A FIRST PRINCIPLES STUDY OF LIGHT IMPURITIES IN SEMICONDUCTORS

Submitted by Paul William Leary to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Physics in the Faculty of Science,

September, 1997

I certify that all material in this thesis which is not my own work has been identified and that no material is included for which a degree has previously been conferred upon me.

Abstract

Presented in this thesis are theoretical investigations into light impurity defects (H and C) in both silicon and germanium. The calculations are performed using the *AIMPRO* code, an *ab initio* local density method applied to large hydrogen terminated clusters.

The silicon self-interstitial (Si_i) is a very important defect in Si, since it mediates dopant diffusion, and we first focus on this defect. Several low energy configurations are found, and the defect is bistable, switching configuration with charge state. The migration of the defect is discussed both in terms of the metastable configurations, and effects of varying charge state. The effects of hydrogen interaction with Si_i are investigated, and the calculations on the IH_2 defect agree well with experimental observations. Preliminary calculations on Si_i aggregates are also discussed.

Carbon defects in silicon are then investigated, and here, several unexplained properties of various centres are addressed. Calculations are performed on the C_s, C_i, and C_i-C_s defects, and our results support the proposed structural models for these centres. The calculated local modes of all of these defects are in good agreement with experiment, and the energies of the later defect in various charge states are consistent with the bistable properties of the centre.

The effects of hydrogen interactions with carbon defects (both C_s and C_i) are then investigated, and a number of new results are found. The results on the C_s-H defect are consistent with experiment, and again, there is evidence for bi-stability of the centre. The C_i-H defect has a low migration energy, and therefore we expect it to be a very reactive species. The interaction of this centre with other impurities, in particular C_s is discussed. The results lead to an identification of the structure of the T-centre defect as C_s-C_iH; comments are also made on several other CH related centres observed by photoluminescence.

Finally, the interaction of intrinsic defects with H, and the C_s defect are investigated in germanium. Parallels are drawn between these results and those in Si, and the thesis ends with preliminary results on the effects of C_s in Si/Ge alloys.

Acknowledgements

I would like to thank both the University of Exeter Department of Physics and the EPSRC for providing the necessary resources and funding for the duration of this PhD.

There are numerous people that I would like to thank, both academically and socially for enriching my life over the last three years.

Firstly, my supervisor Dr. Bob Jones for his guidance, enthusiasm and making all of this possible. Big thanks are also due to the rest of the AIMPRO offspring, starting with Drs. Patrick Briddon, Steve Breuer, Jon Goss, Paul Sitch, Malcolm Heggie, and Chris Latham. Dr. Sven Öberg deserves a special mention for his constant collaboration throughout all of this work. Brothers Chris Ewels, António Resende, Ben Hourahine, and Joachim Elsner have also contributed much towards my understanding, and all of the above have been great friends as well as colleagues. Much of my work has involved collaboration with experimentalists, and I am indebted to the groups at Kings College London and the University of Aarhus, in particular Alexei Safonov, Brian Bech Nielsen, Michael Budde and Lone Hoffmann are thanked for many stimulating discussions.

As well as many of the above, a number of other people have forced me to go to the pub on one too many occasions, and generally made the three years at Exeter amongst the most enjoyable of my life. In particular, Paul Rawlins, Sai Patel, Andy Rundell, Will Stallard, Christopher Townsley, Kevin Lambert, Mike Greally, Mauro Boero, the Turks, Elaine Bloss, have all to certain extents modified my views on Southern drinking habits. My thanks and apologies to my many other friends in Exeter, Newcastle and Whitehaven who have not been specifically mentioned here but have equally contributed to my happiness at various times of my life.

For their constant support, encouragement, and being there I wish to thank all of my family - to whom I dedicate this thesis. The (sometimes hard) times taken to write this thesis have been made so much more bearable by Caroline Hoad, and I'm sure that I'll repay the debt in due course.

Publications list

[1] *Peculiarities of interstitial carbon and di-carbon defects in Si*, R. Jones, S. Öberg, P. Leary, and V. J. B. Torres, Materials Science Forum **196-201**, 785 (1995).

[2] *The di-carbon hydrogen defect in silicon*, A. N. Safonov, E. C. Lightowlers, G. Davies, A. Mainwood, P. Leary, R. Jones, and S. Öberg, Proceedings of the 23rd international conference on the physics of semiconductors, (World Press, Singapore), 2617 (1996).

[3] *Interstitial-carbon hydrogen interaction in silicon*, A. N. Safonov, E. C. Lightowlers, G. Davies, P. Leary, R. Jones, and S. Öberg, Physical Review Letters **77**, 4812 (1996).

[4] *Dynamic properties of interstitial carbon and carbon-carbon pair defects in silicon*, P. Leary, R. Jones, S. Öberg, and V. J. B. Torres, Physical Review B **55**, 2188 (1997).

[5] *Substitutional carbon in germanium*, L. Hoffmann, J. C. Bach, B. Bech Nielsen, P. Leary, R. Jones, and S. Öberg, Physical Review B **55**, 11167 (1997).

[6] *Interstitial carbon-hydrogen defects in silicon*, P. Leary, S. Öberg, P. R. Briddon, and R. Jones, Materials Science Forum (in press).

[7] *The hydrogen-saturated self-interstitial in silicon and germanium*, M. Budde, B. Bech Nielsen, P. Leary, J. Goss, R. Jones, P. R. Briddon, S. Öberg, and S. J. Breuer, Materials Science Forum (in press).

[8] *Substitutional carbon in Ge and* $Si_{(1-x)}Ge_x$, L. Hoffmann, J. C. Bach, J. Lundsgaard Hansen, A. Nylandsted Larsen, B. Bech Nielsen, P. Leary, R. Jones, and S. Öberg, Materials Science Forum (in press).

[9] *Identification of the hydrogen saturated self-interstitials in silicon and germanium*, M. Budde, B. Bech Nielsen, P. Leary, J. Goss, R. Jones, P. R. Briddon, S. Öberg, and S. J. Breuer, accepted for publication in Physical Review B.

[10] *Interaction of hydrogen with substitutional and interstitial carbon defects in silicon*, P. Leary, R. Jones, and S. Öberg, accepted for publication in Physical Review B.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Silicon based integrated circuits are behind the majority of the state of the art devices used in the electronics industry today. As a result, a phenomenal amount of research has been performed on silicon, making it one of the most well studied of all materials around. For purposes of device physics, the concentration of available carriers (both electrons and holes) must be increased to improve the conductance and speed of the device. For this purpose, silicon can be doped, for example, with either boron (acceptor) or phosphorus (donor) creating $p-$ and $n-$ type materials respectively. These defects are examples of where an impurity is intentionally added to the material to improve the performance of a device. Modern silicon chips are fabricated from alternating layers of $n-$ and $p-$ doped material. However the presence of other impurities or defects can be detrimental to device physics since their presence can affect the carrier density when they give rise to deep levels in the forbidden gap.

After the crystalline silicon is implanted with the dopants, a number of electrically active defects are formed, and these are subsequently removed by the thermal annealing of the sample. During the annealing process, impurities, and crucially, the dopants diffuse through the material. The effects of the impurity diffusion is beginning to have a greater effect on the Si electronics industry, due to the continuing efforts to further decrease the size of devices. Advances in the fabrication methods have led to the reduction in size to such an extent that the diffusion length of the dopants will soon limit further progress. Therefore the understanding of the diffusion mechanisms of these and other defects, and whether or not they can be suppressed is now a crucial question.

The diffusion of most impurities (including the dopants) is mediated by the intrinsic point defects: the vacancy V and more importantly the self-interstitial I. Diffusion of substitutional defects like the dopants (B_s and P_s) can only occur at very high temperatures, unless I defects are present. If an impurity X which occupies the substitutional site when grown into the material, is stable as an interstitial species, then the following reaction occurs when X traps an self-interstitial:

$$X_s + I \longrightarrow X_i,$$

even if an impurity Y is not stable as an interstitial species, repeated capture and emission of I can take place:

$$Y_s + I \longleftrightarrow Y_i,$$

and in both cases the interstitial species diffuses with a much lower migration energy $\leq$ 1.0 eV that the substitutional impurity $\geq$ 3.0 eV. The first process is known as interstitial-impurity diffusion, whilst the second is referred to as interstitialcy-impurity diffusion since the impurity becomes substitutional again after the ejection of I.

The predicted vacancy and interstitial concentration is $\sim 10^{13}$ cm^3 around 800°C, and this value is much increased after radiation damage, which includes ion implantation used to locally dope the silicon. Therefore the effects of the self-interstitial defect cannot be ignored. The increased diffusion of dopants via the interstitialcy mechanism is known as *transient enhanced diffusion* (TED), and it is this which is responsible for the large increase in dopant defect distribution after the thermal treatment. It is expected that by the turn of the century, the critical device dimensions will be equal to the dopant diffusion distance during TED for the typical annealing process. Clearly then, the self-interstitial defect, and its reaction processes are of great importance to the Si device industry.

There are two techniques for the growth of monocrystalline silicon, the Czochralski (Cz), and floating zone (FZ) methods. The first step however is the production of elemental Si by the chemical reduction of SiO_2 by carbon. After this, the reaction of the product with HCl, and the reduction of these yields polycrystalline Si. In the Cz method, the polycrystalline Si is first melted in a synthetic SiO_2 crucible. Following this, a seed Si crystal is inserted into the liquid and slowly withdrawn. Crystals grown by the Cz method have a large (10^{18} cm^{-3}) concentration of oxygen. In addition to this, graphitic contaminants or carbonaceous gases present during growth result in a carbon concentration of at least 5×10^{15} cm^{-3} (for example, the reaction of the graphite heating elements with the SiO gas, forming CO which dissolves in the molten Si). In the FZ method, the concentration of oxygen is much lower (10^{16} cm^{-3}) as in this, a molten silicon layer moves between two Si rods, and there is no direct contact between the melt and the crucible. However, the lowest attainable concentration of carbon in FZ material is similar to that in the Cz material. Indeed, the effects of the high oxygen concentration in Cz Si are drastic: if T$\leq$500°C, the so called thermal donors are formed which adversely affect the electrical properties of the material, and if 450$<$T$<$800°C then oxygen begins to precipitate forming SiO_2 regions in the material. This does not only affect the strength of the material, it also leads to the ejection of self-interstitials at some stage to accommodate the local increase in volume (since SiO_2 is a less dense material).

Carbon is another very important defect in Si and is sometimes purposely introduced for a variety of reasons: (i) at lower temperatures its presence leads to a suppression of thermal donor formation, leading instead to C–O precipitates; (ii) the small covalent radius of carbon means that it can help to relieve the strain caused by SiO_2 precipitation at higher temperatures; (iii) it is also able to soak up silicon self-interstitials (Si_i) produced during SiO_2 precipitation; and as a result carbon suppresses TED. Carbon can form a bewildering variety of defect complexes in Si, since both the substitutional and interstitial sites are stable, and it can easily switch from sp^2 to sp^3 hybridisation. The richness of the topic of carbon in silicon stems from this, and also the fact that it is able to complex with most other defects and impurities leading to a large number of defect centres. Therefore both oxygen and carbon are very important contaminants in Si.

Presented in this thesis are first principles calculations of the structure, optical, and electronic properties of several of the above centres in silicon and germanium, using the AIMPRO code. The contents of this thesis are now briefly described.

In chapters 2 and 3, the background theory is described in detail: the reduction of the many body Schrödinger equation to a series of one electron equations by either Hartree-Fock or density functional theory is described. The local density approximation is then discussed, and the first of these chapters then closes with a description of the pseudopotentials used in the code. The AIMPRO methodology; including the basis functions used, the approximations for various energy terms and the practicalities of the calculations is then discussed in the second of these chapters.

Chapter 4 discusses the main experimental methods used in defect physics, and what they are able to determine. Absorption, photoluminescence, infra-red, and deep level transient spectroscopies are first discussed followed by electron paramagnetic resonance. The perturbations of these spectroscopic features and alignment of defects, induced by the application of uniaxial stress are then described. Finally, since bi-stable defects occur continually throughout this thesis, a detailed description of how the experimentalist probes these phenomena is also discussed.

Following the background chapters, four results chapters are then presented. The isolated self-interstitial defect in silicon is first discussed in chapter 5. The energies of low energy configurations are calculated, and the migration of the defect is commented upon. Following this, the effects of the trapping of H by the self-interstitial is considered, and a combination of experimental work, and our theoretical calculations lead to the identification of the IH_2 defect. The chapter concludes with preliminary calculations on the {311} interstitial-aggregate defect, in which the binding energy of an one interstitial to the aggregate is calculated.

Carbon related defects in silicon are then discussed in chapter 6. The substitutional carbon defect is considered first before the more complex carbon centres. When the C_s defect captures a mobile self-interstitial, the isolated interstitial carbon defect C_i is formed. Unlike C_s, this defect is electrically active, and it is inevitable that it will be formed when the material is intentionally carbon doped to suppress TED. Annealing at room temperature results in the migration of C_i until it is trapped at other defects in the material. Several anomalies with the proposed model of the defect and the migration are addressed in the second part of this chapter. The final part of this chapter is concerned with the di-carbon defect, formed when the migrating C_i defect is trapped by C_s. The di-carbon defect is a remarkable centre, being bi-stable, and again electrically active. The energies and their charge state dependence are calculated for this centre, and again, some of the unresolved problems associated with this centre are addressed.

In chapter 7, the interaction between carbon (both substitutional and interstitial) and hydrogen is addressed. When hydrogen migrates to C_s an electrically active defect with low thermal stability is formed. There is then the potential for the formation of a more stable H_2^*-like defect if a second hydrogen atom is captured. Of greater technological importance are the highly stable complexes formed when hydrogen attacks C_i. Here a strong C-H bond is formed, and it is expected that the defects will not break up until the temperatures are much higher $\approx$600°C. The migration of C_i-H is considered, and a number of reaction products involving other carbon and hydrogen atoms is detailed. The results are compared with recent experimental findings of highly stable carbon-hydrogen complexes observed by photoluminescence, leading to the proposition of a structural model for one of the defects - the T-centre. The implications of the work in relation to silicon self-interstitial generation is also described.

Attention is then switched to germanium in chapter 8, where calculations are performed on a number of light impurities. The first half of this chapter is concerned with the interaction between the intrinsic point defects (V and I) and a number of hydrogen atoms. Parallel are drawn between these results, and the corresponding complexes in silicon. Carbon related defects, especially C_s are also expected to be present in germanium, and the second half this chapter focuses on these, as well as C_s in Si/Ge alloys. The inclusion of a small fraction of carbon into this alloy is currently being investigated as a means for the correction of the lattice mismatch between the alloy and pure silicon. This could ultimately enable epitaxial growth of the alloys with low dislocation density on the Si surface, and is therefore an area of considerable importance. The bonding of C_s in Si/Ge is investigated, and comparisons are made with recent experimental findings.

Finally, conclusions to the work presented, and suggestions for further investigations are given in chapter 9.

Chapter 2

Theoretical background

2.1 Introduction

Theoretical modelling of solid state systems is a very diverse area, and many approaches – varying vastly in complexity – are adopted. Each technique has its relative merits for the particular problem in hand, for example the lattice dynamics of solids/surfaces and heterostructures can be modelled accurately with the simplistic bond charge model [1]. More sophisticated potentials such as the Tersoff potential [2], where different coordinations of atoms are taken into account, have also proved very worthy in recent years. For a more detailed microscopic understanding understanding of the system under examination, higher level methods are employed, inevitably involving the solution of the many body Schrödinger equation 2.2 – with some level of approximation. Semi-empirical approaches such as tight binding methods: CNDO, MNDO, MINDO/3 and PM3 etc involve the parameterisation of some of the matrix elements of the Hamiltonian to experimental results, or higher level calculations. Such approaches have proved very successful in the past, but their reliability can be called into question if the application conditions do not resemble the conditions in which the parameters were obtained. With the recent advances in computing performance, yet more complex approaches have became possible on large scale systems. In particular, two first principles techniques, requiring only the atomic numbers and positions of atoms, are now widely used in quantum physics/chemistry. The first of these is Hartree-Fock theory where the variational parameters are the wavefunctions, whereas the electron density $n(\mathbf{r})$ is used in the second – density functional theory. These are discussed in section 2.3.

Considerable savings in time and basis size can be made if the core electrons can be removed from the calculation. In the late 70's and early 80's, pseudopotentials which effectively add a screening term, characteristic of the core electrons, to the Hamiltonian were developed [3] (section 2.4) and their use is common place today. Once the methodology has been decided upon, there are two final choices for the

the solid state theorist: (i) supercells where a basic block of atoms is repeated throughout space, and (ii) large hydrogen terminated clusters, typically having several shells of the bulk material. A plane wave basis is almost exclusively used in the supercell approach, although more recently localised basis sets have been used in supercell calculations; and one has to use a localised basis in the cluster method. Each approach has its advantages/disadvantages when applied to problems in solid state physics, but cluster approaches have advantages as they readily lend themselves to chemistry applications.

After the detailing of the background theory, the method adopted for the calculations throughout this work (AIMPRO) is discussed in chapter 3. AIMPRO, *Ab Initio Modelling PROgram*, is a density functional code, with a localised basis of Gaussian orbitals, with the calculations performed on large hydrogen-terminated clusters.

The success of any theory relates to a good modelling of the known experimental observables, along with good predictive powers. Therefore the ability to take as much as possible from the calculations with regards to experimental observables is a necessity. AIMPRO has the ability to calculate many of the characteristic quantities often found by the experimentalist in defect physics. Once the ground state/low energy structures are determined, we often focus on observables such as the local vibrational modes, the number of gap states and their localisation, as well as isotope effects – all of which can be determined trivially. Constrained relaxations can then be performed to calculate the migration and reorientation energies and pathways of the defects considered. Calculations of the defect-defect binding energies are also possible if a suitably large enough cluster is employed. Comments will be made on these calculations in chapter 3 section 3.3.

2.2 The fundamental problem

For the first principles calculation on an atomic level, where the only input are the atomic numbers, one must revert to the many body Schrödinger equation:

$$\hat{H}\Psi_i = E_i\Psi_i, \tag{2.1}$$

where Ψ_i and E_i are the wavefunction and energy of the i^{th} state, and $\hat{H}$ is the Hamiltonian given below:

$$\begin{aligned}\hat{H} &= \frac{1}{2}\left(-\sum_{i=1}\nabla_i^2 - \sum_{\alpha=1}\frac{m}{M_\alpha}\nabla_\alpha^2\right.\\ &\left.+\sum_{i,j\neq i}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_{i,\alpha}\frac{Z_\alpha}{|\mathbf{r}_i-\mathbf{R}_\alpha|} + \sum_{\alpha\beta\neq\alpha}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|}\right),\end{aligned} \tag{2.2}$$

here i, j and α, β label the electronic and nuclear coordinates respectively. Atomic units are used here, so $\hbar = e = m_e = 4\pi\epsilon_0 = 1$. The atomic unit of energy is then twice the hydrogen ionisation energy (27.212 eV) with the atomic unit of distance being the Bohr radius (0.529 Å). The full Schrödinger equation can then be written as:

$$(T_i + T_e + V_{e-e} + V_{e-i} + V_{i-i} - E)\Psi(\mathbf{r}) = 0. \tag{2.3}$$

Where $\mathbf{r}$ denotes the position and spin of all electrons $(\mathbf{r}_1, s_1, ...)$. The first step in the solution of the Schrödinger equation is to decouple the electronic and nuclear degrees of freedom. This stage is achieved in the Born-Oppenheimer approximation.

2.3 The Born-Oppenheimer approximation

The fact that the atomic nuclei are much more massive that the electrons, $m_{(n,p^+)} \approx 2000 m_e$, results in the motion of the nuclei modulating that of the electronic wavefunction. The total wavefunction can therefore be written as a product of the nuclear and electronic wavefunctions:

$$\Psi_{\mathrm{T}} = \chi(\mathbf{R})\Psi(\mathbf{r}, \mathbf{R}), \tag{2.4}$$

where $\mathbf{R}$ denotes the positions of all nuclei $(\mathbf{R}_1, \mathbf{R}_2, ...)$ and $\mathbf{r}$ is as before. Here $\chi(\mathbf{R})$ is an amplitude associated with the nuclear coordinates alone, and $\Psi(\mathbf{r}, \mathbf{R})$ is the solution of the Schrödinger equation. Substituting this product wavefunction into the Schrödinger equation, multiplying through by $\Psi^*(\mathbf{r}, \mathbf{R})$ and integrating, we arrive at:

$$\begin{aligned}\left(-\sum_a \frac{m}{2M_a}\nabla_a^2 + E(\mathbf{R}) + W(\mathbf{R}) - E_T\right)\chi(\mathbf{R}) = \\ \sum_a \int \Psi^*(\mathbf{r}, \mathbf{R})\frac{m}{M_a}\nabla_a\Psi(\mathbf{r}, \mathbf{R})\nabla_a\chi(\mathbf{R})d\mathbf{r},\end{aligned} \tag{2.5}$$

where $d\mathbf{r}$ is the integral over all electron coordinates $\mathbf{r}_i$, and the summation over all of the spin states is implied. The left hand side of this equation is the Schrödinger equation for the nuclei moving in a potential $E + W$, with W defined by:

$$W(\mathbf{R}) = -\sum_a \frac{m}{2M_a}\int \Psi^*(\mathbf{r}, \mathbf{R})\nabla_a^2\Psi(\mathbf{r}, \mathbf{R})dr. \tag{2.6}$$

The right hand side of equation 2.5 is zero if $\Psi(\mathbf{r}, \mathbf{R})$ is real, and the solution is a non-degenerate ground state. If this is not the case, then it represents a small perturbation, and is usually neglected – although it can have significant effects, for example the breaking of symmetry as in a Jahn-Teller distortion. The decoupling of the electronic and nuclear motion is achieved through neglecting this term.

The effective potential, E($\mathbf{R}$), acting on each electron due to all of the other electrons and nuclei is then constructed, and the forces acting on the nuclei are found for the ground state electron density. The nuclear positions can then be moved until the $E(\mathbf{R})$, the potential energy of the ions averaged over the state $\Psi(\mathbf{r}, \mathbf{R})$, is minimised. At this point, one has achieved the total ground state energy of the system, with respect to both the nuclear positions and the electron charge density.

2.4 Reduction of the problem to single particle equations

A practical approach for the solution of the many body Schrödinger equation is now required, and two approaches, Hartree-Fock theory and density functional theory are in common use today. Along with Hartree theory, these are variational techniques which reduce the many body Schrödinger equation to a series of decoupled linear differential equations. The resulting single particle states are solutions to the one electron Schrödinger equation, and can be found relatively easily. The full many body states can then be constructed using the single particle states if required. In the following subsections, the background of all of these methods is described.

2.4.1 The Hartree approximation

The wavefunction is expanded as a product of single particle states in the Hartree approximation, as shown below:

$$\Psi(\mathbf{r}) = \prod_{i=1}^{N} \phi_i(\mathbf{r}_i), \tag{2.7}$$

where the orbitals $\phi_i(\mathbf{r}_i)$ are orthonormal. Taking the expectation value of $\hat{H}$ with the wavefunction, *ie* $\langle \Psi^*(\mathbf{r}) | \hat{H} | \Psi(\mathbf{r}) \rangle$, it is found that:

$$\begin{aligned} E = -\frac{1}{2}\sum_i \int \phi_i^*(\mathbf{r})\nabla^2\phi_i(\mathbf{r}) - \sum_i \int |\phi_i(\mathbf{r})|^2 \sum_\alpha \frac{Z_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|} d\mathbf{r} \\ + \frac{1}{2}\sum_{i \neq j} \int\int \frac{|\phi_i(\mathbf{r}_1)|^2 |\phi_j(\mathbf{r}_2)|^2}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2. \end{aligned} \tag{2.8}$$

Applying the variational principle with Lagrange multipliers ϵ_{ij} to ensure that the single particle states are orthonormal:

$$\delta\left(E - \sum_i \epsilon_i \int \phi_i(\mathbf{r})\phi_j(\mathbf{r}) d\mathbf{r} \right),$$

leads to the Hartree equations:

$$-\frac{1}{2}\nabla^2\phi_i(\mathbf{r}) + \sum_i \int |\phi_i(\mathbf{r})|^2 \sum_\alpha \frac{Z_\alpha}{|\mathbf{r}-\mathbf{R}_\alpha|} d\mathbf{r} + \sum_{i\neq j} \int \frac{|\phi_j(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' = \sum_j \epsilon_{ij}\phi_i(\mathbf{r}), \quad (2.9)$$

where,

$$\int \phi_i(\mathbf{r})\phi_j(\mathbf{r}) = \delta_{ij}.$$

These single particle equations can then be solved self-consistently, although by construction, the many body wavefunctions here are not antisymmetric under particle interchange. This problem is rectified in the Hartree-Fock approximation.

2.4.2 The Hartree-Fock approximation

Here the wavefunction is now expanded as a Slater determinant of one electron orbitals:

$$\Phi(\mathbf{r}) = \frac{1}{\sqrt{N!}} \det |\phi_i(\mathbf{r}, s)|, \quad (2.10)$$

the two orbitals therefore satisfy:

$$\int \phi_i^*(\mathbf{r}, s)\phi_j(\mathbf{r}, s) dr = \delta_{ij}. \quad (2.11)$$

The construction of the many body wavefunction in this way ensures that the many body wavefunction is antisymmetric under particle exchange. The expectation value of the total energy, $\langle\Psi|\hat{H}|\Psi\rangle$ is then given by the following expression:

$$\begin{aligned} E \;=\;& \sum_n \langle\phi_n^*(\mathbf{r}, s)|\hat{T} + \hat{V}_{e-i} + \hat{V}_{i-i}|\phi_n(\mathbf{r}, s)\rangle + \\ & \frac{1}{2}\sum_{n,m}(\langle\phi_n^*(\mathbf{r}, s)\phi_m^*(\mathbf{r}, s)|\hat{V}_{e-e}|\phi_n(\mathbf{r}, s)\phi_m(\mathbf{r}, s)\rangle - \\ & \langle\phi_n^*(\mathbf{r}, s)\phi_m^*(\mathbf{r}, s)|\hat{V}_{e-e}|\phi_m(\mathbf{r}, s)\phi_n(\mathbf{r}, s)\rangle), \end{aligned} \quad (2.12)$$

throughout, the summation over s is implied. Here, the first term contains the one-particle kinetic, electron-ion and ion-ion matrix elements. The second summation involves four centre integrals relating to the electron-electron interaction. The total energy can then be written in the form:

$$E = -\frac{1}{2}\sum_n \int \phi_n^*(\mathbf{r}, s)\nabla^2\phi_n(\mathbf{r}, s) d\mathbf{r} + \int n(\mathbf{r})V_{e-i} d\mathbf{r} + E_H + E_x + E_{i-i}, \quad (2.13)$$

where E_H, E_x and E_{i-i} are the Hartree, exchange and ion-ion energies respectively, given by:

$$E_H = \frac{1}{2}\int \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2, \quad (2.14)$$

$$E_x = -\frac{1}{2}\sum_{n,m}\int \phi_n^*(\mathbf{r}, s)\phi_m^*(\mathbf{r}, s)\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\phi_m(\mathbf{r}, s)\phi_n(\mathbf{r}, s) d\mathbf{r}_1 d\mathbf{r}_2, \quad (2.15)$$

$$E_{i-i} = \frac{1}{2} \sum_{\alpha \neq \beta} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}, \tag{2.16}$$

and $n(\mathbf{r})$ is the total charge density given by:

$$n(\mathbf{r}) = \sum_{n,s} |\phi_n(\mathbf{r}, s)|^2. \tag{2.17}$$

The total energy E is minimised subject to the orthonormal orbitals $\phi_n(\mathbf{r}, s)$. Using Lagrange multipliers ϵ_{nm}, ϵ_n:

$$E - \sum_{n \neq m} \epsilon_{nm} \int \phi_n^*(\mathbf{r}, s) \phi_m(\mathbf{r}, s) d\mathbf{r} - \sum_n \epsilon_n \left(\int |\phi_n(\mathbf{r}, s)|^2 d\mathbf{r} - 1 \right),$$

with respect to $\phi_n(\mathbf{r}, s)$, E_n and $E_{n,m}$. Performing the minimisation using E from 2.13 we arrive at the Hartree-Fock equations for each orbital n:

$$\left(-\frac{1}{2} \nabla^2 + V_{e-i}(\mathbf{r}) + V^H(\mathbf{r}) + V_n^x(\mathbf{r}) - \epsilon_n \right) \phi_n(\mathbf{r}, s) = \sum_{n \neq m} \epsilon_{nm} \phi_m(\mathbf{r}, s), \tag{2.18}$$

with,

$$\sum_s \int \phi_i^*(\mathbf{r}, s) \phi_j(\mathbf{r}, s) d\mathbf{r} = \delta_{ij}.$$

$V^H(\mathbf{r})$ and $V_n^x(\mathbf{r})$ are the Hartree and exchange potentials respectively, given by:

$$V^H(\mathbf{r}) \phi_i(\mathbf{r}, s) = \frac{\delta E_H}{\delta \phi_i^*} = \int \frac{n(\mathbf{r}') \phi_i(\mathbf{r}, s)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \tag{2.19}$$

$$V_i^x(\mathbf{r}) \phi_i(\mathbf{r}, s) = \frac{\delta E_x}{\delta \phi_i^*} = -\sum_{j s_i} \int \phi_j^*(\mathbf{r}', s) \phi_i(\mathbf{r}', s) \frac{1}{|\mathbf{r}' - \mathbf{r}|} \phi_j(\mathbf{r}, s) d\mathbf{r}', \tag{2.20}$$

where the summation in the exchange term is over occupied orbitals j whose spin state is the same as state i, indicated by s_i.

Now, if ϵ_{nm} is diagonalised, by performing a unitary transformation on the Slater determinant, then the right hand side of 2.18 vanishes. Multiplying the modified equation by $\phi_i^*(\mathbf{r}, s)$, integrating over $\mathbf{r}$, and summing over the occupied spin orbitals i and spin states s gives:

$$T_e + E_{e-ion} + 2E_H + 2E_x + \sum_i \epsilon_i = 0, \tag{2.21}$$

where T_e and E_{e-ion} represent the kinetic energy of the electrons, and the electron-ion potential energy respectively. The total energy is then found by adding the ion-ion term, and removal of the double counted terms, giving:

$$E = \sum_i \epsilon_i - E_H - E_x + E_{ion-ion}, \tag{2.22}$$

where i are occupied orbitals.

Koopmans theorem leads to an interpretation of the energy eigenvalues ϵ_i: consider two electronic configurations, which differ only by the occupation of one electron in an orbital ψ_i: $\Psi_1 = \sum_{\alpha=1}^{N} \phi_\alpha$ and $\Psi_2 = \sum_{\alpha=1}^{N} \phi_\alpha + \phi_i$. Then ϵ_i represents the difference in energy between the two configurations, or -ϵ_i is the ionisation energy for the i^{th} electron. The proof requires that all orbitals other than ϕ_i are identical in both configurations. In practice the introduction of en electron in state ϕ_i will alter all other orbitals, so the theory is not exact.

The Hartree-Fock equations are solved by a self consistent cycle, when complete the sum of occupied orbitals relate to the ground state electronic configuration (charge density) of the system.

The first major draw back of the method is exposed when it is applied to the homogeneous electron gas termed *Jellium*. The orbitals are plane waves, and such a system is obviously conducting (metallic), with a finite density of states at the Fermi level. Once the total energy, and hence the density of states as a function of wavevector $\mathbf{k}$ are determined using Hartree-Fock theory a disturbing result emerges: the theory predicts that $N(E) \rightarrow 0$ as $\mathbf{k} \rightarrow \mathbf{k_F}$. The source of this error is the absence of correlation in Hartree-Fock theory, that is to say the wavefunction does not include the correlated motion of electrons. Correlation can be included exactly, by constructing the wavefunction as a linear combination of Slater determinants, where the coefficients are chosen to minimise to total energy. These are known as configuration interaction (CI) calculations, but their computational demands are so extreme that calculations are limited to a small number of atoms only.

Hartree-Fock calculations generally give very good results when applied to small molecules, with structures and vibrational modes being in good agreement with experiment. However, even without the inclusion of correlation, another draw back is limitation to the system size which the method can readily applied to. A number of four-centre integrals must be repeatedly evaluated and these are very computationally demanding. As a result, the method is impractical for large solid state systems, unless some further approximations are made, removing these four centre integrals. Such approaches (for example CNDO, MNDO, MINDO/3 etc) involve the parameterisation of the troublesome matrix elements to experimental data, although the level of approximation varies in the different methods.

The following section is devoted to density functional theory (DFT), where $n(\mathbf{r})$ is used as the fundamental variable. The Schrödinger equation is again reduced to a series of one electron equations which are solved self-consistently. DFT is the most common first-principles technique employed in solid state physics calculations today including *AIMPRO*.

2.5 Density functional theory

In the previous two formalisms Hartree and Hartree-Fock theory, the electron wavefunction was used to uniquely determine the energy. In density functional theory, the electron charge density, defined by:

$$n(\mathbf{r}) = \sum_i \int \delta(\mathbf{r} - \mathbf{r}_i)|\Psi(\mathbf{r}')|^2 d\mathbf{r}', \tag{2.23}$$

is the fundamental variable, ($\mathbf{r}' = \mathbf{r_1} \rightarrow \mathbf{r_N}$). For such an approach to applicable, it is clear that there must be a one to one correspondence between the Hamiltonian and the charge density for the ground state.

2.5.1 The Hohenberg-Kohn theorem

In 1964, Hohenberg and Kohn showed that for a non-spin polarised non-degenerate ground state, there was indeed a one to one correspondence between $\hat{H}$, $\Psi(\mathbf{r})$ and $n(\mathbf{r})$ [4]. The proof is as follows:

Consider the Hamiltonians H_1 and H_2 giving rise to ground state wavefunctions Ψ_1 and Ψ_2 respectively,

$$H_1 = T + V_1 \rightarrow \Psi_1 \quad \text{and} \quad H_2 = T + V_2 \rightarrow \Psi_2,$$

let us now assume that both potentials V_1 and V_2 give rise to the same ground state electron density $n(\mathbf{r})$. It is clear that if $V_1 \neq V_2$ then $\Psi_1 \neq \Psi_2$. The expectation values of the energies are given by

$$E_1 = \langle\Psi_1|H_1|\Psi_1\rangle \quad \text{and} \quad E_2 = \langle\Psi_2|H_2|\Psi_2\rangle,$$

if the variational principle is now invoked, it is found that

$$E_1 = \langle\Psi_1|H_1|\Psi_1\rangle < \langle\Psi_2|H_1|\Psi_2\rangle,$$

and hence

$$E_1 < E_2 + \int (V_1 - V_2) n(\mathbf{r}) d\mathbf{r},$$

conversely, going the other way, it is found that

$$E_2 < E_1 + \int (V_2 - V_1) n(\mathbf{r}) d\mathbf{r},$$

adding these two equations together gives

$$E_1 + E_2 < E_1 + E_2.$$

There is therefore a contradiction unless the ground state energies are degenerate. Therefore our original assumption that both V_1 and V_2 can lead to the same ground state charge density $n(\mathbf{r})$ was incorrect, the charge density uniquely defines V_{e-ion} and hence $\Psi(\mathbf{r})$. This result is remarkable as it states that the charge density, a function of three spatial variables uniquely defines the wave function which is a function of $N_e\times$(3 spatial + 1 spin) variables.

2.5.2 The variational principle in terms of charge density

Hohenberg and Kohn then went on to show, that if the energy, (E), is expressed in terms of a universal functional $F[n(\mathbf{r})]$, which is independent of $v(\mathbf{r})$, then E has its minimum value for the correct ground state of the system associated with the potential $v(\mathbf{r})$. Hence the charge density can be treated as the variational parameter. Defining $F[n(\mathbf{r})]$, which is a universal function of energy by:

$$F[n(\mathbf{r})] = \langle \Psi \mid \hat{T} + \hat{U} \mid \Psi \rangle, \tag{2.24}$$

the energy functional is then:

$$E_v[n(\mathbf{r})] = \int v(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + F[n(\mathbf{r})]. \tag{2.25}$$

For the correct $n(\mathbf{r})$, it is clear that the $E[n]$ will be equal to the ground state energy E. In fact, the following argument shows that for the correct $n(\mathbf{r})$ the functional $E[n]$ also assumes its minimum value, a more powerful result as it means the charge density can be treated as the variational parameter:

Given two external potentials $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$ that give rise to ground states Ψ_1 and Ψ_2 having charge densities $n_1(\mathbf{r})$ and $n_2(\mathbf{r})$, then we have

$$E_1[n_1(\mathbf{r})] = \int v_1(\mathbf{r}) n_1(\mathbf{r}) d\mathbf{r} + F[n_1(\mathbf{r})] = \langle \Psi_1 | \hat{H}_1 | \Psi_1 \rangle,$$

$$E_2[n_2(\mathbf{r})] = \int v_2(\mathbf{r}) n_2(\mathbf{r}) d\mathbf{r} + F[n_2(\mathbf{r})] = \langle \Psi_2 | \hat{H}_2 | \Psi_2 \rangle,$$

if the energy functional E_1 now operates on the charge density $n_2(\mathbf{r})$ then it is found that,

$$E_1[n_2(\mathbf{r})] = \int v_1(\mathbf{r}) n_2(\mathbf{r}) d\mathbf{r} + F[n_2(\mathbf{r})] = \langle \Psi_2 | \hat{H}_1 | \Psi_2 \rangle > \langle \Psi_1 | \hat{H}_1 | \Psi_1 \rangle.$$

ie E_1 is minimised by Ψ_1 and hence by $n_1(\mathbf{r})$. However, this proof required that for any charge density $n(\mathbf{r})$, an external potential $v(\mathbf{r})$ exists giving rise to this charge density. This condition is not generally true, and is called the v representability problem.

Hohenberg and Kohn then wrote the total energy in terms of a functional of the charge density, where the Hartree energy has been extracted from the universal functional $F[n]$.

$$E_{TOT}[n] = \int n(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \frac{n(\mathbf{r}_1) n(\mathbf{r}_2)}{\mid \mathbf{r}_1 - \mathbf{r}_2 \mid} d\mathbf{r}_1 d\mathbf{r}_2 + G[n] + E_{ion-ion}. \tag{2.26}$$

The first term in this expression is potential energy contribution of the electrons with the external potential. The second term, the Hartree contribution, is the electron electron potential energy term. $G[n]$ is a functional which contains the kinetic energy and exchange-correlation effects. Finally, $E_{ion-ion}$ is the potential energy contribution resulting from the ion-ion interaction.

2.6 The Kohn-Sham equations and local density approximation

In 1965, Kohn and Sham [5] re-wrote the functional, and obtained a set of one electron equations, similar to the Hartree-Fock equations, that had to be solved self consistently. Writing the energy functional in terms of kinetic energy and exchange-correlation functionals,

$$E[n] = T[n] + E_{e-i}[n] + E_H[n] + E_{xc}[n] + E_{i-i}. \tag{2.27}$$

where expressions for $E_{e-i}[n]$, $E_H[n]$ and E_{i-i} are readily available. Equation 2.27 becomes:

$$\begin{aligned} E_{TOT}[n] = T[n] - \int n(\mathbf{r}) \sum_a \frac{Z_a}{|\mathbf{r} - \mathbf{R}_a|} d\mathbf{r} + \frac{1}{2} \int \frac{n(\mathbf{r}_1) n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \\ + E_{xc}[n] + \frac{1}{2} \sum_{a \neq b} \frac{Z_a Z_b}{|\mathbf{R}_a - \mathbf{R}_b|}. \end{aligned} \tag{2.28}$$

However, expressions for $T[n]$ and $E_{xc}[n]$ are required for the evaluation of the total energy in terms of $n(\mathbf{r})$.

2.6.1 The kinetic energy term

In order to obtain an expression for the Kinetic energy functional, Kohn and Sham wrote the charge density $n(\mathbf{r})$ as a summation of orthonormal orbitals:

$$n(\mathbf{r}) = \sum_{i=1}^{N} \mid \psi_i(\mathbf{r}) \mid^2, \tag{2.29}$$

where ψ_i is the wave-function of the i^{th} occupied state. These states are only employed as a tool for expanding the charge density, and strictly speaking cannot be interpreted as single particle states. The kinetic energy is then written as:

$$T = -\frac{1}{2} \sum_{i,s} \int \psi_i^* \nabla^2 \psi_i d\mathbf{r}. \tag{2.30}$$

2.6.2 The exchange-correlation energy

An exact expression for the exchange-correlation energy is not available, as it describes the complicated many-body interactions taking place. So, for the first time in our expansion of E_{tot} in terms of $n(\mathbf{r})$ we must make an approximation. The most common approximations used for this term is the local density approximation (LDA), where E_{xc} is written in terms of the exchange-correlation energy density ϵ_{xc}:

$$E_{xc}[n] = \int n(\mathbf{r})\epsilon_{xc}(n_{\uparrow}, n_{\downarrow})d\mathbf{r}. \tag{2.31}$$

The exchange-correlation energy density is then approximated by:

$$\epsilon_{xc}(\mathbf{r}) = \epsilon_{xc}(n(\mathbf{r})). \tag{2.32}$$

The LDA approximation therefore writes the non-local exchange-correlation potential at a point $\mathbf{r}$ by the value found from the homogeneous electron gas with constant charge density $n(\mathbf{r})$. The exchange-correlation potential is given by the following expression:

$$\mu_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} = \epsilon_{xc}(n(\mathbf{r})) + n(\mathbf{r})\epsilon'_{xc}(n(\mathbf{r})), \tag{2.33}$$

which Kohn and Sham showed to be correct in two limiting cases: (i) slowly varying $n(\mathbf{r})$ and (ii) high density $n(\mathbf{r})$, where the kinetic energy term dominates and any corrections to E_{xc} are negligible.

2.6.3 The exchange-correlation energy density

The final problem, is to find an appropriate expression for $\epsilon_{xc}(n(\mathbf{r}))$. The random phase approximation (RPA) can be used to obtain exact value for $\epsilon_{xc}(n(\mathbf{r}))$ (Gell-Mann and Breuckner 1957) [6], but the results are only applicable for the case of high electron densities. Numerical methods must be used to obtain values for $\epsilon_{xc}(n(\mathbf{r}))$ at the more realistic lower densities. The approximation that is used in our work, is that primarily due to Ceperly and Alder (1980) [7]. They adopted a quantum Monte Carlo approach, and sampled a correlated wavefunction for electrons in a finite volume with periodic boundary conditions. They then extrapolated their result for the energy per electron to the case of infinite volume. Perdew and Zunger (1981) [8] proceeded to parametrise the results of Ceperly and Alder, and matched these to the RPA result at higher densities. The resulting expressions are as follows:

$$\epsilon_{xc}(r_s) = \epsilon_x(r_s) + \epsilon_c(r_s), \tag{2.34}$$

where,

$$\epsilon_x(r_s) = -\frac{0.4582}{r_s},$$

$$\epsilon_c(r_s) = \begin{pmatrix} -0.1423 \\ -0.0843 \end{pmatrix} \left[1 + \begin{pmatrix} 1.0529 \\ 1.3981 \end{pmatrix} \sqrt{r_s} + \begin{pmatrix} 0.3334 \\ 0.2611 \end{pmatrix} r_s \right]^{-1} \qquad r_s \geq 1,$$

$$\epsilon_c(r_s) = \begin{pmatrix} -0.0480 \\ -0.0269 \end{pmatrix} + \begin{pmatrix} 0.0311 \\ 0.0155 \end{pmatrix} \ln r_s - \begin{pmatrix} 0.0116 \\ 0.0048 \end{pmatrix} r_s + \begin{pmatrix} 0.0020 \\ 0.0007 \end{pmatrix} r_s \ln r_s \qquad r_s \leq 1,$$

where the upper coefficients apply to the spin averaged case, and the lower ones to the spin polarised electron gas. r_s is the Wigner-Seitz radius for each electron defined by $r_s = (4\pi n/3)^{-\frac{1}{3}}$. This parameterisation for ϵ_{xc} is used in AIMPRO.

2.6.4 The Kohn-Sham one electron equations

Minimising equation 2.28 subject to the orthonormality constraint,

$$\frac{\delta}{\delta \psi_i^*} \left(E - \sum_i \epsilon_i \int |\psi_i(\mathbf{r})|^2 d\mathbf{r} \right) = 0,$$

one obtains a set of one-electron equations:

$$-\frac{1}{2}\nabla^2 \psi_i + \left[V_{ext}(\mathbf{r}) + \int \frac{n(\mathbf{r}')d\mathbf{r}'}{\mid \mathbf{r} - \mathbf{r}' \mid} + \frac{\delta E_{xc}[n]}{\delta n} \right] \psi_i = \epsilon_i \psi_i, \tag{2.35}$$

where the term in the large square brackets can be considered as the effective potential. These equations are then solved self-consistently. With a similar approach as in Hartree-Fock theory the total energy is then given by:

$$E_{TOT} = \sum_\lambda \epsilon_\lambda - \frac{1}{2} \int \frac{n(\mathbf{r_1})n(\mathbf{r_2})}{\mid \mathbf{r_1} - \mathbf{r_2} \mid} d\mathbf{r_1} d\mathbf{r_2} + E_{xc}[n] - \int n(\mathbf{r}) \frac{\delta E_{xc}[n]}{\delta n} d\mathbf{r} + E_{ion-ion}. \tag{2.36}$$

As mentioned previously, the Kohn-Sham orbitals are not single particle states of the system, and their physical interpretation is therefore unclear. Similarly, the Kohn-Sham eigenvalues were the Lagrangian multipliers used in the minimisation, so they cannot truly represent excitations of the system, which are a result of solving the Dyson equation. One exact result coming from the Kohn-Sham equations is that the highest occupied orbital has an eigenvalue identical to the real chemical potential μ. Despite the uncertainty, the Kohn-Sham eigenvalues and vectors seem to have some physical meaning. For example when investigating semiconductors, there is a clear band gap, and defects give rise to gap states whose pseudowavefunctions appear similar to the true wavefunction localisation of the corresponding state. However the magnitude of the band gaps are often in serious error. These errors are due DFT itself, and not a result of any subsequent approximation.

2.7 Pseudopotential theory

The calculational procedure can be dramatically simplified by the reduction of the total number of electrons considered in any problem – in particular the removal of the core electrons (and states) from the calculation. The justification of this step is simple enough: the core states are highly localised, and have no effect on the chemical properties of atoms. Hence for any practical purpose, only the valence electrons need to be considered. In fact, the inclusion of the core electrons would considerably increase the difficulty of the calculation. To satisfy the wavefunction orthogonality condition, all of the wavefunctions close to the core oscillate wildly making them extremely difficult to model. This in turn would also lead to a rapid oscillation in the valence electron wavefunction in the core region a consequence of the high kinetic energy of the core states. Another potential problem with the inclusion of any core states is that their presence would result in an extremely large value of the total energy. Any errors in the evaluation or the total energy could then make the comparison of energies of similar systems impossible. Another advantage of omitting the core electrons is concerned with heavy atoms. Relativistic effects become increasingly important as the atomic mass increases, and hence it would become necessary to solve the Dirac equation. The valence electrons however continue to behave in a non-relativistic fashion, meaning that at most, a spin-orbit potential needs to be introduced.

The way in which the core states are removed from the calculation is through the use of *pseudopotentials*. The valence electrons are then only considered; the effects of the core electrons on the valence electrons being described via the inclusion of the pseudopotential term in the Hamiltonian. Once the Schrödinger equation is solved, the lowest eigenvalues correspond to those of the valence states only. The construction of the pseudopotentials themselves is not a trivial matter, as there remains the possibility of an exchange interaction between the core and valence electrons.

The most important criteria that any useful pseudopotential must satisfy is the transferability condition, given below in atomic units:

$$-2\pi \left[(r\psi)^2 \frac{d}{d\epsilon} \frac{d}{dr} \ln\psi \right]_R = 4\pi \int_0^R \psi^2 r^2 dr, \tag{2.37}$$

where ψ is the solution to the radial Schrödinger equation at a given energy ϵ, which need not be an eigenenergy. The radial logarithmic derivative of ψ is related to the scattering phase shift. If the normalised pseudo-wavefunction for a valence state, Ψ_{pseudo}, and the corresponding all electron valence wavefunction $\Psi_{all-elec}$ are identical for $R > R_c$, then outside R_c, the integrated charges are identical and the scattering properties have the same variation in energy to first order. A consequence of equation 2.37 is that for two potentials v_1, v_2 having solutions ψ_1, ψ_2 then given that they had the same integrated charge inside some sphere of radius R, then their scattering phase shifts at R around ϵ are identical.

The pseudopotentials used in AIMPRO are the norm-conserving ones due to Bachelet, Hamann and Schlüter, (BHS), (1982) [3]. Norm-conserving pseudopotentials satisfy two conditions: (i) they are nodeless, and (ii) given a core radius R_c, for $R > R_c$ the normalised pseudowavefunction and real wavefunction are identical. The valence and pseudo energy eigenvalues, ϵ^{ps} and ϵ^{true}, are the same. They also yield the exact atomic charge density outside the core, which is essential for self consistent calculations. Therefore by construction, any norm conserving pseudopotential satisfies equation 2.37. BHS succeeded in tabulating reliable pseudopotential parameters for all of the elements between H and Pu - including relativistic effects for the larger atoms.

The first step in the construction of their pseudopotentials was to solve the Kohn-Sham equations for all of the electrons in the atom, giving the true radial wavefunctions $R_l(r)$ and their eigenvalues. At this stage, the potential acting on a given valence electron is from both the core electrons, as well as any other valence electrons. The spin densities and the all electron potential $V^{\text{true}}(r)$ can then be found, where $V^{\text{true}}(r)$ is a sum of the nuclear, Hartree and exchange-correlation potentials:

$$V^{\text{true}}(r) = -\frac{Z}{r} + \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + \mu_{xc}[n(\mathbf{r})]. \tag{2.38}$$

This potential has a coulomb singularity at $r = 0$. A step 1 pseudopotential is then constructed equation 2.39, which eliminates the coulomb singularity.

$$V_l^{(1)}(r) = V^{\text{true}}(r)\left[1 - f\left(\frac{r}{r_{c,l}}\right)\right] + c_l^{\text{true}} f\left(\frac{r}{r_{c,l}}\right). \tag{2.39}$$

The function $f(x)$ is chosen so that it is equal to one at the origin, and rapidly tends to zero when x is greater than one, for example $f(x) = e^{-x^{3.5}}$. The pseudopotential is therefore constant near the origin. The parameter c_l^{true} is adjusted until the lowest energy level for each value of l coincides with the all electron level with the same angular momentum.

The first step pseudo-wavefunction $\Psi^{(1)}(r)$ is then equal to the all electron wavefunction $\Psi^{\text{true}}(r)$ to within a normalisation factor. This follows because $V^{\text{true}}(r)$ and $V_l^{(1)}(r)$ are identical outside a certain radius $r_{c,l}$ which is known as the core radius. In practice, $r_{c,l}$ must be carefully chosen: it cannot be too small as it would lie in the troublesome region where rapid oscillations of $\Psi^{\text{true}}(r)$ occur, but clearly it cannot be too large. Typically, a value of r between the outermost node, and outermost extrema of $\Psi^{\text{true}}(r)$ is chosen for the core radius.

In the second step, $\Psi^{(1)}(r)$ is modified by the introduction of a short ranged term:

$$\Psi^{(2)}(r) = \alpha_l\left[\Psi^{(1)}(r) + \beta_l r^{l+1} f\left(\frac{r}{r_{c,l}}\right)\right], \tag{2.40}$$

and the constants α_l and β_l are chosen so that when normalised $\Psi^{(2)}(r)$ and $\Psi^{\mathrm{true}}(r)$ are identical outside $r_{c,l}$. The pseudopotential $V_l^{(2)}(r)$ which gives rise to $\Psi^{(2)}(r)$ is then found by inverting the Schrödinger equation.

The above pseudopotential is that acting on a given valence electron due to the core electrons, and all of the other valence electrons. The final core pseudopotential is found by descreening, where the exchange-correlation and Hartree energies are calculated for the valence electrons from the electron density. These energy contributions are then removed leaving the bare ion-core potential. The pseudopotentials are *non-local* that is to say angular momentum dependencies are intrinsically built in.

Finally, the resultant potential were parameterised, and written in terms of the following functions: $erf(x)$, e^{-x^2} and $x^2e^{-x^2}$. This choice leads to analytic expressions for the matrix elements of the pseudopotentials available between either Gaussian or plane wave basis sets. Their expression was as follows:

$$V_l(r) = -\frac{Z_\nu}{r}\{\sum_{i=1}^{2} c_i^c \; erf(\sqrt{\alpha_i^c} r)\} + \sum_{i=1}^{3}\{A_{i,l} + r^2 A_{i+3,l}\}e^{-\alpha_{i,l}r^2}, \tag{2.41}$$

where Z_ν is the valence charge. Note here that the first term in this expression in local (l independent). The coefficients of this local term satisfy $c_1^c + c_2^c = 1$, which results in the local part giving rise to a potential $-Z_\nu/r$ for $\alpha_i^c r^2 >> 1$. The second term is dependent on l, and is the non-local part of the pseudopotential. Tables of all of the coefficients in 2.41 are given in [3] for a wide range of atoms. The approach also catered for relativistic atoms where there is some modification to the construction of the pseudopotential. The result of this is a spin-orbit potential term given by:

$$V_l^{so}(r) = \sum_{i=1}^{3}\{B_{i,l} + r^2 B_{i+3,l}\}e^{-\alpha_{i,l}r^2}. \tag{2.42}$$

Finally a practical point about the construction of the pseudopotentials must be noted. In the solid state, higher angular momentum states may play a role, and these were less pronounced in the isolated atom. For the high l states which do not appear in the ground state, an excited state configuration is used in the initial all-electron calculation. For example the d pseudopotential is found from the silicon atom with an excited electronic configuration given by $3s^1 3p^{0.75} 3d^{0.25}$.

Chapter 3

The AIMPRO approach

3.1 Introduction

The basis expansion and approximations used in AIMPRO are now discussed in detail. The wavefunctions of the cluster are expanded in a basis of localised orbitals $\phi_i(\mathbf{r}-\mathbf{R_i})$ such that the wavefunction of the λ^{th} level is given by:

$$\psi_\lambda(\mathbf{r},s) = \chi_\alpha(s)\sum_i c_i^\lambda \phi_i(\mathbf{r}-\mathbf{R}_i), \tag{3.1}$$

where Gaussian orbitals of the form,

$$(x-R_{ix})^{n_1}(y-R_{iy})^{n_2}(z-R_{iz})^{n_3}e^{-a_i(\mathbf{r}-\mathbf{R}_i)^2}, \tag{3.2}$$

are used for the localised orbitals. The angular momentum number of the orbital is then determined by the integers $n_{1,2,3}$, for example $n_1 = n_2 = n_3 = 0$ would give a spherically symmetric s orbital. The Kohn Sham equations are thus converted to matrix equations for c_i^λ.

The charge density for each spin is then expressed in terms of a density matrix $b_{ij,s}$, in the following way:

$$n_s(\mathbf{r}) = \sum_{ij} b_{ij,s}\phi_i(\mathbf{r}-\mathbf{R}_i)\phi_j(\mathbf{r}-\mathbf{R}_j), \tag{3.3}$$

where,

$$b_{ij,s} = \sum_\lambda \delta(s,s_\lambda)c_i^\lambda c_j^\lambda. \tag{3.4}$$

The total charge density is then given by

$$n(\mathbf{r}) = \sum_s n_s(\mathbf{r}). \tag{3.5}$$

This basis is now used to evaluate the total energy using the functional and approximations in the previous section. The choice of Gaussian orbitals ensures that the kinetic and pseudopotential matrix elements are easily found:

$$T_{ij} = -\frac{1}{2}\int \phi_i(\mathbf{r}-\mathbf{R}_i)\nabla^2\phi_j(\mathbf{r}-\mathbf{R}_j)d\mathbf{r}, \tag{3.6}$$

$$V_{ij}^{ps} = \int \phi_i(\mathbf{r} - \mathbf{R}_i) \sum_a V_a^{ps}(\mathbf{r} - \mathbf{R}_a)\phi_j(\mathbf{r} - \mathbf{R}_j) d\mathbf{r}, \tag{3.7}$$

the integrals being analytically performed. The Hartree term poses a problem within the current formalism as this requires the evaluation of $O(N^4)$ integrals (N being the number of basis functions). This represents a phenomenal task computationally and would certainly be impractical for any large scale systems. The exchange-correlation integrals are similarly problematic, hence some approximation scheme, whereby both of these energies are calculated analytically with a more favourable $O(N^3)$ scaling must be adopted. A further requirement of the approximations schemes is that they are analytically differentiable with respect to the nuclear positions, so that the forces acting on each atom may be analytically evaluated. An approximate charge density for each spin $n_s(\mathbf{r})$ is used to achieve this [9, 10]:

$$\tilde{n}_s(\mathbf{r}) = \sum_k c_{k,s} g_k(\mathbf{r}), \quad \tilde{n}(\mathbf{r}) = \sum_s \tilde{n}_s(\mathbf{r}). \tag{3.8}$$

Approximate expressions for both the Hartree and exchange-correlation integrals are now derived using these charge densities.

3.1.1 The Hartree energy

In terms of the true charge density $n(\mathbf{r})$, the Hartree energy is given by:

$$E_H = \frac{1}{2} \int \frac{n(\mathbf{r}_1) n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2. \tag{3.9}$$

We now replace this expression by an approximate value $\tilde{E}_H$, using the approximate charge density $\tilde{n}(\mathbf{r})$, giving:

$$\tilde{E}_H = \int \frac{n(\mathbf{r}_1)\tilde{n}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 - \frac{1}{2} \int \frac{\tilde{n}(\mathbf{r}_1)\tilde{n}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2. \tag{3.10}$$

Clearly, this expression is exact when $n = \tilde{n}$. The next step is to then expand the charge density in terms of a basis set $g_k(\mathbf{r})$,

$$\tilde{n}(\mathbf{r}) = \sum_k c_k g_k(\mathbf{r}), \tag{3.11}$$

and minimise the error in the estimation of the Hartree energy,

$$E_H - \tilde{E}_H = \frac{1}{2} \int \frac{\{n(\mathbf{r}_1) - \tilde{n}(\mathbf{r}_1)\}\{n(\mathbf{r}_2) - \tilde{n}(\mathbf{r}_2)\}}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2,$$

$$E_H - \tilde{E}_H = \sum_{ijk} c_k b_{ij} t_{ijk} - \frac{1}{2} \sum_{kl} c_k c_l G_{kl}, \tag{3.12}$$

with respect to c_k. Differentiating with respect to c_k, one finds that the minimum is described by:

$$\sum_l G_{kl} c_l = \sum_{ij} t_{ijk} b_{ij}, \tag{3.13}$$

where,

$$t_{ijk} = \int \phi_i(\mathbf{r}_1 - \mathbf{R}_i)\phi_j(\mathbf{r}_1 - \mathbf{R}_j) g_k(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2, \tag{3.14}$$

$$G_{kl} = \int g_k(\mathbf{r}_1) g_l(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2. \tag{3.15}$$

the expression for the approximate Hartree energy $\tilde{E}_H$ is therefore given by:

$$\tilde{E}_H = \frac{1}{2} \sum_{kl} c_k c_l G_{kl}. \tag{3.16}$$

The most straightforward choice for the charge density basis g_k are Gaussian orbitals of the form,

$$g_k(\mathbf{r}) = e^{-b_k(\mathbf{r}-\mathbf{R}_k)^2},$$

for sites R_k with exponents of b_k. Typically, these functions are placed at atomic and bond centred sites. All integrals involved in the calculation of the Hartree energy can then be analytically calculated. The time required for the evaluation of the t_{ijk} integrals often makes it necessary to choose a more sophisticated form for the basis $g_k(\mathbf{r})$ that result in a simple analytic form for t_{ijk} which can then be evaluated much more rapidly. This is achieved with functions of the following form,

$$g_k = \{1 - \frac{2b_k}{3}(\mathbf{r} - \mathbf{R}_k)^2\} e^{-b_k(\mathbf{r}-\mathbf{R}_k)^2}, \tag{3.17}$$

giving rise to a Gaussian potential,

$$\int \frac{g_k(\mathbf{r}_1)}{|\mathbf{r} - \mathbf{r}_1|} d\mathbf{r}_1 = \frac{3b_k}{2\pi} e^{-b_k(\mathbf{r}-\mathbf{R}_k)^2}. \tag{3.18}$$

However, additional pure Gaussian functions are necessary to contribute to the total number of electrons when the integrals involving the above form of g_k vanish – as all three Gaussians must overlap to give a non-zero value here.

3.1.2 The exchange-correlation energy

As with the Hartree energy, an intermediate fit $\tilde{n}$ to the charge density is used in the evaluation of the exchange-correlation energy. The fitting functions are not the same as those used for E_H, because there they were chosen to minimise the error in $n - \tilde{n}$ which meant that at a given value of $\mathbf{r}$, n and $\tilde{n}$ may differ significantly.

Additionally, in the Hartree fit, g_k was chosen in such a way that it eased the calculation of t_{ijk}. The exchange-correlation energy E_{xc} is given by:

$$E_{xc} = \int \epsilon_{xc}(n\uparrow, n\downarrow) n d\mathbf{r}, \tag{3.19}$$

is approximated by,

$$\tilde{E}_{xc} = \int \epsilon_{xc}(\tilde{n}\uparrow, \tilde{n}\downarrow) \tilde{n} d\mathbf{r}. \tag{3.20}$$

The approximate charge density $\tilde{n}_s$ is given by:

$$\tilde{n}_s(\mathbf{r}) = \sum_k d_{k,s} h_k(\mathbf{r}), \tag{3.21}$$

where h_k are simple Gaussian functions, and $d_{k,s}$ are chosen to minimise:

$$\int \{n_s(\mathbf{r}) - \tilde{n}_s(\mathbf{r})\}^2 d\mathbf{r},$$

yielding the following equations:

$$\sum_l H_{kl} d_{l,s} = \sum_{ij} u_{ijk} b_{ij,s}, \tag{3.22}$$

where,

$$H_{kl} = \int h_k(\mathbf{r}) h_l(\mathbf{r}) d\mathbf{r}, \tag{3.23}$$

$$u_{ijk} = \int \phi_i(\mathbf{r} - \mathbf{R}_i) \phi_j(\mathbf{r} - \mathbf{R}_j) h_k(\mathbf{r}) d\mathbf{r}. \tag{3.24}$$

Consider now the more simplistic spin averaged case where the spin labels can be dropped. The resulting approximation to the exchange-correlation energy is given by:

$$\tilde{E}_{xc} = \sum_k d_k \int h_k(\mathbf{r}) \epsilon_{xc}(\tilde{n}) d\mathbf{r}. \tag{3.25}$$

The choice of Gaussian orbitals for the functions h_k results in this integral being proportional to the averaged value of ϵ_{xc} under these functions k:

$$\int h_k(\mathbf{r}) \epsilon_{xc}(\tilde{n}) d\mathbf{r} \propto \langle \epsilon_{xc}(\tilde{n}) \rangle_k. \tag{3.26}$$

Since $\epsilon_{xc}(n)$ is slowly varying with n, we make make the following approximation,

$$\langle \epsilon_{xc}(\tilde{n}) \rangle_k \approx \epsilon_{xc}(\langle \tilde{n} \rangle_k), \tag{3.27}$$

where the averaged value of $\tilde{n}$ under k are given by:

$$\langle \tilde{n} \rangle_k = \frac{\sum_l d_l \int h_k h_l d\mathbf{r}}{\int h_k d\mathbf{r}}. \tag{3.28}$$

We have effectively replaced the exact value of $\epsilon_{xc}(n)$ at $\mathbf{r}$ by the homogeneous electron gas value with average density $\langle \tilde{n} \rangle_k$.

This approximation can now be improved upon by expressing the exchange-correlation density as $\epsilon_{xc}(n) = An^{0.30917}$, and by defining a function $f(s)$, given by:

$$f(s) = \ln\Big(\frac{\langle\tilde{n}^s\rangle_k}{\langle\tilde{n}\rangle_k^s}\Big). \tag{3.29}$$

Since we require a value of $s \approx 0.3$, $f(s)$ can be approximated by:

$$f(s) = s(s-1)f(2)/2, \tag{3.30}$$

as $f(0) = f(1) = 0$ and $f(2) > 0$. The approximation requires an expression for $\langle\tilde{n}^2\rangle_k$ which is simply given by:

$$\langle\tilde{n}^2\rangle_k = \frac{\sum_{lm} d_l d_m \int h_k h_l h_m d\mathbf{r}}{I_k}, \tag{3.31}$$

and can be evaluated analytically. Here, I_k is defined by,

$$I_k = \int h_k d\mathbf{r},$$

This then leads us to our final expression - the approximate exchange-correlation energy:

$$\tilde{E}_{xc} = \sum_k d_k \epsilon_k, \tag{3.32}$$

where,

$$\epsilon_k = I_k \epsilon_{xc}(\langle\tilde{n}\rangle_k) e^{f_k}, \tag{3.33}$$

$$f_k = \frac{1}{2}s(s-1)ln\Big(\frac{\langle\tilde{n}^2\rangle_k}{\langle\tilde{n}\rangle_k{}^2}\Big). \tag{3.34}$$

A modified expression has been derived for the spin polarised case (Lister and Jones 1988 [11]) which follows the above approach, and results in similar equations with additional summations over the spin states.

3.1.3 The Kohn-Sham eigenvalue matrix equations

Using the expression and approximations derived in the last subsection, the total energy can now be written as:

$$E = \sum_{ij}\{T_{ij} + V_{ij}^{ps}\}b_{ij} + \tilde{E}_H + \tilde{E}_{xc} + E_{ion-ion}, \tag{3.35}$$

This energy is then minimised with respect to the orthonormal wavefunctions ψ_λ, ie subject to the constraint that,

$$\sum c_i^\lambda c_j^\mu S_{ij} = \delta_{\lambda\mu},$$

where S_{ij} is the overlap matrix. Lagrange multipliers E_λ are introduced for this purpose,

$$\sum_{ij\lambda} c_i^\lambda \{T_{ij} + V_{ij}^{ps} - E_\lambda S_{ij}\} c_j^\lambda + \tilde{E}_H + \tilde{E}_{xc} + E_{i-i}. \tag{3.36}$$

Performing the minimisation with respect to c_i^λ, we obtain our matrix equivalent of the Kohn-Sham equations:

$$\sum_j \{T_{ij} + V_{ij}^{ps} + V_{ij}^H + V_{ij,s_\lambda}^{xc} - E_\lambda S_{ij}\} c_j^\lambda = 0. \tag{3.37}$$

Here, we have introduced the Hartree and exchange-correlation potentials:

$$\frac{\partial \tilde{E}_H}{\partial c_i^\lambda} = \sum_j V_{ij}^H c_j^\lambda, \quad V_{ij}^H = \sum_{kl} G_{kl} c_l \frac{\partial c_k}{\partial b_{ij}}, \tag{3.38}$$

$$\frac{\partial \tilde{E}_{xc}}{\partial c_i^\lambda} = \sum_j V_{ij,s_\lambda}^{xc} c_j^\lambda, \quad V_{ij,s}^{xc} = \sum_k \{\epsilon_{k,s} + \sum_l d_{l,s} \frac{\partial \epsilon_{l,s}}{\partial d_{k,s}}\} \frac{\partial d_{k,s}}{\partial b_{ij,s}}. \tag{3.39}$$

Using equations 3.13 and 3.22, these expressions can be simplified to:

$$\sum_l G_{kl} \frac{\partial c_l}{\partial b_{ij}} = t_{ijk}, \quad \sum_l H_{kl} \frac{\partial d_{l,s}}{\partial b_{ij,s}} = u_{ijk}. \tag{3.40}$$

Equation 3.37 can now be written in terms of the generalised eigenvalue problem:

$$\sum_j (H_{ij} - ES_{ij}) c_j = 0, \qquad \text{or} \qquad (H - ES)c = 0,$$

Choleski decomposition is then used to write the overlap matrix in terms of the upper triangular matrix U, ie $S_{ij} = U^t U$, followed by the inversion of the matrix U. Both processes involve $O(N^3)$ operations. Defining $Uc = d$ and $(U^{-1})^t H U^{-1} = V$, converts the problem into the standard eigenvalue problem,

$$(V - E)d = 0.$$

The eigenvalues are found by the normal Householder method, and an inverse iteration procedure is used to calculate the eigenvectors for the occupied and lowest unoccupied states.

3.2 Calculational procedure

3.2.1 Clusters and basis sets

Now that all of the background theoretical method has been outlined, the practical implementation of the method is now discussed in detail. The first step in tackling the problem of modelling defects in semiconductors is the choice of suitable cluster for use with the localised basis set. The bulk region typically consists of between 35 and 150 atoms in a highly symmetric configuration. The atoms at the surface of this bulk material will therefore have dangling bonds whose 'surface states' give rise to levels in the pseudo-band gap which can then interact with any defect related levels. In many problems it is required that all atoms are relaxed to obtain a feel for the strain induced by a defect, the undercoordination at the surface of the starting structure would have drastic effects if this were done making the results unreliable.

Inevitablely, the approach adopted here for the modelling solid state systems involves large hydrogen terminated clusters. The hydrogen atoms are added to the undercoordinated surface atoms along the bond direction(s), and lead to the complete removal of all surface states from the pseudo band gap. Despite the fact that DFT under estimated band gaps, the calculated band gaps of materials using DFT-cluster methods are generally greater than the observed values. Gaps of 3.5 eV for Si and Ge are typical, compared the experimental values of 1.17 and 0.75 eV respectively, although the calculated values decrease slowly with increasing cluster size.

Next, a suitable cluster for the problem to be investigated must be chosen. It is essential that once a defect is inserted into a cluster, the symmetry of the whole cluster is the same as the point group symmetry of the defect. So, for example a tetrahedral atom centred cluster could be used for defects with C_{2v} and C_{3v} symmetry. Another important point is the size of cluster used. Defect surface interactions may effect the results, so in general the results should be checked in different sized clusters to ensure consistency.

Two sets of basis functions are used in AIMPRO, one to expand the wavefunction, and the second for the charge density. Pure Gaussians of the form,

$$(x - R_{ix})^{n_1}(y - R_{iy})^{n_2}(z - R_{iz})^{n_3} e^{-a_i(\mathbf{r}-\mathbf{R}_i)^2},$$

are used for the wavefunction expansion, at both atom and bond centred sites. The choice of various combinations of $n_i = 0, 1$ for example, results in s and p_x, p_y, and p_z orbitals. The coefficients a_i are determined by minimising the energy of the pseudo-atom. This minimisation also generates the coefficients c_i^λ for each valence state λ. The basis used is termed *minimal* when a fixed linear combination is used for the coefficients, along with different exponents. If the coefficients are treated as variational parameters in the calculation, then the basis is referred to as *big*. In large runs, only the defect and surrounding atoms are treated in big basis, the terminating hydrogens are always in the minimal basis. An example is as follows: Consider n_i=0,1 as above, then we have both s and p orbitals. If four different exponents are chosen then there is a total of 16 Gaussians representing the wavefunction of the atom for a big basis. The minimal basis is obtained when these 16 orbitals are contracted into four linear combinations of four Gaussians. This results in the number of orbitals per atom being reduced from 16 to 4.

For the charge density basis, either pure or modified Gaussians are used for the expansion. The charge density exponents b_i are found by maximising the approximate Hartree energy $\tilde{H}$ (recalling that this approximation is always bounded above E_H).

Finally the notation adopted throughout this thesis for describing the basis used on an atom X is as follows: X{n,m} simply means that n and m exponents were used to model the wavefunction and charge density respectively.

3.2.2 Self consistency of the charge density

The matrix Kohn-Sham equations 3.37 must now be solved self consistently, so that the input and output charge density coefficients c_k^{in}, $d_{k,s}^{\mathrm{in}}$ and c_k^{out}, $d_{k,s}^{\mathrm{out}}$ are equal to within a specific tolerance. Often the initial charge density is that of the neutral atoms, and will be vastly different from the final charge density. A weighted average of the input and output charge densities is then used as the input for the next iteration:

$$c_k'^{\mathrm{in}} = c_k^{\mathrm{in}} + w(c_k^{\mathrm{out}} - c_k^{\mathrm{in}}). \tag{3.41}$$

The same procedure is adopted for the $d_{k,s}$'s. The process of obtaining the output charge density is denoted by the non-linear operation:

$$c_k^{\mathrm{out}} = L_k(c_k^{\mathrm{in}}), \tag{3.42}$$

using the above weighting, it is found that,

$$c_k'^{\mathrm{out}} = L_k(c_k^{\mathrm{in}} + w(c_k^{\mathrm{out}} - c_k^{\mathrm{in}})),$$

providing our weighting factor is small enough, then we can approximate this by the linear equation,

$$c_k'^{\mathrm{out}} = L_k(c_k^{\mathrm{in}}) + w\sum_l D_{kl}(c_{kl}^{\mathrm{out}} - c_{kl}^{\mathrm{in}}),$$

$$c_k'^{\mathrm{out}} = c_k^{\mathrm{out}} + w\sum_l D_{kl}(c_{kl}^{\mathrm{out}} - c_{kl}^{\mathrm{in}}).$$

Where D_{kl} is given by:

$$D_{kl} = \frac{\delta L_k}{\delta c_l},$$

Self consistency is achieved when:

$$c_k'^{\mathrm{out}} = c_k'^{\mathrm{in}},$$

or

$$c_k^{\rm in} + w(c_k^{\rm out} - c_k^{\rm in}) = c_k^{\rm out} + w\sum_l D_{kl}(c_{kl}^{\rm out} - c_{kl}^{\rm in}),$$

which gives:

$$(w^{-1} - 1)(c_k^{\rm in} - c_k^{\rm out}) = \sum_l D_{kl}(c_{kl}^{\rm out} - c_{kl}^{\rm in}). \tag{3.43}$$

The right hand side of this expression is determined by the choice of a small value of w, w_1, followed by the evaluation of the output charge density, giving:

$$c_{1k}^{\rm out} = c_k^{\rm out} + w_1\sum_l D_{kl}(c_l^{\rm out} - c_l^{\rm in}), \tag{3.44}$$

and therefore,

$$\sum_l D_{kl}(c_l^{\rm out} - c_l^{\rm in}) = (c_{1k}^{\rm out} - c_k^{\rm out})/w_1, \tag{3.45}$$

Equations 3.43 and 3.45 are then used to give an expression for w which is solved by a least squares method. The difference between the sides of the equation are written in terms e_k,

$$e_k = (1 - w)(c_k^{\rm in} - c_k^{\rm out})/w - (c_{1k}^{\rm out} - c_k^{\rm out})/w_1, \tag{3.46}$$

where the electrostatic energy of the 'charge density' e_k, defined by:

$$\sum_k e_k g_k(\mathbf{r}),$$

is minimised giving the value for w. In other words, the energy defined by,

$$\frac{1}{2}\sum_{kl} e_k G_{kl} e_l,$$

is made as small as possible. The above procedure is then generalised so that the predicted charge density is constructed from those of several of the previous iterations:

$$c_k' = \sum_l \{c_{lk}^{\rm in} + w_l(c_{lk}^{\rm out} - c_{lk}^{\rm in})\}. \tag{3.47}$$

The self consistent energy converges exponentially, in general taking anywhere between four and ten iterations to achieve the convergence. Problems can arise however if there exist two levels, one occupied and the other unoccupied, which are very close in energy. In this case, the Hubbard term resulting from the occupation of the lower level results in its energy increasing to a value above the second level. This process continually repeats successively emptying one level and occupying the next after each iteration. This so called *charge sloshing* often results in the non-convergence of the charge density. clearly this is a problem as many defects can have partially occupied degenerate levels.

The way in which this problem is circumvented is by the inclusion of a *finite temperature*, so the electron charge is *smeared out* over the near by levels, by an amount governed by Fermi statistics, or by forcing partial occupancies of the troublesome levels. Considering the finite temperature approach: if the level E_λ is occupied by f_λ electrons, then the energy to be minimised if now the free energy F subject to the constraint that the total number of electrons is M, ie

$$F = E + k_B T \sum_\lambda \{f_\lambda \ln f_\lambda + (1 - f_\lambda)\ln(1 - f_\lambda)\} - \mu\{\sum_\lambda f_\lambda - M\}. \tag{3.48}$$

The sum is over all orbitals λ, minimising expression this with respect to f_λ and μ gives:

$$f_\lambda = \frac{1}{e^{(E_\lambda - \mu)/k_B T} + 1}, \qquad \text{and} \qquad \sum_\lambda f_\lambda = M. \tag{3.49}$$

The expression for the density matrix must now be generalised to:

$$b_{ij,s} = \sum_\lambda \delta(s, s_\lambda) f_\lambda c_i^\lambda c_j^\lambda. \tag{3.50}$$

Typically, a value of $k_B T = 0.04$ eV is chosen for the energy. It should be noted that the use of this temperature term is only an aid to the convergence of the charge density, and is not an attempt to simulate the material at finite temperatures. Secondly, care must be taken in the analysis of the results when the finite temperature option is used, as this can lead to the incorrect structures, for example it would remove the energy saving from a Jahn Teller distortion.

3.2.3 Structural optimisation

Once we have obtained the self consistent charge density, we have the lowest electronic energy for the particular atomic configuration. This however will not be the ground state energy of the whole system. This is because the atoms themselves will not be in their ground state configuration as the forces on them may be non zero. We now have to move the atoms to lower the energy of the system until the forces on each atom are essentially zero. This is again an iterative scheme, and the self consistent charge density must be found for each new set of atomic positions. First of all, an analytic expression for the forces on each atom must be found. The atoms are then moved in the direction of the forces by a conjugate gradient algorithm until the the equilibrium structure is found. It must be noted that the ground state configuration is not found once the energy of a particular configuration is minimised, and all forces are zero. This is because the defect may be bistable or meta-stable, possessing several low energy configurations, all of which need to be examined.

Evaluation of the atomic forces

The force acting on an atom a, in the direction l is given by:

$$f_{la} = -\frac{\partial F}{\partial R_{la}} = -\frac{\partial E}{\partial R_{la}}. \tag{3.51}$$

Inserting this into the previous expression (equation 3.36) gives an expression for the change in energy when the atoms are moved from their positions R_{la} by ΔR_{la}. This gives:

$$\Delta E = \sum_{ij} b_{ij}\Delta\{T_{ij} + V_{ij}^{ps}\} + \sum_{ij}\{T_{ij} + V_{ij}^{ps}\}\Delta b_{ij} + \Delta\tilde{E}_H + \Delta\tilde{E}_{xc} + \Delta E_{i-i}, \tag{3.52}$$

where the changes in the Hartree and exchange correlation energies are given by:

$$\Delta\tilde{E}_H = \sum_{kl} c_k G_{kl}\Delta c_l + \frac{1}{2}\sum_{kl} c_k c_l \Delta G_{kl}, \tag{3.53}$$

$$\Delta\tilde{E}_{xc} = \sum_{k,s} \epsilon_{k,s}\Delta d_{k,s} + \sum_{k,s} d_{k,s}\Delta\epsilon_{k,s}. \tag{3.54}$$

Equations 3.13 and 3.22 can now be used to evaluate the terms Δc_k and $\Delta d_{k,s}$, giving,

$$\sum_l G_{kl}\Delta c_l = \sum_{ij}\{t_{ijk}\Delta b_{ij} + b_{ij}\Delta t_{ijk}\} - \sum_l c_l \Delta G_{kl}, \tag{3.55}$$

and

$$\sum_l H_{kl}\Delta d_{l,s} = \sum_{ij}\{u_{ijk}\Delta b_{ij,s} + b_{ij,s}\Delta u_{ijk}\} - \sum_l d_{l,s}\Delta H_{kl}. \tag{3.56}$$

We can now write the terms containing Δb_{ij} as:

$$\Delta E' = \sum_{ij}\{T_{ij} + V_{ij}^{ps} + V_{ij}^{H}\}\Delta b_{ij} + \sum_{ijs} V_{ij,s}^{xc}\Delta b_{ij,s}. \tag{3.57}$$

and reduce this expression to the Pulay force in the following way. Now, from the expression for the Kohn-Sham equations 3.37 we find that this term can be re written as:

$$\Delta E' = \sum_{ij\lambda} E_\lambda S_{ij}\Delta c_i^\lambda c_j^\lambda = \sum_\lambda E_\lambda \Delta\{\sum_{ij} c_i^\lambda c_j^\lambda S_{ij}\} - \sum_{ij\lambda} E_\lambda c_i^\lambda c_j^\lambda \Delta S_{ij}. \tag{3.58}$$

The first term in the final expression vanishes, leaving the change in energy with displacement given by:

$$\Delta E' = -\sum_{ij\lambda} E_\lambda c_i^\lambda c_j^\lambda \Delta S_{ij}, \tag{3.59}$$

and hence an analytic expression of the forces has been derived:

$$\Delta E = \sum_{ij} b_{ij}\Delta\{T_{ij} + V_{ij}^{ps}\} - \sum_{ij\lambda} E_\lambda c_i^\lambda c_j^\lambda \Delta S_{ij} + \Delta\tilde{E}_H + \Delta\tilde{E}_{xc} + \Delta E_{i-i}, \tag{3.60}$$

and,

$$f_{la} = \frac{\Delta E}{\Delta R_{la}}. \tag{3.61}$$

Conjugate gradient optimisation

Now that we have an expression for the forces on each atom, an method must be found to move the atoms in the direction of the forces until the structure is optimised. The approach adopted in AIMPRO uses a standard conjugate gradients technique to move the atoms. This approach is very efficient as it used knowledge of previous search directions in moving the atoms along with the latest set of forces. Denoting the force on each atom a by f'_{la} in the l direction, then the atoms are moved along a *conjugate* direction, d'_{la} so that the new positions are given by:

$$R'_{la} = R_{la} + w d'_{la}, \tag{3.62}$$

where w is chosen such that the Free energy in equation 3.48 is minimised, via a cubic of quadratic interpolation scheme. The new search direction is related to the new forces f'_{la} as well as the previous forces f_{la} and search directions d_{la} by:

$$d'_{la} = f'_{la} - \frac{\sum_{la} f'_{la}(f'_{la} - f_{la})}{\sum_{la} f_{la}^2} d_{la}. \tag{3.63}$$

For the first iteration, all atoms are simply moved in the direction of the forces. This method results in a quadratic convergence to the minimum energy configuration.

It is important to constrain the symmetry of the defect during the optimisation. The constraints can be removed afterwards to ensure that the results are physical. This is necessary as many symmetries can give rise to degenerate local vibrational modes and gap levels, and if the full symmetry is not constrained, then the resulting calculations will not yield results with the correct degeneracies. There is also the added advantage of a potentially large saving in computational time when high symmetries are exploited.

3.3 The calculation of experimental observables

We are now in a position to calculate the ground state properties of defects, and relay to the experimentalist the structure of the ground state. This is of course a valuable result in itself but we must go further. For example this result may be of little use when two proposed models for a defect are very close in energy. We could not say with certainty which was the ground state structure if the energy difference of between structures A and B was as little as 0.25 eV. There is also the question of bi-stability which may further complicate the issue. We really need to provide more information directly related to experimental observables to discriminate between structures. This remaining part of this chapter is associated with just that: the calculation of the observables associated with any particular structure investigated.

The observables themselves are discussed in detail in the next chapter, here I limit myself to a few brief comments regarding their nature, and then go on to describe how we model them theoretically.

3.3.1 Migration, reorientation and binding energies

Other experimental observables that are relatively easily to calculate (at least an upper bound to!) are the migration, reorientation and binding energies of a defect. It is important to be able estimate the energy at which a defect becomes mobile, as well as the migration path as once a defect starts to migrate, it will interact with other defects and impurities present creating more complex secondary defects. The calculated binding energy of the two initial defects in the secondary defect is also a valuable quantity, and is relatively straight forward to calculate. There are three basic approaches that can be taken to estimate the migration or reorientation path, each technique has its relative merits when applied to a specific problem. All of these involve constraining the relaxation procedure in some way to force the structure away from the ground state, and therefore will give an upper bound to the saddle point energy. These are best illustrated with simple examples, and are discussed in turn below. The method adopted to calculate the defect binding energy is then discussed.

Symmetry constraints

If one has a very good idea for what the saddle point is, and the saddle point has a certain point group symmetry, then it is possible to force the defect to adopt the saddle structure by imposing a symmetry constraint.

Consider the restructuring of a vacancy hydrogen complex (VH), where the hydrogen moves and bonds to one of the other bulk neighbours (illustrated in Figure 3.1 below).

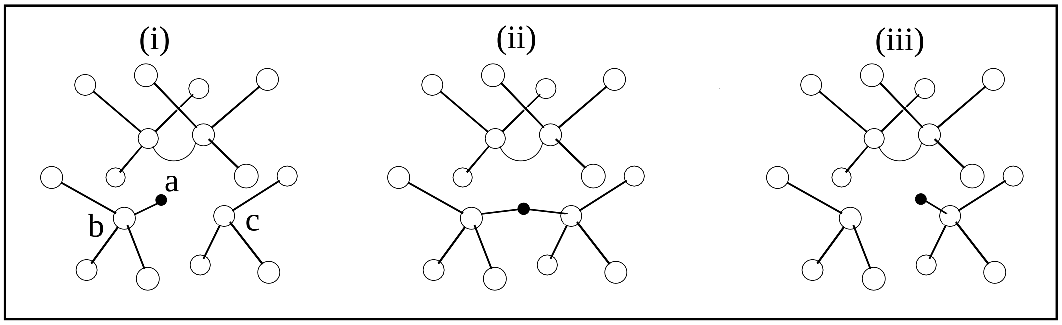


Figure 3.1: The likely saddle point for the reorientation of a VH complex.

The initial and final points are (i) and (iii) respectively, equivalent configurations possessing (almost) C_{3v} symmetry. When H reorients, the most likely saddle point is one where the H lies mid-way between two bulk atoms (ii). This configuration possesses C_{2v} symmetry - and therefore constraining the symmetry to be exactly this, and relaxing all atoms will give an estimation of the saddle point energy. This result is a first approximation and should then be checked using one of the other diffusion methods.

The relative bond length constraint

The example of H restructuring in VH is used again to highlight the approach of this method. The restructuring now involves the breaking of the bond $a-b$ and the formation of a new bond $a-c$. A suitable constraint would then involve the relative bond lengths, $|\mathbf{R}_a - \mathbf{R}_b|$ and $|\mathbf{R}_a - \mathbf{R}_c|$, and all atoms would be relaxed using this constraint. The actual constraint used is:

$$x = (\mathbf{R}_a - \mathbf{R}_b)^2 - (\mathbf{R}_a - \mathbf{R}_c)^2 = \text{constant}. \tag{3.64}$$

This results in a linear equation for one of the Cartesian components l of atom a which can be solved for any value of x,

$$R_{la} = \frac{2\sum_{k\neq l}(R_{kb} - R_{kc})R_{ka} + \sum_k (R_{kb}^2 - R_{kc}^2) - x}{2(R_{lb} - R_{lc})}, \tag{3.65}$$

if the atoms are subsequently moved, the change in energy is given by:

$$\Delta E = -\sum_{m,d} f_{md}\Delta R_{md}. \tag{3.66}$$

Using the vector identity,

$$(\mathbf{R}_c - \mathbf{R}_b).\Delta\mathbf{R}_a = (\mathbf{R}_a - \mathbf{R}_b).\Delta\mathbf{R}_b - (\mathbf{R}_a - \mathbf{R}_c).\Delta\mathbf{R}_c, \tag{3.67}$$

the term ΔR_{la} in the expression for ΔE can be written in terms of $\Delta R_{ka}, k \neq l$, $\Delta\mathbf{R}_b$ and $\Delta\mathbf{R}_c$ and hence be eliminated from the expression. This results in a modification of the forces on the atoms a, b, c (for $k \neq l$) but the changes are straightforward to evaluate. A value of l is chosen to maximise the value of the denominator in the expression for R_{la}. A number of constraints can be used simultaneously for more complex problems.

The orthogonal relaxation method

The procedure for determining the migration/reorientation barrier for defects can also be found using a constraint which forces the relaxation along orthogonal directions to fractional vectors of the initial and final configurations. The method should be applied to problems where the saddle point for the possible migration paths are not clear. Vector displacements of equivalent atoms from the initial and final structures are determined, $\mathbf{v}_i = \mathbf{R}_i^1$ - $\mathbf{R}_i^2$, where $\mathbf{R}_i^1$ and $\mathbf{R}_i^2$ refer to the staring and ending atomic configurations. All atoms were then moved along the vector $\mathbf{v}_i$ by a fraction α to give the intermediate points in the migration, and then all atoms were allowed to relax from these positions subject to a further constraint that the vector of further displacements are perpendicular to $\mathbf{v}_i$, until the optimisation was complete. The result of this is a profile of the energy as a function of α for the restructuring of the defect, and the barrier for migration/reorientation corresponds to the maximum value of energy obtained, at the corresponding saddle configuration.

Defect binding energies

Following the migration of defect A and subsequent trapping to defect B, it is useful to be able to say something about the stability of the new complex AB. The calculation of the binding energy of two defects is straightforward enough: the ground state energy of AB is calculated in a large cluster, and compared to a calculation of the energy in an identically sized cluster, where A and B are now in their isolated ground state configurations and separated by several bulk nearest neighbours. These calculations can be very time consuming as typically clusters of 200-300 atoms need to be used to obtain a sufficiently large defect-defect separation.

3.3.2 Local vibrational modes

One of the most common features observed directly associated with a defect are the local atomic vibrations of the involving the defect and surrounding atoms. These are often well above the Raman edge in both Si and Ge and are easily observed by techniques such as IR spectroscopy and PL.

To model the vibrations of the cluster, we need to construct the dynamical matrix, which contains the information about all of the force constants. Where here we have expanded the energy $E(R)$ in a Taylor series about the equilibrium values R_0, ie:

$$E(R) = E(R_o) + \frac{1}{2}\sum_{la,mb} E_{la,mb}\Delta R_{la}\Delta R_{mb} + \text{higher order terms}, \tag{3.68}$$

where $E_{la,mb}$ is the dynamical matrix of the cluster is defined by:

$$E_{la,mb} = \frac{1}{\sqrt{M_\alpha M_\beta}}\Big(\frac{\partial^2 E}{\partial R_{la}\partial R_{mb}}\Big). \tag{3.69}$$

Here, R_{xy} denotes an atom at site y displaced in the direction x. We now make the harmonic approximation, and solve the following eigenvalue equation to obtain the normal mode frequencies ω^α and their associated eigen vectors v^α_{mb}:

$$\sum_{mb} E_{la,mb} v^\alpha_{mb} = \omega^2_\alpha v^\alpha_{la}, \tag{3.70}$$

We now need some method of calculating all of the force constants of the cluster and setting up the dynamical matrix. We first consider the atoms in and around the defect core. As indicated by the expression for the dynamical matrix, the second derivative of the energy E with respect to atomic displacements R_{la} and R_{mb} must be calculated. Atom a is now displaced along a Cartesian direction $\pm l$ by an amount ϵ and the self consistent charge densities found. Typically a value of ϵ=0.25 a.u. is used. The forces on each atom are then re-evaluated, with the movement of atom a creating a non-zero force of magnitude $f^{\pm}_{mb}(l,a)$ on atom b. The second derivatives are then approximated by the finite difference formula:

$$D_{la,mb} = (f^+_{mb}(l,a) - f^-_{mb}(l,a))/2\epsilon, \tag{3.71}$$

up to second order in ϵ. The frequencies generated from this method are termed *quasi harmonic*, because $D_{la,mb}$ is not an infinitesimal derivative. This is a very time consuming step, involving repeatedly moving atoms and re-calculating the self consistent charge density.

For the rest of the atoms in the cluster, a valence force potential due to Musgrave and Pople [12] is used to construct the elements of the dynamical matrix. The potential for atom a is given by:

$$\begin{aligned} V_a = \quad & 1/4 \sum_b k_r^{(a)} (\Delta r_{ab})^2 + r_0^2/2 \sum_{b>c} k_\theta^{(a)} (\Delta\theta_{bac})^2 + r_0 \sum_{c>b} k_{r\theta}^{(a)} (\Delta r_{ab} + \Delta r_{ac}) \Delta\theta_{bac} \\ & + \sum_{c>b} k_{rr}^{(a)} \Delta r_{ab} \Delta r_{ac} + r_0^2 \sum_{d>c>b} k_{\theta\theta}^{(a)} \Delta\theta_{bac} \Delta\theta_{cad}, \end{aligned} \tag{3.72}$$

where Δr_a and $\Delta\theta_{bac}$ represent the distance between atoms a and b and the $b-a-c$ bond angles respectively. The sums are over the nearest neighbours of atom a only. Table 3.1 gives the coefficients $k_r^{(i)}$, $k_\theta^{(i)}$, $k_{r\theta}^{(i)}$, $k_{rr}^{(i)}$ and $k_{\theta\theta}^{(i)}$ for the materials investigated in this thesis.

Table 3.1: Musgrave Pople parameters (eV/Å^2) for the bulk semiconductors considered in this thesis. Values for r_0 are 2.35 and 2.41 Å for Si and Ge respectively.

Atom type	k_r	k_θ	$k_{r\theta}$	k_{rr}	$k_{\theta\theta}$
Si	10.187	0.290	0.103	0.072	0.000
Ge	7.661	0.323	-0.075	0.911	0.105

This approximation is justifiable for two reasons, firstly and most importantly, the local modes of a defect are generally well above the Raman peak, the highest bulk modes are well reproduced and disagreement at lower frequencies is irrelevant. However for local modes close to the Raman edge the motion of surrounding bulk atoms is significant and should be taken into account. Secondly, as mentioned before it would be impractical to calculate double derivatives of all atoms. The mass of all terminating hydrogens is set to 1000 atomic numbers so that they do not partake in any vibrations, ie this is a boundary condition that the surface atoms are fixed.

For the case of light impurities, for example hydrogen in semiconductors, calculated (quasi-harmonic) local modes can be significantly in error when compared to the experimental values. In these cases, the amplitude of the atomic vibrations is larger (a consequence of the lighter mass), and therefore terms higher than quadratic order in the vibronic potential become important. That is to say anharmonic effects are appreciable, with the asymmetry shifting the modes and modifying the selection rules. The way in which we treat anharmonicity is somewhat crude: first of all the quasi harmonic modes are found along with their eigenvectors, the atomic positions are then stepped along the eigenvector directions, and the total energy recalculated. The resultant curve is the variation in energy with the amplitude of atom displacements for a given local mode - each local mode is therefore treated separately and information about mode coupling is lost. The value of the local mode including the anharmonic terms is found by solving the 1D Schrödinger equation for the vibronic potential. An example where the method has been applied successfully is the C-H stretch mode in GaAs [13].

3.3.3 Effective charges and IR absorption ratios

Leigh and Szigoti (1967) [14] derived an expression for the integrated intensity of a local vibrational mode in terms of the induced effective charge η , the velocity of light c, the refractive index n, and M and ρ the mass and concentrations of the impurity atoms present. This expression is as follows:

$$I = \frac{2\pi^2\rho}{ncM}\eta^2, \tag{3.73}$$

and both I and η can be measured experimentally. The effective charge is related to the induced dipole moment P_i in the direction of the polarisation of the field, and the corresponding normal coordinate Q_α of the mode by:

$$\eta^2 = \sum_\beta M \left(\frac{\partial P_i}{\partial Q_\alpha^\beta}\right)^2. \tag{3.74}$$

To calculate η then, the change in the adiabatic dipole moment of the cluster is found when the atoms are displaced by their normal coordinates,

$$\pm\frac{Q_\alpha u_{ij}^\alpha}{\sqrt{M}}, \tag{3.75}$$

which were obtained from the dynamical matrix 3.3.2. The induced dipole moment $\mathbf{P}$ is then evaluated from the changes in the dipole moment of the cluster with these displacements via the following expression:

$$\mathbf{P} = \sum_i Z_i \mathbf{R}_i - \int \mathbf{r} n(\mathbf{r}) d\mathbf{r}. \tag{3.76}$$

η depends on the nature of the local mode itself, the polarisation of the electric field $\mathbf{E}$, and the orientation of the defects. Therefore η^2 must be averaged if the defects giving rise to a particular mode are randomly oriented.

3.3.4 Gap levels and pseudowavefunction localisation

The final property of defects in semiconductors that we are able to model are associated with the gap states that the defect give rise to. Experimentally, the gap level positions can be determined, along with information regarding the symmetry and localisation of states. There is then the possibility of electronic transitions between states, where one or both of the states is associated with the defect. As was stated previously 2.6.4, the Kohn-Sham wavefunctions Ψ_{KS} cannot be interpreted as single particle states, but in spite of this the KS eigenvectors and values do seem to have some physical meaning. Accepting that information about real systems can be obtained from the KS levels, even if the reasons for this are somewhat obscure, we can begin to make predictions/comparisons in relation to the electronic activity of a defect.

If a defect is know to give rise to several gap levels, then the calculations often show the same number of levels, with correct ordering, although the actual positions of the levels is often in serious error. There are several post calculational approaches that can be used to try to correct the errors in gap level and the size of the calculated gap. The first is to simply scale the calculated gap to the experimental band one, so any gap levels maintain their relative positions; and these rarely agree with experiment. Another approach is the so called scissor operator where the unoccupied states of the *perfect* cluster are forced upwards by an amount V, where V is chosen to give the correct band gap. The method adopted here to calculate transition energies, as well as the positions of donor/acceptor levels is the transition method due to Slater (1960) [15].

The method proceeds by determining the defect ionisation energy using a fractional occupancy of the levels involved. The theory is exact, in the sense that it is formulated in terms of the total energies of two identical structures which have N and $N = 1$ electrons respectively, and therefore their total energies are given correctly. The assumptions here are that the structure is valid for both charge states, and that the cluster size is large enough so that effects of the finite cluster size are small.

The proof consists of two parts, the first of which is due to Slater [15], who derived an expression for the ionisation energy, given by:

$$E_I = E(N+1, \mathbf{R}) - E(N, \mathbf{R}), \tag{3.77}$$

in terms of semi-occupied levels. Expanding the above expression as a Taylor series about an intermediate state with occupations such that both levels contain $\frac{1}{2}e$, the ionisation energy can be expressed as:

$$E_I = \left. \frac{\partial E(m, \mathbf{R})}{\partial m} \right|_{m=N+\frac{1}{2}} . \tag{3.78}$$

In order to evaluate this ionisation energy within the LDA, an expression for this partial derivative is necessary. Janak [16], was able to derive an expression for this derivative, by expanding the charge density in the following way:

$$\rho(\mathbf{r}) = \sum_{i}^{N} n_i |\psi_i(\mathbf{r})|^2, \tag{3.79}$$

accounting for fractional occupancies. The partial derivative of the total energy with respect to n_i, reduced to the following trivial expression:

$$\frac{\partial E}{\partial n_i} = \epsilon_i. \tag{3.80}$$

Equation 3.80 is known as Janaks theorem.

Using a combination of equations 3.78 and 3.80, the defect ionisation energy can be expressed in terms of the Kohn-Sham eigenvalues of the intermediate state ie.

$$E_i = \epsilon_{N+1}\left(\frac{1}{2}\right) - \epsilon_N\left(\frac{1}{2}\right). \tag{3.81}$$

A defect acceptor (-/0) level, relative to E_c is therefore estimated by occupying both the defect level and the conduction band level by $\frac{1}{2}e$, the transition energy being given by the difference. An identical approach is used for the donor (0/+) levels, relative to E_v, but here the donor and conduction band levels are each occupied by half a hole. In the calculations, effects due to structural relaxation of the differently charged defects were ignored, that is to say, the optimised positions for one charge state (often neutral) were used in all calculations. One problem encountered is due to the magnitude of the calculated band gap: for both Si and Ge, the cluster approach yields values between 3 and 4 eV for the gap. The approach adopted here is to linearly scale the levels down to the experimental gap for the material under investigation. An estimate for the accuracy of the method is given in [17] where the calculated levels of Au in Si are within 0.2-0.3 eV of those observed.

Our attention is now turned to the nature of the gap states themselves. In many cases it is necessary to obtain information about the localisation and symmetries of gap states. The symmetries enable us to detail which electronic transitions are dipole allowed, and the localisations enable us to say whether a given transition will have an isotope shift, as well as describing the the magnitude of any hyperfine interaction expected in a magnetic resonance experiment. This information is obtained by simply plotting Ψ_{KS} in the relevant planes. One problem is that the amplitude of Ψ_{KS} is very small in the region close to the nucleus, a consequence of the pseudopotential. Mulliken populations m_i^λ described by:

$$\sum_s \int \psi_\lambda^2(\mathbf{r}, s) d\mathbf{r} = \sum_{ij} c_i^\lambda c_j^\lambda S_{ij} = \sum_i m_i^\lambda, \tag{3.82}$$

where

$$m_i^\lambda = c_i^\lambda \sum_j S_{ij} c_j^\lambda,$$

are used to obtain more accurate information about the localisation of any state.

Finally, AIMPRO is also able to obtain an order of magnitude estimate for the radiative lifetime of a particular transition. The rate of electronic dipole transitions between two states λ and μ is described by the following expression (Svelto 1976) [18]:

$$\frac{1}{\tau_{\lambda\mu}} = \frac{4n\omega^3}{3c^3\hbar} \frac{e^2}{4\pi\epsilon_o} |\mathbf{r}_{\lambda\mu}|^2, \tag{3.83}$$

where

$$\mathbf{r}_{\lambda\mu} = \sum_s \int \psi_\lambda(\mathbf{r}, s) \mathbf{r} \psi_\mu(\mathbf{r}, s) d\mathbf{r},$$

is the dipole matrix element. ω, c, n and e are the transition frequency, speed of light, refractive index and electronic charge respectively. Recent applications of the method are described in [19].

Chapter 4

Experimental investigations of defects

4.1 Introduction

The introduction of a defect or impurity into a semiconductor has three basic effects on the properties of the host. Firstly, electronic levels of varying depth (< 0.1 eV and > 0.1 eV from the nearest band edge termed shallow and deep respectively) are often introduced into the forbidden gap, secondly the phonon density of states is affected, and thirdly, there may be magnetic effects associated with the defect.

The first of these - the introduction of gap levels plays an important role in semiconductor technology. Silicon for example is often doped with boron (p-type) or phosphorus (n-type), resulting in shallow acceptor and donor levels respectively, and improving the conductance of the material. The origins of deep levels is complex, and they can behave as carrier traps, or as generation-recombination centres if around mid-gap. Their effects are often detrimental to device physics as they tend to compensate the shallow levels by capturing the free carriers and hence reducing the effective doping density. In addition to this, as will be seen later, kinetic defect reactions can also degrade devices. However, a high defect concentration can also be useful: for example GaAs can be made semi-insulating, a necessity for a semiconductor substrate. The deep levels can also be specifically used to compensate unintentionally introduced shallow levels. The presence of deep levels is a useful way to provide a pathway for the generation and recombination of electron-hole pairs across the whole of the band gap. This can have a detrimental effect as it can limit LED and laser efficiency, but can also be helpful as it provides a way of controlling minority carrier lifetimes. Electronic transitions can occur between defect levels and the conduction or valence bands, or between two defect levels themselves. Clearly then this leads to absorption and emission of photons with below band-gap energy, $E < E_g$. These can either be electronic or magnetic dipole associated transitions.

An acceptor level (-/0) is defined as the difference in total energy of the defect when an additional electron is inserted into the empty conduction band state E_c, and when this electron occupies the lowest empty level of the defect, E_d. Similarly, the donor level (0/+) represents the difference in energy when a hole is promoted to the defect level from the valence band, E_v. Mathematically, the acceptor level position relative to E_c is given by equation 4.1:

$$\mathrm{E_{TOT}}[E_d(e), E_c(0)] - \mathrm{E_{TOT}}[E_d(0), E_c(e)] \tag{4.1}$$

and the donor level relative to E_v is given by:

$$\mathrm{E_{TOT}}[E_d(0), E_v(h)] - \mathrm{E_{TOT}}[E_d(h), E_v(0)] \tag{4.2}$$

A point defect destroys the lattice symmetry, changing the normal modes of vibration. Three cases exist, (i) all the introduced modes are below the Raman peak, and (ii) some modes appear above the maximum lattice modes and are termed local, as they cannot propagate through the lattice, and (iii) gap modes can exist, between the acoustic and optic bands in compound semiconductors. The frequencies of the introduced modes depend on the mass differences between the defect and host atoms as well as the local force constants, dependent on the bonding. Roughly speaking, if the difference between masses and force constants is small (with the defect atom being more massive than the host), then all of the modes are below the Raman peak. Conversely, for light impurities or large force constants (relative to the host), the modes lie above the Raman edge (local modes). For light impurities such as hydrogen and carbon in silicon or germanium the latter are particularly important. Under certain conditions though, pseudo-local vibrational modes can arise as resonances of sub-Raman defect related vibrational modes.

This chapter is concerned with the theoretical description of several experimental techniques employed to detect the above features and hence categorise a particular defect. Optical experiments (absorption, photoluminescence and infra-red spectroscopies [20]) are first described, followed by deep level transient spectroscopy [21], and finally electron paramagnetic resonance [22]. After outlining each method, the effects of uniaxial stress are briefly discussed as it allows the experimentalist to determine more information regarding the particular defect. For example, monitoring the splittings of spectral features from a given centre with applied uniaxial stress provides information about the symmetry of the defect. Finally the experimental analysis of multi/bi-stable defects in silicon is discussed in detail.

In general, for a complete description of a defect, several techniques must be employed. For example, annealing studies used to correlate a spectra observed in EPR with a gap level seen in DLTS. These can also be correlated with PL lines and IR modes, leading to a complete picture of the defect - the structural model. Several features can be correlated when they appear and disappear at the same temperatures during annealing, with further strength being provided from isotope effects linking a common impurity atom to the signals. Theoretical modelling is vital when the experimental data is inconclusive, and often, a combination of experiment and theory can lead to the structural model. Observing the loss of the signal associated with a given defect upon annealing gives information about the migration or dissociation energy of a defect. Exactly which process is taking place can be determined from the products of a given defect reaction once identified.

4.2 Absorption and photoluminescence spectroscopy

4.2.1 Introduction

This section is concerned with the response of a typical semiconductor after illumination with light of frequency ν, and the subsequent electron-photon interaction. A material is transparent to light of a particular frequency if $\hbar\nu > E_g$ where E_g is the band gap. Therefore insulators and (ideal) large band gap semiconductors are transparent to visible light, whereas silicon and germanium are transparent at shorter (IR) wavelengths. Of course the situation is somewhat different in reality, the defects present in the material give rise to levels which lie in the forbidden gap, and these result in selective absorption and subsequent emission of light with $\hbar\nu < E_g$. Therefore, the influence of a particular defect can be probed by optical measurements, and these can lead to information on the concentration of a defect, as well as the nature of the electronic states involved in the transition. Photoluminescence is sensitive to down to defect concentrations of around $\approx 10^{12}$ cm^{-3}.

4.2.2 Absorption

The process of absorption is the optical excitation of an electron from the ground state structure into an excited state, here, specifically involving a defect related level. Optical absorption measurements probe the strength of absorption (given by the absorption coefficient μ) as a function of the incident photon energy. In terms of the radiation intensity I, the absorption coefficient is defined as:

$$\mathrm{d}I = -\mu I \mathrm{d}t, \tag{4.3}$$

where dI is the fractional decrease in intensity in a region of thickness dt. Therefore,

$$\mu = \frac{2.303}{t}\log_{10}\left(\frac{I_0}{I}\right). \tag{4.4}$$

This is in turn related to the oscillator strength f by:

$$\int \mu(E)\mathrm{d}E = \frac{2\pi^2 e^2 \hbar}{nmc^2}\left(\frac{E_0}{E_{\text{eff}}}\right)^2 Nf, \tag{4.5}$$

where N is the defect concentration per unit volume, E_0 the average electric field in the material, and E_{eff} is the local electric (effective) field at the defect. The solution of equation 4.5 is of the form:

$$Nf = \text{constant} \times \frac{10^{17} n}{(n^2+2)^2}\mu_m W, \tag{4.6}$$

where n is the refractive index of the semiconductor, μ_m the band peak absorption, and W is the band width at half the peak height. Therefore if the oscillator strength of the transition is known, the defect concentration can be computed from μ. Provided that the optical transition is dipole allowed (equation 4.12) then absorption reveals all of the excited states of the system under observation. The absorption is measured by subtracting the transmitted spectrum from the original incident continuum. The transmitted light through a sample of thickness d is then given by:

$$I = I_0 \frac{(1-R)^2 \mathrm{e}^{-\mu d}}{1 - R^2 \mathrm{e}^{-2\mu d}}, \tag{4.7}$$

where I_0 is the incident intensity of the light, μ is the absorption coefficient, and R is the reflectivity given by $R = [(n-1)/(n+1)]^2$, assuming that the extinction coefficient is approximately zero.

To understand the selection rules in the optical transition between two states (absorption considered below, but the same holds for luminescence), one has to consider the optical cross section given by:

$$\sigma(h\nu) = \frac{\beta}{h\nu}\left|\langle\psi_f|\left(-i\hbar\frac{\partial}{\partial z}\right)|\psi_i\rangle\right|^2 \delta(E_f - E_i - h\nu), \tag{4.8}$$

where $\beta = (4\pi^2 e^2 \hbar/m^2 c)$ and z is the component of the position vector $\mathbf{r}$ along $\mathbf{A_0}$ - the vector potential of the incident radiation. In fact the vector potential is given by $\mathbf{A} = \mathbf{A_0}\cos(\mathbf{g}.\mathbf{r} - \omega t)$. A transition is therefore forbidden if this matrix element vanishes:

$$\langle\psi_f|\left(-i\hbar\frac{\partial}{\partial z}\right)|\psi_i\rangle = 0, \tag{4.9}$$

this condition can be written in terms of the component of the particle vector position $\mathbf{r}$ using the following relationship:

$$[H, z] = -\frac{\hbar^2}{2m}\frac{\partial}{\partial z}. \tag{4.10}$$

The matrix element condition for the transition is then reduced to:

$$-\frac{2m}{\hbar^2}\langle\psi_f|z|\psi_i\rangle(E_f - E_i) = 0, \tag{4.11}$$

and therefore for the transition to be allowed, the irreducible representation (containing z), is contained in the product:

$$\Gamma(\psi_i) \times \Gamma(\psi_f), \tag{4.12}$$

of the representations of the initial and final states.

4.2.3 Photoluminescence (PL)

Following the electronic excitation from the absorption, the electron can subsequently return to the ground state, losing energy by a combination of photon and phonon emission. The processes of absorption and luminescence are shown schematically along with the configurational coordinate diagram, figure 4.1.

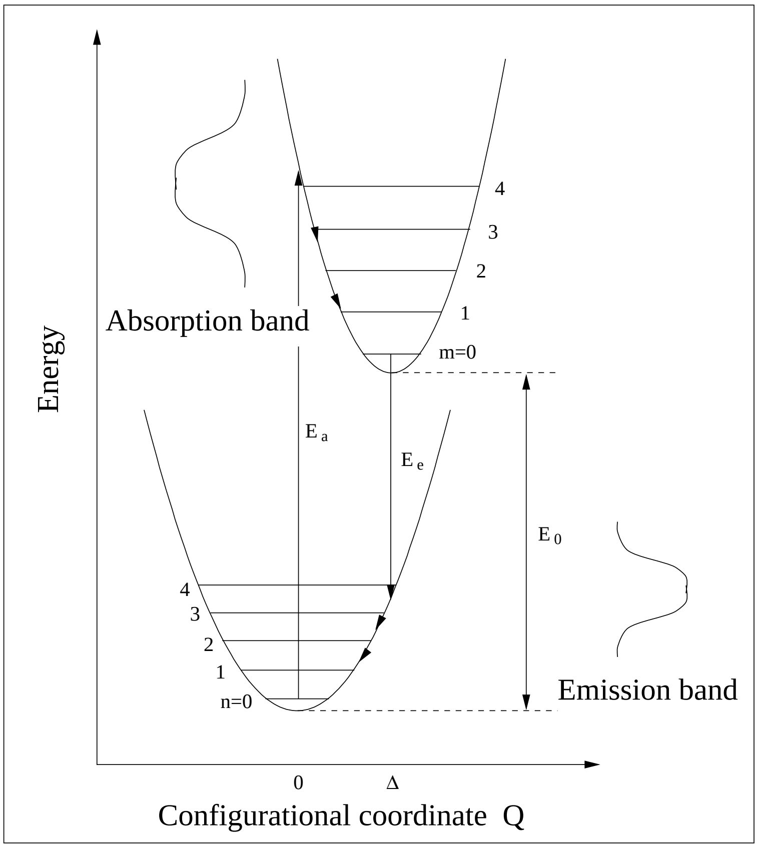


Figure 4.1: the configurational coordinate diagram associated with the absorption and emission transitions, with the $Q = 0$ representing the ground state, and $Q = \Delta$ the excited state.

In a typical photoluminescence experiment, electrons hole pairs are created by illumination with above band gap light ($h\nu > E_g$). Alternatively, electric fields (electroluminescence) or a beam of incident electrons (cathodoluminescence) can be used as the excitation source. The relaxation to the ground state, electro-hole recombination, occurs via both radiative and non-radiative mechanisms, generally involving the defect states. The selection rules involving the many-body wavefunctions of the relevant energy levels, $\Psi_f \rightarrow \Psi_i$, are exactly as indicated by equation 4.12, for the electronic wavefunctions ψ_i. Clearly then this imposes strict selection rules on the electronic transition, and therefore a transition between an B_1, and a B_2 state in C_{2v} symmetry for example is dipole forbidden. Note that a transition may be weakly allowed if the the integral in equation 4.12 is non-zero for higher order (*eg.* quadrupole) operators, although the transition probability will be extremely small meaning that the radiative lifetime is long.

Photoluminescence is a complex process, depending on temperature, defect concentration, capture cross sections, electron-phonon coupling; and in general cannot be used to determine the concentration of a defect. It is possible to estimate this under certain conditions: namely in the case of free exciton donor and acceptor luminescence. If this is the case, the donor and acceptor concentrations are simply proportional to the ratio of the zero phonon line peak height to the free exciton recombination peak height.

Information concerning the chemical species of a defect can sometimes be observed in photoluminescence by isotopic substitution. This can have the effect of shifting the zero-phonon line by a small amount (a result of changing the zero point energies figure 4.1) and hence changing the transition energy. This is not always the case, and in some cases can prove to be misleading - it is dependent on the nature of the electron-phonon coupling, and although extensively used for chemical categorisation, it does not always reflect the atomic structure of the centre.

More accurate information concerning the chemical constituents of a defect come from phonon-sidebands sometimes observed along with the zero-phonon line. The luminescence peak is much broader than those observed in atomic spectra, because of the influence of the surrounding lattice. In fact, the energy difference between the the peaks in the absorption and emission bands is a consequence of the excitation of lattice modes in the transition process. Referring to the configurational coordinate diagram (figure 4.1), based on the adiabatic Born-Oppenheimer approximation, the energies of the atomic vibrations in terms of ν_g and ν_e for the ground and excited states are given by:

$$E_g = (n + \frac{1}{2})\hbar\nu_g, \tag{4.13}$$

and,

$$E_e = (m + \frac{1}{2})\hbar\nu_e + E_0, \tag{4.14}$$

where n and m are quantum numbers, and E_0 is the difference between the two energy levels when $n = m = 0$. Note that here, for simplicity, the vibronic energy surface is shown for *one* phonon only in the electronic ground and excited states. Different phonons would be presented by other parabolae centred on $Q = 0$ and $Q = \Delta$ for the electronic ground and excited states respectively. Generally, the many body wavefunction in the excited state is more diffuse than that of the ground state, and therefore the position of the minimum of the nuclear coordinate occurs at a different position $Q = \Delta$ than the ground state potential. Invoking the Frank-Condon principle that the lattice does not respond in the time that the electronic transition takes place, these are then represented by vertical lines in the configurational coordinated diagram. The transition energy varies as the lattice vibrates resulting in broadened absorption and emission lines. The lattice relaxes to $m = 0$ in the excited state before luminescence takes place. Therefore $E_a > E_e$ the difference being known as the *Stokes shift*, and the Stokes shift is bigger if a larger number of phonons are involved in a particular transition.

When an appreciable overlap exists between $\psi_f(Q)$ and $\psi_i(Q)$ then it becomes possible to excite the transition, and observe emission with energy E_0, without emission of phonons - the so called *zero phonon line*. The probabilities of transitions between ψ_0 and ψ_m involving S phonons is given by:

$$P_{0m} = \frac{e^{-s} S^m}{m!}, \tag{4.15}$$

and a similar expression holds for luminescence in terms of n. Clearly then, if S (known as the Huang-Rhys factor) is greater, then the probability of observing the zero phonon line is reduced and the number of observed phonon side bands is increased. For example in absorption, for large S the optical line shape is a Gaussian centred on $S = m$. In photoluminescence, the sideband structure accompanying the zero phonon line is directly related to the localised vibrational modes of the defect. Therefore when observed, the replicas can be used as a more precise probe of the chemical nature of the defect because of the isotope effects on the localised modes themselves. It must be noted that extreme care must be taken in these experiments as the replica must be measured relative to the zero phonon line, which can itself shift with varying isotope. In addition to this, *only* vibrational modes with A_1 symmetry can be observed in the photoluminescence experiment. This is because the transition is governed by the following matrix elements:

$$\langle \psi_f \chi_0 | z | \psi_i \chi_1 \rangle, \tag{4.16}$$

where the transition is from the vibronic ground state χ_0 of the electronic excited state ψ_f to the χ_1 phonon level of the electronic ground state ψ_i. Now, the vibronic ground state of any electronic state has A_1 symmetry, and for a transition to take place the following condition must hold:

$$\Gamma(A_1) \times \Gamma(\text{phonon}) \supset \Gamma(A_1), \tag{4.17}$$

this being the case, the symmetry of the excited phonon mode must also be A_1.

Finally, more information can be obtained from the photoluminescence experiment in the presence of an external magnetic field. If this is the case, for transitions involving a defect in which $S \neq 0$ a Zeeman splitting of the involved levels results in a number of luminescence lines. Information can then be obtained on the spin state, charge state (if the constituents are known) and possibly the symmetry of the defect, from and analysis of the splittings (or lack of them).

4.3 Infra-red spectroscopy (IR)

4.3.1 Introduction

In section 4.2, the photon-electron interaction gave rise to selective absorption and emission, therefore allowing photoluminescence to primarily monitor electronic properties of the defect. In some cases however phonon modes could be detected as phonon replicas of the zero phonon line. We now consider the direct interaction between a photon and a phonon in the semiconductor.

If the pure crystal is considered, then both the crystal momentum and energy must be conserved. Therefore single-phonon processes can only occur close to $q = 0$ since the photon momentum is small. Processes involving more than one phonon can occur if the sum of their momenta is close to zero. In bulk monopolar semiconductors, no first order absorption occurs since there is no induced dipole (shortly we shall see that this is a necessary condition for absorption). This is not the case in compound semiconductors where the large induced dipole results in a strong absorption near $\omega_{\max}$ known as the Reststrahl band. Of interest here are the local vibrational modes that light impurities introduce into semiconductors, and in particular the most commonly employed spectroscopic technique for their direct detection. The direct excitation of a phonon means that contrary to PL where the mode was detected as a phonon replica associated with A_1 symmetry, modes of all symmetries can now be detected. The sensitivity of the technique is dependent on the sample size, but typically the lowest detectable defect concentrations are $\approx 10^{-15}$ cm^{-3}.

4.3.2 Infra-red absorption

Infra-red absorption measurements proceed by monitoring the differences in the absorptions in a variety of samples, with differing defect concentrations. In doing this, compensation for different sample sizes must also be included in the analysis.

The local vibrational modes are infra-red active when the vibration leads to an induced dipole moment (equations 4.18 and 4.19), which itself is frequency dependent. In addition to these, the band modes are perturbed due to the loss of translational symmetry. Multi-phonon absorption can also occur due to anharmonicity, but are generally less important than the single-phonon modes. Consider the coupling of the electronic vector of the incident radiation to the dipole moment P. Writing the equilibrium dipole moment as P_e, the dipole moment for small displacements U can be written as:

$$P = P_e + \left(\frac{dP}{dU}\right)_e U + \left(\frac{d^2P}{dU^2}\right)_e \frac{U^2}{2} + ... \tag{4.18}$$

The probability of a transition is proportional to $|\langle\Psi_i|P|\Psi_f\rangle|^2$ and therefore there is no coupling of P_e with the vibrational states. For one phonon transitions, the intensity is proportional to $[(dP/dU)_e]^2$ and hence the magnitude of $(dP/dU)_e$ - the apparent charge, η - can be determined experimentally. It must also be stated that local field corrections increase the absorption by a factor of $\wedge = (n^2+2)^2/9$, giving an enhancement of around 20 for silicon. The apparent charge η is then related to the effective charge e^* by $\eta^2 = \wedge e^{*2}$. The integrated absorption for a particular phonon mode is given by:

$$I = \int \alpha(\omega)d\omega = \frac{2\pi^2\rho\eta^2}{ncM}. \tag{4.19}$$

Here, c, and n have their usual meanings, ρ and M are the concentrations and masses of the impurity present, and hence the intensity depends on the square of the apparent charge. Note the similarities between this equation and 4.5 where the absorption involves electronic rather than purely vibronic states. An identical expression to equation 4.7 also holds for the for the transmission intensity, but here α replaces μ. Therefore, the symmetry of a centre cannot be determined from the ratio of intensities of several local modes. Although they will have degeneracies relating to the point group of the defect, their induced dipoles and hence effective charges could vary substantially from mode to mode. Differing defect charge state can also effect the position of the local modes as the structure changes with charge state. The effect is most pronounced for defects with deep levels, where the wavefunction is highly localised and hence varying its occupation is more likely to alter the structure than if it were a more diffuse shallow level.

Absorption at well defined wavelength occurs for a local mode ($\omega > \omega_{\max}$) and the peaks tend to be much broader when a resonant mode is detected. The chemical constituents of the defect can again be examined using different isotopes, and observing their affects on the local vibrational modes. To first order, assuming the local force constants k', which are in general different to the bulk force constants, the LVM frequency, ω_{L}, is given by:

$$\omega_{\mathrm{L}}^2 = 2k' \left[\frac{1}{m_{\mathrm{imp}}} + \frac{1}{\chi M_{\mathrm{H}}} \right], \tag{4.20}$$

where m_{imp} and M_{H} are the masses of the impurity and host atoms respectively, and χ is a parameter depending on the local bond bending and stretching force constants. The shifts in LVM with isotopes can then be approximated by:

$$2\frac{\delta\omega_{\mathrm{L}}}{\omega_{\mathrm{L}}} = \frac{\delta m_{\mathrm{imp}}}{m_{\mathrm{imp}}}. \tag{4.21}$$

Therefore the mode decreases with increasing nuclear mass as expected. The above only holds for simple defects (for example substitutional defects where only one mode is present), since in more complex defects, a large number of modes exist and substantial coupling leading to anti-crossings can occur. Ultimately this can even lead to the anomalous upwards shift of a local mode with heavier isotope. In a semiconductor where the host atom has a number of common isotopes, the small splittings in a local mode can reflect the symmetry of the defect.

4.4 Deep level transient spectroscopy

4.4.1 Introduction

DLTS is a well established and extremely powerful spectroscopic technique, used in the study of deep (> 0.1) eV levels. It is based upon the variations in capacitance of a Schottky barrier or *pn*-junction and their relation to the number of carriers trapped in deep levels in the space charge (depletion) region. DLTS provides important information such as the trap activation energy E_T, concentration N_T, and capture cross section σ which gives information on the ability of a trap to capture carriers. DLTS has a much greater sensitivity than any of the 'single shot' experimental methods such as thermally stimulated capacitance (TSCAP), because it employs signal averaging of repeated forward bias injection and quiescent reverse bias stages, which ultimately result in a sensitivity of around 10^{12} defects per cubic cm.

4.4.2 Carrier traps and capacitance transients

Consider a deep level in the forbidden gap, introduced by a defect. If the trap concentration is denoted by N, and the capture and thermal emission rates for electrons and holes are given respectively by c_1, e_1 and c_2, e_2 then the steady state occupation of the levels can be given by:

$$n_1 = \frac{c_1}{c_1 + c_2}N, \quad \text{and} \quad n_1 = \frac{e_1}{e_1 + e_2}N, \tag{4.22}$$

for the electron concentration during the capture and emission process, and,

$$n_2 = \frac{c_2}{c_1 + c_2}N, \quad \text{and} \quad n_2 = \frac{e_2}{e_1 + e_2}N, \tag{4.23}$$

is applicable to the hole concentration. These are shown schematically in 4.2a. The capture rate c_n is related to the capture cross section of the defect σ, n_c the density of free carriers, and their mean thermal velocity $\langle v^{\mathrm{th}} \rangle$ by:

$$c_n = \sigma n_c \langle v^{\mathrm{th}} \rangle, \tag{4.24}$$

and the emission rates are very small compared to the capture rates during the filling period. The thermal emission rate from the traps are governed by Boltzmann statistics and are given by:

$$e_n = \sigma_{\mathrm{ap}} N_D \langle v_{\mathrm{th}} \rangle \exp\left(\frac{-E_T}{kT}\right), \tag{4.25}$$

where σ_{ap} is the apparent capture cross section, N_D is the effective density of states in the appropriate carrier band and E_T is the thermal activation energy for the emission of the carrier. Equations 4.24 and 4.25 are applicable to both electrons and holes and strictly speaking should be written as $c_{i,n}$ and $e_{i,n}$, where $i = 1, 2$, along with subscripts on the constants.

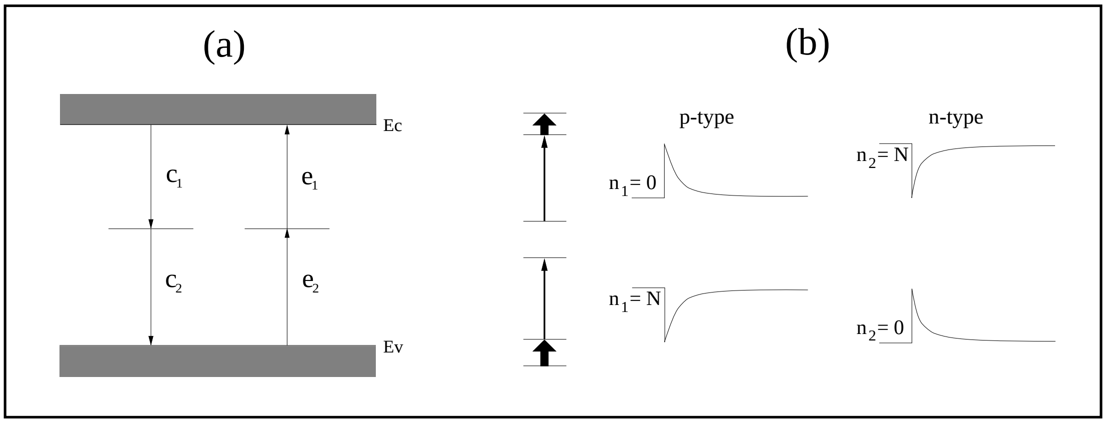


Figure 4.2: (a) Carrier capture and emission processes involving a deep level: c_1, e_1 and c_2, e_2 are the capture and emission rates for electrons and holes respectively, and (b) Occupation of minority and majority carrier traps following the injection pulse in both n- and p- type material.

In the reversed bias quiescent state, the observable traps are in the depletion region and the capture rates are zero. Therefore, the occupation of the trap is governed by e_1 and e_2. The variation in trap concentration after a injection of carriers is shown in figure 4.2b. An electron trap tends to be empty of electrons $n_1 = 0$, $n_2 = N$ whereas a hole trap tends to be full of electrons $n_1 = N$, $n_2 = 0$. Therefore an electron trap has $e_1 >> e_2$, with $e_2 >> e_1$ being applicable to the hole trap.

The changing occupation of the traps back towards thermal equilibrium (illustrated in figure 4.2 leads to a capacitance transient, ΔC. This transient is also an exponential function of time, with the rate constant given by $e_1 + e_2$ and one of these usually dominates. The capacitance transient is always positive for a minority carrier trap, and negative for a majority carrier trap, and is similar in appearance to trap occupations shown above. Therefore a given trap is only observed after the injection of the appropriate carrier. The non equilibrium conditions are achieved by the injection of carriers with an appropriate biasing pulse, which fills the trap. Following this, emission takes place in the reversed bias quiescent state leading to the changes in capacitance, which are detected in the experiment.

Both $p - n$ junctions and Schottky barriers are used to perform the DLTS measurement. Consider the case of a Schottky barrier, the depletion region W for this system is given by:

$$W = \left[\frac{2\epsilon(V_d - V)}{qN}\right]^{\frac{1}{2}}, \tag{4.26}$$

where N is as defined above, ϵ is the permitivity of the semiconductor, q the electronic charge, and V_d and V are the diffusion and applied potentials, the latter being positive for forward bias. The barrier acts as a parallel plate capacitor with capacitance C given by:

$$C = \frac{\epsilon A}{d} = \left[\frac{q\epsilon N A^2}{2(V_d - V)}\right], \tag{4.27}$$

so a reverse bias leads to a decrease in W and therefore an increase in the capacitance, consistent with the increase in minority carrier concentration.

4.4.3 DLTS measurement techniques

The main feature of the DLTS method is the rate window in which the capacitance transient is measured. If the transient is measured at times t_1 and t_2 after the initial injection pulse, and at varying temperatures, then a profile similar to that in figure 4.3 is observed. The change in capacitance occurring within the rate window is temperature dependent, so both the rate window and sample temperature are varied in the experiment.

If the rate window is defined as $\tau_{\max}^{-1}$, then the maximum of $C(t_1) - C(t_2)$ against T is is found in the following way: firstly, a normalised signal $S(T)$ is defined, with

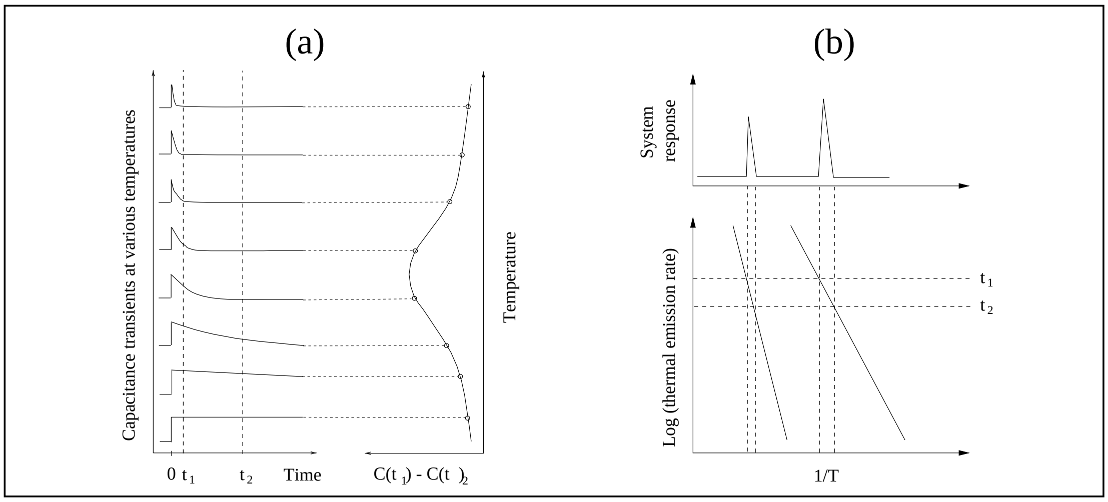


Figure 4.3: (a) The capacitance transient variation with temperature for a given rate window, and (b) schematic diagram of the system response for the same rate window (two traps shown here).

respect to a $\Delta C(0)$ – the change in capacitance due to the pulse at $T = 0$.

$$S(T) = [C(t_1) - C(t_2)]/\Delta C(0). \tag{4.28}$$

Assuming exponential transients, this is given by:

$$S(T) = [\exp(-t_1/\tau)] - [\exp(-t_1/\tau)] = \exp(-t_1/\tau)[1 - \exp(-\Delta t/\tau)], \tag{4.29}$$

with $\Delta t = t_2 - t_1$, and finally, the maxima is given by:

$$\frac{dS(T)}{d\tau} = 0 \longrightarrow \tau_{\max} = (t_1 - t_2)[\ln(t_1/t_2)]^{-1}, \tag{4.30}$$

and therefore the position of this peak is a precisely defined quantity in the measurement, and varies with the rate window. Since one of the emission processes usually dominates, $\tau_{\max}^{-1} = e_i$. The schematic diagram in figure 4.3 show the variation of the transient and the system response with temperature for a given rate window.

An estimate for the trap or defect concentration comes from the height of the induced capacitance transient ΔC created by the pulse, and the quiescent reverse bias capacitance C by the following equation:

$$N_T = 2\frac{\Delta C}{C}\frac{\lambda^2}{N_c W^2}, \tag{4.31}$$

where N_c is the free carrier density of states in the relevant band, and λ is the distance from the metal point to the position in the depletion region where the trap crosses E_F.

The trap activation energy can be determined from the temperature variation in the thermal emission rate of the trap, by plotting the logarithm of this rate against

$1/T$. This of course assumes that the prefactors in equation 4.25 are temperature independent. This is not the case, as $v \propto T^{\frac{1}{2}}$ and $N_D \propto T^{\frac{3}{2}}$, and therefore to omit these effects, an Arrhenius plot of the form:

$$\ln\left[\frac{e_i}{T^2}\right] \quad \text{v} \quad \frac{1000}{T}, \tag{4.32}$$

is constructed. Note that:

$$\ln\left[\frac{e_i}{T^2}\right] = \chi - \frac{E_T}{kT}, \tag{4.33}$$

where χ is a constant related to $\ln\sigma$. The value of the trap energy is generally determined to within a few percent using this method.

The final quantity which can be determined by DLTS is the trap capture cross section σ. A rough estimate for σ can be determined from χ, the y-axis intercept ($T = \infty$) of the Arrhenius plot, equation 4.33. However, this method is not very accurate as it can give an apparent capture cross section which is often very different from the real cross section. The following more accurate approach is adopted in the determination of σ: the diode is forward biased for a short time resulting in a partial occupation of the (majority carrier) traps. Traps with large σ capture more carriers in the bias period than those with a smaller value of σ. The injection time is then increased until the value of ΔC is constant and therefore the trap is completely filled during the pulse time. Once this saturation has been achieved, and the characteristic filling time τ_{fill} has been determined, σ can be found from:

$$\sigma = \frac{1}{\tau_{\text{fill}} v n}, \tag{4.34}$$

where n is the effective doping density. Therefore, an Arrhenius plot of peak height against pulse length is used in the determination of σ.

4.4.4 Related methods

There are a number of techniques based on DLTS, these are now briefly discussed.
Double DLTS: (DDLTS) is essentially the same as DLTS with the exception that *two* filling pulses are used to fill the traps. The magnitude of the voltage bias is different for the two pulses, meaning that they fill traps in different regions ($\lambda - \lambda_1$ and $\lambda - \lambda_2$), and so there are two sets of transient that need to be analysed. Processing of the transients allows an investigation of only the traps in the region $\lambda_2 - \lambda_1$. By adjusting the bias and pulse length, it is possible to analyse traps from a narrow zone in the depletion region. Profiles of the whole depletion width, or the field dependence of the trap can be investigated in this manner.
Optical DLTS: (ODLTS) is a technique in which the filling of the levels is controlled

by optical excitation. Illumination with a below band gap laser results in photoionisation of states, and when the light source is removed, an emission transient is observed in the usual way. Improving on this, deep level optical spectroscopy (DLOS) uses an illumination source of a well defined wavelength to selectively fill a trap of known optical activation energy.

Minority carrier transient spectroscopy (MCTS) is able to probe minority carrier traps as well as the majority carrier traps observed in the standard DLTS. MCTS uses optical excitation, but is different from ODLTS and DLOS in that excess carriers are generated with a flooding illumination. The excitation energy is therefore above band gap here, and is incident on a sample held under reverse bias. Some of the minority carriers created in this process are then captured by minority carrier traps and upon cessation of the excitation, are thermally emitted and give rise to a capacitance transient. A combination of both DLTS and MCTS can be used to characterise majority and minority carrier traps in the material.

4.5 Electron paramagnetic resonance (EPR)

4.5.1 Introduction

EPR is an extremely powerful technique, based on the splittings of defect related levels in magnetic fields, due to both the electron and nuclear spins. The defect must possess a non zero electronic spin, and hence a magnetic dipole moment if it is to be EPR active. The centre is termed *paramagnetic* if the defects are present in low enough concentrations so that they do not interact. In the EPR experiment, excitation of electrons between the Zeeman split levels is achieved by exposure to electromagnetic radiation of frequency ν. Resonant absorption is achieved when the incident photon energy $h\nu = \Delta E$, the Zeeman splitting of the magnetic dipoles. The technique is sensitive to the defect symmetry, the localisation the defects highest occupied wavefunction (along with its symmetry), and the nuclear spin of all atoms in the defect an its immediate vicinity. EPR is also a very sensitive method, being able to detect defects with concentrations as low as 10^{13} cm^{-3}.

4.5.2 Zeeman splitting of the levels

Denoting the electron spin, orbital angular momentum and nuclear spin by **S**, **L**, and **I** respectively, then under the influence of an applied magnetic field, the Hamiltonian can be written as:

$$\begin{aligned} H &= H_0(1 \text{ eV}) + H_{\mathbf{LS}}(10^{-2} \text{ eV}) + H_{\mathbf{SS}}(10^{-4} \text{ eV}) + H_{\mathbf{SI}}(10^{-6} \text{ eV}) + \\ &\quad H_{\mathbf{LI}}(10^{-6} \text{ eV}) + H_{Ze}(10^{-4} \text{ eV}) + H_{Zn}(10^{-7} \text{ eV}), \end{aligned} \tag{4.35}$$

where $H_{\mathbf{LS}}$ is the contribution of the spin-orbit interaction to the total energy and so on. The relative importance of these terms is illustrated by normalising the unperturbed electron energy ($\mathbf{H_0}$) to 1 eV.

The electronic Zeeman interaction

Consider now an applied magnetic field $\mathbf{H}$ on the levels of a paramagnetic defect. Defining $\mathbf{J} = \mathbf{L} + \mathbf{S}$, then the Zeeman Hamiltonian can be written as:

$$H_{Ze} = g\beta \mathbf{H} \cdot \mathbf{J}, \tag{4.36}$$

where β is the Bohr magneton defined by:

$$\beta = \frac{e\hbar}{2mc}, \tag{4.37}$$

and g is the well known Lande factor (equal to g_0=2.0023 for a free electron). The shifts in energy are then given in terms of m_j, ($m_j = -J, -(J-1), ..., J$) by:

$$\Delta E_{Ze} = g\beta|\mathbf{H}|m_j, \tag{4.38}$$

and therefore the level splits into $(2J+1)$ levels, with equal spacings of $g\beta|\mathbf{H}|$. Normally the angular momentum is quenched, so that its component along $\mathbf{H}$ is zero, $\langle\psi|L_z|\psi\rangle = 0$, and the spin Hamiltonian reduces to:

$$\Delta E_{Ze} = g\beta|\mathbf{H}|m_s. \tag{4.39}$$

In the case of an electron bound to a defect, g deviates from the free electron value and the Hamiltonian (quenched) can be written as:

$$H_{Ze} = \beta\mathbf{H}\cdot(g'\mathbf{S}), \tag{4.40}$$

where $g' = g_0 + \Delta g'$. In fact because the anisotropy g' is a tensor, and it is the analysis of this which gives the defect symmetry, since it reflects the environment of the defect. Generally the g tensor is an ellipsoid with principle axes g_1, g_2 and g_3 dependent on the local symmetry. In an EPR measurement on a tetrahedral semiconductor, the sample is rotated about $\langle 110\rangle$ chosen to be perpendicular to $\mathbf{H}$. This enables g to be obtained along the all crystallographic directions ($\langle 100\rangle$, $\langle 110\rangle$, and $\langle 111\rangle$) within a 90° rotation of the sample. The principle axis of the defect can then be trivially deduced.

The hyperfine interaction

For a paramagnetic defect, a further splitting known as the hyperfine splitting occurs, a result of the interaction between the electron spin and the spin of neighbouring nuclei. This is given by:

$$H_{HF} = \mathbf{S}\hat{A}\mathbf{I} = \frac{8\pi}{3}g\beta g_n\beta_n|\psi(0)|^2\mathbf{I}\cdot\mathbf{S} - g\beta g_n\beta_n\left[\frac{(\mathbf{I}\cdot\mathbf{S})}{r^3} - \frac{3(\mathbf{I}\cdot\mathbf{r})(\mathbf{S}\cdot\mathbf{r})}{r^5}\right], \tag{4.41}$$

where $\hat{A}$ is a scalar (tensor) satisfying $\hat{A} < g\beta\mathbf{H}$, and $\psi(0)$ is the wavefunction associated with the unpaired electron. The first term in equation 4.41 is isotropic whereas the later term is not. This results in a further splitting of the levels by:

$$\Delta E_{HF} = A m_s m_i, \tag{4.42}$$

where m_i is the nuclear azimuthal quantum number which takes the values ($m_i = -I, -(I-1), ..., I$). These hyperfine structures are very sensitive to the defects wavefunction, as well as the atoms involved in the defect. Since the nuclear spin is characteristic to a particular atom, the spectrum provides a signature for the atoms in the defect, as well as giving the character of the wavefunction. The character of the wavefunction is obtained from the fact that the coupling of the electron and nuclear magnetic moments consists of two terms (equation 4.41).

The nuclear Zeeman interaction

Finally, there is the interaction between the nuclear magnetic moment $\mathbf{I}$ and the applied magnetic field giving rise to the nuclear Zeeman splitting:

$$H_{Zn} = -g_n\beta_n\mathbf{H}\cdot\mathbf{I}, \tag{4.43}$$

which leads to splittings of:

$$\Delta E_{Zn} = -g_n\beta_n|\mathbf{H}|m_i. \tag{4.44}$$

Figure 4.4 shows the splittings of a level for which $\mathbf{S} = \mathbf{I} = \frac{1}{2}$, here the interaction is with the nucleus of *one* atom. In reality, the spectra are extremely complex as the splittings due to interactions with all surrounding nuclei are super imposed upon one another.

Achieving spin resonance

Transitions between the Zeeman split levels (equation 4.38) are governed by the selection rule $\Delta m_j = \pm 1$, and are driven by photons of energy:

$$h\nu = g\beta|\mathbf{H}|, \tag{4.45}$$

where the resonance is observed. Without photon excitation, the occupancy of the levels is simply governed by Maxwell-Boltzmann statistics, whereas the absorption of photons tends to equilibriate the population of the levels. A resonance can be detected if there is a population difference between the levels. The sample is placed in an alternating magnetic field to observe the resonance, and the incident photon frequency is constant. The absorption is often very weak, and superimposed on

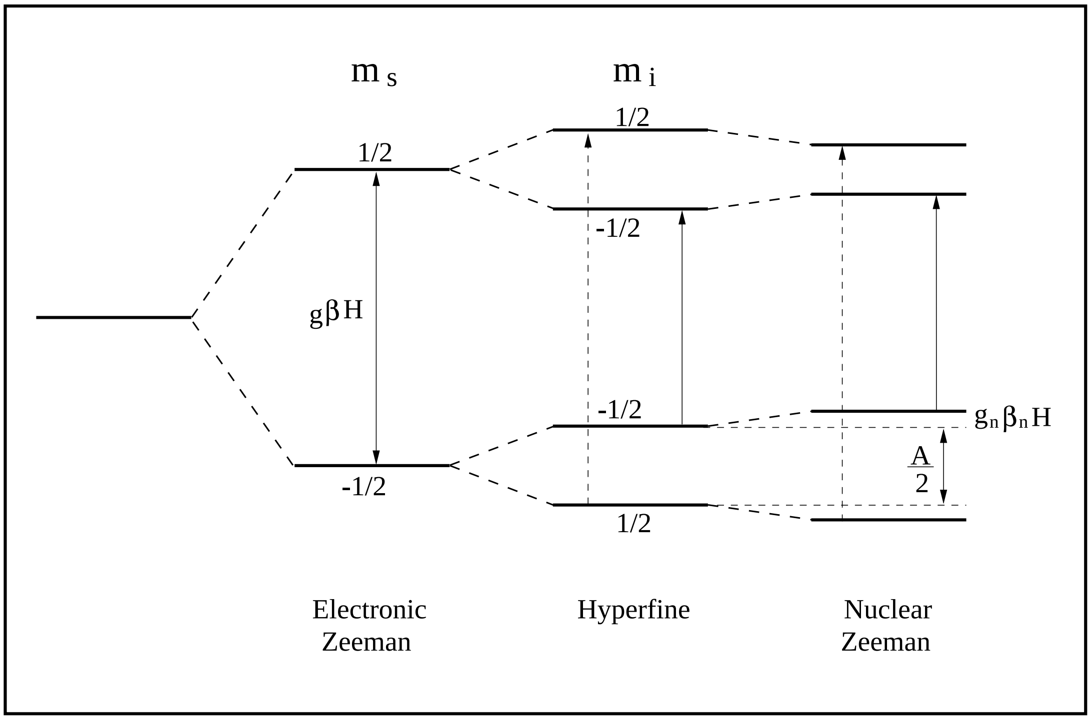


Figure 4.4: Splitting of energy levels for the various interactions for a defect with $\mathbf{S} = \mathbf{I} = \frac{1}{2}$.

a large slowly varying background signal, to get around this, a second alternating magnetic field is applied to enhance the signal. It should also be noted that it is possible to probe defects at different spatial positions in the sample by the application of a linearly increasing magnetic field across the sample. Defects in different positions will be brought into resonance at different times in the experiment in a similar way to magnetic resonance imaging in medical physics.

A closely related technique to EPR is optical detection of magnetic resonance (ODMR). Here, a defect which may not be EPR active is driven into an excited state by optical excitation, and transitions are then induced between the Zeeman levels of the excited state. These transitions are not detected in the same way as in EPR, but instead by the changes in luminescence intensity or polarisability, and the experimental set up is simpler than in EPR.

4.5.3 Electron nuclear double resonance (ENDOR)

Further information may be obtained by ENDOR, a resonant technique involving both the electron and nuclear spins. Referring to figure 4.4 the nuclear Zeeman term causes a splitting of $g_n\beta_n|\mathbf{H}|$, depending on the sign on m_i. It is this nuclear magnetic resonance which is observed in ENDOR, the technique proceeds as follows: microwave excitation is used to equalise the populations of the levels at the frequency of resonance, which leads to the cancellation of the EPR signal.

Radio-frequency excitation at,

$$h\nu_n^{(1)} = \frac{A}{2} - g_n\beta_n|\mathbf{H}|, \tag{4.46}$$

is used to induce a transition between the two levels at $m_s = \frac{1}{2}$. Since this modifies the populations of the levels, the EPR spectrum re-emerges. Conversely transitions between the $m_s = -\frac{1}{2}$ level can be driven by photons of energy,

$$h\nu_n^{(2)} = \frac{A}{2} + g_n\beta_n|\mathbf{H}|. \tag{4.47}$$

If the EPR signal can be saturated in the way described above, then ENDOR can provide a direct measurement of the nuclear magnetic moment and the hyperfine constant, given by $h(\nu^{(2)} - \nu^{(1)}) = 2g_n\beta_n|\mathbf{H}|$, and $h(\nu^{(2)} + \nu^{(1)}) = A$ respectively.

4.6 The application of stress

4.6.1 Introduction

The application of uniaxial stress to a defect is a common and extremely useful technique which can be used to determine two things. The application of stress lifts the oriental degeneracies of a symmetric defect and hence (i) it splits the spectral lines (IR, PL, and DLTS, when observed) - examination of these splittings can be used to determine a defects symmetry along with the polarisation properties of transitions, and (ii) for an anisotropic defect, the application of stress at high temperatures followed by cooling results in the preferential alignment of the defects - allowing the kinetics of reorientation to be determined by a symmetry sensitive technique such as EPR.

4.6.2 Determination of the defect symmetry

EPR is the main tool employed to determine the symmetry of a defect, however, this is not possible for a great many centres either because they are not EPR active, or their associated EPR spectra has not been detected. There are other methods which can directly give the point group of a centre, however the use of applied uniaxial stress is one of the most common methods used to determine the symmetry of a defect. The point group symmetry of a defect can be determined by monitoring the splittings in spectral features observed in IR of PL (LVM frequencies or the zero phonon line), with the application of uniaxial stress. Here, the uniaxial stress removes part or all of the degeneracies associated with a particular defect. When stresses are applied along the simple crystallographic directions, the number of split components along with their relative intensities and polarisations can be understood in terms of simple geometrical arguments. Kaplyanskii [23], obtained the following expression for the stress induced splitting of a given set of spectral features (a and b) say:

$$\begin{bmatrix} a & v \\ v & b \end{bmatrix}, \tag{4.48}$$

where,

$$a = A_1\sigma_{zz} + A_2(\sigma_{xx} + \sigma_{yy}) + 2A_3\sigma_{xy} + 4A_4(\sigma_{yz} - \sigma_{zx}), \tag{4.49}$$

$$b = B_1\sigma_{zz} + B_2(\sigma_{xx} + \sigma_{yy}) + 2B_3\sigma_{xy} + 4B_4(\sigma_{yz} - \sigma_{zx}), \tag{4.50}$$

$$v = V_1\sigma_{zz} + V_2(\sigma_{xx} + \sigma_{yy}) + 2V_3\sigma_{xy} + 4V_4(\sigma_{yz} - \sigma_{zx}). \tag{4.51}$$

A_i and B_i are the splitting parameters for the spectral features (which may be a zero-phonon line and its excited state, or two local modes), and V_i are the parameters accounting the interaction between them. In the usual notation σ_{ij} are the stress tensor components for the applied stress, representing the jth component of the force per unit area on a plane whose outward normal is parallel to the positive x_i direction. The uniaxial stress measurements are generally made with compressions along the main crystallographic directions ($\langle 100\rangle$, $\langle 110\rangle$, and $\langle 111\rangle$), and the splittings fitted to the above. The values of the parameters enables the determination of the symmetry of the centre giving rise to a and b.

Consider the case of a defect with C_{2v} symmetry, giving rise to a spectral feature S, which can be a local mode or zero phonon line. Then, under applied uniaxial stress, for this point group the shift in S is given by:

$$\Delta S = A_1\sigma_{zz} + A_2(\sigma_{xx} + \sigma_{yy}) + 2A_3\sigma_{xy}, \tag{4.52}$$

and therefore S splits into 2, 3, and 2 components for stresses along $\langle 100\rangle$, $\langle 110\rangle$, and $\langle 111\rangle$ respectively. The symmetry of the local modes themselves can also be derived through analysis of the splittings with polarisation.

Another effect of the uniaxial stress on photoluminescence lines comes from the lowering of the defect symmetry and splitting of the levels. The dipole matrix elements depend on the polarisation, and hence different transitions can be induced depending on whether the excitation is applied parallel or perpendicular to the stress. As a result of this, information on the defects symmetry can be derived.

4.6.3 Determination of the reorientation energy

The stress induced alignment experiment proceeds as follows: in the unstressed structure, defects aligned with their principle axis along all crystallographic directions allowed by its point group are all equally likely and there is a random distribution of the centres. Now under applied stress, the alignment for a symmetric defect is no longer random, certain directions are preferred. The preferential alignment can be frozen in by lowering the temperature of the sample with applied stress, followed by the removal of the stress at low temperatures (relative to the temperature at which the defect reorients). For an anisotropic defect in a tetrahedral semiconductor, the change in energy for an applied stress σ is written as:

$$\Delta E_i = Tr[\mathbf{A}.\sigma_i], \tag{4.53}$$

or in terms of the strain ϵ:

$$\Delta E_i = Tr[\mathbf{B}.\epsilon_i], \tag{4.54}$$

where i refers to the orientation of the defect and σ_i or ϵ_i are the tensors for the applied stress or strain, resolved in the direction of the principal axis of the defect (i). The tensors **A** and **B** are known as the *piezo-spectroscopic* tensors, and these describe the coupling properties of the defect to the stresses or strains in the defects principal axis system. Under applied stress, when the defect can reorientate, a Maxwell-Boltzmann distribution of orientations is achieved, given by:

$$\frac{n_i}{n_j} = \exp\left[\frac{(\Delta E_j - \Delta E_i)}{kT}\right]. \tag{4.55}$$

Therefore by applying stresses along a few crystallographic directions, the traceless components of the piezo-spectroscopic tensor can be determined from equation 4.53 or 4.54 after the diagonalisation of the measured symmetric matrix. The defects principle axis are then defined by the diagonal matrix elements $A_i, i = 1, 2, 3$ (after diagonalisation).

Next, a symmetry sensitive technique (most often EPR) is used to monitor the relative populations of each orientation. This is straight forward in EPR, as the anisotropy of the g tensor or fine structure terms with the different populations is reflected in the different signal amplitudes. The kinetics of reorientation are then found from the recovery of the random alignment of defects at varying temperatures, these follow:

$$n_j(t) - n_{j\mathrm{eq}} = [n_j(0) - n_{j\mathrm{eq}}]\exp\left(\frac{-t}{\tau}\right), \tag{4.56}$$

for a first-order recovery process, and $n_{j\mathrm{eq}}$ is the equilibrium value of n_j. The reorientation energy U is then extracted from the time constant τ by:

$$\tau^{-1} = \nu_0 \exp\left(\frac{-U}{kT}\right), \tag{4.57}$$

with ν_0 being the attempt frequency for the process.

4.7 Analysis of bistable defects in semiconductors

4.7.1 Introduction

In general, a defect adopts one specific configuration, and this the so-called ground state structure lies well below O(1eV) the next lowest energy configuration. When the ground state is closer to the first excited state configuration (the lowest energy metastable structure), it becomes possible to probe the metastability experimentally. Bistability (or even multistability) arises when there exists one (or more) metastable configurations, separated by an energy barrier, and the ground state structure *switches* with varying charge state.

Consider first an isolated interstitial defect X_i in a tetrahedrally bonded semiconductor, where X can be either a host or impurity atom, often formed by the reaction between X_s and I (self-interstitial):

$$X_s + I \rightarrow X_i$$

There are many possible sites: tetrahedral T_d, hexagonal H, bond centred (BC), anti-bonding (AB) or split-interstitial for this interstitial and hence there is the potential here for bistability. In-fact this is the case for H in Si [24], where H resides at the T site for $H_i^{0,-}$, whereas the BC site is favoured in $H_i^{+,0}$. Evidently though, there is a small number of possible configurations that the defect can adopt.

More complex defects are formed by the reaction between a migrating X_i, and a second, substitutional impurity Y_s:

$$X_i + Y_s \rightarrow X_i Y_s$$

Clearly, there is a vast number of configurations the pair can adopt, combinations of the interstitial configurations of X at varying distances from the Y atom. Note also that there is the possibility of interstitial exchange to X_sY_i. There are two basic mechanisms by which the formation of a pair defect is favoured: (i) electrostatic attraction; for example donor-acceptor pairs such as Fe_i-Al_s [25], and C_i-P_s [26] in silicon where a Coulomb attraction exists between the positively charged ionised donor and the negative acceptor (applicable to the neutral defect), and (ii) reduction of strain energy; for example C_i-C_s [27] where no long range Coulomb attraction exists between the pair, instead the energy reduction is a result of the relieving of strain around both defects. It is often the case that the migration/reorientation energy of the isolated interstitial is (<1 eV), and this figure will only be slightly modified – at least for reorientation – by the formation of the pair. Therefore, the barrier for conversion between two XY configurations will be around this value also, and the barrier can be surmounted in a variety of ways experimentally – leading to the observation of the very interesting meta/bi-stability phenomena.

So far a few examples of meta/bi-stability in silicon have been mentioned. Before going on to describe how the different defect structures are probed experimentally, it should be pointed out that investigations into bi-stability are of great current interest and technological importance in a variety of semiconductors. Other observed bi-stable centres in silicon include the early thermal donors (TDs) [28], and a boron related centre [29]. Observed bistable centres in other semiconductors include EL2 in GaAs [30], DX centres in $Al_xGa_{(1-x)}As$ and $GaAs_{(1-x)}P_x$ [31], Fe related centres in InP [32], and thermal donors in Germanium [33]. In some cases (in particular H_i, Si_i) there is little or no experimental information on these defects and metastable configurations due to their low migration energies. The low migration of these species will inevitably come from meta/bistability effects, possibly including rapid capture or emission of one or more electrons. Much information on the structure of these defects (especially Si_i) has therefore been obtained by theoretical calculations, and will be discussed later on in chapter 5.

EPR first identified structurally multistable defects [34] in studies of the oxygen-vacancy pair in Si. Other techniques such as photocapacitance and ODMR have also proved successful in monitoring bistable centres. However, DLTS is the most common technique employed in the monitoring of meta/bi-stable centres in Si, via the monitoring of the donor or acceptor levels associated with the different configurations. In general DLTS is able to obtain all of the required information to construct a full configurational coordinate diagram, discussed in section 4.7.2. These defects can be cycled between their different configurations experimentally, and in this respect, should be compared to preferential alignment studies discussed earlier 4.6.3, where the oriental degeneracy of a defect was removed with the application of applied uniaxial stress.

Some of the earliest identified metastable and bi-stable defects, the Fe_i-Al_s pair and the Jellison centre in Si, are then discussed in detail in 4.7.6, along with the electronic properties of the C_i-P_s centre to illustrate the methods employed in the investigations of multistability. More will be said about the C_i-C_s and C_i-P_s defects in chapter 6.5 where theoretical modelling of the phenomena is also performed.

4.7.2 The configurational coordinate diagram

Consider a defect which has a ground state configuration Q_1 in both the m and n charge states. The meta-stable structure Q_2 lies ΔE_m and ΔE_n ($\Delta E_m > \Delta E_n$) above the ground state in the m and n charge states respectively. Finally, the barrier for structural conversion between Q_2 and Q_1 is E_B^m (m charge state) and

E_B^n (n charge state). The configurational coordinate diagram for such a defect is shown in Figure 4.5a.

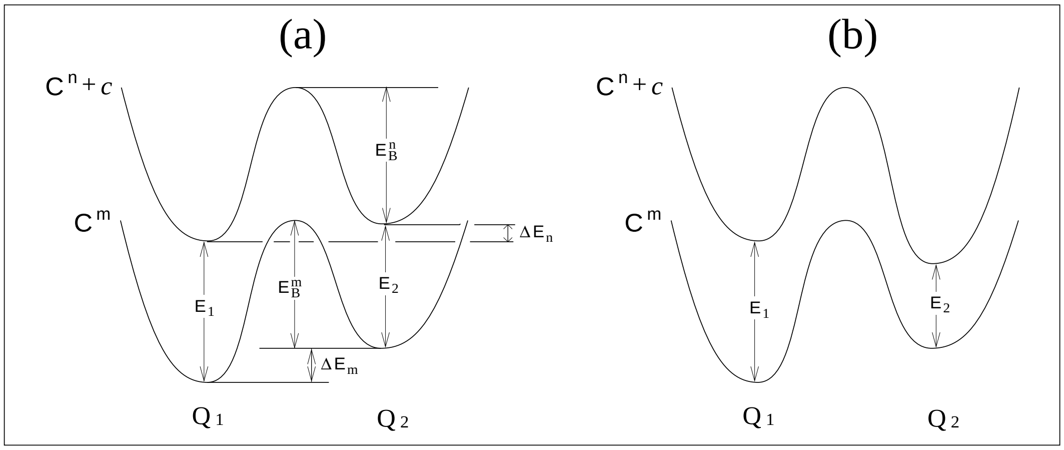


Figure 4.5: The configurational coordinate diagram for (a) a meta-stable defect, and (b) a bistable defect.

In the case of a bistable defect, where the ground state configuration *changes* with charge state, and Q_2 is the stable structure after the emission of a carrier (figure 4.5b). The ionisation energies for $\mathrm{C}^m \to \mathrm{C}^n + c$ in the Q_1 and Q_2 configurations is given by E_1 and E_2 respectively, and it is these that are monitored in DLTS. Here $m = n \pm 1$ and c represents the emitted carrier (e for acceptor and h for donor levels). The case for n=0 is detailed in Table 4.1 below.

Table 4.1: Acceptor and donor transitions for the configurational coordinated diagram in figure 4.5, and n =0.

	Transition	Level Q_1	Level Q_2
Acceptor (-/0)	$\mathrm{C}^- \to \mathrm{C}^0 + e$	E_c-E_1	E_c-E_2
Donor (0/+)	$\mathrm{C}^+ \to \mathrm{C}^0 + h$	E_v+E_1	E_v+E_2

After formation of our defect C, in equilibrium conditions, the relative populations of the defect configurations Q_1 and Q_2 are given by equations 4.58 and 4.59 for the m and n charge states respectively, when Q_2 is metastable in both charge states.

$$f(Q_2^m) \sim \exp\left[\frac{-(\Delta E_m \text{ eV})}{kT}\right], \qquad f(Q_1^m) = 1 - f(Q_2^m), \tag{4.58}$$

$$f(Q_2^n) \sim \exp\left[\frac{-(\Delta E_n \text{ eV})}{kT}\right] \qquad f(Q_1^n) = 1 - f(Q_2^n). \tag{4.59}$$

Therefore, referring to Figure 4.5a, under equilibrium conditions, the number of C defects in the Q_2 configuration is greater for C^n than it is for C^m because there $\Delta E_m > \Delta E_n$. Crucially if ΔE_n is small, then an appreciable fraction of the defects will be in the metastable configuration in the n charge state. If the defect is bistable as in figure 4.5b, then equation 4.59 should be replaced by equation 4.60 below:

$$f(Q_1^n) \sim \exp\left[\frac{-(\Delta E_n \text{ eV})}{kT}\right] \qquad f(Q_2^n) = 1 - f(Q_1^n). \tag{4.60}$$

It should be noted for meta/bistable defects that if the emission energies satisfy the condition E_1, $E_2 < E_B$, then it would be expected that the metastable phenomena should be detected, however if E_1, $E_2 > E_B$ then it is unlikely that E_2 would be detected in DLTS because most centres prepared in the Q_2 configuration will convert to the Q_1 configuration before charge transfer at $Q = Q_2$ is detected. Similar problems would not be encountered in DLOS measurements. More will be said regarding these complications in section 4.7.4.

4.7.3 Structural excitation of bi/metastable defects

There are three basic approaches in which conversion of some or all of the C defects from the Q_1 to the Q_2 configurations is achieved. These are listed below:
(1) Cooling down under reverse bias,
(2) Minority carrier injection, and
(3) Photoexcitation Nd:YAG (1.16 eV) light.
The first of these, is commonly used in DLTS investigations of meta/bistable when the energy difference between Q_1 and Q_2 either: (i) becomes small, $kT \approx \Delta E$ - the energy difference between the configurations (metastable), or (ii) reverses (bistable) with varying *charge state*. Consider the case of the (-/0) acceptor transition in a *metastable defect* in n-type Si, illustrated in Figure 4.6.
Initially, in the lightly irradiated Si, the Fermi level is at E_F and hence all C defects are in the negative charge state and the ratio N_{Q_2}/N_{Q_1} is essentially zero. However when a reverse-bias is applied the quasi-Fermi level for the electrons decreases to E_F', resulting in the formation of the neutral defect C^0. In thermal equilibrium, the ratio N_{Q_2}/N_{Q_1} can be appreciable, since energy difference between the two configurations (ΔE_0) is much less for the neutral defect (Figure 4.6). Hence if the sample is cooled to low temperatures under the reverse bias, and then the bias removed, an appreciable number of C^- defects in the metastable configuration Q_2 will remain – *ie* a non equilibrium situation has been achieved. In the case of a bistable defect, cooling down to low temperatures under reverse bias will result in essentially all of the defects occupying Q_2^n. Removal of the bias at low temperatures

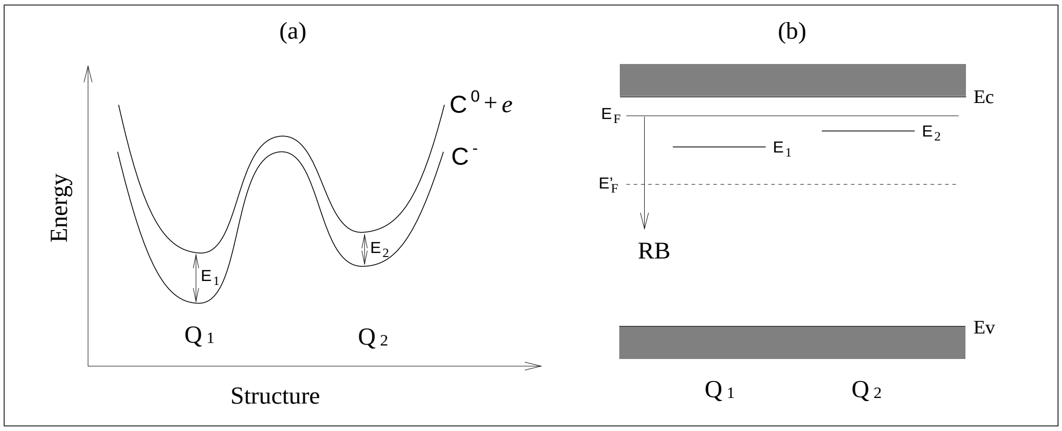


Figure 4.6: (a) the configurational coordinate diagram for a meta-stable defect,and (b) the gap states and Fermi level in lightly irradiated n type Si.

will result in all centres being in the metastable Q_2^m charge state, and the detection of the bistability.

Both the electrical carrier injection and the photoexcitation generate generate excess electrons and holes. An electronic transition may then occur between a carrier and the defect, with the energy released as a photon (PL), by multiple phonon emission (MPE) or a combination of the two. If the e and h form a free-exciton which is subsequently trapped at a defect, then Auger excitations are also possible. It is then possible that these processes and can result in the structural change of the defect by two mechanisms (i) recombination enhanced conversion, and (ii) an excitonic Auger emission structural excitation. The structural conversion resulting from these processes are independent of the cooling conditions and can occur at low temperatures.

The first of these processes, recombination enhancement [35] is a process in which the activation energy of the defect transition is lowered (or even becomes athermal) as a result of carrier recombination at the defect. As described in the section on PL, the different charge states of a defect are described by a number of energy surfaces and any recombination can involve photons and/or a number of phonons. If the defect local vibrational modes are excited as a result of the recombination, then since it takes a while until this energy is dissipated to the surrounding crystal, the energy might be added to the thermal energy to overcome the potential barrier for the transition. The excited state of the defect has therefore enhanced the reaction rate. Roughly speaking the magnitude of the contribution to the barrier reduction is related to the localisation of the electronic states involved in the transition. In the case of deep levels, where the carrier is bound to within a few Å of the defect, there is a strong electron-phonon coupling, and any energy supplied to the defect as a result of a transition is provided in a region close to the defects reaction saddle point. For radiative transitions therefore, there is generally no enhancement – unless a phonon sideband is also excited. For multi-phonon emission, probability $\exp(-E_D/\hbar\omega_l)$ (where $\hbar\omega_l$ is the energy of excited phonons and E_D the defect energy level), then the enhancement is also possible.

In most cases of reaction kinetics a diffusion process is the rate limiting step. The reaction rate R_T for a transition is given by equation 4.61:

$$R_T = \nu \frac{\exp(\Delta S/k)}{N_j} \exp[-E_T/kT], \tag{4.61}$$

where ν is the attempt frequency $\sim 10^{13}$ s^{-1}, $\exp(\Delta S/k)$ is the entropy associated with the number of equivalent saddle points (or bond softening) and N_j is the number of diffusion jumps required for the specific process. These factors are determined experimentally by monitoring the reaction rate against temperature.

The above analysis is not applicable to the case of recombination enhanced diffusion, because here equilibrium conditions do not apply to the excited defect. However two simple expressions have been derived for the recombination enhanced reaction rate E_{RE} and these are as follows:

$$R_{RE} = \mu \frac{\exp(\Delta S/k)}{N_j} R_R \exp\left[-\frac{(E_T - E_R)}{kT}\right], \quad E_R << E_T, \tag{4.62}$$

$$R_{RE} = \mu \frac{\exp(\Delta S/k)}{N_j} R_R, \quad E_R >> E_T, \tag{4.63}$$

where E_R is the vibrational energy introduced into the defect as a result of the transition, μ is an efficiency factor relating to the fraction of recombination events resulting in a successful reaction, and R_R is the recombination rate. The effect of the recombination enhancement is to lower the reaction barrier height. Referring to equation 4.62 at high temperatures, the thermal activation E_T will be the dominant effect, whilst at lower temperatures the effective activation energy is $E_{RE} \approx E_T - E_R$. If $E_R > E_T$ then invoking equation 4.63, the reaction rate is independent of temperature in all but the recombination rate E_R.

The excitonic Auger capture process is somewhat different to the recombination enhanced diffusion, but again provides a mechanism for configurational conversion of metastable defects in semiconductors. Here, a free exciton interacts with the defect and recombines. The recombination energy is transferred to an electron, which is emitted to the conduction band. This hot Auger electron then reaches the bottom of E_c in a short time (typically a matter of picoseconds), and in doing so liberates its energy to the defect. It is this energy which excites the defect into the metastable configuration. This process therefore takes place athermally, and in some cases the Auger electron can be detected. This process has recently been shown to be the case in for the S-Cu complex in Si [36], where the defect conversion rate was observed to be dependent on the free exciton concentration, and temperature independent. In addition to this, the Auger electron was also detected.

4.7.4 DLTS investigations of bistable defects

In the case of a bistable defect, two or more configurations exist in a particular charge state and it is possible to cycle between these configurations as discussed above. A change in configuration can then be detected by a corresponding change in the optical and electronic properties of the defect, and DLTS can be used to monitor the absolute concentration of defects in each configuration. DLTS is superior to other junction capacitance techniques (for example TSCAP) in that it has a greater sensitivity and resolution. As discussed earlier 4.4 the junction capacitance spectroscopies detect defect related trap states in the gap via a change in capacitance ΔC which occurs upon carrier emission. DLTS achieves a greater sensitivity by repeated bias changes, with a forward bias pulse to fill the traps (pulse width t_p), followed by a quiescent reverse bias stage during which the carrier emission is detected (with t_r denoting the time between majority carrier pulses). The whole process is repeated many times often with varying temperature and the signal averaged; it is this repetition and subsequent measurement of the transient which yields a greater sensitivity. The DLTS system time constant τ_{DLTS} can be varied independently of t_r, provided $t_r \geq \tau_{\mathrm{DLTS}}$.

If a defect can adopt one of two configurations, then two energy levels E_1 and E_2 exist, and can be detected after the emission of the carrier if certain conditions are met: namely that carrier occupied metastable configuration lives long enough at sufficiently low temperatures, and the experimental method allows the preservation of this metastable state until the carrier has been emitted and detected. The time constant for the emission of a carrier is of the form:

$$\tau = a\exp[(-b)/kT], \tag{4.64}$$

where a is a constant, and b is the emission energy.

Monitoring the carrier emissions under different conditions (biasing, annealing, biased annealing) leads to information on the properties of the defect in each configuration as well as the kinetics associated with the structural change. To do this, the relative magnitudes of the DLTS peak heights (proportional to the concentrations of each defect structure) are monitored after various excitations to promote a given charge state or structure.

Complications arise in the analysis of the DLTS results in the investigation of bi/metastable defects. This is because in the experiment itself, the quasi-Fermi level changes during the biasing stages. Therefore the charge states of the defect will vary during the filling pulse and quiescent stages, and in the case of bistable defects the *ground state* structure will also change, following the variations in charge state. Under certain conditions, the configurational transformations may then be induced by the measurement itself and recombination enhancement may also have an important effect.

Consider a defect which has two possible structures Q_1 and Q_2, the former being the ground state configuration when the defect is occupied by a carrier and the later being the more stable structure after the emission of the carrier. This situation represents a bistable defect (illustrated in figure 4.5 b) with Q_2 and Q_1 being the metastable configurations for the carrier occupied and unoccupied defects respectively. Now, in the course of the DLTS run, the *bias changes* affect the *relative stability* of the two configurations in such a way that:
(i) during the trap filling pulses (width t_p) the majority carrier injection promotes $Q_2 \rightarrow Q_1$ conversion,
(ii) whilst $Q_1 \rightarrow Q_2$ is favoured during the quiescent reverse bias stage (width t_r).
The discussion presented here concerns DLTS and electrical injections. The results are however applicable to optical experiments when DLOS is used. It should be noted that the analysis is concerned with the thermal processes *ie.* the energy required to overcome the barrier for structural conversion is predominantly due to the temperature, although the role of phonons leading to recombination enhancement may be important. Therefore the difficulties in the investigations of bistable defects by DLTS are most likely to arise if the barrier E_B is small, that is to say $E_B \sim kT$.

Generally speaking, the DLTS peak height is proportional to ΔC which in turn is proportional to the defect concentration. In DLTS of bistable defects therefore, the peak heights during each pulse interval are determined by the fraction of defects in each configuration at the end of the preceding trap filling pulse. For bistable defects where $E_B >> kT$ these fractions are determined by the initial conditions only. However the picture is different when $E_B \sim kT$, here this fraction arises from a combination of the initial conditions and the configurational transformations which occurred during the previous pulse and interval cycle. To properly analyse the DLTS peak height data, and obtain information regarding the defect concentrations, an expression was derived [37] which gives the fraction of defects in each configuration at the end of the filling pulse as a function of experimental parameters and the transformation rates. These transformation rates are calculated from measured fractions, which, under appropriate conditions can be determined from the ratio of the observed peak to maximum peak height.

Levinson assumed that only exponential transients were present, in the derivation of the expression for the number of defects. Consider the bistable defect with structures Q_1 and Q_2 with fractions f_{Q_1} and f_{Q_2} respectively, after the end of a trap filling pulse. Denoting the temperature dependent $Q_1 \rightarrow Q_2$ kinetics term by $\theta_{Q_1Q_2}(t_r)$ and the $Q_2 \rightarrow Q_1$ by $\theta_{Q_2Q_1}(t_p)$ and assuming that the defect transformation kinetics were first order in defect concentrations, the expressions can now be derived. $\theta_{Q_1Q_2}(t_r)$ and $\theta_{Q_2Q_1}(t_p)$ are determined from the differential equations which govern the time evolution of the fractions during the interval and pulsing sequences respectively. At time t_r at the end of the quiescent reverse bias interval, the fraction of defects in Q_1 is given by:

$$f_{Q_1}(t_r) = 1 - f_{Q_2}(t_r) = \theta_{Q_1Q_2} f_{Q_1}(0), \tag{4.65}$$

at the end of the following filling pulse, this expression for the fraction in Q_1 is:

$$f_{Q_1}(t_p + t_r) = 1 - f_{Q_2}(t_p + t_r) = 1 - \theta_{Q_2Q_1} f_{Q_2}(t_r). \tag{4.66}$$

Assuming constant $\theta_{Q_1Q_2}(t_r)$ and $\theta_{Q_2Q_1}(t_p)$ and therefore constant temperature, after n interval-pulse cycles, the resultant expression for f_{Q_1} is:

$$f_{Q_1}[n(t_p + t_r)] = (\theta_{Q_1Q_2}\theta_{Q_2Q_1})^n f_{Q_1}(0) + (1 - \theta_{Q_2Q_1}) \sum_{j=1}^{n-1} (\theta_{Q_1Q_2}\theta_{Q_2Q_1})^j. \tag{4.67}$$

As $n \rightarrow \infty$, with $\theta_{Q_1Q_2}$ and $\theta_{Q_2Q_1}$ bound by the conditions $0 \leq \theta_{Q_1Q_2} \leq 1$ and $0 \leq \theta_{Q_2Q_1} \leq 1$, then assuming that $\theta_{Q_1Q_2}$ and $\theta_{Q_2Q_1}$ are not both equal to one, 4.67 reduces to:

$$f_{Q_1}^{\infty} = 1 - f_{Q_2}^{\infty} = (1 - \theta_{Q_2Q_1})(1 - \theta_{Q_1Q_2}\theta_{Q_2Q_1})^{-1}. \tag{4.68}$$

If the transformations are governed by a single rate limiting step, then $\theta_{Q_1Q_2}$ and $\theta_{Q_2Q_1}$ are simple decaying exponentials:

$$\theta_{Q_1Q_2} = e^{-R_{Q_1Q_2}t_r}, \qquad \theta_{Q_2Q_1} = e^{-R_{Q_2Q_1}t_p}, \tag{4.69}$$

here, $R_{Q_1Q_2}$ and $R_{Q_2Q_1}$ represent the $Q_1 \rightarrow Q_2$ and $Q_2 \rightarrow Q_1$ transformation rates respectively, using these 4.68 becomes:

$$f_{Q_1}^{\infty} = 1 - f_{Q_2}^{\infty} = (1 - e^{-R_{Q_2Q_1}t_p})[1 - e^{-(R_{Q_1Q_2}t_r + R_{Q_2Q_1}t_p)}]^{-1}. \tag{4.70}$$

The DLTS time constant τ^* is defined as the time required for the condition $(\theta_{Q_1Q_2}\theta_{Q_2Q_1})^n = e^{-1}$ to be satisfied. Since $\tau^* = n(t_r + t_p)$, then:

$$\tau^* = -\frac{t_t + t_p}{\ln(\theta_{Q_1Q_2}\theta_{Q_2Q_1})} = \frac{t_t + t_p}{R_{Q_1Q_2}t_r + R_{Q_2Q_1}t_p}. \tag{4.71}$$

Therefore, if the transformation $Q_1 \rightarrow Q_2$ proceeds much further in the time step t_r than $Q_2 \rightarrow Q_1$ does in t_p then the signal will be dominated by the Q_1 emission E_1. Equations 4.68 and 4.70 can be used to analyse the DLTS data when τ^* is small *ie.* if at least one of the transformation rates are fast compared to the heating rate of the sample. If this is true, any variations in the sample temperature during the measurement do not affect the result, and the ratio of the peak heights to their maxima reflect the fractions f_{Q_1} and f_{Q_2}. If $t_r >> t_p$ then from 4.70 the effect is to increase f_{Q_1} at the expense of f_{Q_2}, and similarly f_{Q_2} is favoured if $t_p >> t_r$. If the transformation rates are slow and therefore τ^* large, then the fractions are not significantly affected by the measurement.

There are two mechanisms by which the DLTS run can influence the structure of the defects: in the first, since Q_2 is the stable configuration during the quiescent reverse bias, the magnitude of E_2 simply depends on the f_{Q_2} at the start of that interval. However since Q_1 is metastable during t_r, and assuming that $Q_1 \rightarrow Q_2$ is charge state controlled and takes place after the emission of a carrier, then the normal ΔC is observed at the beginning of the *first* pulse interval. However, since the emission resulted in the structural transformation, the transient in subsequent intervals *will* be affected. Unlike the first effect where transients in proceeding intervals was affected, the second mechanism leads to a configurational conversion *during* the pulse interval and therefore continually affects the fractions. These involve carriers emitted after the $Q_1 \rightarrow Q_2$ transition takes place and therefore occur at E_2 in the same pulse cycle. This therefore requires that the defect is initially occupied by more than one carrier, and depending on the relative times for the configurational transformation and subsequent emission effects such as multiple carrier emissions or complex non-exponential decays can be observed.

Assuming now that accurate information pertaining to the fraction of defects in each configuration can be derived, the determination of other properties such as the barrier for conversion and the absolute difference in energy between ground state configurations is now discussed. Again, in the following analysis refer to the generalised configurational coordinated diagrams in figure 4.5. To determine the barriers for the conversion between the Q_1 and Q_2 configurations in the m and n charge states, the following approach is adopted: the barriers are determined from the thermally activated structural transformation kinetics. Consider first the $Q_1 \rightarrow Q_2$ transition for the carrier occupied defect (charge state m), the sample is cooled from a given temperature under *reverse bias* to promote the occupation of the Q_2 structure. After the initial concentrations f_{Q_1} and f_{Q_2} have been determined, the diode is then shorted, and the sample annealed at a temperature T for a time t. Following this, the sample is rapidly cooled to low temperatures and the DLTS measurements performed to observe the changes in the fractions f_{Q_1} and f_{Q_2}. This procedure is repeated at various temperatures, and the normalised DLTS pulse height given by equation 4.72 is plotted against the anneal time for the various temperatures T.

$$\text{Normalised peak height} = \frac{f_{Q_2}(t) - f_{Q_2}(\infty)}{f_{Q_2}(0) - f_{Q_2}(\infty)}. \tag{4.72}$$

For the structural transformation in the n charge state, the reverse transition $Q_2 \to Q_1$ can be studied in a similar way by cooling the sample down with the bias off, followed by a series of reverse bias anneals. Alternatively, before the anneals the relative fractions f_{Q_1} and f_{Q_2} can be altered by one of the other excitation methods to achieve the non-equilibrium conditions. In general the barriers will be different in the differing charge states, and hence the conversion rates will have different temperature dependencies. In the n charge state where Q_2 is metastable, the functional form of the recovery of the fraction f_{Q_2} from the non-equilibrium conditions is given by equation 4.73:

$$f_{Q_2}(t) = [f_{Q_2}(0) - f_{Q_2}(\infty)]\exp(-Rt) + f_{Q_2}(\infty), \tag{4.73}$$

where R is a transformation rate of the form:

$$R_{Q_i \to Q_j} = \alpha \exp[-\beta/kt], \tag{4.74}$$

here, α and β are constants the former being the pre-exponential factor and the latter the barrier for the structural change. A pre-exponential factor of $O(10^{13})$ is indicative of a 'single jump' process reflecting a meta-stable transition. In fact this can be confirmed by spatially resolved DLTS experiments with various applied biases. To ensure that no charge state changes occurred during the $Q_1 \to Q_2$ transition, TSCAP experiments can be performed – a change in charge state being detected by any steps in the data.

Finally, the absolute differences in energy between the two configurations in any charge state can be determined. Consider charge state n where Q_1 is metastable (figure 4.5). There is always a residual occupancy of the Q_1 structure, denoted by $f_{Q_1}(\infty)$ which is temperature dependent given by equation 4.59, arising from Maxwell-Boltzmann statistics. Experimentally, this fraction is determined using appropriate biasing conditions in the following way:

$$\frac{f^n_{Q_2}}{f^n_{Q_1}} = \frac{f^n_{Q_2}(\mathrm{RB})}{f^n_{Q_1}(\mathrm{RB})} = \frac{f^n_{Q_1}(\mathrm{ZB}) - f^n_{Q_1}(\mathrm{RB})}{f^n_{Q_1}(\mathrm{RB})}, \tag{4.75}$$

where (ZB) refers to zero bias and (RB) to reverse bias. A similar expression can be obtained for the case of metastable defects, and for other charge states. This ratio is obtained for a range of quenching temperatures, and the log of the fraction plotted against the reciprocal temperature. The gradient of the resultant line is then ΔE_n.

4.7.5 EPR and PL analysis of meta/bistable defects

Although most investigations into bistable centres utilised DLTS, the bistability of a defect can also be detected by other techniques such as EPR and PL. An illustration of the way in which bistability (or indeed metastability) is detected using these methods is given below. Consider a bistable defect (as in figure 4.5), and let us assume for simplicity that $n = 0$, and the neutral defect is not paramagnetic. Further more, the neutral defect (Q_2) gives rise to a PL line at X meV, with no PL signal from any other configuration in any other charge state. Therefore one would expect an EPR signal from both Q_1 and Q_2 when $m = \pm 1$. For simplicity, the EPR spectra from Q_1 and Q_2 are labelled S_1^m and S_2^m respectively. Now because Q_1 is the ground state in both charge states, when the EPR experiment is performed (with E_F positioned such that all defects C are in the m charge state, irrespective of configuration), then EPR detects almost all of the defects in Q_1 by the spectrum S_1. However, if optical excitation of the defects configuration is induced by an injection of fixed intensity laser light, $S_1 \rightarrow S_2$ conversion is detected. The conversion of the spectrum is an exponential function of the illumination time, with the sum of S_1 and S_2 signal intensities being constant. It is worth noting that as there is no observed loss of the EPR signal, the charge state is constant during the conversion *ie.* Q_1^m + light$\rightarrow Q_2^m$. A pictorial representation of the process is given in Figure 4.7a. Conversion back to the S_1 signal is achieved by heating the sample in the dark. The resultant time constant for this return process provides a second method of determining the conversion barrier E_B^m, and the results can be compared to those from DLTS. Further irradiation of the sample can move E_F to a position where it coincides with the S_1 acceptor (n-type) or the S_2 donor (p-type). Here it is assumed that both the Q_2 donor and acceptor levels are shallower than those of Q_1 – although similar arguments can be followed when this is not the case. The result of this is that under thermal equilibrium, part of the defects will be in the Q_1^m charge state with the others occupying Q_2^0. Under illumination $Q_2^0 + e^-$(photo-generated)$\rightarrow Q_2^-$ takes place, as well as the photo-generated $Q_1 \rightarrow Q_2$ conversion, and the resultant

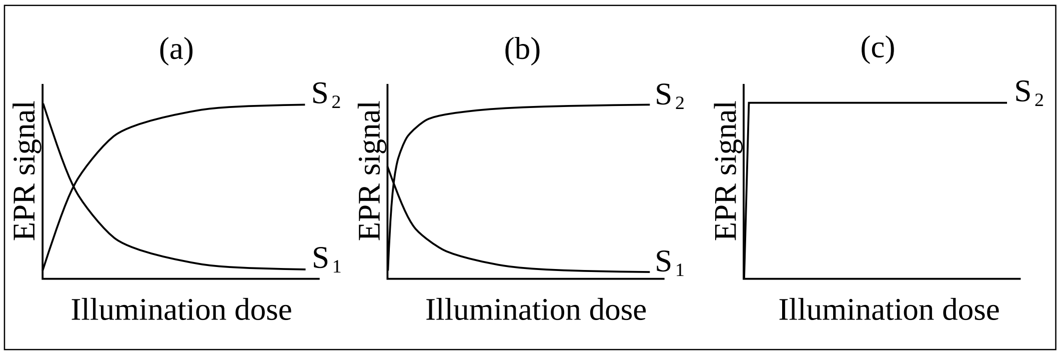


Figure 4.7: Schematic illustration of the effect of low temperature photo-injection on the intensities of EPR signals S_1 and S_2 for various positions of the Fermi level.

EPR signal is as shown in figure 4.7b. Finally, for large irradiation doses, the Q_2^0 state is adopted from the outset and upon photo-excitation the EPR spectrum is as shown in figure 4.7c. When this is the case, and Q_2^0 captures a photogenerated carrier becoming Q_2^m followed by detection in EPR at low temperatures, the ionised defect is unstable. Therefore when the temperature is raised $Q_2^m \rightarrow Q_2^0 + c$ occurs, and the temperature dependence of the loss of S_2 can be measured and will be equivalent to that of the carrier emission in DLTS. The EPR measurement can thus complement the DLTS measurements and such a procedure can serve to strengthen the correlation of a DLTS signal E_i with an EPR spectrum S_i.

In a similar manner to the EPR experiments, if Q_2^0 is the only configuration to give rise to a PL signal as described above, the bistability will also be manifested in the PL. Again photoexcitation of Q_1^m to Q_2^m followed by carrier emission is required when all defects are charged, and therefore the PL signal will be seen to rise at a rate $(1-e^{-t/\tau})$, where t is the illumination time and τ the conversion rate time constant. Further irradiation until all the bistable defects are neutral implies that the PL signal should be observed straight away having constant intensity with illumination time. Locking E_F in such a way can provide a further check as to which charge state of a defect gives rise to a PL, normally achieved through Zeeman measurements. Under thermal equilibrium, the ratio of neutral defects $f_{Q_1}^0/f_{Q_2}^0$ is given exactly by the following expression 4.76:

$$\frac{f_{Q_1}^0}{f_{Q_2}^0} = \exp(-\Delta E_0/kT)\left[\frac{1+\exp(-|E_1 - E_F|)/kT}{1+\exp(-|E_2 - E_F|)/kT}\right], \tag{4.76}$$

This expression is of course generally applicable to the various, even mixed, charge states, and is more accurate than equations 4.58, 4.59, and 4.60, where the position of E_F was neglected. By monitoring this ration as the Fermi level moves this ratio (observed in PL) can be used to estimate one of ΔE_0, E_1 or E_2 if two of the others have been determined by a different method. Indeed the approached discussed here for EPR and PL analysis of bistable defects were utilised in conjunction with DLTS in a complete analysis of the di-carbon centre in Si [27].

4.7.6 The Fe$_i$-Al$_s$ defect in Si

An example of metastability is silicon is now discussed in some detail, following the arguments of Chantre [25]. The iron atoms are usually present in Si as electrically inactive precipitates, and form Fe$_i$ with an deep donor level at E_v+0.38 eV upon fast cooling or quenching. The Fe atom resides at the T_d site and has a migration energy of 0.68 eV. Therefore, in p-type Si negatively charged acceptors Ac$^-$ attract the positively charged Fe ions forming the Fe$^+$-Ac$^-$ pair. Early EPR work [38] found that the TM$_i^+$-Ac$_s^-$ pairs adopted a $\langle 111 \rangle$ configuration with the TM at the T_d site closest to the Ac$^+$ atom. The exception to this was Fe$_i$In$_s$, which showed $\langle 100 \rangle$ symmetry, indicating that strain effects become more important as the size of the acceptor atom increases.

Fe$^+$-Ac$^-$ pairs again form donor levels in the lower half of the forbidden gap, associated with Fe$^+$-Ac$^-$ $\rightarrow$ Fe^{2+}-Ac$^-$ transitions. Consider Fe$_i$Al$_s$, two peaks are observed in DLTS labelled H1 and H2, at E_v+0.20 and E_v+0.13 eV respectively. When the sample was cooled to 77 K down under zero bias H1 was the dominant peak, implying [H1]>>[H2]. On the other hand, cooling down under reverse bias to 77 K from successively higher temperatures leads to an increase in the H2 signal to a point where [H2]/[H1] is appreciable on removal of the bias before the DLTS run. The phenomena was reversible, and implied that the behaviour of the H1 and H2 peaks was consistent with a metastable reaction involving a simple rearrangement of the defect structure.

The kinetics of the configurational change was studied by thermally activating the structural reorientation from the non-equilibrium situation. To obtain the barrier in the positive charge state, cooling under RB to promote the number of metastable centres was first performed. Removal of the bias and monitoring [H2]/[H1] during annealing gave a value of 0.5 eV for the thermally activated change in structure. A similar approach was taken for the neutral charge state, but here zero bias cooling giving ([H1]>>[H2]) was performed before a series of reverse bias anneals, leading to a barrier of 0.64 eV. During these conversions, it was observed that the sum of the DLTS peak heights was constant, again implying a metastable reaction.

The residual H2 peak height associated with the positive defect was never zero, and was found to be temperature dependent (a result of the thermal occupancy of the metastable structure during the anneals). The absolute energy difference between the ground state and metastable state was found to be 0.14 eV from analysis of this dependence. The configurational coordinate diagram resulting from these DLTS measurements is shown in Figure 4.8.

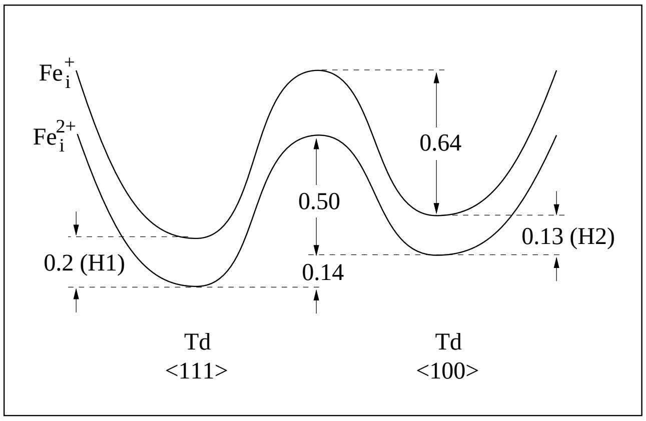


Figure 4.8: The configurational coordinate diagram for the metastable Fe_i-Al_s pair defect in Si.

A simple ionic model was employed to explain the metastability, and obtain the structures of both configurations of the pair. It was first noted that the binding energy of the pair was around 0.5 eV, very close to the energy of two electrostatically bound point charges: taking $E_b = e^2/\epsilon r_1$, where ϵ is the Si dielectric constant, and r_1=2.35 Å the distance between the T_d and substitutional sites, they obtained a value of 0.52 eV for the binding energy. They then argued that if the Fe ion was to move forming a $\langle 100 \rangle$ configuration with the Fe_i occupying the the next nearest T_d site, then the distance between the pair increases to r_2=2.72 Å. The difference in energy between the $\langle 100 \rangle$ and $\langle 111 \rangle$ structures is then given by equation 4.77:

$$\Delta E_b = \frac{e^2}{\epsilon}\left[\frac{1}{r_1} - \frac{1}{r_2}\right], \tag{4.77}$$

giving an energy difference of 0.071 eV. The structural models for the $\langle 111 \rangle$ and $\langle 100 \rangle$ configurations are shown in figure 4.9.

The number of FeAl pairs in the $\langle 100 \rangle$ configuration at temperature T is given by Maxwell-Boltzmann statistics 4.78:

$$f \sim \exp[-(0.071eV)/kT]. \tag{4.78}$$

Another prediction of the ionic model is that the energy level of the $\langle 111 \rangle$ pair would shift when the Fe ion moved, leading to a second gap level associated with the $\langle 100 \rangle$ defect. To explain the positioning of the level, a configurational coordinate diagram was constructed based on electrostatic arguments, shown in Figure 4.10.

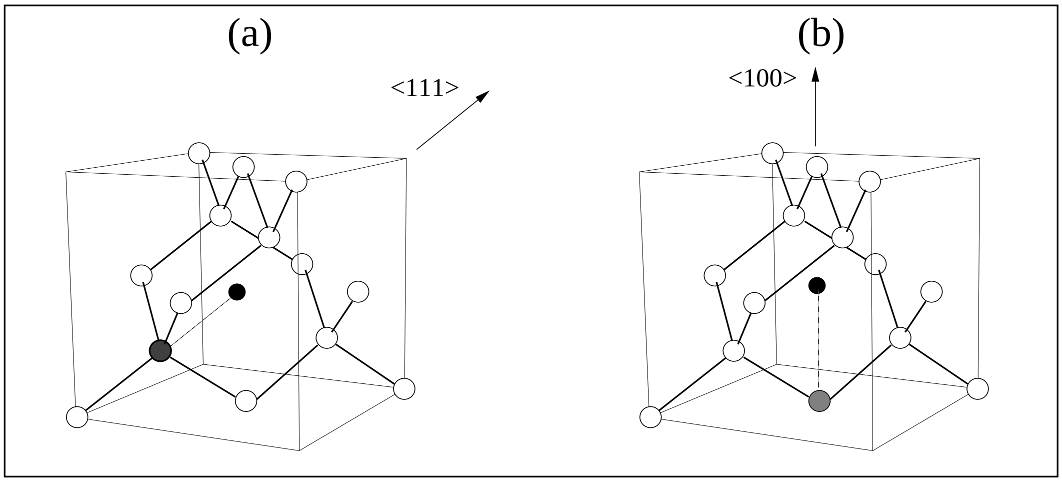


Figure 4.9: The proposed structural models for the Fe_i-Al_s pair defect (a) the $\langle 111\rangle$ configuration, and (b) the $\langle 100\rangle$ configuration.

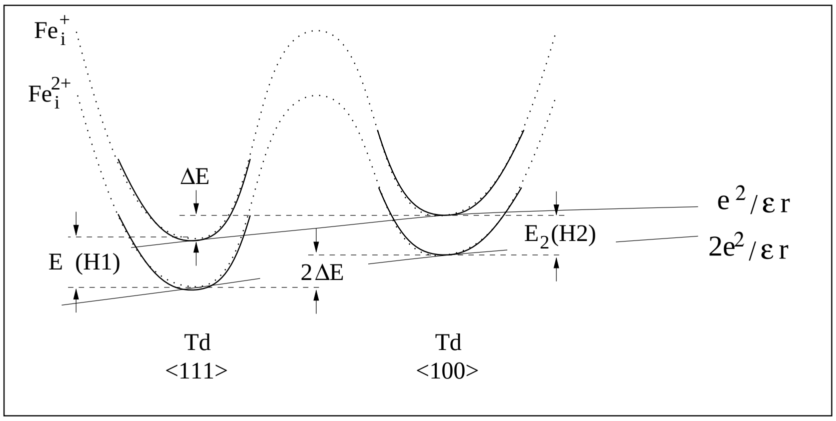


Figure 4.10: The configurational coordinate diagram for the metastable Fe_i-Al_s pair defect in Si, based on the coulomb model.

The energy difference between the two configurations of the for the Fe^{2+} defect is double the difference of that when the Fe ion is singly charged. Therefore, the energy of the second donor level, is related to the first by:

$$E_2 = E_1 - \Delta E = E_1 - 0.071\mathrm{eV}. \tag{4.79}$$

The agreement of between the results of the ionic model, and the experimental results is extremely good: the ionic model predicts and energy difference of $2\Delta E$=0.14 eV between Fe^{2+}-Al^- whereas the observed difference is 0.14 eV. Secondly, the observed donor levels H1 and H2 at E_v+0.2 and E_v+0.13 eV have precisely the energy difference predicted in 4.79. The smaller energy difference between the two configurations in the neutral charge state is entirely consistent with the observations of increasing [H2] after reverse-bias cooling form 250 K to 77 K. The ionic model has therefore lead to a complete understanding of the configurational instability of this centre, coming from a charge state dependent, electrostatic partial dissociation of the pair. It was also noted that the metastability arised from the low (0.68 eV) migration energy of the Fe^{+} ion, the barriers being slightly modified due to the presence of Al and differing charge states. The confirmation of the ionic model and assignment of the stable and metastable structures on the grounds of the electronic properties only was a remarkable feat, confirmation of the structural models by EPR [39] appeared around the same time as the work of Chantre *et al.*

4.7.7 The Jellison bi-stable centre

In 1982 Jellison observed an electron trap at E_c-0.17 eV in n-type Si, and more significantly, this trap converted to a faster component – shallower E_c-0.11 eV trap under certain pulsing conditions in the DLTS experiment [40]. This defect is very different to the FeAl defect because (i) it is an example of a bi-stable defect and (ii) the measurement itself could induce the configurational transition indicating that the barrier is relatively small, and the effects of the defect local vibrational modes on the conversion may be significant. The E_c-0.17 eV level coincides with that of the A-centre (VO) complex and therefore the bi-stable centre giving rise to the levels was originally thought to be VO and VO* [25]. Later work [27] (see chapter 6.5) showed that the Jellison centre was in fact the di-carbon C_i-C_s defect.

Now, special pulsing conditions in DLTS could achieve the $E(0.17)\rightarrow E(0.11)$ conversion, again denoting the time between majority carrier pulses as t_r and the pulse width as t_p these conditions were:

$$t_r >> \tau_r = 7.5 \times 10^{-13}\exp[(0.174 \text{ eV})/kT] \quad \text{s}, \tag{4.80}$$

$$t_p << \tau_p = 7.3 \times 10^{-13}\exp[(0.145 \text{ eV})/kT] \quad \text{s}. \tag{4.81}$$

Consistent with equation 4.70 the peak height of metastable B configuration at E_c-0.11 eV was given by the following expression:

$$E(0.11)/E(0.11)_{MAX} = (1 - e^{-t_r/\tau_r})e^{-t_p/\tau_p}, \tag{4.82}$$

with the condition that $(R_{Q_1Q_2}t_r + R_{Q_2Q_1}t_p)>>1$ and the time constants given by $\tau_r^{-1}=R_{Q_1Q_2}$ and $\tau_p^{-1}=R_{Q_2Q_1}$.

The charged ground state configuration was termed the A-form and the neutral ground state as the B-form. The configurational coordinate diagram was constructed to explain this bistability [25]. This is shown below in figure 4.11.

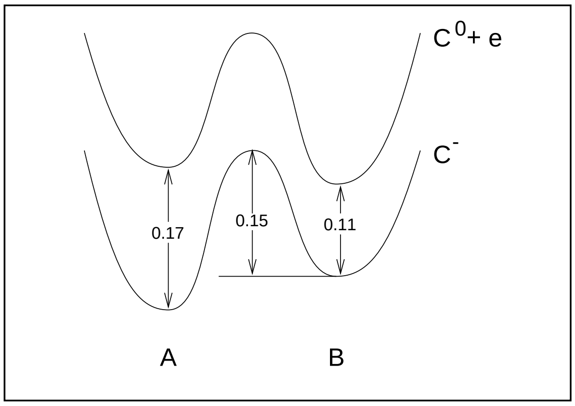


Figure 4.11: The configurational coordinate diagram for the Jellison centre.

The time constant τ_p is associated with the majority carrier injection pulse and hence the B$^-$ to A$^-$, $E(0.11) \rightarrow E(0.17)$ conversion. Similarly, the quiescent reverse bias time constant τ_r is associated with the $E(0.17) \rightarrow E(0.11)$, A^0 to B^0 transition.

Another observation was not all of the $E(0.17)$ level was converted to the $E(0.10)$ level. It was later demonstrated [25] that after zero bias cooling only $E(0.17)$ (A$^- \rightarrow A^0$) was observed, but both $E(0.17)$ and $E(0.11)$ ($B^- \rightarrow B^0$) were observed after reverse bias cooling – consistent with a bi/metastable reaction. The observations were explained with the aid of the configurational coordinate diagram as follows: Cooling down under zero bias led to all of the defects being in the A$^-$ state, alternatively cooling down under reverse bias led to the generation of B^0 at low enough temperatures. Upon electron capture at the start of the DLTS run, B$^- \rightarrow$A$^-$ conversion takes place at a rate τ_p^{-1} before the re-emission of the electron (a result of the low barrier height), leading to a reduction of the B$^- \rightarrow$B^0 signal. B$^- \rightarrow$A$^-$ conversion is slow during the refilling periods t_p and takes place until completion – leading to full restoration of the $E(0.17)$ signal. Therefore, in order to observe the $E(0.11)$ signal, the time between pulses, t_r, should be long enough so that B^0 is as large as possible, and the pulse width t_p short enough to limit the B$^- \rightarrow$A$^-$ conversion. The significance of equation 4.80 is that the A$^0 \rightarrow$B^0 barrier is lower than the activation energy for electron emission from A$^-$. The barrier for thermal recovery, B$^- \rightarrow$A$^-$ was determined by Jellison to be $\sim$0.15 eV while the barrier for A$^- \rightarrow$B^0 was $\sim$0.17 eV – identical to the A$^-$ emmission. It is therefore possible that the barrier for A$^0 \rightarrow$B^0 conversion is $\leq$0.17 eV, so that the A$^- \rightarrow$B^0 conversion is limited by the A$^- \rightarrow A^0$ emmission, which contrls the transformation.

4.7.8 Electronic properties of the multi-stable CP pair

The final example of a metastable centre is that of the C_i-P_s pair, observed in e-irradiated n-type Si. The identifiaction of the components of the defect (C and P) will be discussed later, and here the only the electronic properties of the defect are discussed. The pair is formed when C_i migrates to and is subsequently trapped by P_s.

The C_i-P_s is a remarkable example of a bistable defect, having a total of *four* metastable configurations. The defect was first observed in the mid 80's by DLTS [41, 42], after electron irradiation followed by annealing at 350 K for 30 minutes. In the ground state configuration, labelled IA, the defect gives rise donor and acceptor levels at E_v+0.48 and E_c-0.38 eV respectively.

DLTS with electrical injection

Excitation to two of the metastable configurations (IB, III) can be achieved by minority carrier injection at various temperatures, and the last two (IIB, IIA) forms being generated from an anneal of the III configuration. The generation procedure of the metastable configurations, and the positions of their donor acceptor levels were studied in detail [41, 42], and are summarised in Figures 4.12, 4.2. Silicon doped with either As or Sb displayed similar metastability to that observed in Si:P.

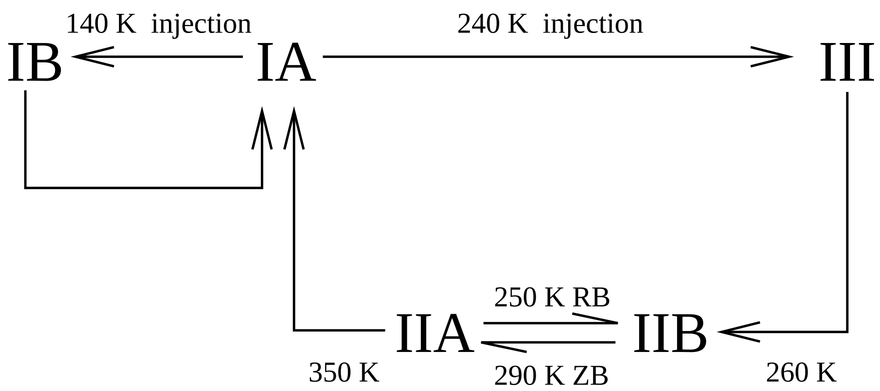


Figure 4.12: The formation processes for the generation of the metastable C_i-P_s configurations.

The positions of the donor and acceptor level positions of the various defect configurations is given below in Table 4.2.
The observed increase of the donor/acceptor levels from their position in the IA configuration was explained in terms of a Coulomb model. When C_i binds to P_s, two acceptor levels of C_i are pushed deep into the gap - a result of the Coulomb interaction between C_i and P_s^+. The positions of the gap levels will therefore increase

Table 4.2: Acceptor and donor transitions for the C_i-P_s defect in silicon (eV).

Configuration	Acceptor (-/0) level	Donor (0/+) level
IB	E_c-0.07	E_c-0.39
IA	E_c-0.38	E_v+0.48
IIA	E_c-0.26	Mid gap
IIB	—	E_c-0.32
III	—	E_c-0.23

as the distance between the C and P atom increases, consistent with the observations in DLTS [26]. Estimations of C_i-P_s separation in all configurations were made using the Coulomb model, these were 1.9 (IA), 3.8 (IB), 2.4 (IIA), 5.0 (IIB) and 8.5 Å(III) respectively.

DLTS with optical excitation

Experiments using optical excitation were also performed in DLTS. The optical excitation has advantages over electrical injection where the applied forward bias means that the Fermi level is undefined. If reverse bias is applied to the diode, E_F is lowered in the gap corresponding to increased irradiation dose, therefore enabling a monitoring of the charge states of the defects in the DLTS experiments [26]. In a lightly irradiated sample $E_F \approx E_c$-0.17 eV, optically induced conversion to the IB configuration was only possible under reverse bias conditions, indicating a charge state dependence for the conversion between these structures. The same result is observed in EPR [26].

Optical injection at higher temperatures (T≈240 K), where IA→III were also performed under differing bias conditions. Under reverse bias, no conversion was observed, instead a slow decrease in the IA signal with an increase in the E_c-0.10 eV due to C_i indicated breakup of the defects. With zero bias at 240 K, optically induced conversion from IA→IIA,IIB was observed. Configuration III could only be obtained by electrical injection, and III→IIA,IIB back conversion was subsequently achieved at lower temperatures (T<185 K) than were observed for this process under electrical injection.

The (extremely complex) configurational coordinate diagram for the defect was then constructed, and on the basis of the DLTS and later EPR work structural models for IA, IB and IIA forms of the defect were proposed. The arguments leading to the models for the metastable C_i-P_s pairs was as follows: the Coulomb model indicated that the C-P separation was least in the the IA configuration, and increases in the metastable configurations suggesting that one of the impurity atoms moves away from the other. The barrier height for the process was similar to that of C_i migration which suggested that the carbon atom moved away from the phosphorus. The changes in electronic structure were consistent with the formation of a C_i^- defect, with a near by P_s^+ (applicable to the neutral defect), with the Coulomb interaction pulling the C-related levels down as C_i approaches P_s.

From the observed behaviour of the defect under minority carrier injection in DLTS, it was argued that recombination-enhanced migration of C_i plays an important role in the formation of the metastable configurations. This is the case because the the barriers for conversion between the configurations is in general large $\approx$0.5 eV for IA to IB, and between 0.8-1.3 eV for the barriers between III, IIB, IIA and IA. These results are reflected in the injection and annealing temperatures required for the conversion. Above room temperature, electrical injection leads to the migration of C_i and breakup of the defect; whereas below this temperature, the metastable defects (IIA, IIB, III) are formed - which return to IA upon annealing. The authors argued that the metastable configurations were C_i-P_s defects where C_i migrated a short distance from the P_s before its trapping at a near by lattice site. Metastable configuration IB is somewhat different in that it could be formed under injection and subsequently returned to IA at significantly lower temperatures that the others.

Chapter 5

Self-interstitial related defects in Si

5.1 Introduction

The isolated self-interstitial remains an elusive defect to the experimentalist, and most of the current knowledge comes from theoretical modelling. The defect is believed to be highly mobile, and is rapidly trapped by other impurity atoms. Of particular interest is the reaction between Si_i and H, since this may facilitate the understanding of the isolated defect. Here we consider the Si_iH and Si_iH_2 defects, and compare these to recent experimental findings. Low temperature proton implantation of Si results in the formation of native defects complexed with hydrogen. Two hydrogen related local vibrational modes at 1989.4, 1986.5, 748.0, and 743.1 cm^{-1} in Si, and these are attributed to stretch and wag modes of the $\langle 100 \rangle$ oriented self-interstitial complexed with two hydrogen atoms. We present calculations on the structure of the self-interstitial, followed by the structures and vibrational modes of the singly and doubly hydrogenated interstitials in both Si. These results compare favourably with the observations confirming the assignment. Finally, interstitial aggregates are discussed, and it is shown that the binding energy of Si_i to a $\{311\}$ like chain is large, consistent with the high stability of this defect.

5.2 Calculational details

For the self-interstitial split configurations, and their interaction with hydrogen, a 132-134 atom cluster $Si_{72}H_{60+n}$ was used, where $n \in \{0, 1, 2\}$. The additional host atom and n hydrogen atoms were added to the centre of the tetrahedral 131 atom cluster which represents the perfect crystal. The following basis set was used for both the wave function and charge density in these calculations: eight Gaussian $s-$ and $p-$ orbitals of different exponents were centred on the nuclei, and a fixed linear combination eight $s-$ and $p-$ Gaussian orbitals was used on the remaining silicon atoms. Four $s-$ and $p-$ Gaussian orbitals of different widths were used on the n defect hydrogen atom, while a linear combination of three of each of these sets of orbitals was used on the terminating hydrogen atoms. In addition to this, three $s-$ and $p-$ Gaussian orbitals were placed at the centre of every bond in and around the defect core.

For the calculations of I in various configurations and charge states, an additional Si atom was added to a tetrahedral 297 atom cluster, $Si_{181}H_{116}$, and all Si atoms were allowed to relax. A similar basis to that described above was used in these calculations with a larger number of Si atoms having the differing exponents. For

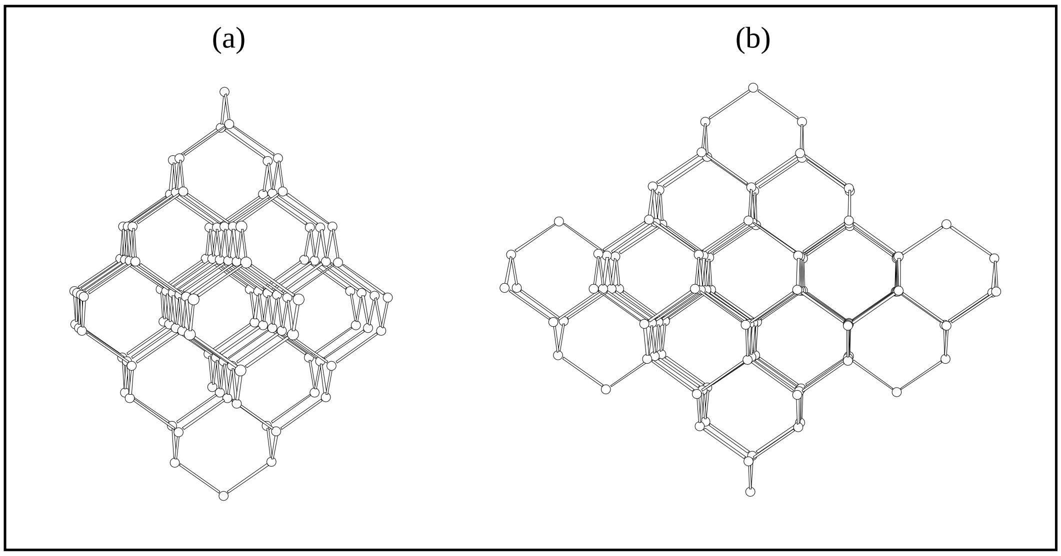


Figure 5.1: The 180 atom cluster $Si_{100}H_{80}$ used in the {311} defect calculations, (a) viewed in the plane defined by the [100] and [011] directions, and (b) the plane defined by the [100] and $[0\bar{1}1]$ directions. The terminating H-atoms are not shown.

the {311} defect, a 180 atom cluster $Si_{100}H_{80}$, elongated along $\langle 110 \rangle$ was used. This cluster is shown in Figure 5.1. For these calculations, Gaussian orbitals were placed at the centre of every bond, and the number placed on each nucleus was reduced to four. Again, Gaussian orbitals of different exponents were used on a large number of Si atoms. Four additional Si atoms were inserted into the centre of the cluster, and the atomic positions set up to mimic the proposed structure for the agglomerate.

Full symmetry constraints were used, and all atoms were relaxed until the equilibrium geometry was achieved. For the IH and IH_2 defects, the second derivative of the energy with respect to atomic displacement was calculated directly for the inner $6+n$ atoms. The silicon Musgrave-Pople potential was used for the remaining atoms, the terminating hydrogens were assumed to be infinitely massive, and the local modes were calculated.

5.3 The self-interstitial (Si_i)

5.3.1 Previous work

The vacancy and the self-interstitial defect are the fundamental intrinsic point defects in crystalline silicon. Both defects are created by neutron, electron or ion irradiation. In addition, the self-interstitial defect can be "grown in" during single crystal growth, or created in agglomeration of oxygen into SiO_2 regions, either in bulk or by surface oxidation. The vacancy has been extensively studied [43] and references therein, its migration energy is low, around 0.3 eV, and it is detected in large numbers in irradiated material at 4.2K. Consequently, the vacancy is one of the best characterised point defects in any crystalline material. In contrast to this, the self-interstitial has never been detected directly. Its presence can however be inferred by the generation of interstitial impurity defects: such as boron, carbon, aluminium, and gallium to name a few, following irradiation. As expected the net concentration of these defects is close to that of the vacancy concentration. It is believed that the interstitials are highly mobile during irradiation process, and the interstitial impurities are created by a kick-out mechanism. The migration of the impurities (X) like those listed above, and transition metals is often significantly enhanced in the presence of Si_i, in the following way:

$$\mathrm{Si}_i + \mathrm{X} \rightarrow \mathrm{X}_i + \mathrm{Si} \tag{5.1}$$

and the X_i has a significantly lower migration energy than the substitutional impurity. The X_s defect is re-formed upon the reverse "kick-out" of the interstitial.

As a result of this, most of the present knowledge of the self-interstitial comes from the numerous theoretical calculations on the defect, some of which are discussed below. The defect structure and migration mechanism are currently of great theoretical and technological interest, the determination of the ground state structures are first discussed.

The structure of Si_i

Even in the case of the isolated self-interstitial in tetrahedrally bonded semiconductor like Si, there are several possible configurations for the self-interstitial. Amongst these, are the high symmetry structures: $\langle 100\rangle$ split-interstitial possessing D_{2d} symmetry (this will Jahn-Teller distort to the lower C_2 symmetry), the $\langle 110\rangle$ split-interstitial with C_{2v} symmetry, the bond centred interstitial with D_{3d} symmetry (the symmetry is lowered to C_{1h} if the Si_i bows out of the centre of the bond), the tetrahedral interstitial with T_d symmetry, and finally the hexagonal interstitial which has C_{3v} symmetry. These configurations are illustrated in Figure 5.2(a-e) respectively. A number of lower symmetry configurations may exist, being perturbations of the above structures. There has been a large number of calculations on the ground state structure of the self-interstitial in silicon. The majority of these (all of the first principles calculations) find that the ground state structure of Si_i^0 is the $\langle 110 \rangle$ split-interstitial, see for example [44, 45, 46, 47, 48, 49, 50, 51], although not all of these were first principle calculations. Several of the above authors also explicitly calculated the formation energies of the vacancy and self-interstitial, and obtained values in the region 3-4 eV for both of these defects. In addition to the high symmetry defects, the calculations performed by Nastar *et al* [48] found an extended interstitial configuration labelled I_e as the ground state; and Clark *et al* found a low energy metastable state at 0.16 eV above the $\langle 110 \rangle$ configuration which was a distorted bond-centred-like defect. The results of Chadi [44], based on 32 atom DFT supercell calculations were as follows: the $\langle 110 \rangle$ was the ground state configuration in all of the charge states investigated (2+,+,0,-). A number of other low energy configurations were found, among these the hexagonal interstitial was around 0.12 eV above the ground state for $Si_i^{(0,+)}$, and the BC configuration was within 0.4 eV of the ground state in all charge states investigated. The energies of the $\langle 100 \rangle$ split-interstitial were consistently above the $\langle 110 \rangle$ in all charge states. Regarding the electronic structure of the defect, it was found that the $\langle 100 \rangle$ configuration gives rise to almost degenerate single and double acceptor states at E_v+1.0 eV. These arise from p_x and p_y orbitals on the interstitial atoms. The threshold energy for optical excitation from one of these levels to the conduction band was estimated to be 0.5 eV.

Mainwood [46] considered the self-interstitial defect using an approximate molecular orbital theory which incorporated Car-Parinello like dynamics. Finite clusters of between 70 and 100 atoms were used, terminated by sp^3 hybrids which were not allowed to relax. As well as the ground state energies, the speed of the method allowed calculations into the migration barrier and pathway of the defect. The results of the calculations are given in Table 5.1. It is interesting to note that here, in contradiction to the results of Chadi [44], the defect is expected to be bi-stable, the ground state configuration varying with charge state. As will be demonstrated in the next section, this can have an important effect on the migration of the defect.

Nastar *et al* [48] considered the self-interstitial defect in the neutral charge state, focusing on the low energy structures and saddle point configurations, using a Stillinger-Weber potential. For the ground state configuration, a new structure called the extended configuration and labelled I_e was found as the ground state, in

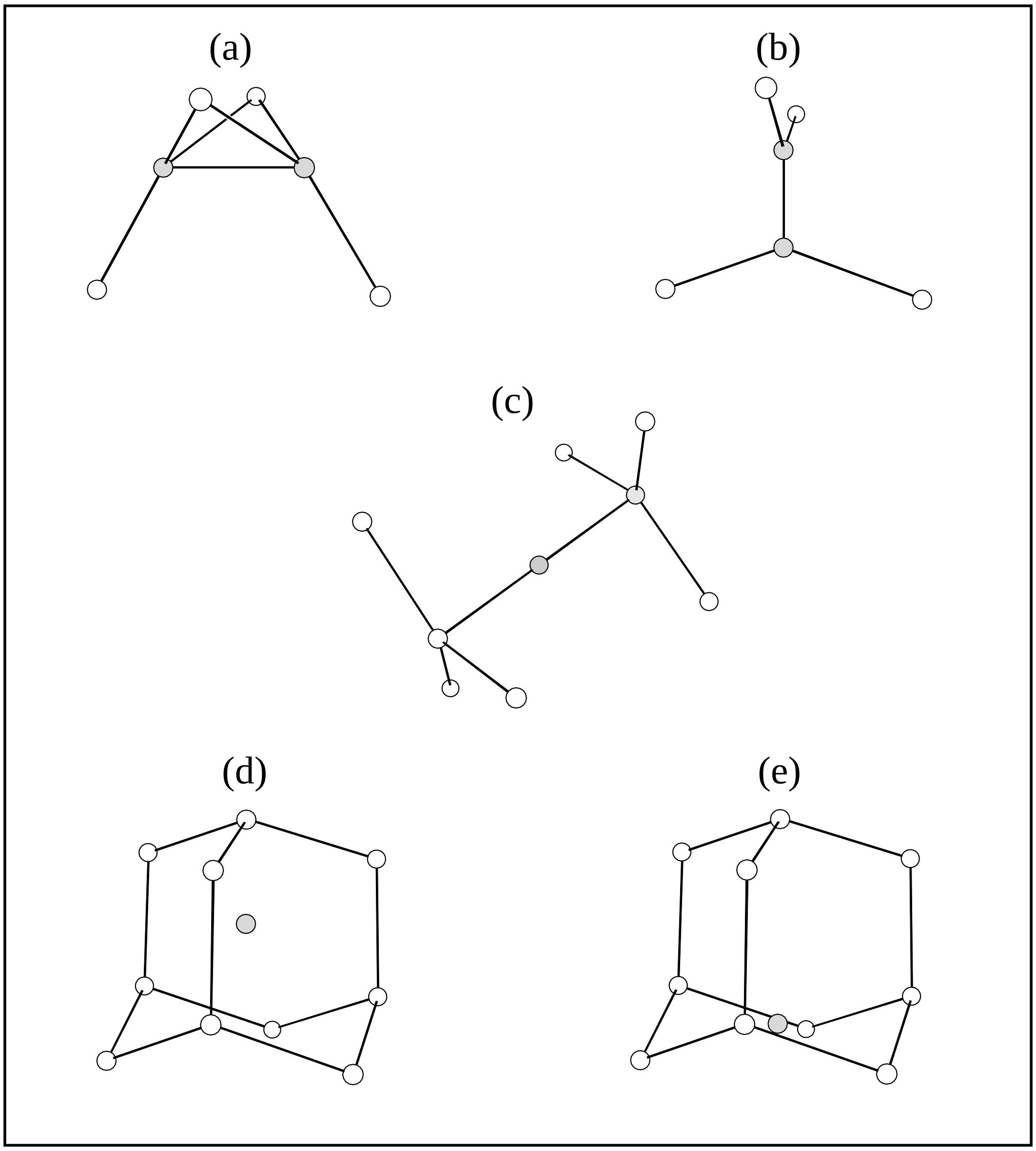


Figure 5.2: The high symmetry configurations for the self-interstitial defect in Si. The configurations are as follows: (a) $\langle 110 \rangle$, (b) $\langle 100 \rangle$, (c) bond centred, (d) tetrahedral and (e) hexagonal. In (a) (b) and (c), the hatched atoms are at a single lattice site; in (d) and (e) the hatched atom is the interstitial in the open cage region of the crystal. All other atoms are perturbed by a lesser amount form their bulk positions.

Table 5.1: Energies (eV), relative to the ground state, of various charge states of the Si_i defect, from [46].

Configuration	Si_i^{2+}	Si_i^{+}	Si_i^{0}	Si_i^{-}
$\langle 110 \rangle$	1.5	0.0	0.0	0.0
$\langle 100 \rangle$	3.2	1.7	2.0	0.7
Bond centred	3.5	2.1	2.0	1.3
Hexagonal	0.0	0.2	1.6	0.5
Tetrahedral	0.4	0.5	2.3	1.8

contradiction with the majority of other calculations. The formation energy of this structure was 3.91 eV, and the energies (eV) of the other well known configurations, relative to the ground state are as follows: $\langle 110 \rangle$ 0.77, $\langle 100 \rangle$ 1.68, bond centred 2.09, and hexagonal 3.05. In fact, the last two of these were unstable and relaxed to a lower energy configuration, and were referred to as second and third order saddle point configurations respectively.

The results of Zhu *et al* [49] suggest almost identical formation energies for the vacancy and self-interstitial, with values of 3.65 and 3.70 eV respectively. These calculations were LDA based supercell calculations using 32 and 64 atom supercells. Again, of the structures investigated the $\langle 110 \rangle$ oriented defect was lowest in energy. Close in energy to this was the hexagonal interstitial, which was 0.1 eV above the ground state. The tetrahedral interstitial was much higher in energy, being 1.4 eV above the $\langle 110 \rangle$ configuration.

In summary, the majority of calculations, and indeed, all of the recent first-principles ones find that the ground state structure of the neutral Si_i defect is a $\langle 110 \rangle$ split-interstitial. However, there is some disagreement over the lowest energy structure in various charge states, as well as the relative energies of the metastable configurations.

The migration of Si_i

Experimentally, the data on Si self-diffusion show that:

$$D(T) = D_0 \exp(-E_a/kT), \tag{5.2}$$

and thus, the coefficient exhibits Arrhenius behaviour. The activation energy for both intrinsic defects was in the range 4-5 eV, and are summarised as follows:

$$C_I^* d_I = 914 \exp(-4.84/kT) \quad \mathrm{cm}^2/\mathrm{sec}, \tag{5.3}$$

$$C_V^* d_V = 0.6 \exp(-4.03/kT) \quad \mathrm{cm}^2/\mathrm{sec}, \tag{5.4}$$

for the interstitial and vacancy respectively [52]. Here, $d_{I,V}$ and $C^*_{I,V}$ are the single defect diffusivities and their equilibrium concentrations respectively. The activation energy here includes the formation energy, and therefore the barrier for migration of Si_i will be $\approx$1 eV. At low temperatures, vacancies dominate the self-diffusion, because of their lower migration energies, whilst at higher temperatures, the interstitial mechanism dominated due to the large prefactor compared to that of the vacancy (equations 5.3 and 5.4). It must be stated however that due to problems determining the absolute concentration of self-interstitials, estimates of the migration barrier for Si_i range between 1.0-3.5 eV [53].

Presently, calculations for the migration of the neutral self-interstitial generally yield values around 1-1.5 eV, typical for the breaking and re-making of one or more Si-Si bonds. It is therefore difficult to understand the observation that the Si_i defect appears to be extremely mobile even at 4.2K, since this would imply that the migrations barrier barrier was $< 4\times10^{-4}$ eV! Such a low barrier is inconsistent with high-temperature diffusion results, and the theoretical calculations.

The most likely explanation for the low temperature migration is that of electron assisted transport, originally proposed in [54], which also accounts for the migration of the silicon vacancy. In this mechanism, although the barriers between two configurations may be high, if the migration barriers and/or ground state configurations are charge state dependent (the later represents bi-stability), then the migration can in principle proceed in a complicated way involving various saddle points, and the rapid capture and emission of electrons [55]. The barrier is therefore effectively lowered by cyclic changes in charge state during the migration which lead to the athermal diffusion. It may also be the case that if the Si_i^0 and Si_i^{2+} defects have different ground state configurations, then the Si_i may be a negative-U defect, not existing in the singly charged states [56].

In light of the significance of the charge state on the migration of the defect, a number of workers have calculated the relative energies of the various interstitial configurations in a number of charge states. Early work [55, 56, 57, 58] using *ab initio* supercell techniques, found a charge-state dependence on the structure of Si_i, and a configurational coordinated diagram was constructed based on the theory. For Si_i^{2+}, the T_d site was most stable, lying between 0.5-0.7 eV below the BC site. For Si_i^{+} and Si_i^{0}, the T_d site is no longer the ground state, instead, there was a number of low energy configurations including the BC, hexagonal, and $\langle 110\rangle$ split-interstitial. Negative-U behaviour was also predicted for the defect [57, 58] by both of these groups.

Blöchl *et al* calculated the self-diffusion constants in silicon, and found that the dominant process was due to Si_i. Here, the diffusion constants for the vacancy and concerted exchange mechanisms proved to be 3 and 4 orders of magnitude less than that due to Si_i. From the energies and entropies of formation, the authors concluded that the concentrations of Si_i and vacancies were in the regions of 10^{12}-10^{13} and 10^8-10^9 cm^{-3} at 1000°C. Although not explicitly calculated, the migration energy of Si_i was estimated to be in the range 0.7-1.5 eV.

Calculations of the migration barrier and pathway of the neutral defect, not involving any charge switching mechanism yield values in the range 0.2-1.5 eV [46]. None of these were first principles calculations which would prove much too time consuming considering the number of possible pathways and structures involved in the process. For example the calculations of Nastar *et al* find that the BC, and hexagonal saddle points at 2.09 and 3.05 eV above the ground state can link more than one pair of low energy configurations. Of course alternative paths exist, and they were able to show that the $\langle 110\rangle$ split configuration can diffuse through out the lattice with an energy of 0.65 eV; the diffusion of the ground state, I_e, was able to proceed via the $\langle 110\rangle$ configuration with an activation energy of 1.62 eV. Recent tight-binding molecular dynamics studies of self-diffusion resulted in values of 0.1 and 1.37 eV for the vacancy and interstitial respectively [50]. In these calculations, long range diffusion pathway of Si_i involved $\langle 110\rangle$-T_d-$\langle 110\rangle$ structural changes. The I_e configuration was unstable in the tight-binding calculation, spontaneously relaxing into the $\langle 110\rangle$ configuration.

5.3.2 Aimpro results

Before investigating all of the interstitial configurations in various charge states and high symmetry configurations, the neutral split-interstitial configurations were first considered in a 132 atom cluster, $Si_{72}H_{60}$. The symmetry was enforced to C_{2v} and (C_2, or C_{1h}) for the $\langle 110\rangle$ and $\langle 100\rangle$ structures respectively. The point group of the unrelaxed $\langle 100\rangle$ defect is in-fact D_{2d}, and the defect possesses a partially occupied e (doublet) level around mid-gap. Therefore the structure is a candidate for a Jahn-Teller distortion, and lifts the tetragonal symmetry, splitting the partially occupied doublet level into two singlets. The C_2 structure which could prevail as a result of the distortion proved unstable, spontaneously relaxing into the $\langle 110\rangle$ structure. The C_{1h} structure proved to be stable, here, there is a large displacement of one of the two core atoms, which moves along the [110] direction. This structure is illustrated in Figure 5.3. The $\langle 110\rangle$ proved to be the lowest energy configuration of the two, lying 0.31 eV lower in energy than the $\langle 100\rangle$ derived structure. The atomic positions for the $\langle 110\rangle$ Si_i defect from our 132 atom calculations are given in Table 5.2, and are in extremely good agreement with the structure obtained in [47].

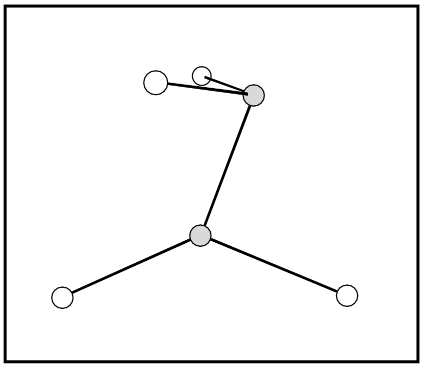


Figure 5.3: The C_{1h} split-interstitial $\langle 100 \rangle$ derived configuration.

Table 5.2: Coordinates for the $\langle 110 \rangle$ split-interstitial defect in in Si, relative to cube unit cell of length 5.43 Å, (a). The results of the previous calculation by [47], (b) are included for comparison.

Si I		(a)			(b)	
	x	y	z	x	y	z
Si_1	0.147	0.147	-0.089	0.157	0.157	-0.124
Si_2	-0.147	-0.147	-0.089	-0.157	-0.157	-0.124
NN1	-0.255	0.255	-0.272	-0.262	0.262	-0.262
NN2	-0.284	-0.284	0.283	-0.267	-0.267	0.271
NN3	0.255	-0.255	-0.272	0.262	-0.262	-0.262
NN4	0.284	0.284	0.283	0.267	0.267	0.271

To investigate other high symmetry configurations of Si_i, and the dependence of the structure on charge state a larger cluster was necessary, and a 298 atom cluster $Si_{182}H_{116}$ was used. Before detailing the results, a few points should be made regarding the calculations: (a) because the $\langle 100 \rangle$ configuration was consistently higher in energy than the $\langle 110 \rangle$ split in both the previous calculations and our 132 atom results, we did not consider it further, and (b) before the introduction of the interstitial atom, the 291 atoms cluster used possessed T_d symmetry; in the runs, the defect symmetry was forced to be as high as possible when the cluster did not allow the perfect representation of the defect point group. It is therefore plausible that any high energy configuration found could in principle relax to a lower energy configuration on the removal of the constraint.

The results of the calculations are given in Table 5.3. We first note that the bond-centred C_{3v} structure is consistently $\geq$1 eV above the *ground state* structure and will not be discussed further. The first striking result is the defect adopts the $\langle 110 \rangle$ structure in the neutral and negative charge states, whereas the tetrahedral configuration is the most stable in the singly and doubly positive charge states.

This is in agreement with several of the previous calculations discussed above, and is extremely important as it is consistent with the suggestion that a charge state cycling mechanism can account for the migration of the defect at 4.2K. The defect is therefore yet another example of a bi-stable centre in silicon. The results for Si_i^0 closely resemble those of [44, 49], with a small energy difference existing between the $\langle 110 \rangle$ and the hexagonal configurations.

Table 5.3: Energies (eV), relative to the ground state, for the Si_i defects investigated in the 298 atom cluster $Si_{182}H_{116}$ using AIMPRO.

Configuration	Si^{2+}	Si^{+}	Si^{0}	Si^{-}
$\langle 110 \rangle$	1.99	0.72	0.00	0.00
BC	0.92	0.99	1.19	1.42
T	0.00	0.00	0.45	0.79
H	1.02	0.29	0.02	0.52

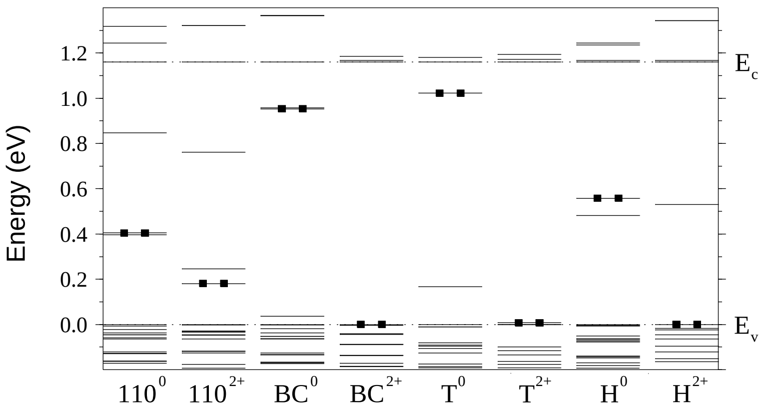


Figure 5.4: The Kohn-Sham electronic levels of the neutral and double positive charge states of Si_i, in its various configurations. Note that the eigenvalues have been scaled to give the Si band gap correctly.

All of the interstitial configurations are electrically active, and the Kohn-Sham levels for the neutral defect are shown in Figure 5.4. Note that in this diagram, and in all other diagrams of gap levels presented in this chapter, the Kohn-Sham band gap is scaled to the experimental gap of 1.16 eV. In agreement with Chadi [44], we find that the neutral $\langle 110 \rangle$ oriented defect has a fully-occupied doublet level in the gap. The levels for the charged defects are qualitatively the same, with the removal or addition of a number of electrons. There are subtle differences though, for example in the 2+ charge state, the $\langle 110 \rangle$ defect is electrically active whereas the tetrahedral, hexagonal, and bond-centred configurations are not. Indeed, this could explain the increasing stability of the other configuration over the $\langle 110\rangle$ with increasing positive charge, and in particular, the caged interstitial sites..

Within the cluster approach, we cannot calculate the formation energy of the defect, and it is difficult to ascertain whether it is a negative-U centre. In the neutral charge state, the hexagonal configuration is almost degenerate in energy with the $\langle 110\rangle$ structure, and therefore it is likely that the migration pathway(s) will involve the $\langle 110\rangle$, tetrahedral, and hexagonal configurations. Although we cannot come to any conclusion regarding the migration barrier if the charge state is varied, we can comment on the migration of the defect in a given charge state even though we did not attempt to calculate the barrier: in the neutral charge state for example, the migration is likely to occur between the $\langle 110\rangle$ and the tetrahedral configurations. The energy difference of 0.45 eV between these two does not represent the migration barrier, but is a lower bound to it; the tetrahedral configuration is metastable, and a barrier $\geq$ 0.45 eV will exist for the structural change. A pathway connecting the tetrahedral and hexagonal sites, also with energy $\geq$ 0.45 eV should also be possible. For Si_i^{2+}, the migration pathway is likely to be between the tetrahedral and hexagonal sites only, with an energy $\geq$ 1.02 eV.

5.4 Hydrogen interacting with Si_i

The high mobility and reactivity of the Si_i defect means that it is unlikely that it will be detected experimentally as an isolated species. Therefore, in order to obtain the micro-structural information concerning the defect, it is necessary to consider stable reaction products that closely resemble Si_i: namely Si_iH and Si_iH_2, formed when H interacts with the dangling bonds of a split interstitial. It is well known that H interacts strongly with dangling bonds in Si [59] where the vacancy-hydrogen complexes were identified. The Si-H bonds can form in open parts of the lattice, and as a result will not exert an appreciable strain on the surrounding lattice, and therefore the re-arrangement of the interstitial following the trapping of a H atom should be moderate.

5.4.1 Previous work

Experimental results

In proton implanted Si, a number of high frequency modes (around 20) are observed by IR spectroscopy in the range 1800-2250 cm^{-1}, and 700-850 cm^{-1}. These are now known to correspond to Si-H stretch and wag modes respectively, since the high modes shift by a factor of $\approx 1/\sqrt{2}$ when deuterons are implanted. Stein [60] first observed a line around 1980 cm^{-1}at room temperature, which proved to be a doublet, with two lines being observed at 1987 and 1989 cm^{-1} at low temperatures (10K). These lines were observed in hydrogen doped Si following irradiation, and their identical annealing behaviour showed that they arised from the same defect. Xie *et al* [61] originally assigned the lines to the VH_2 complex, however recent work [59] identified local modes at 2121 and 2144 cm^{-1} as the stretch modes of VH_2.

In recent work [62] four local modes were observed at 743.1, 748.0, 1986.5, and 1989.4 cm^{-1}, and the annealing studies showed that they were due to the same defect. It should be noted that these lines occur well below those of the VH_n complexes in Si, which lie in the range 2100-2250 cm^{-1}. The four lines appeared together, always with the same relative intensities, and annealed out together at 225°C. When the sample was implanted with deuterons, the stretch modes shifted downwards to 1446.1 and 1448.2 cm^{-1}. In addition to these, three additional lines are observed at 745.7, 1447.3, and 1987.8 cm^{-1}, again with the same annealing conditions as those described above. These are due to the isotopically mixed configurations of the same defect, and the observation of only one additional stretch and bend mode in the mixed case indicated that the defect contained two equivalent H atoms.

With the application of uniaxial stress, the stress induced splittings of a local mode, as well as the dependency of its intensity with polarisation can yield a vast amount of information on the defect [23]. If the monoclinic-II C_2 symmetry is considered for example, then the stress induced splittings, and intensities of the LVMs, is given in Table 5.4.

Table 5.4: The first-order stress pattern for a centre with monoclinic-II symmetry [23]. The dipole moment is parallel to the C_2 axis for the $|A\rangle \rightarrow |A\rangle$ transition. For the $|A\rangle \rightarrow |B\rangle$ transition, the dipole moment is perpendicular to the C_2 axis, with the angle ϕ to a $\langle 100\rangle$ direction in this plane. The full table of splitting given below was derived in [62].

F $\parallel$	Frequency shift (σ)	$\lvert A\rangle \rightarrow \lvert A\rangle$			$\lvert A\rangle \rightarrow \lvert B\rangle$		
[100]		$I_{\parallel}$	$I_{\perp}$		$I_{\parallel}$	$I_{\perp}$	
	A_1	4	0		0	2	
	A_2	0	2		$4\sin^2\phi$	$2\cos^2\phi$	
	A_3	0	2		$4\cos^2\phi$	$2\sin^2\phi$	
[111]		$I_{\parallel}$	$I_{\perp}$		$I_{\parallel}$	$I_{\perp}$	
	$\frac{1}{3}(A_1 + A_2 + A_3 + 2A_4)$	2	2		$2+2\sin 2\phi$	$2-2\sin 2\phi$	
	$\frac{1}{3}(A_1 + A_2 + A_3 - 2A_4)$	2	2		$2-2\sin 2\phi$	$2+2\sin 2\phi$	
[110]		$I_{[110]}$	$I_{[001]}$	$I_{[1\bar{1}0]}$	$I_{[110]}$	$I_{[001]}$	$I_{[1\bar{1}0]}$
	$\frac{1}{2}(A_2 + A_3 + 2A_4)$	0	2	0	$1+\sin 2\phi$	0	$1-\sin 2\phi$
	$\frac{1}{2}(A_2 + A_3 - 2A_4)$	0	2	0	$1-\sin 2\phi$	0	$1+\sin 2\phi$
	$\frac{1}{2}(A_1 + A_2)$	2	0	2	$2\sin^2\phi$	$4\cos^2\phi$	$2\sin^2\phi$
	$\frac{1}{2}(A_1 + A_3)$	2	0	2	$2\cos^2\phi$	$4\sin^2\phi$	$2\cos^2\phi$

Here it was assumed that the stress-induced perturbation is linear in the components of σ_{ij} of the stress tensor. The electrical dipole moment involved in the optical transition is either parallel or perpendicular to the C_2 axis ($\langle 100 \rangle$), and the results obtained for both polarisations are shown in the above table. The stress induced splittings of the line are consistent with the monoclinic-II shifts, and from the data, the azimuthal angle was estimated to be ϕ=0±8°. Moreover, the symmetry of the stretch modes were determined, with the B mode at 1986.5 cm^{-1}, and an A mode at 1989.4 cm^{-1}. The polar and azimuthal angles, θ and ϕ are defined in figure 5.5.

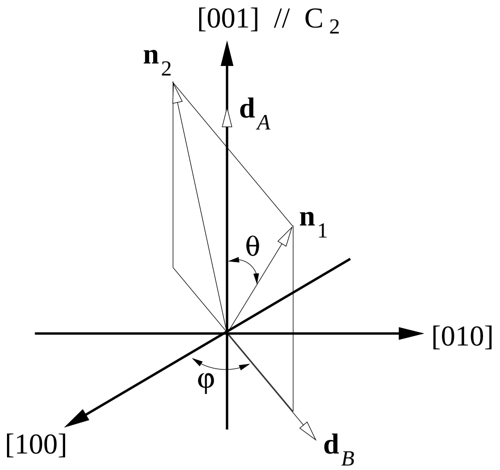


Figure 5.5: The definition of the polar and azimuthal angles θ and ϕ that specify the directions of the unit vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ for a specific orientation of the defect. The direction of the induced dipole moments for the A and B modes are also shown by $\mathbf{d}_A$ and $\mathbf{d}_B$.

The intensity ratio of A and B mode intensities at zero stress was shown to be:

$$\frac{I^A}{I^B} = \cot^2\theta, \tag{5.5}$$

and hence the polar angle of the bond dipoles was determined from the intensities of the absorption lines with zero stress. The two stretch modes yield a polar angle of θ=48±1°, whilst the bend modes gave a value of θ=43±2°.

The most simple defect satisfying all of the above criteria, is the $\langle 100 \rangle$ oriented IH_2 defect shown in Figure 5.6. This defect contains two nearly perpendicular, and equivalent Si-H bonds which are almost aligned along the [101] and [$\bar{1}$01] directions, with the C_2 axis being parallel to the [001] direction. The four local modes [H] were therefore assigned to the IH_2 defect. In-fact, the experimental determination of the micro-structure of the IH_2 defect will provide a credible check for the theoretical studies of the IH_n defects, where n =0,1,2.

Previous theoretical modelling

The idea of one or more hydrogen atoms saturating a dangling bond of the interstitial defect is not a new one, and calculations on such centres were performed as early as 1991 [63]. These calculations were semi-empirical parameratised ones (MINDO/3) using a 32 atom cyclic cluster to model the pure crystal. Only ⟨100⟩ derived structures were considered, for both the IH defect and IH_2 defects. For the IH defect, the hydrogen binding energy was found to be 1.35 eV, and the defect gave rise to a local vibrational mode at 2197 cm^{-1} which scaled to 1960 cm^{-1} when the systematic error in the calculated local mode is accounted for. For IH_2 the calculations show that the defect gives rise to two high local modes at 2210 and 2213 cm^{-1}, which scale to 1972 and 1974 cm^{-1} respectively, in good agreement with the modes observed by Budde *et al.*

More recent calculations into the self-interstitial, and its interaction with hydrogen were performed by Van de Walle *et al* [47]. For the IH defect, it was found that a perturbed ⟨110⟩ structure was favoured over any other configuration investigated. The stretch mode of the defect was calculated to be 1870 cm^{-1}, and the binding energy of H to the defect was found to be 1.34 eV (relative to bond centred H), and is smaller than the calculated binding energy of H to V. The defect was found to have a gap level at E_v+0.4 eV, which is occupied by a single electron in the neutral charge state. The IH_2 complex was then considered, and the equilibrium structure was closer to that of a ⟨100⟩ split, rather than a ⟨110⟩ split. Again, the calculated binding energy for the hydrogen atoms is around 1.3 eV, and the addition of the second H atom rendered the defect electrically inactive.

The possible configurations for the singly and doubly H-terminated split-interstitial defects are shown in Figure 5.6a-d, and it is these we focus on in 5.4.2.

5.4.2 Aimpro results

For the investigations into the IH and IH_2 defects, the defect was inserted into the centre of the tetrahedral 131 atom cluster $Si_{71}H_{60}$. We only considered the defects formed when one or more hydrogen atoms attack the dangling bonds of the split-interstitial configurations.

The IH defect

For the starting configurations, we used the isolated ⟨100⟩ and ⟨110⟩ defect atomic positions, and attached one H atom to one of the two equivalent core atoms. Both new structures possess C_{1h} symmetry, and the relaxations were carried out with this symmetry enforced. The energy difference between the two different configurations

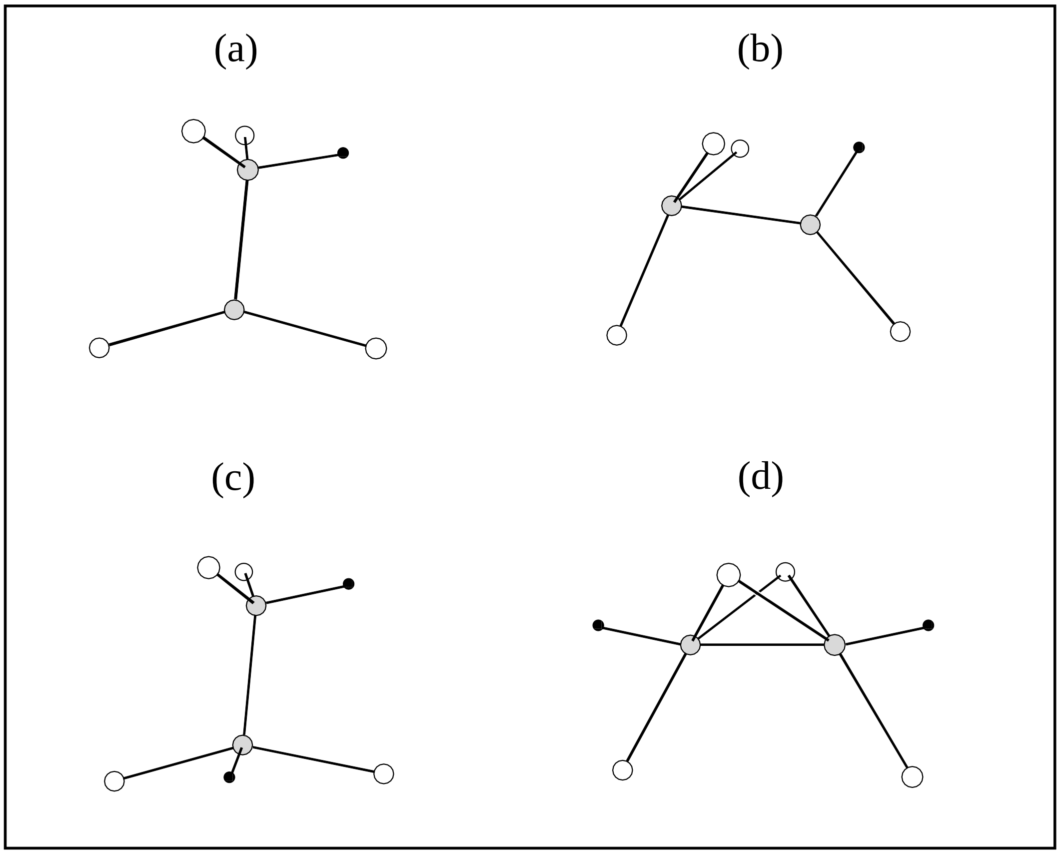


Figure 5.6: The singly and doubly hydrogenated split-interstitial defects in Si: (a) $\langle 100\rangle$ IH, (b) $\langle 110\rangle$ IH, (c) $\langle 100\rangle$ IH_2, and (d) $\langle 110\rangle$ IH_2. Again the large white atoms are Si, and the small black atom(s) are H; the speckled Si, and the H atom(s) occupy a single lattice site.

proved to be small, with $IH_{\langle 100\rangle}$ being 0.24 eV lower in energy than $IH_{\langle 110\rangle}$. This result is in disagreement with the other recent calculations [47], where $IH_{\langle 110\rangle}$ was found to be the ground state structure, and their structures were similar to the one we found for this defect. Our $\langle 100\rangle$ structure is similar to that found by Deak *et al* [63]. However, this energy difference is close to the accuracy of our calculations, and we therefore discuss both configurations. The relaxed atomic positions for the IH defects are given Table 5.5, and those for the $\langle 110\rangle$ IH defect calculated in [47] are given in Table 5.6.

Although we did not investigate the migration barrier of the IH defect, it is likely that the presence of the hydrogen stabilises the defect, and therefore its local modes should be observed at low temperatures. We therefore calculated the modes and isotope shifts for both defects, and these are given in Table 5.7. So far, there has been no experimental reports of these modes, the stretch frequencies at 2166.9 cm^{-1} for $IH_{\langle 100\rangle}$, and 2190.8 cm^{-1} for $IH_{\langle 110\rangle}$ are considerably higher than the $\langle 110\rangle$ configuration [47], but agrees reasonably with the unscaled value of 2217cm^{-1} for

Table 5.5: Coordinates for the IH defects in Si, relative to cube unit cell of length 5.43 Å.

Si IH	⟨100⟩			⟨110⟩		
	x	y	z	x	y	z
Si_1	0.202	0.045	-0.045	0.153	0.153	-0.004
Si_2	-0.191	0.006	-0.006	-0.126	-0.126	-0.092
NN1	-0.310	0.284	-0.284	-0.269	0.242	-0.269
NN2	-0.294	-0.275	0.275	-0.279	-0.279	0.265
NN3	0.306	-0.284	-0.276	0.242	-0.269	-0.269
NN4	0.306	0.276	0.284	0.322	0.322	0.321
H	0.234	0.239	-0.239	0.260	0.260	-0.231

Table 5.6: Coordinates for the ⟨110⟩ IH defect in Si from [47], relative to cube unit cell of length 5.43 Å.

Si IH	⟨100⟩		
	x	y	z
Si_1	0.118	0.118	-0.073
Si_2	-0.186	-0.186	-0.156
NN1	-0.260	0.265	-0.258
NN2	-0.256	-0.256	0.260
NN3	0.265	-0.260	-0.258
NN4	0.274	0.274	0.282
H	0.269	0.269	-0.269

the ⟨100⟩ configuration [63]. In fact our method in general overbinds the hydrogen bonds, and therefore the calculated stretch modes for H related defects are generally ≈7% to high [13, 59]. These defects are paramagnetic in the neutral charge state, and possess a deep singly occupied level, Figure 5.7.

The IH_2 defect

When two hydrogen atoms are added to the ⟨100⟩ and ⟨110⟩ split-interstitials, the resultant structures have C_2 and C_{2v} symmetries respectively. When the relaxation is carried out, maintaining these symmetries, the ⟨100⟩ IH_2 defect is the ground state structure, being 2.56 eV lower in energy than the relaxed ⟨110⟩ derived configuration. Upon removal of the symmetry constraint, the ⟨110⟩ structure rotates to the ⟨100⟩ one. Therefore there is no doubt that the low energy structure IH_2 is the ⟨100⟩ oriented defect with C_2 symmetry, and therefore it is this structure that one would expect to detect experimentally. Since all dangling bonds are now satisfied, the defect should be stable against migration, and should be detected until

Table 5.7: Local Vibrational Modes, cm^{-1}, of IH defect.

Si	⟨100⟩			⟨110⟩	
	H	D		H	D
A	2166.9	1556.3	A	2190.8	1574.1
B	743.8	578.5	A	882.0	649.7
A	724.2	578.8	B	749.0	539.3
A	603.6	604.0	A	630.8	610.1
B	577.2	547.4	A	550.4	550.3

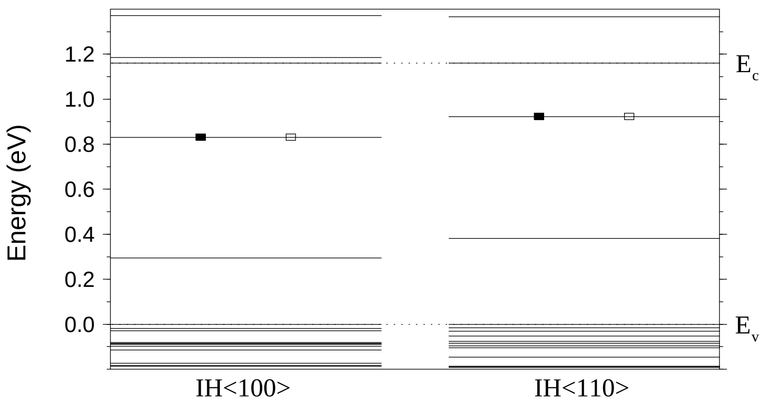


Figure 5.7: The Kohn-Sham electronic levels of the neutral charge state of the IH defect, in ⟨100⟩ and the ⟨110⟩ configurations, scaled to the Si band gap.

the temperature is high enough to boil-off the hydrogen atoms, at which point the constituents become mobile. Our structure resembles closely that from the previous calculations [63, 47], and the structural information is given in Table 5.8 below. The polar angle θ measured from our structure is 46°, and the azimuthal angle ϕ is 10° (refer to figure 5.5). These values are in excellent agreement with the experimental results [62], and along with the stability of the defect, support their assignments of the four local modes to IH_2.

Further support for the assignment comes from the calculated local vibrational modes for IH_2. The calculated values for the modes of IH_2, IHD, and ID_2 are given in Table 5.9, along with those observed experimentally. Again, there is extremely good agreement with the experimental results when the overbinding is taken into account, and the small splittings of the stretch and wag modes is reproduced fairly accurately. The isotopic shifts of the stretch modes with D is also reproduced very accurately. In both experimental observations and theoretical modelling [59, 62, 63], the Si-H stretch modes are consistently higher in the VH_n complexes, than they are in the IH_m complexes. This suggests that in the case of the IH_m defects, that

Table 5.8: Coordinates for the ⟨110⟩ IH_2 defect in in Si, relative to cube unit cell of length 5.43 Å, (a). The results of the previous calculation by [47], (b) are included for comparison.

Si I	(a)			(b)		
	x	y	z	x	y	z
Si_1	0.197	0.054	-0.042	0.195	0.063	-0.059
Si_2	-0.197	0.054	0.042	-0.198	0.063	0.059
NN1	-0.313	0.280	-0.287	-0.300	0.276	-0.285
NN2	-0.295	-0.279	0.271	-0.284	-0.282	0.272
NN3	0.295	-0.279	-0.271	0.284	-0.282	-0.270
NN4	0.313	0.280	0.287	0.300	0.276	0.285
H_1	0.237	0.244	-0.237	0.225	0.263	-0.262
H_2	-0.237	0.244	0.237	-0.225	0.263	0.262

Table 5.9: Local vibrational modes, cm^{-1}, of the IH_2 in Si.

Si	H-H	H-D	D-D
A	2144.7	2143.8	1540.2
B	2142.9	1540.1	1539.9
A	774.7	771.4	590.2
B	768.1	727.3	582.6
B	736.4	589.8	564.3
A	717.5	579.8	555.0
observed			
A	1989.4	1987.8	1448.2
B	1986.5	1447.3	1446.1
B	748.0	745.7	
A	743.1		

primarily the p-orbitals of the innermost (defect) silicon atoms contribute to the bonding of the hydrogen. The splitting in the mixed H, D case shows that the two atoms are close together and equivalent. This is also consistent with the view that the defects are interstitial and not vacancy like.

The overbinding is a result of both the use of density functional theory, and anharmonic effects. The results therefore support the assignment of the four observed local modes (H) to the IH_2 defect. However, the ordering of the wag modes is given incorrectly by the theory. For a more detailed analysis of the mode orderings and splittings, a full analysis of other effects, including anharmonicity is needed, and is beyond the scope of the current work.

Finally, some comments are made on the electrical activity of the defect. In contradiction to the previous calculation [47], and what one might expect, we find that the addition of the second hydrogen atom does not completely free the band gap of defects associated levels. The Kohn-Sham eigenvalues for the defect are shown in figure 5.8. The origin of these levels is still unclear, but it is interesting to contrast the position of the level with that of the isolated interstitial defect.

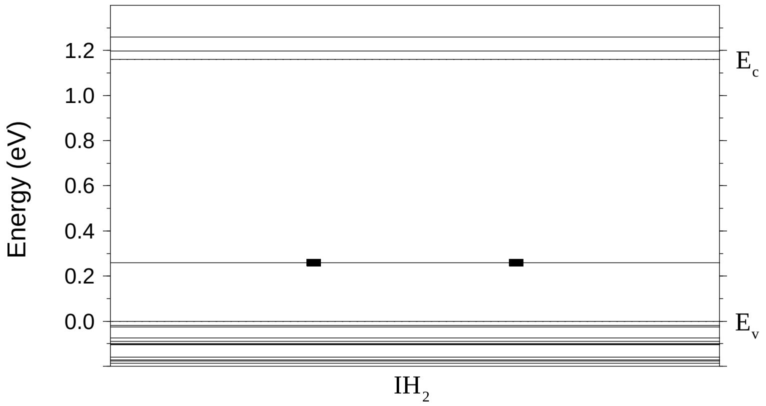


Figure 5.8: The Kohn-Sham electronic levels of the ⟨100⟩ oriented IH_2 defect, scaled to the Si band gap.

5.5 Interstitial aggregates

At higher temperatures, the effects of the migration of interstitials can have devastating effects with regards to manufacturing of devices, and the continued advancements in the reduction in their size. This is a result of their effect on the diffusion of dopants such as boron in Si. As the device size decreases further, towards the diffusion length of the dopants, then this quantity becomes a much more important as it sets a lower bound on the size attainable. The diffusion length of these dopants are greatly increased in the presence of interstitials, and this phenomena is known as transient enhanced diffusion (TED). The basic mechanism for the enhancement is as follows: The substitutional dopant, X_s, is highly stable against migration, with a barrier of several eV for the process. However on capturing a mobile self-interstitial, a highly mobile ($E_{mig} < 1$ eV) interstitial-dopant related defect X_i is formed. The diffusion process is complete when the Si_i atom is ejected leaving X_s. For example, B implantation, followed by rapid thermal annealing at 700°C for 15 minutes results in a broadening of the implanted boron profile by thousands of Å. TED lasts seconds at 950°C, but proceeds for hours when the temperature is reduced to 670°C [53].

The excess interstitials result from surface oxidation of the implantation procedure itself. Cowern *et al* [64] were able to show that the diffusion of Boron was mediated by the self-interstitial, and furthermore, that the migration proceeds through a *kick out* mechanism as described above. Anomalously, the diffusion lengths here were observed to *increase* with *decreasing* temperature.

5.5.1 Previous work

Self-interstitial defects cannot be directly detected experimentally when isolated, but can be detected as an agglomerate. Aggregates of self-interstitial defects are believed to be the constituents of the rod-like defects observed in Cz-Si crystal growth [65]. These aggregates can be created by various processes, and more importantly, those used in the fabrication of electronic devices. Agglomeration of Si_i defects can also occur as a result of the presence of impurity atoms such as O and C at temperatures as low as 350°C. Takeda *et al* [66] observed the agglomeration of self-interstitials at 450°C by high resolution transmission electron microscopy. The interstitial aggregates grow along the $\langle 110 \rangle$ direction, and extend along the $\{311\}$ plane. Information regarding the atomic configuration of the defect was deduced, and it was found that there is a large reconstruction of the lattice around the interstitials shown in Figure 5.9, the pure unrelaxed atomic positions are also shown for comparison.

The current line of thought [64, 67] is that the interstitials which result in the TED, are emitted when the $\{311\}$ rod-like defects (interstitial aggregates) dissociate at high temperatures. This result implies that the binding energy of Si_i to the aggregate is high, around 3.0 eV.

Theoretical calculations into the mechanism of TED [49] have shown that the binding energy of B to Si_i is around 1 eV. The interstitial mediated kick out mechanism was modelled, and it was found that the barrier for the kick out was 1 eV. Once formed, the defect was able to migrate through the hexagonal-tetrahedral-hexagonal pathway with an activation energy of 0.3 eV, before the reverse kick out of the interstitial which had a barrier of 0.6 eV. Since the value for the reverse kick out was much larger than the B_i diffusion barrier, the boron can migrate quite some distance before returning to the substitutional site. These values are consistent with the anomalous temperature dependence of the diffusion length observed in [64].

5.5.2 Aimpro results

Here, we focus on the structure of the $\{311\}$ defect, and demonstrate that it is indeed stable. For this a 180 atom cluster $Si_{100}H_{80}$ was used, shown in Figure 5.9a. To model the $\{311\}$ defect, which in practice contain thousands of atoms, we added *four* interstitial atoms to the centre of this cluster, along the $\langle 110 \rangle$ direction. The atomic positions of the central atoms in the cluster were arranged in such a way that they reproduced the core structure of the {311} defect: that is to say pairs of interstitial atoms were added along the hexagonal channel, and the atomic positions of the neighbouring atoms were moved so that the large scale reconstruction and unusual bonding around the core was reproduced. The structure was then relaxed with no imposed constraints, and the fully relaxed structure is shown in Figure 5.9.

The structure proposed by Takeda *et al* [66] proved to be stable, and gap levels associated with the Si_i atoms at the end of the chain were found. All other atoms, including the Si_i atoms in the middle of the chain were fully coordinated. We therefore expect that a large scale version of this structure - the {311} defect - would be electrically active. A second feature of the aggregate is its high stability. To estimate the binding energy of an Si_i defect to the {311} chain, we removed one of the atoms from our I_4-{311}-like chain and placed it at a nearby split-interstitial site, Figure 5.10. This lead to an increase in energy of around 4.16 eV, consistent with the large binding energy of the defect observed experimentally.

5.6 Discussion

In agreement with the majority of the previous theoretical calculations, we find the $\langle 110 \rangle$ configuration as the ground state of the neutral Si_i defect. This is also the stable configuration in the negative charge state, but the defect adopts the tetrahedral configuration in the singly and doubly positive charge states. A number of low meta-stable configurations always exist, although the migration barrier was not explicitly calculated, this suggests that it could be low, a value of around 1 eV in all charge states is plausible. Furthermore, it is significant that we find the defect to be bistable, in agreement with some of the other previous calculations. This is consistent with the arguments of Bar-Yam *et al* [55, 56] who suggested that cyclic changes in charge state could result in an extremely low effective migration barrier, and explaining why isolated Si_i is not observed even at temperatures as low as 4.2K.

We then considered the interaction of the self-interstitial defect with one or more H atoms. For Si_iH, we conclude that the ground state configuration was derived from the $\langle 100 \rangle$ defect. This is in disagreement with recent theoretical calculations where a $\langle 110 \rangle$ oriented IH defect was favoured. Nevertheless, the energy differences between the two were small, and it is possible that the defect is mobile at higher temperatures, but would be *frozen in* at lower temperatures. In spite of this, the defect as thus far eluded detection experimentally. An immobile and stable defect is formed when two H atoms completely passivate the dangling bonds of Si_i, and the resulted IH_2 defect has a $\langle 100 \rangle$ orientation, and possesses C_2 symmetry. The

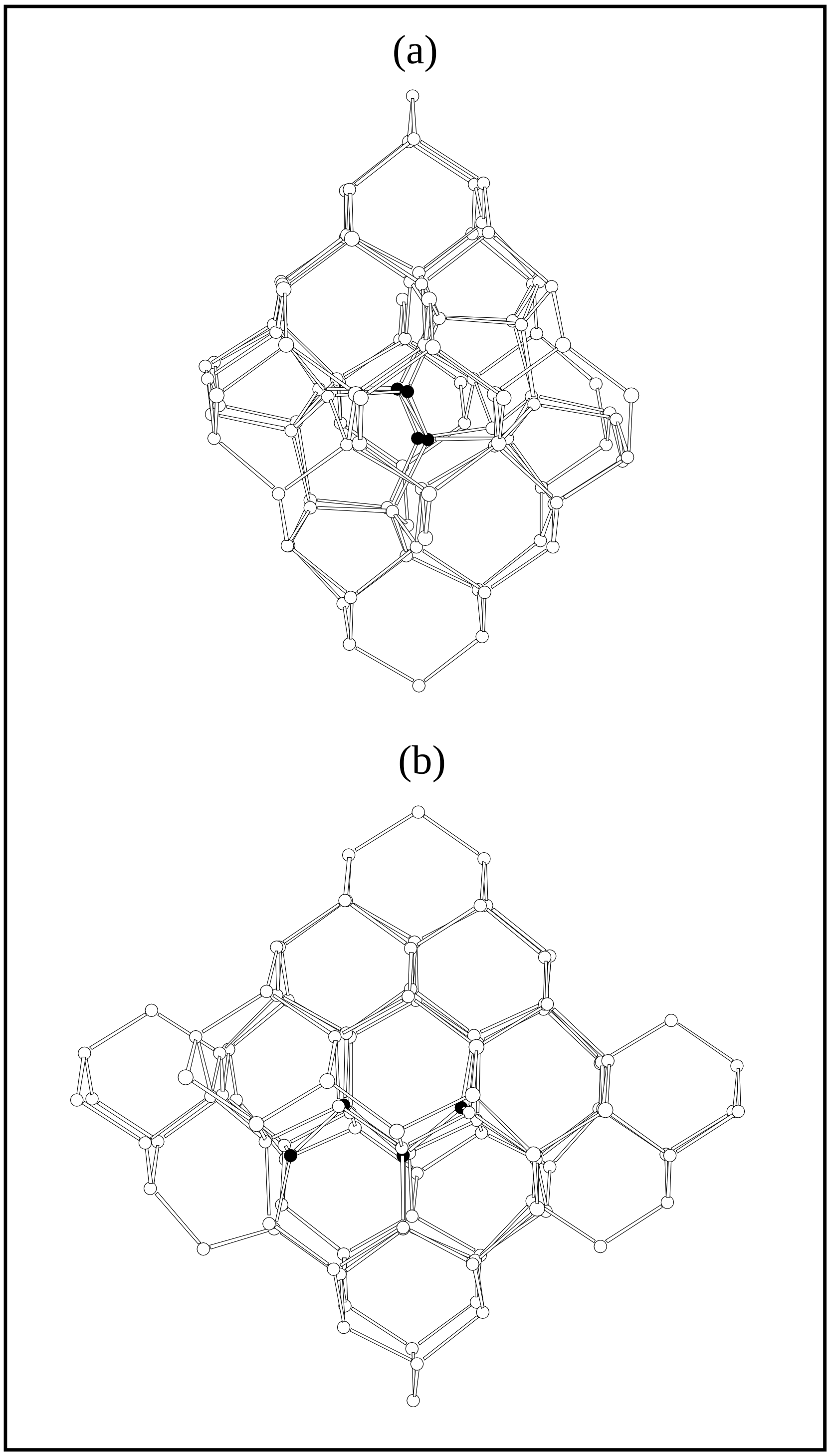


Figure 5.9: The 184 atom cluster $Si_{104}H_{80}$ used in the {311} defect calculations, (a) viewed in the plane defined by the [100] and [011] directions, and (b) the plane defined by the [100] and [0$\bar{1}$1] directions. The interstitial atoms are shown in black here for clarity, and the terminating H-atoms are not shown.

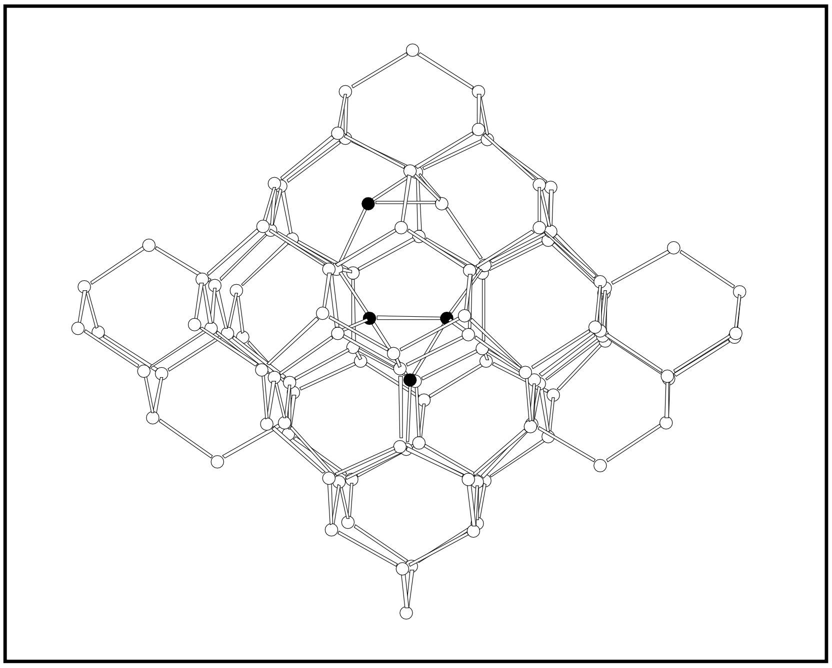


Figure 5.10: The 184 atom cluster $Si_{104}H_{80}$ used in the {311} defect calculations, here one Si_i atom has been removed from the chain to calculate the binding energy. The defect is viewed in the plane defined by the [100] and $[0\bar{1}1]$ directions, and again the interstitial atoms are shown in black.

calculations support the recent assignment of this defect as being the one giving rise to local modes at 743.1, 748.0, 1986.5, and 1989.4 cm^{-1} in H implanted Si.

Finally, preliminary calculations on the structure of the {311} defect were performed. It was demonstrated that the proposed structure was indeed stable, and furthermore, the binding energy of an interstitial to such a chain was extremely large (around 4 eV). This is consistent with the break up of these extended defects after rapid thermal annealing at T>700°C, leading to TED. The presence of other defects such as carbon, which also sinks for Si_i defects (see following chapter) suppresses TED since the {311} defects are not formed in C-rich material.

Chapter 6

Isolated carbon defects in Si

6.1 Introduction

Interstitial carbon, C_i, defects in Si exhibit a number of interesting and unexplained features. The C_i defect in the neutral charge state of gives rise to two almost degenerate vibrational modes at 920 and 931 cm^{-1} whose 2:1 absorption intensity ratio naturally suggests a trigonal defect in conflict with uniaxial stress measurements. The di-carbon, C_s-C_i, defect is bi-stable, and the energy difference between its A and B forms is surprisingly small even though the bonding is very different. In the B-form appropriate to the neutral charge state, a silicon interstitial is believed to be located near a bond centred site between two C_s atoms. This must give rise to vibrational modes which involve the motion of both C atoms in apparent conflict with the results of photoluminescence experiments. We calculate the structure and vibrational modes of these defects and find that the ratio of the absorption intensities of the local modes of C_i is in reasonable agreement with experiment even though the structure of the defect is not trigonal. We also show that modes in the vicinity of those detected by PL for the B-form of the di-carbon centre involve independent movements of the two C atoms. Finally, the trends in the relative energies of the A and B-forms in three charge states are investigated.

6.2 Calculation details

Before detailing the other defects in this chapter, and our results, details of the cluster sizes and basis are now given. A carbon atom was inserted into the central region of a 71 atom tetrahedral cluster $Si_{35}H_{36}$ to investigate the structure of C_i. The migration energy of C_i was found by inserting a C atom into a trigonal 86 atom cluster $Si_{44}H_{42}$, oriented along $\langle 111 \rangle$. A central Si atom was then replaced by C to investigate the di-carbon defect. The electronic wave functions were expanded in s- and p- Gaussian orbitals, centred on nuclei and the centre of all bonds. The basis used here for the structural and local mode calculations was as following: eight Gaussian fitting functions of different exponents were used on the carbon atom(s) and inner silicon atoms, and a fixed linear combination of eight Gaussian orbitals were placed on the remaining Si atoms. A linear combination of three Gaussian orbitals were used for the terminating hydrogens. Additionally, three $s-$ and $p-$ Gaussian orbitals were placed at the centre of every C-Si and Si-Si bond. The electron charge density was fitted in exactly the same way as the wavefunctions. The second derivatives of the energy with respect to atomic positions were calculated for the carbon atom(s) and their neighbouring Si atoms. For the binding energy of C_i to C_s a 181 atom cluster $C_2Si_{99}H_{80}$ was used.

A larger basis is required for the accurate determination of η so in this calculation the number of Gaussian orbitals used to fit the wavefunction and charge density of C and Si was increased to ten. It is the effective charge which controls the intensity ratio of the two IR absorption lines. In this work, η is calculated for the two highest local vibrational modes of C_i, having A_1 and B_1 symmetries. These modes largely involve motion of the carbon atom along [001] and [110] respectively together with movements of the Si neighbours.

6.3 The substitutional carbon defect

6.3.1 Experimental background and previous theoretical modelling

Carbon is a common and important impurity in silicon, typically occurring in concentrations of around 5×10^{15}-4×10^{17} cm^{-3}, the latter being its solubility at the melting point of silicon. Carbon is always present in Si, even if not purposely implanted, being introduced from graphitic contaminants or carbonaceous gases present during growth of crystalline silicon. The carbon atom predominantly occupies substitutional sites, C_s when in-grown with no subsequent treatment of the sample. In this configuration, all atoms are fully coordinated, and the defect is electrically inactive. The substitutional carbon defect is shown in figure 6.1, and possesses T_d symmetry. The C_s defect was first identified in the 1960's [68, 69], with the detection of an LVM at 607 cm^{-1} by IR spectroscopy, which due to the high defect symmetry is a triply degenerate T_2 mode. The effective charge η for this mode was found to be $2.4e$ [69, 70]. In isotopically enriched samples, this LVM decreased to 589 and 573 cm^{-1} for ^{13}C and ^{14}C respectively confirming the presence of carbon. X-ray studies show that the C-Si bonds are 15% shorter than the bulk Si-Si bonds [71], this, along with the fact that the C atom is much lighter than Si results in this mode being well above the Raman peak for Si which occurs

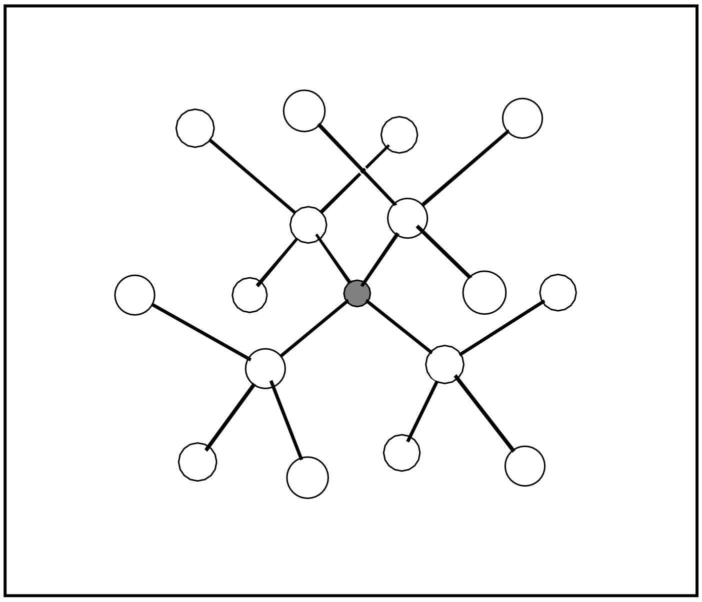


Figure 6.1: The substitutional carbon defect in silicon.

at 523 cm^{-1}. The defect has a high migration barrier (around 3.1 eV) [72] due to the fact that the defect resides at the substitutional site, making four very strong bonds. Although electrically inactive, C_s is a very important defect, because of the strain it induces in the lattice. As a result of this, C_s will act as a trap for defects which have a local compressive strain on the lattice, for example the self-interstitial Si_i, or oxygen interstitials O_i. Although the trapping of a self-interstitial leads to an electrically active defect, which is detrimental to device fabrication, recent work [73] illustrates the uses of employing carbon as a sink for interstitials: when silicon with a higher carbon content is used, the Si_i atoms are trapped by C_s forming the C_i defect, described in the next section. Therefore, the diffusion rates of the dopants such as B will be reduced to their thermal values, and transient enhanced diffusion suppressed, since the {311} defects are no longer formed. Without the carbon sinks, their enhanced diffusion can be up to one hundred times greater that of the thermal diffusion rate; also carbons presence limits the interstitial driven clustering of these dopants in silicon. However, purposely including carbon in the material for such purposes inevitably introduces the defects described in the following sections.

Theoretical modelling of C_s showed that the local modes are very sensitive to the C-Si bond length. All calculations found that the C-Si bond lengths were approximately 85% of the bulk Si-Si bonds. The calculated frequencies were in the range 561-684 cm^{-1} [74, 75] with a corresponding effective charge η of 0.85, compared to the experimental value of 2.50. MNDO calculations gave 625 cm^{-1}

for the local vibrational mode [76].

6.3.2 Aimpro results

The central atom of the 71 atom cluster, $Si_{35}H_{36}$, was replaced by C and all atoms allowed to relax. Carbon pulled the neighbouring Si atoms inwards, making C-Si bonds of length 1.96 Å. Thus the bond lengths are 84% of the Si-Si bond length, in excellent agreement with the X-ray studies and previous calculations. The calculated local mode (T_2) was found at 662.2 cm^{-1}, in fair agreement with the experimental value. This mode shifted downwards by 19.8 and 36.9 cm^{-1} with ^{13}C and ^{14}C respectively, in excellent agreement with the experimental shifts of 19 and 34 cm^{-1}.

6.4 The isolated interstitial carbon defect

6.4.1 Experimental background

More complex carbon defects are formed in radiation damaged silicon. Upon irradiation, silicon interstitials (Si$_i$) are created, and diffuse rapidly through the lattice. These self-interstitials can be captured by C$_s$, relieving the strain on both defects, and forming the interstitial carbon defect C$_i$ [34, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86]. The neutral defect was first identified by Bean and Newman (1970), and is characterised by two local vibrational bands at 920 and 931 cm^{-1}, which increase in intensity as the 607 cm^{-1} LVM anneals out. As well as correlating the emergence of these lines with the decrease of the LVM due to C$_s$, carbon was shown to be a constituent of the defect giving rise to these lines in ^{13}C and ^{14}C enriched silicon. Here, the two LVMs occurred at 892, 904 and 867, 881 for the respective isotopes. In the irradiation experiment, since the Fermi level was around mid-gap, these local modes were attributed to the neutral defect. The magnitude of these C-related modes are 300 cm^{-1} above those due to C$_s$ indicating that the carbon atom is in a more restricted environment. The absorption intensities are in a ratio of approximate 2:1 leading the earliest workers to suggest that C$_i$ was a trigonal defect [77] figure 6.2a, explaining the doubly degenerate local mode.

Later work with lightly irradiated p-type Si was able to correlate an electron paramagnetic resonance, EPR, spectrum, labelled Si-G12, with the LVMs observed in infra-red absorption. The EPR showed that the defect giving rise to G12 possessed C_{2v} symmetry [79], having a $\langle 100 \rangle$ orientation. The hyperfine splitting of G12 with ^{13}C showed that the highest occupied state consisted of a localised p-orbital centred on the C atom, and the isotope effects showed that only one carbon atom was involved in the centre. EPR studies on the preferential alignment of the defects under uniaxial stress were performed, and gave a barrier of 0.88 eV for the reorientation of the defect, as well as the piezospectroscopic parameters. Annealing studies found that the Si-G12 spectrum disappears after a 30 minute isochronal anneal at 65 °C, the same temperature range as the two LVMs associated to C_i^0 annealed out. DLTS measurements gave a donor level at E_v+0.28 eV, this level having very similar annealing characteristics to G12. These results are consistent with the assignment of G12 and the donor level to C_i^+, and suggested a model of the defect shown in figure 6.2b. More recent EPR and DLTS investigations on n-type

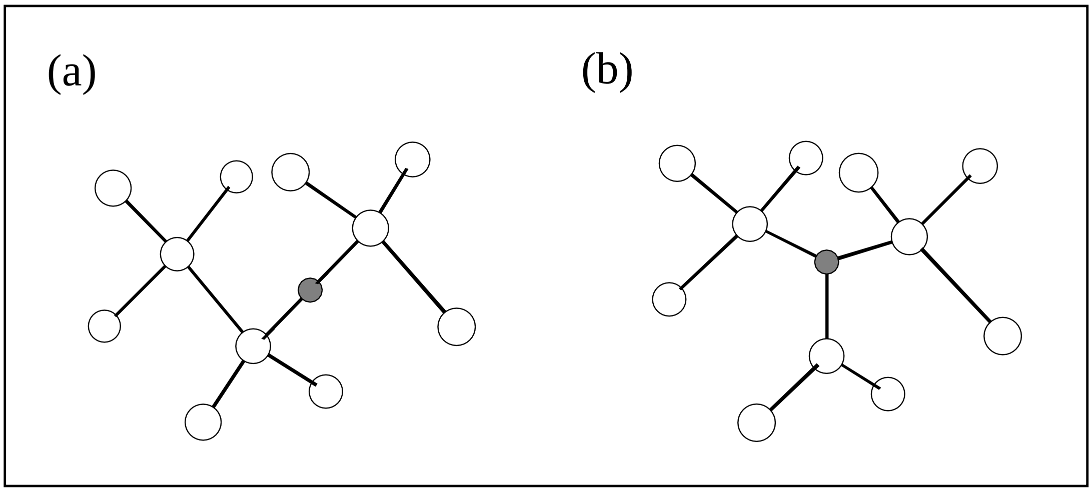


Figure 6.2: Possible models for the interstitial carbon defect in silicon.

material [80] correlated the EPR spectrum Si-L6 with an acceptor level at E_c-0.10 eV. The symmetry of the defect giving rise to L6 was C_{2v}, and the similarity of its annealing characteristics to G12 lead to an assignment of L6 to C_i^-. Preferential alignment experiments using the EPR spectrum found a reorientation energy of 0.77 eV, similar to that of C_i^+. The similarity between the re-orientation barriers of the defects in both charge states ($\approx$ 0.8 eV), and their piezospectroscopic tensors, indicated that they possessed the same structure. However, no hyperfine splitting of the EPR line due to C_i^- with ^{13}C or ^{29}Si was observed, possibly because the partially occupied acceptor state is extended over a number of Si atoms and has only a small overlap with carbon.

The neutral defect is associated with a 856 meV photoluminescence line [81, 87], having an annealing activation energy of 0.87 eV, similar to that of the charged defect. The lack of any perturbations of this line in magnetic field confirmed that it was from the neutral defect; also consistent with this was the higher irradiation dose used in these measurements, the effect of which is to drive E_F to mid-gap and imply a neutral defect. Uniaxial stress measurements revealed that the symmetry of the defect giving rise to the PL signal was also C_{2v} (monoclinic I) [81]. It is just possible that the neutral defect possesses trigonal symmetry and reorients to C_{2v} in a charged state when a carrier is thermally trapped before radiative recombination occurs. However, unambiguous confirmation that the symmetry of the neutral defect is indeed C_{2v} and not trigonal came from the response of the two LVMs of the neutral defect to uniaxial stress [82]. These experiments showed that the symmetries of the 920 and 931 cm^{-1} LVMs were A_1 and B_1 respectively. Thus it appears that the defect adopts the structure shown in figure 6.2b in all three charge states and consists of a C and Si atom sharing a single lattice site with a $\langle 100 \rangle$ orientation [79, 80, 82]. However, this model gives no natural explanation of the near degeneracy of the local vibrational modes of the defect, or their 2:1 ratio in absorption intensities. The migration energy of C_i^0 was measured by Tipping and Newman [88] using long range annealing experiments, giving a migration energy of 0.87 eV.

6.4.2 Previous theoretical calculations

Several theoretical calculations have been performed on the C_i defect, all finding the $\langle 100 \rangle$ structure to be the ground state configuration [83, 84, 85, 86, 46]. The calculated vibrational modes of the various defect configurations [84] provides further evidence for the C_{2v} configuration in figure 6.2a. For this model they were 1005 cm^{-1} and 853 cm^{-1} for the A_1 and the B_1 modes respectively, in fair agreement with experiment. None of the three possible C_{3v} symmetry configurations possessed two high frequency modes. The localisation of the pseudo-wavefunctions arising from previous calculations [83, 86] are also in agreement with the EPR data, with the highest occupied state of C_i^0 being a non-bonding p-orbital localised on C, lying below a more diffuse p-pike orbital centred on the silicon atom.

The energy barrier to reorientation of the defect among the equivalent $\langle 100 \rangle$ orientations has been calculated by Tersoff [85]. He found this to be 0.7 eV for a saddle point where the C atom lies at a bond centred (BC) site along $\langle 111 \rangle$. This is in good agreement with experimental reorientation and migration barriers around 0.8 eV [79, 80, 88].

More recent calculations on the migration/reorientation barriers of the defect have calculated these to be 0.51 eV via a saddle point with C_2 symmetry [86] and 0.77 eV via a H-site [46]. The migration/reorientation pathway considered by Capaz *et al* [86] involved a low energy, 0.51 eV, four-fold coordinated carbon atom at the saddle point. This path involves C moving through the lattice and at the same time reorienting and hence explains why the barriers to migration and reorientation are

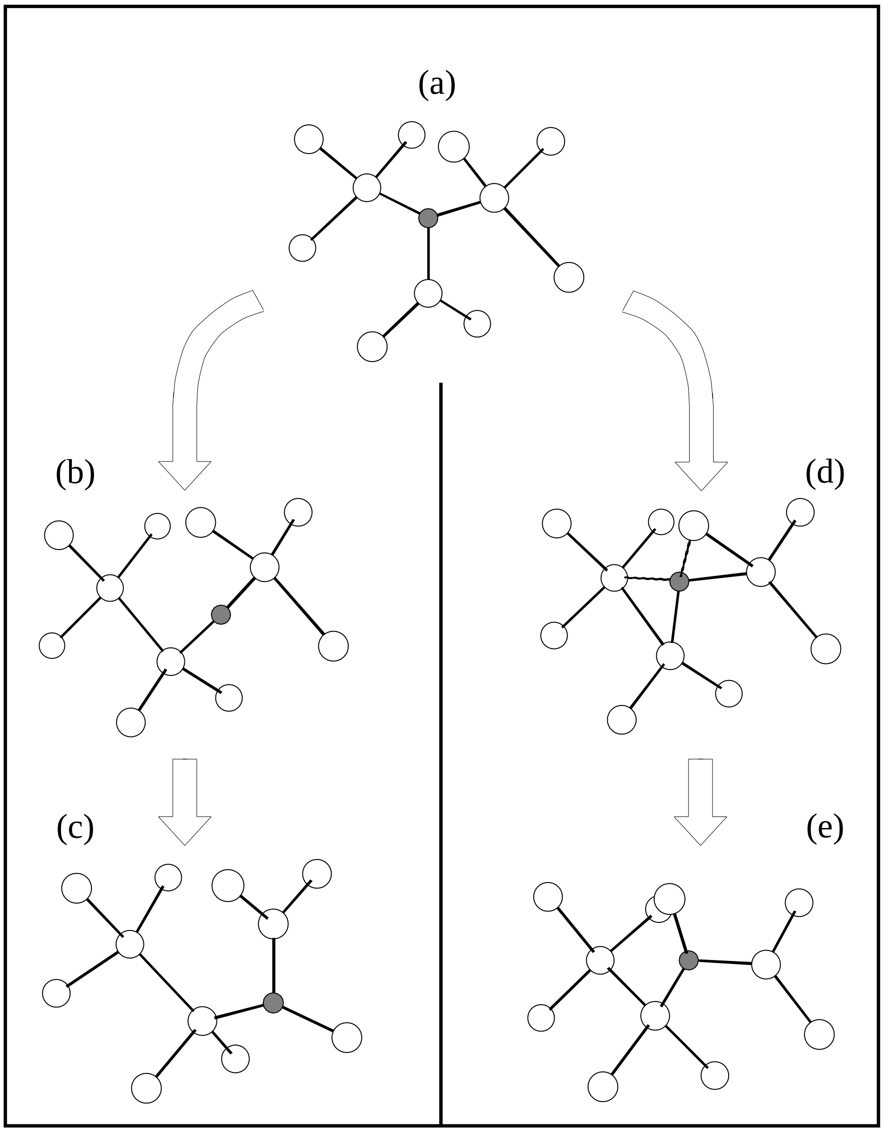


Figure 6.3: Migration pathways for the C_i defect, (b) is the BC saddle, considered in [85, 86], and (d) is the C_2 saddle found in [86].

the same. The path via a BC defect was also considered but they found it involve a 2.5 eV barrier which is considerably greater than that found by Tersoff. The structural data from this recent *ab initio* supercell calculation [86] for the neutral defect is given in table 6.1. The migration paths are illustrated in figure 6.3, where steps (b) and (c) represent the saddle and end point positions for the bond centred path, with (d) and (e) being the saddle and end points for the C_2 pathway.

6.4.3 Aimpro results

Previous theoretical modelling studies [83, 84, 85, 86] have found that the ground state configuration is the [100] C-Si split-interstitial, with C_{2v} symmetry, as illustrated in figure 6.2b. We therefore take this structure as the starting point in our calculations, and relaxed all atoms of the 72 atom cluster, C $Si_{35}H_{36}$, containing the defect in various charge states. The symmetry was constrained to C_{2v} for the 72 atom cluster runs. An 87 atom cluster containing C_i, $CSi_{44}H_{42}$, is used for the migration/reorientation barrier calculations.

Details of the structure found for C_i are given in Table 6.1, and compare favourably with a previous *ab initio* calculation [86]. For the neutral defect, the C atom made strong bonds with the Si atom, labelled Si_3, along the [100] axis and slightly weaker bonds with its neighbouring Si atoms, labelled Si_4, along [011], where the subscripts denote the atomic coordination. Nevertheless, the C atom pulls in the Si_4 neighbours along this direction leaving the Si_4-Si bonds along [011] slightly stretched at 2.352 Å. The Si_4-C-Si_4 and Si_4-C-Si_3 bond angles are 137° and 125° respectively.

Table 6.1: Structural data of the C_i and C_i-C_s defects in Å . Here, the subscripts indicate the coordination of the atoms involved, with reference to Figures 1-3.

	C_3-Si_3	C_3-Si_4	Si_3-Si_4	$\theta_{Si_4-C_3-Si_4}$	$\theta_{Si_4-Si_3-Si_4}$
C_i ref. [86]	1.77	1.80 (2)	2.22 (2)	157°	122°
C_i this work	1.728	1.820 (2)	2.300 (2)	137°	125°
	C_4-Si	C_4-Si_3	C_3-Si_3	C_3-Si	Si_3-Si
C_i-C_s A-form	2.007 (2) 2.075	1.846	1.708	1.855 (2)	2.352
	C-Si	C-Si_2	θ_{C-Si_2-C}		
C_i-C_s B-form	1.960 (2) 2.330 2.043 (2) 1.970	1.831 1.819	126°		

We find two high frequency modes at 922.2 (A_1) and 866.9 (B_1) cm^{-1}, which decrease by 27.6 and 27.8 cm^{-1} respectively with ^{13}C. These frequencies are in fair agreement with the experimental results (Table 6.2) of 921 and 930 cm^{-1} – although the ordering of the modes is given incorrectly. The calculated isotope shifts are in very good agreement with the experimental results, with a decrease of 28.8 and 25.8 cm^{-1} for the A_1 and the B_1 modes respectively. These results are similar to those of [84], where again, the ordering of the LVMs were given incorrectly. The

Table 6.2: Experimental and calculated LVMs, cm^{-1}, of the C_i defect, and their downwards isotope shifts

		A			B	
	^{12}C	^{13}C	^{14}C	^{12}C	^{13}C	^{14}C
Experiment	921.0	29	54	930.3	26	49
Theory	922.2	28	52	866.9	28	52

separation of the two vibrational modes is much greater than experiment and is very sensitive to the C-Si bond lengths [89]. If the length of the C-Si_4 bond is reduced by 2% then the modes become 926 cm^{-1} (A_1) and 950 cm^{-1} (B_1) and the ordering is now in agreement with experiment. In fact, the 1% difference in the experimental frequencies suggests that the C-Si bond lengths can differ by no more than 0.03% [13].

We then investigated the structure of the negatively charged defect. Adding an electron and relaxing all the atoms of the cluster, without any restriction, causes the three C-Si bond lengths to become almost equal: the C-Si_3 and C-Si_4 bonds becoming 1.788 and 1.805 (2) Å respectively. The Si_4-C-Si_4 bond angle decreased slightly to 136.0°, and the Si_4-Si bond length decreased slightly to 2.238 Å . The resulting vibrational modes have now crossed over with an ordering in agreement with experiment. The A_1 mode has decreased to 840.98 cm^{-1} with a corresponding increase of the B_1 mode to 901.78 cm^{-1}. So, once again the near degeneracy of these modes seems to arise from an equality in the three Si-C bonds. In this case, if the C bond angles were 120°, the C atom and its three Si neighbours would then make a molecule with a C_3 axis along $[01\bar{1}]$. This would then automatically give a two-fold degenerate E vibrational mode [89]. This symmetry would, however, also imply that the infra-red intensities of the two modes are equal which is contrary to experiment. The effective charges for the highest two local modes of the fully relaxed neutral defect were then calculated from the change in the cluster dipole moment with the atoms displaced according to the normal coordinates. The effective charges were found to be η_{A_1}=0.44$|e|$ and η_{B_1}=0.34$|e|$. The IR intensity ratio of the local modes is given by $(\eta_{A_1}/\eta_{B_1})^2$, and results in a ratio in the range 1.5-1.7:1, in reasonable agreement with the experimental ratio of 2:1 with the A_1 mode being the more intense in agreement with observation. For the negatively charge defect, we find the intensity ratio to be in the range 1.4-1.5:1. Thus although the symmetry of the core of the defect is almost trigonal and the two LVMs transform as E, the departure from symmetry is sufficient to lead to quite distinct induced dipole moments.

We now discuss the electronic structure of the defect. Although the energy levels are only given approximately by the theory, we use them along with the pseudo-wavefunctions for a qualitative analysis of the zero phonon transition. The Kohn Sham eigenvalues (figure 6.4a) indicated that there are two, possibly three defect related levels in the band gap. The highest occupied state has b_1 symmetry, and in agreement with EPR [79] it corresponds to a non-bonding $p-$ orbital on carbon figure 6.4c. This lies above an a_1 level whose wave-function, figure 6.4b, has little amplitude on C, and is mainly localised over the Si atoms neighbouring C atoms. The first unoccupied state shown in figure 6.4d has b_2 symmetry, and is diffusely localised over the three Si neighbours of C, predominantly on Si_3 but with little amplitude on carbon. Again, this is in agreement with the EPR data on the ionised donor and acceptor states [80]. The level above the b_2 level has a_1 symmetry, and is delocalised and probably part of the conduction band. The 856 meV zero phonon line transition might be either due to an electron in the upper a_1 state recombining with a hole trapped in the b_1 state associated with the non-bonding p-orbital on C, or, alternatively a electron in the upper b_2 state recombining with a hole in the lower a_1 state. Most workers have considered it to be the former process, involving the b_1 state localised on carbon [90]. This seems to us as unlikely because the PL has no fine structure, and the zero phonon line is unaffected by substitution of ^{13}C [87]. This implies that the bonding of carbon is unchanged as a result of the transition. If the C atom relaxed as a consequence of a hole occupying the C-related b_1 state - as indicated by the calculations - then an isotope shift would be expected. We therefore suggest that the observed PL involves the latter transition, which involves the diffusely localised states. This would imply that the dipole for the PL transition is at right angles to the $p-$orbital on the carbon atom.

A Slater transition calculation of the donor and acceptor levels of the C_i defect yielded values of E_v+0.29 and E_c-0.33 eV respectively, in fair agreement with the experimental results of E_v+0.28 and E_c-0.10 eV. These energies were calculated in a 298 atom cluster $CSi_{181}H_{116}$, and the calculated energies have been scaled by a factor (1.16/3.51), the numerator being the experimental value of the Si band gap, and the denominator is the calculated gap in a pure 297 atom Si cluster $Si_{181}H_{116}$ as in [17].

The energies for migration/reorientation of the defect were also calculated for the trajectories considered by Capaz *et al* using a relaxation procedure with con-

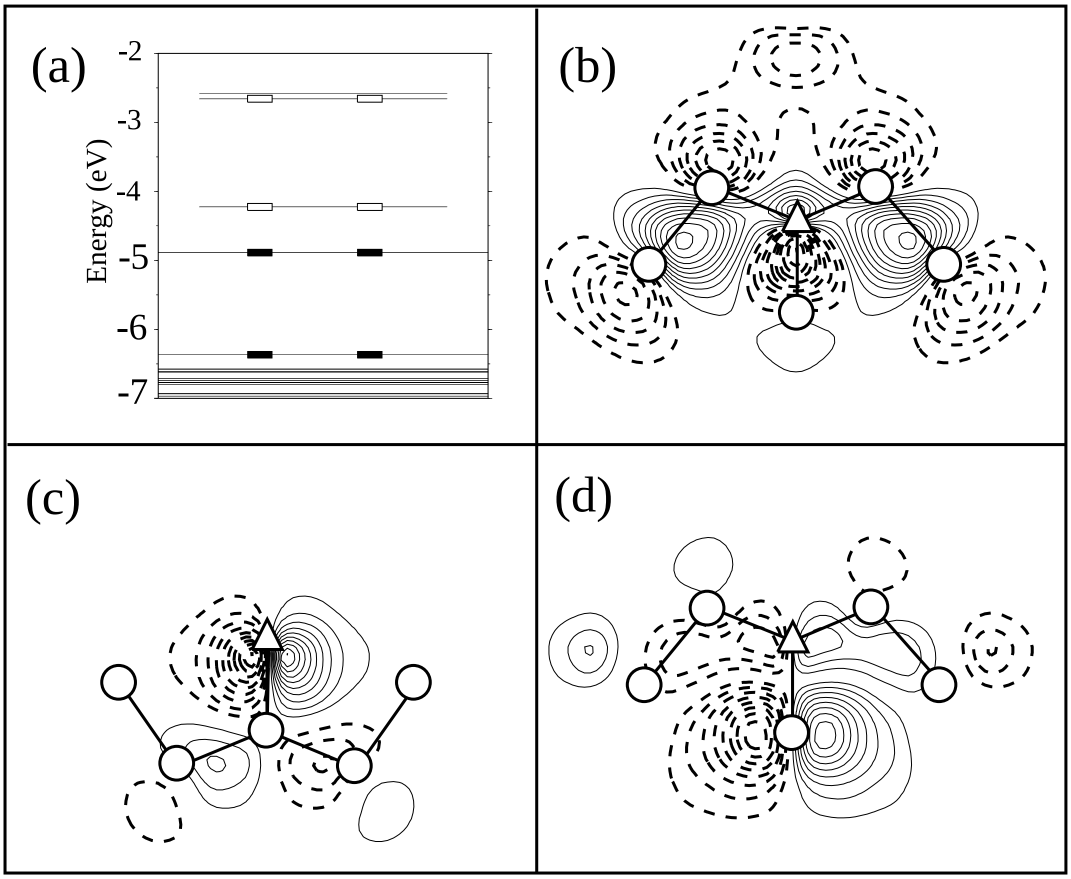


Figure 6.4: The Kohn Sham electronic levels and wavefunctions of C$_i$.

straints. In contradiction to this work, we find that both the C_2 and bond centred saddle points have almost equivalent barrier heights, with an energy of 1.10 eV. We therefore conclude that both are viable paths for the migration/rotation of C$_i$, figure 6.3. The 1.10 eV barrier is in good agreement with the experimental results of between 0.73-0.87 eV for both migration and re-orientation.

In summary, the calculations yield a structure which has approximately equal C-Si bonds and gives two almost degenerate modes around 920 cm^{-1}. The near degeneracy in modes being attributed to an approximate trigonal symmetry at the defect core. However, the deviation from this symmetry cause the two modes to have different absorption strengths. The migration energy of the defect is relatively low (1.1 eV) with the C atom being either two- (BC) or four–fold (C_2) coordinated at the saddle point.

6.5 The di-carbon defect, C_i-C_s

6.5.1 Experimental background

The di-carbon defect C_i-C_s is formed when Si:C, containing a low oxygen-content, is e-irradiated at room temperature creating mobile C_i defects which are subsequently trapped by C_s [34]. The defect is bi-stable taking the A-form in the charged states and the B-form in the neutral one; table 6.3 summarising the main experimental results is presented at the end of this section. The conversion to the metastable configuration can be achieved by the sudden injection of carriers with a light pulse, or by a switch in the bias potential. EPR and DLTS studies [27] showed that $(C_i$-$C_s)^{\pm}$ defects possessed C_{1h} symmetry, and they gave rise to acceptor and donor levels at E_c-0.17 and E_v+0.09 eV respectively. In fact, these are the two levels reported by Jellison [40], originally attributed to the O-V complex. The EPR data also found that the carbon atoms were inequivalent, and the g-tensors were perturbed by stress in a very similar way to those of the charged C_i defect, implying that the EPR spectra were from different charge states of the same defect. The model proposed from the experimental data is very similar to that of C_i, the symmetry being lowered by the presence of a second carbon atom at a substitutional site shown in figure 6.5a. This configuration of the C_i-C_s defect is known as the A-form, and is the stable configuration for both singly ionised states.

For high e-irradiation fluences, the Fermi level is around mid gap and the neutral B-form of the defect is formed [27, 40]. A zero phonon line at 969 meV had already been observed as early as 1970 [77], and gradually in following years it was linked to a defect containing two carbons [91], from annealing studies. The 969 meV line was unambiguously linked to carbon by examining its shifts with different carbon isotopes [92, 93], and was then linked to a defect containing two carbons [94, 90]. The dominant isotope effect on the zero phonon line with ^{13}C, was only governed by one of the two carbon atoms, and the effect was much greater with when different silicon isotopes where used [92, 93]. Hence the nature of this defect is consistent with two carbon atoms and one unique silicon atom. Again, the lack of any perturbation in magnetic field meant that the defect giving rise to the 969 meV transition must be in its neutral charge state.

Although this state was diamagnetic, an excited triplet state B^* has been observed by optically detected magnetic resonance (ODMR) [95]. This revealed that the symmetry of the B-form is C_{3v} for $T > 30$ K, and C_{1h} otherwise. Hyperfine interactions on ^{13}C detected by ODMR indicated that the carbon atoms are nearly equivalent. DLTS studies on the metastable $B^{\pm}$ form, which has trapped a carrier, show that has shallow donor and acceptor levels at E_v+0.07 and E_c-0.11 eV respectively. It has also been possible to observe EPR from the metastable B^- form, but not from B^+. B^- is also found to have C_{3v} symmetry for $T > 15$ K and again C_{1h} below this temperature. The ODMR results suggest that the two C atoms lie close to neighbouring substitutional sites, with a silicon interstitial, Si_i, close to a BC site lying between them. This configuration is illustrated in figure 6.5b. A detailed analysis into the structures and bistability of the neutral and singly charged di-carbon centre is presented in [27], and the configurational-coordinate diagram they constructed for the centre is shown in figure 6.6.

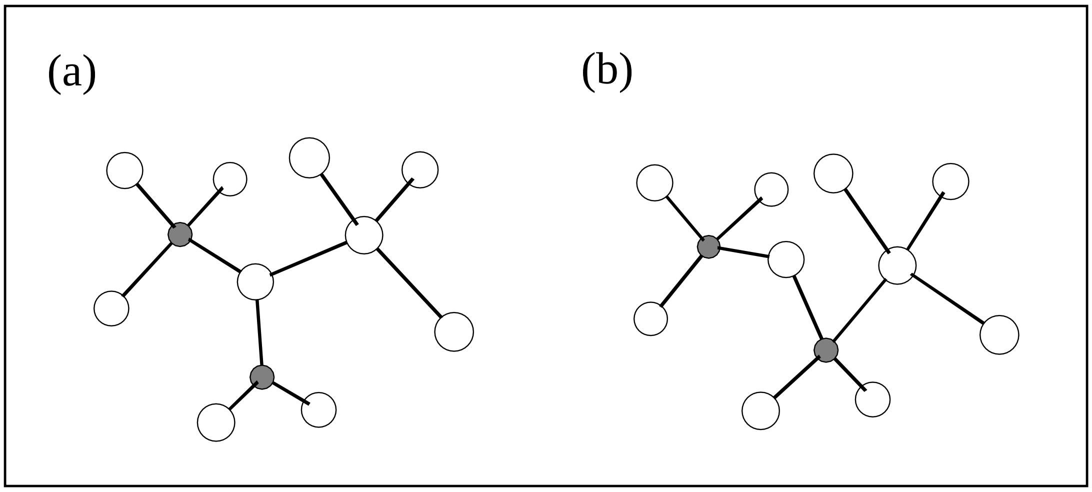


Figure 6.5: The proposed models for the di-carbon defect in silicon.

There are three peculiar features with this model. Firstly, if Si_i lay at the BC site, then the resulting D_{3d} symmetry would give a partially occupied e-level. This would probably be split through a Jahn-Teller distortion resulting in Si_i moving away from the BC site. However, the distortion cannot be too great if the reorientation energy around the C-C axis is to remain small. Nevertheless, the observed donor and acceptor levels are split by 0.97 eV, whereas a small distortion would lead to a difference in donor and acceptor levels of at most 0.5 eV being an estimate for the Hubbard U-correlation energy. On the other hand, if the distortion is not small, then the reorientation energy might be substantial and the symmetry of the defect would not be C_{3v}.

A second difficulty comes from the vibrational modes of the defect. The neutral B-form gives rise to a zero phonon luminescence line at 969 meV. The fine structure associated with this line reveals two local vibrational modes at 543.0 and 579.5 cm^{-1} (^{12}C) which shift by 10.1 and 14.9 cm^{-1} respectively for ^{13}C samples. However, in a sample containing a mixture of ^{12}C and ^{13}C, additional lines are observed which are shifted only by 0.1 and 0.6 cm^{-1}. Indeed, the original PL investigations failed

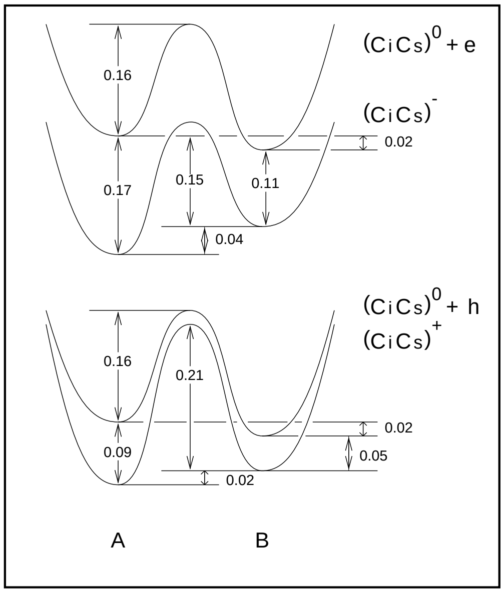


Figure 6.6: The configurational coordinate diagram for the di-carbon defect.

to find distinct modes in the mixed case, and incorrectly suggested that the defect only contained one C atom. These tiny shifts confirmed that two carbon atoms were present in the defect, but implied that they are almost dynamically decoupled [96, 93]. It has been suggested that the this decoupling arises because the two C-Si$_i$ bonds are almost orthogonal to each other [27]. Even if this was the case, it would require that the C-Si$_i$ stretch modes lie around 550 cm^{-1}. This is unlikely since the reduced coordination of Si$_i$ will result in a shorter C-Si$_i$ bond, and hence the C-Si$_i$ stretch should be larger than the highest mode of C$_s$ which occurs at 607 cm^{-1} [97].

A third peculiarity is that the positive and negative charge states of the B-form are metastable, with energy differences with the A form of only a few hundredths of an eV in all three charge states. Similarly, the barriers between the forms are only 0.1-0.2 eV. As the bonding in the two forms is quite different and bond energies are of the order of 1 eV, it is surprising that these energy difference are so small.

In recent work [98] the spin-1 EPR spectrum labled PT1 due to the di-carbon centre has also been observed, and the spectrum is closely related to those obtained in ODMR. At low temperatures, the spectra has C_{1h} symmetry, and this converts to a trigonal symmetry above 12K. The activation energy for the rotation about equivalent C_{1h} sites was found to be 0.0086 eV. Between 30-50K, a new EPR spectra WL1 was detected, arising from a defect with C_{1h} symmetry. The authors suggested that the WL1 spectra arises from an excited triplet state of the C$_i$C$_s$ defect in another configuration.

Table 6.3: Experimental results for the three charge states of the di-carbon defect.

Defect	Method	Ground state	Metastable state
CC^+		A	B
	DLTS	E_v+0.09 eV	E_v+0.05 eV (est.)
	EPR	G11 C_{1h}	
CC^-		A	B
	DLTS	E_c-0.17 eV	E_c-0.11 eV
	EPR	G17 C_{1h}	L7 (C_{3v} T>15K) C_{1h}
CC^0		B	A
	PL	969 meV	
	ODMR	(C_{3v} T>30K) C_{1h}	

6.5.2 Previous theoretical results

Theoretical modelling of the di-carbon defect is still in its early stages, and to date three groups have examined the centre [84, 46, 89]. The first two focused on the relative energies of the A and B forms, and in our previous work, we analysed local modes of the B form only. The findings of the previous calculations relating to the defect energies are summarised in Table 6.5. In addition to the two accepted configurations, other models in which a C-C bond was formed were considered, and it was found that these were higher in energy than those where the carbons were separated by at least one Si neighbour.

6.5.3 Aimpro results

All atoms of an 87 atom cluster $C_2Si_{43}H_{42}$ containing two substitutional C atoms between which a Si_i is sited, were allowed to relax. This corresponds to the B-form of the di-carbon defect shown in Fig. 3. The Si_i moved away from the bond centred site, with the carbon atoms moving slightly from their substitutional sites until the equilibrium structure was obtained. The bonding in the equilibrium structure is given in Table 6.1. The C-Si_i-C angle is 126° showing that the Si_i has moved well-away from a bond centred site. This leads to a great splitting in the two gap Kohn-Sham levels and is consistent with the large donor-acceptor level difference.

The local vibrational modes of both configurations of the defect are given in Table 6.4. It is clear that the B-form must give rise to six local modes - two of which are A modes due to asymmetric and symmetric stretch of the C-Si_i. These modes must be of high frequency, simply because the reduced coordination of Si_i must result in stronger C-Si_i bonds. The remaining four modes involve predominantly the motion of the C atoms together with their Si neighbours. If the Si_i was at the bond centred site, these would correspond to two degenerate E modes in a D_{3d} defect, but are split into four with the lower C_{1h} symmetry resulting from the displacement of Si_i.

Table 6.4: Calculated and experimental LVMs, cm^{-1}, and their downwards isotope shifts of A- and B-forms of the C_i-C_s defect.

Sym.	^{12}C-^{12}C	^{12}C-^{13}C	^{13}C-^{12}C	^{13}C-^{13}C
Calc. A				
A	987.6	986.8 (0.8)	961.2 (26.4)	960.1 (27.5)
B	776.9	776.7 (0.2)	752.4 (24.5)	752.2 (24.7)
A	739.2	717.9 (21.3)	737.0 (2.2)	716.0 (23.2)
A	600.0	584.0 (16.0)	599.8 (0.2)	583.8 (16.2)
B	598.7	582.7 (16.0)	598.6 (0.1)	582.6 (16.1)
A	531.3	531.3 (0.0)	531.3 (0.0)	531.3 (0.0)
Calc. B				
A	837.9	823.8 (14.1)	831.6 (6.3)	815.8 (22.1)
A	714.8	704.1 (10.7)	701.0 (13.8)	691.7 (23.1)
B	648.7	648.7 (0.0)	630.3 (18.4)	630.3 (18.4)
A	582.3	570.1 (12.2)	580.0 (2.3)	568.2 (14.1)
B	551.8	543.2 (8.6)	551.7 (0.1)	543.0 (8.8)
A	543.2	543.0 (0.2)	538.5 (4.7)	538.5 (4.7)
A	532.2	531.9 (0.3)	531.8 (0.4)	531.7 (0.5)
Obs. B [90]				
A	579.5	0.6	-	14.9
A	543.0	0.1	-	10.1

The local vibrational modes frequency of the B-form is now discussed in detail: we find six modes above the Raman peak, ranging from 543 to 838 cm^{-1}. The C-Si_i stretch modes occur at 838 and 715 cm^{-1} and these shift by 22 and 23 cm^{-1} with pure ^{13}C, but pronounced shifts occur in the mixed isotope case. The shifts of these modes in the mixed case reflect the fact that the two carbon atoms are almost equivalent in this configuration. The magnitude of these modes, and the fact that they exhibit mixed splitting clearly rules them out as being responsible for the modes observed by PL.

The next four modes involve mainly the C atoms and their back bonded neighbours, with little motion of Si_i. The 649 and 552 cm^{-1} modes are decoupled with one carbon moving along $[0\bar{1}1]$ in each mode. However, these modes are of B symmetry, and therefore would not be detected by PL. The final two local modes are of A symmetry, occurring at 582 and 543 cm^{-1}, and represent vibrations in which each C atom moves independently against its Si neighbours, but not including Si_i. These modes shift by 14 and 5 cm^{-1} with pure ^{13}C, but the additional modes in the mixed isotopic case are all within 2 cm^{-1} of the pure modes. Therefore these modes are essentially decoupled, and are in excellent agreement with the those found by PL. This gives two A modes at 579 and 543 cm^{-1} with shifts 15 and 10 cm^{-1} for pure ^{13}C. No further vibrational modes are observed experimentally, which raises the question as to why the highest two calculated A modes are not detected. Now, PL does not necessarily detect all A modes of a defect, for example in the C$_i$ defect, discussed above, the A mode at 921 cm^{-1} is not detected as a replica of the 856 meV zero phonon line. However, IR studies on the di-carbon defect have also failed to find any local modes. A possible explanation as to why these higher modes are not detected is that they have very short lifetimes. The energy of these modes $\approx$ 0.1 eV, is comparable to the activation barrier for the conversion from the B-form to the A-form (0.16 eV). Since the configuration energy curves must have zero gradient at the barrier height, it suggests that the curves cannot be described by a harmonic potential much beyond 0.1 eV. We therefore suggest that anharmonic effects are considerable, leading to particularly large three phonon scattering processes which have very short lifetimes.

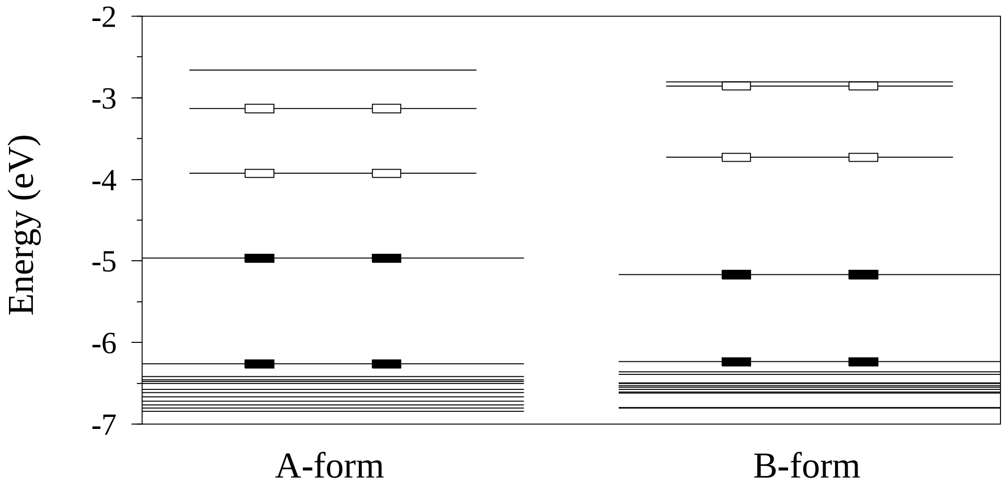


Figure 6.7: The Kohn Sham electronic levels for the two configurations of the di-carbon defect.

Investigations were also made into the A-form of the defect, shown in Fig. 6.5, the equilibrium bond lengths are given in Table 6.1. The energy differences between the two forms in the neutral charge state was found to be 0.35 eV, the A-form being lower, – a figure very much greater than the experimental difference of about 0.04 eV. It seems that accurate energy differences require a very much larger cluster as the strain fields of the two defects are very different. Although there is an offset of around 0.3 eV between our results and experiment, it is interesting to investigate the trends in energy difference between the two configurations with charge state. The

A-form was found to be lower in energy by 0.43 and 0.50 eV for both positive and negative charge states respectively. These results are in good qualitative agreement with experiment, showing that when charged, the relative stability of the A-form over the B-form increases by 0.08 and 0.15 eV, compared to the experimental values of 0.04 and 0.06 eV for the positive and negative charge states respectively. These results are given in Table 6.5, along with the experimental values [27] and the previous theoretical calculations [84, 46].

Table 6.5: Relative total energies (eV) for the A and B forms of the di-carbon defect with varying charge state.

	A^+	B^+	A^0	B^0	A^-	B^-
Experiment [27]						
	0.000	0.020	0.020	0.000	0.000	0.040
Previous calculation [84]						
MINDO/3-Si_{38}	0.000	0.621	0.337	0.000	—	—
MINDO/3-Si_{41}	0.000	0.383	0.508	0.000	—	—
MNDO-Si_{38}	NF	0.000	0.000	1.151	—	—
MNDO-Si_{41}	0.000	1.086	3.501	5.127	—	—
PM3-Si_{38}	0.596	0.000	1.903	0.000	—	—
PM3-Si_{41}	0.737	0.000	1.816	0.000	—	—
Previous calculation [46]						
	—	—	0.800	0.000	—	—
Present results						
	0.000	0.428	0.000	0.346	0.000	0.500

Other configurations were, considered, of these the $\langle 100\rangle$ oriented defect with a C-C pair occupying a single lattice site proved to be particularly low in energy. This configuration lying between -0.26, 0.31, and -0.33 eV in energy relative to the A-form, in 87, 132, and 177 atom clusters respectively. The relative energies are oscillating with cluster size, suggesting that the C-C pair may be stable, however there is no evidence for this experimentally. Finally, we estimated the binding energy of C_i to C_s in a 180 atom cluster (neutral charge state). The formation of the defect resulted in a lowering of the total energy by 0.97 eV.

6.5.4 Further experimental results

In a recent paper [99], a number of photoluminescence lines, arising from centres containing carbon, and in some cases oxygen are detailed. Among these, the G-line has been re-investigated and the authors claim to have identified a new local mode of this defect at 731.1 cm^{-1}, which *does not* display any mixed isotope splittings, and shifts by 21.7 cm^{-1} with ^{13}C. Clearly if this assignment is correct, there a problem with the B-form structure, as it does not give rise to an uncoupled local mode around 730 cm^{-1} (table 6.4). This is the case for the A-form, but there is no mode around 540 cm^{-1} which has a large carbon isotope shift. The authors also considered the C-line 789.6 meV (C_iO_i) and report some local modes for this centre. On annealing for 30 minutes at 350°C, the F-line at 949.9 meV is formed, and this has several modes similar to those of both the C-line and the G-line. It was therefore concluded that the F-line defect is a C_3O defect formed when the *mobile* G-line defect is captured by the C-line defect. This is an entirely different view to the conventional one of the G-centre breaking up around 300°C. Upon further annealing this time at 450°C, the H-line is created, and it is suggested that this centre is formed when the F-line defect captures another oxygen atom.

Some years earlier, Kwok [100] considered the carbon and oxygen related point defects formed when the G-line is annealed for a few minutes at 300°C. Here, sharp lines at 949.9, 945.6, 935.1, 932.8, 924.1, 919.8 and 902.8 meV were observed. The first of these is coincident with the F-line, and the fourth with the T-line. A number of phonon side-bands were also observed, in particular, one at 878.7 meV, exactly the energy of the replica assigned in reference [99] as the new mode of the G-line defect. Kwok's interpretation however was that this feature was a local mode of the 949.9 meV line and not the G-centre. Another confusing result was the ^{13}C isotope shift of the 878.7 meV side band which was found to be 10 cm^{-1}, half the value found in [99]. In fact Safonov [101] also finds many modes related to the F-line defect, and in particular one at exactly 878.7 meV, which shifts by 8 cm^{-1} with ^{13}C very close to the result of Kwok. Given the coincidence in energy between the alleged local mode of both centres, and the vast differences in annealing treatment, it seems hard to reconcile both sets of results, and further work is necessary.

6.6 Conclusions

The calculations presented here have successfully explained some of the remaining difficulties with the structural models of the carbon and di-carbon defects. Our calculated LVMs for the C_i defect are in reasonable agreement with experiment, and the near equality of the three C-Si lengths results in the two almost degenerate local vibrational modes. The calculated infra-red intensity ratio of the two LVMs gave a ratio of 1.5-1.7:1 in fair agreement with the experimental result of 2:1. The calculated energies for the migration/reorientation barriers lead to an upper bound of 1.10 eV, in good agreement with experiment.

The B-form of the neutral C_i-C_s defect gave rise to six local modes, only two of which have been detected experimentally by PL. Two of these have very high frequency, and probably could not be detected because of large anharmonicities which would lead to a very broadened absorption. Two of these modes having A_1 symmetries lay around 560 cm^{-1}. These involved almost independent movements of the C atoms with there neighbouring Si atoms, excluding the interstitial. The isotope shifts of the vibrational modes are in excellent agreement with experiment. Investigations into the relative energies of the configurations of the di-carbon defect with charge state confirms that the stability of the A-form over the B-form increases with charging. The binding energy of the di-carbon centre was calculated to be around 1 eV.

Recent results have seemingly cast doubts over the configuration of the defect giving rise to the 969.4 meV zero-phonon transition. However, as described above, many spectral features are observed, and the evidence is so far not conclusive. Clearly further work is necessary before this potential problem is cleared up.

Chapter 7

The interaction of H with C in silicon

7.1 Introduction

In this chapter, the interaction between both substitutional, C_s, and interstitial, C_i, carbon defects in silicon with hydrogen are considered. We find the binding energy of neutral H with C_s is 1.01 eV and the defect to be bistable. In the positive and neutral charge states H lies near the centre of a C-Si bond and is anti-bonded to C in the negative charge state. A second H atom can be trapped in a H_2^* like defect. H forms stronger bonds with interstitial C_i. In the C_i-H defect, the binding energy of H is 2.8 eV and two low energy structures have almost degenerate energies. These consisted of a bond centred Si-(C_i-H)-Si defect and a $\langle 100 \rangle$ oriented C_i-Si split-interstitial with H bonded to C_i. The calculated barrier for conversion between the two stable structures is very low ~ 0.3 eV implying that the defect migrates rapidly, and readily reacts with other defects or impurities present. Two possible reactions are considered: the first is with another H and the second with C_s. The defect is completely passivated in the former while the stable form of the latter consists of a $\langle 100 \rangle$ C-C di-carbon interstitial where one radical is passivated by H. The calculated symmetry and the local vibrational modes are in excellent agreement with those experimentally observed for the T-photoluminescent centre. Finally, a further reaction involving the T-centre and a second H atom is considered and is found to lead to the elimination of electrical activity.

7.2 Calculational details

A variety of different clusters were used in the calculations presented in this chapter. For most calculations C and H atoms were added to an 86 atom trigonal cluster, $Si_{44}H_{42}$, representing the perfect material. A central Si atom was replaced by C and another H atom added to model the C_s-H defect. In a similar way additional atoms were added to investigate interstitial carbon-hydrogen defects. To explore dissociation and reorientation energies, C and H atoms were added to a larger 131 atom cluster $Si_{71}H_{60}$ cluster representing perfect material. In addition to these, for a select few runs, a 296 atom cluster $Si_{180}H_{116}$ was used, with the appropriate defects inserted. The electronic wavefunctions were expanded using a basis of s- and p- Gaussian orbitals centred on the nuclei, and at the centre of selected bonds. Eight Gaussian orbitals of different exponents were used on the C atom(s), and the innermost Si atoms. A fixed linear combination of eight Gaussian orbitals were used on all other Si atoms. Three Gaussian orbitals with different exponents were sited on the defect related H atoms, and a linear combination of these Gaussian orbitals were used for the terminating H atoms. Additionally, three s- and p-Gaussian orbitals with different exponents were placed at the centre of every bond in and around the defect core. The charge density was expanded in a basis of s-Gaussian functions sited at the same sites as above with the same number of different exponents. For calculations of the local vibrational modes of the defects, the second derivative of the energy with respect to position were calculated for the carbon and defect hydrogen atom(s), and all nearest neighbour silicon atoms.

The procedure for determining the migration barrier for the defects investigated here is as follows: The saddle point for the possible migration paths are not clear for the defects investigated in this work, so to estimate it the orthogonal relaxation procedure was taken (see chapter 3.3.1), after the vector displacements between the initial and final configurations $\mathbf{v}_i = \mathbf{R}_i^1 - \mathbf{R}_i^2$, and $\mathbf{R}_i^1$ and $\mathbf{R}_i^2$ were found. The result of this is a profile of the energy as a function of α for the restructuring of the defect, and the barrier for migration/reorientation corresponds to the maximum value of energy obtained, at the corresponding saddle configuration.

7.3 The substitutional carbon-hydrogen defects

7.3.1 Experimental background

Recent DLTS experiments [102, 103, 104, 105] have suggested that hydrogen can be trapped by C_s forming the E3 defect with an 0/+ level at E_c-0.15 eV. The presence of both C and H in the defect was inferred from the variation in the trap concentration with [C] and [H]. It is believed that the defect contains only one C and one H atom and does not contain oxygen although there is no direct evidence for this constitution. The stability appears to depend strongly on charge state. [C-H]$^+$ anneals out at 80°C with a rate $\nu e^{-E/kT}$ where the activation energy, E, is 0.7 eV and $\nu \sim 10^7$ s^{-1}. This value of ν suggests that the dissociation of the defect is controlled by the capture of electrons resulting in a less-stable neutral defect. This is supported by observations [105] that light enhances the dissociation but only outside a depletion layer where photo-generated electrons can be trapped. These experiments show that the neutral defect anneals out at 20°C. In p-Si, a hole trap (H1), stable to 100°C, has been correlated with C and H and this dissociates with E and ν being 1.7 eV and 10^{19}s^{-1}. This last figure seems far too large and suggests that dissociation is not governed by an Arrhenius behaviour. The -/0 level is placed at E_v+0.33 eV.

The similarities in the behaviour and their spatial distribution led Kamiura *et al.* [105] to suggest that E3 and H1 arise from a bond centred and anti-bonding configuration of a C_s-H defect respectively, implying that the defect is bistable.

Uniaxial stress studies performed on the E3 centre, revealed that the defect symmetry was indeed trigonal, consistent with the proposed model of the bond-centred C-H complex [106]. The results suggested that the defect becomes unstable in the neutral charge state, after capture of an electron from the conduction band, and a configurational coordinate diagram was constructed to this effect. Here, there minima in the energy surface for the formed and dissociated $(C_s\text{-H})^0$ defect were degenerate, suggesting that the H was no longer bound.

7.3.2 Previous theoretical modelling

Before detailing the theoretical modelling of the defect to date, the possible C_s-H configurations are described. The four most obvious configurations are (a) H lying near a bond centre between C and a Si neighbour but bonded to C. We shall refer to this configuration as (C_sH_{BC})Si, (b) H anti-bonded to C, $(H_{AB}C_s)$Si, (c) H bond centred between C and Si but bonded to Si, $C_s(H_{BC}Si)$, and finally (d) H anti-bonded to Si, $C_s(SiH_{AB})$. These are illustrated in Figure 7.1 (a-d). Two further configurations are considered which are related to (b) and (d) but the H atom lies at a T_d site and the C-Si bond is reformed. Early theoretical modelling of the defect [107], using Hartree-Fock theory, concluded that there two almost degenerate neutral ground state configurations. These correspond to structures $C_s(H_{BC}Si)$ and $(H_{AB}C_s)$Si with the later being slightly higher in energy (0.05 eV). The binding energy of H to C in this defect was found to be 1.42 eV, and both of these configurations gave rise to defect related states in the band gap. The $C_s(SiH_{AB})$ configuration was found to be 0.94 eV above the ground state with the tetrahedral sites being even higher in energy.

More recent calculations on the neutral defect [108] using LDF theory found (C_sH_{BC})Si to be the ground state energy configuration, in contradiction with the previous calculation. The discrepancy between the two calculations was resolved

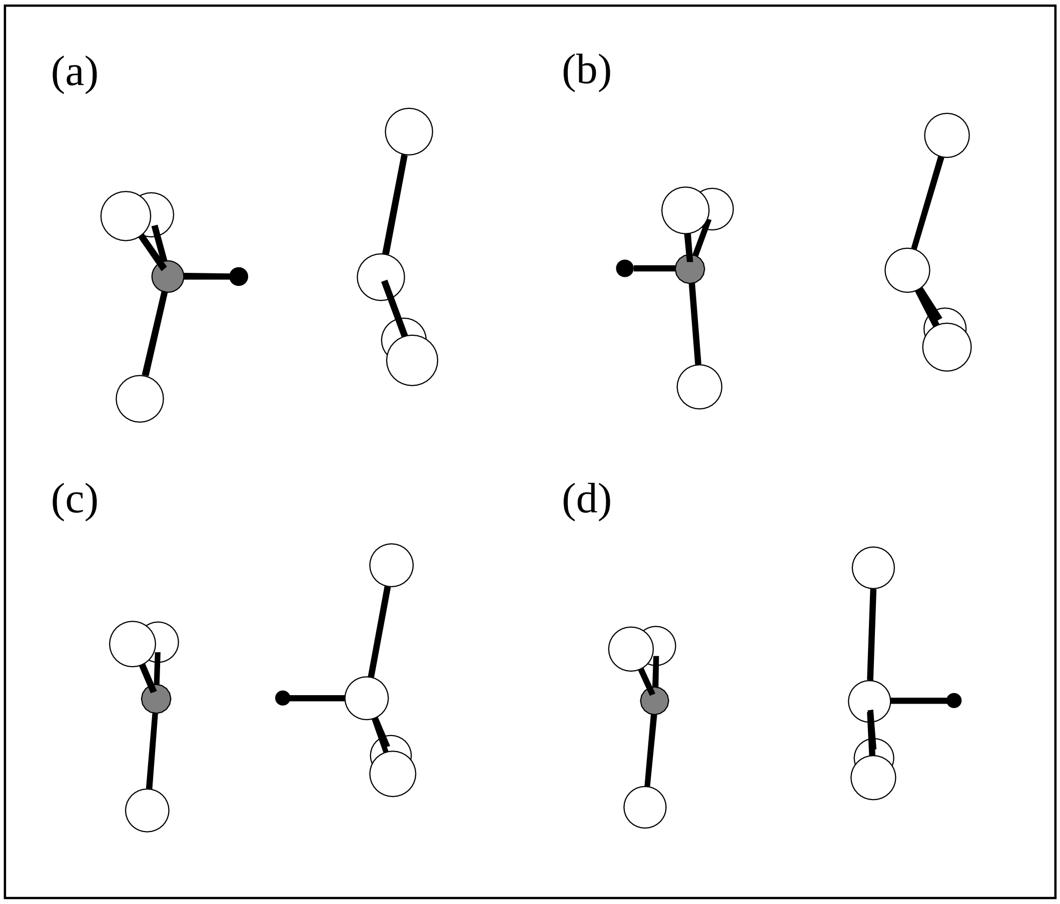


Figure 7.1: The possible C_s-H defects in Si: (a) the (C_sH_{BC})Si configuration, (b) the $(H_{AB}C_s)$Si configuration, (c) the $C_s(H_{BC}Si)$ configuration, and (d) the $C_s(SiH_{AB})$ configuration.

after further work [109] using standard and post-Hartree Fock techniques on molecular clusters, as well as (DFT) calculations in clusters and supercells. The two bond centred structures (C_sH_{BC})Si and $C_s(H_{BC}Si)$ were found to be the lowest energy configurations with LDF calculations predicting (C_sH_{BC})Si as the lowest energy configuration in agreement with [108]. When correlation corrections to Hartree Fock were included, the lowest energy configuration switched from $C_s(H_{BC}Si)$ to (C_sH_{BC})Si, but the energy difference between these two forms was found to be only a few meV.

In the light of the experimental data which suggests that the ground state structure changes with charge state, LDF calculations have been performed on the C_s-H defect in the three charge states utilising 9 atom supercells [110]. The stable structure was (C_sH_{BC})Si in all charge states with a H-binding energy of 1.20 eV for the neutral one. In the negative charge state, however, the dissociated configuration where H lies at a tetrahedral T_d site to one of the Si neighbours of C has an energy within 0.06 eV of the ground state. The closeness of these energies suggests that the negatively charged defect is barely stable. The calculations for the relative energies of the metastable neutral structures differed from the previous results. The lowest energy metastable configuration was $C_s(SiH_{AB})$ lying 0.57 eV above the ground state. Almost degenerate in energy with this structure was $(H_{AB}C_s)Si$ which was 0.61 eV above the ground state.

These results therefore indicated that trapping of one or more electrons by $(C_s$-$H)^+$ could lead to a rapid dissociation and such a change would explain the loss of the E_c-0.15 eV state after illumination. For clarity, the main results are summarised, along with the results of our calculations described below in Table 7.1.

7.3.3 Aimpro results

The C and H atoms were introduced into the centre of the 86 atom cluster with atomic positions and bond lengths chosen to form all four possible structures outlined in section Ia (Figure 7.1 a-d). All atoms were then allowed to relax with the symmetry constrained to C_{3v}. Table 7.1 shows the energies of the various configurations. In agreement with all recent theoretical calculations, $(C_sH_{BC})Si$ was found

Table 7.1: Relative energies (eV) for the C_s-H defect in Si with varying charge state. NF indicates that the structure was not found as a stable or metastable configuration.

	Previous work [108]			Present work		
Structure	+	0	-	+	0	-
$(C_sH_{BC})Si$	0.00	0.00	0.00	0.00	0.00	0.04
$(H_{AB}C_s)Si$	0.70	0.60	0.30	0.49	0.33	0.00
$C_s(H_{BC}Si)$	NF	NF	NF	NF	NF	NF
$C_s(SiH_{AB})$	0.71	0.57	0.06	1.97	1.12	0.79

to be the most stable structure in the neutral charge state. Our trends in energies against charge state are in agreement with those of ref. [110], although we disagree on the metastable structures. These trends are illustrated in Figure 7.2, and given along with the results of the previous calculation in Table 7.1. The energy differences between the stable and metastable configurations decrease as the charge state changes from positive to negative in both the present results, and the previous calculation. The defect is bistable with H lying at a BC site but bonded with C, the $(C_sH_{BC})Si$ configuration (Fig. 7.1a), being the most stable one in the neutral and positive charge states. In the negative charge state however, the configuration where H lies at an AB site while still bonded with C, ie. $(H_{AB}C)Si$, in Fig. 7.1b is the most stable.

For all charge states, $C_s(H_{BC}Si)$ is unstable and H moves spontaneously along

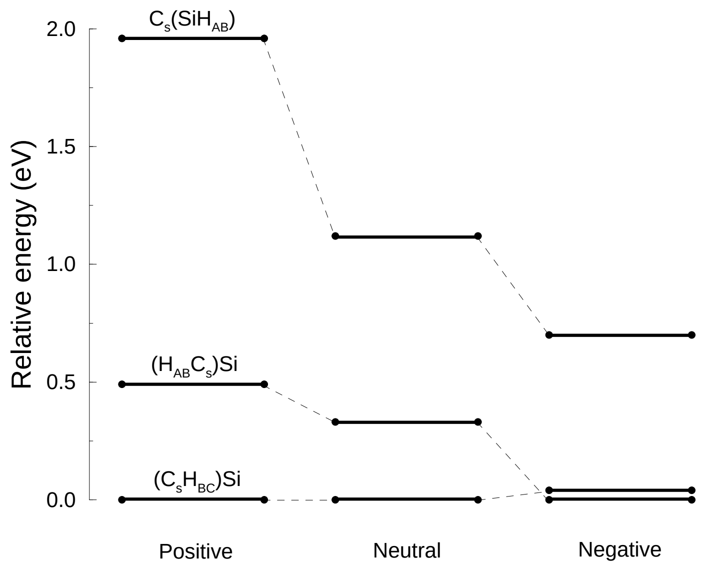


Figure 7.2: Trends in relative energies of the various configurations with charge state.

⟨111⟩ bonding with C to form the (C_sH_{BC})Si structure (Fig 7.1a). This agrees with the results of Kaneta et al [108] but is in disagreement with the results of Maric *et al* [107, 109] who found $C_s(H_{BC}Si)$ to be metastable. Kaneta *et al* [108] found that the $C_s(SiH_{AB})$ in the negative charge state to be almost degenerate with (C_sH_{BC})Si but we found a difference of energies of 0.75 eV.

The optimised bond lengths for the three stable structures and their variation in charge state are given in Table 7.2. These results show that H moves almost to the the T_d site in the negative charge state of $C_s(SiH_{AB})$. This was the only tetrahedral configuration found. Our result is in disagreement with the previous study [110] where it was found that the tetrahedral defect formed in the negative charge state, similar to isolated hydrogen. We now consider the energy levels of

Table 7.2: Bond lengths of the three stable configurations of C_s-H in Si, in Å.

	(C_sH_{BC})Si			$(H_{AB}C_s)$Si			$C_s(SiH_{AB})$		
Charge	+	0	-	+	0	-	+	0	-
C-H	1.09	1.10	1.10	1.08	1.08	1.09	3.72	5.23	4.95
Si-H	2.21	2.19	2.11	4.83	4.66	4.44	1.61	1.51	1.52
C-Si	3.31	3.29	3.21	3.74	3.57	3.35	2.11	3.73	3.43
C-Si (3)	1.99	1.97	1.94	1.99	1.98	1.96	1.99	1.89	1.86

the different structures in the neutral charge state. All three stable configurations gave rise to a deep level which is singly occupied as shown in Figure 7.3, here the Kohn-Sham eigenvalues are scaled to the band gap of silicon. The theory is not able to predict accurately the donor/acceptor levels but it is clear from the Kohn Sham eigenvalues that in all cases when H complexes with C_s an electrically active defect remains. These results are consistent with the tentative assignment of E3 and H1 to the (C_sH_{BC})Si and $(H_{AB}C_s)$Si structures respectively [105].

The binding energy of neutral H with C_s is found to be 1.01 eV by comparing the energies of the complex with its dissociated constituents using a 132 atom cluster. In the dissociated form a neutral H lies at a BC site remote from C_s. This value for the binding energy is similar to those found in previous calculations. This binding

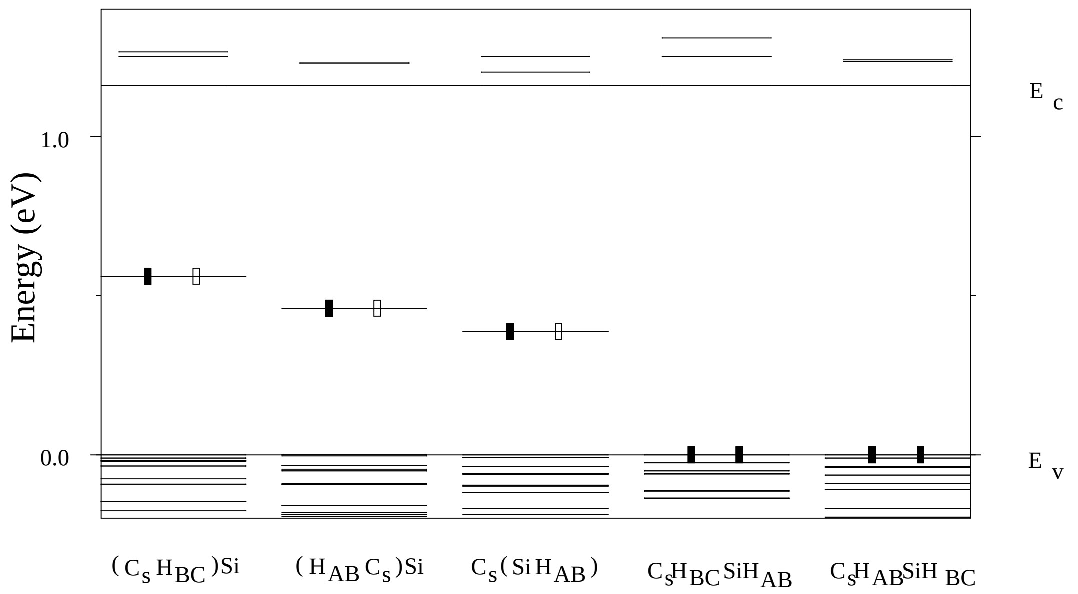


Figure 7.3: The Kohn-Sham levels of the substitutional CH defects.

energy is inconsistent with the configurational coordinate diagram constructed in [106] where it was suggested that the energies of the nearest and second nearest neighbour sites were degenerate in the neutral charge state, leading to the instability of the complex. However, there it was assumed that the neutral defect simply dissociated. The resolution of the problem associated with the binding energy can be resolved if further reaction paths are considered. One possibility, following the arguments of Endros [102] is that the breakup occurs in the neutral charge state due to the formation of H_2 molecules in the material, the energy of the C_s and a nearby H_2 being lower than the C_s-H defect close to H_i. Another possibility is that the neutral defect could be negative-U, and the capture of two electrons leads to a structural change - consistent with our calculated trends in energy. In fact, referring to figure 7.2, and noting that the $C_s(SiH_{AB})$ structure is close to the dissociated one, then our results show that the binding energy of H decreases with electron capture. There is also the enhanced possibility of the formation of the H_2^* like C_s-H_2 defect in the neutral charge state. Therefore, another energy site does exist for the dissociated neutral defect, as was shown in the configurational coordinate diagram, but this is a result of the interaction between the defect and a second isolated hydrogen atom, or because of negative-U properties. The calculated barrier for the dissociation of the neutral defect in the absence of other impurities was 1.68 eV.

The predicted concentrations of XY centres is now discussed for the following reaction, under equilibrium conditions.

$$\mathrm{X} + \mathrm{Y} \longleftrightarrow \mathrm{XY}, \tag{7.1}$$

and the concentration of each component is given by:

$$n_{\mathrm{X}} = N\exp\left[\frac{-(E_{\mathrm{X}} - \mu_{\mathrm{X}})}{kT}\right], \tag{7.2}$$

$$n_{\mathrm{Y}} = N\exp\left[\frac{-(E_{\mathrm{Y}} - \mu_{\mathrm{Y}})}{kT}\right], \tag{7.3}$$

$$n_{\mathrm{XY}} = N\exp\left[\frac{-(E_{\mathrm{XY}} - \mu_{\mathrm{XY}})}{kT}\right], \tag{7.4}$$

where all concentrations are determined by the formation energies of the defects (E) and their related chemical potential (μ), and N is the number of available lattice sites $\sim 10^{23}$ cm^{-3}. Under equilibrium conditions, the chemical potential satisfies the following equation:

$$\mu_{\mathrm{XY}} = \mu_{\mathrm{X}} + \mu_{\mathrm{Y}}. \tag{7.5}$$

Therefore, the concentration of XY-centres will be given by:

$$n_{\mathrm{XY}} = N\exp\left[\frac{-(E_{\mathrm{XY}} - \mu_{\mathrm{X}} - \mu_{\mathrm{Y}})}{kT}\right], \tag{7.6}$$

which reduces to:

$$n_{\mathrm{XY}} = \frac{n_{\mathrm{X}} n_{\mathrm{Y}}}{N}\exp\left[\frac{-(E_{\mathrm{XY}} - E_{\mathrm{X}} - E_{Y})}{kT}\right]. \tag{7.7}$$

The numerator of the exponential term in equation 7.7 is simply the binding energy of the defect, and therefore the binding energy can be estimated from the experimental data in the following way: if the lower limit for detection of the C_s-H centre is around 10^{12} cm^{-3}, the concentrations of C and H being 10^{17} cm^{-3}, and 10^{15} cm^{-3} respectively, then since the dissociation occurs around 100°C, equation 7.7 yields a value of ≈0.2 eV for the defect binding energy. This is clearly in disagreement with the calculated value of around 1 eV. However, as discussed above, we believe that the resolution of these problems is either due to the formation of hydrogen molecules, negative-U properties of the centre, or the enhancement of C_s-H_2 production.

7.3.4 The C_s-H_2 defect

The unstable nature of C_s-H as demonstrated by its low dissociation energy suggests that it readily complexes with other impurities and especially a second H atom. One possibility is that a H_2 molecule would be formed and this might lie close to C_s attracted by the polarisation field. A second possibility is that a H_2^* defect bound to carbon is formed. Since the defect contains CH bonds it is expected to be more stable than a solitary H_2^* defect which disappears at 200°C [111].

There are two likely structures for this defect: (a) the (C_sH_{BC})Si structure of C_s-H with the second hydrogen at the anti-bonding silicon site, and (b) the $(H_{AB}C_s)$Si configuration with the second hydrogen bonded to Si, and at the bond centred site. These are labelled $(C_sH_{BC})(SiH_{AB})$ and $(H_{AC}C_s)(H_{BC}Si)$ respectively, and are illustrated in Figure 7.4a,b. Their total energies are almost identical, the

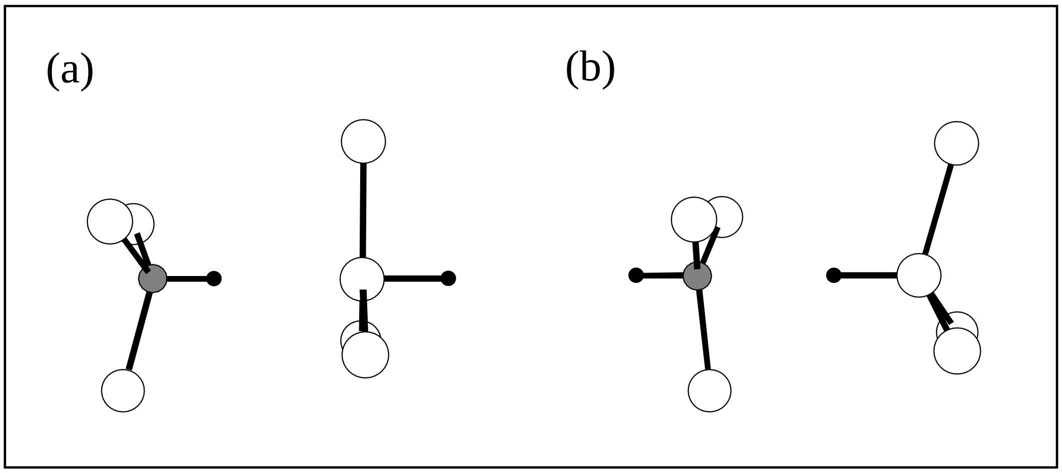


Figure 7.4: The possible C_s-H_2 configurations: (a) the $(C_sH_{BC})(SiH_{AB})$ configuration, and (b) the $(H_{AC}C_s)(H_{BC}Si)$ configuration.

first configuration being lower in energy by 0.02 eV. Details of their structure are given in in Table 7.3. The Kohn-Sham eigenvalues are illustrated in Figure 7.3, shows that they are electrically inactive. The LVMs for both configurations have been calculated, and are given, along with their symmetries, in Table 7.4. There have been no reports of these modes. The highest modes of each configuration

Table 7.3: Bond lengths of the two possible configurations of C_s-H_2 in Si, in Å.

	$(C_sH_{BC})(SiH_{AB})$	$(H_{AC}C_s)(H_{BC}Si)$
Charge state	0	0
C-H	1.097	1.084
Si-H	1.501	1.484
C-Si (3)	1.976	1.984
Si-Si (3)	2.324	2.312

Table 7.4: Calculated local vibrational modes of the two C_s-H_2 configurations along with their isotope shifts.

Symmetry	^{12}C H H	^{12}C {H,D} mix		^{12}C D D	^{13}C H H
$CH_{BC}SiH_{AB}$					
A_1	3004.3	3004.3	2204.3	2203.6	2995.7
A_1	2135.1	1532.3	2134.4	1532.3	2135.1
E	1232.1	1232.1	925.0	915.3	1228.6
E	920.0	655.1	911.0	654.6	920.0
E	642.3	641.2	617.4	617.4	627.0
$SiH_{BC}CH_{AB}$					
A_1	3053.2	3051.5	2420.8	2234.4	3045.1
A_1	2412.3	1732.4	2227.1	1731.9	2412.3
E	1074.8	1074.7	791.4	791.2	1073.1
E	664.8	664.3	641.5	641.1	648.9
E	629.1	429.1	629.0	429.1	628.7

correspond to C-H and Si-H stretch respectively. It is likely that these modes are overestimated by about 10% as similar calculations give this overestimate for both VH_n defects in Si [59] and C-H defects in GaAs [13]. The compressed Si-H bond in the second configuration results in a significant increase of its stretch mode compared to that of the first. The C-H and Si-H wag modes are much higher than those in the second configuration. Both configurations give rise to a carbon related E mode around 650 cm^{-1}. The large differences in the wag and stretch modes of these C_s-H_2 defects suggests that their measurement might enable the models of the defect to be distinguished.

7.4 Interstitial carbon-hydrogen defects

7.4.1 Experimental background

Minaev and Mudyri [112] observed several sharp PL lines from Cz-Si annealed between 450°C and 600°C. The presence of oxygen and carbon in many of these was deduced from their presence in the material. Deliberate introduction of H and C, followed by *e*- or *n*-irradiation can dramatically affect the intensities of some of these PL lines. Isotopic analysis leads to shifts in the lines showing that many of these defects contain carbon (eg the P and H enters) and both carbon and hydrogen (I, M, and T centres) [113, 114, 115, 116, 117]. The I, M, and T lines are believed to contain one hydrogen atom, due to the splitting of their zero-phonon lines into doublets when both H and D are present in the material [117]. Since these lines were also observed in material which was thought to be nominally H free, their presence showed that H is a common unintentional impurity in Si, a very important result [114, 116].

The T-centre

Of particular interest here is the the T-line (935.1 meV) which has been studied most extensively. The T-line is observed in either FZ or CZ silicon after radiation damage and subsequent thermal annealing in the temperature range 400-600°C. In addition to this, the centre can be created by thermal treatment alone in C rich CZ material, and a small fraction of FZ samples. This indicated that interstitials must be generated before the centre forms, and therefore it is likely that the T-centre is an interstitial related defect. The phonon-assisted region at lower energy contained a broad sideband of perturbed lattice modes, typical of carbon-containing defects [90], as well as a number of local modes. Isotope shifts in this zero-phonon line with ^{13}C [113] confirmed the presence of carbon in the defect, and further studies [115] showed that the defect contained H and possessed C_{1h} symmetry.

There is no evidence from the splitting of the zero phonon line in mixed isotopic cases, for more than one C or H atom. However, C-isotopic shifts of an LVM, observed as a phonon replica to the T-line, demonstrated that the defect contained two inequivalent C atoms. This defect is therefore an example of the failure of the isotopic splittings of the zero phonon line itself in the chemical characterisation of a defect. The T-centre spectrum is shown in figure 7.5, the T-line is at 935.1 meV, and a transition occurs from an excited state 1.75 meV above the T-line. A number of local modes were observed as phonon replicas, labelled L_2 to L_5 – it is certain that these are features associated with the T-line since above 10K, the excited state is mirrored in the replicas. The observed local modes and their isotope shifts are detailed in Table 7.5. A detailed analysis of uniaxial stress and magnetic field

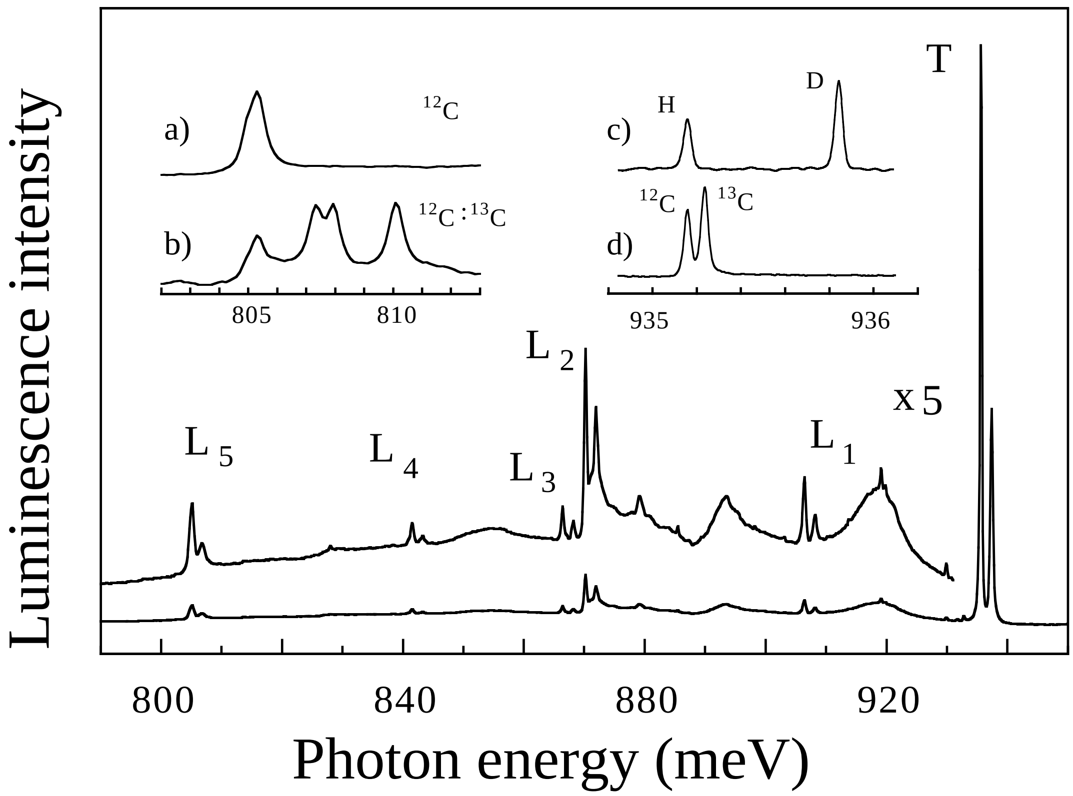


Figure 7.5: The T-line (935.1 meV) and its phonon side bands recorded at 35K. The first insert shows the local mode L_5 with (a) ^{12}C and (b) mixed ^{12}C and ^{13}C showing that the defect contains two inequivalent carbon atoms. The second insert shows the effect of isotopic substitution on the zero phonon line with (c) D and (d) ^{13}C.

perturbations led to the conclusion that the defect gives rise to a deep acceptor level at E_c-0.2 eV [115, 118]. In addition to this, analysis of the magnetic data showed that the defect in its neutral ground state has an uncoupled electron with spin $\frac{1}{2}$ and a g-factor close to 2. The paramagnetic nature of the centre is also consistent with the presence of one H atom in the defect. In recent work [118] (chapter 7.4.4) this and other LVMs along with their isotope shifts have been detailed, and these results, in conjunction with *ab initio* calculations have elucidated the microscopic structure of the defect. Only local modes of A symmetry are detected in PL, and in addition to this, detector limitations meant that modes in the region 2000-3000 cm^{-1} could not be observed in these experiments. This consists of a $\langle 100 \rangle$ oriented C-CH pair occupying a single lattice site, the structural model is discussed in more detail in section 7.4.4. This defect is isoelectronic with C$_i$-P$_s$ and is also the first C related defect known to contain a C-C bond and should be contrasted with the di-carbon centre discussed above. If the H concentration is large, with respect to C, the T-line disappears. We shall discuss possible reasons for this in section 7.6.

Table 7.5: Observed local vibrational modes, cm^{-1}, and their downward shifts with C-isotopes, of the T-centre defect [118].

Symmetry	Mode	^{12}C-(^{12}C-H)	mixed C	13,13
A	L_5	1056.0	18.0, 27.0	45.0
A	L_4	796.0		25.0
A	L_3	567.5	3.0, 7.5	10.0
A	L_2	531.5	2.0, 4.0	5.5
Symmetry	Mode	^{12}C-(^{12}C-D)	mixed C	13,13
A	L_5	1052.0	16.5, 20.5	38.0
A	L_4	759.0		16.5
A	L_3	558.5	2.5, 5.3	7.7
A	L_2	528.0	-1.8, 1.3	2.3

The I-centre

The I-line (965.0 meV) is observed in heat treated CZ Si, with [C]$\approx 2\times 10^{17}$cm^{-3} and [0]$\approx 1\times 10^{18}$cm^{-3}, along with the T-centre, after electron irradiation, and subsequent annealing in the temperature range 450-500°C. Isotope shifts of the zero-phonon line with ^{13}C and D confirmed the presence of carbon and hydrogen in the centre, and the shifts of local modes show that the centre contains two inequivalent C atoms and one H atom [119]. These results closely resemble those of the T-centre, and the local modes are given in the Table 7.6 below. As yet there is no spectroscopic

Table 7.6: Local vibrational modes of the I-centre (cm^{-1}), and their ^{13}C isotope shifts. For comparison, the local modes of the T-centre are shown in brackets.

Mode	^{12}C-(^{12}C-H)	^{12}C-(^{13}C-H)	^{12}C-(^{13}C-H)	^{13}C-(^{13}C-H)
L_2	543.1 (531.5)	(4.0)	5.7 (2.0)	(3.8)
L_3	586.9 (567.5)	4.4 (7.5)	7.2 (3.0)	11.1 (10.0)
L_5	(1056.0)	(18.0)	(27.0)	(45.0)
Mode	^{12}C-(^{12}C-D)	^{12}C-(^{13}C-D)	^{12}C-(^{13}C-D)	^{13}C-(^{13}C-D)
L_2	536.6 (528)	(1.3)	5.5 (2.5)	(7.7)
L_3	578.3 (558.5)	4.2 (5.3)	5.8 (2.5)	9.4 (7.7)
L_5	1054.5 (1056.0)	12.2 (16.5)	29.1 (20.5)	44.6 (38.0)

evidence for the incorporation of oxygen in the centre, however the defect is the dominant luminescence line in CZ Si, and is not observed in FZ Si; suggesting that O might be a constituent of the defect. The T-centre is stable up to 600°C, and the I-centre is created in the temperature range 450-500°C - the temperature range in which O_i defects become mobile. Furthermore the similarity of the electronic structure and local modes of the T and I-lines suggests that the I-line is merely the T-centre defect perturbed by a nearby oxygen atom, The paramagnetic activity of the complex is consistent with the presence of an odd number of hydrogen atoms. Uniaxial stress measurements reveal that the symmetry of the centre is C_{1h}, and hence the oxygen - if part of the defect - must lie in the T-centres mirror plane. The authors concluded that the T-line defect trapped an oxygen atom creating the I-centre [119].

The M-centre

Recently, the defect giving rise to the M-line (760.8 meV) has been studied in detail [120]. The centre appears in FZ material in the temperature range 300-600°C, and again is known to be carbon and hydrogen related. Several local modes of the centre are observed, and confirm the presence of one hydrogen, and two or more carbon atoms in the defect core. Similarities were drawn between some of the local modes of the M and G line defects, with the suggestion that the M-line complex was formed when hydrogen attacked the G-centre. Uniaxial stress measurements showed that the defect had low C_1 symmetry, and the acceptor level position was estimated to be E_c-0.37 eV. Magnetic perturbations revealed that the centre was paramagnetic, again consistent with the centre containing one hydrogen atom.

7.4.2 The C_i-H defect

The isolated C_i defect shown in Figure 6.2b possesses two undercoordinated atoms (C and Si) and therefore H could readily bond to either of these atoms forming a $\langle 100 \rangle$ oriented defect with H bonded to C (Figure 7.6a), or with H bonded to Si (not illustrated). These configurations are labelled $(C_iH)_{\langle 100 \rangle}Si$ and $C_i(SiH)_{\langle 100 \rangle}$. There is also the possibility of a structural change to a BC form with either the C-H unit at the centre of a Si-Si bond $(C_iH)_{BC}Si$ (Figure 7.6b), or with a Si-H unit at the centre of a C_s-Si bond $C_s(Si_iH)_{BC}$. All four of these configurations possess C_{1h} symmetry. The four configurations were inserted into a tetrahedral 131 atom cluster, and all atoms relaxed. The defects retained their C_{1h} symmetry throughout the relaxation. In the neutral charge state, two structures possessed almost degenerate energies in the neutral charge state: the BC defect described as $(C_iH)_{BC}Si$ possessed a energy 0.03 eV above the $(C_iH)_{\langle 100 \rangle}Si$ structure. In the BC structure, carbon made bonds of length 1.078 and 1.797 and 1.811 Å with hydrogen and both silicon neighbours respectively. In the $\langle 100 \rangle$ structure, the bonding (in Å) was as follows C_4-H 1.092, C_4-Si_4 1.944 (2), C_4-Si_3 1.861, and Si_3-Si_4 made two inequivalent bonds of length 2.249, 2.252. Here the subscript refers to the coordination of the atoms, and any number in brackets indicated the multiplicity of the bond, The other structures

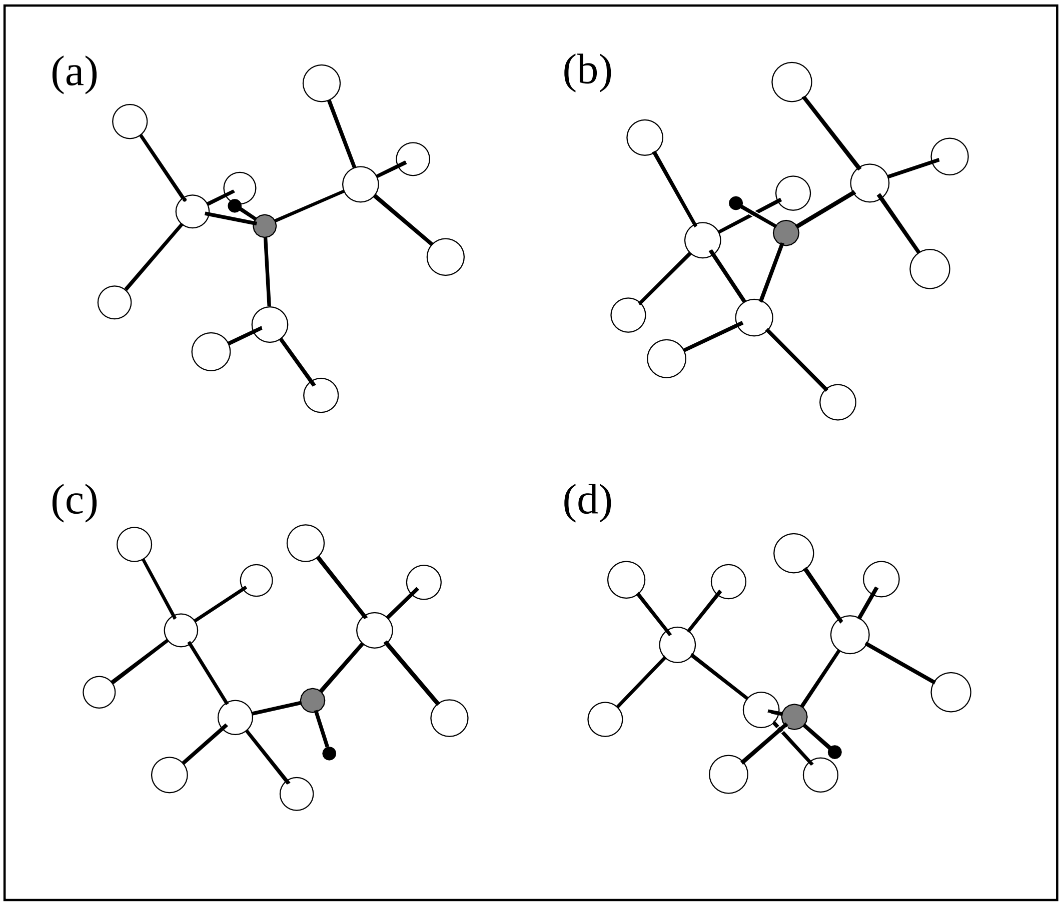


Figure 7.6: The possible configurations of the C_i-H defect in Si, and the defect migration path: (a) the $(C_iH)_{\langle 100\rangle}$Si configuration, (b) the $(C_iH)_{BC}$Si configuration, (c) the $(C_iH)_{BC}$Si configuration with the rotated CH unit, and (d) reformation of the $(C_iH)_{\langle 100\rangle}$Si structure.

were less stable by at least 0.47 eV. The binding energy of H with C_i in the $\langle 100\rangle$ form was evaluated in the same way as described above for C_s-H. This gave a value of 2.8 eV which is, as expected, considerably greater than the binding of H with C_s.

The near degeneracy in the energies of the $\langle 100\rangle$ and BC forms is surprising in view of the very different bonding patterns. These two configurations possessed very different energies for the C_i defect and it is the presence of H which stabilises the BC form. It is interesting to note that carbon can again, as in the di-carbon defect, form two configurations of the same defect, with a small absolute energy difference between the structures, and the barrier for conversion between the two is again likely to be low as only one bond is broken followed by the formation of a new bond in this process. The results suggest the defect could rapidly diffuse. This is because a long range diffusion path can be constructed connecting these forms. The saddle point for this migration path was found by the following procedure. A number of constrained relaxations were carried out as described in section 3.3.1.

The initial cluster coordinates in each run were linearly interpolated between the relaxed ⟨100⟩ and ⟨111⟩ bond centred forms illustrated in Fig. 7.6a-b. Thus these coordinates are simply $(1-\alpha)R_1+\alpha R_2$ where R_i describes the ⟨100⟩ and ⟨111⟩ forms respectively. The energies of each run versus α are shown in Fig. 7.7 (α_1). As the defect moves away from the ⟨100⟩ configuration into the ⟨111⟩ BC form, the energy rises until $\sim$ 0.27 eV peaking around α=0.6 and then dropping as the BC form is realised. This value of α corresponds to the barrier for inter-conversion between the two low energy forms of the defect. There is then the possibility of a ±120° rotation of the bond-centred C_i-H unit about the ⟨111⟩ axis, the BC defect having three equivalent structures. A second constrained run with R_i representing the BC_1 and BC_2 configurations (figs 7.6b,c) was performed, and the results are shown in figure 7.7 (α_2). The 0.16 eV barrier for the rotation of the bond centred CH unit is lower than the barrier for interconversion between the two different structures, and peaks around α=0.25 and α=0.75. There is a slight reduction in energy at α=0.5 to 0.11 eV, where the C_i-H unit passes through a C_{1h} mirror plane (corresponding to a rotation through 60°). Following the rotation of the BC defect, the migration process is completed with the reformation of the ⟨100⟩ oriented defect (figure 7.6d). From these results, it is clear that the migration energy for this defect is extremely

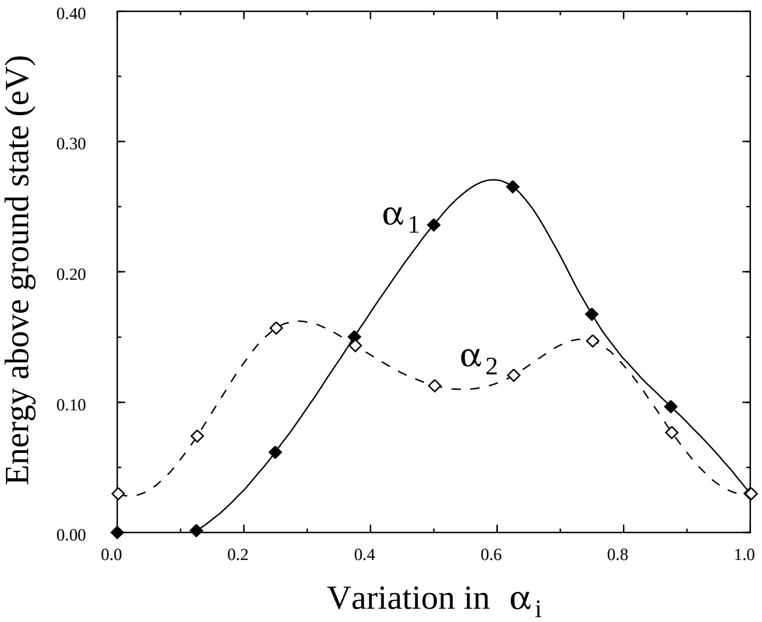


Figure 7.7: The energy as a function of α for the restructuring and rotation of C_i-H.

low, around 0.3 eV. This is considerably lower than the 1.1 eV calculated for the migration of C_i [121] where the observed value is 0.8 eV. It may seem surprising that the C_i-H defect diffuses easier than C_i. However, we think this reflects the low energy of the BC configuration which is stabilised by the H atom: without H, the C atom would possess two dangling bonds. We therefore expect the C_i-H defect to be an extremely reactive unit, rapidly migrating to and complexing with other defects present in the material.

The defect gives rise to a gap level which contains one electron. For $(C_iH)_{\langle 100\rangle}$Si this was above mid-gap, and the level appeared below mid-gap for the $(C_iH)_{BC}$Si structure. Although we cannot exactly place the defect levels accurately, it is clear that this positioning of these levels is consistent with the highest occupied state being a single electron occupying either a Si or C related p-orbital. Hence the defects are paramagnetic in the neutral charge state with $S = \frac{1}{2}$. The Kohn Sham levels for all four of these configurations, as well as their dependence on charge state are illustrated in Figure 7.8.

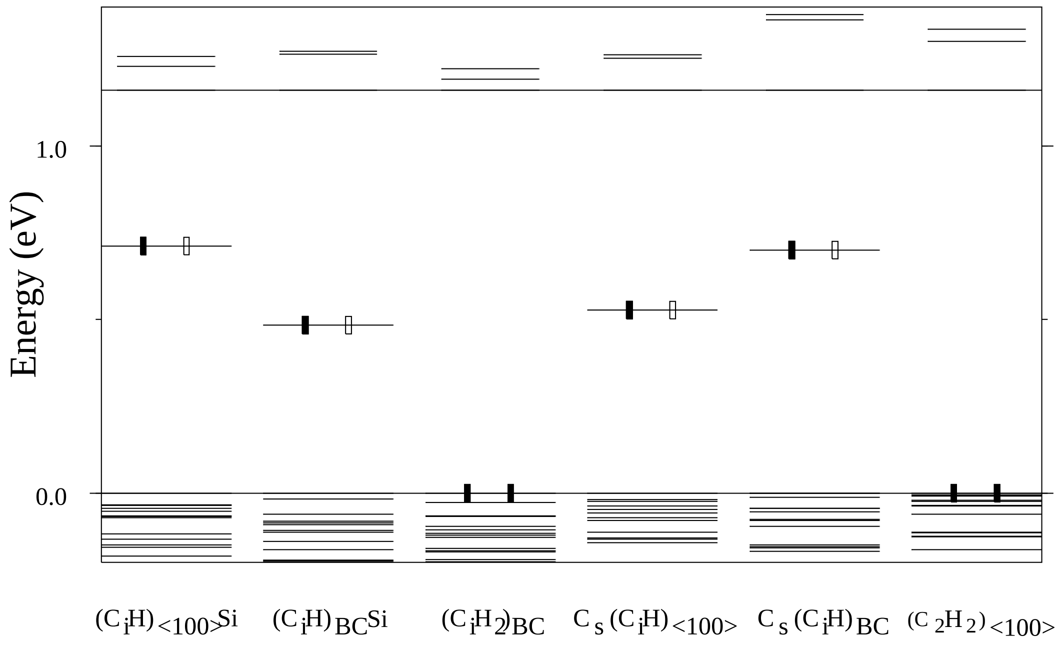


Figure 7.8: The Kohn Sham levels of the interstitial CH defects.

We now investigate the charged C_i-H defects. Their structures are almost identical to the neutral defect. It is clear that the difference in energy between the two forms of the charged defect is much greater than the neutral case. In the negative charge state, the BC structure proved to be lowest in energy — being 0.55 eV below the $\langle 100\rangle$ structure. Conversely, in the positive charge state the $\langle 100\rangle$ oriented defect lay 0.72 eV below the BC configuration. The greater difference in the energies of the two forms must result in a greater diffusion energy for the charged defects and these must be at least 0.55 and 0.72 eV respectively if the same paths are followed. These are comparable with the migration energy for C_i. The LVMs

Table 7.7: Calculated local modes of C_i-H configurations, cm^{-1}.

Symmetry	^{12}CH	^{12}CD	^{13}CH
BC			
A	3050.0	2234.6	3041.6
A	1142.8	772.0	1134.7
A	952.4	1010.1	928.1
B	627.8	549.8	622.0
A	573.7	569.6	569.4
⟨100⟩			
A	3029.2	2219.5	3020.8
B	1038.0	797.9	1035.2
A	974.3	786.5	969.3
A	727.1	649.9	710.0
B	683.6	652.7	664.7

are given in Table 7.7 for the neutral defect. Anharmonic effects will probably be considerable as the highest modes and as the energies for these excitations are close to the migration energy of the neutral defect, they are unlikely to be detected. This is similar to the situation for the di-carbon defect [121].

7.4.3 The C_i-H_2 defect

The rapid diffusivity of C_i-H suggests that it is likely to react easily with other impurities and we now consider the effects of the addition of a second H which would render the defect electrically inactive and probably immobile. Once again, there are two likely configurations for the C_i-H_2 defect arising from the saturation of the radical in C_i-H : the ⟨100⟩ oriented CH-SiH split interstitial (Figure 7.9a) labelled $(C_iH_2)_{\langle 100\rangle}$ and a (C_i-H_2) unit lying at a bond centre between two silicon atoms (Figure 7.9b), $(C_iH_2)_{BC}$. The first of these does not possess any symmetry, and the latter has C_{1h} symmetry. An 88 atom cluster $CSi_{43}H_{44}$ was used to investigate these forms. The BC structure proved to be lowest in energy, being 0.58 eV below the ⟨100⟩ structure. This reflects the strong C-H bond strength. Fig. 7.8 shows that the second H has completely passivated the defect. The addition of the second hydrogen atoms modifies the bond lengths from those in the $(C_iH)_{BC}$Si defect, slightly increasing them to 1.090 (2), 1.860 and 1.872 for the C_4-H and C_4-Si bonds respectively. Table 7.8 gives the calculated vibrational modes of the defect. Since this defect is electrically inert and possesses high thermal stability, it is surprising that there have been no reports of its associated vibrational modes. Its Si_iH_2 analogue has been detected in low temperature proton implanted samples [62] and has

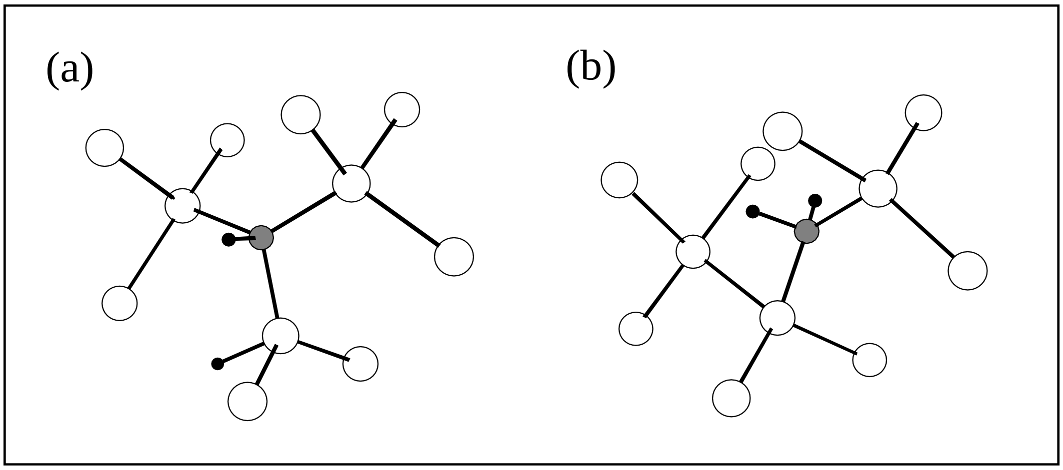


Figure 7.9: The possible C_i-H_2 configurations: (a) the $(C_iH_2)_{\langle 100\rangle}$ configuration, and (b) the $(C_iH_2)_{BC}$ configuration.

Table 7.8: Calculated local modes of the $(C_iH_2)_{BC}$ configuration, cm^{-1}.

Symmetry	^{12}CHH	^{12}CHD	^{12}CDD	^{13}CHH
$(C_iH_2)_{BC}$				
A	3092.5	3076.4	2266.9	3087.4
B	3062.5	2252.3	2235.3	3051.4
A	1448.0	1288.9	1066.3	1443.8
B	1121.4	754.6	798.6	1121.4
A	1093.3	1088.8	728.1	1085.8
A	850.6	904.1	923.1	827.9
B	692.6	634.7	601.9	683.3
A	576.0	570.4	567.7	570.5

been the subject of theoretical calculations [63, 122, 59]. This centre was discussed in chapter 5.

7.4.4 The C_s-(C_i-H) defect

Another possible reaction which C_i-H might undergo is trapping by C_s. As well as the capture of C_iH by C_s, a C_s-(C_i-H) defect could be formed by the capture of H by the di-carbon defect. These structures will also be investigated, and in the next section, we will outline the most likely formation mechanism of all of these defects.

For the C_s-(C_i-H) defect, a combination of the 86 and 131 atom clusters were used, modified to contain the defect. There are of course many possible configurations of such defects but the most likely ones are: (a) The A-form of the di-carbon defect with H bonded to either of the undercoordinated atoms, (b) the B-form of the di-carbon defect with H bonded to the Si_i, (c) a $\langle 100 \rangle$ oriented C-(CH) interstitial defect with these atoms occupying a single lattice site (Figure 7.10a) labelled $C_s(C_iH)_{\langle 100 \rangle}$, (d) a bond centred C_i-H unit with a neighbouring C_s atom (Figure 7.10b sites 1,2) labelled $C_s(C_iH)_{BC}$, (e) a bond centred C_i-H unit with a substitutional carbon atom in the next nearest neighbour site (Figure 7.10b sites 3,4), and finally (f) a $\langle 110 \rangle$ oriented split-interstitial defect with the C-CH occupying a single lattice site. All of the above configurations possessed C_{1h} symmetry. The PL centre

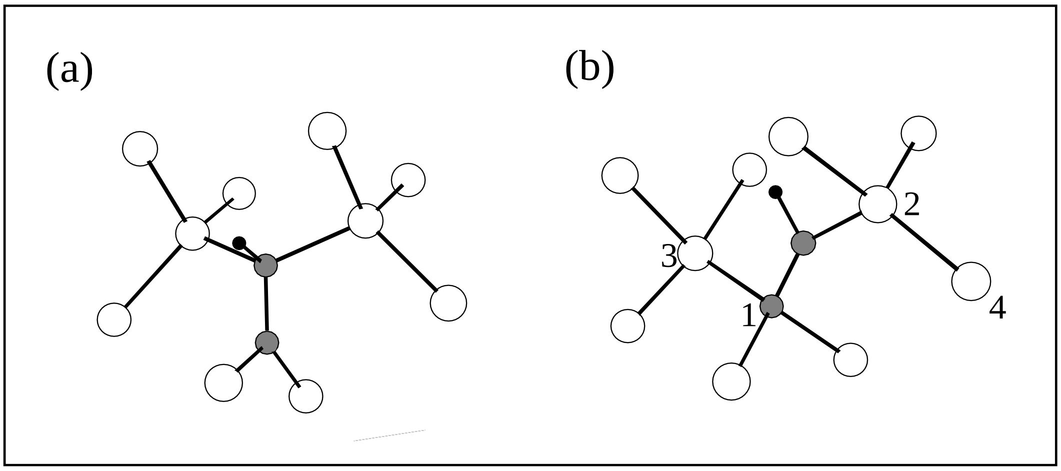


Figure 7.10: The low energy C_2-H defect configurations: (a) the $C_s(C_iH)_{\langle 100 \rangle}$ configuration (T-centre defect), and (b) the $C_s(C_iH)_{BC}$ configuration.

known called T is also known to contain two carbon atoms and one hydrogen atom.

We first note that the C-related L_5 mode at 1056 cm^{-1} lies within 130 cm^{-1} of modes due to C_i and more than 400 cm^{-1} above those of C_s. This suggests that the T-centre contains C_i. We now argue that the observed C-isotope shifts of the L_5 mode suggest the presence of a C-C bond. This follows as the reduced mass of an isolated C-C unit can be expressed as:

$$\frac{\delta \nu}{\nu} = \frac{\delta(1/M_{CH} + 1/M_C)}{2(1/M_{CH} + 1/M_C)} \tag{7.8}$$

and implies that its stretch frequency would shift downwards by 20, 20 and 41 cm^{-1} for ^{13}C-^{12}C, ^{12}C-^{13}C, and ^{13}C-^{13}C respectively. These shifts roughly agree with the observed values of 18, 27 and 45 cm^{-1} for the H case, and 16.5, 20.5 and 38 cm^{-1} for D. The different shifts in the two mixed C-isotope cases imply that the C atoms are inequivalent, and the distinct shifts in the H and D cases suggest that H is bonded with one of the C atoms. Thus we look for models with C_{1h} symmetry containing C_i and C-C-H units, although, in addition to such structures, we considered all of those listed above. It is unlikely that both C atoms are interstitial for such a model would not yield the observed isotope structure of L_5. In addition, such a model would give rise to two high frequency C-related modes and only one has been observed.

The defects were inserted into 86 atom clusters, forming $C_2HSi_{43}H_{42}$, and all atoms relaxed with a full symmetry constraint until the equilibrium geometries were determined. The ground state configuration is that described in (c), where H saturates a C radical in the $\langle 100 \rangle$ oriented C-C split interstitial. A second configuration $C_s(C_iH)_{BC}$ has an energy 0.28 eV higher. The energies of the other configurations, relative to the ground state are as follows: A-C_sC_iH, the A-form of C_i-C_s with H bonded to the C_i 0.55 eV; B-C_sC_iH, the B-form of C_i-C_s with H bonded to the Si_i 1.81 eV; C_s+$(C_iH)_{BC}$. the C-H BC model with the Si labelled 3,4 (Figure 7.10b) 0.68 eV. It is interesting to note that both low energy configurations possess a C-C bond, contrary to the case for the isolated di-carbon centre.

The Kohn-Sham eigenvalues are shown in Fig. 7.8, and indicate that the defect gives rise to one gap level, containing one electron. This occurs around mid-gap, and is related to a non-bonding p-orbital on the undercoordinated carbon atom. An estimate for the donor and acceptor level positions was obtained using the Slater transition method, and scaling the results to the experimental band gap. This gave the acceptor level at E_c-0.30 eV, in reasonable agreement with the observed value of E_c-0.21 eV. The donor level was found to be at E_v+0.28 eV, and has thus far escaped detection experimentally. Again, the results were scaled down to the Si band gap as described in [17]. The acceptor level position can then be explained by a large Hubbard-U term of at least 0.4 eV. The highest occupied wavefunction consisted of a singly occupied non-bonding p-orbital on the under-coordinated carbon atom, denoted C_3. Therefore, only small isotropic hyperfine couplings with C_4 and H are to be expected. The neutral ground state structure is paramagnetic, with $S=\frac{1}{2}$ and is isoelectronic with the IA configuration of the C_i-P_s defect which has been investigated experimentally [123].

The defect has been assigned to the 0.9351 eV PL centre, or T-centre. The structure and vibrational modes of and their relationship with the modes observed for the T-centre will be described below [118]. Analysis of the Zeeman splitting of the PL line confirms that the neutral state of the T-line defect is paramagnetic with $S=\frac{1}{2}$ consistent with the singly occupied Kohn Sham level from the calculation, illustrated along with the BC configuration levels in Figure 7.8.

The calculations show that the C_3-C_4, C_4-H, C_3-Si and C_4-Si bond lengths are 1.46, 1.11, 1.91 (2) and 2.05 (2) Å , respectively. The calculated vibrational modes and their isotope shifts are given in Table 1. Now, PL only detects modes of A-symmetry and we concentrate on those at 2913, 1180, 1098, 744, 558 and 542 cm^{-1}. The first four modes correspond predominantly to C_4-H stretch, C_4-H wag, C_3-C_4 stretch and C_3-Si stretch respectively. The last two modes at 558 and 542 cm^{-1} are close to the Raman frequency, and involve predominantly the motion of C_3 and C_4, respectively, as well as their Si neighbours. There is close agreement between the modes L_2 to L_5 observed by PL and the calculated vibrational modes which enables them to be identified (Table 1). The two highest energy modes at 2913 and 1180 cm^{-1} are undetected by PL. However, the intensities of the phonon-assisted transitions are critically dependent on the exciton-phonon coupling, and not necessarily all local modes can be detected by PL. In addition, there is an experimental problem in observing the highest energy local mode since the Ge detector used is insensitive in the region where the position of the corresponding luminescence line is expected. In general, these are close to the observed ones and allow us to identify L_2 to L_5. The C_3-C_4 stretch mode lies close to L_5 and has carbon isotope shifts 19, 21, 41 cm^{-1} in the H case and 17, 24, 42 cm^{-1} in the D case. These shifts are in good agreement with the experimental values 18, 27, 45 cm^{-1} and 16.5, 20.5, 38 cm^{-1}, respectively, observed for L_5. We therefore identify the L_5 mode with C_3-C_4 stretch. The C_3-Si stretch mode at 744 cm^{-1} involves the movement of H as well, but has little amplitude on C_4. As a consequence, this mode displays large isotope shifts with both ^{13}C and D — 23 cm^{-1} and 30 cm^{-1} respectively — but essentially no additional modes appear in the mixed C-isotopic cases. These results allow us to identify the mode with L_4 at 796 cm^{-1}, which decreases by 25 and 37 cm^{-1} for ^{13}C and D, respectively, and involves the motion of only one of the C atoms. Finally, the two lower A modes at 558 and 542 cm^{-1} (H) are in good agreement with L_3 and L_2, at 567.5 and 531.5 cm^{-1}, respectively. These modes are more delocalised and have only small shifts with D and the various combinations of the C-isotopes. Configuration (d) is a dissociated form of the defect where the C_s and C_i-H units are separated by one silicon atom. The energy of this configuration, and other clusters involving the dissociated units suggests that the binding energy of C_i-H with C_s is between $\sim$ 0.6-1.0 eV.

The binding energy of the T-centre can be estimated from the experimental results using identical arguments discussed in the section on C_s complexes. Again, if we assume that the detectable limit of the defects is 10^{12} cm^{-3}, the concentrations of C_i-H and C_s are 10^{15} cm^{-3}, and 10^{17} cm^{-3} respectively, then inserting these values along with kT at the dissociation temperature into equation 7.7, the binding energy is found to be approximately 0.5 eV, in good agreement with the calculated value.

However, it is likely that the thermal dissociation of the defect occurs when the H 'boils off' at high temperatures, followed by the dissociation of the C-C unit, or alternatively, the C_i-H unit is driven off. Thus the T-centre is a thermally stable defect with an undercoordinated C atom and may trap a second H atom. This

Table 7.9: Calculated local vibrational modes, cm^{-1}, and their downward shifts with C-isotopes, of the $\langle 100 \rangle$ C_s-(C_iH) (T-centre) defect [118].

Symmetry	Mode	^{12}C-(^{12}C-H)	12,13	13,12	13,13
A		2913.6	8.3	0.0	8.3
A		1180.4	1.8	5.5	6.9
A	L_5	1097.8	19.2	20.7	40.8
B		870.7	1.7	0.0	1.7
A	L_4	743.6	1.0	22.0	23.1
A	L_3	558.0	4.9	3.4	7.6
A	L_2	542.4	3.7	0.2	3.8
Symmetry	Mode	^{12}C-(^{12}C-D)	12,13	13,12	13,13
A		2138.3	12.1	0.0	12.2
A		892.6	4.5	8.1	12.5
A	L_5	1102.2	17.1	24.3	42.3
B		643.6	4.0	0.2	4.2
A	L_4	713.7	0.9	16.6	17.3
A	L_3	552.7	4.0	2.7	6.1
A	L_2	539.2	2.5	0.1	2.5

passivated di-carbon centre will be discussed in the next section.

7.4.5 The C_2-H_2 defect

In samples with large concentrations of H, the intensity of the T-line is greatly reduced. One possible explanation is that the C_i-H precursor defect traps a second H atom, becoming immobile thus blocking the formation of the T-centre. A second possibility is that the T-centre rapidly traps a second H atom.

Here, we investigated three possible configurations of the C_2-H_2 defect, using a 134 atom tetrahedral cluster, $C_2Si_{70}H_{62}$ in which all atoms were relaxed. The configurations examined were: (a) the $\langle 100 \rangle$ oriented defect labelled $(C_2H_2)_{\langle 100 \rangle}$, (b) a BC defect, with a C_i-H_2 unit between C_s and a Si neighbour, labelled $(C_2H_2)_{BC}$, and finally (c) $(C_2H_2)_{\langle 110 \rangle}$, where two (C-H) units share a lattice site this orientation. The symmetry of these three configurations was constrained to C_2, C_{1h} and C_{2v} respectively. The relaxed structures of these defects are illustrated in Figure 7.11(a-c). The most stable defect is $(C_2H_2)_{\langle 100 \rangle}$ where H saturates the C radical in the T-centre, the other structures were 0.99 eV $(C_2H_2)_{BC}$ and 3.59 eV $(C_2H_2)_{\langle 110 \rangle}$ higher in energy. In this defect, both C atoms are equivalent, making bonds of 1.085, 1.500, 2.014, and 2.019 Å for the C-H, C-C and two C-Si bonds respectively. The Kohn-Sham eigenvalues indicated that the stable defect is electrically inactive. The calculated LVMs are given in Table 7.10. The defect is analogous with the

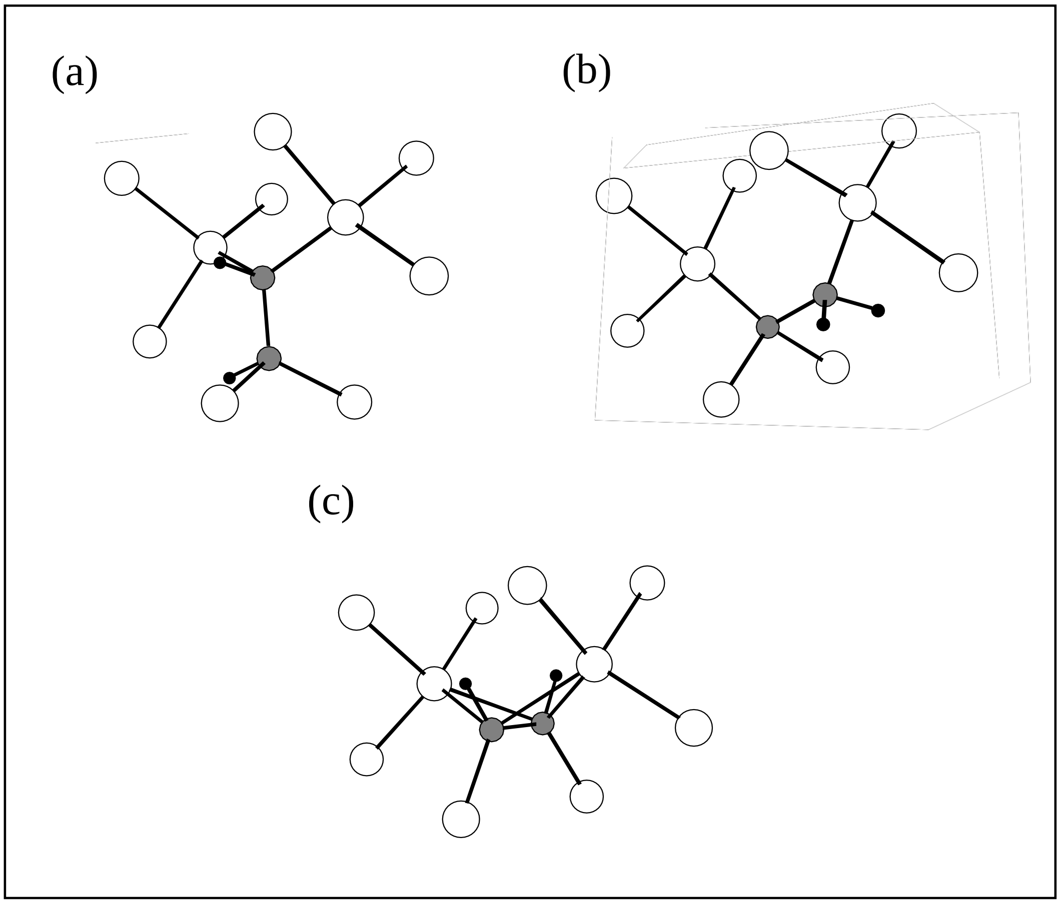


Figure 7.11: The possible structures for the C_2-H_2 split interstitial defect: (a) the $(C_2H_2)_{\langle 100\rangle}$ configuration, (b) the $(C_2H_2)_{BC}$ configuration, and (c) the $(C_2H_2)_{\langle 110\rangle}$ configuration.

hydrogenated self-interstitial defect (Si_iH_2) which also has a $\langle 100\rangle$ orientation [122, 59] with C_2 symmetry. For comparison, the calculated and experimentally observed modes for this defect are also given. The C_2-H_2 defect possesses many localised modes lying above the Raman frequency. In addition to C-H stretch and wag modes around 3000 and 1200 cm^{-1}, there is a high frequency C-C mode at 1040 cm^{-1} as well as local modes from the C-Si back bonds.

7.4.6 The I-line

Two configurations of the I-line defect were investigated, with the oxygen atom being placed in either of the Si-Si back-bonds adjoining the C_3 atom, preserving the C_{1h} symmetry. The configurations differed in energy by 0.2 eV, the lowest energy structure being the one with the larger separation between the H and O atoms. The local vibrational modes of both configurations were calculated, along with their D, ^{13}C, and ^{18}O isotope shifts - these were very similar for both structures,

Table 7.10: The calculated vibrational modes cm^{-1}, and isotope shifts for the $\langle 100 \rangle$ C_2-H_2 interstitial defect.

Symmetry	$^{12}C^{12}CHH$	$^{12}C^{12}CHD$	$^{12}C^{12}CDD$	$^{12}C^{13}CHH$	$^{13}C^{13}CHH$
A	2995.6	2992.6	2196.8	2993.6	2987.1
B	2989.6	2196.1	2195.5	2982.6	2980.8
A	1226.6	1223.7	926.1	1224.3	1221.4
B	1220.9	895.6	882.9	1218.6	1216.8
B	1039.8	1043.1	1046.4	1022.6	1006.8
A	1000.0	993.7	743.0	998.7	997.6
B	976.7	741.3	722.3	973.8	969.1
A	633.5	616.1	601.7	628.1	615.2
B	631.4	612.8	599.9	619.2	613.1
A	560.2	557.4	555.3	556.1	552.8

the results for the ground state configuration are given in Table 7.11. The A modes are labelled L_a-L_f, and it is immediately clear that L_c, L_f, and L_h are all perturbed T-centre like modes and agree well with those observed for the I-line. However, the assignment is not conclusive, and detection of further modes - in particular the O_i stretch (L_d) or the L_e mode, the perturbed L_4 mode of the T-centre, would strengthen this model considerably.

The Slater transition argument was used to estimate the position of the I-lines acceptor level to compare its position to that of the T-centre, since the electronic properties of the defect should be similar. The I- and T-lines occur at 965.0 and 935.1 meV respectively, and therefore it is expected that the I-line acceptor level will be slightly shallower than that of the T-centre. The acceptor level was estimated using a negatively charged cluster, with a fractional of the defect level and conduction band with half an electron each. The resultant energy was then scaled by 3.5 eV, the calculated band gap from a 297 atom pure Si cluster. Similar to the T-centre, the defect gave rise to one level in the gap, and the calculated position of the acceptor level was E_c-0.27 eV. This is therefore 27 meV shallower than the calculated position of the T-centres level, and is in excellent agreement with the 29.9 meV difference in energy between the zero phonon lines of the two defects.

7.5 Other theoretical modelling

Recent work using CNDO methods [124] has considered the formation of CCH centres via a different mechanism namely the capture of H by the di-carbon centre ie.

$$C_i + C_s \longrightarrow (C_i - C_s) \quad \textbf{or} \quad (C_s - \mathrm{Si} - C_s),$$

Table 7.11: LVMs, cm^{-1}, of the proposed structure for the I-line defect, along with their downward isotope shifts. The atomic masses of each each constituent defect atom (C C O H) is given. The hydrogen atoms is bonded to the first listed carbon atom.

Mode/symmetry	12 12 16 1	12 12 18 1	13 12 16 1	12 13 16 1	13 13 16 1
L_a A	3177.2	3177.2	3168.6	3177.2	3168.6
L_b A	1245.2	1245.1	1243.9	1237.1	1235.7
L_c A	1093.2	1089.3	1078.9	1074.1	1061.1
			(14.3)	(19.1)	(32.2)
B	1020.2	1020.2	1019.0	1020.1	1018.9
L_d A	1008.4	963.1	1005.1	1005.8	1000.7
L_e A	893.0	892.5	892.4	866.5	865.8
L_f A	583.6	582.9	578.9	580.7	576.5
			(4.7)	(2.9)	(7.1)
B	583.3	583.3	581.4	574.8	573.1
L_g A	569.8	567.6	568.2	569.7	568.0
B	546.9	545.6	545.9	546.3	545.1
L_h A	544.0	543.7	540.5	543.7	540.2
			3.5	0.3	3.8
Mode/symmetry	12 12 16 2	12 12 18 2	13 12 16 2	12 13 16 2	13 13 16 2
L_a A	2326.6	1326.6	2314.2	2326.6	2314.2
L_b A	986.4	987.2	977.4	969.4	960.3
L_c A	1111.5	1108.5	1101.8	1085.8	1076.2
			(9.7)	(25.7)	(35.3)
B	740.8	740.8	738.1	740.5	737.9
L_d A	1010.5	962.9	1009.9	1007.5	1005.6
L_e A	807.3	807.2	806.3	796.0	795.4
L_f A	580.7	579.9	576.8	578.2	574.7
			(3.9)	(2.5)	(6.0)
B	580.9	580.8	579.6	572.8	571.8
L_g A	568.6	566.6	567.6	568.5	567.4
B	546.5	545.2	545.6	545.8	546.5
L_h A	540.7	540.3	538.4	540.4	538.2
			(2.3)	(0.3)	(2.5)

followed by the capture of H,

$$(\mathrm{C}_i - \mathrm{C}_s) + \mathrm{H}_i \longrightarrow \mathrm{C} - \mathrm{C} - \mathrm{H},$$

or

$$(\mathrm{C}_s - \mathrm{Si} - \mathrm{C}_s) + \mathrm{H}_i \longrightarrow \mathrm{C} - \mathrm{C} - \mathrm{H}.$$

In the alternative model, discussed in the last section, C_i captures H_i forming a rapidly mobile unit, which is then trapped at C_s forming the defect:

$$\mathrm{C}_i + \mathrm{H}_i \longrightarrow (\mathrm{C} - \mathrm{H}_i),$$

followed by,

$$(\mathrm{C} - \mathrm{H}_i) + \mathrm{C}_s \longrightarrow \mathrm{C} - \mathrm{C} - \mathrm{H}.$$

The resulting low energy structure was one where H was situated at a carbon anti-bonding site of the B-form. This structure was found to be lower in energy than the T-centre configuration. Accepting that the two formation mechanisms are possible, then a number of CH related defects are expected to be observed and this is the case experimentally. It may be that if a different formation mechanism occurs, a metastable form may well be observed if the barrier for the reconfiguration is high. The calculations lead to a lowering in energy of 2.7 eV, consistent with the binding energy of 2.8 eV of H to C_i in our calculations. However, it is difficult to believe that H would lie at an anti-bonding site, rather than saturate one of the dangling bonds of the defect, but nevertheless different C-C-H defects may exist if the formation mechanisms differed. Note again that this conclusion is consistent with the analogies drawn with the C-P centre, where no less than four metastable configurations have been identified. Our calculations on the structure found in [124] show that the configuration is unstable, the defect spontaneously relaxing into the hydrogenated A-form structure, with an energy $\approx$0.5 eV greater than the T-centre structure.

7.6 Discussion and conclusions

It is clear that many defects can arise when H complexes with carbon. Substitutional carbon can weakly bind a single H atom leading to a bistable electrically active centre and which can be associated with the E3 and H1 DLTS lines. The H atom lies in a bond centred site for the neutral and positive defects but is anti-bonded to C in the negative charge state. The binding energy of neutral H to C_s was calculated to be around 1.01 eV, in good agreement with other theoretical calculations, and is lower in the charged defects. This result seems to conflict with experimental results that have been interpreted in terms of a low activation energy for dissociation of the neutral species. On the other hand, if the neutral defect was a negative-U one, then two electrons would be trapped and lead to a structural change. Alternatively, the formation of H_2 molecules, or a H_2^* like defect would also lead to the loss of the neutral centre.

We then considered the effect of adding a second hydrogen atom which leads to a stable electrically inactive H_2^* centre. The two most likely configurations, $(C_sH_{BC})(SiH_{AB})$ and $(H_{AC}C_s)(H_{BC}Si)$, possessed almost identical energies, although their LVMs are quite different. There have been no reports of this centre.

We then considered the defects formed when H interacts with C_i. The binding energy of H with C_i is very large, ~ 2.8 eV, and consequently such defects should be readily formed in irradiated material or annealed CZ-Si. The neutral defect exists in two almost degenerate structures consisting of a $\langle 100 \rangle$ oriented and bond centred pair. The bond centred form has a similar structure to O_i. Its migration barrier is 0.4 eV and this low value implies that the defect would easily migrate through the lattice readily forming complexes. The centre is electrically active with a singly occupied level in the band gap. The negative charge centre is is stable in the bond centred form while the $\langle 100 \rangle$ form is favoured in the positive charge state.

The calculations presented here favour the $\langle 100 \rangle$ oriented C_3-C_4-H split interstitial as a candidate for the T-centre (935.1 meV), although additional support would arise from observations of the C-H stretch and wag modes at 2914 cm^{-1} and 1180 cm^{-1}, as well as the mode of B-symmetry. The presence of a C-C bond is surprising as the C_i-C_s defect, which is stable to about 300 °C, only contains C-Si bonds [27]. Our recent calculations have shown that the C-C configuration of the di-carbon centre may also be a low energy structure, however, there is no experimental evidence for this. We suggest that the T-centre is formed when C_i traps H in the BC or $\langle 100 \rangle$ sited defect. This in turn diffuses to C_s where conversion to the T-centre takes place. A calculation shows the trapping of C_i-H by C_s leads to an energy reduction of 0.6 eV [121]. This formation mechanism is more likely than the trapping of H by the di-carbon defect as the latter is not stable at 400-600°C. The calculated local vibrational modes and their isotope shifts are in good agreement with observations on this centre. The T-centre can be observed when CZ-Si is annealed to 450C, or when H soaked FZ-Si containing C is irradiated and subsequently annealed around 450°C. The annealing is probably necessary to break up H molecules and silicon interstitial aggregates.

The identification of this defect, and other related defects, are important for several reasons. It is commonly produced in FZ and CZ silicon samples by irradiation and/or thermal treatment, showing that H is present as an unintentional impurity in silicon [114, 116]. The defect structure and the mechanisms of formation provide further evidence that thermal treatment of CZ Si leads to a generation of Si_i, which can be subsequently trapped by C_s to produce mobile C_i. The stability of the T-centre over an important temperature range (400-600°C) where thermal donors are generated may enable it to be used to monitor Si_i production. The results reported in this chapter show that the presence of hydrogen can lead to the formation of unusually stable electrically and optically active defects. In conclusion, a combination of high resolution photoluminescence experiments, Zeeman and uniaxial stress studies, and *ab initio* theory have successfully elucidated the structure and properties of an interstitial carbon-hydrogen defect in silicon.

The C_i-H centre can be passivated by further H trapping. Such a centre would be be immobile and its formation would block the production of T-centres. Alternatively, T-centres themselves could trap a second H atom forming the inert C_s-C_i-H_2 defect. Both of these defects have been considered, and local modes calculated, to date they have evaded detection experimentally.

The high mobility and reactivity of C_i-H defects suggests that complexes could be formed with many other impurities. There are of course many other possible defects formed, which are beyond the scope of this current work, among these are the reaction of the migrating C_i-H with O, N and other CH defects. It seems likely that CH forms very similar defects to those of isoelectronic impurities such as N or P. For example, we have already pointed out that the C-CH pair is isoelectronic with C_i-P_s. More recent work [125] has showed that a N_iO_{2i} defect is a candidate for the shallow thermal donor, and if N is replaced by CH, there is little change in the structural or electronic properties. At the moment there is no definitive evidence that such complexes exist but there are indications. A shallow H-donor called NL10(H) has H lying along a [110] axis is stable to 520°C [126]. This seems a distinct defect from the H-passivated thermal donors [127] which are stable up to 200°C. However, there is no evidence for the involvement of carbon at the present time.

Chapter 8

Light impurities in Germanium

8.1 Introduction

Due to the chemical similarities between Si and Ge, calculations are now reported on a number of light-impurity related defects in germanium. First, the interaction between hydrogen and intrinsic defects (V and I) is considered, and the calculations support the recent assignment of four local modes at 1883.5, 1881.8, 705.5, and 700.3 cm^{-1} in proton implanted material to the IH_2 defect. The local modes of the VH_n defects are tabulated and a systematic experimental investigation into these defects is now necessary. Attention is then turned to carbon in germanium, and the calculations support the assignment of a local vibrational mode at 531 cm^{-1} to the C_s defect, local modes of the C_i defect are also given. Finally preliminary work on C_s in $Si_{(1-x)}Ge_x$ alloys is presented, it is found that carbon preferentially bonds to Si in this material. The calculated local modes of the defect with varying x is satisfactorily reproduced.

8.2 Calculation details

A tetrahedral 71 atom cluster $Ge_{36}H_{35}$ was used in the calculations on the VH_n complexes, with $n \in \{1, 2, 3, 4\}$. Here the central germanium atom was removed, and n hydrogen atoms were introduced. Four s- and $p-$ Gaussian orbitals of different exponents were placed on the innermost Ge atoms, and a fixed linear combination of four Gaussian orbitals was used on the remaining Ge atoms. Four $s-$ and $p-$ Gaussian orbitals of different widths were used on the n defect hydrogen atom, while a linear combination of three of each of these sets of orbitals was used on the terminating hydrogen atoms. Additionally, two $s-$ and $p-$ Gaussian orbitals were placed at the centre of every Ge-Ge bond in the cluster.

For the self-interstitial split configurations, and their interaction with hydrogen, a 132-134 atom cluster $Ge_{72}H_{60+n}$ was used, where $n \in \{0, 1, 2\}$. The additional host atom and n hydrogen atoms were added to the centre of the tetrahedral 131 atom cluster which represents the perfect crystal. The following basis set was used for both the wave function and charge density in these calculations: eight Gaussian $s-$ and $p-$ orbitals of different exponents were centred on the nuclei, and a fixed linear combination eight $s-$ and $p-$ Gaussian orbitals was used on the remaining silicon atoms. The hydrogen basis was the same as in the VH_n complex. In addition to this, three $s-$ and $p-$ Gaussian orbitals were placed at the centre of every bond in and around the defect core.

To investigate the C_s defect in Ge, the central Ge of a tetrahedral 72 and 131 atom clusters Ge $_{35}H_{36}$, $Ge_{71}H_{60}$ forming $CGe_{34}H_{36}$ and $CGe_{70}H_{60}$ respectively. A full T_d symmetry constraint was imposed for the relaxation. A Ge atom was re-inserted into these clusters at the split-interstitial site, for the calculations on C_i. Eight Gaussian fitting functions of different exponents were used on the carbon atom, and four Gaussian orbitals with differing exponents on the inner 5 and 17 Ge atoms for the 71 and 131 atom clusters respectively. A fixed linear combination of four Gaussian orbitals was used on the remaining Ge atoms. The basis used for the terminating hydrogens was as above. For the runs on the alloys, the 71 atom cluster was used for the local modes and the 131 atom cluster for the binding energy runs - the Si and Ge atom basis was increased to eight Gaussian orbitals here. Three $s-$ and $p-$ Gaussian orbitals were placed at the centre of every Ge-Ge bond in the cluster.

Full symmetry constraints were used in all runs, and all atoms were relaxed until the equilibrium geometry was achieved. For the VH_n defect, the second derivative of the energy with respect to atomic positions was calculated explicitly for the inner 4+H atoms. For IH and IH_2 defect, the second derivative of the energy with respect to atomic displacement was calculated directly for the inner 6+H atoms. Finally, the second derivatives were calculated for the carbon atom and its nearest neighbours, for the C-related defects in Ge and $Si_{(1-x)}Ge_x$. The germanium Musgrave-Pople potential was used for the remaining atoms, and atomic mass used for Ge throughout was 74; the terminating hydrogens were assumed to be infinitely massive, and the local modes were calculated.

8.3 Hydrogen interacting with intrinsic defects in germanium

In chapter 5, we considered the self-interstitial in silicon, and its interaction with hydrogen. The interaction between H with vacancies in silicon had been addressed previously in [59]. In this section, the interaction with hydrogen with native defects in germanium is discussed for the first time.

Due to the similarities in the chemical properties of Si and Ge, it is reasonable to expect that many of the simple defects observed in Si (for example hydrogen related or C_s defects) should also be present when Ge is ion implanted with the appropriate element. The literature to date suggests that hydrogen can behave differently in Ge than it does in Si, for example, the amount of deuterium incorporated into the material during plasma treatment increases with increasing temperature in Si, whereas the converse is true for Ge [128]. In Ge, the passivation of acceptors is is only partially complete [129, 130], and there are no reports of the passivation of donors. In Si however, the complete passivation of both donor and acceptor species is commonplace.

To my knowledge, there has been no reports of local modes being assigned to a specific VH_n centres in Ge, however, recently modes due to the other intrinsic-hydrogen defects have been observed, namely H_2^* [131], and IH_2 [62]. Higher local modes, invariably from VH_n complexes are observed, but these have yet to be studied in detail.

In proton-implanted germanium, a similar series of absorption lines to those seen in Si are observed, and these occur in the range 1750-2100 cm^{-1} (Ge-H stretch) and 650-800 cm^{-1} (Ge-H wag). Of special interest are four local vibrational modes were observed at 1883.5, 1881.8, 705.5, and 700.3 cm^{-1}, and these resemble very closely the results described in chapter 5 for the IH_2 defect in silicon. In fact, within 0.3%, the frequencies of the Ge-H modes scale to the four Si-H modes described earlier. Note that the stretch and bend modes here lie about 100, and 40 cm^{-1} below those of the Si modes, consistent with the reduced bond strength, and the proposal was that in each material the same centre (IH_2) was giving rise to the modes. The 1883.5 and 1881.8 cm^{-1} lines shift downwards to 1359.0 and 1357.6 cm^{-1} with D, however no Ge-D wag modes are detected. In the mixed isotopic case, three additional lines are observed at 703.0, 1358.4, and 1882.8 cm^{-1}, consistent with the assignment of these lines to IH_2. These four lines all annealed out together, disappearing at 200°C in Ge, a slightly lower temperature than in Si, and is also comparable to the difference in annealing temperatures of the H_2^* defect in both materials [111, 131]. The Ge lines were more sensitive to applied uniaxial stress than the Si ones, and as was the case in Si, the symmetry of the centre giving rise to the four lines was found to be C_2. Using identical arguments as before, the polar angle of the bond dipoles was found to be θ=45±1°, and the azimuthal angle was close to 0°, these angles have been defined in chapter 5.

As expected, the results closely resemble those in Si for the both the VH_n $n \in \{1,2,3,4\}$ and the IH_m $m \in \{0,1,2\}$ defects, with both the observed and calculated modes occurring at lower frequencies in Ge than Si due to the slightly weaker X-H bond in the former material. This result is consistent with the lower atomisation energy of GeH_4 when compared to SiH_4 [132].

In the rest of this section, we consider the VH_n complexes in Ge, and then examine the split-self-interstitial configurations, and their interaction with hydrogen.

8.3.1 Vacancy-hydrogen complexes

The vacancy-hydrogen complexes investigated here are illustrated in Figure 8.1 below. All atoms of the cluster $H_nGe_{35}H_{35}$ were allowed to relax until the equilibrium

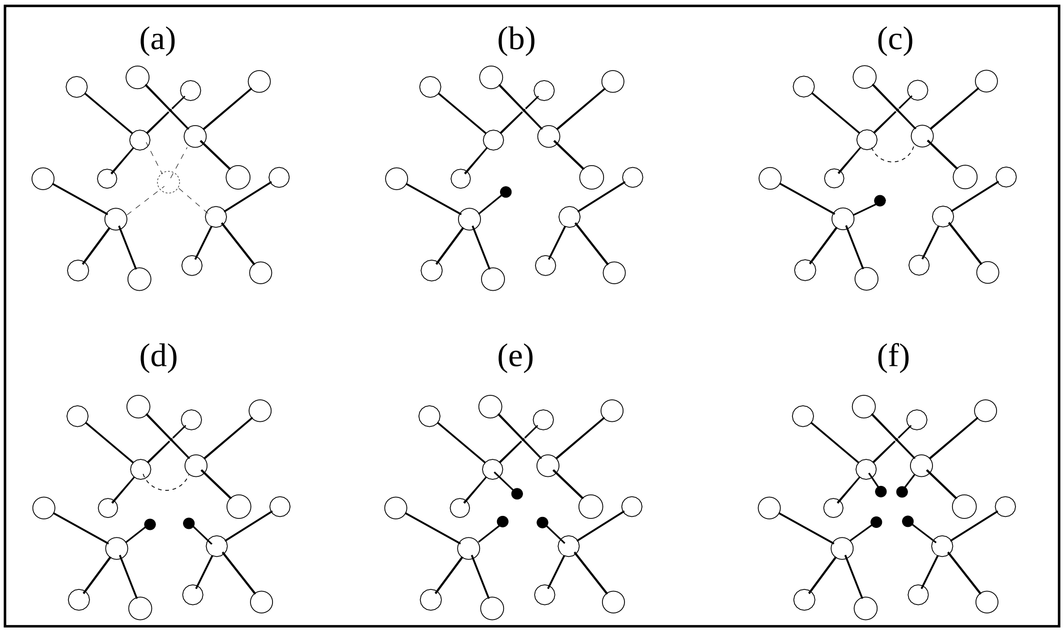


Figure 8.1: The VH_n complexes in Ge: (a) shows the the vacancy, the missing atom being shown by the dashed circle, (b) the C_{3v} VH complex, (c) the VH complex after the Jahn Teller distortion, (d) VH_2, (e) VH_3, and (f) VH_4.

positions were achieved. In all cases, the symmetry was fully constrained to that of the ideal point group of the defect. The resultant double derivatives were then calculated. The scaled Kohn-Sham energy levels for all of the VH_n complexes are shown in figure 8.2.

VH

After relaxation, H make a strong bond of length 1.493 Å with Ge, resulting in a high local vibration mode at 2122 cm^{-1}, which shifts downwards to 1509 cm^{-1} with D. All local modes of the neutral defect along with their isotope shifts are presented in Table 8.1. It must be pointed out that the symmetry was constrained to C_{3v} in

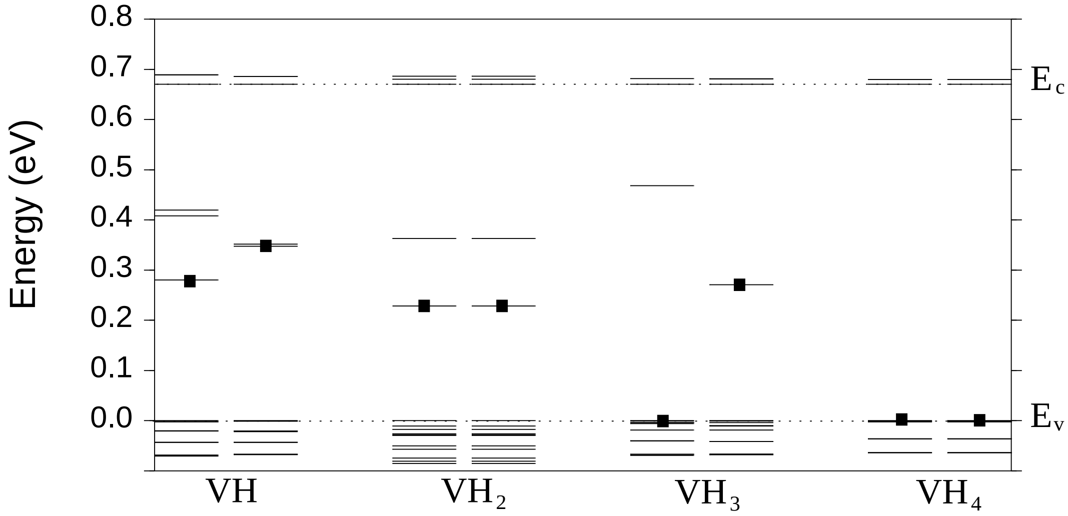


Figure 8.2: The Kohn-Sham levels for the neutral VH_n complexes in Ge. Note that the levels have been scaled to the Ge band-gap of 0.67 eV.

Table 8.1: LVMs, cm^{-1}, of the VH defect in Germanium.

Ge:H	Ge:D
2122	1509
657	469
657	469

this calculation. It turns out that this is a somewhat unphysical constraint as far as the lowest energy configuration is concerned, for the following reason: referring to the Kohn-Sham levels for the defect, Figure 8.2, as expected in the neutral charge state, the highest occupied level in the gap is one with e symmetry. The neutral ground state is paramagnetic with $S = \frac{1}{2}$. The energy can therefore be lowered by a Jahn-Teller distortion, which result in the lowering of the defect symmetry to C_{1h}, and the splitting of the e-level. The distortion is likely to be modest (for example H moves 5° of the $\langle 111 \rangle$ axis in Si:VH, according to theory [59]) and therefore our calculated stretch mode would be largely insensitive to the distortion, which would lead to a small splitting of the wag modes. As a result of the distortion, the two atoms either side of the mirror plane move towards each other by a small amount, forming a weak bond. The driving force for this distortion is lost in VH$^+$ which possesses C_{3v} symmetry.

VH_2

The addition of the second hydrogen atom results in a neutral centre possessing C_{2v} symmetry. The Ge-H bonds are slightly longer at 1.513 Å than they were in VH, and the defect now gives rise to two stretch modes at 2209, and 2147 cm^{-1} having A and B symmetry respectively. Both of these are higher than the stretch mode observed for VH. The isotope shifts of these, and the wag modes are fully detailed in table 8.2. The lowering of the symmetry results in two widely split singlet levels

Table 8.2: Calculated LVMs, cm^{-1}, of the VH_2 defect in Germanium.

Ge:HH	Ge:HD	Ge:DD
2209	2176	1567
2147	1547	1528
704	655	501
646	644	462
641	461	460
578	443	415

in the band gap, the lower of which is fully occupied, and the highest is unoccupied (figure 8.2). The un-occupied level is associated with the dilated bond formed between the two under-coordinated atoms of the vacancy, which attract each other in a similar manner as in the VH complex.

VH_3

The VH_3 defect possesses trigonal symmetry, and Ge-H bonds of length 1.513 Å are formed. The neutral defect gives rise to a singly occupied a_1 level in the band-gap, and hence is paramagnetic with $S = \frac{1}{2}$. All local modes of the centre are given in table 8.3. In keeping with the trends of increasing stretch frequency with increasing n, VH_3 has its highest modes at 2283 and 2171, which possess A and E symmetry respectively.

VH_4

Finally, the vacancy can be completely passivated by the addition of a four hydrogen atoms, saturating all of the dangling bonds. The neutral defect has perfect tetrahedral symmetry, with four Ge-H bonds of length 1.506 Å. As a result four stretch modes are found at 2454 and 2282 (T_2) cm^{-1}, and again the frequencies have increased with n.

Table 8.3: Calculated LVMs, cm^{-1}, of the VH_3 defect in Germanium.

Ge:HHH	Ge:HHD	Ge:HDD	Ge:DDD
2283	2248	2211	1622
2171	2171	1593	1545
2171	1568	1545	1545
745	740	706	529
745	681	636	529
669	646	521	480
617	593	462	441
589	477	451	422
589	440	424	422

Table 8.4: Calculated LVMs, cm^{-1}, of the VH_4 defect in Germanium.

Ge:HHHH	Ge:HHHD	Ge:HHDD	Ge:HDDD	Ge:DDDD
2454	2416	2376	2331	1743
2282	2282	2282	1706	1623
2282	2281	1675	1623	1623
2282	1648	1624	1623	1623
806	782	765	721	571
806	782	742	721	571
709	709	693	540	505
709	674	656	540	505
709	674	531	505	505
631	631	488	469	452
631	487	487	469	452
631	487	461	452	452

8.3.2 Self-interstitial-hydrogen complexes

The isolated self-interstitial in germanium

To our knowledge, the self-interstitial in germanium has not been investigated theoretically previously. Here, we focus on the $\langle 100 \rangle$ and $\langle 110 \rangle$ split-interstitial configurations only. For illustrations of the complexes, as well as the possible configurations formed when the defect captures one or two hydrogen atoms, the reader is referred to chapter 5. All atoms were relaxed with the symmetry constrained to the point group of the initial configuration.

As was the case in Si, the most stable configuration for the defect in Ge was found to be a $\langle 110 \rangle$ split-interstitial, with C_{2v} symmetry, where two Ge atoms share a single lattice site oriented along this crystallographic direction. The optimised coordinates of the inner six atoms are given in Table 8.5. The Kohn-Sham energy levels are given in Figure 8.3. They show that the defect has two almost degenerate fully occupied deep donor levels, and unoccupied acceptor levels but the theory is not able to determine these levels accurately. Again, the C_2 $\langle 100 \rangle$ configuration proved unstable relaxing to the $\langle 110 \rangle$ ground state, and although the C_{1h} $\langle 100 \rangle$ structure was stable, it was a local minima at 0.51 eV above the $\langle 110 \rangle$ form.

Table 8.5: Coordinates for the isolated split-interstitial defects in in Ge, relative to cube unit cell of length 5.66 Å respectively.

Ge I	$\langle 100 \rangle$			$\langle 110 \rangle$		
	x	y	z	x	y	z
Ge_1	0.133	0.006	0.006	0.157	0.157	-0.134
Ge_2	-0.222	0.130	0.130	-0.157	-0.157	-0.134
NN1	-0.288	0.260	-0.269	-0.265	0.265	-0.269
NN2	-0.288	-0.269	0.260	-0.263	-0.263	0.263
NN3	0.283	-0.268	-0.268	0.265	-0.265	-0.269
NN4	0.314	0.270	0.270	0.263	0.263	0.263

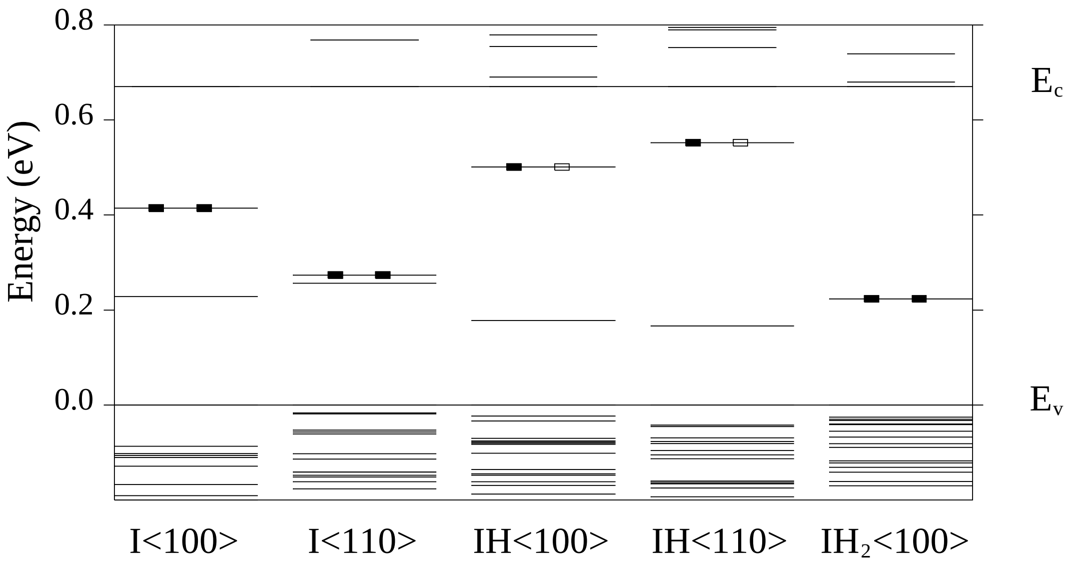


Figure 8.3: The Kohn-Sham levels for the neutral IH_n complexes in Ge. Note that the levels have been scaled to the Ge band-gap of 0.67 eV.

The IH defect in germanium

One hydrogen atom was then attached to one of the dangling bonds of the split-interstitial defect, forming a defect with C_{1h} symmetry for both the $\langle 100 \rangle$ and $\langle 110 \rangle$

derived configurations. All atoms were then relaxed with this symmetry constraint, and the optimised positions are given in Table 8.6. These structures are again

Table 8.6: Coordinates for the IH defects in in Ge, relative to cube unit cell of length 5.66 Å respectively.

Ge IH		⟨100⟩			⟨110⟩	
	x	y	z	x	y	z
Ge_1	0.203	0.043	-0.043	0.122	0.122	-0.054
Ge_2	-0.183	0.007	-0.007	-0.176	-0.176	-0.136
NN1	-0.304	0.278	-0.278	-0.253	0.249	-0.252
NN2	-0.284	-0.268	0.268	-0.259	-0.259	0.257
NN3	0.299	-0.278	-0.270	0.249	-0.253	-0.252
NN4	0.299	0.270	0.278	0.282	0.282	0.280
H	0.225	0.233	-0.233	0.252	0.252	-0.249

very similar to those derived for the IH defect in silicon, however there are slight differences in the calculated total energies - here the ground state is degenerate. The calculated local vibrational modes and isotope shifts of both configurations are given in Table 8.7. One result is immediately clear: if the magnitude of this calculated stretch mode in either of the IH configurations is compared to those of VH_n, it is found that the IH modes always lie *below* those of the VH_n complexes. This is consistent with the fact that in the IH complex, the bonding with H primarily involves p-orbitals, whereas s-orbitals are involved in the VH_n centres. To date, there has been no reports of these modes experimentally. The electronic structure

Table 8.7: Local Vibrational Modes, cm^{-1}, of IH defect.

Ge ⟨100⟩	H	D	⟨110⟩	H	D
A	2079.7	1478.9	A	2073.7	1474.8
A	734.5	521.0	A	800.5	567.3
B	729.9	517.8	B	762.2	541.4

of the neutral defect is shown in figure 8.3. These reveal that the IH defects are paramagnetic in the neutral charge state and possess a deep singly occupied level.

The IH_2 defect in germanium

Finally, we consider the IH_2 defect in germanium. The ⟨100⟩ configuration possessing C_2 symmetry was the lowest energy configuration, lying 2.18 eV below the

$\langle 100 \rangle$ C_{2v} configuration. If the symmetry constraint was removed from the later structure, it relaxed towards the ground state. The atomic positions for the IH_2 defect are given in Table 8.8. From this structure, the polar angle θ was determined

Table 8.8: Coordinates for the IH_2 defect in in Ge, relative to cube unit cell of length 5.66 Å respectively.

Ge IH_2		$\langle 110 \rangle$	
	x	y	z
Ge_1	0.195	0.049	-0.037
Ge_2	-0.195	0.049	0.037
NN1	-0.296	-0.277	0.269
NN2	0.296	-0.277	-0.269
NN3	-0.315	0.277	-0.284
NN4	0.315	0.277	0.284
H_1	0.228	0.236	-0.229
H_2	-0.228	0.236	0.229

to be 46°C, with the azimuthal angle ϕ being 8°C. These values are in reasonable agreement with the experimental results of 45±1° and 0° for θ and ϕ respectively. The calculated local modes of the IH_2 defect add further weight to the assignment

Table 8.9: Local Vibrational Modes, cm^{-1}, of IH_2 defects.

Ge	H-H	H-D	D-D
A	2056.7	2055.1	1462.5
B	2053.3	1461.3	1460.1
B	787.4	784.9	558.8
A	784.7	555.9	555.7
A	725.0	712.4	514.7
B	694.6	503.7	493.1
observed			
A	1883.5	1882.8	1359.0
B	1881.8	1358.4	1357.6
	705.5	703.0	
	700.3		

of the four observed modes to this defect. These are shown together with the experimental results in Table 8.9. The calculated local modes and isotope shifts are in good agreement with the observed values, accepting the overbinding of the H to Ge, again the stretch modes are consistently below those of the VH_n complexes. The anomalous electrical activity of the IH_2 centre is again present in germanium, the defect giving rise to a fully occupied level deep in the band gap. The origin of this level is still unclear.

8.4 Substitutional and interstitial carbon defects in germanium

8.4.1 Experimental results

At first glance one would expect simple defects like C_s to be present in Ge due to the similar chemical properties with Si. The dissociation energy of the C-Ge bond is high $\approx$4.7 eV [133], compared to the 2.8 eV that is required to break the Ge-Ge bond. This increased bond strength favours the formation of C-Ge bonds, however the induced strain does not tend to favour the formation of a Ge-C bond in the bulk. This is because the ideal Ge-C bond length in the simple molecule CH_3GeH_3 is around 1.9 Å- much shorter than the Ge-Ge bulk bond length of 2.41 eV. Therefore C_s would introduce a large strain field in Ge. The balance between these competing effects is crucial, and it may be that a more complex defect, like an interstitial carbon is formed to relieve the strain on the defect. In fact the solubility of C in Ge is very low, around 10^8-10^{10}cm^{-3} [134], eight orders of magnitude lower than in Si, indicating that strain in playing an important role.

The C_s defect has only recently been identified in Ge [135] after carbon was ion implanted. Following the implantation with ^{12}C, and annealing at 450°C, an absorption line was observed at 531 cm^{-1}. This line shifts down to 512 cm^{-1} when ^{13}C is implanted, the ratio of these frequencies being very close to $\sqrt{13/12}$ expected for simple arguments. No new modes appear in the mixed ^{12}C, ^{13}C implanted case, implying that the mode originates from a defect containing one C atom. Ion channelling experiments implied that only 31$\pm$3% of the carbon atoms were located at substitutional sites, the rest of which were located randomly - possibly as precipitates although no modes were reported in the region 1200-1500cm^{-1}. Isochronal annealing studies of the material showed that the C_s mode begins to appear at 350°C, and reaches its maximum intensity at 450°C. The intensity decrease with anneals at higher temperature until the peak is eventually lost at 700°C. The loss of the 531 cm^{-1} line correlates with the depopulation of the substitutional site measured in the channelling studies. This serves to confirm that the mode results for a single substitutional carbon atom. The concentration of C_s in theses studies was 10^{20} cm^{-3}, orders of magnitude above any other impurity. The value of η for the 531 cm^{-1} was (3.4$\pm$0.5)e, significantly higher than the value obtained in silicon.

8.4.2 Aimpro results

The 71 and 131 atom clusters yielded almost identical structures, with the C-Ge bonds being within 1.3% of each other. Consequently, the results of the 131 atom cluster are discussed here, the calculated local modes of the 71 atom cluster are given in Table 8.10. As with substitutional carbon in silicon, C_s in Ge exerts a large tensile strain on the surrounding atoms. The bulk Ge-Ge bond length is 2.407 Å and carbon makes C-Ge bonds of length 2.046 Å, a 18% reduction from the bulk value, and a larger percentage difference than was the case in Si. As a result of this, the Ge-Ge back bonds are increased by 1.6% to 2.447 Å. The reduced bond length and the lighter defect atom, mean that the highest local vibrational mode occurs at 516 cm^{-1} in comparison to the experimental value of 530 cm^{-1}. This triply degenerate localised mode involved the vibration of the C atom, and to a very much lesser extent, its four Ge neighbours. Below this mode was the Raman peak, there were no increases in LVMs due to Ge around the Raman edge (304 cm^{-1}) as the strain due to carbon was tensile. The calculated local mode drops by 19.3 and 36.4 cm^{-1} with ^{13}C and ^{14}C respectively, and compares extremely well with the experimental reduction of 19 cm^{-1} with ^{13}C. The experimental and theoretical data concerning the local vibrational modes is given in Table 1. The Kohn-Sham eigenvalues indicated that the neutral defect was electrically inactive consistent with the complete coordination of all atoms.

Table 8.10: Calculated bond lengths (in Å) and local vibrational mode frequencies (cm^{-1} for C_s in germanium. The experimental local mode frequencies are given for comparison.

	Ge-C	Ge-Ge	^{12}C	^{13}C	^{14}C
Experiment [135]			530	511	—
Theory (71 atom cluster)	2.018	2.415	563	542	524
Theory (131 atom cluster)	2.046	2.447	516	497	480

Calculations were also performed on the interstitial carbon defect, (C_i), in germanium. Both 72 and 132 atom clusters were used and the structures for the ⟨100⟩ split-interstitial were as follows: in the smaller cluster, C-Ge bonds of length 1.766, and 1.899 (2) Å were made, and these lengths were modified slightly to 1.801, and 1.822 Å in the larger cluster. The C_{3v} bond centred configuration was also investigated in the larger cluster, and proved to be much higher >3 eV than the C_{2v} structure. The migration of C_i in Ge is therefore expected to proceed by a different pathway than in Si, and the barrier for this process may be significantly higher in germanium. The calculated local modes of the C_i defect occur at 842.8 (A), and 736.6 (B) cm^{-1}, and shift downwards by 30.4, and 27.1 cm^{-1} respectively with ^{13}C.

The neutral defect gave rise to two distinct deep gap levels, the lower of which is fully occupied, and the highest empty.

8.5 Substitutional carbon in $Si_{(1-x)}Ge_x$ alloys.

Currently, there is a great deal of interest in C, Si, and Ge alloys, due to the potential for band-gap engineering, and lattice matching to Si. In particular much attention has been payed to $Si_{(1-x)}Ge_x$ alloys doped with high concentrations of carbon.

In contrast to $Si_{(1-x)}Ge_x$ [136], $Si_{(1-x-y)}Ge_xC_y$ layers of arbitrary thickness, and low dislocation density can be grown on the pure Si (100) surface [137, 138, 139], although the the temperature and fraction (x) window in which this growth can be achieved is small. The reason for this is that addition of a small amount of carbon into the material can compensate for the 4% lattice mismatch between Si and Ge. Therefore thicker layers of the alloy with low defect concentrations may be grown on Si. This therefore make C_s and very interesting and important defect in $Si_{(1-x)}Ge_x$ alloys. Although the work in the previous section demonstrated that C_s can be formed in Ge, it is unclear whether C-Si or C-Ge bonds are favoured in the alloy. Evidence was found in [139] of a repulsive interaction between C and Ge, which lead to an absence of C-Ge bonds for y <0.2, implying that C-Si bonds were preferentially formed.

8.5.1 Experimental background

When carbon is implanted into relaxed epitaxial MBE-grown $Si_{(1-x)}Ge_x$ alloys the position of the local vibrational modes due to C_s depend on the fraction x. Clearly when x=0, the mode occurs at 607 cm^{-1} and decreases with increasing x. At x=1, the absorption is observed at 531 cm^{-1}, with all intense modes, and the majority of others being bound by those of the pure crystalline materials. For example, in $Si_{0.65}Ge_{0.35}$, two broad absorption lines occur at 551 and 592 cm^{-1}, measurements in materials co-implanted with ^{12}C and ^{13}C revealed that these were local modes of defects containing one carbon atom. The experimental results were consistent with C binding to both Si and Ge in the alloy. The total intensities of the various lines remains constant with annealing temperatures up to 900°C, however the peak heights of the various modes in the alloy do change during the anneals. The T_2 mode of C_s in Si broadens and shifts downwards as the Ge concentration is increased from 0-50%. Additional modes begin to appear around 550 cm^{-1} when the Ge fraction is around 0.15. For example, in ^{13}C implanted material, there is a mode at ~542 cm^{-1} (x=0.15), which decreases to ~525 cm^{-1} when x=0.50. Lower modes also exist at

$\sim$522 and $\sim$514 cm^{-1} for x=0.35 and 0.5 respectively (^{13}C). The observed modes for the various fractions are shown in Figure 8.4.

The case of $Si_{0.85}Ge_{0.15}$ was considered for a qualitative understanding of the local mode structure. Here, it was suggested that two defects prevail C_s-Si_4 and C_s-$GeSi_3$. The first of these gives rise to a triply degenerate T_2 mode, whilst the later with C_{3v} symmetry will give rise to a doubly degenerate E mode resulting from Si-C motion, and an A mode predominantly involving the Ge and C atom. Here it is assumed that the surroundings are isotropic (tetrahedral symmetry). Therefore, in ^{13}C, the 542 cm^{-1} (low intensity), 582 cm^{-1} (high intensity), and the $\sim$600 cm^{-1} (low intensity), modes were assigned to A, T_2 and E symmetry modes respectively. The intensity of the A and E modes, as well as other modes from C-Ge_2Si_2, C-Ge_3Si, and C-Ge_4 are expected to increase as the Ge fraction increases. When C has two Si and two Ge neighbours, the defect will have C_{2v} symmetry and will give rise to two B modes, and one A mode. Therefore the downward shift in the local vibrational mode frequencies can be qualitatively understood in terms of the increasing lattice constant as the Ge fraction is increased from 0-50%.

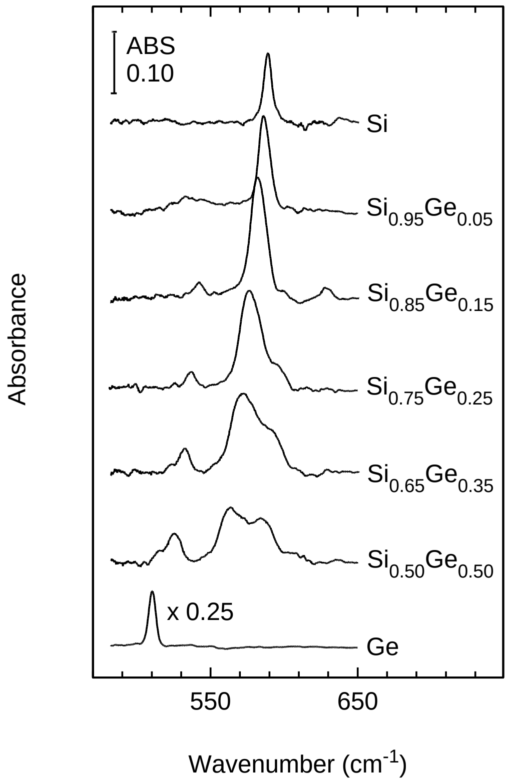


Figure 8.4: Absorbance spectra measured on ^{13}C implanted $Si_{(1-x)}Ge_x$ alloys, ($0 \leq x < 1$) after annealing at 700°C. The absorbance from the Ge sample was measured after annealing at 550°C [140].

8.5.2 Aimpro results

The approach adopted in the modelling of C_s was as follows: the substitutional carbon defect was first placed in the centre of the 71 atom tetrahedral cluster making $CSi_{35}H_{35}$, and the optimised positions and local vibrational modes found. The Si neighbours of C were then replaced one at a time by a Ge atom, and the structure re-optimised and the new local modes found. This therefore gave us the C-Si_4, C-Si_3Ge, C-Si_2Ge_2, C-SiGe_3, and C-Ge_4 local modes in a bulk *silicon* cluster. The above process was then carried out in a 71 atom germanium cluster $CGe_{35}H_{35}$, with the nearest neighbours being replaced in turn by silicon atoms, to obtain the above modes in bulk *germanium*.

Before describing the results in detail, the approach used is briefly outlined:

- The calculated local modes of C_s in Si with 0-4 Ge neighbours are scaled such that when C_s has four Si neighbours, the local mode is at 607 cm^{-1} coincident with the experimental result.
- The LVM of C_s with four germanium neighbours, calculated in the silicon cluster, is higher than the mode for C_s in pure Ge, due to the compressional effect of the silicon host. To correct for this, the modes are multiplied by a factor of $[1 - 0.05x]$, which scale the modes with increasing lattice constant.
- Evidence exists that C preferentially binds to Si in the alloy, this is also taken into account in the analysis, by weighting the probabilities that C has either a Si or Ge neighbour in the appropriate way.
- The calculated local modes were then broadened by Gaussians of set width to assist the comparison between theory and experiment. A linear scaling factor is also used so that the pure C_s mode in Ge appears four times as intense as the corresponding mode in Si.

The optimised bond lengths are given in Table 8.11 below. Refer to Figure 8.5 for the bond labelling used in the Table. It is clear from the table, that the C-Ge bond is longer than the Si-C bond. Also, the bonds are somewhat shorter in the pure silicon cluster when compared to the corresponding bonds in the pure germanium cluster. This is a result of the smaller lattice constant of Si (2.349 Å) compared to that of Ge (2.407 Å).

The local vibrational modes of the C_s defect were then calculated in both of the above clusters, and the results are given in Table 8.12. Also given, are the local modes from the bulk silicon cluster after they have been scaled linearly so that the

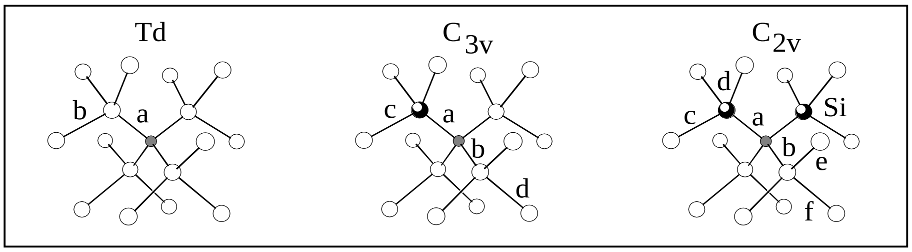


Figure 8.5: Bond labelling for the C_s defect in a T_d, C_{3v} and C_{2v} environment.

Table 8.11: The bond lengths (Å) for C_s in bulk Si and Ge, with varying numbers of silicon and germanium nearest neighbours. Note that the Si-C bonds are always shorter than the Ge-C bonds. Refer to Figure 8.5 for the bond labelling.

		a	b	c	d	e	f
Si							
C-Si_4	T_d	1.962	2.334				
C-Si_3Ge	C_{3v}	2.043	1.954	2.366	2.334		
C-Si_2Ge_2	C_{2v}	1.947	2.033	2.331	2.333	2.363	2.364
C-SiGe_3	C_{3v}	1.938	2.027	2.331	2.363		
C-Ge_4	T_d	2.018	2.363				
Ge							
C-Si_4	T_d	1.977	2.374				
C-Si_3Ge	C_{3v}	2.065	1.966	2.409	2.373		
C-Si_2Ge_2	C_{2v}	1.957	2.053	2.370	2.370	2.406	2.406
C-SiGe_3	C_{3v}	1.951	2.044	2.369	2.405		
C-Ge_4	T_d	2.035	2.403				

C-Si_4 mode is equal to the experimental result of 607 cm^{-1}. The calculated local modes of each C-$Si_m Ge_n$ ($m \in \{1, 2, 3, 4\}, n = 1 - m$) are consistently higher in the silicon cluster, a result of the shortening of the bonds due to the larger compressive strain on the defect. Note here that the calculations for the local modes for C_s in the pure materials is 662 cm^{-1} for Si and 555 cm^{-1} for Ge. The calculations also suggest that modes outside the region 531-607 cm^{-1} may be present, since there are many modes outside the calculated range of 555-662 cm^{-1}.

The fact that the lattice constant is changing with x complicates matters. For example if $x \approx 0$, then the results from the Si cluster must be considered, conversely if $x \approx 0$ the local modes should be taken from the germanium cluster. A weighted average is necessary when relatively large amounts of both materials are present. The scaled Si values are now considered in the analysis of the local modes; and to incorporate the lattice constant effects, these values are scaled by a factor of

Table 8.12: Calculated local vibrational for modes (cm^{-1}) for C_s in bulk Si and Ge, with varying numbers of silicon and germanium nearest neighbours. In the final column, the results from bulk Si are rescaled to yield the experimental value of the T_2 mode of C_s in Si.

# Si neighbours	Bulk Si cluster	Bulk Ge cluster	Scaled Si
4	662 662 662	638 638 638	607 607 607
3	612 673 673	588 651 651	561 617 617
2	594 642 684	568 617 663	545 589 627
1	598 598 675	573 573 649	548 548 619
0	605 605 605	579 579 579	555 555 555

$[1 - 0.05x]$, so that the T_2 mode appear at 607 cm^{-1} when $x = 1$, and 528 cm^{-1} when $x = 0$. Calculations of the induced dipoles of all of the above modes indicated that they are all of equal intensity, therefore only the changes in the refractive index of the alloy with varying x will affect the peak intensity.

Assuming that both Si and Ge are equally likely to bond to carbon, then the probability distribution for n germanium neighbours, and m silicon neighbours, for concentrations x is given by:

$$p(n, m) = x^n(1 - x)^m \frac{4!}{n!m!}, \tag{8.1}$$

the resultant profile of the variation of $p(n, m)$ with x is shown in Figure 8.6a. The expected absorption profile is shown in Figure 8.6b, here Gaussians are placed over our calculated frequencies to mimic the experimental broadening. In addition to this, the peak height is increased linearly from 1-4 as the fraction x varies from 1-0. Clearly, our results differ significantly from those observed in [140].

Straight away, it can be seen that the C-Si_4 T_2 mode fades away too quickly with x fraction, when compared to the experimental results. The change in peak height with anneals can be explained if configurations with higher number of Si neighbours are favourable. In fact the experimental data seem to suggest this; there is a monotonic decrease of the T_2 mode due to C-Si_4 as x is increased from 0.0 to 0.5, and even at the latter value the T_2 peak remains the dominant feature.

To check this, we first of all calculated the difference in ground state energy of two $CGeSi_{69}H_{60}$ clusters when a Ge-C bond is formed, and when the Ge is in a Si-Si back-bond. The calculation clearly favoured the formation of Si-C bonds, the cluster with a Ge-C bond present being the higher in energy by 0.20 eV. To further check this, the same calculation was performed in a two $CSiGe_{69}H_{60}$ clusters, and again, there was a saving in energy of 0.25 eV when the C-Si bond was formed. Clearly then the probabilities should be weighted to this effect.

Using an energy cost ΔE, the weightings can be achieved by first calculating

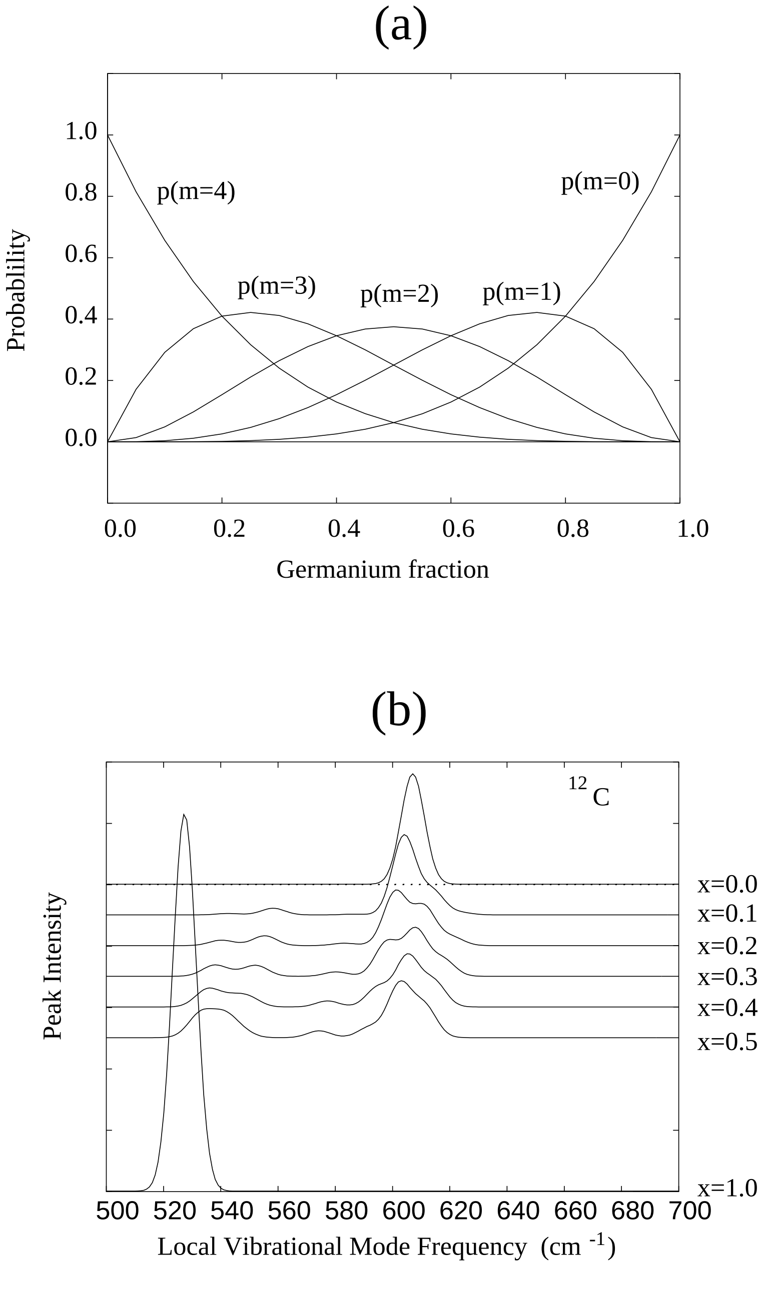


Figure 8.6: The variation with Si fraction x of (a) the probability distribution $p(n, m)$, and (b) the positions and intensities of the various local modes.

the probabilities as in equation 8.1, and then modifying the probabilities by a given exponential factor. Defining $p(n,m)$ as the un-altered probabilities, and $p'(n,m)$ as the weighted ones, then the latter can be given by:

$$p'(n,m) = p(n,m)\exp[\alpha n \Delta E], \tag{8.2}$$

where $\alpha = 1/kT$, and then re-normalising,

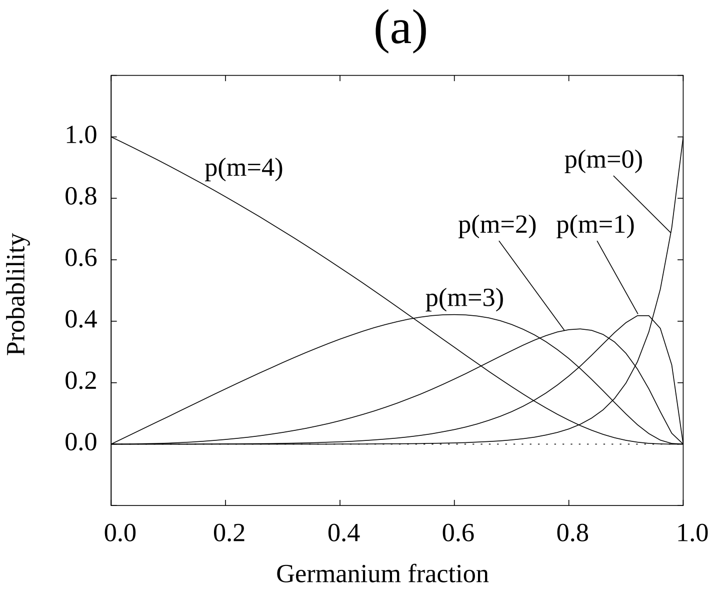
$$p'(n,m) = \frac{p'(n,m)}{\sum p'(n,m)}. \tag{8.3}$$

Using this distribution, the modified probability functions, and calculated absorbance was found for $\alpha\Delta E$=1.5. The results are shown in Figure 8.7, the probability profile in is given in (a), and the absorption profile is in shown in (b).

Using the basic assumption that C-Si bonds are preferentially formed, and accounting for the changes in the lattice constant, the trends in our results are now in good agreement with the experimental data presented in [140]. For the $Si_{0.65}Ge_{0.35}$ alloy (^{13}C), our calculations give a large peak at 596 cm^{-1} and a smaller peak at 545 cm^{-1} in fair agreement with the broad peaks at 591 and 551 cm^{-1} respectively. However, the absolute values of our modes are only in fair agreement with experiment, a direct comparison between the calculations and the experiment is given in Figures 8.8 and 8.9. For example, the splitting of the two intense peaks around 580 cm^{-1} with x=0.5, is only half of the experimental observed one. Clearly in the experiment the C_s-Si_4 mode is decreasing more rapidly than our calculations predict. Nevertheless, these preliminary results are encouraging.

8.6 Discussion and conclusions

In the first half of this chapter, the interaction between hydrogen and the intrinsic defects (V and I) was considered in germanium. The results are essentially the same as in silicon, and can be summarised as follows: the VH_n defects give rise to high stretch modes in the region 2122-2454 cm^{-1}, as well as wag modes in the region 578-806 cm^{-1}. These stretch modes increase with increasing n, a trend which was found both experimentally and theoretically for the vacancy-hydrogen complexes in silicon. The corresponding stretch modes in the interstitial-hydrogen complexes occurred at lower frequencies, 2050-2080 cm^{-1} indicating the $p-$ orbital nature of the bonding, and again this trend was observed in the corresponding defects in Si. The wag modes of the IH_n complexes were in the range 690-740 cm^{-1}. The calculated LVMs of the IH_2 complex support the recent assignment of four H-related local modes at 1883.5, 1881.8, 705.5, and 700.3 cm^{-1} in proton implanted material to this complex. It must be pointed out that the overbinding of H to Ge in the



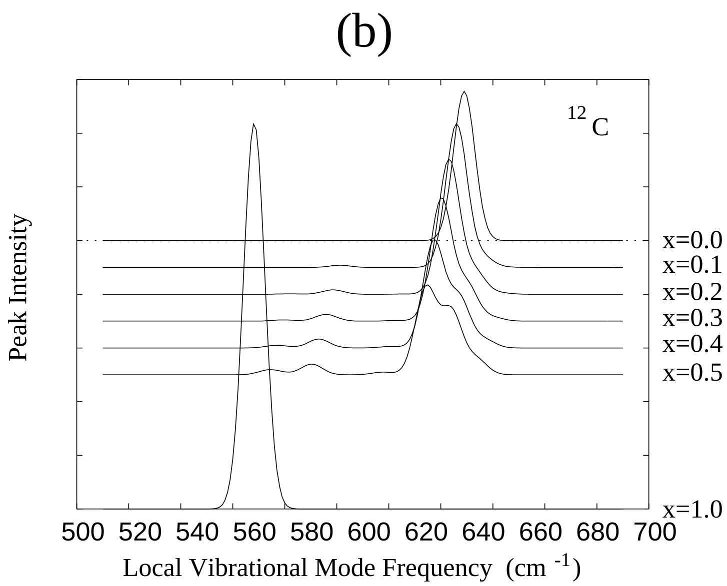


Figure 8.7: The variation with Si fraction x, using $\alpha\Delta E$=1.5, of (a) the probability distribution $p'(n,m)$, and (b) the positions and intensities of the various local modes.

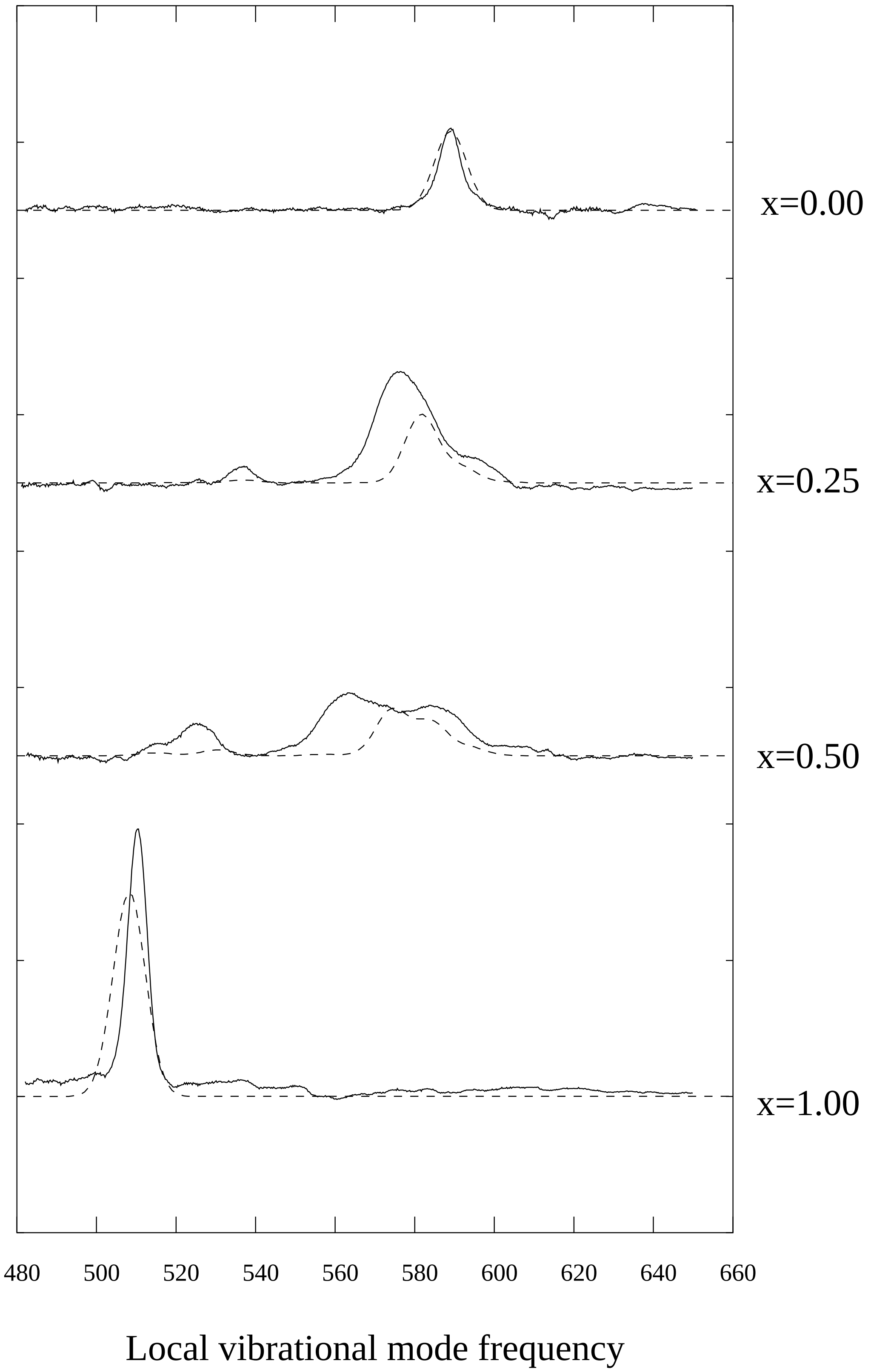


Figure 8.8: Comparison between the theory, and the experimental local modes of C_s (^{13}C) in $Si_{(1-x)}Ge_x$ for x=0.0, 0.5, 0.75 and 1.0. The solid and dashed lines represent the experimental and theoretical results respectively.

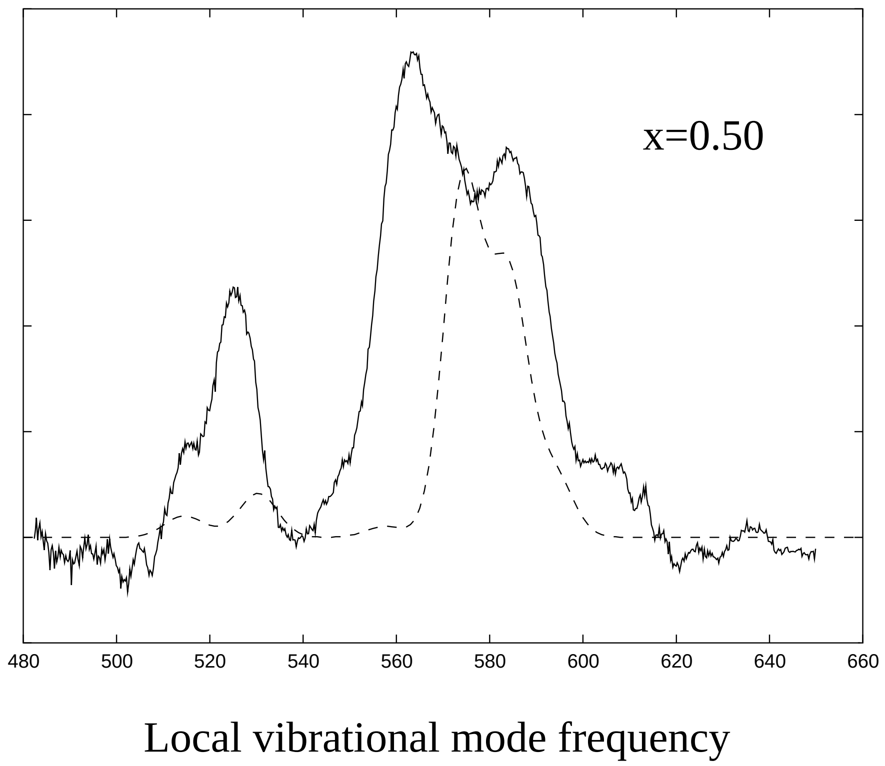


Figure 8.9: Comparison between the theory, and the experimental local modes of C_s (^{13}C) in $Si_{(1-x)}Ge_x$ for x=0.5. The solid and dashed lines represent the experimental and theoretical results respectively. Here, it is apparent that all experimental modes are accounted for, although their exact positions and intensities are given incorrectly.

calculations results in typical errors of up to 7% (the value being too high) for the H related modes. The electronic levels of the VH_n complexes are very easy to understand, the effect of adding each H is to populate one of the vacancy levels with an electron, lowering its position in the gap. The addition of the fourth H fully passivates the complex. More difficult to understand are the levels of the IH_n complexes where, when n=2, the Kohn-Sham eigenvalues indicate that the defect is not fully passivated.

Simple carbon defects in Ge and $Si_{(1-x)}Ge_x$ were also considered. In bulk germanium, the T_2 local mode of the C_s defect was found to be in the range 516-563 cm^{-1}, in good agreement with the experimental value of 531 cm^{-1}. The calculated ^{13}C isotope shifts of this defect were also in extremely good agreement with the experimental results. The local modes of the C_i defect has been undetected so far in Ge, but calculations on its structure and local modes have been performed for future reference. The defect gives rise to two high local modes at 842.8 (A), and 736.6 (B) cm^{-1}, far lower than the corresponding modes in Si. Substitutional carbon was also investigated in the silicon-germanium alloys, and here we find that carbon preferentially bonds to silicon. The variation in the calculated local modes of C_s in the alloy with x is fairly well reproduced when various factors such as the preferential binding and the changes in crystal volume are accounted for. In agreement with experiment, the calculations show that no local modes are expected outside the regions bound by C_s in bulk Si and Ge, 531-607 cm^{-1}.

Chapter 9

Conclusions

The results of this thesis are summarised in detail at the end of chapters 4-7, here, further comments are made; in particular the relative merits and frailties of the method as applied to each problem is discussed. Suggestions for further work, both theoretically and experimentally are also outlined.

Chapter 5 considered the application of the method to the very important self-interstitial defect in silicon, and in particular, the total energies of various configurations of the Si_i defect were calculated. Although we did not explicitly calculate the migration pathway(s) and barrier of the defect, comments were made on these, as well as the possibility of enhancement of the migration through changes in the defect charge state. Such calculations are still beyond *ab initio* techniques, but in time these questions should be addressed; nevertheless, the results presented here can account for all of the intriguing properties of the defect. A calculation of the total energies of the various configurations of the neutral and charged Si_i defects using the recently developed supercell version of AIMPRO could also prove worthy. Here an investigation into the relative importance of defect-surface and defect-defect interactions in the clusters and supercells respectively could be performed; and the subsequent convergence of the energies would increase ones confidence of the accuracy of any calculation. It is inevitable though that less computationally demanding methods should be used for calculations on large scale Si_i clusters since these will remain impossible using *ab initio* techniques for the foreseeable future. However, these simplistic models have proved unreliable in the past, especially in cases of unusual bonding (present in the {311} defect. It seems certain that the supercell AIMPRO code has a future role in the studies of extended defects like the {311} interstitial aggregate. Our calculations are in broad agreement with the majority of other theoretical work in this area, the role of theory is particularly important in the study of this defect as it still eludes the experimentalist.

The calculated properties of the IH_2 defect agree extremely well with the experimental assignment of the two local modes in the region 1900-2000 cm^{-1} to this defect, leading to the unambiguous identification of this centre for the first time.

Carbon is also a very common and important impurity in silicon, and a variety of complexes are formed when Si_i complexes with C_s, followed by subsequent thermal annealing at various temperatures. In general, these centres are well understood, with structural models proposed from the vast amount of experimental data gathered on these defects. A few problems did exist with regards to the accepted models, and it is these which were addressed in chapter 6. The isolated C_i defect possessed two close local modes whose 2:1 IR intensity ratio could not be explained using the standard model. The closeness of the modes is relatively east to understand, since the four atoms of the defect core resembles a D_{3d} centre which would have a degenerate mode - this splits for the lower symmetry. The calculations found this, and furthermore, the induced dipoles which these modes were significantly different, and satisfactorily accounted for the observed intensity ratio. These results highlighted the value of the code when applied to a wealth of different properties probed by experiment. As well as these optical properties, the Kohn-Sham eigenvalues and vectors describe the electronic properties of the defect extremely well. Two levels were found in the gap, and the localisation of the pseudowavefunctions was in excellent agreement with the results of EPR. The calculated position of these levels was in fair agreement with experiment when the Slater transition argument was invoked. Indeed, the recent results of Resende *et al* have shown that the *position* of the gap levels of transition metals in Si can be calculated with good accuracy using such an approach. Given the importance of the exact positioning of a defects gap level(s), and to date, the lack of an accurate theoretical model to describe them - the development of a predictive tool in this area is of great benefit. The di-carbon centre is a well documented case of a bistable centre in silicon, existing in two configurations which depended on the defect charge state. Although the calculated total energies from the cluster code did not find the low energy structure switching with charge state, the trends in the relative energies of configurations with varying charge state were well reproduced. The long standing problem concerning the isotope shifts of the local modes of the neutral defect was seemingly resolved, however recent experimental results have cast further doubt in this area. Clearly, further calculations now need to be performed, and more experimental work is necessary on the G-line. The use of very large clusters or better still, supercells could well yield more accurate energies. Regardless of this many of the carbon centres are extremely well understood, and the calculations presented further support the assignments.

Attention was then turned to the interaction between hydrogen and carbon related complexes in silicon, and the results for these centres was presented in chapter 7. Experimentally, it is observed that the C_s-H^0 defect spontaneously dissociates when the positive defect captures an electron. This suggestion is in disagreement with our calculations, as well as previous ones in which a considerable binding energy is found. The apparent disagreement can however be explained by the formation of H_2 molecules in the neutral charge state resulting in the loss of energy, a reaction path not amenable to the charged defects. The local vibrational modes of C_s-H_n $\{n \in 1, 2\}$ were also presented, and these are currently being sought experimentally. The chemistry of C_i-H defects in Si is much richer than the immobile C_s-H, which also had low thermal stability. Our work is the first theoretical investigation of this and related centres, and a variety of interesting phenomena are predicted for the defect. Most notably is the bistable nature of the centre, and the degenerate ground state configurations through which a long range migration pathway can be constructed leading to an extremely low barrier of 0.3 eV. As a result the C_iH unit is likely to be highly reactive and will rapidly migrate to and complex with other defects and impurities present. Investigations were then made into the trapping of the defect to other defects, and in conjunction with an experimental work, identified the ground state structure of the T-centre defect as a $\langle 100 \rangle$ oriented split-interstitial. In-fact, a large number of other highly stable photoluminescence centres are known to be CH related and more work needs to be done to determine the microscopic configurations of these centres. Some other C_i-H complexes were considered but we mainly focused on the interaction with C_s and the passivation of the above centres with a second hydrogen atom. These CH defects are all highly stable, being created and destroyed in the temperature ranges 300-450°C and around 600°C respectively. This temperature range is of great technological importance since it is here when the O_i defect becomes mobile and firstly forms thermal donors and subsequently, the shallow thermal donors. The role if Si_i in these reactions is still controversial, and therefore future experiments may utilise highly stable C_i related centres as a monitor for Si_i generation, as oxygen precipitates. The presence of C in large concentrations hinders the formation of thermal donors although the exact mechanism is unclear; if these defects contained C_i, then the formation of C_i defects would lead to a loss reduction in their concentration, alternatively O_i could precipitate around C_s due to the respective compressive and tensile strains the defects induce in the lattice. Even more importantly, recent calculations by Ewels *et al* using AIMPRO, have shown that a C_i-H unit complexed with oxygen, may be the core of the early shallow thermal donors. The importance of C_i-H defects in Si has therefore grown considerable in stature over recent years and is a very active area of research at the moment. The richness of these defects stems from its low migration energy, an the similarity of the CH unit to many other defects and impurities in Si (notably N for the shallow thermal donor). Parallels with phosphorus have also been drawn in this thesis, since the T-centre is isoelectronic with the CP pair.

Finally, light impurities in germanium are discussed in chapter 8. The interaction between H and native defects was first considered. The local vibrational modes of the VH_n and IH_m defects were presented, and good agreement between the calculated modes of the IH_2 centre, and two experimentally observed IR modes lead to the assignment of this defect. A detailed analysis of VH_n complexes has not been performed to date. The substitutional carbon defect in Ge, and Si/Ge alloys was then investigated; implanting the alloy in particular with carbon is an area of current investigation due to the tensile strain it induces in the lattice. Therefore it has been proposed that small C concentrations may reduce the lattice mismatch, and consequently alloys with high crystalline quality and low defect concentration may be grown on the pure Si surface. These crystalline layered materials have long been of interest due to their potential in device manufacturing an band gap engineering. Our results clearly showed the C preferentially bonds to Si in the alloy, and the trends in positions of the local modes with varying Si/Ge concentrations were fairly well reproduced.

The AIMPRO cluster code has therefore been able to accurately account for the observed optical, magnetic, and electronic properties of several well characterised defects in silicon. As well as these basic properties of defects, the binding and migration energies of defects can also be calculated. The recent success of AIMPRO in conjunction with the Slater transition argument in accounting for the position of a defects gap level represents a promising advancement in the theoretical modelling of the electronic properties of defects. The application of the cluster code to defects in semiconductors has distinct advantages, but there are also flaws, notably, the difficulty in determining the formation energy of defects. In future, with the advent of the super-cell version of AIMPRO a more systematic set of results will be possible with the combination of both methods to deal with extended defects. Although the calculations in this thesis have been exclusively on monopolar semiconductors, the AIMPRO code has also been applied to compound semiconductors. Both the cluster and the supercell codes have been applied successfully to the properties of, and defects in important compound semiconductors such as GaAs and GaN.

Another strength of the cluster code is that it can be applied to a much wider range of problems outside solid state physics, including molecular chemistry (both fullerenes and molecular electronics have been considered) and the modelling of biological molecules. As computer power increases the scale of accessible problems increases, and the potential of such a versatile code is large.

Of course, the role of theoretical work is not to simply strengthen the assignments made by experimentalists - although this is a very worthy task - but ultimately of greater value is the use of theory as a predictive tool. The self-interstitial defect in Si and isolated H in semiconductors are examples where much of the understanding of the defect comes from theory. The work on C_i-H defects presented in chapter 6 represents the first theoretical calculations on such centres. Such work has helped experimentalists to assign phonon replicas. Here predictions are made: about the bistability and migration of the defect, and ultimately about the formation of some of the centres observed in PL. Again this is very worthy since a complete understanding of possible reaction paths is necessary to understand the evolution of defects in silicon. The complexity of the problem of cataloguing defects in silicon increases with increasing temperature as well as the number of constituent atoms which open up more reaction paths. Much of the work presented here, in particular, the work on the T-centre represents a prime example of the productiveness of the joint experimental/theory collaboration. Such work is necessary before a good understanding of defects in silicon, and their interactions are more completely understood, particularly in the elevated temperatures >400°C at which the samples are thermally processed to reduce the damage resulting from irradiation - necessary to include dopants at specific regions in the crystal. It seems inevitable though that collaboration is the key to more rapid progress, with the joint efforts of several experimental and theoretical groups resulting in the understanding of a given defect, but it seems that in answering one set of questions we simply ask new ones. The field of defects in silicon is an extremely challenging, rewarding and yet frustrating area with a long past, and I'm sure a long future.

Bibliography

[1] H. M. Tütüncü, PhD thesis, Exeter University, (1997).

[2] J. Tersoff, Phys. Rev. B **39**, 5566 (1989).

[3] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[4] P. Hohenberg, and W. Kohn, Phys. Rev. B **136** 864 (1964).

[5] W. Kohn, and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).

[6] M. Gellmann, and K. A. Brueckner, Phys. Rev. **106**, 364 (1957).

[7] D. M. Ceperly, and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[8] J. P. Perdew, and A. Zunger, Phys. Rev. B **23**, 548 (1981).

[9] B. I. Dunlap, W. J. Connolly, and J. R. Sabin, J. Chem. Phys. **71**, 8024 (1979).

[10] R. Jones, and A. Sayyash, J. Phys. C: Solid st. Phys. **19**, L653 (1986).

[11] G. M. S. Lister, and R. Jones, Unpublished (1988).

[12] M. J. P. Musgrave, and J. A. Pople, Proc. Roy. Soc. **A268**, 474 (1962).

[13] R. Jones, J. P. Goss, C. Ewels, and S. Öberg, Phys. Rev. B **50**, 8378 (1994)

[14] R. S. Leigh, and B. Szigeti, Proc. Roy. Soc. A **301**, 211 (1967).

[15] J. C. Slater, *Quantum Theory of Atomic Structure*, Vol. 2, McGraw Hill, New York (1960).

[16] J. F. Janak, Phys. Rev. B **18**, 7165 (1978).

[17] A. Resende, J. Goss, P. R. Briddon, S. Öberg, and R. Jones, unpublished.

[18] O. Svelto, *Principles of lasers*, Plenum Press, New York.

[19] J. P. Goss, R. Jones, S. J. Breuer, P. R. Briddon, and S. Öberg, Phys. Rev. Lett. **77**, 3041 (1996).

[20] See for example: B. Henderson, 'Defects in crystalline solids', Edward Arnold Publishers Ltd., (1972); J. Bourgoin and M. Lannoo, 'Point defects in semiconductors II', Springer-Verlag (New York), (1983); R. C. Newman, 'Infra-red studies of crystal defects', Taylor and Francis (London), (1973); R. C. Newman, 'Localised vibrational mode spectroscopy', ed. R. A. Stradling and P. C. Klipstein, Adam Hilger (Bristol and New York), (1990).

[21] See for example: D. V. Lang, J. Appl. Phys. **45**, 3023 (1974); J. Bourgoin and M. Lannoo, 'Point defects in semiconductors II', Springer-Verlag (New York), (1983); D. W. Palmer, 'Characterisation of semiconductors by capacitance methods', ed. R. A. Stradling and P. C. Klipstein, Adam Hilger (Bristol and New York), (1990).

[22] See for example: J. Bourgoin and M. Lannoo, 'Point defects in semiconductors II', Springer-Verlag (New York), (1983); R. C. Newman, 'Point defect studies using electron paramagnetic resonance', ed. R. A. Stradling and P. C. Klipstein, Adam Hilger (Bristol and New York), (1990).

[23] A. A. Kaplyanskii, Optics and Spectroscopy **16**, 329 (1964).

[24] For example, various articles in: Phil. Trans. R. Soc. Lond. A, **350** 169-333, *Ed.* S. F. J. Cox, A. M. Stoneham, and M. C. R. Symons, (1995).

[25] A. Chantre, and D. Bois, Phys. Rev. B **31**, 7979 (1985).

[26] X. D. Zhan, and G. D. Watkins, Phys. Rev. B **47**, 6363 (1993).

[27] L. W. Song, X. D. Zhan, B. W. Benson, and G. D. Watkins, Phys. Rev. Lett. **60**, 460 (1988); L. W. Song, X. D. Zhan, B. W. Benson, and G. D. Watkins, Phys. Rev. B. **42**, 5765 (1990).

[28] L. F. Makarenko, V. P Markevich, and L. I. Murin, Sov. Phys. Semicond. **18**, 1192 (1985); A. Chantre, Phys. Rev. Lett. **50**, 1500 (1987); T. Hallberg, and J. L. Linström, Appl. Phys. Lett. **68**, 3458 (1996).

[29] S. K. Bains, and P. Banbury, J. Phys. C: Solid state phys. **18**, L109 (1985).

[30] D. V. Lang, in *Deep centers in semiconductors*, ed. S. T. Panteledes (Gordon and Breach, New York), chapter 6 (1986).

[31] D. V. Lang, in *Deep centers in semiconductors*, ed. S. T. Panteledes (Gordon and Breach, New York), chapter 7 (1986).

[32] M. Stavola, M. Levinson, J. L. Benton, and L. C. Kimberling, Phys. Rev. B **30**, 832 (1984); M. Levinson, M. Stavola, P. Besomi, and W. A. Bonher, Phys. Rev. B **30**, 5817 (1984).

[33] P. Clauws, F. Callens, F. Maes, J. Vennick, and E. Boesmam, Phys. Rev. B **44**, 3665 (1991).

[34] G. D. Watkins in, *Radiation Damage in Semiconductors*, Ed P. Barach, (Dunod,Paris 1964), 97.

[35] L. C. Kimberling, Sol. State. Elec. **21**, 1391 (1978).

[36] C. M. Chen, J. H. Svensson, E. Janzen, B. Monemar, A. Henry, A. M. Frens, M. T. Bennebroeck, and J. Schimdt, Phys. Rev. Lett. **71**, 416 (1993).

[37] M. Levinson, J. Appl. Phys. **58**, 2628 (1985).

[38] G. W. Ludwig, and H. H. Woodbury, in *Solid state physics*, ed. H. Ehrenreich, Frederick Seitz, and David Turnbull, (Acandemic, New York), Vol. **13**, 223 (1962).

[39] J. Van Kooten, G. A. Weller, and C. A. J. Ammerlaan, Phys. Rev. B **30**, 4564 (1984).

[40] G. E. Jellison, J. Appl. Phys. **53**, 5715 (1982).

[41] L. W. Song, X. D. Zhan, B. W. Benson, and G. D. Watkins, Phys. Rev. B **33**, 1452 (1986).

[42] A. Chantre, and L. C. Kimberling, Appl. Phys. Lett. **48**, 1000 (1986).

[43] G. D. Watkins, in *Materials Science and Technology* **4**, edited by W. Schröter (VCH, Weinheim, 1991).

[44] D. J. Chadi, Phys. Rev. B **46**, 9400 (1992).

[45] P. E. Blöchl, E. Smargiassi, R. Car, D. B. Laks, W. Andreoni, and S. T. Panteledes, Phys. Rev. Lett. **70**, 2435 (1993).

[46] A. Mainwood, Mater. Sci. Forum **196-201**, 1589 (1995).

[47] C. G. Van de Walle, and J. Neugenbauer, Phys. Rev. B **52**, 14320 (1995).

[48] M. Nastar, V. V. Bulatov, and S. Yip, Phys. Rev. B **53**, 13521 (1996).

[49] J. Zhu, T. Diaz de la Rubia, L. H. Yang, C. Mailhiot, and G. H. Gilmer, Phys. Rev. B **54**, 4741 (1996).

[50] M. Tang, L. Colombo, J. Zhun, and T. Diaz de la Rubia, Phys. Rev. B **55**, 14279 (1997).

[51] S. J. Clark, and G. J. Ackland, Phys. Rev. B **56**, 47 (1997).

[52] U. Gösele, A. Plössl, and T. Y. Tan, in *Process Pysics and Modelling in Semiconductor Technology*, edited by G. R. Srinivasan, C. S. Murthy, and S. T. Dunham (Electrochemical Society, Pennington, NJ), 309 (1996).

[53] D. J. Eaglesham, *Dopants, defects and diffusion*, Physics world, bf41 (Nov 1995).

[54] J. C. Bourgoin, and J. W. Corbett, Phys. Rev. Lett **38**, 35 (1972).

[55] Y. Bar-Yam, and J. D. Joannopolous, Phys. Rev. B **30**, 1844 (1984).

[56] Y. Bar-Yam, and J. D. Joannopoulus, Phys. Rev. B **30**, 2216 (1984).

[57] Y. Bar-Yam, and J. D. Joannopolous, *Proc. 13th Int. Conf. on Defects in Semiconductors*, edited by L. C. Kimberling, and J. M. Parsey Jr. (Warrendale, PA: American Institute of Mechanical Engineers), 251 (1985).

[58] R. Car, P. J. Kelly, A. Oshiyama, and S. T. Panteledes, *Proc. 13th Int. Conf. on Defects in Semiconductors*, edited by L. C. Kimberling, and J. M. Parsey Jr. (Warrendale, PA: American Institute of Mechanical Engineers), 251 (1985).

[59] M. Bech Nielsen, L. Hoffmann, M. Budde, R. Jones, J. Goss, and S. Öberg, in *Defects in Semiconductors 18*, **196-201** of Materials Science Forum, edited by M. Suezawa and H Katayama-Yoshida (Trans-Tech, Aedermannsdorf, Switzerland) 933 (1995).

[60] H. J. Stein, J. Elecectronic Mat. **4**, 159 (1975).

[61] L. M. Xie, M. W. Qi, and J. M. Chen, J. Phys.: Condens. Matter **3**, 8519 (1991).

[62] M. Budde, B. Bech Nielsen, P. Leary, J. Goss, R. Jones, P. R. Briddon, S. Õberg, and S. J. Breuer, accepted for publication in Phys. Rev. B.; M. Budde, B. Bech Nielsen, P. Leary, J. Goss, R. Jones, P. R. Briddon, S. Õberg, and S. J. Breuer, Mater. Sci. Forum (in press).

[63] P. Deak, L. C. Snyder, M. Heinrich, C. R. Ortiz, and J. W. Corbett, Physica B **170**, 253 (1991).

[64] N. E. B. Cowern, K. T. F. Janssen, G. F. A. van de Walle, and D. J. Gravesteijn, Phys. Rev. Lett. **65**, 2434 (1990); N. E. B. Cowern, G. F. A. van de Walle, D. J. Gravesteijn, and C. J. Vriezema, Phys. Rev. Lett. **67**, 212 (1991); N. E. B. Cowern, G. F. A. van de Walle, P. C. Zalm, and D. J. Oostra, Phys. Rev. Lett. **69**, 116 (1992).

[65] S. Takeda, M. Kohyama, and K. Ibe, Philos. Mag. A **70**, 287 (1994).

[66] S. Takeda, and T. Kamino, Phys. Rev. B **51**, 2148 (1995).

[67] D. J. Eaglesham, P. A. Stolk, H. -J. Gossman, and J. M. Poate, Appl. Phys. Lett. **65**, 2981 (1994).

[68] R. C. Newman, and J. B. Willis, J. Phys. & Chem. Solids **26**, 373 (1965).

[69] R. C. Newman, and R. S. Smith, J. Phys. & Chem. Solids **30**, 1943 (1969).

[70] S. P. Chappell, M. Claybourn, R. C. Newman, and K. G. Barraclough, Semicond. Sci. Technol. **3**, 1047 (1988).

[71] Windisch D, Becker P, Phil. Mag. A **58**, 435 (1988).

[72] F. Rollert, N. A. Stolwijk, and H. Mehrer, Mater. Sci. Forum. **38-41**, 753 (1989).

[73] P. A. Stolk, D. J. Eaglesham, H. J. Gossman, and J. M. Poate, Appl. Pjys. Lett. **66**, 1370 (1995).

[74] R. Jones, J. Phys. C. **20**, L713 (1987).

[75] R. Jones, and S. Öberg, Semicond. Sci. & Technol. **7**, 27 (1992).

[76] G. G. DeLeo, W. B. Fowler, and G. D. Watkins, Phys. Rev. B **29**, 3193 (198 4).

[77] A. R. Bean, and R. C. Newman, Solid State Commun. **8**, 175 (1970).

[78] R. C. Newman, and A. R. Bean, Rad. Eff. **8**,189 (1971).

[79] G. D. Watkins, and K. L. Brower, Phys. Rev. Lett. **36**, 1329 (1976).

[80] L. W. Song, and G. D. Watkins, Phys. Rev. B. **42**, 5759 (1990).

[81] R. Wooley, E. C. Lightowlers, A. K. Tipping, M. Claybourn, and R. C. Newman, *Mat. Sci. Forum* **10-12**, 929 (1986).

[82] J. F. Zheng, M. Stavola, and G. D. Watkins, *The Physics of Semiconductors*, ed. D. J. Lockwood, World Scientific, Singapore, (1994) p 2363 .

[83] M. Besson, and G. G. DeLeo, Phs. Rev. B **43**, 4028 (1991).

[84] M. J. Burnard and G. G. DeLeo, Phys. Rev. B **47**, 10217 (1993).

[85] J. Tersoff, Phys. Rev. Lett. **64**, 1757 (1990).

[86] R. B. Capaz, A. Dal Pino,Jr., and J. D. Joannopoulos, Phys. Rev. B **50**, 7439 (1994).

[87] K. Thonke, A. Teschner, and R. Sauer, Solid State Commun. **61**, 241, (1987).

[88] A. K. Tipping, and R. C. Newman, Semicond. Sci. Technol. **2**, 315, (1987).

[89] R. Jones, S. Öberg, P. Leary, and V. J. B. Torres, Materials Science Forum **196-201**, 785 (1995).

[90] G. Davies, and R. C. Newman, in *Handbook on Semiconductors*, **3**, ed. by S. Mahajan, Elsevier (1994), p 1557.

[91] J. R. Noonan, C. G. Kirkpatrick, and B. G. Streetman, J. Appl. Phys. **47**, 3 010 (1976).

[92] G. Davies, and M. C. do Carmo, J. Phys. C. **14**, l687 (1981).

[93] K. Thonke, H. Klemisch, J. Weber, and R. Sauer, Phys. Rev. B **24**, 5874 (1981).

[94] K. M. Lee, K. P. O'Donnell, J. Weber, B. C. Cavenett, and G. D. Watkins, Phys. R ev. Lett. **48**, 37 (1982).

[95] K. P. O'Donnell, K. M. Lee, and G. D. Watkins, Physica B **116**, 258 (1983).

[96] G. Davies, and M. C. do Carmo, J. Phys. C. **14**, 1687 (1981).

[97] R. C. Newman, and R. S. Smith, J. Phys. Chem. Solids **30**, 1492 (1969).

[98] M. M. Afanasjev, R. Laiho, L. S. Vlasenko, and M. P. Vlasenko, Solid State Commun. **102**, 595 (1997).

[99] E. C. Lightowlers, and A. N. Safonov, Mater. Sci. Forum (in press).

[100] T. K. Kwok, Phys. Rev. B **51**, 17188 (1995).

[101] A. N. Safonov, private communication.

[102] A. Endrös, Phys. Rev. Lett. **63**, 70 (1989).

[103] Y. Kamiura, M. Yoneta, Y. Nishiyami, and F. Hashimoto, J. Appl. Phys. **72**, 3394 (1992).

[104] Y. Kamiura, M. Tsutsue, Y. Yamashita, and F. Hashimoto, J. Appl. Phys. **78**, 4478 (1995).

[105] Y. Kamiura, M. Tsutsue, M. Hayashi, Y. Yamashita and F. Hashimoto, Materials Science Forum **196-201**, 903, (1995).

[106] Y. Kamiura, N. Ishiga, S. Ohyama, and Y. Yamashita, Mater. Sci. Forum (in press).

[107] D. M. Maric, P. F. Meier, and S. K. Estreicher, Phys. Rev. B **47**, 3620 (1993).

[108] C. Kaneta, and H. Katayama-Yoshida, in *The Physics of Semiconductors*, edited by D. J. Lockwood (World Scientific, Singapore), (1995).

[109] Y. Zhou, R Luchsinger, P. F. Meier, H. U. Suter, Dj. Maric, and S. K. Estreicher, Materials Science Forum **196-201**, 891, (1995).

[110] C. Kaneta, and H. Katayama-Yoshida, Y. Kamiura, M. Tsutsue, M. Hayashi, Y. Yamashita and F. Hashimoto, Materials Science Forum **196-201**, 897, (1995).

[111] J. B. Holbech, B. Bech Nielsen, R. Jones, P. Sitch, and S. Öberg, Phys. Rev. Lett. **71**, 875 (1993).

[112] N. S. Minaev, and A. V. Mudryi, Phys. Stat. Solidi (a) **68**, 561 (1981).

[113] E. Irion, N. Burger, K. Thonke, and R. Sauer, J. Phys. C: Solid State Phys. **18**, 5083 (1985).

[114] A.N. Safonov and E.C. Lightowlers, Mat. Sci. Forum **143-147**, 903 (1994).

[115] A. N. Safonov, E. C. Lightowlers, and G. Davies, Materials Science Forum **196-201**, 909, (1995).

[116] E. C. Lightowlers, R. C. Newman, and J. H. Tucker, Semicond. Sci. and Tech. **9**, 1370 (1994).

[117] E. C. Lightowlers, Mat. Sci Forum **196-201**, 817 (1995).

[118] A. N. Safonov, E. C. Lightowlers, G. Davies, P. Leary, R. Jones, and S. Öberg, Phys. Rev. Lett. **77**, 4812 (1996).

[119] J. Gower, G. Davies, E. C. Lightowlers, and A. N. Safonov, Mater. Sci. Forum (in press).

[120] A. N. Safonov, and E. C. Lightowlers, Mater. Sci. Forum (in press).

[121] P. Leary, R. Jones, S. Öberg, and V. J. B. Torres, Phys. Rev. B. **55**, 2188 (1997).

[122] C. G. Van de Walle, and J. Neugebauer, Phys. Rev. B **52**, 14320 (1995).

[123] X. D. Zhan, and G. D. Watkins, Phys. Rev. B **47**, 6363 (1993).

[124] A. Mainwood, Mater. Sci. Forum (in press).

[125] C. P. Ewels, R. Jones, S. Oberg, J. Miro, and P. Deak Phys. Rev. Lett. **77**, 865 (1996).

[126] R. C. Newman, J. H. Tucker, N. G. Semialtianos, E. C. Lightowlers, T. Gregorkiewicz, I. S. Zevenbergen, C. A. J. Ammerlaan, Phys. Rev. B, **54**, 6803 (1996).

[127] J. Weber, and D. I. Bohne, in *Early Stages of Oxygen Precipitation in Silicon,* ed. R. Jones, NATO ASI Series (3. High Technology) Vol. 17, Kluwer Academic Publishers, Dordrecht, 123 (1996).

[128] M. Stutzmann, J. B. Chevrier, C. P. Herrero, and A. Breitschwert, Appl. Phys. A **53**, 47 (1991).

[129] K. Kohler, J. W. Coburn, D. E. Horne, E. Kay, and J. H. Keller, J. Appl. Phys. **57**, 59 (1985).

[130] S. J. Pearton, J. W. Corbett, and T. S. Shi, Appl. Phys. A **43**, 153 (1987).

[131] M. Budde, B. Bech Nielsen, R. Jones, J. Goss, and S. Öberg, Phys. Rev. B **54**, 5485 (1996).

[132] R. T. Sanderson, *Chemical Bonds and Bond Energy*, 2nd edition, (Academic, New York), 113 (1976).

[133] A. G. Gaydon, *Dissociation Energies and Spectra of Diatomic Molecules*, 3rd edition (Chapman and Hall, London), (1968).

[134] R. I. Scace, and G. A. Slack, J. Chem. Phys. **30**, 1551 (1959).

[135] L. Hoffmann, J. C. Bach, B. Bech Nielsen, P. Leary, R. Jones, and S. Öberg, Phys. Rev. B **55**, 11167 (1997).

[136] *GeSi strained layers and their applications*, edited by A. M. Stoneham, and S. C. Jain, (Institure of Physics Publishing, Bristol), (1995).

[137] K. Eberl, S. S. Lyer, S. Zollner, J. C. Tsang, and F. K. Le Goues, Appl. Phys. Lett. **60**, 3033 (1992).

[138] B. Dietrich, H. J. Osten, H. Rücker, M. Methfessel, and P. Zaumseil, Phys. Rev. B **49**, 17185 (1994).

[139] H. Jacobsson, J. Xiang, N. Herbots, S. Whaley, P. Ye, and S. Hearne, J. Appl. Phys. **81**, 3081 (1997).

[140] L. Hoffmann, J. C. Bach, J. Lundsgaard Hansen, A. Nylandsted Larsen, B. Bech Nielsen, P. Leary, R. Jones, and S. Öberg, Mater. Sci. Forum (in press).