 6.2 have now elongated into long white smudges; two of these have been outlined. Each STM scan line takes $\frac{1}{25}$th of a second. *Image courtesy of James Owen.*

The imaging of a moving hydrogen atom by a scanning tip is an interesting problem; a phenomenological model was created to study this. As explained above, each dimer is scanned about four times in the course of the STM scanning down the selected area. In the model, the hydrogen started at a random site, and was given a certain probability of hopping each time the tip scanned (e.g. every 1/50th of a second). The length of available dimer row (e.g. 10 dimers, which represents the relatively high defect density on the surface, and an assumption that the hydrogen cannot hop over a defect which is justified later in this chapter) was fixed, and a given number of smudges for those initial conditions generated. A set of 10 smudges is given below for 10 dimers, with 4 line scans per dimer and a hopping probability of 0.2 per line scan; a 1 represents conjunction of the STM tip and the H, while a 0 represents no such conjunction. Each result is broken up

into units of four, each representing a dimer.

```
Smudge  1: 1000 0001 0000 1001 1001 1011 1110 0010 0000 1111
Smudge  2: 0000 0000 0000 1111 0000 0010 0000 0000 0000 0000
Smudge  3: 0000 0001 0010 0000 0000 0000 0000 0011 1000 0000
Smudge  4: 0011 0000 0000 0000 0000 0000 0000 0000 0000 0000
Smudge  5: 0000 1100 0000 0000 0000 0000 0010 0000 0000 0000
Smudge  6: 0000 0000 0000 0000 0000 0000 0000 0001 0000 1100
Smudge  7: 0000 0000 0000 1001 0000 0000 0000 0000 0000 0000
Smudge  8: 1111 1000 0000 0000 0000 0110 0000 0000 0000 0000
Smudge  9: 1111 1101 0000 0000 1111 0110 0000 0000 0010 0111
Smudge 10: 0000 0000 0000 0000 0000 0000 0000 0000 1100 0000
```

From this kind of model, and the statistics generated by longer runs similar to the above, it was found that the hydrogen starts to smudge with a hopping probability of about 0.2 per line scan. The onset of smudging is seen as a slight fuzziness of the hydrogens, as seen in Fig. 6.2. This value of 0.2 has been used, appropriately modified to reflect a hopping rate, as a measured hopping rate in the plot shown in Figure 6.4, and is labelled with squares. The overall set of results is shown in Figure 6.4, as a plot of hopping *rate* against 1/T. The Arrhenius plot assumes that the hopping probability, h, depends on an attempt frequency, A, and the barrier E_b, in the following manner: $h = A\mathrm{e}^{-E_b/\mathrm{kT}}$. The attempt frequency can critically affect the barrier height calculated from the plot; a value of 10^{13} s^{-1} is a reasonable figure from the literature for the Si-H bond vibration, and yields a barrier height of 1.68 eV. This is shown as a solid line in Fig. 6.4. The maxima given by the low temperature data (labelled with arrows pointing down), and the minima given by the high temperature data (labelled with arrows pointing up), indicate that this value of 10^{13} s^{-1} is a *minimum*. A fit with an attempt frequency of 10^{14} s^{-1} is shown with a dashed line in Fig. 6.4. This gives a barrier height of 1.8 eV; it is important to stress that both of these barriers are within experimental errors. The whole procedure is discussed in more detail by Owen *et al.* (1996).

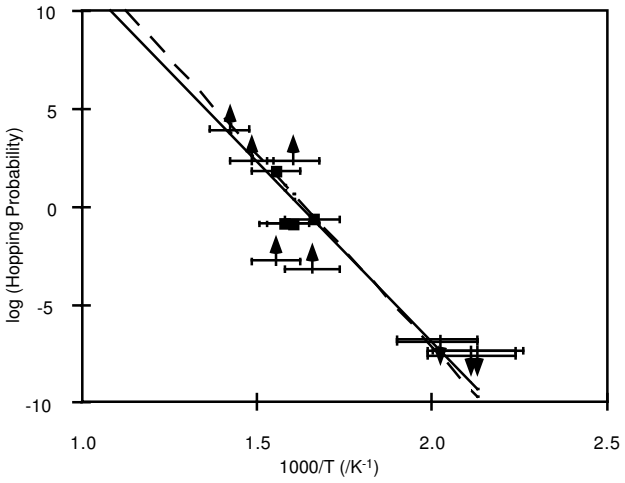

Figure 6.4: An Arrhenius plot of the hopping rate against inverse temperature for the motion of single hydrogen atoms on the Si(001) surface. There are three different types of data point: minima, measured hopping rate and maxima, which are discussed more fully in the text. The solid line is fitted to a pre-exponential rate of 10^{13} s^{-1}, and gives a hopping barrier of 1.68 eV. The dashed line is fitted to a pre-exponential rate of 10^{14} s^{-1}, which gives a hopping barrier of 1.80 eV. *Picture courtesy of James Owen.*

6.3 Atomistic modelling of single atom diffusion

This problem has been approached from two directions: that of LDA and GGA, and that of tight binding. The former have been used for this problem essentially to validate the tight binding calculations, and for comparison with experiment; once the tight binding parameters have been validated, larger systems than can be tackled with LDA may be addressed, such as steps and defects. These will be described in Sections 6.4 and 6.5.

6.3.1 DFT calculations

The hydrogen diffusion was modelled using LDA to obtain coordinates and charge densities, and both LDA and *post hoc* GGA to obtain the energies, using the standard unit cell. The hydrogen atom which was to diffuse was placed on one end of a dimer in one row, and to prevent the system from having an unfilled band, a hydrogen atom was also placed on the other side of the other dimer row. This hydrogen atom played no part in the diffusion (except to obviate the need for a subspace rotation at every step to correctly half-fill one band). The only constraint applied to the system was to force the hydrogen atom to remain in a plane of constant y (in Figure 6.5(b), the y axis runs vertically up the page), which was moved along the dimer row at each point in the calculation. The energies from both LDA and GGA calculations are shown in Figure 6.5(a).

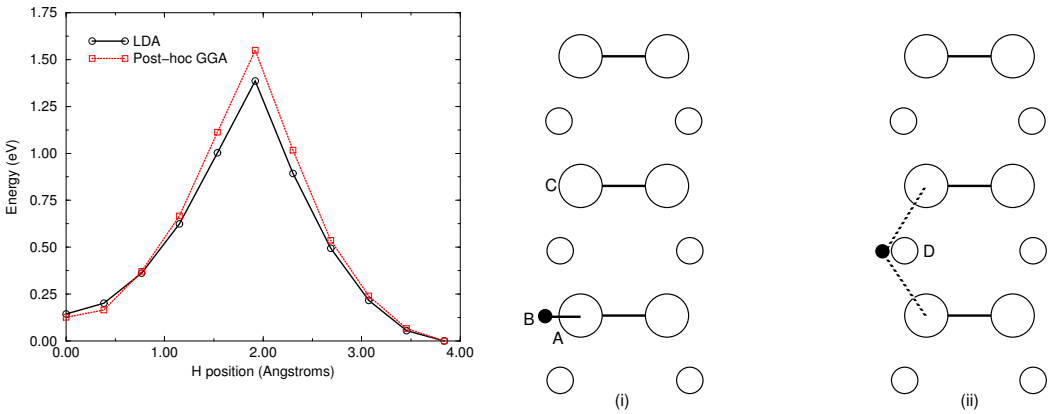

Figure 6.5: (a) The energy for a hydrogen atom diffusing on the Si(001) surface, calculated using LDA and post-hoc gradient corrections (GGA). The barriers are 1.55 eV from GGA and 1.39 eV from LDA. (b) The process indicated schematically. The letters A,B and C refer to Fig. 6.7, and letter D is for reference in the discussion of the tight binding results.

The mid-point ((ii) in Fig. 6.5(b)) is highly unstable, as shown by the sharp peak in the graph. This is not unreasonable, as the hydrogen is bonded to two silicon atoms at this point, which is not an energetically favourable state for a $1s^1$ atom to be in. The distances between the hydrogen and the two silicon atoms are shown in the next section, in Fig. 6.7; these show that once the hydrogen is past the mid-point it bonds strongly to the atom towards which it has been diffusing, and less strongly to the atom on which it started. The energies obtained (1.55 eV from GGA and 1.39 eV from LDA) are a little smaller than the experimental number (1.68 eV), but bring home the point that gradient corrections can be important in obtaining the correct value of barrier heights, for the GGA result agrees to within experimental error. The barrier is increased to 1.6 eV with the addition of spin to the calculation, which brings the calculated value well within experimental errors. Another point worth raising here is that of attempt frequencies: the experimental number of 1.68 eV was found from *hopping rates* and assuming an attempt frequency of 10^{13} s^{-1}, while another frequency which fits perfectly well (10^{14} s^{-1}) gives a barrier of 1.80 eV. It is certainly accepted that LDA cannot predict attempt frequencies accurately (Smith and Jónsson (1996) state that a factor of 5 is too sensitive for LDA to correctly model; see Blöchl, van de Walle and Pantelides (1990)), and so it has been decided that the prefactor relevant for comparing LDA/GGA hopping rates should be taken as 10^{13} s^{-1}. This fits well with values from the literature found using, for example, experimental vibrational mode studies. This is the first *ab initio* study of this system to predict a barrier which agrees with the

experimental value.

6.3.2 TB calculations

To calculate the energy of a hydrogen atom diffusing on the Si(001) surface using tight binding is a difficult task. Particularly, the bonding at the saddle point will be highly non-equilibrium and therefore test the fitting of the parameterisation enormously. The fitting employed in the parameterisation developed during this work has been described in Chapter 3; here I will describe the application to a problem of diffusion.

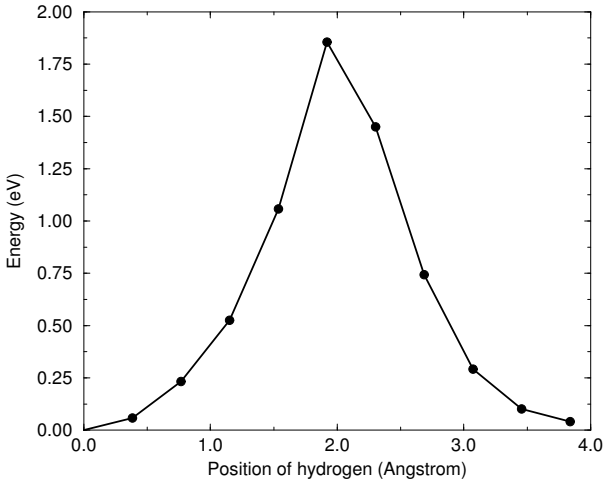

Figure 6.6: The TB energy for a hydrogen atom diffusing on the Si(001) surface. The barrier is 1.96 eV (see text for a discussion of a correction to be applied to this). The process is illustrated schematically in Fig. 6.5(b).

The standard unit cell was used for this calculation. The only constraint applied to the system was the same as for the LDA modelling, i.e. constraining the hydrogen to lie in planes of constant y. The hydrogen was only allowed to bond to two silicon atoms - the one it starts on and the one it ends on. The other atom to which it could bond is the second layer atom to which the two dimers are bonded (marked as 'D' in Fig. 6.5(b)). Tests showed that if this atom was allowed to bond to the hydrogen, the bond energy was almost zero, but the repulsive energy was high; this was clearly unphysical and solved by removing the interaction. Examining surfaces of constant charge density from the LDA calculations showed that there was almost no charge between these two atoms, indicating that any interaction is negligible and that the above approximation is good. The start and end points are not exactly symmetrical, as the cell is p(2x2), but the energy difference is small.

The energy for the diffusion reaction is shown in Figure 6.6, and the bond lengths between the hydrogen and the two atoms to which it bonds are shown, along with the equivalent LDA lengths, in Figure 6.7. This latter figure displays the biggest problem which tight binding has: when the hydrogen is just within bonding range of the second dimer, instead of a slight attractive interaction (as seen in LDA) there is a slight repulsive one (shown by the knee in the TB curve, and caused by the same effect as that alluded to above for the second layer Si atom - a high repulsive term where there is a low electron density). This is due to the fact that TB cannot take into account screening - the physical interactions represented by the TB repulsive term are very small, but this is not reflected in the relatively large repulsive interaction. The error induced by the repulsion has been estimated as 0.2 eV; this was found by taking the unit cell at the mid-point, and comparing the energy for the cell both as it is and after re-relaxing **without** the interaction between the hydrogen and the second atom (C). This makes the overall barrier 1.76 eV (with an error of up to $\pm$ 0.15 eV), which is much more in line with experiment. This correction will be applied throughout the thesis where hydrogen diffusion processes are modelled. More importantly, however, it should be noted that this anomalous repulsive interaction will be reasonably independent of whatever diffusion process is being modelled, and that energy *differences* between processes will reflect reality extremely accurately. This will

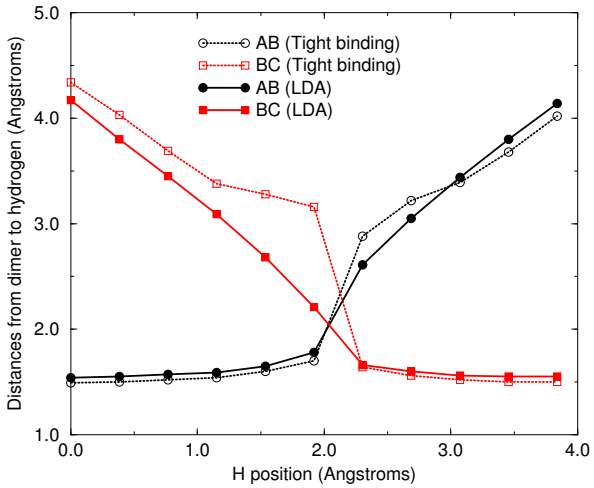

Figure 6.7: The distance from the two Si atoms which bond to the hydrogen, as calculated by tight binding and LDA, as the hydrogen moves. All distances in Ångström. The letters refer to the atoms shown schematically in Fig. 6.5(b).

be important in the remaining sections of this chapter. It is extremely encouraging that, with the correction applied, TB predicts a barrier in good agreement with experimental and LDA results.

6.4 Diffusion on the saturated surface

It has been shown above that the diffusion of single hydrogen atoms on a clean silicon (001) surface can be understood both by experiment and theory; however, the growing surface is likely to be saturated with hydrogen, and it is necessary to understand diffusion in this situation as well. As mentioned in the introduction to this chapter, it is energetically favourable for two hydrogen atoms to adsorb on the same dimer on the silicon surface, and once there is more than about 0.2 ML of hydrogen, this occurs. Once the coverage reaches 0.8 ML or above, most of the surface will be covered with paired hydrogens, and it becomes helpful to think of the clean dimers as the minority species. While these dimers do not migrate as such, the effect of two hydrogens moving in one direction will be to produce a vacancy moving in the other. It is particularly easy to image these vacancies in STM in empty states, as the π^* orbital stands out; it is a beautiful image of an anti-bonding orbital, which has a node in its middle. An example of the hydrogen saturated surface is shown in Figure 6.8.

The saturated surface does not start to show any signs of movement until 600 K, at which point changes occur on the timescale of minutes, and changes between sequential images are not visible until 650 K (as compared to the clean surface with single hydrogens where changes between sequential images were visible at 600 K). Once movement occurs, it occurs in pairs. That is, the π^* orbitals appear to move, which is actually caused by pairs of hydrogen moving. The tight binding parameterisation described in Chapter 3 and validated above in Section 6.3.2 has been used to investigate the cause of the barrier for paired diffusion being higher than for single atom diffusion, and the mechanism underlying paired diffusion.

The standard unit cell was used, with five of the dimers saturated with hydrogen, and one dimer clean. There are two obvious ways for the pairs of hydrogen to move: in a concerted motion, with both moving simultaneously; and for one to diffuse, creating two single hydrogens briefly, and for the second then to follow (or, equivalently, for the first to return). Both of these possibilities were modelled, applying different constraints. For the concerted motion, both atoms were forced to remain in the same plane of constant y, which was moved along the dimer row at each point in the calculation. In Figure 6.9(b), the y axis lies vertically up the page. For the separate motion, the first of the diffusing atoms was constrained to lie in a plane of constant y, while the second was left free; once the first atom had moved to the new dimer, the same process was applied to the second atom. As was true for single hydrogen diffusion, the second layer atoms labelled 'D' were not allowed to bond to the hydrogen. The results are shown in Figures 6.9 and 6.10

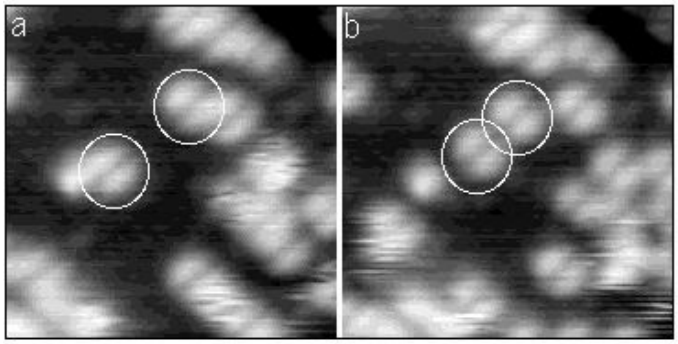

Figure 6.8: STM images showing paired hydrogen diffusion. The dark background is silicon dimers saturated with hydrogen. Clean dimers appear as a bright feature, with a node down the middle. The two clean dimers circled in (a) appear to have moved at least one dimer closer together in (b) - an effect caused by diffusion of paired hydrogen. *Images courtesy of James Owen.*

for the separate and concerted motions respectively.

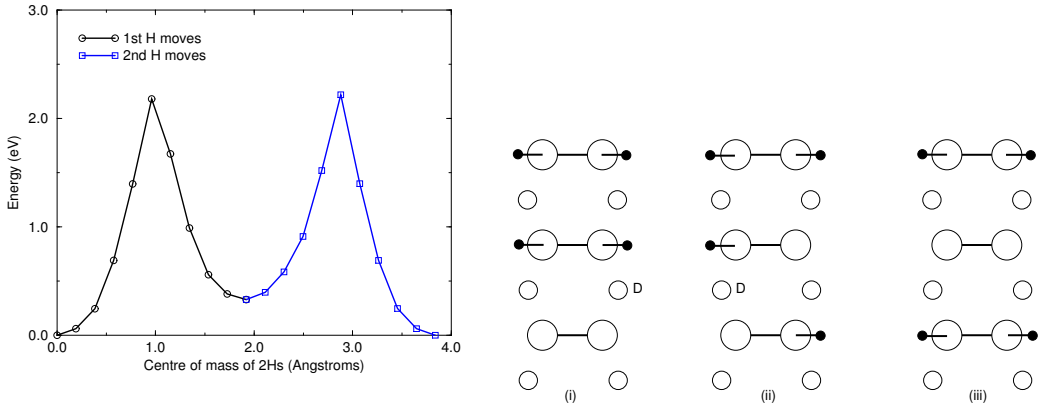

Figure 6.9: (a) The energy for a pair of hydrogen atoms diffusing on the Si(001) surface separately, calculated using TB. The barriers are 2.18 eV from the end to the middle, and 1.89 eV from the middle to the end. (b) The process indicated schematically. (i) corresponds to 0.0 Å; (ii) to the minimum at 1.92 Å; and (iii) to the end point at 3.84 Å.

As can be seen, while the energy barrier for the concerted motion (Fig. 6.10) is less than twice the energy for a single hydrogen diffusing, it is still 3.6 eV, which is far too high for any diffusion to occur, even at 650K (even applying the correction mentioned above yields a barrier of 3.2 eV). The separate case (Fig. 6.9) is more interesting. The initial barrier of 2.18 eV (1.98 eV corrected) is higher than that found for the clean surface (1.96 eV or 1.76 eV corrected), but the barrier for the second hydrogen to diffuse is significantly lower than the initial barrier (1.89 eV or 1.69 eV corrected). The corrected barrier for the first atom to diffuse (1.98 eV) matches extremely well with the experimental value of 1.95 $\pm$ 0.2 eV (Owen, Bowler, Miki and Briggs 1997) and 1.69 eV for the second hydrogen means that the second diffusion event will occur extremely fast at 600 or 650 K. This makes the diffusion mechanism clear: the rate-limiting step is the breaking up of the paired hydrogen, which is then followed extremely quickly by the repairing, with a 50% probability of reforming without having moved. This means that at 600K, the second hydrogen is about 300 times more likely to diffuse than the first, and at 700K it is about 150 times more likely. So, at an experimental temperature at which the vacancies are observed to be hopping once every 10 seconds or so (i.e. once per image on average) the mean hopping time for the first atom will be 10 seconds, while for the second atom it will be 1/20th of a second. This leads to the appearance of concerted motion in STM.

This investigation is a clear case of where the experimental/theoretical interaction is so vital: without the experimental images of the vacancies moving, the question to address would have been unclear; without

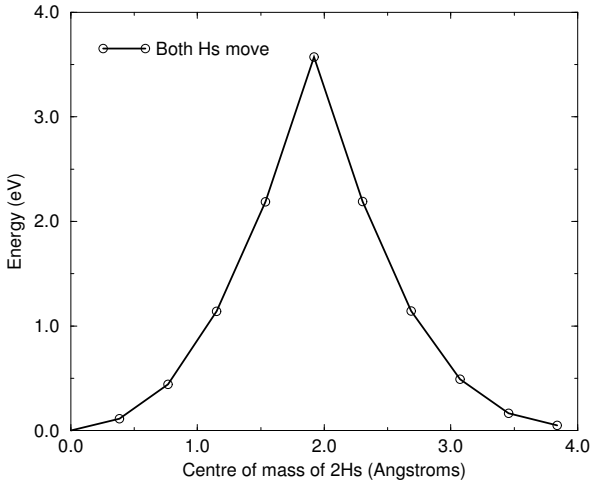

Figure 6.10: The energy for a pair of hydrogen atoms diffusing on the Si(001) surface in a concerted manner, calculated using TB. The barrier is 3.6 eV.

the theoretical modelling, however, the mechanism could not have been elucidated. This is one way in which the collaboration proceeded: experiment finds an interesting, but inexplicable, phenomenon, and theory is guided to the explanation.

6.5 Behaviour of hydrogen at steps and defects

As has already been discussed in Chapter 4, the bonding at steps and defects is different to that on the perfect surface. It is important to understand how this affects both diffusion and stability, as these will have a strong influence on growth if different.

6.5.1 Saturation of steps and relation to growth

The question of growth of silicon will be covered in more detail than it is here in Chapter 7; however, it is worth discussing the question of denuded zones and steps here. In solid-source MBE, large areas of clean terrace are seen both above and below B-type steps on the growing surface - there are no islands and very few atoms in these areas. The explanation for this is that the B-type step is a nucleation site for growth, and all atoms which are within a certain distance of the step bond there, resulting in step-flow growth and double steps, eventually. In the gas source growth experiments performed, however, (Owen *et al.* 1997a,b) none of these denuded zones were found below 700K, and islands nucleated close to the step edge. The only explanation available for this is that the rebonded edges of the step are being saturated with hydrogen, which is stopping growth.

To determine whether this is correct, two questions must be answered: is the step a good site for hydrogen to adsorb at, and how does it get onto the step edge ? The first question can be answered by finding the energies for adsorbing hydrogen at different step edges in different configurations; these are given in Table 6.1 from TB calculations, which used a unit cell fourteen dimers long and two dimer rows wide. The hydrogens at the top of the step are adsorbed on the dimer at the edge, while the hydrogens at the bottom are adsorbed on the three-coordinate second-layer atoms. See Fig. 4.8 for a diagram of the B-type step. The energies are given for pairs of hydrogen, as this is the configuration found on the growing surface.

These data reveal a variety of results: there is little interaction between dimer rows (compare the columns Saturated Bottom and Isolated Pair); the top of a step is less stable for a pair than the clean surface; the rebonded step is a significantly better site for a pair of hydrogens than the clean surface, whereas the unbonded step is a poor site. This implies that, if it can get there, hydrogen will stick onto a rebonded B-type step.

System	Saturated Bottom	Saturated Top	Isolated Pair
Clean	-5.98	-5.98	-5.98
Rebonded	-6.16	-5.92	-6.11
Unbonded	-5.78	-5.56	-5.76

Table 6.1: The energy for adsorbing pairs of hydrogens on a clean surface, and at a B-type step (both rebonded and unbonded) from TB. All energies are in eV. "Saturated" refers to complete coverage of the step; "Isolated" means that there is at least one clean dimer row between neighbouring sites.

The most obvious manner for a hydrogen to get onto a rebonded B-type step is by diffusion down from the terrace above; diffusion along the step will be blocked by any other hydrogen already present, and may well have a higher barrier due to a more stable start point (by analogy with the saturated surface diffusion). The diffusion from the top of the step onto the base has been modelled with TB, and the results are presented in Figure 6.11(a) below, along with a schematic diagram showing the process in Fig. 6.11(b). As with all processes modelled in this chapter, the atom labelled 'D' was not allowed to bond to the hydrogen.

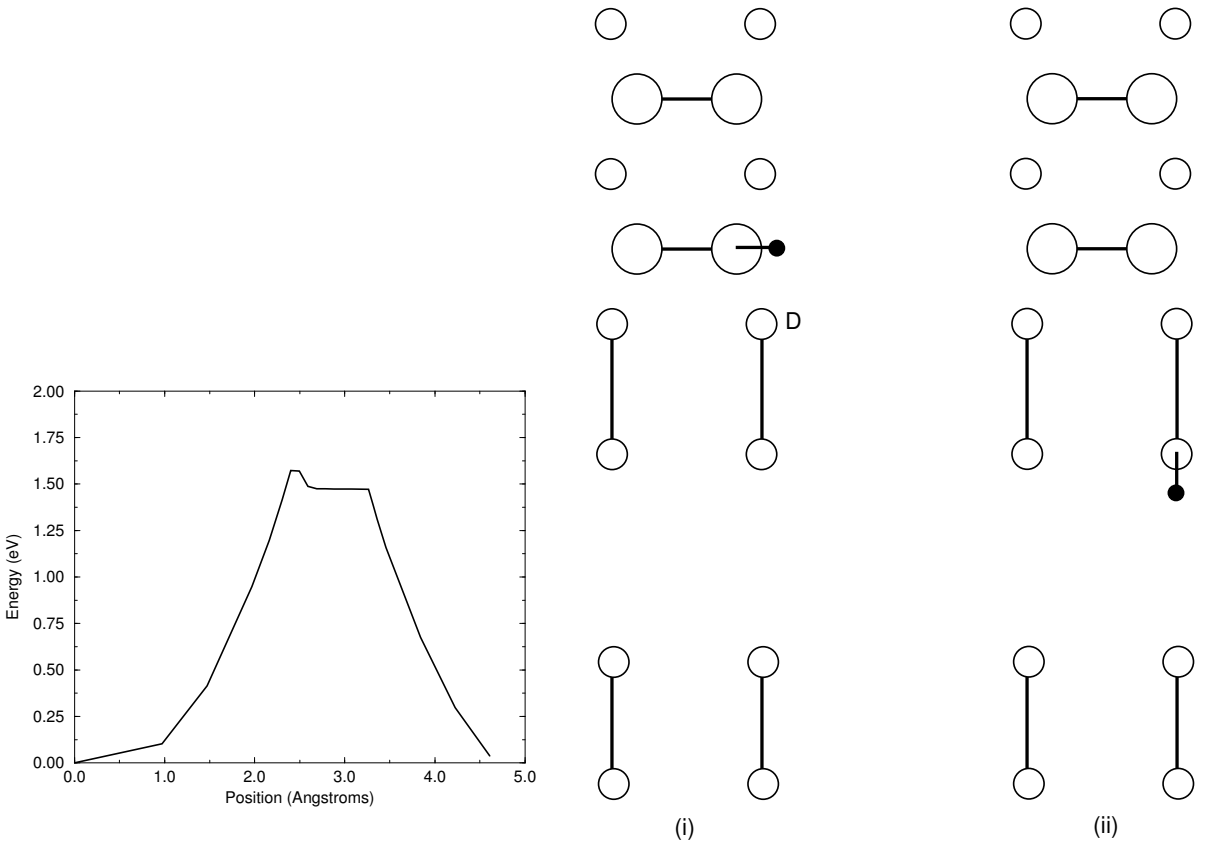

Figure 6.11: (a) The energy for a hydrogen atom diffusing down a B-type step, from the upper terrace to one of the rebonded atoms below. The barrier is 1.57 eV. (b) The process indicated schematically. (i) is the start point and (ii) is the end point.

The diffusion was modelled using TB with a unit cell in which there were 5 dimers in the upper terrace, and the equivalent of 5 in the lower (as discussed in Chapter 4, this is the smallest unit cell for a rebonded step which has no clear interactions between steps), with the bottom layers of the cell terminated in hydrogen and frozen in bulk-like positions in the usual manner. The diffusing hydrogen was constrained in planes of constant y (lying *along* the dimer row in the *upper* terrace - the y axis runs vertically up the page in Fig. 6.11(b)), and only permitted to bond to the atoms it starts and finishes on - specifically not to atom D. The barrier found is 1.57 eV, which is reduced to 1.37 eV taking into account the correction. This is 0.4 eV lower than the barrier on the clean surface, and means that by the time the hydrogen is diffusing on the clean

surface, it will diffuse extremely fast down the steps (3000 times as fast as on the clean surface at 600K). The plateau just to the right of the peak is an unusual feature, and is caused by the geometry of the system. Whereas for diffusion on the clean surface the mid-point is unstable, and the hydrogen is only bonded to one atom either side of the peak, the step geometry allows the hydrogen to remain bonded to both atoms by increasing its height at each point away from the step edge. Eventually, the distance to the top of the step becomes too high, and the drop starts. In the context of saturating the bottom of a step with hydrogen, it is worth noting that the barrier is equally low for the hydrogen to *climb* the step as to descend it. There are, however, other factors that will combine to maintain the saturation: the extra stability gained from pairs of hydrogens at the bottom; and the population of hydrogen on the upper terrace - if the site at the top of the step is occupied (as it may well be, given the amount of hydrogen on the surface), then the diffusion will not occur.

To sum up, the evidence for hydrogenated B-type steps during growth is as follows: a pair of hydrogens is 0.2 eV more stable at a rebonded B-type step than on a clean dimer; the barrier for diffusing down a step is 1.37 eV; hydrogenating a step does not stop enhancement in STM at low voltages (see Chapter 4); experimental images show a lack of denuded zones. Altogether, this offers a strong case for believing that the steps are hydrogenated, and this is what causes the edges to stop being a nucleus for dimer growth.

6.5.2 Hydrogen at defects: nano-mirrors

While the bonding of the 1DV has been discussed in Chapter 4, and shown to be such that it is *less* reactive that an equivalent dimer site, the bonding of the dimers either side of it has not been investigated. If they pin - or alternatively repel - hydrogen, then this will have an effect on the mobility of hydrogen and hence growth. As was the case for steps, the first test is to investigate stabilities for adsorbing hydrogen, and these are shown in Table 6.2. These energies come from TB calculations, with a unit cell consisting of two standard unit cells side-by-side, with one defect in the system.

Clean		Defect	
No of Hs	Energy (eV)	No of Hs	Energy (eV)
0	0.00	0	0.00
1	-2.56	1	-2.69
2	-5.98	2	-5.98
3	-8.53	3	-8.63
4	-11.95	4	-11.92

Table 6.2: The energy for adsorbing hydrogen at a single missing dimer defect, compared to a clean surface.

There is little conclusive to draw from these data, except, perhaps, that the defect is a sink for a single hydrogen. This can be understood by considering the bonding: the defect is a strained bonding situation, and the energy gained by putting a hydrogen on the dimer must be offset against the cost of flattening the dimer slightly (the angle drops from about 15° to about 3°); this becomes exacerbated when two hydrogens are placed on the dimer, and it must become flat. However, the energy gained by saturating the surface state means that, overall, the two situations are equal. More information about the effect of the defect on hydrogen can be gained by considering the diffusion away from the defect, and the energy for this calculated using TB is shown in Figure 6.12(a), along with a schematic diagram indicating the diffusion cell modelled in (b). The unit cell used was six dimers long and two dimer rows wide, with a defect in one of the rows. The constraint applied was, as above, forcing the hydrogen to remain in planes of constant y, with the y axis running vertically up the page in Fig. 6.12(b). As usual, atom D played no part in the diffusion reaction. The defect has pulled the dimers next to it towards each other, resulting in a longer distance for the hydrogen to diffuse.

This figure shows that the defect acts as both a sink for hydrogen atoms (for once they are adsorbed next to it, they require a large amount of energy to get off) and also a mirror (it is as hard to get on as to get off). As with the step, it is the bonding of the situation which is responsible for the barrier height. The dimer which the hydrogen starts on is next to a 1DV, and is pulled away from the normal position in the surface. This means that the distance between the dimer the hydrogen starts on and the one it is going towards is increased - as can be seen by comparing Figs. 6.6 and 6.12. This increased distance means that at the mid-point, the hydrogen atom has much longer bonds to both dimers, giving an increased barrier relative to diffusion on the clean surface. The barrier is remarkably symmetrical, which implies that the

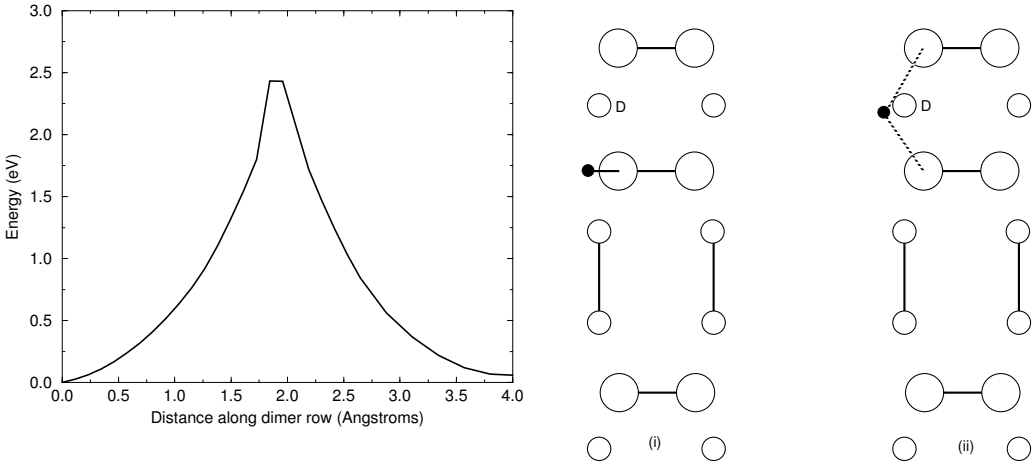

Figure 6.12: (a) The energy for a hydrogen atom diffusing away from a 1DV, starting next to the defect and moving away. The barrier is 2.5 eV. (b) The process indicated schematically. (i) shows the start point and (ii) the midpoint.

defect may well act as a mirror - any hydrogen adsorbed *two* dimers away from a defect will be far more likely to diffuse away from the defect than towards it, due to the increased barrier of 0.5 eV.

This resolves a question from experiment (Owen 1996, p.72 and Fig. 4.7); there were several observations of a smudge trapped between two defects, which then disappeared at high temperature (700 K). The only explanation given the above result is that the hydrogen atom hopped *across* dimer rows (an event seen experimentally at 700 K, though very rarely), rather than across the defect.

Conclusions The behaviour of hydrogen has been investigated, using tight binding and LDA. The diffusion on a clean surface has been modelled, and shows good agreement with experiment; the tight binding parameterisation has proved reliable. The mechanism for diffusion on the saturated surface has been found using experimental results to produce a path, and theoretical modelling to explain the behaviour along this path. The behaviour of hydrogen near steps and defects has been modelled, and an explanation provided for the lack of denuded zones in gas source growth. The rebonded B-type step proves to be a sink for hydrogen, while the 1DV will act as a mirror due to the high barrier for diffusion onto dimers adjacent to it.

Chapter 7

Growth of Si(001) from disilane

In Architecture *as in all other* Operative *Arts, the* end *must direct the* Operation. *The* end *is to build well.*

Sir Henry Wotton, *Elements of Architecture*

It is clear that silicon has become a substance of vital importance to Western society over the last 30 years, and is increasingly important in the developing world; the production of devices on silicon and the growth of silicon and new substrates (such as Ge/Si heterostructures and SiGe/Si devices) is of paramount importance for future technological development. However, there are many problems associated with these new substances. As was mentioned in Chapter 5, various undesirable effects can arise during growth due to the strain between the two different substances, such as Ge segregation during $Si_{1-x}Ge_x$ growth on Si, or 3D islanding of Ge on Si. The use of hydride precursors for SiGe/Si growth has been shown to reduce Ge segregation during growth (Ohtani *et al.* 1994); to understand this phenomenon, and better control it, requires an understanding of growth from hydride precursors. This thesis, along with James Owen's thesis (Owen 1996), gives details of the atomistic processes involved in growth of silicon from disilane, which is a necessary prerequisite for understanding growth of SiGe alloys.

The CVD growth of silicon (001) from disilane (Si_2H_6) has been studied using a variety of surface science techniques (Gates 1988; Gates, Greenlief and Beach 1990; Gates 1992; Liu *et al.* 1992; Mokler *et al.* 1992; Kolasinski *et al.* 1992), which have given an understanding of the process on a macroscopic scale, including the rate and results of growth. Disilane is a promising candidate for low temperature CVD growth because it has a relatively high sticking coefficient, and a low activation energy for breaking the Si-Si bond. This means that it can be used for growth without the presence of an activator such as chlorine, as is necessary in the case of, for instance, SiH_4 (Gates 1988).

Scanning tunneling microscopy (STM) studies (Boland 1991; Lin *et al.* 1992; Bronikowski *et al.* 1993; Wang, Bronikowski and Hamers 1994; Owen *et al.* 1997a,b) have begun to give a real-space understanding of the detailed chemistry of disilane on silicon, and the initial stages of adsorption, dissociation and epitaxial growth. Boland (1991) studied the role of hydrogen on the silicon surface in the adsorption and dissociation of disilane, and showed that growth at high hydrogen coverage is governed by the reactions of disilane with the hydrogenated surface. He also observed that the result of annealing a disilane-saturated surface was strips of epitaxial dimer rows. Lin *et al.* (1992) studied adsorption at 350°C, and observed the formation of anisotropic islands, which they attributed to silicon dimers saturated with hydrogen. While they found that the fragments from disilane adsorption were randomly distributed, they did not identify the nature of these fragments. Bronikowski *et al.* (1993) identified the initial product as adsorbed SiH_3, which quickly breaks down to form SiH_2, with a hydrogen left on the surface. After postulating four candidate sites for the SiH_2 adsorption (shown in Fig. 7.2), they identified the actual bonding site from the geometry of the adsorbed fragment. A typical image of the Si(001) surface after adsorption of a small quantity of disilane is shown in Figure 7.1. SiH_2 refers to a fragment resulting from disilane adsorption bonded in the intra-row configuration (C in Figure 7.2), SiH_3 is the precursor to this, NRMH is a non-rotated (i.e. non-epitaxial) monohydride dimer, and defects and paired hydrogens have been discussed in previous chapters. The images are at different biases, which enables the defects (which show an enhancement) to be differentiated from the paired hydrogen (which don't).

In a later study, Wang, Bronikowski and Hamers (1994) investigated the initial stages of epitaxial growth which proceed from this adsorbed fragment. They suggested a mechanism whereby two SiH_2 groups in

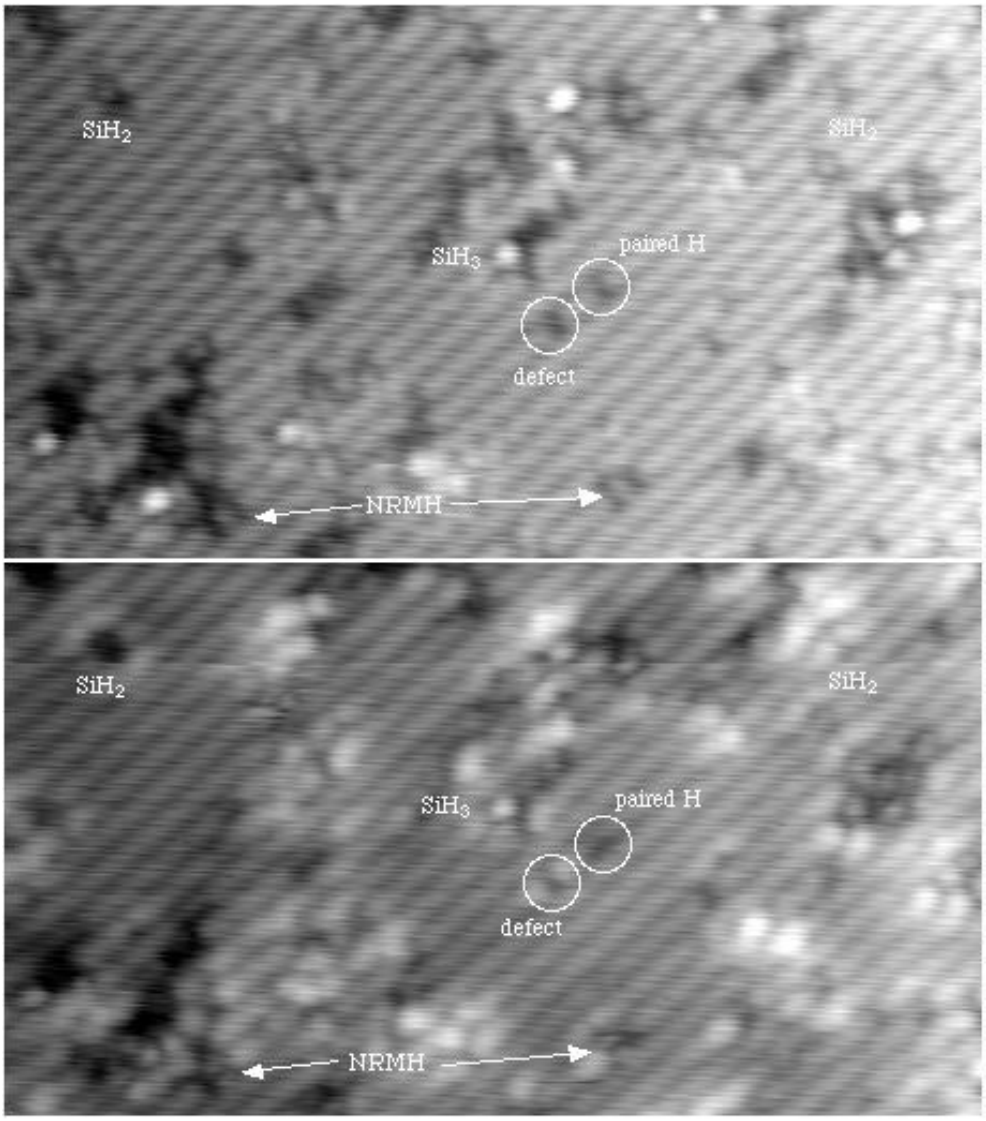

Figure 7.1: STM images showing the fragments formed from disilane adsorption. Labels are explained in the text. Voltages used were -2V and -1V. *Images courtesy of James Owen.*

position D in Figure 7.2 on adjacent rows bond to form a hydrogenated dimer, and lose an H_2 molecule. The only hydrogenated dimers which they observed were non-rotated, that is, parallel to the dimers in the surface.

7.1 SiH_2 fragments

In this section, the total energies calculated using LDA for an SiH_2 fragment in the four possible bonding sites in Fig.7.2 are presented, two of which prove to be very close in energy (Bowler and Goringe 1996). The energy for a second adsorbate, and therefore the interaction energies implied by this, have also been evaluated. The results imply that epitaxial hydrogenated dimers cannot form directly from SiH_2, but that the mechanism proposed by Wang, Bronikowski and Hamers (1994) for the formation of non-rotated hydrogenated dimers is plausible; it will be considered in Section 7.3.

A standard unit cell was used, with one SiH_2 fragment on the top surface. A vacuum slab of at least 7 Å was used to ensure decoupling of periodic images. A disilane molecule was relaxed in isolation for purposes of comparison.

The energies for the four different structures, and details of bond lengths and angles are given in Table 7.1.

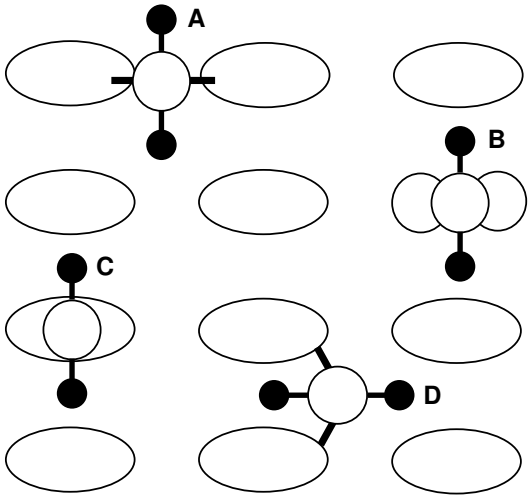

Figure 7.2: A schematic plan view of the possible bonding sites for an SiH_2 fragment: (a) between two dimer rows (inter-row); (b) inserting into a dimer bond (in-dimer); (c) on top of a dimer (on-dimer); (d) between two dimers in the same row (intra-row).

As can be seen, both the intra-row structure and the on-dimer structure are plausible candidates for the lowest energy structure.

		Bond lengths (/Å)		Bond angles (/°)	
Fragment	Energy(/eV)	$Si_{(a)}$-$Si_{(d)}$	$Si_{(d)}$-$Si_{(d)}$	H-$Si_{(a)}$-$Si_{(d)}$	$Si_{(d)}$-$Si_{(a)}$-$Si_{(d)}$
(a) Inter-row	0.313	2.37	3.92	109	105
(b) In-dimer	0.187	2.31	3.29	113	91
(c) On-dimer	0.004	2.32	2.43	118	63
(d) Intra-row	0.000	2.41	2.40	121	92

Table 7.1: Energies and bond lengths and angles for the different adsorption sites. Energies are relative to the intra-row site; subscript (a) refers to the silicon adsorbate, while subscript (d) refers to a dimer atom. The structures are illustrated in Fig. 7.2.

The inter-row structure is not stable with the substrate dimer bond intact, because the bond lengths required are too great. The figures given in Table 7.1 for A are with the dimer bond broken, and it can be seen that this structure is not energetically favourable. The in-dimer structure is found to be energetically unfavourable compared to the on-dimer structure, as was anticipated by Bronikowski *et al.* (1993), because the dimer σ-bond is broken.

There are two structures which have energies that make them plausible candidates for the stable adsorption site: the on-dimer and the intra-row. The on-dimer structure was discarded by Bronikowski *et al.* (1993), on the grounds that the angles in the structure (63° in Table 7.1) were sterically unfavourable. However, as shown in Table 7.1, the energy is nearly comparable with the intra-row structure. A possible reason for this stability is that the intra-row structure breaks the π-bonds on two dimers, and saturates only one dangling bond on each, whereas the on-dimer structure breaks the π-bond on only one dimer, and saturates both dangling bonds on that dimer. Bond angles close to sixty degrees are seen elsewhere in silicon chemistry : in the angular distribution for liquid silicon, a clear peak can be seen at 60 degrees (Horsfield and Clancy 1994); the bond angle formed between the silicon dimer and the Si-C bond when C_2H_2 and C_2H_4 adsorb on silicon is 73 and 75 degrees respectively (Andrew Fisher, Private Communication); and with a silicon interstitial defect, many of the angles formed are about 60 degrees.

The silicon-hydrogen bonds are, in all cases, within 1% of the length in the isolated Si_2H_6 molecule (1.53Å), which suggests that there has been very little change in the nature of the bonding. The adsorbate-dimer bond lengths in Table 7.1 can also be compared with Si_2H_6, in which the silicon-silicon bond length is found to be 2.31 Å. In the intra-row structure, this bond has been lengthened by about 4%; this is due to the geometric strain, and is a further reason for the relative stability of the on-dimer structure. The dimer

bonds are slightly lengthened when compared to the clean dimer length (2.37 Å) as a result of the partial breaking of the π-bond referred to above.

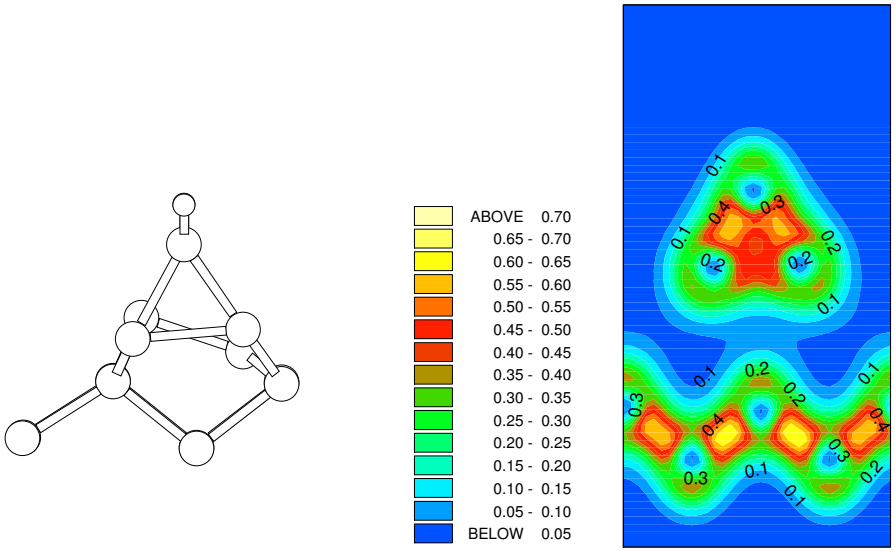

Figure 7.3: The atomic positions and charge density for an SiH_2 fragment in the on-dimer position on the Si(001) surface. Charge density is in units of electrons per $Å^3$.

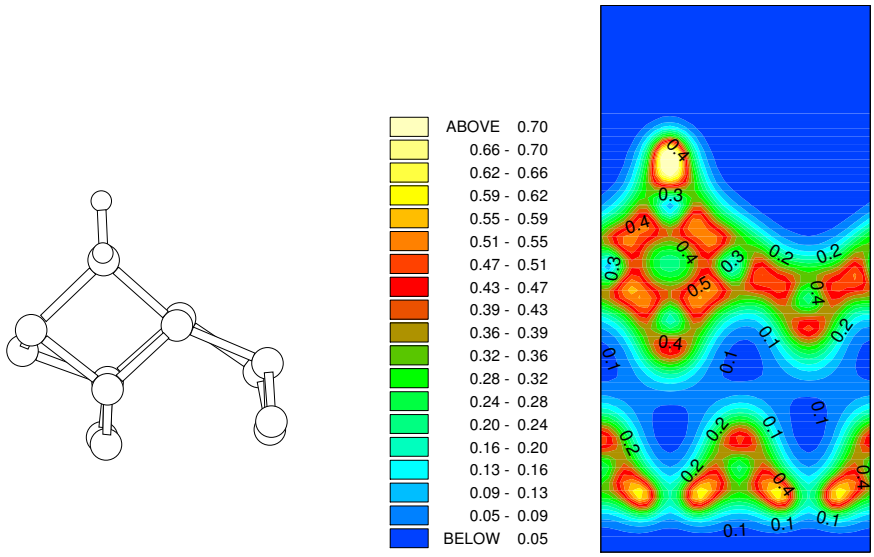

Figure 7.4: The atomic positions and charge density for an SiH_2 fragment in the intra-row position on the Si(001) surface. Charge density is in units of electrons per $Å^3$.

The structure and charge densities for the two minimum energy sites, with a clean dimer displayed for comparison, are shown in Fig.7.3-7.5. For the on-dimer structure (Fig.7.3), the bonds between the adsorbed silicon and the dimer atoms are as strong as a dimer bond despite the unfavourable angles, while the bond between the dimer atoms is weakened and lengthened (c.f. Table 7.1). The intra-row structure (Fig.7.4) has a four-membered ring. The bonds between the adsorbed silicon and the dimer atoms are as strong as the back-bonds, and indeed resemble them rather closely. In each case, the dimer bond length is increased somewhat compared to the clean dimer (2.37 Å). This is characteristic of the removal of the dimer π-bond.

Bronikowski *et al.* (1993) noted that the SiH_2 fragment did not cause the dimers in its immediate vicinity

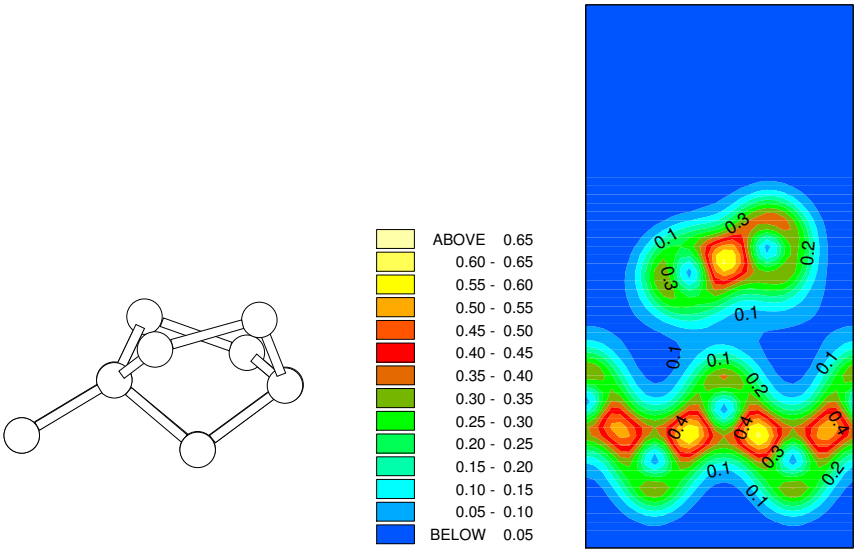

Figure 7.5: The atomic positions and charge density for a clean dimer in the Si(001) surface. Charge density is in units of electrons per $Å^3$.

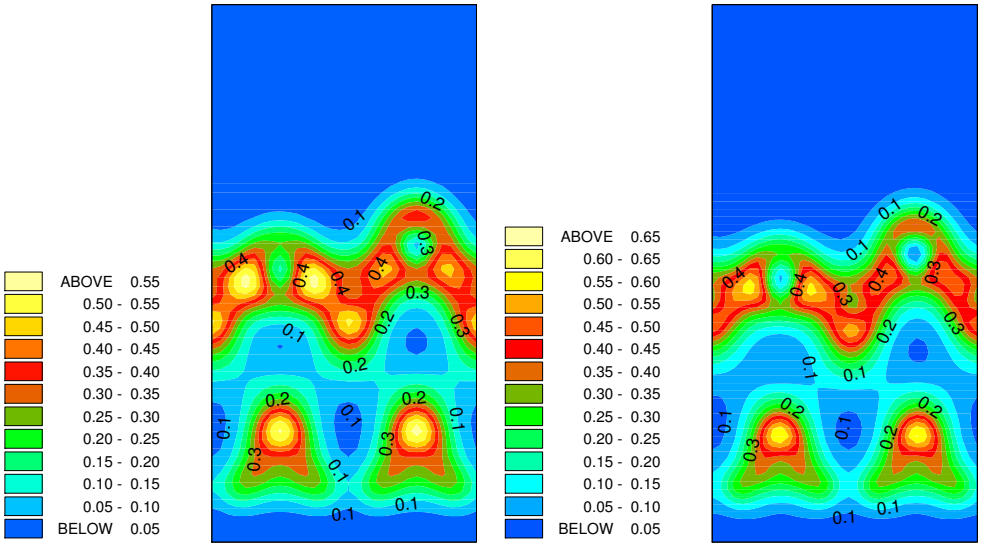

Figure 7.6: The charge densities for the ends of dimers (a) on the clean surface and (b) for dimers with an intra-row structure adsorbed on the other end. Note that the two are almost indistinguishable. Charge density is in units of electrons per $Å^3$.

to be pinned. Their explanation for this involved the electronic states of the dimers on which the fragment was adsorbed. Inspection of Fig.7.4 (the intra-row structure) suggests an alternative explanation: as the ends of the dimers on which the fragment is adsorbed (at the front) are at essentially the same height, and the other ends of the dimers (at the back) differ in height, there is no reason for the bare ends not to flip as clean dimers do. Indeed, in Figure 7.6 a contour plot of the charge density at the clean end of the dimers with the SiH_2 fragment adsorbed is shown with the same plot for the clean surface shown for comparison. The charge densities are almost identical, suggesting that there is a degree of delocalisation around the dimer-SiH_2-dimer system, causing the dimers to be linked and flip in concert. This will mean that the clean dimers in the vicinity are not pinned.

The next step in the growth of silicon from disilane is for two fragments to form a single dimer of monohydride, losing H_2 in the process (Wang, Bronikowski and Hamers 1994). For this to take place, two

SiH_2 groups must first be adsorbed on adjacent sites. An interaction energy of two such groups can be defined as the energy of two groups adsorbed on adjacent sites minus the energy of two groups adsorbed independently. For the on-dimer structure, this interaction energy is +0.20 eV (a positive energy indicates a repulsive interaction) per adsorbed fragment. This is very similar to the case of C_2H_2 and C_2H_4 on silicon, where the molecules adsorb on alternate dimers (Fisher, Blöchl and Briggs 1997; Mayne *et al.* 1996), due to a positive interaction energy, though the source of the repulsion may well be completely different. This large positive interaction energy will prevent adsorption on adjacent sites for the on-dimer structure except at very high coverages. By contrast, the interaction energy for the intra-row structure is very close to zero (magnitude less than 0.01 eV) so there will be no such effect.

Geometrically, an epitaxial hydrogenated dimer could form from the two adjacent SiH_2 fragments in the inter-row, in-dimer or on-dimer structures (see Fig.7.2). The first two of these structures are not energetically favourable even in isolation, and the repulsive interaction energy for the on-dimer structure prevents the adsorption of two SiH_2 groups on adjacent dimers. Therefore, epitaxial hydrogenated dimers cannot form from SiH_2 fragments.

A non-epitaxial hydrogenated dimer could form from two SiH_2 fragments in the intra-dimer structure adsorbed on adjacent dimer rows. The mechanism which has been proposed by Wang, Bronikowski and Hamers (1994) involves the formation of a Si-Si bond and the elimination of H_2, and is discussed in Section 7.3.

7.2 SiH_2 diffusion

In solid source MBE of silicon, the species arriving at the surface is silicon adatoms. These are free to diffuse about the surface after deposition (with a low barrier of 0.7 eV (Swartzentruber 1996)), and quickly form long, thin islands of epitaxial silicon dimers. The same is not true for gas source MBE, where the species left adsorbed after deposition are SiH_2 and H. The SiH_2 groups are adsorbed randomly on the surface after adsorption (Wang, Bronikowski and Hamers 1994; Owen *et al.* 1997a), with no correlation in position. By 450 K, these fragments have rearranged to lie next to each other on adjacent dimer rows: clearly, they must have started to diffuse. There are at least three possible diffusion processes which could go on: on-dimer to on-dimer (along the top of the dimer rows); intra-row to intra-row (along the side of the dimer rows); and on-dimer to intra-row or *vice versa* (diagonally across the dimer rows). As there is such freedom of movement, the energy surface for the SiH_2 group to diffuse along the dimer row with any of these process was evaluated using LDA, and is shown in Figure 7.7(a). The area of the surface considered is indicated in Fig. 7.7(b) with a dashed rectangle. A unit cell which was half of the standard unit cell (i.e. one dimer row wide) was used, with a single SiH_2 fragment on the surface.

The surface shows classic saddle-point behaviour, with a minimum barrier of 1.1 eV to go from on-dimer to intra-row or *vice versa* (which is nicely in line with the temperature found experimentally at which these groups start to move). There is another possibility for diffusion from on-dimer to on-dimer, with a barrier of 1.4 eV (with Fig. 7.7(a) showing only the first half of the process) but this will be less likely. This suggests that, once sufficient thermal energy is available, the groups will diffuse up and down the rows. Two on-dimer groups on adjacent dimers will repel each other, and it is unlikely that two intra-row groups could reside on the opposite sides of the same pair of dimers; while energetically favourable, it would be sterically hindered. The monohydride formation discussed in the next section, where the reaction proceeds between two intra-row groups on adjacent dimer rows, will be the most likely formation reaction.

7.3 Monohydride dimer formation

Once the SiH_2 groups have been observed to rearrange on the surface, the next feature observed in STM is a non-rotated (i.e. non-epitaxial) monohydride group, across the trench between dimer rows. It is highly likely that these form from two SiH_2 groups, which sit on adjacent dimer rows in the intra-row position. The reason that other orientations are not seen has been explained above - the on-dimer position has a repulsive interaction for two neighbouring groups, and the monohydride dimer on top of a dimer row is probably sterically hindered - the second group could not diffuse to sit in the appropriate position. As a point of comparison with clean dimers, the stabilities of monohydride dimers in the four positions available (on the dimer row and over the trench in both epitaxial and non-epitaxial positions) were modelled. These are shown in Table 7.2, and provide another interesting example of how vital the experimental-theoretical interaction can be: total energy calculations suggest that the most stable monohydride dimer, and thus the

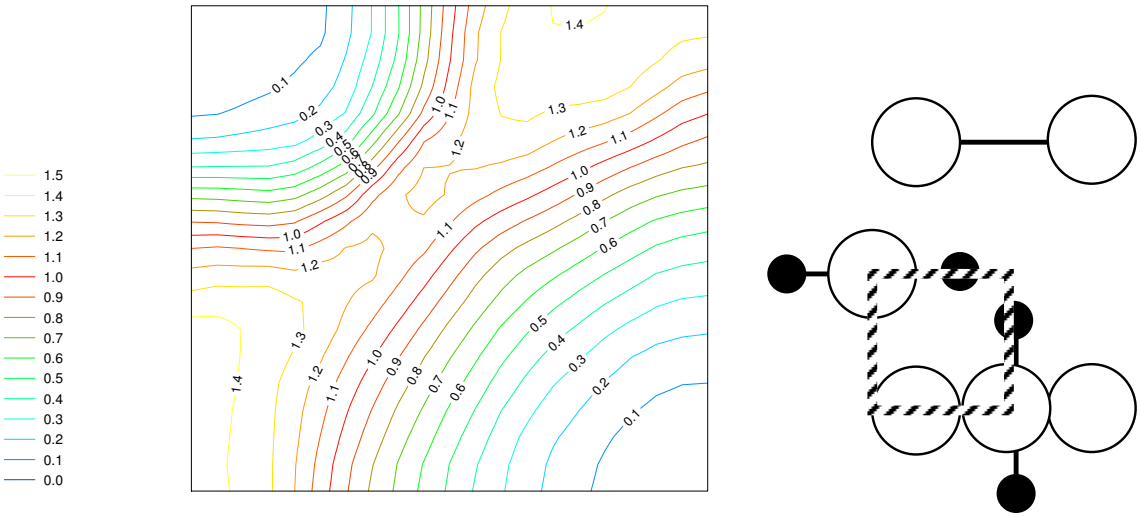

Figure 7.7: (a)The energy surface for the diffusion of an SiH_2 group along a dimer row. Energies are in eV. The bottom right hand corner represents the on-dimer position, and the top left corner the intra-row position, as indicated schematically in (b), where the dashed rectangle covers the same area as (a). The barrier along the diagonal is just over 1.1 eV. The interpolation method was the native bilinear interpolation of UNIMAP.

first candidate for modelling, is the epitaxial row dimer; only from experiment is it clear that the formation of the non-epitaxial trench dimer, third most stable structure, should be modelled.

Structure	Relative energy (/eV)
Row Dimer (E)	0.000
Row Dimer (N)	0.159
Trench Dimer (N)	0.680
Trench Dimer (E)	1.238

Table 7.2: The energies for a monohydride dimer in different orientations, relative to the most stable position. E refers to epitaxial (i.e. perpendicular to the substrate dimers), N to non-epitaxial (i.e. parallel to the substrate dimers).

The monohydride formation reaction was modelled in LDA, with a standard unit cell. Two SiH_2 fragments were placed on the top surface, in the intra-row position on adjacent dimer rows. For each point on the graph, the distance between the two Si atoms in the SiH_2 groups was fixed and the rest of the cell (apart from the termination) was allowed to relax. The Si-Si distance was decreased between calculations in order to force the formation reaction to happen. At each point, therefore, the whole SiH_2-SiH_2 system was free to move and rotate, but the Si-Si distance remained constant.

The reaction proceeds in three stages (shown schematically in Figure 7.8), for which the driving force is the steric repulsion of adjacent hydrogens. As the groups are pulled together initially, the SiH_2 groups undergo little change, except for a slight rotation to bring the hydrogens upright (Fig. 7.8(a)). Once the separation reaches 3.8 Å, one of the groups undergoes a drastic rearrangement, rotating so that one hydrogen points directly into the trench between dimer rows, while the other continues to move so that the hydrogens are upright (Fig. 7.8(b)). This conformation is held as the groups are pulled together, until a silicon-silicon bond is formed and two of the hydrogens (one from each group) react to give an H_2 molecule which is lost to the vacuum (Fig. 7.8(c)) in the calculation, or, in the STM experiment, to the gas phase. This is indicated with the letter A in Fig. 7.9.

The energy for these processes, along with that for stretching the dimer (Fig. 7.8(d)), is plotted in Figure 7.9. The reaction starts at the right hand side of the graph (as the separation is being *decreased*) and proceeds to the left. As can be seen, it is an endothermic reaction; however it is also *irreversible*, as the H_2 molecule is lost to the vacuum once it is formed. The three stages of the formation reaction are

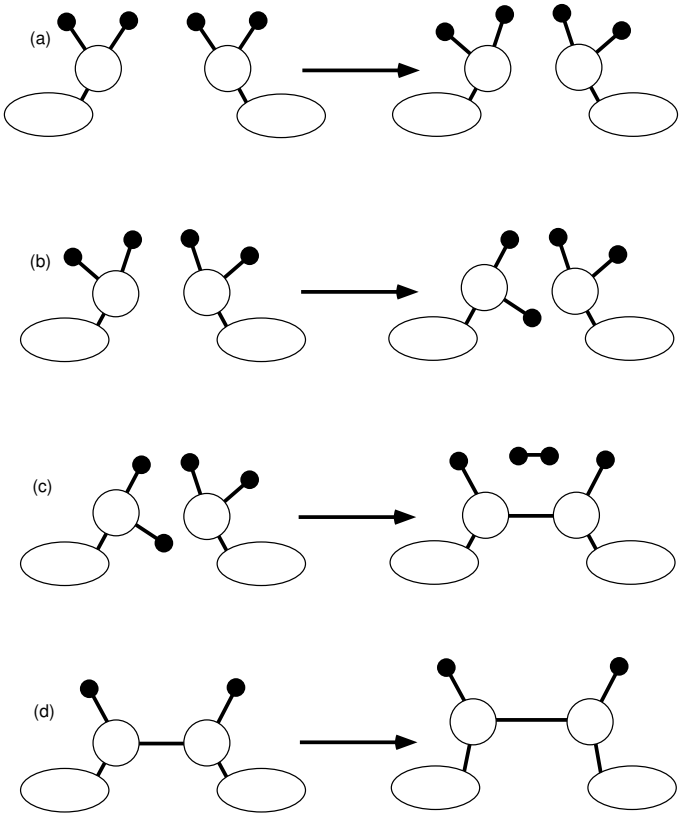

Figure 7.8: The four stages of the reaction, indicated schematically. (a) Initial stage: little rearrangement. (b) Central stage: one SiH_2 group undergoes a large rearrangement. (c) Final stage: the Si-Si bond forms, and two hydrogens are given off as an H_2 molecule. (d) Separation of the monohydride: exploring the vibrational frequency of the Si-Si bond in the monohydride dimer.

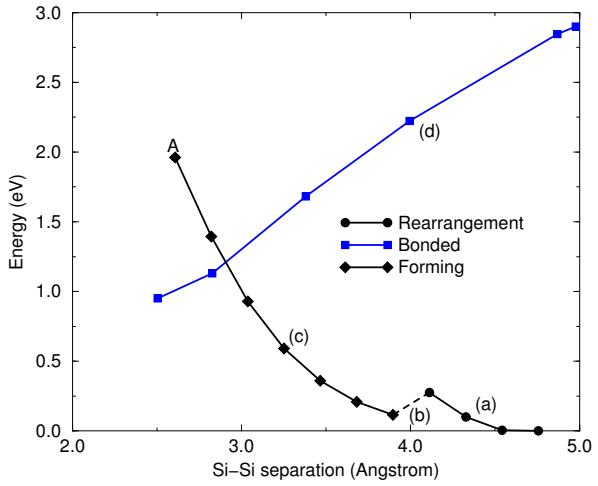

Figure 7.9: The energy for the formation of a monohydride ad-dimer in the non-epitaxial sense from two SiH_2 fragments.

indicated with circles and a solid line (for part (a) in Fig. 7.8), a dashed line (for part (b)), diamonds and a solid line (for part (c)), while the stretching mode (part (d)) is indicated with squares and a blue line. The

reaction proceeds with a maximum barrier of 1.9 eV; however, there may well be a lower energy path with a barrier closer to the cross-over between the two processes, in the region of 1.4 eV. From these graphs, and the experimental evidence, the true barrier is estimated as 1.4±0.2eV, though all that can be conclusively said at present is that the barrier is between 1.2 and 1.9 eV. The difficulty is that the rearrangement during the formation of the Si-Si bond (Fig. 7.8(c)) is an extremely slow process, requiring many hundreds of steps in LDA, which is costly in computational terms. This is hard to explore correctly; with more resources, and a parallel implementation of the nudged elastic band method mentioned in Chapter 2, the modelling of the barrier could be improved. Nevertheless, the process appears to be reasonable given the available evidence.

7.4 Rotation and dehydrogenation of ad-dimers

The rotation of a silicon dimer on top of the dimer rows has been observed and modelled in the past (Swartzentruber, Smith and Jónsson 1996); in this section, the DFT modelling of the rotation over the *trench* between dimer rows is discussed. This may have relevance to the process of conversion between the non-epitaxial monohydride dimers which are formed as discussed in the previous section and clean dimers which form the new layer of silicon, and this is touched on briefly at the end of the section. All of the calculations in this section used the standard unit cell, with a dimer (either clean or hydrogenated) placed over the trench between dimer rows.

7.4.1 Rotation of the clean dimer

The clean ad-dimer has three distinct structures, each of which is at least metastable. Initially in the calculations, it was placed in between the dimer rows, in the non-epitaxial position, as shown in Figure 7.10(a). This is stable with respect to rotations of up to about 30° (b). Once it rotates past this point, it migrates spontaneously into the least stable of the structures (c). This structure can be viewed as a transition stage between the two more stable structures, and is not placed symmetrically between the four neighbouring substrate dimers, but is shifted along the row. Each of the atoms in the ad-dimer are in a stable configuration; the dimer atom at the top of the figure in Fig. 7.10(d) is in the position it will occupy in the epitaxial ad-dimer, while the atom at the bottom is in the position it will occupy in the non-epitaxial ad-dimer.

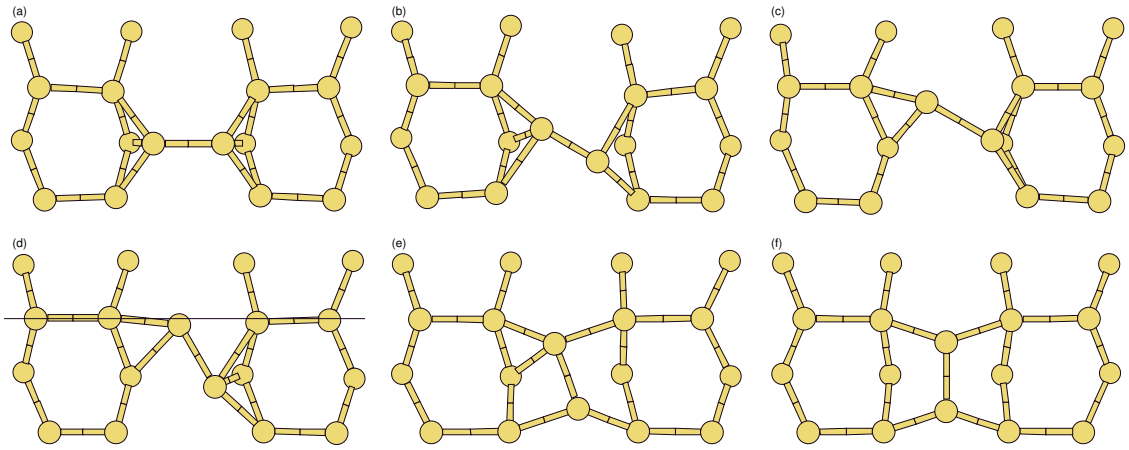

Figure 7.10: The atomic positions for the rotating, dehydrogenated dimer at 0°, 30°, 40° (shifted), 60° (shifted), 70° and 90°. The dimer is the central pair of atoms. The line on the 60° (shifted) schematic refers to Fig. 7.12.

If this metastable, transitional structure is rotated further, it again spontaneously transforms into the epitaxial ad-dimer (which is metastable with respect to the non-epitaxial dimer) once the angle is taken past 65° (e). Thus the three domains of stability for the different structures are from 0°-30°, the non-epitaxial dimer; from 30° to 60°, the transitional structure; and from 60° to 90°, the epitaxial dimer. The energy barrier for the process is shown in Fig. 7.11, where the central lines denote the energy for the transitional structure, and the outer lines the energies for the non-epitaxial and epitaxial dimers.

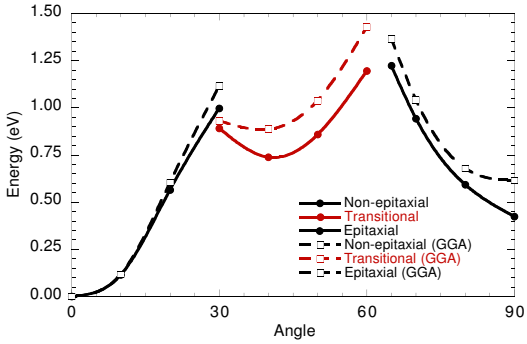

Figure 7.11: The energy for a clean dimer rotating over the trench from LDA (solid lines) and GGA (dashed lines).

The barrier to go from the non-epitaxial position to the transitional structure is 0.95 eV, and the barriers to go on from there are 0.2 eV to return to the non-epitaxial position, and 0.6 eV to go on into the epitaxial position. However, it seems more likely that the dimer will go straight from the non-epitaxial position to the epitaxial position (with a barrier of 1.2 eV) via the transitional structure, rather than residing in the transitional structure. The barrier in the opposite direction, i.e. to go from the epitaxial position to the transitional structure, is 0.95 eV and will not be changed by having to leave the transitional structure. The effect of gradient corrections is relatively small - only raising the overall barrier by 0.2 eV - though they do make the transitional structure less stable with respect to the non-epitaxial position.

The bonding at the top end of the dimer is quite peculiar. Figure 7.12 shows charge density contours taken in a plane containing the top pair of substrate dimers in Fig. 7.10, for three different angles. This plane is indicated with a thin line on the 60° (shifted) structure in Fig. 7.10(d). The process seen in the transition from the top to the bottom panel of Fig. 7.12 is the formation of a second bond between the rotating dimer and the substrate dimers; in the top panel, the ad-dimer atom (C) is at the same height as the substrate dimers (A, B and D, E), which would mean that it would form a pair of bonds with an angle between them of 180°. This would be highly unusual bonding for silicon, and indeed does not happen: despite the fact that the distances are almost identical, there is no bond between the ad-dimer and the right-hand substrate dimer. As the rotation proceeds, and the ad-dimer rises out of the plane of the substrate and moves back down the dimer row, the bond can be seen to be forming, until in the bottom panel the two bonds to the substrate are of equal strength.

The overall barriers to rotation (1.4 eV from non-epitaxial to epitaxial and 1.15 eV in the other direction) indicate that this process can occur at the substrate temperatures where diffusion and dimer string formation are starting to happen. Therefore, the orientation of the ad-dimers seems not to matter. The rotation may also have relevance to a diffusion mode *across* dimer rows recently observed by Borovsky, Krueger and Ganz (1997).

7.4.2 Rotation of the hydrogenated dimer

The hydrogenated dimer is less complex in its behaviour than the clean dimer, as the hydrogen reduces the degrees of freedom in bonding for the ad-dimer by soaking up the charge which was free. The non-epitaxial dimer can rotate smoothly to the epitaxial position without migrating along the dimer row (so that the centre of mass of the ad-dimer lies between the four neighbouring substrate dimers at all times), and while there is a lower energy path with the dimer shifted along the row (analogous to the clean case), this state is neither metastable (as it is in the clean case) nor reached spontaneously; the barrier to go from the symmetric structure to the shifted structure is not known. The physical positions of the atoms are shown in Figure 7.13.

The energy barriers for the two processes are shown in Fig. 7.14, indicating clearly that the barrier to rotation in the shifted position is lower, if it can be reached. It is not surprising that the significant atomic rearrangement seen in the clean case cannot occur here, given the smaller number of degrees of freedom for the system. The barrier will lie between 1.6 and 1.9 eV, depending on how easy it is to reach the shifted position; however, these energies correspond to a temperature far higher than the highest temperature at which monohydride dimers are still seen on the surface (Owen 1996), making this rotation mode far less

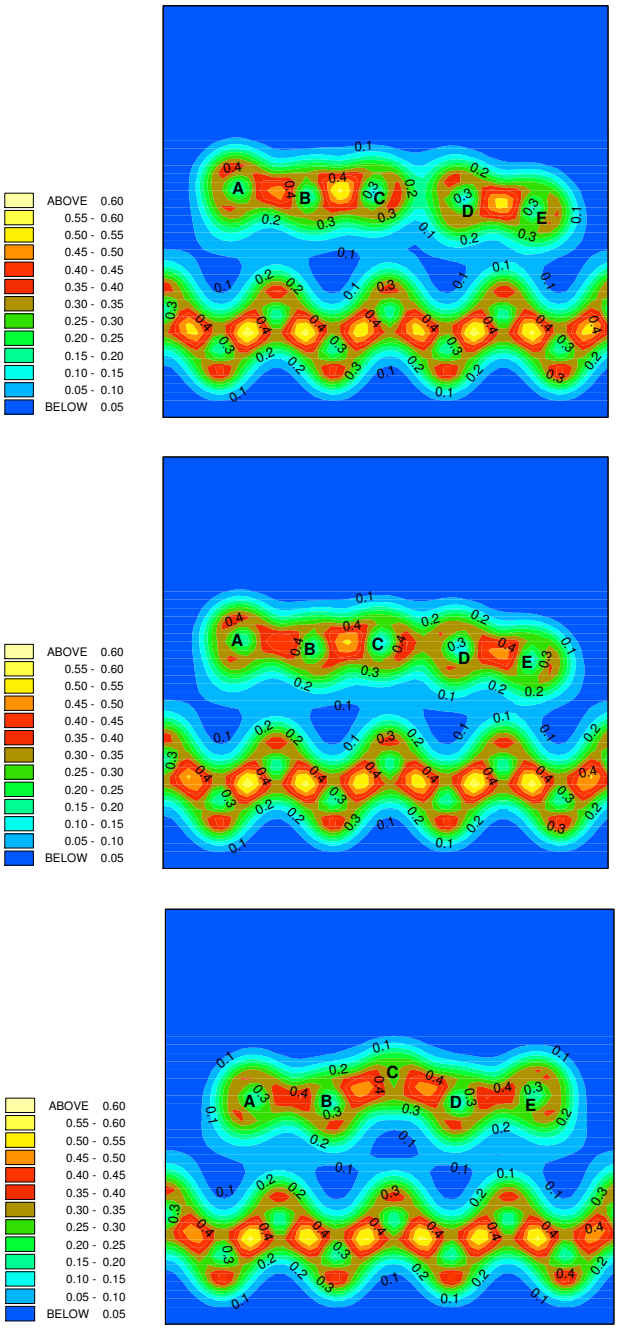

Figure 7.12: Contour plots of charge density for one end of the rotating, clean ad-dimer. Compare the bond strengths (as shown by regions of high charge density) between the central ad-dimer and the substrate dimers. The plane in which these contour plots were made is shown with a line on the 60° structure in Fig. 7.10. The letters indicate the positions of the atoms in the substrate dimers (A,B; D,E) and in the rotating ad-dimer (C). Charge density is in units of electrons per $Å^3$.

relevant overall than the clean rotation mode. As the *post hoc* GGA corrections are only likely to raise these barriers, they were not evaluated. The height of the barrier does explain why the only hydrogenated dimers imaged are non-epitaxial. The area where the rotation becomes interesting is in a dehydrogenation mechanism, which will be described in the next section.

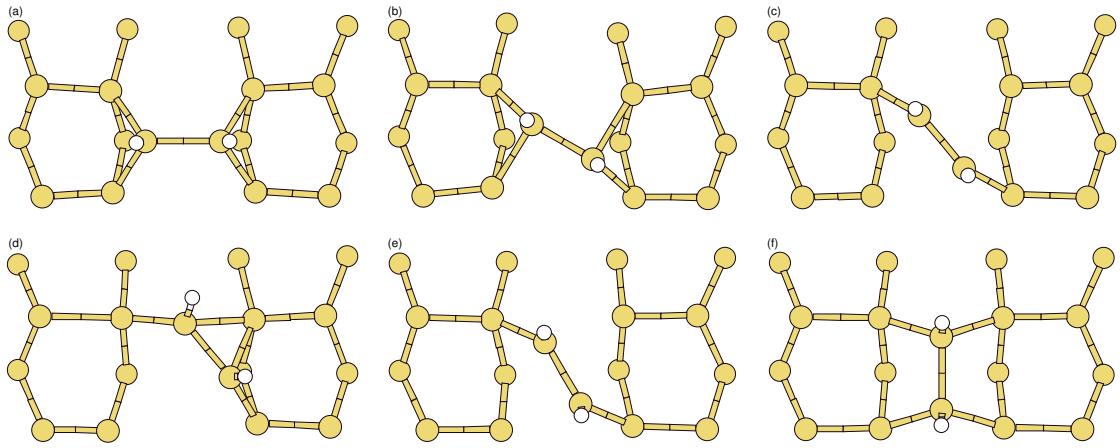

Figure 7.13: The atomic positions for the rotating, hydrogenated dimer at 0°, 30°, 50°, 50° (shifted), 60° (shifted) and 90°. The dimer is the central pair of atoms with hydrogen (small white dots) attached.

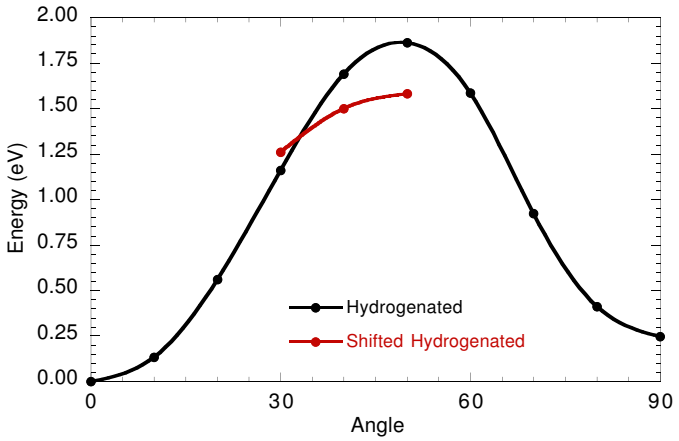

Figure 7.14: The energy for a hydrogenated dimer rotating over the trench from LDA.

7.4.3 A possible dehydrogenation mechanism

The question of dehydrogenation of the non-epitaxial monohydride dimers which are formed is a very thorny one. The transition from hydrogenated dimers to clean dimers is seen at a temperature (420 K) which is much lower than either the lowest possible hydrogen desorption temperature (at least 700 K) or the temperature at which hydrogen becomes mobile on the clean surface (550 K) (Owen 1996). If the rotation mode for the hydrogenated dimer had included a transitional state which was entered with a barrier in the region of 1 to 1.5 eV, then that state could have played a part in the dehydrogenation mechanism (Owen 1996); however, as shown above, this is not the case. The situation which remains is that the rotation of the hydrogenated dimer probably plays a part in the dehydrogenation, but this is not well defined. To fully understand the mechanism would require a series of simulations for a diffusion event where a hydrogen atom moves off the ad-dimer onto the substrate, each with the dimer constrained at a different angle; this would have to be done using DFT, as tight binding cannot model the rotation, and would require a prohibitively large amount of computer time.

As a test of the feasibility of the diffusion off the dimer, a very simple calculation has been performed, using tight binding. In order to obtain the correct coordinates for the ad-dimer, the DFT positions for the 40° (symmetric) dimer were used, and all atoms apart from the hydrogen were constrained to remain in a fixed position. The hydrogen was forced to remain in vertical planes parallel to the substrate dimer rows, but was allowed to relax within those planes. The unit cell, as for the calculations above, was two dimer rows wide each of which was two dimers long, with five layers of silicon terminated as for all the DFT calculations. As with the diffusion calculations in Chapter 6, the hydrogen was only allowed to bond to the atoms where

it started (the ad-dimer) and finished (a substrate dimer). The energy for this process is shown in Figure 7.15(a), and the pathway is indicated schematically in Fig. 7.15(b).

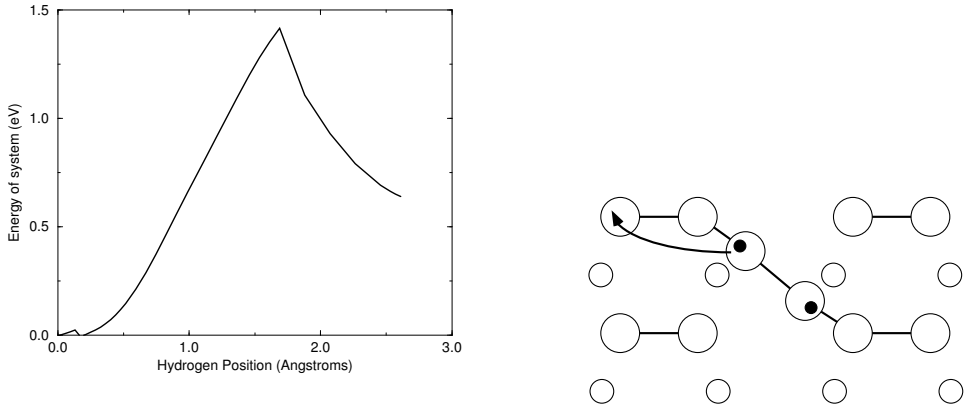

Figure 7.15: (a) The energy for a single hydrogen atom diffusing off a hydrogenated ad-dimer. The calculation was performed, using tight binding, by freezing the entire cell except the hydrogen. (b) The process indicated schematically.

The barrier found, 1.4 eV, is extremely reasonable given the temperatures at which the dehydrogenation is seen to occur, and suggests that this may be the mechanism. As the cell is frozen, it is not clear whether the correction of Chapter 6 should be applied. However, this mechanism is not conclusive, and more comprehensive calculations using DFT, with full relaxation of the cell at each stage, would be required for certainty.

7.5 Dimer diffusion along the trench between dimer rows

Experimental observations show that there is a large population of clean dimers over the trench after dosing with disilane above about 450 K. The next stage in growth requires these dimers to move, so that they can combine to form islands. The diffusion on top of the row has already been modelled (Yamasaki, Uda and Terakura 1996) and observed (Swartzentruber 1996) with reasonable agreement between the barriers. Recently dimers have been observed diffusing along the trench between dimer rows, with a measured barrier of 1.21±0.09 eV (Borovsky, Krueger and Ganz 1997). This diffusion has been modelled using LDA (Goringe and Bowler 1997), partially as an extension of the clean ad-dimer rotation mechanism, where the dimer moves along the row a short distance.

The single ad-dimer was placed between the dimer rows, over the trough in the non-epitaxial orientation (i.e. parallel to the underlying dimers), as shown schematically by a dotted line in Fig. 7.16b. In order to investigate all possible diffusion mechanisms, the energy surface (Figure 7.16a) was evaluated. For each point on the surface, the distance along the diffusion direction of the two atoms was fixed, and all other coordinates allowed to relax. The standard unit cell was used, which was the smallest cell which allowed no substrate dimer to be bonded to more than one ad-atom.

Line (AE) on Figure 7.16 represents the dimer diffusing intact and in its original configuration from a start site to an end site. This mode of diffusion gives a barrier of 1.8eV, which would be prohibitively high at temperatures below 800K; however, diffusion has been measured at temperatures as low as 450K (Borovsky, Krueger and Ganz 1997). There is an alternative path, shown as line (ABCDE) in Fig. 7.16(a), which gives a maximum barrier of 1.15 eV - in excellent agreement with the measured barrier of 1.21 eV.

This presents a method for diffusion (which is very similar to that predicted for diffusion *on top* of the dimer row (Yamasaki, Uda and Terakura 1996)) by break-up of the dimer (AB - barrier 1.15 eV), diffusion of one adatom (BC - barrier 0.36 eV), diffusion of the other (CD - barrier 0.58 eV) and reforming of the dimer (DE - barrier 0.23 eV). Since the energy barrier to the diffusion of adatoms (CD) is significantly lower than the initial breaking of the dimer (AB) (0.6 eV compared to 1.15 eV), it is likely that once a dimer breaks and the first adatom diffuses, the subsequent adatom diffusion will occur over a considerably shorter timescale. However, as the barrier is still considerably greater than kT, the second adatom will not always

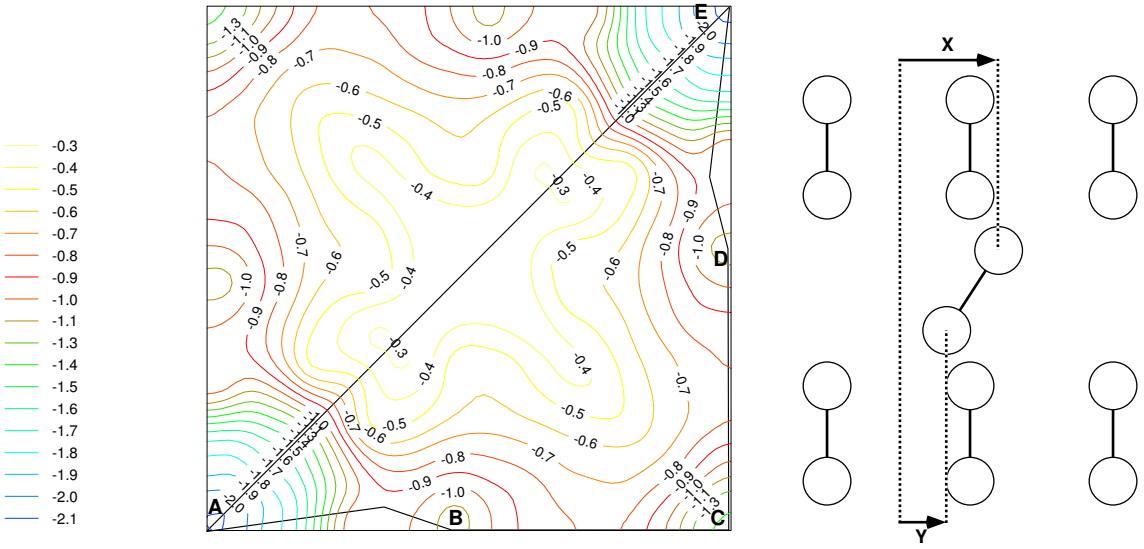

Figure 7.16: (a) The energy surface for a dimer diffusing over the trench on the Si(001) surface. Energies are in eV. The x- and y-axes are the coordinates along the diffusion direction of the two atoms, as illustrated in (b). Letters on (a) refer to the positions indicated schematically in (b): A is the initial position for the dimer, indicated with a dotted line; B is the metastable state during dimer break up; C is two adatoms on neighbouring sites; D and E are the analogues of A and B during reforming of the dimer. The lines added are as a guide to the eye only. The interpolation method was the native bilinear method of UNIMAP.

diffuse in the same direction as the first; it is equally likely that the first adatom will diffuse back to its original position. Indeed, given that the energy barrier (CD) is not significantly different from the adatom diffusion barrier in the absence of another adatom, it is almost equally likely that the second adatom will diffuse in the opposite direction.

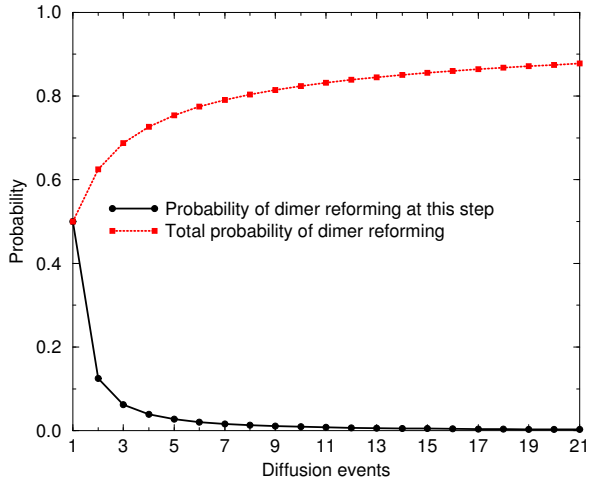

Figure 7.17: The probability for a dimer to reform after a given number of diffusion events. Because the dimer can only reform after an odd number of events, even number are omitted. The probability for reforming at a *particular* diffusion event is shown in black, and the cumulative probability of reforming at some point up to a given diffusion event in red.

The effect of this uncertainty in diffusion direction, and the typical distance the adatoms will move before the dimer reforms, can be investigated with a simple statistical model of the diffusion of two particles.

Assuming diffusion of adatoms only along the trench, the state of the system can be described in terms of $S(n,t)$, the probability that the adatoms are separated by n dimers after t adatom diffusions have taken place, and $D(d,t)$, the probability that the centre of mass of the two adatoms has moved $d/2$ dimers after t adatom diffusions. At any timestep one adatom diffuses at random, so the separation is equally likely to increase or decrease; $S(n,t) = \frac{1}{2}S(n-1,t-1) + \frac{1}{2}S(n+1,t-1)$, and the centre of mass is equally likely to move in either direction; $D(d,t) = \frac{1}{2}D(d-1,t-1) + \frac{1}{2}D(d+1,t-1)$. It is important to note that the S and D probability distributions are independent of one another.

The initial boundary condition must now be added: at $t = 0$, one adatom has hopped by one dimer in a random direction ($S(1,0) = 1$; $D(1,0) = 0.5$ and $D(-1,0) = 0.5$); and an additional condition that if the two adatoms come back together they rebind and are removed from the system ($S(1,t) = \frac{1}{2}S(2,t-1)$).

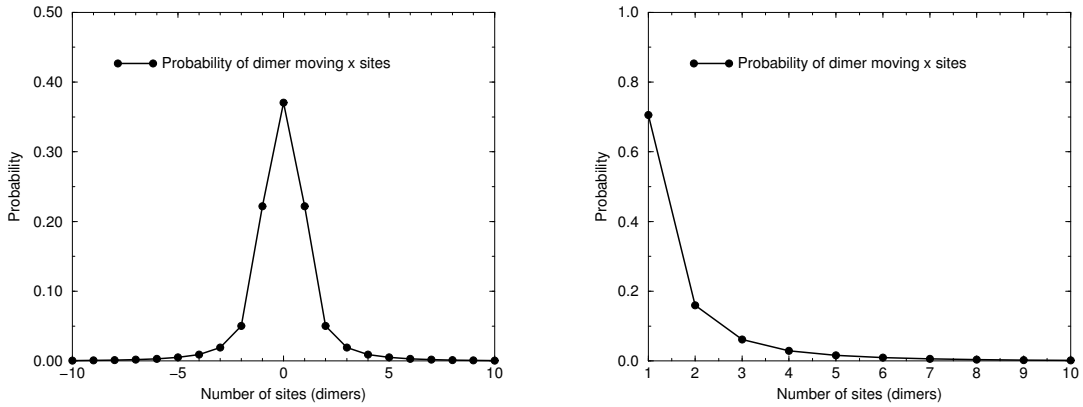

Figure 7.18: (a) The probability for a dimer to reform a given distance away from its start point. (b) The probability for a dimer to reform a given distance away from its start point given that it diffuses a non-zero distance.

The evolution of these probabilities can be numerically calculated extremely easily. The probability at any particular time step for reforming the dimer is then simply given by $S(0,t)$, and the cumulative probability of reforming the dimer by a given time is $\sum_n S(0,t)$. These are plotted in Figure 7.17. Of rather more interest is the distance that the dimer has moved when it reforms; the distribution of dimer diffusion distances is plotted in Figure 7.18(a). Since a diffusion distance of zero represents an event which will not be observed, Figure 7.18(b) shows the probability of diffusion distances given a non-zero distance. The majority (70%) of dimers in this model diffuse only one unit of distance along the dimer row, and only 14% diffuse more than two. This diffusion mode gives a barrier in extremely good agreement with experiment, and shows how the dimers which are formed from the disilane fragments can diffuse. Borovsky, Krueger and Ganz (1997) also observed dimers diffusing *across* dimer rows; this will be important in the formation of large islands at later stages of growth.

7.6 A new feature: the square

As has been mentioned above, clean dimers are formed on a Si(001) surface which has been dosed with disilane once the surface reaches 470K. These single epitaxial dimers amalgamate to form strings which grow anisotropically perpendicular to the substrate dimer rows (Boland 1991), by preferential adsorption of dimers at the B-type step edge at the end of the string (Pearson *et al.* 1996). The smallest length of epitaxial dimer string observed is three dimers; this is because it is the shortest length which allows rebonded step edges at each end. However, spontaneous formation of a three-dimer string from three isolated dimers is kinetically unlikely. The nucleation mechanism for these dimer strings is therefore unknown; this section describes a previously unknown feature on the Si(001) surface which may play a role in nucleation.

While this thesis is in the main about gas source epitaxial growth, there is enough contact with solid source (where no hydrogen is present) that it is worth reviewing the area. There has been considerable effort to determine the metastable structures that form as the nuclei of dimer strings in solid source MBE.

The primary candidates seem to be diluted-dimer rows (Bedrossian 1995), and a pair of epitaxial dimers, one on top of the dimer row, and one in the adjacent trench, which will be notated as TR (Yamasaki, Uda and Terakura 1996). Diluted-dimer strings are ad-dimers which are parallel to the underlying dimer rows occupying only trench sites, and attract other adatoms to form long chains of trench dimers. These chains can then react with further adatoms to form epitaxial strings (Brocks and Kelly 1996) and can also form long chains of diluted dimers terminated in ad-dimers, as well as larger structures(van Dam *et al.* 1996). They are not usually seen in GSMBE (Wang, Bronikowski and Hamers 1994). The primary reason for this is likely to be that the hydrogenated ad-dimers which form during GSMBE do not attract SiH_2 groups, and so long diluted-dimer chains do not form.

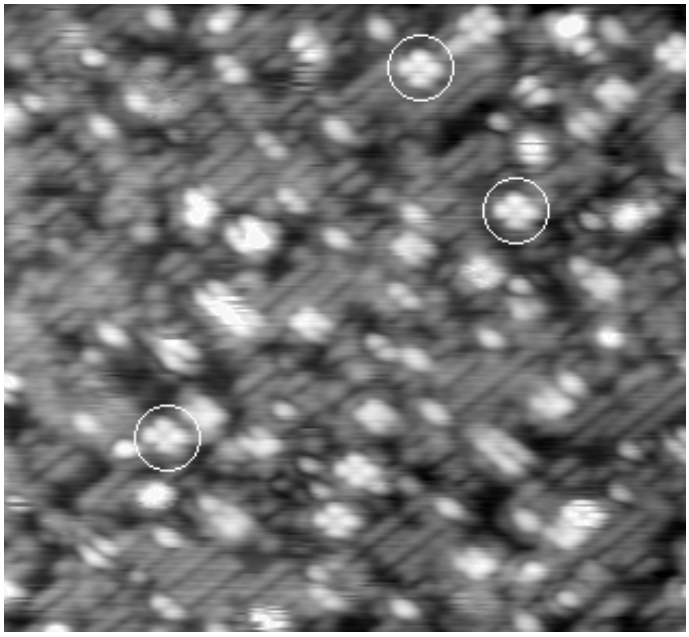

Figure 7.19: An STM image 200 Å square, taken at +2 V. Several examples of the features of interest (squares) are circled. They sit between two dimers and across two dimer rows. *Image courtesy of James Owen.*

7.6.1 Identification

An image of the disilane-covered Si(001) surface at 550 K is shown in Fig. 7.19. This is an empty-states picture, in which the epitaxial dimer rows appear as two bright lines, with a dark node running down the middle. The background dimer rows appear grey; missing dimers and hydrogenated dimers appear black. The bright features are mostly clean rotated epitaxial dimers. The new feature is circled, and will be referred to as a square (Owen *et al.* 1996). It is situated between two dimer rows, and between two dimers (a schematic picture can be seen in Fig. 7.22a). These squares are seen after dosing with disilane at a substrate temperature above 400 K and appear to be stable up to 600 K. In filled states images (Fig. 7.20a), the square has two perpendicular nodes, running vertically and horizontally across the picture, which are at 45° to the underlying dimer rows. In empty states (Fig. 7.20b), the nodes are weak and are orientated parallel and perpendicular to the underlying dimers. By comparison, the epitaxial dimer row above it (A) has a much stronger node along its centre.

Images at different negative sample biases were taken in order to probe the electronic structure. A pair of pictures is shown in Fig. 7.21. As the sample bias is reduced, the contrast of the square changes relative to the background. At -1.7 V (a) the square is brighter than the background, but as the voltage is reduced to -1.0 V (b), it darkens until it is the same apparent level as the dimer rows. From such variable-voltage imaging, and careful comparison with other features on the surface, it is clear that the square is not a pair of epitaxial dimers. This is important, as a pair of epitaxial dimers situated on neighbouring dimer rows (e.g. the lower pair of dimers in Fig. 7.20(b)) looks similar to a square in empty states.

The structure initially conjectured, four silicon adatoms arranged over the trench between dimer rows in

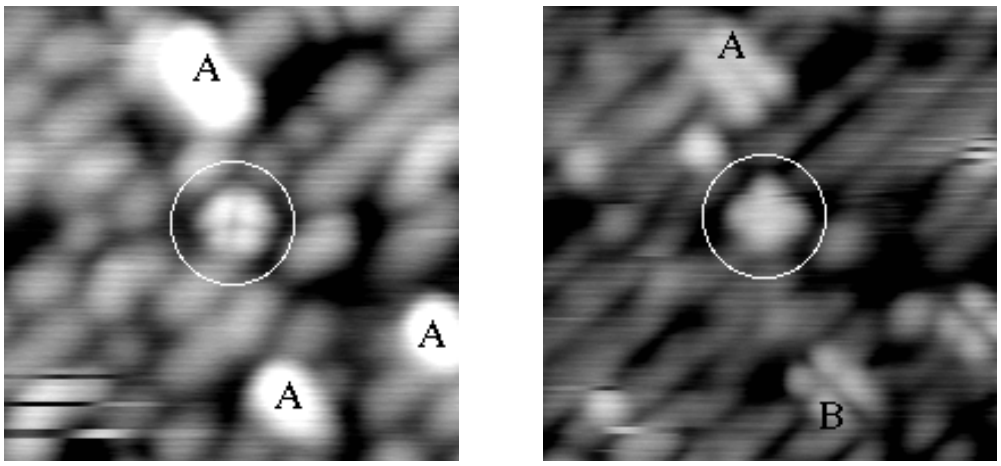

Figure 7.20: STM images 60 Å square, taken at (a) -2V and (b) +2V. At the top and bottom of picture (a) are white features (A) which are epitaxial dimer rows. In the centre (circled) is an example of a square, showing the nodes at 45° to the underlying dimer rows. At the top and bottom of picture (b) are doubled white features (A,B) which are epitaxial dimer rows. In the centre (circled) is an example of a square, showing the nodes lying parallel and perpendicular to the underlying dimer rows. *Images courtesy of James Owen.*

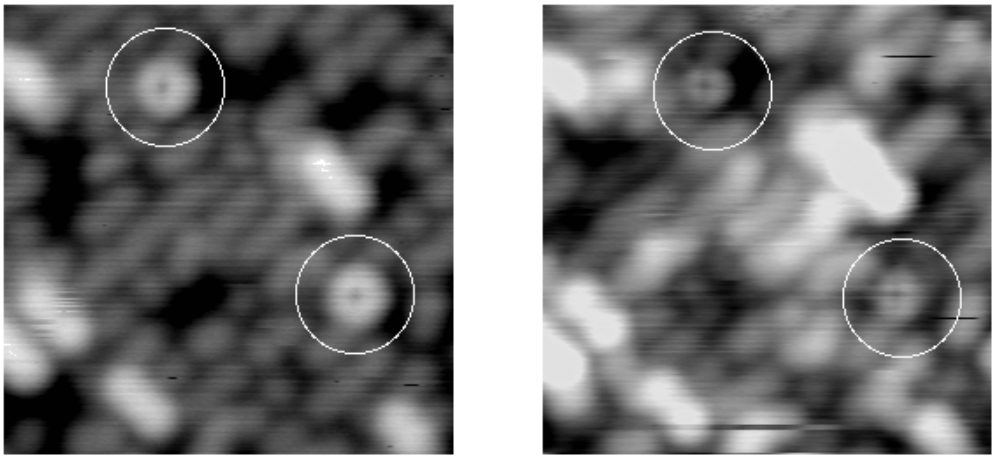

Figure 7.21: An STM image showing a square feature at (a) -1.7V and (b) -1.0V. The squares become darker and smaller relative to the dimer rows as the voltage is brought closer to the Fermi level.*Images courtesy of James Owen.*

a square ring, proved to be unstable in tight-binding, distorting to form a rhombus-like structure with two opposite corners higher than the other two. This is shown schematically in Fig. 7.22(a). The unit cell used consisted of two standard unit cells side-by-side (making a unit cell two dimer rows wide, each six dimers long) with the square placed between the dimer rows. The STM images, however are symmetric. Modelling showed that the square structure can oscillate back and forth between two mirror images, so that the average, as seen by the STM, is a square. The energy barrier to this flipping was calculated to be 0.15eV. At room temperature it will flip millions of times a second, which will result in a symmetric, square appearance in the STM. A schematic average picture of the states which the STM will image is shown in Fig. 7.22(b).

The tight-binding calculations (using two standard unit cells side-by-side) indicated that this rhombus-like structure was 1.1 eV more stable than two isolated ad-dimers on the surface, in the lowest energy position. However, tight-binding underestimates the energy of the most stable ad-dimer, so this is likely to be an overestimate of the relative stability of the distorted square structure. When energy comparisons were made with two ad-dimers on the clean surface using LDA, the square was found to be 0.18 eV more stable than two isolated ad-dimers. These calculations were performed using the standard unit cell.

The alternative candidate for the most stable structure of four atoms, the TR structure modelled by Yamasaki, Uda and Terakura (1996), is found to be 0.0 eV more stable than two isolated ad-dimers by Brocks and Kelly (1996), 0.32 eV more stable by Yamasaki, Uda and Terakura (1996), and 0.49 eV more stable in these calculations. There is clearly some uncertainty about the stability of this structure, but the square feature is likely to be a little less stable than it. The square is, however, seen in great profusion in STM images (Owen *et al.* 1997a), suggesting that it is common, and that the formation pathway may well be kinetically easier for the square than for the TR structure. This will be dealt with in more detail in the next section.

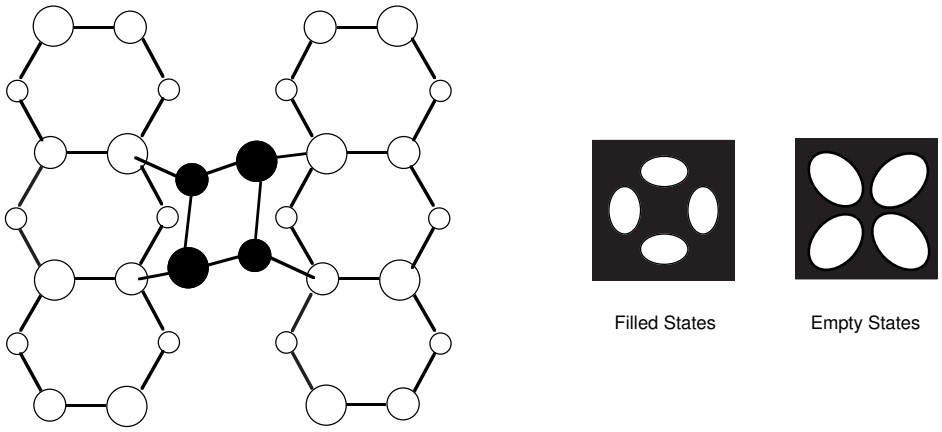

Figure 7.22: The proposed structure for the square. A schematic ball-and-stick model is shown in (a). The four atoms which make up the distorted square are shown by filled circles, the substrate dimer atoms are shown by open circles. A schematic diagram explaining the appearance in STM is shown in (b). For the two pictures in (b), the black square represents the background while the four white ellipses represent the electronic states imaged in STM. The diagram has been symmetrised for ease of presentation. On the left, filled states, showing that in STM dark lines (i.e. no electron density) will run diagonally across the square. On the right, empty states, showing that in STM the dangling bonds create the dark lines running horizontally and vertically across the square.

There is some question as to whether the bonding of the square to one end of a dimer will cause the π-bonds of the dimers to be broken, thus giving a bright appearance to these dangling bonds, as is seen in the case of hydrogen adsorbing on Si(001). However, in a manner similar to that shown above in section 7.1, the dangling bonds in adjacent dimers bonded to the square transfer charge between themselves, and thus give an appearance indistinguishable from the clean surface.

The stability and symmetrical appearance of the square have been explained by tight binding and LDA modelling. However, the behaviour with varying voltage (i.e. becoming darker relative to the dimer rows) is not - in fact, the opposite occurs, and the square becomes brighter in modelling. This suggests that, in the case of the square, the simple approach described in Chapter 1 and used in Chapter 4 is insufficient. The development of a better approximation must be left to others.

7.6.2 Formation and Transformation

While the square feature is clearly a common and stable feature on the silicon surface during growth, it is not clear from the above investigation how it forms. Given that the dimer seems to be the basic unit of growth in GSMBE, there are two possible arrangements of dimers which would allow the square to form. These are shown schematically in Figure 7.23. The arrows which are on this diagram reflect the results of the modelling of these processes, which offer a path for the formation of the TR structure.

The formation of the square from two non-epitaxial row dimers was modelled using tight binding by

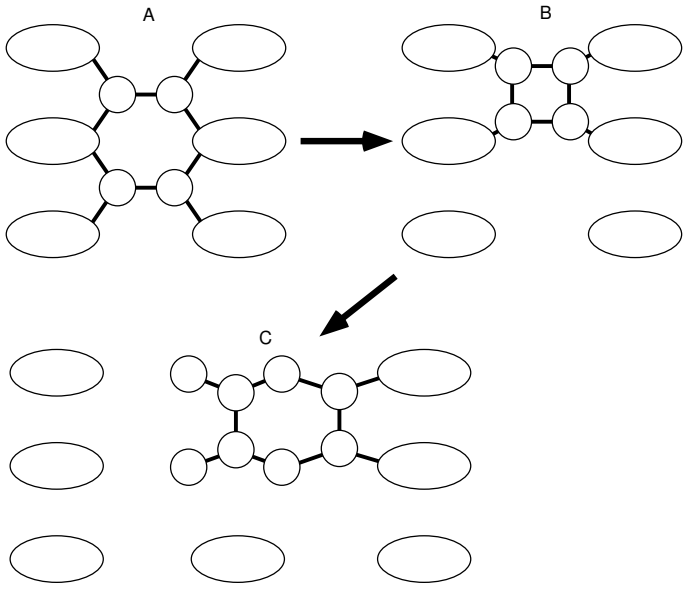

Figure 7.23: A schematic diagram showing the formation route for the square (B) from two dimers (A), and a possible alternative to nucleation of a TR (C).

fixing the separation of the centres of mass of the two dimers in the direction along the dimer rows (shown in Fig. 7.23) at each step on the pathway. The unit cell used was two dimer rows wide, each of which was four dimers long, with the usual termination. The energy barrier for the process is shown in Figure 7.24(a). The reaction proceeds from the right starting with two dimers (labelled A on the graph), as the distance between the dimers is being decreased, and reaches the square at a separation of 2.31 Å, labelled B. The net barrier is only 0.6 eV - extremely small, and indicative that the square will form from these two dimers very soon after they diffuse to adjacent sites at growth temperatures (the frequency of formation, assuming an attempt frequency of 10^{12} in an Arrhenius formula, is 6,000,000 times per second at 600K). The sharp drop at the left of the graph (leading to the formation of the square) is due to the strength of the bonds formed in the square. Given that clean dimers are formed from the disilane fragments deposited on the surface, and that there is a diffusion pathway along the trench between dimer rows, this pathway accounts for the large numbers of squares seen experimentally.

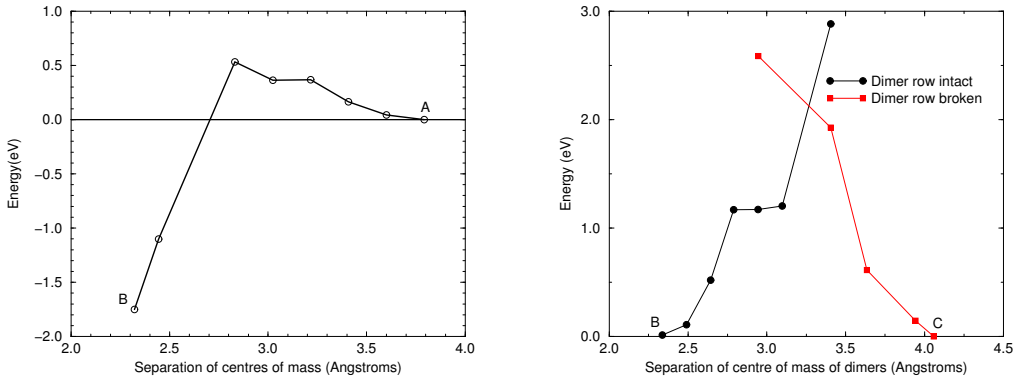

Figure 7.24: (a) A graph showing the energy required to form as square as indicated schematically in Fig. 7.23. (b) A graph showing the energy required to form a TR from a square, as indicated schematically in the lower half of Fig. 7.23.

The next stage in the growth process is to transform the square into something else - either into a TR feature, as shown in the second part of Fig. 7.23, or as described in the next section. The TR is the most

stable arrangement for four silicon atoms on the surface (Yamasaki, Uda and Terakura 1996, Owen *et al.* 1997) but is sterically unlikely to form from two ad-dimers, except from the cross-row diffusion mechanism. The decomposition of the square is a possible pathway for the formation of this feature (which would then have one "sticky" side for the formation of a small string of dimers). The reaction was modelled in tight binding by fixing the separation of the left and right pairs of atoms in the square in the direction perpendicular to the dimer rows (as shown in Fig. 7.23). The unit cell used was three dimer rows wide, each of which were four dimers long, and terminated in the usual manner. The energy for the reaction is shown in Fig. 7.24(b), and there are two parts to the graph depending on whether the substrate dimers are broken (labelled "dimer row broken" in red with squares) or intact (labelled "dimer row intact" in black with circles). The reaction starts from the square (labelled B) and proceeds to the right, arriving at the TR structure (labelled C). The barrier at the crossing point between these two curves is 2.1 eV, though this is, as in the next section, uncertain, and requires better modelling techniques on more powerful computers. The plateau on the "dimer row intact" curve represents the breaking of one of the square bonds, which gives a very soft vibrational mode to the system, allowing the flatness of the curve. However, the high barrier to formation may explains the relative infrequency of these features seen in experiment, though there may be a lower energy pathway based on the cross-row diffusion mode observed by Borovsky, Krueger and Ganz (1997).

7.7 Nucleation of dimer strings from the square

The square feature has the energetic advantage, in terms of growth, that it has two rebonded edges. If a trench dimer were to land next to either of these edges (as shown schematically in Figure 7.25(b)), it would be likely to stick (this position is at least 0.3 eV more stable than elsewhere on the surface in the same configuration) and then the whole ensemble could rearrange to give a string of three dimers (shown in the second part of Fig. 7.25(b)).

The reaction for formation of the string of three dimers was modelled in tight binding by constraining the centres of mass of the right-hand pair of atoms in a square (labelled as D1 in Fig. 7.25(b)) and an epitaxial dimer over the neighbouring trench (labelled as D2). The unit cell used was four dimer rows wide, each of which was four dimers long, and terminated in the usual manner. The energy for the reaction is shown in Figure 7.25(a). The plot starts on the right hand side, with the square and the dimer isolated (as shown at the top of Fig. 7.25(b) and labelled as B in (a)), and proceeds to the left. There are two curves, as was the case for the TR formation; the right hand one, referred to as "square intact" and labelled with circles in black, has the central substrate dimers and the square intact for most of its length, whereas the left hand one, referred to as "initial dimer formed" and labelled with squares in blue, has the left hand pair of square atoms forming the left-hand dimer of the string, the central substrate dimer broken and the middle pair of atoms of the square slowly forming the central dimer of the string. There must be a crossover point between these curves somewhere, but it is as yet not found. This puts the barrier at 1.5 eV as a minimum.

As with the formation of the TR, the use of the nudged elastic band method on a parallel computer would enable an accurate evaluation of this barrier. Nevertheless, it presents the energetics of the formation of a short string quite believably, and in good agreement with experiment: the value of 1.5 eV as a minimum is a little lower than would correspond to the temperature at which these strings are seen to form (Owen 1996); the barrier to adding a dimer onto a short string is 1.7 eV (Pearson *et al.* 1996). This evidence suggests that the model is reasonable, and that the nucleation process for dimer strings has been described.

7.8 The path from adsorbed disilane fragments to dimer strings

Having spent the last 7 sections discussing the details of different reactions, it seems reasonable to step back and review the entire process. This also forms a useful link to the next section, which concerns the defects which form once islands start to meet.

The entire pathway is summarised in Figure 7.26, which displays energy barriers for those processes modelled in this thesis. The sequence of events starts with the decomposition of disilane (in some manner) to yield SiH_2 groups (1) and adsorbed hydrogen. The SiH_2 groups diffuse around the surface (2) until two of them react to form a monohydride dimer (3). This then loses hydrogen (in a reaction involving partial rotation) to give a clean dimer (4). The clean dimer can diffuse around the surface (5), and two of them, once they are on adjacent sites, form a square(6). This then produces short strings of dimers, of length either two or three, by decomposition (7).

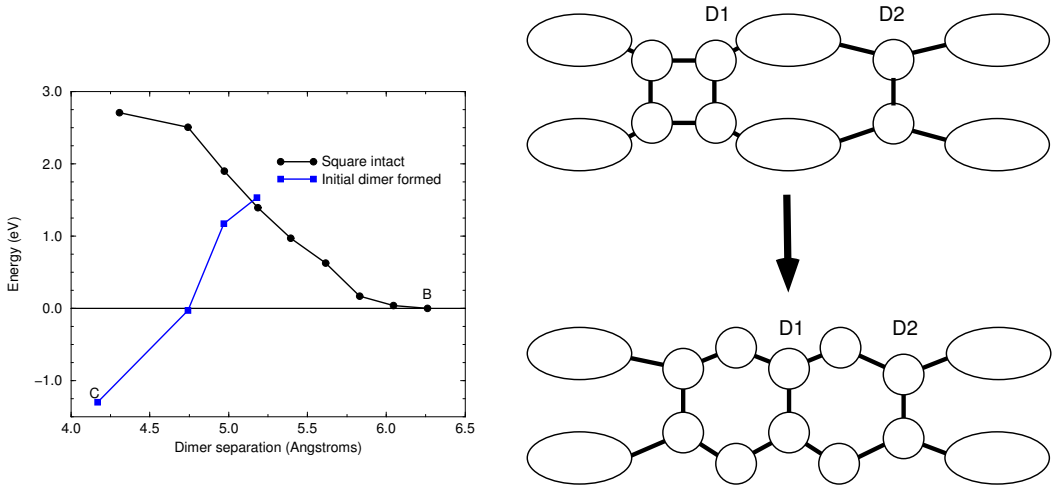

Figure 7.25: (a) The energy barrier for the formation of a string of three dimers from a square and a dimer. The x axis is the separation between the centre of mass of the dimer and the centre of mass of the right-hand pair of atoms in the square (shown in (b)). The reaction proceeds from right to left on the graph. (b) the reaction indicated schematically.

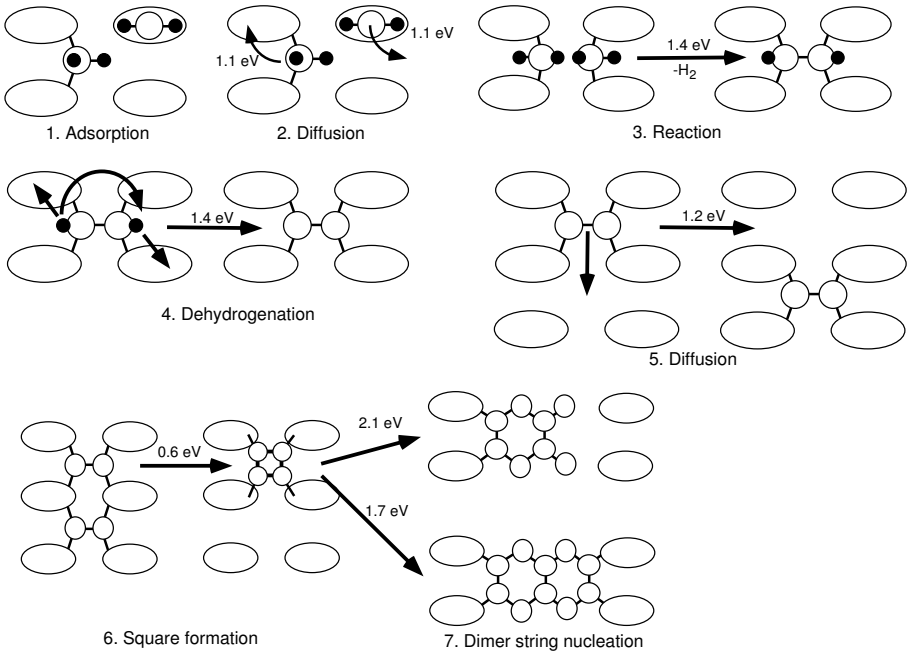

Figure 7.26: A schematic diagram showing the entire pathway from the adsorbed fragments, their diffusion and reaction, up to the formation of clean strings of dimers on the surface. The diffusion mechanism calculated above is not shown as there are many possible paths for diffusion; nevertheless, this mode is very important.

The behaviour once strings of dimers have formed is rather simple. These strings grow by addition at the ends, and by nucleation of a new string at the side(Pearson *et al.* 1996). The sequence of events will follow the pathway, resulting in short strings of dimers forming. Once this has occurred, these will grow into islands of new silicon in different places on the substrate. Unlike in SSMBE, where denuded zones are seen both above and below B-type step edges, these are not seen in GSMBE (Owen *et al.* 1997a,b). The reason for this is that the reactive rebonded B-type step edges are passivated with hydrogen, as discussed in Chapter 6. The next section covers what happens when two of these islands which are forming meet in anti-phase on the growing surface.

7.9 Anti-phase boundaries

In the previous sections the atomistic reactions leading to the formation of short dimer strings on the clean surface have been discussed. After the formation of these strings, the development of the surface is merely a matter of formation of new dimers from newly deposited SiH_2 groups and dimer diffusion. Once islands become large enough to start meeting as they grow in size, a new phenomenon occurs: the anti-phase boundary.

An anti-phase boundary (APB) can occur on the silicon surface during growth, due to the 2x1 (or higher) periodicity of the dimer reconstruction (Hamers, Köhler and Demuth 1989). An APB forms when the **trenches between dimer rows** in one island are **in phase** with the **dimer rows** on another island. It can serve as a nucleus for growth of a new layer of silicon (Bronikowski *et al.* 1993), and in gas source growth APBs are seen to nucleate islands when islands are not nucleating on clean terraces (Owen 1996). By analogy with steps on the silicon surface, there are two kinds of APB possible: one where the dimer rows are perpendicular to the APB (which will be referred to as a B-type APB, rather than AP2 as in Bronikowski *et al.* (1993)) and one where the dimer rows are parallel to the APB (the A-type APB). The B-type is considerably more interesting to the study of growth, as it is small enough in its most stable form to allow growth of a string of dimers along it, whereas the A-type APB is too wide. This is an experimental result - nucleation is seen to occur on B-type APBs but not on A-type.

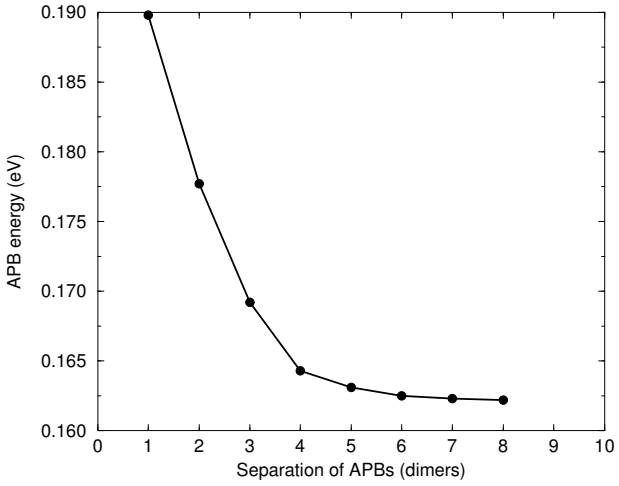

Figure 7.27: A plot of APB energy against separation of APBs for two 0DV B-type APBs in a unit cell.

The interest in APBs occurs because they are high energy sites, and can therefore lower the barrier for growth of strings (a string on an APB will remove it, reducing the energy of the system). The nucleation of a string of dimers on an APB will concern us later in this section; first, the structure must be elucidated. As with the study of steps and kinks, it is important to determine the appropriate size of supercell to use for isolation of APBs. Figure 7.27 shows the energy per APB against separation for a unit cell containing two 0DV type-B APBs (the notation is explained in the next paragraph). The calculations were performed with TB, using cells one dimer row wide and of varying length (from two to sixteen dimers long) and terminated in the usual manner. The graph shows that almost complete convergence is attained at a separation of eight dimers, though convergence to meV accuracy is reached at a separation of five dimers. In the calculations of structures, the separation of eight dimers will be used, while in the larger unit cells used for kinking and dimer string nucleation calculations a separation of five dimers will be used. This is again a compromise between isolation and computational speed - the large cells have over 2,000 atoms in them.

Now that isolation has been addressed, the most stable structure must be found. The APB could have dimers missing between the two different phases, and the structures considered have been named by analogy with the missing dimer defects (so that an APB with one dimer missing is a 1DV APB). In this work, structures with up to two dimers missing in the APB have been considered, as well as that proposed by Hamers, Köhler and Demuth (1989) from an experimental (STM) observation. Schematics of the structures

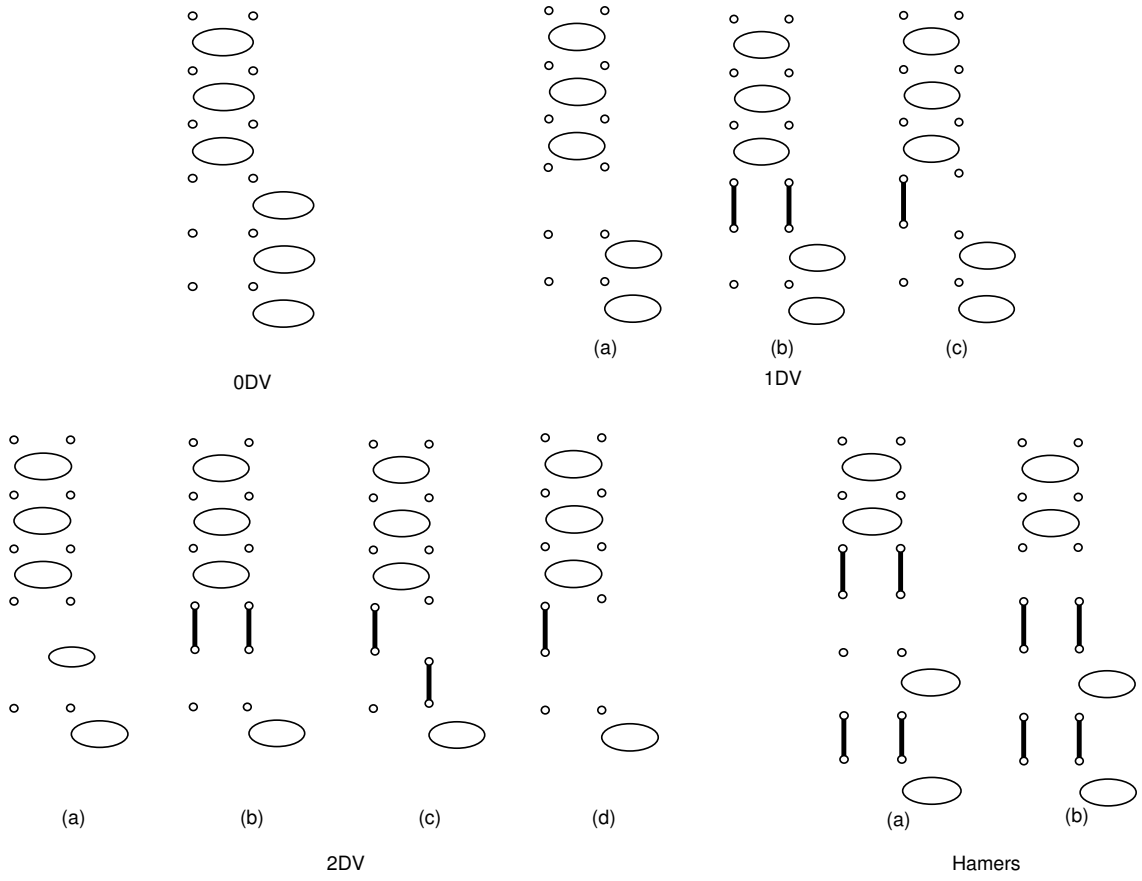

Figure 7.28: A schematic diagram of all the APB structures modelled

with no dimers missing (0DV), one dimer missing (1DV), two dimers missing (2DV) and the experimentally observed (Hamers) are shown in Figure 7.28, and their energies relative to the most stable APB (the 0DV) are given in Table 7.3. These energies were found using TB, with unit cells sixteen dimers long and one dimer wide, terminated in the usual manner.

Type	Energy (eV)
0DV	0.1622
1DVa	1.8137
1DVb	0.4798
1DVc	1.4798
2DVa	1.8961
2DVb	1.0400
2DVc	1.8404
2DVd	2.2410
HamersDV	1.1522
HamersDV(b)	2.6611

Table 7.3: Energies for different types of anti-phase boundary. Type 0 has no rows of atoms missing, while types 1 and 2 have 1 and 2 rows missing respectively. The Hamers structure is seen in Hamers *et al.* (1989). Schematics for the APBs can be seen in Fig. 7.28. The same correction as was applied to step energies in Sect. 4.3 for exposed areas with only four layers before the bulk-like termination has been applied.

From the table it is clear that the most stable structure for the B-type APB is the 0DV, followed by the 1DVb, the 2DVb and the Hamers structure. This requires some interpretation, for all APBs observed in STM either have at least one dimer missing, or have a small island nucleated on top of them. It may well

be that the 0DV B-type APB is hard to form (as it would be sterically difficult for dimers to reach the gap of a 1DV), and that once formed, it is a sufficiently good nucleation site that an island grows on top of it. Given that the experimental conditions in which the APBs are observed are at a temperature and disilane flux that promotes growth, this interpretation seems reasonable. The behaviour of a dimer string on top of an APB will be shown below.

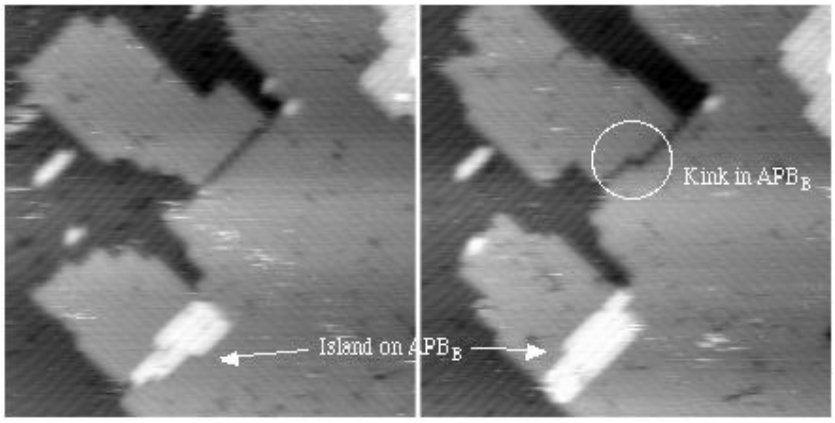

Figure 7.29: Two STM images taken a short time apart, showing some of the features associated with a B-type APB. (a) Indicated in the lower left hand corner is an APB with an island nucleated on top of it. (b) The APB in the middle of the image has formed a kink, as circled. *Image courtesy of James Owen.*

Figure 7.29 shows two STM images of the same area a short time apart. Both (a) and (b) show an APB in the lower left hand corner with an island nucleated on top of it. Fig. 7.29(b) shows a kink in an APB which has formed in the interval between the pictures. The kinking energy for a 0DV B-type APB has been modelled using TB in a unit cell ten dimer rows wide, with two APBs separated by five dimers each way. Forming a kink in the 0DV necessitates the removal of two atoms, but yields a kinking energy of 0.36 eV, which is not prohibitively high at 600 K. As all experimentally observed APBs are at least 1DVs, which will kink more easily (as no atoms must be removed, and there is more room for accomodation of strain), the structures found seem to be reasonable, and behave in a manner which does not contradict experimental results.

As mentioned above, the real interest in APBs is their ability to nucleate new islands of the second layer. Figure 7.30 shows the energy gained by adsorbing dimers onto the ends of strings of different lengths on a clean APB (a) and next to a string already on an APB (b). The energies were found with TB, using a unit cell ten dimers long and ten dimer rows wide, terminated in the usual manner. The differential energy is the important factor in dimer string growth. Fig. 7.30 (a) shows that the average energy gain for adding a dimer to a string on an APB is 0.1 eV higher than for adding to a string on a clean surface, and that the difference between adding to an even-numbered string and an odd-numbered string is much smaller than on the clean surface. On the clean surface, a string of an even-numbered length has one end unbonded and one rebonded, while an odd-numbered string has both ends rebonded and is more stable. On an APB there is no such effect, as the dimer rows either side are out of phase.

Indeed, it would seem that the energy for adding to both should be the same. However, as is clear from Fig. 7.30(a), the energy varies, and the reason why is illustrated in Figure 7.31. When there are an even number of dimers in the string, the twist forced by the underlying structure is the same at both ends of the string, and all the dimers twist in the same way. When there are an odd number of dimers in the string, however, the twist is in an opposite sense, and the central dimer is straight. This effect is enough to cause the slight variation of energy seen.

Fig. 7.30(b) shows the energy for adding a dimer to a string *next* to a string already on an APB. There is a small difference between this behaviour and that of adding to a string on a clean surface. The two results show why APBs are good nucleation sites for second layer islands, and indicate that a small amount of strain energy can have a large effect in growth processes.

Conclusions This chapter has presented the results for modelling all of the important processes involved in growth of the Si(001) surface from disilane, Si_2H_6. The pathway from the adsorbed fragments to clean

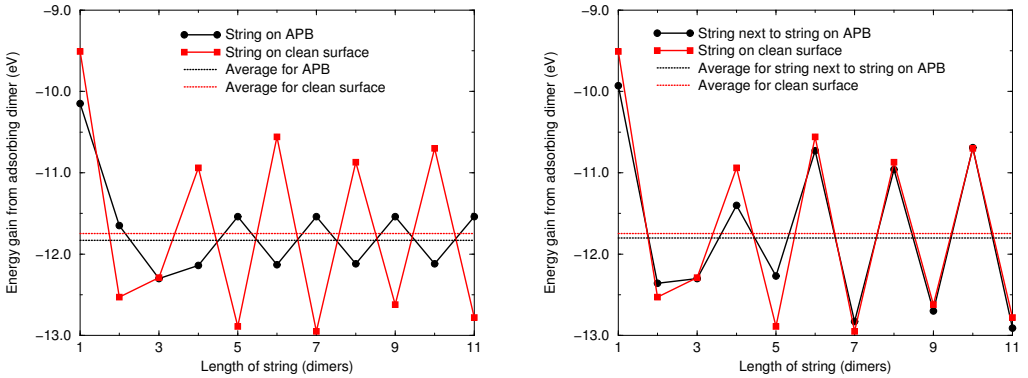

Figure 7.30: The energy for adsorbing a dimer onto a dimer string of different lengths, on a 0DV APB and a clean surface. (a) On a clean APB (b) Next to a string on an APB.

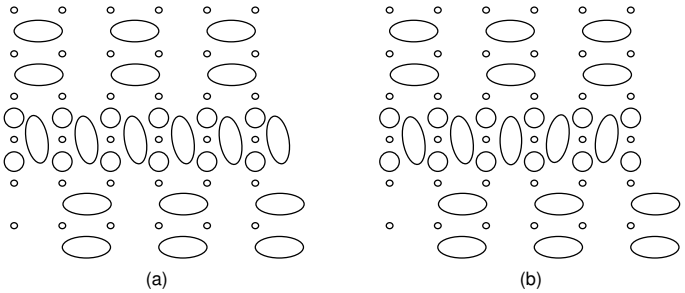

Figure 7.31: The structure of (a) an even-numbered length string and (b) an odd-numbered length string on an APB, showing the different twists forced on the strings.

dimers and small strings has been shown, and the nucleation of a second layer at high energy sites has been modelled. The results are in good agreement with experiment, and mean that, along with the experiments of Owen (1996), the atomistic aspects of the growth process on this surface have been comprehensively understood.

Chapter 8

Conclusion

Were I to await perfection, my book would never be finished

History of Chinese writing, Tai T'ung, 13th Century

This thesis has described computational modelling of various reactions involved in the growth of the Si(001) surface from disilane: defects and steps on the surface; the behaviour of bismuth (a surfactant) and two of the structures formed by it on the surface after annealing; the behaviour of hydrogen (which can act as a diffusion blocker and a surfactant, at different temperatures) at low and high coverages; and the specific reactions of disilane and its adsorption fragments in the growth process from adsorption to dimer strings.

Over the course of the work, I have developed two key themes: for many of the calculations, the tight binding approximation has provided extremely accurate and useful predictions; and without collaboration and interaction with experiments much of the work would have been impossible. I shall expound these two, and point the way to the future in the remainder of this conclusion.

8.1 Tight Binding is Good Enough

The tight binding approximation described in Chapter 2 takes some liberties with the quantum mechanical framework of bonding in solids, though the approximations made are, in the main, justifiable. The form used in this thesis is the crudest possible (only the nearest atoms are allowed to bond to each other, and the basis functions are assumed to be orthogonal) but it has vindicated itself admirably. It has been extremely useful, and successful, throughout the thesis, and here I list the areas where it has been used:

1. Step energies and kinking behaviour. Good agreement with experiment has been found.
2. Long straight lines of bismuth. The structure has been identified, and a growth mode proposed, which agrees with available experimental observations.
3. A c(4x4) bismuth reconstruction. A structure has been proposed by careful comparison with many different experimental sources.
4. Hydrogen diffusion. The diffusion of a single atom on the clean surface has been reproduced in good agreement with experiment and DFT results. The diffusion mode on the saturated surface has been explained, the saturation of steps during growth confirmed, and the behaviour of a defect as a mirror described.
5. The dehydrogenation mechanism for an ad-dimer has been explored, and a barrier which seems reasonable found.
6. The formation of a square, along with its decomposition to form short dimer strings, have been shown. This is a key step in the growth of silicon from disilane.
7. The structure of anti-phase boundaries has been explored, and their behaviour as nucleation sites for silicon islands explained.

Through these simulations, the accuracy and transferability of the parameterisations developed herein have been proved. The approaches used for fitting should be extended to new areas where tight binding will be able to provide answers, particularly as interest in features on surfaces turns to larger length scales. The use of linear scaling tight binding has proved invaluable for exploring non-periodic features such as steps and defects.

There is a growing trend towards the meeting of tight binding and density functional theory. The use of so-called *ab initio* tight binding to obviate the need for fitting of parameters is extremely exciting, and will provide the ability to keep up with experiments much more easily, as simulations will not need to wait for months while the parameterisation is developed and validated.

The role which tight binding fills, in being a fast quantum mechanical method, will expand as simulations become larger and timescales longer. The links between length scales which are being emphasised in the MML will be strengthened as the energetics and barriers which tight binding has been shown to be able to predict will allow Monte Carlo methods, which use the barriers, to achieve simulations for timescales of up to seconds for microns of material.

8.2 Every theorist needs an experimentalist

When the Royal Oak pub (on the Woodstock Road) came up for sale a few years ago, several of us in the lab thought that the Department should bid for it. Over the last three years, many discussions have taken place there, often aided by beer and cider, and much of the most fruitful work in this thesis has stemmed from such conversations and the trust and respect built by them.

While this may seem a slightly frivolous idea, it is far from it. The level of interaction between James Owen and myself is, I am told, rare, and in our case has proved most fruitful, yielding two theses, at least seven papers and (most importantly) a deep understanding of the growth process of the Si(001) surface from disilane. This understanding has already provoked interest in two commercial growth centres (DRA, Malvern, and Hewlett-Packard, Palo Alto).

What this interaction actually produces covers several levels: on the day-to-day level, ideas can be examined from different perspectives; in the long term, a mutual trust and understanding of techniques result; and in specific investigations, discoveries which require both disciplines can be made (for instance the defect structure in Chapter 4 or the square in Chapter 7) and expectations overturned (for instance, the formation of monohydride dimers discussed in Chapter 7, where the structure formed is at least 1 eV less stable than the most stable structure, but has a formation pathway which makes it far more likely to be formed). Certainly this thesis would have been much poorer without the STM experiments going on in the Department, and I hope to continue the interaction in the future. Without experimental results to anchor the models, theoretical calculations can become meaningless; equivalently, without theoretical calculations to explain and interpret the experimental results, experiments can become meaningless.

8.3 Now we can do germanium

As electronics becomes more and more a part of our society, the demands and expectations made on circuits become higher. At present, all high frequency applications (above 2 GHz or so) require the use of transistors and circuits fabricated on GaAs (or other expensive III-V materials) , which increases the cost and difficulty of manufacture. One of the growth areas of research is SiGe alloys, and Ge structures on Si substrates, which offer the possibility of high frequency applications and tunable electronic properties (through different layered structures), all within the same growth technology as is used for silicon.

Now that the silicon growth process has been successfully investigated, germanium can take center stage. As well as the desirable properties which Ge has for manufacturers, there are problems in growth which are both scientifically intriguing and technologically frustrating. Firstly, germanium tends to segregate to the surface during growth, though this problem is being addressed through the use of surfactants such as bismuth and hydrogen as described in Chapter 5. Secondly, germanium is a larger atom than silicon, and this introduces strain into the system; the ways in which this strain is relieved are many and varied. At low concentrations, it is dissipated in defects, though at larger concentrations and thicknesses macroscopic effects appear: hut pits and clusters, ripples and slip planes. Hut clusters are possibly the most spectacular of these, and an example of them is shown in Figure 8.1.

The formation and growth of hut clusters, and the distribution of strain around their base, are fascinating problems, as is the peculiar propensity for them to have (501)-facets as their sides. I shall be investigating

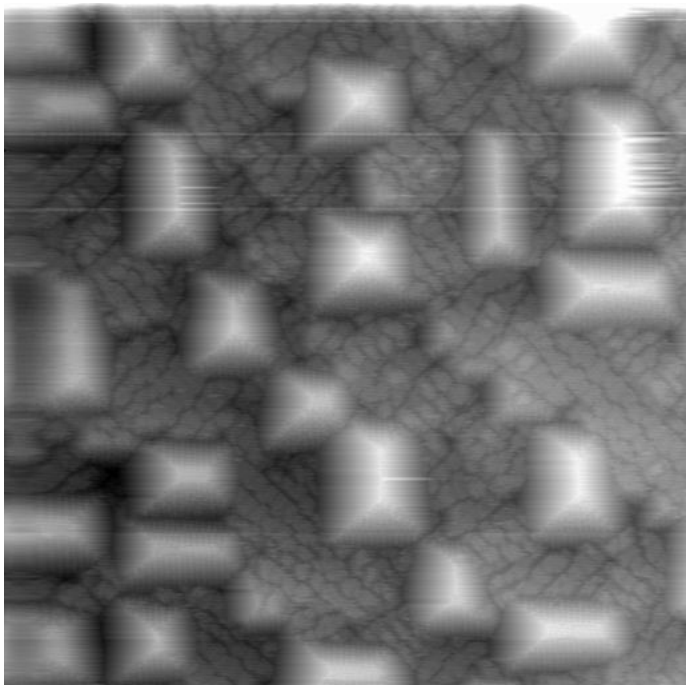

Figure 8.1: An STM image of "hut" clusters formed when more than about 6 layers of germanium are grown on silicon. The sides run along (100)-like directions (which are elastically soft), and the faces are all (501)-like. *Image courtesy of Ilan Goldfarb*

these questions using a linear-scaling density functional code, which is similar in many ways to the density matrix code used in this thesis. To model a hut cluster in LDA properly will the require the ability to simulate several thousand atoms, which should be achieved sometime in the next year and a half; this simulation will be the logical conclusion of many of the themes of this thesis.

References

Lesser artists borrow, great artists steal.

Igor Stravinsky

Chapter 1

Battaile C.C., Srolovitz D.J. and Butler J.E. (1997) Submitted to *J. Appl. Phys.*.
Binnig G., Rohrer H., Gerber Ch. and Weibel E. (1982) *Appl. Phys. Lett.* **40**, 178.
Christian J.W. (1965) *The Theory of Transformations in Metals and Alloys* (Pergamon, Oxford).
Clarke A.R.H., Pethica J.B., Nieminen J.A., Besenbacher F., Laegsgaard E. and Stensgaard I. (1996) *Phys. Rev. Lett.* **76**, 1276; Clarke A.R.H., D.Phil Thesis (Oxford University).
Fisher A.J., Blöchl P.E. and Briggs G.A.D. (1997) *Surf. Sci.* **374**, 298.
Gates S.M. (1988) *Surf. Sci.* **195**, 307.
Glasstone S., Laidler K.J. and Eyring H. (1941) *The Theory of Rate Processes* (McGraw-Hill, New York).
Many A., Goldstein Y. and Grover N.B. (1965) *Semiconductor Surfaces* (North and Holland: Amsterdam).
Ness H. and Fisher A.J. (1997) *Phys. Rev. B* **55**, 10081.
Owen J.H.G. (1996) *The Adsorption, Diffusion and Reaction of Small Adsorbates on the Si(001) surface*, D.Phil Thesis (Oxford University).
Tersoff J. and Hamann D.R. (1983) *Phys. Rev. Lett.* **50**, 1998; *Phys. Rev. B* **31**, 805.
Vineyard, G.H. (1957) J. Phys. Chem. Solids **3**, 121.
Zangwill, A. (1988) *Physics at Surfaces* (CUP:Cambridge).

Chapter 2

Aoki M. (1993) *Phys. Rev. Lett.* **71**,3842.
Aoki M. and Pettifor D.G. (1993) in *Physics of Transition Metals*, edited by P.M.Oppeneer and J. Kübler (World Scientific, Singapore).
Bachelet G.B., Hamann D.R. and Schlüter M. (1982) *Phys. Rev. B* **26**, 4199.
Bates S.P., Kresse G. and Gillan M.J. (1997) In preparation.
Batista V. and Coker D.F. (1996) *J. Chem. Phys.* **105**, 4033.
Becke A.D. (1988) *Phys. Rev. A* **38**, 3098.
Beer N.and Pettifor D.G. (1984) *Structure and Phase Stability of Alloys*, edited by P.Phariseau and W.Temmerman (New York:Plenum Press).
Bigger J.R.K., McInnes D.A., Sutton A.P., Payne M.C., Stich I., King-Smith R.D., Bird D.M. and Clarke L.J. (1992) *Phys. Rev. Lett.* **69**, 2224.
Bird D.M., Clarke L.J., King-Smith R.D., Payne M.C., Stich I. and Sutton A.P. (1992) *Phys. Rev. Lett.* **69**, 3785.
Bird D.M., Clarke L.J., Payne M.C. and Stich I. (1993) *Chem. Phys. Lett.* **212**, 518.
Bird D.M. and Gravil P.A. (1997) *Surf. Sci.* **377**, 555.
Born M. and Oppenheimer J.R. (1927) *Am. J. Phys.* **84**, 451.
Bowler D.R., Aoki M., Goringe C.M., Horsfield A.P. and Pettifor D.G. (1997) *Modelling Simul. Mat. Sci. Eng.* **5**, 199.
Ceperly D.M. and Alder B.J. (1980) *Phys. Rev. Lett.* **45**, 566.
Cohen M.L. and Heine V. (1970) *Solid State Physics* **24**, 37.
Coker D.F. and Xiao L. (1995) *J. Chem. Phys.* **102**, 496.
Daw M.S. (1993) *Phys. Rev. B* **47**,10895.
DeVita A., Gillan M.J., Lin J.S., Payne M.C., Stich I. and Clarke L.J. (1992) *Phys. Rev. Lett.* **68**, 3319.
DeVita A., Stich I., Gillan M.J., Payne M.C. and Clarke L.J. (1993) *Phys. Rev. Lett.* **71**, 1276.
Dirac P.A.M. (1930) *Proc. Cambridge Philos. Soc.***26**,376.
Ducastelle F. and Cyrot-Lackmann F. (1970) *J. Phys. Chem. Solids* **31**, 1295.
Feynman R.P. (1939) *Phys. Rev.* **56**, 340.
Foiles S.M. (1993) *Phys. Rev. B* **48**, 4287.
Foulkes W.M.C. (1987) Ph.D. Thesis (University of Cambridge).
Friedel J. (1954) *Adv. Phys.* **3**, 446.
Galli G. and Parrinello M. (1992) *Phys. Rev. Lett.* **69**,3547.
Gillan M.J. (1989) *J. Phys.:Condens. Matter* **1** 689.
Goedecker S. and Colombo L. (1994) *Phys. Rev. Lett.* **73**,122.
Goedecker S. and Teter M. (1995) *Phys. Rev. B* **51**, 9455.
Glanville S., Paxton A.T. and Finnis M.W. (1988) *J.Phys F* **18**, 693.
Goringe C.M. (1995) D.Phil Thesis, Oxford University; Goringe C.M. (1997) To be submitted to *Modelling Simul. Mat. Sci. Eng.*.

Goringe C.M., Bowler D.R. and Hernández E.H. (1997) *Rep. Prog. Phys., in press.*
Goringe C.M., Clark L.J., Lee M.H., Payne M.C., Stich I., White J.A., Gillan M.J. and Sutton A.P. (1997) *J. Phys. Chem. B* **101**, 1498.
Hamann D.R., Schlüter M. and Chiang C. (1979) *Phys. Rev. Lett.* **43**, 1494.
Hammer B., Jakobsen K.W. and Nørskov J.K. (1993) *Phys. Rev. Lett.* **70**, 3971.
Harris J. (1985) *Phys. Rev. B* **31**, 1770.
Haydock R. (1980) *Sol. St. Phys.* **35**, 216.
Head-Gordon M. and Tully J.C. (1995) *J. Chem. Phys.* **103**, 10137.
Heine V. (1980) *Sol. St. Phys.* **35**, 1.
Hellmann H. (1937) *Einfuhrung in die Quantumchemie* (Leipzig: Franz Deutsche 1937).
Hohenberg P. and Kohn W. (1964) *Phys. Rev.* **136**, B864).
Horsfield A.P. (1996) *Mat. Sci. Engin. B* **37**, 219.
Horsfield A.P., Bratkovsky A.M., Fearn M., Pettifor D.G. and Aoki M. (1996) *Phys. Rev. B* **53**, 12694.
Inoue J. and Ohta Y. (1987) *J.Phys.C* **20**, 1947.
Joannopoulos J.D., Starkloff T. and Kastner M.A. (1977) *Phys. Rev. Lett.* **38**, 660.
Jones R.O. and Gunnarsson O. (1989) *Rev. Mod. Phys.* **61**, 689.
Jones R. and Lewis M.W. (1984) *Philos. Mag. B* **49**, 95.
Jónsson H. and Mills G. (1997) Submitted to *J. Chem. Phys.*.
Kay M., Darling G.R., Holloway S., White J.A. and Bird D.M. (1995) *Chem. Phys. Lett.* **245**, 311.
Kerker G. (1980) *J. Phys. C* **13**, 189.
Kleinman L. and Bylander D.M. (1982) *Phys. Rev. Lett.* **48**, 1425.
Kohn W. and Sham L.J. (1965) *Phys. Rev.* **140**, A1133.
Kohn W. (1959) *Phys. Rev.* **115**, 809.
Kress J.D. and Voter A.F. (1995) *Phys. Rev. B* **52**,8766 (1996).
Lanczos C. (1950) *J. Res. Natl. Bur. Stand.* **45**,225.
Lee C., Yang W. and Parr R.G. (1988) *Phys. Rev. B* **37**, 785.
Levy M. (1982) *Phys. Rev. A* **26**, 1200).
Li X.-P.,Nunes R.W. and Vanderbilt D. (1993) *Phys. Rev. B* **47**,10891.
Lieb E.H. (1983) *Int. J. Quantum. Chem.* **24**, 243.
McWeeney R. (1960) *Rev. Mod. Phys.* **32**,335.
Mauri F.,Galli G. and Car R. (1993) *Phys. Rev. B* **47**,9973.
Mills G., Jónsson H. and Schenter G.K. (1995) *Surf. Sci.* **324**, 305.
Monkhorst H.J. and Pack J.D. (1976) *Phys. Rev. B* **13**, 5188.
Nachtigall P., Jordan K.D., Smith A. and Jónsson H. (1996) *J. Chem. Phys.* **104**, 148.
Ordejon P.,Drabold D.A.,Grumbach M.P. and Martin R.M. (1993) *Phys. Rev. B* **48**,14646.
Parr R.G. and Yang W. (1989) *Density Functional Theory of Atoms and Molecules* (OUP:Oxford).
Paxton A.T., Sutton A.P. and Nex C.M.M. (1987) *J.Phys C* **20**, L263.
Paxton A.T. and Sutton A.P. (1989) *Acta Metall.* **37**, 1693.
Payne M.C., Teter M.P., Allan D.C., Arias T.A. and Joannopoulos J.D. (1992) *Rev. Mod. Phys.***64**, 1045.
Perdew J.P. (1985) *Phys. Rev. Lett.* **55**,1665.
Perdew J.P. and Yue W. (1986) *Phys. Rev. B* **33**, 8800.
Perdew J.P. and Yang W. (1991) Unpublished, and Perdew J.P. (1991) in *Electronic Structure of Solids '91*, ed. P.Ziesche and H.Eschrig (Akadamie Verlag, Berlin, 1991).
Pettifor D.G. (1977) *J.Phys. F:Met. Phys.* **7**, 613.
Pettifor D.G. (1989) *Phys. Rev. Lett.* **63**,2480.
Pettifor D.G. and Aoki M. (1991) *Philos. Trans. R. Soc. London A* **334**, 439.
Pugh S. and Gillan M.J. (1994) *Surf. Sci.* **320**, 331.
Redondo A., Goddard W.A. III and McGill T.C. (1977) *Phys. Rev. B* **15**, 5038.
Slater J.C. and Koster G.F. (1954) *Phys. Rev.* **94**, 1498.
Starkloff T. and Joannopoulos J.D. (1977) *Phys. Rev. B* **16**, 5212.
Stechel E.B., Williams A.R. and Feibelman P.J. (1994) *Phys. Rev. B* **49**,10088.
Stich I., Payne M.C., King-Smith R.D., Lin J.S. and Clarke L.J. (1992) *Phys. Rev. Lett.* **68**, 1351.
Stich I., Payne M.C., DeVita A., Gillan M.J. and Clarke L.J. (1993) *Chem. Phys. Lett.* **212**, 617.
Sutton A.P., Finnis M.W., Pettifor D.G. and Ohta Y. (1988) *J. Phys. C***21**, 35.
Vanderbilt D. (1990) *Phys. Rev. B* **41**, 7892.
Voter A.F., Kress J.D. and Silver R.N. (1996) *Phys. Rev. B* **53**, 12733.
White J.A. and Bird D.M. (1993) *Chem. Phys. Lett.* **213**, 422.
White J.A., Bird D.M., Payne M.C. and Stich I. (1994) *Phys. Rev. Lett.* **73**, 1404.
White J.A., Bird D.M. and Payne M.C. (1996) *Phys. Rev. B* **53**, 1667.
Xiao L. and Coker D.F. (1995) *J. Chem. Phys.* **102**, 1107.
Yin M.T. and Cohen M.L. (1982) *Phys. Rev. B* **25**, 7403.
Zunger A. and Cohen M.L. (1979) *Phys. Rev. B* **20**, 4082.
Chapter 3

Bowler D.R., Aoki M., Goringe C.M., Horsfield A.P. and Pettifor D.G. (1997) *Modelling Simul. Mat. Sci. Eng.* **5**, 199.
Chadi D.J. (1977) *Phys. Rev. B* **16**, 790.
Chadi D.J. (1979a) *Phys. Rev. B* **19**, 2074.
Chadi D.J. (1979b) *Phys. Rev. Lett.* **43**,43.
Chadi D.J. (1979c) *J. Vac. Sci. Technol.***16**, 1290.
Chadi D.J. and Cohen M.L. (1975) *phys. stat. solidi* **68**, 405.
Chadi D.J. and Martin R.M. (1976) *Solid State Commun.* **19**, 643.
Froyen S. and Harrison W.A. (1979) *Phys. Rev. B* **20**, 2420.

Goodwin L.,Skinner D.R. and Pettifor D.G. (1989) *Europhys. Lett.* **9**, 701.
Goringe C.M. (1995) D.Phil Thesis, Oxford University; Goringe C.M. (1997) To be submitted to *Modelling Simul. Mat. Sci. Eng.*.
Harrison W.A. (1980) *Electronic Structure and the Properties of Solids* (W.H.Freeman and Co., San Francisco).
Horsfield A.P., Godwin P.D., Pettifor D.G. and Sutton A.P. (1996) *Phys. Rev. B* **54**, 15773.
Kress J.D. and Voter A.F. (1995) *Phys. Rev. B* **52**,8766 (1996).
Kwon I., Biswas R., Wang C.Z., Ho K.M. and Soukoulis C.M. (1994) *Phys. Rev. B* **49**,7242.
Min B.J., Lee Y.H., Wang C.Z., Chan C.T. and Ho K.M. (1992) *Phys. Rev. B* **45**, 6839.
Panzarini G. and Colombo L. (1994) *Phys. Rev. Lett.* **73**, 1636.
Paxton A.T., Sutton A.P. and Nex C.M.M. (1987) *J. Phys. C* **20**, L263.
Sankey O.F. and Niklewski D.J. (1989) *Phys. Rev. B* **40**, 3979.
Slater J.C. and Koster G.F. (1954) *Phys. Rev.* **94**, 1498.
Vogl P., Hjalmarson H.P. and Dow J.D. (1983) *J. Phys. Chem. Solids* **44**, 365.
Chapter 4

Andersohn L. and Köhler U. (1993) *Surf. Sci.* **284**, 77.
Bartelt N.C., Tromp R.M. and Williams E.D. (1994) *Phys. Rev. Lett.* **73**, 1656.
Bogusławski P., Zhang Q.-M., Z.Zhang and Bernholc J. (1994) *Phys. Rev. Lett.* **72**, 3694.
Carter L.E. and Carter E.A. (1995) *Surf. Sci.* **323**, 39-50.
Chadi D.J. (1979) *J. Vac. Sci. Technol.***16**, 1290.
Chadi D.J. (1987) *Phys. Rev. Lett.* **59**, 1691.
Dabrowski J. and Scheffler M. (1992) *Appl. Surf. Sci.* **56-58**,15.
Hamers R.J. and Köhler U.K. (1990) *J.Vac. Sci. Tech. A* **7**, 2854.
Hamers R.J., Tromp R.M. and Demuth J.E. (1986) *Phys. Rev. B* **34**, 5343.
Harrison W.A. (1980) *Electronic Structure and the Properties of Solids* (W.H.Freeman and Co., San Francisco).
Ihara S., Ho S.L., Uda T. and Hirao M. (1990) *Phys. Rev. Lett.* **65**, 1909.
Kitamura N., Swartzentruber B.S., Lagally M.G. and Webb M.B. (1993) *Phys. Rev. B* **48**, 5704.
Northrup J.E. (1985) *Phys. Rev. Lett.* **54**, 815.
Owen J.H.G., Bowler D.R., Goringe C.M., Miki K. and Briggs G.A.D. (1995) *Surf. Sci. Lett.* **341**, L1042.
Pandey K.C. (1985) in *Proceedings of the Seventeenth International Conference on the Physics of Semiconductors*, eds D.J.Chadi and W.A.Harrison, (Springer-Verlag, NY), p55.
Pearson C., Borovsky B., Krueger M., Curtis R. and Ganz E. (1995a) *Phys. Rev. Lett.* **74**, 2710.
Pearson C., Krueger M., Curtis R., Borovsky B., Shi X. and Ganz E. (1995b) *J. Vac. Sci. Technol. A***13**, 1506.
Poppendieck T.D., Ngoc T.C. and Webb M.B. (1978) *Surf. Sci.* **75**, 287.
Ramstad A., Brocks G. and Kelly P.J. (1995) *Phys. Rev. B* **51**, 14504.
Roberts N. and Needs R.J. (1990) *Surf. Sci.* **236**, 112.
Roland C. and Gilmer G.H. (1993) *Phys. Rev. B* **47**, 16286.
Schlier R.E. and Farnsworth H.E. (1959) *J. Chem. Phys.* **30**, 917.
Swartzentruber B.S., Mo Y.-W., Kariotis R., Lagally M.G. and Webb M.B. (1990) *Phys. Rev. Lett.* **65**, 1913.
Swartzentruber B.S., Kitamura N., Lagally M.G. and Webb M.B. (1993) *Phys. Rev. B* **47**, 13432.
Swartzentruber B.S. and Schacht M. (1995) *Surf. Sci.* **322**, 83.
Tromp R.M., Hamers R.J. and Demuth J.E. (1985) *Phys. Rev. Lett.* **55**, 1303.
Wang J., Arias T.A. and Joannopoulos J.D. (1993) *Phys. Rev. B* **47**, 10497.
Wolkow R.A. (1992) *Phys. Rev. Lett.* **68**, 2636.
Wu C.J., Ionova I.V. and Carter E.A. (1993) *Surf. Sci.* **295**, 64.
Wu F., Chen X., Zhang Z. and Lagally M.G. (1995) *Phys. Rev. Lett.* **74**, 574.
Zandvliet H.J.W., Elswijk H.B., van Loenen E.J. and Dijkkamp D. (1992) *Phys. Rev. B* **45**, 5965.
Zandvliet H.J.W., Elswijk H.B. and van Loenen E.J. (1992) *Surf. Sci.* **272**, 264.
Chapter 5
Avery A.R., Goringe C.M., Holmes D.M., Sudijono J.L. and Jones T.S. (1996) *Phys. Rev. Lett.* **76**, 3344.
Copel M., Reuter C., Kaxiras E. and Tromp R.M. (1989) *Phys. Rev. Lett.* **63**, 632.
Franklin G.E., Tang S., Woicik J.C., Bedzyk M.J., Freeman A.J. and Golovchenko J.A. (1995) *Phys. Rev. B* **52**, R5515.
Fujita K., Fukatsu S., Yaguchi H., Igarashi T., Shiraki Y. and Ito R. (1990) *Jpn. J. Appl. Phys.* **29**, L1981,
Fukatsu S., Usami N. and Shiraki Y. (1993) *Jpn. J. Appl. Phys.* **32**, 1502.
Gavioli L., Betti M.G. and Mariani C. (1997) *Surf. Sci.* **377-379**. 215.
Hanada T. and Kawai M. (1991) *Surf. Sci.* **242**, 137.
Katayama M., Nakayama T., Aono M. and McConville C.F. (1996) *Phys. Rev. B* **54**, 8600.
Kim K.-S., Takakuwa Y., Mori Y. and Kono S. (1996) *Jpn. J. Appl. Phys. (Part 2)* **35**, L1695.
Kitamura N., Lagally M.G. and Webb M.B. (1993) *Phys. Rev. Lett.* **71**, 2802.
Lyman P.F., Qian Y., Lee T.-L. and Bedzyk M.J. (1996) *Physica B* **221**, 426.
Miki K., Owen J.H.G., Bowler D.R., Briggs G.A.D. and Sakamoto K. (1997) In preparation.
Naitoh M., Shimaya H., Nishigaki S., Oishi N. and Shoji F. (1997) *Surf. Sci.* **377-379**, 899.
Noh H.P., Park C., Jeon D., Cho K., Hashizume T., Kuk Y. and Sakurai T. (1994) *J. Vac. Sci. Technol. B* **12**, 2097.
Nörenberg H., Bowler D.R. and Briggs G.A.D. (1997) Submitted to *Appl. Surf. Sci.*
Ohtani N., Mokler S., Xie M.H., Zhang J. and Joyce B.A. (1994) *Jpn. J. Appl. Phys. (Part 1)* **33**, 2311.
Park C., Bakhtizin R.Z., Hashizume T. and Sakurai T. (1994) *J. Vac. Sci. Technol. B* **12**, 2049.
Qian Y.L., Bedzyk M.J., Lyman P.F., Lee T.L., Tang S.P. and Freeman A.J. (1996) *Phys. Rev. B* **54**, 4424.
Sakamoto K., Matsuhata H., Kyoya K., Miki K. and Sakamoto T. (1993) *Jpn. J. Appl. Phys. (Part 2)* **32**, L204.
Sakamoto K., Matsuhata H., Kyoya K., Miki K. and Sakamoto T. (1994) *Jpn. J. Appl. Phys. (Part 1)* **33**, 2307.
Tang S. and Freeman A.J. (1994) *Phys. Rev. B* **50**, 1701.
Uhrberg R.I.G., Northrup J.E., Biegelsen D.K., Bringans R.D. and Swartz L.-E. (1992) *Phys. Rev. B* **46**, 10251.
Wang Y., Hamers R.J. and Kaxiras E. (1995) *Phys. Rev. Lett.* **74**, 403.

Watanabe H. and Ichikawa M. (1997) *Phys. Rev. B* **55**, 9699.
Yamasaki T., Uda T. and Terakura K. (1996) *Phys. Rev. Lett.* **76**, 2949.
Zhang Z., Kulakov M.A. and Bullemer B. (1996) *Surf. Sci.* **369**, 69.
Chapter 6

Blöchl P.E., Van de Walle C.G. and Pantelides S.E. (1990) *Phys. Rev. Lett.* **64**, 1401.
Boland J.J. (1991a) *Phys. Rev. B* **44**, 1383.
Boland J.J. (1991b) *Phys. Rev. Lett.* **67**, 1539.
Boland J.J. (1992) *J. Vac. Sci. Technol. A* **10**, 2458.
Boland J.J. (1993) *Advances in Physics* **42**, 129.
Eaglesham D.J., Unterwald F.C., Luftman H., Adams D.P. and Yalisove S.M. (1993) *J.Appl.Phys.* **74**, 6615.
Mokler S.M., Ohtani N., Xie M.H., Zhang J. and Joyce B.A. (1993) *J.Vac. Sci. Tech. B* **11**, 1073.
Owen J.H.G., Bowler D.R., Goringe C.M., Miki K. and Briggs G.A.D. (1996) *Phys. Rev. B* **54**, 14153.
Owen J.H.G., Bowler D.R., Miki K. and Briggs G.A.D. (1997) Submitted to *Phys. Rev. Lett.* .
Owen J.H.G., Miki K., Bowler D.R., Goringe C.M., Goldfarb I. and Briggs G.A.D. (1997a) *Surf. Sci.* In press.
Owen J.H.G., Miki K., Bowler D.R., Goringe C.M., Goldfarb I. and Briggs G.A.D. (1997b) *Surf. Sci.* In press.
Sinniah K., Sherman M.G., Lewis L.B., Weinberg W.H., Yates, J.T. Jr. and Janda K.C. (1989) *Phys. Rev. Lett.* **62**, 567.
Smith A.P. and Jónsson H. (1996) *Phys. Rev. Lett.* **77**, 1326.
Stucki F., Schaefer J.A., Anderson J.R., Lapeyre G.J. and Gopel W. (1983) Solid State Commun. **47**, 795.
Vittadini A., Selloni A. and Casarin M. (1993) *Surf. Sci. Lett.* **289**, L625.
Widdra W., Yi S.I., Maboudian R., Briggs G.A.D. and Weinberg W.H. (1995) Phys. Rev. Lett. **74**, 2074.
Wu Y.M. and Nix R.M. (1994) *Surf. Sci.* **306**, 59.
Wu C.J. and Carter E.A. (1991) *Chem. Phys. Lett.* **185**, 172.
Wu C.J. and Carter E.A. (1992) *Phys. Rev. B* **46**, 4651.
Wu C.J., Ionova I.I. and Carter E.A. (1994) *Phys. Rev. B* **49**, 13488.
Chapter 7
Bedrossian P.J. (1995) *Phys. Rev. Lett.* **74**, 3648.
Boland J.J. (1991) *Phys. Rev. B* **44**, 1383.
Borovsky B., Krueger M. and Ganz E. (1997) *Phys. Rev. Lett.* **78**, 4229.
Bowler D.R. and Goringe C.M. (1996) *Surf. Sci. Lett.* **360**, 489.
Brocks G. and Kelly P.J. (1996) *Phys. Rev. Lett.* **76**, 2362.
Bronikowski M.J., Wang Y. and Hamers R.J. (1993) *Phys. Rev. B* **48**, 12361.
Bronikowski M.J., Wang Y., McEllistrem M.T., Chen D. and Hamers R.J. (1993) *Surf. Sci.* **298**, 50.
Fisher A.J., Blöchl P.E. and Briggs G.A.D. (1997) *Surf. Sci.* **374**, 298.
Gates S.M. (1988) *Surf. Sci.* **195**, 307.
Gates S.M. (1992) *J. Phys. Chem.* **96**, 10439.
Gates S.M., Greenlief C.M. and Beach D.B. (1990) *J. Chem. Phys.* **93**, 7493.
Goringe C.M. and Bowler D.R. (1997) *Phys. Rev. B , in press.*
Hamers R.J., Köhler U.K. and Demuth J.E. (1989) *Ultramicroscopy* **31**, 10.
Horsfield A.P. and Clancy P. (1994) *Modelling Simul. Mat. Sci. Eng.* **2**, 277.
Kolasinski K.W., Shane S.F. and Nare R.N. (1992) *J. Chem. Phys.* **96**, 3995.
Lin D.S., Hirschorn E.S., Chiang T.C., Tsu R., Luben D. and Greene J.E. (1992) *Phys. Rev. B* **45**, 3494.
Liu W.K., Mokler S.M., Ohtani N., Roberts C. and Joyce B.A. (1992) *Surf. Sci.* **264**, 301.
Mayne A.J., Goringe C.M., Smith C.W. and Briggs G.A.D. (1996) *Surf. Sci.* **348**, 209.
Mokler S.M., Liu W.K., Ohtani N.and Joyce B.A. (1992) *J. Cryst. Growth* **120**, 290.
Ohtani N., Mokler S., Xie M.H., Zhang J. and Joyce B.A. (1994) *Jpn. J. Appl. Phys. (Part 1)* **33**, 2311.
Owen J.H.G. (1996) *The Adsorption, Diffusion and Reaction of Small Adsorbates on the Si(001) surface*, D.Phil Thesis (Oxford University).
Owen J.H.G., Bowler D.R., Goringe C.M., Miki K. and Briggs G.A.D. (1997) *Surf. Sci. Lett.* In press.
Owen J.H.G., Miki K., Bowler D.R., Goringe C.M., Goldfarb I. and Briggs G.A.D. (1997a) *Surf. Sci.* In press.
Owen J.H.G., Miki K., Bowler D.R., Goringe C.M., Goldfarb I. and Briggs G.A.D. (1997b) *Surf. Sci.* In press.
Pearson C., Krueger M., Ganz E. (1996) *Phys. Rev. Lett.* **76**, 2306.
Swartzentruber B.S. (1996) *Phys. Rev. Lett.* **76**, 459.
Swartzentruber B.S., Smith A.P. and Jónsson H. (1996) *Phys. Rev. Lett.* **77**, 2518.
van Dam A., van Wingerden J., Haye M.J., Scolte P.M.L.O. and Tuinstra F. (1996) *Phys. Rev. B* **54**, 1557.
Wang Y., Bronikowski M.J. and Hamers R.J. (1994) *Surf. Sci.* **311**, 64.
Yamasaki T., Uda T. and Terakura K. (1996) *Phys. Rev. Lett.* **76**, 2949.