Boron related point defects in silicon

Submitted by James Adey to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Physics.

December, 2004

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other university.

Abstract

The results of density functional theory calculations performed to model the properties of boron related defects in silicon are presented. The interaction of boron with self-interstitials that may be produced via irradiation or ion implantation, is considered in detail as is the interaction of these boron-interstitial complexes with common impurities. The theoretical modelling described here allows the explanation on the atomic scale of all of the boron-interstitial clusters observed experimentally to date with the exception of the recently observed extended boron-interstitial clusters. In most cases the formation and annihilation mechanisms are also deduced. A study of biaxially strained silicon reveals the important role that the change to the Fermi-energy with strain has on the stability of defects in such material. Throughout this thesis a strong emphasis is placed on the cooperation of theory and experiment.

To Becky and my mother, Lynn.

Aknowledgements

I am extremely grateful for the opportunity to work under the supervision Prof. R. Jones and for his excellent guidance and motivation I thank him. I also wish to thank the rest of the AIMPRO family, particularly my colleagues Casper, Chris, Jon, Ben, José, Thomas, Steve and Luke who have not only assisted in this work but who have made working in Exeter a pleasure. I thank Dr. J. M. Rowe for his maintenance of our local system. Dr. P. R. Briddon has worked tirelessly on the maintenance and development of the code and to him we are all indebted. I would also like to thank Prof. S. Öberg for his help in computing matters and Prof. D. W. Palmer for the enthusiasm and knowledge he has brought to this work and for his careful proof reading of this thesis. Finally I wish to thank the EPSRC whose funding made this research possible.

Contents

List of Tables

List of Figures

List of Publications

1. J. Adey, R. Jones, P. R. Briddon, *Enhanced dopant solubility in strained silicon*, to appear in Physica Status Solidi (C).

2. J. Adey, R. Jones, P. R. Briddon, *Enhanced dopant solubility in strained silicon*, Journal of Physics: Condensed Matter **16**, 9117 (2004).

3. J. Schmidt, K. Bothe, D. Macdonald, J. Adey, R. Jones and D. W. Palmer, *Mechanisms of Light-Induced Degradation in c-Si Solar Cells*, to appear in the Proceedings of The 4th International Symposium on Advanced Science and Technology of Silicon Materials (Hawaii, 2004).

4. J. Adey, R. Jones, D. W. Palmer, P. R. Briddon and S. Öberg, *Degradation of boron doped Cz-Si solar cells*, Physical Review Letters **93**, 055504 (2004).

5. J. Adey, J. P. Goss, R. Jones and P. R. Briddon, *Interstitial boron defects in Si*, Physica B: Condensed Matter **340-342**, 505 (2003).

6. J. Adey, R. Jones, P. R. Briddon and J. P. Goss, *Optical and electrical activity of boron interstitial defects in Si*, Journal of Physics: Condensed Matter **15**, S2851 (2003).

7. J. Adey, R. Jones and P. R. Briddon, *Formation of B_iO_i, B_iC_s, and $B_iB_sH_i$ defects in e-irradiated or ion-implanted silicon containing boron*, Applied Physics Letters **83**, 665 (2003).

8. J. Adey, J. P. Goss, R. Jones and P. R. Briddon, *Identification of boron clusters and boron-interstitial clusters in silicon*, Physical Review B **67**, 245325 (2003).

Introduction

Silicon was discovered in Sweden in 1824 by Jöns Jacob Berzelius and since then has become one of the most widely studied materials known to man. To control the electrical properties of silicon the crystal must be doped. Silicon is a group IV element and crystallises into a face centred cubic lattice with a two atom basis set (one atom at each lattice site (0,0,0) and a second at (1/4,1/4,1/4)). From a simple one electron picture, it can be understood how replacing some of the silicon atoms by a group V element (forced to be four-fold coordinated) will result in an excess of free electrons that can act as charge carriers making the material n-type. Alternatively a group III element may be used and this will result in an excess of free holes making the material p-type. The boundary between a p and n-type region is known as a p-n junction and this is the fundamental building block of the transistor and indeed any other electronic device. It is therefore critical that the properties of commonly used dopants in silicon are well understood. By far the most common p-type dopant is boron and due to interest in interest in the phenomena of transient enhanced diffusion and boron interstitial cluster formation, discussed below, the properties of boron in silicon are the main topic of this thesis. Of the n-type dopants phosphorus and arsenic are common choices. Arsenic is the only n-type dopant discussed here, where its solubility is considered in a novel form of silicon in chapter 7.

The latest generation of computer chips feature transistors that are just 50 nm in size. Although silicon can be doped during growth the highly nonuniform doping distribution required by device manufacturers makes this method of doping impractical. Instead the highly p and n-type regions are generated by implanting dopants into the silicon. Only this technique allows accurately controlled doped regions on the nano-meter scale. However there are some serious problems associated with ion-implantation stemming from the fact that implantation generates much damage including many self-interstitials and self-interstitial clusters. The diffusion of substitutional boron requires self-interstitials since boron diffuses via the kick-out mechanism (Zhu *et al.*, 1996) or according to recent calculations by an interstitialcy mechanism (Windl *et al.*, 1999; Hakala *et al.*, 2000; Jeong and Oshiyama, 2001). Studying electron irradiated silicon at low temperature, Watkins (1975) observed that the production rate of the boron-interstitial defect from substitutional boron is comparable to that of the vacancy implying that almost all generated interstitials (that do not recombine) are trapped by boron. The boron interstitial defect is extremely mobile; it diffuses with an activation energy of 0.6 eV (Watkins, 1975) and therefore becomes mobile well below room temperature (Tipping and Newman, 1987). The implication of this is that any process which generates self-interstitials at above room temperature is likely to lead to the diffusion of substitutional boron. Following ion-implantation when the wafer is annealed at around 800°C to remove damage the smaller, less stable self-interstitial clusters dissociate which leads to a supersaturation of self-interstitials. This excess of self-interstitials promotes the rapid diffusion of boron known as transient enhanced diffusion (TED). During TED boron diffuses over distances far greater than that expected for thermal diffusion alone. For example, Stolk *et al.* (1995*b*) observed that during a 15 minute, 810°C anneal of boron doped silicon implanted with silicon, boron diffuses around 200 nm. For thermal diffusion alone (*i.e.* in the absence of a supersaturation of self-interstitials) the boron would have only diffused 3-5 nm. The occurrence of TED within a 50 nm device would clearly be catastrophic. A second problem associated with TED is that of boron interstitial cluster (BIC) formation. BICs form in regions of high boron concentration and since clustered interstitial boron atoms cannot act as shallow acceptors the electrically active fraction of the implanted dopant is greatly reduced (Cowern *et al.*, 1990).

Many techniques have now been demonstrated to reduce the effect of TED. For example the use of lower energy ion implants (Agarwal *et al.*, 1997), preamorphising the surface (Roth *et al.*, 1997), co-implanting boron and fluorine (Downey *et al.*, 1998), using rapid thermal anneals with high ramp rates (Mannino *et al.*, 2001) or spike anneals. Commonly a combination of these methods are used. It is however very much desirable to understand the interaction of boron with self-interstitials and with other impurities since parameters such as the binding energies are important input parameters for simulations of manufacturing techniques. Moreover, as devices sizes continue to decrease rapidly the acceptable limits of diffusion and clustering must also decrease quickly and so a fundamental understanding of the properties of boron in silicon is imperative.

Most of the chapters within this thesis are focused upon understanding the interaction of boron with self-interstitials (chapters 3 and 4) or with impurities common to silicon (chapter 5) by means of first principles calculations (described in chapter 1). An important problem associated with a particular device, the degradation of solar cells due to the generation of a carrier lifetime degrading defect with illumination, is discussed in chapter 6. Finally chapter 7 deals with the properties of the dopants boron and arsenic in silicon that has been subject to a biaxial strain to alter its properties. Throughout this thesis results of calculations are closely compared to the results of experiment. Such an approach allows confidence in the modelling as well as giving a greater understanding to what is observed experimentally. A description of most of the experimental techniques discussed within this thesis is given in chapter 2 while chapter 1 describes the nature of the calculations performed.

Chapter 1

Theoretical Framework

1.1 The many-body problem

Any quantum mechanical system may be described by the Schrödinger equation

$$\hat{\mathcal{H}}\Psi = E\Psi \tag{1.1}$$

where E is the energy of the system and the operator $\mathcal{H}$ is known as the Hamiltonian. For the solid state the wavefunction Ψ that describes the many-body problem is a function of the coordinates of the α nuclei and i electrons, $\mathbf{R}_\alpha$ and r_i respectively. Note that r_i denotes both the position ($\mathbf{r_i}$) and the spin (s_i) of the i-th electron.

The non-relativistic Hamiltonian used to solve equation 1.1 is

$$\hat{\mathcal{H}} = -\frac{1}{2}\sum_{i}^{N}\nabla_i^2 - \sum_{\alpha}\frac{1}{2M_\alpha}\nabla_\alpha^2 + \frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{I}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_{i,\alpha=1}^{N,I}\frac{Z_\alpha}{|\mathbf{r}_i-\mathbf{R}_\alpha|} + \frac{1}{2}\sum_{\substack{\alpha,\beta=1\\ \alpha\neq\beta}}^{I}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|} \tag{1.2}$$

where M_α and Z_α are the mass and charge of the α-th nucleus respectively and N and I are the number of electrons and nuclei (ions) respectively. Note that all quantities herein are expressed in terms of atomic units unless stated otherwise. This measurement system sets $\hbar$, e, m and $4\pi\epsilon_0$ to unity where $\hbar$ is $h/2\pi$ and h is Planck's constant, e is the electron charge, m electron mass and ϵ_0 is the permittivity of vacuum. One unit of length is then 0.529 Å and a unit of energy is 27.211 eV. This Hamiltonian (equation 1.2) is simply,

$$\hat{\mathcal{H}} = \hat{\mathcal{T}}_e + \hat{\mathcal{T}}_n + \hat{\mathcal{V}}_{e-e} + \hat{\mathcal{V}}_{e-n} + \hat{\mathcal{V}}_{n-n} \tag{1.3}$$

where $\hat{\mathcal{T}}_n$ and $\hat{\mathcal{V}}_{n-n}$ are the kinetic and potential energy operators of the nuclei, $\hat{\mathcal{T}}_e$ and $\hat{\mathcal{V}}_{e-e}$ are the kinetic and potential energy operators of the electrons and $\hat{\mathcal{V}}_{e-n}$ is the operator giving the energy due to the interactions between electrons and nuclei.

Solving equation 1.1 using the Hamiltonian given in equation 1.2 is practically impossible for any system larger than the hydrogen atom and so to make the problem tractable it is necessary to introduce some approximations.

1.2 The Born-Oppenheimer approximation

Since the mass of an electron is several orders of magnitude less than that of a nucleus it is reasonable to assume that the electrons react instantaneously to the movement of the nuclei. Born and Oppenheimer (1925) thus proposed that the electrons' wavefunction $\Psi(r, \mathbf{R})$ is modulated by a function $\chi(\mathbf{R})$ whose amplitude depends only upon the nuclei. This gives the total wavefunction,

$$\Psi_T(r, \mathbf{R}) = \chi(\mathbf{R})\Psi(r, \mathbf{R}). \tag{1.4}$$

Using this and the Hamiltonian given in equation 1.2 we may rewrite the Schrödinger equation as,

$$\begin{aligned} \hat{\mathcal{H}}\chi(\mathbf{R})\Psi(r, \mathbf{R}) &= E_e\chi(\mathbf{R})\Psi(r, \mathbf{R}) + \Psi(r, \mathbf{R})(\hat{\mathcal{T}}_n + \hat{\mathcal{V}}_{n-n})\chi(\mathbf{R}) \\ &- \sum_\alpha \frac{1}{2M_\alpha}[\chi(\mathbf{R})\nabla_\alpha^2\Psi(r, \mathbf{R}) + 2\nabla_\alpha\Psi(r, \mathbf{R}) \cdot \nabla_\alpha\chi(\mathbf{R})] \end{aligned} \tag{1.5}$$

where

$$E_e\Psi(r, \mathbf{R}) = (\hat{\mathcal{T}}_e + \hat{\mathcal{V}}_{e-e} + \hat{\mathcal{V}}_{e-n})\Psi(r, \mathbf{R}). \tag{1.6}$$

Multiplying equation 1.5 by $\Psi^*(r, \mathbf{R})$ and integrating over r (which means integrating over $\mathbf{r}$ then summing over all spins) gives,

$$
\begin{aligned}
E\chi(\mathbf{R}) &= (\hat{\mathcal{T}}_n + \hat{\mathcal{V}}_{n-n} + E_e)\chi(\mathbf{R}) + W(\mathbf{R})\chi(\mathbf{R}) \\
&- \sum_{\alpha}^{I} \frac{1}{M_\alpha} \int \Psi^*(r, \mathbf{R}) \nabla_\alpha \Psi(r, \mathbf{R}) d\mathbf{r} \cdot \nabla_\alpha \chi(\mathbf{R})
\end{aligned}
\tag{1.7}
$$

where

$$
W(\mathbf{R}) = -\sum_{\alpha}^{I} \frac{1}{2M_\alpha} \int \Psi^*(r, \mathbf{R}) \nabla_\alpha^2 \Psi(r, \mathbf{R}) d\mathbf{r}. \tag{1.8}
$$

The last term in equation 1.7 vanishes when $\Psi(r, \mathbf{R})$ is real signifying a non-degenerate ground-state and is otherwise a small perturbation. This small perturbation can however be important, sometimes giving rise to the Jahn-Teller effect (Stoneham, 1975). The term $W(\mathbf{R})$ (equation 1.8) is an almost always negligible and hence is neglected in equation 1.7. Neglecting these terms delivers decoupling of the electron and nuclear motions and allows the electron eigenequation (equation 1.6) to be solved separately. This will be the focus of much of the remainder of this chapter.

1.3 Hartree-Fock Theory

A useful method for solving equation 1.6 is to consider the many-electron wavefunction $\Psi(r)$ as a product of N single electron wavefunctions or spin-orbitals $\psi_N(r_N)$ (Fock, 1930; Roothaan, 1951; Slater, 1965; Stoneham, 1975). To take the Pauli exclusion principle into account the product of the single electron functions must be antisymmetric with respect to the exchange of two particles. A convenient way to express this anti-symmetrised product of one electron wavefunctions is in the form of a Slater determinant (Slater, 1929),

$$
\Psi_{\mathrm{HF}}(r) = \frac{1}{\sqrt{N!}} \det \begin{bmatrix} \psi_1(r_1) & \dots & \psi_N(r_1) \\ \vdots & \ddots & \vdots \\ \psi_1(r_N) & \dots & \psi_N(r_N) \end{bmatrix}. \tag{1.9}
$$

The determinant is guaranteed to be anti-symmetric since exchanging two of the single electron spin-orbitals will change $\Psi(r)$ by a factor of -1 while the presence of two identical spin-orbitals will result in $\Psi(r) = 0$.

The expectation value of equation 1.6 when the wavefunction is taken to be Ψ_{HF} is,

$$E = \sum_{i}^{N} H_i + \frac{1}{2}\sum_{i,j=1}^{N}(J_{ij} - K_{ij}) \tag{1.10}$$

where

$$H_i = \int \psi_i^*(r)\left[-\frac{1}{2}\nabla_i^2 - \sum_{\alpha=1}^{I}\frac{Z_\alpha}{|\mathbf{r_i} - \mathbf{R}_\alpha|}\right]\psi_i(r)dr \tag{1.11}$$

$$J_{ij} = \iint \psi_i^*(r_1)\psi_j^*(r_2)\frac{1}{|\mathbf{r_1} - \mathbf{r_2}|}\psi_i(r_1)\psi_j(r_2)dr_1dr_2 \tag{1.12}$$

$$K_{ij} = \iint \psi_i^*(r_1)\psi_j^*(r_2)\frac{1}{|\mathbf{r_1} - \mathbf{r_2}|}\psi_j(r_1)\psi_i(r_2)dr_1dr_2. \tag{1.13}$$

From equations 1.2 and 1.3 it can be seen that H_i which is known as the one electron integral can be expressed

$$H_i = \langle i|\hat{\mathcal{T}}_e + \hat{\mathcal{V}}_{e-n}|i\rangle. \tag{1.14}$$

Similarly J_{ij} and K_{ij} which are known as the Coulomb or Hartree and exchange integrals respectively can be expressed as

$$J_{ij} = \langle ij|\hat{\mathcal{V}}_{e-e}|ij\rangle \tag{1.15}$$

and

$$K_{ij} = \langle ij|\hat{\mathcal{V}}_{e-e}|ji\rangle. \tag{1.16}$$

The four centre, non-local integrals for J_{ij} and K_{ij} describe respectively electron-electron repulsion and an exchange component which represents spin-correlation effects. Since $J_{ii} = K_{ii}$ it is not necessary to exclude the situation $i = j$ in the two electron summation within equation 1.10.

Minimisation of equation 1.10, subject to the orthogonality condition $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, gives the Hartree-Fock equations (Mc Weeney, 1989; Thijssen, 1999),

$$\hat{\mathcal{F}}\psi_i(r) = \sum_{j=1}^{N}\epsilon_{ij}\psi_j(r) \tag{1.17}$$

where $\hat{\mathcal{F}}$ is the Fock operator given by,

$$\hat{\mathcal{F}} = \hat{h} + \hat{j} - \hat{k} \tag{1.18}$$

with $\langle\psi_i|\hat{h}|\psi_i\rangle = H_i$ (see equations 1.11 and 1.14),

$$\hat{j}f(r_1) = \sum_{k=1}^{N} \int \psi_k^*(r_2)\psi_k(r_2)\frac{1}{|r_1 - r_2|}dr_2 f(r_1) \tag{1.19}$$

and

$$\hat{k}f(r_1) = \sum_{k=1}^{N} \int \psi_k^*(r_2)f(r_2)\frac{1}{|r_1 - r_2|}dr_2 \psi_k(r_1) \tag{1.20}$$

where $f(r)$ is an arbitrary function. The matrix ϵ in equation 1.17 is composed of Lagrange multipliers from the minimisation procedure. The diagonal elements are given by,

$$\epsilon_i \equiv \epsilon_{ii} = \langle\psi_i|\hat{\mathcal{F}}|\psi_i\rangle = H_i + \sum_{j=1}^{N}(J_{ij} - K_{ij}) \tag{1.21}$$

which after summation and comparison to equation 1.10 gives,

$$E = \sum_{i}^{N} \epsilon_i - \frac{1}{2}\sum_{i,j=1}^{N}(J_{ij} - K_{ij}). \tag{1.22}$$

A simplification to the unrestricted open-shell Hartree-Fock method described above can be made when the number of electrons is even. In this case the N spin-orbitals in equation 1.10 are replaced by $N/2$ spin-up orbitals and $N/2$ spin-down orbitals. Equation 1.10 then become,

$$E = 2\sum_{i}^{N/2} H_i + \sum_{i,j=1}^{N/2}(2J_{ij} - K_{ij}) \tag{1.23}$$

where H_i, J_{ij} and K_{ij} are no longer spin dependant and so integrals over r may be replaced by integrals over $\mathbf{r}$. Note that since summations are now over $N/2$ instead of N electrons, matrices are quartered in size. This form of Hartree-Fock theory is known as the restricted Hartree-Fock method.

Koopmans (1934) demonstrated the physical meaning of the eigenvalues ϵ_i. An important consequence of Koopmans' theorem is the ability to calculate excited states within the Hartree-Fock method.

Theorem 1 (Koopmans' Theory) *Assuming that the eigenstates do not vary after removal of one electron from the system, the ionisation energy I_m of the m-th electron is given by $I_m = -\epsilon_m$.*

A major short-coming of Hartree-Fock theory is that it predicts zero density of states at the Fermi energy when applied to a homogeneous electron gas. This short-coming is due to the fact that electron-correlation is not included within the theory (the spin orbitals do not have a functional dependence upon $r_i - r_j$). The Configuration Interaction (CI) method overcomes this problem by expressing Ψ as a linear combination of Slater determinants. This leads to a correction to the total energy which is called the correlation term and in principle the many-electron eigenvalues are reproduced exactly. However, the computational requirements of the CI method usually make it impractical.

1.4 Variational principle

The wavefunction Ψ that is a solution to the Schrödinger equation with a given Hamiltonian may be found computationally via either grid or variational methods. Grid methods use discretisation to find Ψ while variational methods depend upon the variational principle as described here.

An approximation of Ψ is built from a sum of basis functions which form a subspace of Hilbert space $\{\phi_1, ..., \phi_M\}$,

$$\Psi_0 \approx \Psi_{app} = \sum_{i}^{M} c_i \phi_i(r). \tag{1.24}$$

The expectation value of the total energy is then estimated to be,

$$E \approx E_{app} = \frac{\langle \Psi_{app} | \hat{\mathcal{H}} | \Psi_{app} \rangle}{\langle \Psi_{app} | \Psi_{app} \rangle} = \frac{\sum_{i,j=1}^{M} c_i^* c_j H_{ij}}{\sum_{i,j=1}^{M} c_i^* c_j S_{ij}} \tag{1.25}$$

where $H_{ij} = \langle \phi_i | \hat{\mathcal{H}} | \phi_j \rangle$ and $S_{ij} = \langle \phi_i | \phi_j \rangle$, the Hamiltonian and overlap matrix elements respectively. For the ground-state the derivative of E_{app} with c_i should vanish implying

$$\sum_{j=1}^{M} (H_{ij} - E_{app} S_{ij}) c_j = 0, \text{ for } i = 1, .., M. \tag{1.26}$$

This condition is equivalent to the eigenequation

$$\mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{c} = E_{app} \mathbf{c} \tag{1.27}$$

where the matrix **H** is the Hamiltonian and **S** is the overlap matrix. The lowest eigenvalue of equation 1.27 will be greater than or equal to the true ground-state energy E_0. E_{app} may be made to converge to E_0 by increasing the number of basis functions. This is known as the Rayleigh-Ritz variational principle (Kemble, 1932; MacDonald, 1933).

Theorem 2 (Variational principle) *The energy E_{app} computed from an approximate wavefunction Ψ_{app} is an upper bound to the true ground-state energy E_0. Full minimisation of the functional E_{app} with respect to all allowed basis functions gives the true ground-state Ψ_0 and energy $E_0 = E_{app}$; that is*

$$E_0 = \min_{\Psi} E_{app}(\Psi). \tag{1.28}$$

The variational principle is applied to Hartree-Fock theory first by expanding the spin-orbitals into linear combinations of basis functions where the basis functions are usually atomic orbitals,

$$\psi_j(r_1) = \sum_i^M c_{ij}\phi_i(r). \tag{1.29}$$

This is used to construct the Fock equation in matrix form which leads to the generalised eigenvalue equation,

$$\mathbf{F} \cdot \mathbf{c} = \epsilon \cdot \mathbf{S} \cdot \mathbf{c} \tag{1.30}$$

known as the Roothaan equation (Roothaan, 1951). Again **S** is the overlap matrix $S_{ij} = \langle \phi_i | \phi_j \rangle$. **F** is the Fock Matrix given by,

$$F_{ij} = H_{ij} + \sum_{\nu\mu} \left(\sum_a c_{\nu a} c_{\mu a} \right) (2\langle \phi_i \phi_j | \hat{\mathcal{V}}_{e-e} | \phi_\nu \phi_\mu \rangle - \langle \phi_i \phi_\nu | \hat{\mathcal{V}}_{e-e} | \phi_\mu \phi_j \rangle \tag{1.31}$$

where

$$H_{ij} = \langle \phi_i | \hat{\mathcal{T}}_e + \hat{\mathcal{V}}_{e-n} | \phi_j \rangle. \tag{1.32}$$

After making an initial guess for Ψ_{app} these equations are solved self-consistently until the changes in Coulomb and Hartree potentials (see equations 1.15 and 1.16) with each self-consistent cycle become sufficiently small. Thus the variational principle allows the ground-state wavefunction and eigenvalues to be found. However, due to the computationally demanding four-centre integrals (equations 1.12 and 1.13) even without the inclusion of CI Hartree-Fock theory only allows rather small systems to be modelled. To tackle larger systems further simplifications are required.

1.5 Density Functional Theory

Density Functional Theory (DFT) (Hohenberg and Kohn, 1964; Kohn and Sham, 1965), which has been extensively reviewed (Lundqvist and March, 1983; Parr and Yang, 1989; Jones and Gunnarsson, 1989) considers the total energy as a functional of electron charge density $n(\mathbf{r})$ rather than a composition of single electron wavefunctions as is the case for Hartree-Fock theory.

Hohenberg and Kohn (1964) demonstrated that the ground-state charge density is uniquely defined by the external potential and the number of electrons N and since the charge density is

$$n(\mathbf{r}) = N \int |\Psi(r)|^2 dr \tag{1.33}$$

it follows that the ground-state wavefunction will be defined also. Note that the charge density is a function of $\mathbf{r}$ only, in contrast to the wavefunction which is dependant upon r (position and spin).

Theorem 3 (First Hohenberg-Kohn theorem) *The external potential is determined, within a trivial additive constant, by the electron density* $n(\mathbf{r})$ *(Hohenberg and Kohn, 1964).*

The total energy of the system is given by,

$$E[n] = F[n] + \int v_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r} \tag{1.34}$$

where v_{ext} is the external potential experienced by electrons due to, for example, ion-electron interactions and F is a system-independent functional that accounts for the electronic kinetic, electron correlation and exchange-correlation energies.

Theorem 4 (Second Hohenberg-Kohn theorem) *For a trial density $\tilde{n}(\mathbf{r})$ such that $\tilde{n}(\mathbf{r}) \geq 0$ and $\int \tilde{n}(\mathbf{r})d\mathbf{r} = N$ (Hohenberg and Kohn, 1964)*

$$E_0 \leq E[\tilde{n}]. \tag{1.35}$$

This theorem is the basis of a variational principle with

$$E_0 = \min_{\tilde{n}} E[\tilde{n}] \tag{1.36}$$

in analogy to equation 1.28.

Despite the huge simplification that comes with using the charge density as the variational variable no further approximations are made. However the exchange-correlation contribution to the functional F is still non-local and generally unknown.

1.5.1 Kohn-Sham equations

Under the frame work of DFT the ion-electron system is described by a set of one-electron Schrödinger equations (Kohn and Sham, 1965),

$$\left[-\frac{1}{2}\nabla^2 + \int \frac{n(r')}{|\mathbf{r}-\mathbf{r}'|}dr' - \sum_{\alpha}^{I} \frac{Z_\alpha}{|\mathbf{r}-\mathbf{R}_\alpha|} + \frac{\delta E_{\mathrm{xc}}[n]}{\delta n(r)}\right]\psi_\lambda(r) = \epsilon_\lambda \psi_\lambda(r) \tag{1.37}$$

where the charge density n is obtained by summing up all N occupied spin-orbital states,

$$n(r) = \sum_{\lambda=1}^{N}\sum_{s} |\psi_\lambda(\mathbf{r}, s)|^2. \tag{1.38}$$

The first three terms of equation 1.37 describe the kinetic energy, Hartree energy and the external potential due to the nuclei respectively. The fourth term encompasses all remaining many-body contributions to the Hamiltonian in the form of an exchange-correlation functional. The strength of DFT comes from the fact that there exists a universal E_{xc} density functional that depends only upon the charge density $n(r)$ allowing the *exact* ground-state charge-density and total energy of a system to be obtained.

The total energy of the many-body system described by the above equations is (Parr and Yang, 1989)

$$E[n] = \sum_{\lambda=1}^{N} \epsilon_\lambda - J[n] + E_{\text{xc}}[n] - \int V_{\text{xc}}[n]n(r)dr \tag{1.39}$$

where J and V_{xc} are the Hartree energy and exchange-correlation potentials respectively. They are given by,

$$J[n] = \frac{1}{2}\int \frac{n(r)n(r')}{|\mathbf{r}-\mathbf{r}'|}drdr' \tag{1.40}$$

and

$$V_{\text{xc}}[n] = \frac{\delta E_{\text{xc}}[n]}{\delta n(r)}. \tag{1.41}$$

Equations 1.37 and 1.38 above are known as the Kohn-Sham equations. They are solved self-consistently by employing the variational principle described by equation 1.35. However, the precise form of the exchange-correlation functional E_{xc} remains unknown and so one of the standard approximations currently available must be used.

1.5.2 The exchange-correlation functional

A common approximation that attempts to avoid the problem of the unknown exchange-correlation functional E_{xc} is the local density approximation (LDA) or for systems having non-zero total spin, the local spin density approximation (LSDA) (Kohn and Sham, 1965; von Barth and Hedin, 1972; Perdew and Zunger, 1981). The exchange-correlation energy is assumed local and is usually separated into separate exchange and correlation contributions,

$$E_{\text{xc}}[n_\uparrow, n_\downarrow] = E_{\text{x}}[n_\uparrow, n_\downarrow] + E_{\text{c}}[n_\uparrow, n_\downarrow] \tag{1.42}$$

using the LSDA notation. The exchange functional of a homogeneous electron gas has an analytic form (von Barth and Hedin, 1972),

$$E_{\text{x}}[n_\uparrow, n_\downarrow] = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3}\left(n_\uparrow^{4/3} + n_\downarrow^{4/3}\right). \tag{1.43}$$

The correlation term is more complicated with perturbation theory yielding one expression, valid for the high density regime (Perdew and Zunger, 1981) and the Green function Monte

Carlo method yielding another expression valid for the low density regime (Ceperley, 1978; Ceperley and Alder, 1980). A parameterised functional form of the exchange functional is derived by fitting to the numerical results. The most common of these parameterised functionals are Perdew and Zunger (1981) (PZ), Vosko *et al.* (1980) (VWN) and Perdew and Wang (1992) (PW).

In principle a more accurate estimate of the exchange-correlation energy is obtained by considering a first order expansion of E_{xc} in charge density, thus including terms dependent upon the charge density gradient ∇n (Perdew, 1991; Perdew *et al.*, 1996*a*,*b*). This method of obtaining E_{xc} is known as the generalised gradient approximation (GGA).

1.6 Pseudopotentials

The chemical properties of an atom are almost entirely dependent upon the atom's valence electrons and their interaction with neighbouring atoms while the core states are relatively independent of the atom's environment. The core states can then be treated as a screening or pseudopotential which acts upon the valence electrons. However, such a pseudopotential will contain terms arising from the interaction of core and valence electrons and hence will not be transferable to different chemical environments. A transferable ionic potential, $\mathcal{V}^{\mathrm{ps}}_{\mathrm{ion}}$ is constructed by subtracting the potential (Coulombic and exchange-correlation) arising from the valence states from the pseudopotential $\mathcal{V}^{\mathrm{ps}}$,

$$\mathcal{V}^{\mathrm{ps}}_{\mathrm{ion}}(\mathbf{r}) = \mathcal{V}^{\mathrm{ps}}(\mathbf{r}) - \int \frac{n^{\mathrm{ps}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' - E_{\mathrm{XC}}[n^{\mathrm{ps}}(\mathbf{r})] \tag{1.44}$$

where

$$n^{\mathrm{ps}}(\mathbf{r}) = \sum_{\lambda} |\psi^{\mathrm{ps}}_{\lambda}(\mathbf{r})|^2 \tag{1.45}$$

and λ denotes all occupied valence eigenstates of the pseudopotential $\mathcal{V}^{\mathrm{ps}}$. Thus n^{ps} is the charge density of the valence states.

In employing the pseudopotential formalism a number of approximations are made. First a one-electron picture is used to divide the electrons into core and valence states. The second assumption, that the core states are independent of the atoms environment, is known as the *frozen core approximation*. Finally it is assumed that the valence and core states do not overlap significantly and hence we may write $E_{\text{XC}}(n_{core} + n_{valence}) = E_{\text{XC}}(n_{core}) + E_{\text{XC}}(n_{valence})$. This is known as the *small core approximation*. In cases where the overlap is significant it may be necessary to apply *non-linear core corrections* to the pseudopotential (Louie *et al.*, 1982).

The pseudopotential calculation has huge benefits over an all-electron calculation. The greatest benefit is that the number of basis functions required to fit all states is greatly reduced allowing larger systems to be considered. A positive side effect of the pseudopotential approach is a reduction in total energy with respect to an all-electron calculation. This means that numerical errors when comparing the total energy of similar systems are reduced.

The calculations presented in the proceeding chapters employ mainly Bachelet-Hamann-Schlüter (BHS) (Bachelet *et al.*, 1982) and Troullier-Martins (TM) (Troullier and Martins, 1991) pseudopotentials although Hartwigsen-Goedecker-Hutter (HGH) (Hartwigsen *et al.*, 1998) pseudopotentials have also been used on occasion. These pseudopotentials are norm conserving which means that outside a given radius they exactly replicate the real all-electron wavefunction.

Detailed information on the pseudopotential method has been given by several authors (Harrison, 1966; Brust, 1968; Stoneham, 1975; Heine, 1970; Pickett, 1989).

1.7 Boundary conditions

Since DFT is able to deal only with what is essentially a small number of atoms it is important to consider the boundary conditions at the edge of the atoms being treated. Two common choices of boundary conditions exist when treating large systems such as semiconductor crystals which can be considered infinite in respect of the bulk properties. The cluster method considers a cluster or molecule of atoms. At the boundary of the cluster, for calculations of a bulk semiconductor, the dangling bonds are terminated by hydrogen atoms to avoid any dangling bond or surface effects. However, if the cluster is not large enough or the defect is not placed at the centre of the cluster, defect-surface interactions may occur.

The second method avoids defect-surface interactions and is known as the supercell method. Here rather than terminating the dangling bonds at the boundary, the supercell is repeated infinitely in all directions and thus each supercell is considered as a unit cell with associated Brillouin zone (BZ). Any effects due to the surface are removed since the crystal is now infinite, but still there are a number of potential problems associated with this method that must be borne in mind. If we consider a supercell containing a defect, then what is being modelled is actually an infinite array of the defects separated by a_0 where a_0 is the lattice constant of the supercell considered cubic in this simple argument. Hence if a_0 is not large enough then the array of defects will interact with each other via elastic, Coulombic, dipolar or quadrapolar interactions. It is therefore essential that, whichever boundary conditions are used, the calculated properties are converged with respect to the cluster/supercell size. Calculations presented in the proceeding chapters all employ the supercell method. Supercells of 64 or 216 atoms where tested and found to give converged results.

1.8 Brillouin zone sampling

When the supercell method is employed the calculation of physical quantities such as the charge density require integration over the BZ. The integrand function $f(\mathbf{k})$ is periodic in reciprocal space and has no simple analytic form. Hence to avoid integration over a dense mesh of points in $\mathbf{k}$-space a number of schemes have been developed for choosing sets of N_k *special* $\mathbf{k}$-points from which the average value of f over the BZ may be calculated (Baldereschi, 1973; Chadi and Cohen, 1973; Monkhorst and Pack, 1976; Pack and Monkhorst, 1977). The average value $\bar{f}$ of the integral f over the BZ of volume $(2\pi)^3/\Omega$ is then,

$$\bar{f} = \frac{\Omega}{(2\pi)^3} \int f(\mathbf{k}) d\mathbf{k} \approx \frac{1}{N_k} \sum_{i}^{N_k} f(\mathbf{k_i}). \tag{1.46}$$

Throughout this thesis the calculations presented have used the sampling scheme of Monkhorst and Pack (MP) (Monkhorst and Pack, 1976; Pack and Monkhorst, 1977). The main advantages of this scheme are its simplicity and the fact that convergence is easily verifiable.

The MP special **k**-points are a grid of $I \times J \times K$ points in reciprocal space given by,

$$\mathbf{k}(i,j,k) = u_i\mathbf{g_1} + u_j\mathbf{g_2} + u_k\mathbf{g_3} \tag{1.47}$$

where $\mathbf{g_1}$, $\mathbf{g_2}$ and $\mathbf{g_3}$ are the reciprocal space unit-vectors and u_i, u_j and u_k are given by,

$$u_i = (2i - I - 1)/2I, \qquad (i = 1, \ldots I) \tag{1.48}$$

$$u_j = (2j - J - 1)/2J, \qquad (j = 1, \ldots J) \tag{1.49}$$

$$u_k = (2k - K - 1)/2K, \qquad (k = 1, \ldots K) \tag{1.50}$$

with I, J and $K \geq 1$. Often when $I = J = K$ the sampling scheme is referred to as MP-I^3. Convergence can easily be ensured by increasing the values of I, J and K until the calculated value or property of interest does not change significantly.

In a supercell with high enough symmetry some of the **k**-points may be equivalent due to the symmetry operations associated with that supercell. In this case the redundant **k**-points may be removed and the equivalent one that is left will be weighted accordingly.

For the calculation presented in the proceeding chapters a 2^3 Monkhorst-Pack sampling scheme was found to offer sufficient for convergence.

1.9 Basis functions

The AIMPRO LSDA DFT code used throughout this thesis employs a real-space Gaussian type basis set (Jones and Briddon, 1999). One problem with the supercell approach however is that the requirement of integration over the BZ makes an expansion of the wavefunction in reciprocal space necessary.

Bloch basis functions $B_{\mathbf{k}i}(\mathbf{r})$ are constructed from Cartesian Gaussian functions ϕ_i centred at the atomic sites $\mathbf{R_i}$ with N_L lattice vectors $\mathbf{L_n}$,

$$B_{\mathbf{k}i}(r) = \frac{1}{\sqrt{N_L}} \sum_{L_n} \phi_i(\mathbf{r} - \mathbf{R_i} - \mathbf{L_n}) \mathrm{e}^{i\mathbf{k}.\mathbf{L_n}} \tag{1.51}$$

where $\mathbf{k}$ is a reciprocal space vector within the BZ. The localised orbitals are given by,

$$\phi_i(\mathbf{r} - \mathbf{R_i}) = (x - R_{ix})^{l_1} (y - R_{iy})^{l_2} (z - R_{iz})^{l_3} \mathrm{e}^{-\alpha(\mathbf{r}-\mathbf{R_i})^2} \tag{1.52}$$

with $l_i \geq 0$ and where s, p or d-like orbitals correspond to $\sum_i l_i = 0$, 1 or 2 respectively.

The Kohn-Sham orbitals $\psi_{\mathbf{k}\lambda}$ can then be expressed in terms of the $B_{\mathbf{k}i}$ basis functions,

$$\psi_{\mathbf{k}\lambda} = \sum_i c^{\mathbf{k}}_{\lambda i} B_{\mathbf{k}i}(\mathbf{r}) \tag{1.53}$$

resulting in a charge-density,

$$n(\mathbf{r}) = \sum_{i,j,\mathbf{k}} b_{ij}(\mathbf{k}) B^*_{\mathbf{k}i}(\mathbf{r}) B_{\mathbf{k}j}(\mathbf{r}) \tag{1.54}$$

with

$$b_{ij}(\mathbf{k}) = \sum_\lambda f_{\mathbf{k}\lambda} (c^{\mathbf{k}}_{\lambda i})^* c^{\mathbf{k}}_{\lambda j} \tag{1.55}$$

where $f_{\mathbf{k}\lambda}$ is the occupancy of the $\mathbf{k}\lambda$ state. This is either 2 for a filled state, 0 for an empty state or 1 for a half-filled state in a spin averaged calculation while for a spin-polarised calculation the filling of each state is either 1 or 0. If Fermi smearing is used the levels may be partially occupied, according to the Fermi-function. If the filling is metallic then $f_{\mathbf{k}\lambda}$ will be independent of $\mathbf{k}$.

Note that the basis functions $B_{\mathbf{k}i}(\mathbf{r})$ satisfy Bloch's theorem since

$$B_{\mathbf{k}i}(\mathbf{r} + \mathbf{L_n}) = e^{i\mathbf{k}\mathbf{L_n}} B_{\mathbf{k}i}(\mathbf{r}). \tag{1.56}$$

Thus $B_{\mathbf{k}i}$ and $B_{\mathbf{k}'j}$ are orthogonal and the Hamiltonian and other translational symmetric functions have matrix elements diagonal in $\mathbf{k}$-space.

The basis functions for the same $\mathbf{k}$ are not orthogonal and the overlap between them is $S^0_{ij}(\mathbf{k})$ where

$$\begin{aligned} S^{\mathbf{g}}_{ij}(\mathbf{k}) &= \int B^*_{\mathbf{k}i}(\mathbf{r})\mathrm{e}^{-i\mathbf{gr}}B_{\mathbf{k}j}(\mathbf{r})d\mathbf{r} \\ &= \sum_{ijn} \mathrm{e}^{i\mathbf{k}.\mathbf{L_n}} \int \phi_i(\mathbf{r}-\mathbf{R}_i)\mathrm{e}^{-i\mathbf{gr}}\phi_j(\mathbf{r}-\mathbf{R}_j-\mathbf{R}_L)d\mathbf{r}. \end{aligned} \tag{1.57}$$

The Kohn-Sham orbitals are normalised when the coefficients satisfy

$$\sum_{ij} c^{\mathbf{k}}_{\lambda i}{}^* c^{\mathbf{k}}_{\lambda j} S^0_{ij}(\mathbf{k}) = 1. \tag{1.58}$$

The advantage of Gaussian-like functions is that their integrals can be found analytically. Furthermore, in contrast to another popular choice of basis functions namely plain waves, they can be made to vanish quickly away from $\mathbf{R_i}$ (Jones and Briddon, 1999). However, localised basis sets can result in over-completeness which may lead to numerical instabilities.

1.10 Reciprocal-space basis

Although the Kohn-Sham orbitals are obtained in terms of a real-space basis set, the periodic boundary conditions of the supercell method allow the use of reciprocal space. The Hartree (including the local pseudopotential) and the exchange correlation energies are more efficiently evaluated in $\mathbf{k}$-space. To do this the charge-density must be Fourier transformed, the quality of the transformation being dependent upon the grid of $\mathbf{g}$-vectors used. In practice a uniform grid of $\mathbf{g}$-vectors inside a sphere of radius g_{cut} is chosen. This defines the cut-off energy as

$$E_{cut} = \frac{1}{2} g^2_{cut}. \tag{1.59}$$

A value of E_{cut} for which the calculated properties of interest are converged must be selected.

1.11 Ewald summations

Series like $1/r^p$ with $p \leq 2$ converge very slowly as r goes to infinity and so to calculate Coulombic and dipolar interactions it is necessary to evaluate slowly converging sums. To solve this problem Ewald (1921) proposed a method that splits the series into fast and slow converging parts,

$$\sum_L \frac{1}{r_L^p} = \sum_L \frac{\mathrm{erfc}(\alpha r_L)}{r_L^p} + \sum_L \frac{\mathrm{erf}(\alpha r_L)}{r_L^p}. \tag{1.60}$$

Evaluation of the first term is straight forward as $\mathrm{erfc}(x)$ converges to zero when x tends to infinity. The second term which is slow-converging in real space is Fourier transformed and consequently in reciprocal space, its terms become short-ranged. The parameter α controls the transition between the real-space and reciprocal space summations. Although the analytical result does not depend upon the choice of α in practice this is not the case. This is a consequence of the summations being over only a finite set of lattice vectors. An extensive treatment of this method has been given by Leeuw *et al.* (1980). This method is used in AIMPRO to separate and include the local and non-local contributions of the pseudopotential.

1.12 Calculation of observables

DFT is a ground-state theory. This is a consequence of the fact that the Kohn-Sham eigenstates are not electron-wavefunctions but are rather a set of basis functions from which the charge density is expanded. Hence excited state properties (including unoccupied electron energy levels) are beyond the reach of DFT. There are however many properties that DFT can successfully model with the aim of predicting or explaining experimental observations. This section explains how some of the most important of these properties are derived and to what observations they may be related.

1.12.1 Fundamental properties

Local structure - Forces

Once the Kohn-Sham equations have been solved to find the self-consistent charge density for a given arrangement of atoms it is extremely desirable to know the force acting on each individual atom. The force on an atom F_α is simply $-\nabla_\alpha E$ and so by displacing each atom by a small amount and calculating ΔE the force acting on each atom can be found.

Varying the ion positions changes the charge-density in two ways: a change in $b_{ij}(\mathbf{k})$ and a change in the basis functions $B_{\mathbf{k}i}$.

Thus

$$\Delta n = \sum_{ij\mathbf{k}} b_{ij}(\mathbf{k})\Delta\{B^*_{\mathbf{k}i}(\mathbf{r})B_{\mathbf{k}j}(\mathbf{r})\} + \sum_{ij\mathbf{k}} B^*_{\mathbf{k}i}(\mathbf{r})B_{\mathbf{k}j}(\mathbf{r})\Delta b_{ij}(\mathbf{k}). \tag{1.61}$$

The changes to E_e and E_{XC} are

$$\Delta E_e = \frac{1}{N}\int \frac{(n(\mathbf{r}_1)-\rho(\mathbf{r}_1))(\Delta n(\mathbf{r}_2)-\Delta\rho(\mathbf{r}_2))}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 - \Delta E_{l\mathbf{r}} \tag{1.62}$$

and

$$\Delta E_{XC} = \frac{1}{N}\int \mu^{XC}(\mathbf{r})\Delta n d\mathbf{r}. \tag{1.63}$$

The Fourier expansion of $1/\mathbf{r}$ and $\mu^{XC}(\mathbf{r})$ can then be used to deduce

$$\begin{aligned}
\Delta E_e &= 4\pi\Omega \sum_{\mathbf{g}'} \frac{(n_{-\mathbf{g}} - \rho_{-\mathbf{g}})((\Delta n)_{\mathbf{g}} - (\Delta\rho)_{\mathbf{g}})}{\mathbf{g}^2} && (1.64)\\
(\Delta n)_{\mathbf{g}} &= \frac{1}{N_L\Omega}\int e^{-i\mathbf{g}\cdot\mathbf{r}}\Delta n(\mathbf{r})d\mathbf{r}\\
&= \frac{1}{N_L\Omega}\sum_{ij\mathbf{k}}\{S^{\mathbf{g}}_{ij}(\mathbf{k})\Delta b_{ij}(\mathbf{k}) + b_{ij}(\mathbf{k})\Delta S^{\mathbf{g}}_{ij}(\mathbf{k})\} && (1.65)\\
\Delta S^{\mathbf{g}}_{ij} &= \int e^{-i\mathbf{g}\cdot\mathbf{r}}\Delta\{B^*_{\mathbf{k}i}(\mathbf{r})B_{\mathbf{k}j}(\mathbf{r})\} && (1.66)\\
\Delta E_e &= \frac{1}{N_L}\sum_{ij\mathbf{k}} V^H_{ij}(\mathbf{k})\Delta b_{ij}(\mathbf{k})\\
&+ 4\pi\sum_{\mathbf{g}'}\frac{(n_{-\mathbf{g}} - \rho_{-\mathbf{g}})}{\mathbf{g}^2}(\frac{1}{N_L}\sum_{\mathbf{k}} b_{ij}(\mathbf{k})\Delta S^{\mathbf{g}}_{ij}(\mathbf{k}) - \Omega(\Delta\rho)_{\mathbf{g}}) && (1.67)\\
(\Delta\rho)_{\mathbf{g}} &= \frac{1}{N_L\Omega}\int e^{-i\mathbf{g}\cdot\mathbf{r}}\Delta\rho(\mathbf{r})d\mathbf{r}\\
&= \frac{1}{\Omega}\sum_a \Delta\int e^{-i\mathbf{g}\cdot\mathbf{r}} Z_a\left(\sqrt{c_a/\pi}\right)^{3/2} e^{-c_a|\mathbf{r}-\mathbf{R}_a-\mathbf{R}_L|^2}d\mathbf{r}\\
&= \frac{1}{\Omega}\sum_a Z_a e^{-\mathbf{g}^2/4c_a}\Delta e^{i\mathbf{g}\cdot\mathbf{R}_a} && (1.68)
\end{aligned}$$

whereas

$$\Delta E_{XC} = \frac{1}{N_L}\sum_{ij\mathbf{k}} V^{XC}_{ij}(\mathbf{k})\Delta b_{ij}(\mathbf{k}) + \frac{1}{N_L}\sum_{ij\mathbf{g}\mathbf{k}} b_{ij}(\mathbf{k})\mu^{XC}_{-\mathbf{g}}\Delta S^{\mathbf{g}}_{ij}(\mathbf{k}). \tag{1.69}$$

The other terms are

$$\frac{1}{N_L}\sum_{ij\mathbf{k}}\{T_{ij}(\mathbf{k}) + V^{nonl}_{ij}(\mathbf{k},\mathbf{k}')\}\Delta b_{ij}(\mathbf{k}) + \frac{1}{N_L}\sum_{ij\mathbf{k}} b_{ij}(\mathbf{k})\Delta\{T_{ij}(\mathbf{k}) + V^{nonl}_{ij}(\mathbf{k},\mathbf{k}')\} + \Delta E_{ion}. \tag{1.70}$$

Terms in $\Delta b_{ij}(\mathbf{k})$ are gathered and using the KS equations and normalisation of $c^{\mathbf{k}}_{\lambda_i}$,

$$\frac{1}{N_L}\sum_{ij\mathbf{k}}\{T_{ij}(\mathbf{k}) + V^{nonl}_{ij}(\mathbf{k},\mathbf{k}') + V^H_{ij}(\mathbf{k}) + V^{XC}_{ij}(\mathbf{k})\}\Delta b_{ij}(\mathbf{k}) = -\frac{1}{N_L}\sum_{ij\mathbf{k}} b^e_{ij}(\mathbf{k})\Delta S^0_{ij}(\mathbf{k}) \tag{1.71}$$

where

$$b^e_{ij}(\mathbf{k}) = \sum_\lambda f_{\lambda\mathbf{k}}(c^{\mathbf{k}}_{\lambda i})^* c^{\mathbf{k}}_{\lambda j} E_{\lambda\mathbf{k}}. \tag{1.72}$$

Hence

$$\begin{aligned}
\Delta E &= -\frac{1}{N_L}\sum_{ij\mathbf{k}} b^e_{ij}(\mathbf{k})\Delta S^0_{ij}(\mathbf{k}) + \Delta E_{ion} \\
&+ \frac{1}{N_L}\sum_{ij\mathbf{k}} b_{ij}(\mathbf{k})\Delta\{T_{ij}(\mathbf{k}) + V^{nonl}_{ij}(\mathbf{k},\mathbf{k}')\} \\
&+ \frac{4\pi}{N_L}\sum_{ij\mathbf{g}\mathbf{k}'} b_{ij}(\mathbf{k})\frac{(n_{-\mathbf{g}} - \rho_{-\mathbf{g}})}{\mathbf{g}^2}\Delta S^{\mathbf{g}}_{ij}(\mathbf{k}) \\
&+ \frac{1}{N_L}\sum_{\mathbf{g}} \mu^{XC}_{-\mathbf{g}} b_{ij}(\mathbf{k})\Delta S^{\mathbf{g}}_{ij}(\mathbf{k}) - 4\pi\Omega\sum_{\mathbf{g}'}\frac{(n_{-\mathbf{g}} - \rho_{-\mathbf{g}})}{\mathbf{g}^2}(\Delta\rho)_{\mathbf{g}} \qquad (1.73)
\end{aligned}$$

and

$$\begin{aligned}
\Delta T_{ij}(\mathbf{k}) &= \Delta\int B^*_{\mathbf{k}i}(\mathbf{r})\left(-\frac{1}{2}\right)\nabla^2 B_{\mathbf{k}j}(\mathbf{r})d\mathbf{r} \\
&= \sum_L e^{i\mathbf{k}\cdot\mathbf{R}_L}\Delta\int \phi_i(\mathbf{r}-\mathbf{R}_i)\left(-\frac{1}{2}\right)\nabla^2\phi_j(\mathbf{r}-\mathbf{R}_j-\mathbf{R}_L)d\mathbf{r}. \qquad (1.74)
\end{aligned}$$

The ΔE_{ion} term is given by

$$\Delta E_{ion} = \frac{1}{2}\sum_{baL}\Delta\frac{Z_b Z_a}{|\mathbf{R}_b - \mathbf{R}_a - \mathbf{R}_L|}\mathrm{erfc}(c_a c_b|\mathbf{R}_b - \mathbf{R}_a - \mathbf{R}_L|/(c_a + c_b)). \qquad (1.75)$$

Once the contributions to ΔE have been summed and the force on each individual atom is known the atoms can be moved according to the forces. This is done in an iterative manner until forces and the change in total energy with each iteration are insignificant. The structure is then said to be *relaxed*. The AIMPRO code makes use of a conjugate gradient algorithm which means that the atoms are moved in the direction that includes a component from the previous forces as well as the current forces. The distance an atom is moved along this direction is chosen by fitting a cubic or quadratic function to the rate of change of energy with position. One caveat of such a simple approach to finding the most stable structure is that although the relaxed structure will be located at a local minima on the total-energy surface, there is no guarantee that this will be the *global* minima. To ensure, as far as is possible, that the relaxed structure is that at the global minima it is necessary to start the structural optimisation from several different initial configurations.

Total Energy

One of the most fundamental properties of a system is its total energy E_T given previously in equation 1.39. The total energy is the energy of a supercell or cluster and is often used to compare the energy of similar systems when this is equivalent to comparing the formation energies, E_f. In this case a similar system is that containing the same number of atoms of each species (where a species of atom is defined by a specific pseudopotential) and the same overall charge. If these conditions are not met then it is necessary to compare formation energies instead. Since the formation energy takes the charge and the number of atoms of different species into account via the atomic and electron chemical potentials, the formation energy is a well defined physical property.

1.12.2 Derived Properties

Formation Energy

The chemical potential μ_s of a species s is defined as the derivative of the Gibbs free energy $G = E + PV - TS$ with respect to the number of atoms of that species n_s (Reif, 1965; Flynn, 1972). As well as referring to a species of atom, s could equally well refer to the electrons to give the electron chemical potential. Under thermodynamic equilibrium μ_s is equal over the entire system, regardless of any differences in phase and hence is equivalent to the free energy per particle. Neglecting the term PV which is small for solid state reactions and the term $-TS$ (an approximation strictly valid only at low temperature), the formation energy is given by,

$$E_f = E_T + q\mu_e - \sum_s n_s\mu_s \tag{1.76}$$

where E_T is the total energy of the system, q is the nett positive charge of the system in terms of the fundamental electron charge (*i.e.* the system's charge state), n_s is the number of atoms of species s and μ_s is the chemical potential corresponding to that species. The electron chemical potential μ_e is usually taken to be $E_F + E_v$ where E_F is the Fermi energy relative to E_v and E_v is the energy of the highest occupied orbital, usually taken from the Kohn-Sham levels. Alternatively E_v can be obtained by comparing $E_T(q = -1)$ to $E_T(q = 0)$. To obtain accurate values for E_f it is necessary to correct the energy for the interaction between charged supercells, not discussed here. Furthermore, due to finite size effects E_v may differ artificially between different supercells. When comparing the formation energy of different supercells the back-ground potential should thus be aligned.

The formation energy is a quantity that provides access to a great deal of information. By taking the chemical potentials into account it is possible to compare supercells of different size (to check for convergence with respect to cell size) and to compare the stability of defects containing different numbers of each species (useful for the calculation of binding energies, discussed below). Comparing the formation energy of a given defect in different charge states gives the electrical levels (the stable charge state as a function of Fermi-level). The formation energy can also be used to calculate the solubility of a defect as discussed in chapter 7.

Binding Energies

When a defect can be considered to be formed by two or more primary defects it is often useful to know what the cohesive energy is between the constituents. One way to calculate the cohesive or binding energy E_B of a complex C formed by two constituents A and B, is to compare the formation energy of the complex E_f^C to the formation energies of the constituents,

$$E_B = E_f^A + E_f^B - E_f^C \tag{1.77}$$

where E_f^A and E_f^B are the formation energies of the two constituents that make up the complex C.

An alternate way to calculate the binding energy is to construct a series of supercells in which the constituents A and B are first close together (in the form of the complex C) and then at various larger separations. Comparing the energy of these supercells gives the energy as a function of separation of A and B. If the energy has reached an asymptotic limit for the largest separation then the binding energy is given by the difference in energy between the supercell containing separated A and B and the supercell containing C. This method is often more accurate than the formation energy method, especially when the supercells are all equally charged as is often the case, since to first order the strain and quadrupole interactions between neighbouring cells will be equal for each separation and hence cancel out.

Migration Energy

In principle calculating the migration energy of an atom or complex is rather simple. Within AIMPRO there are two methods commonly used, Method 2 and Method 4. Both methods simply provide a way to perform a series of constrained relaxations that will transform the initial structure to the final structure. If the constraints are chosen carefully the transformation should occur via the lowest energy route.

The constraint used by Method 4 is just the vector $\mathbf{v}$ describing the direction in which the atom should move. The user then provides a set of distances along this vector that the atom is to move. After moving each step the force on the diffusing atom is modified so that it has zero component along the vector $\mathbf{v}$ (*i.e.* $\mathbf{F}_{atom} \cdot \mathbf{v} = 0$) and the structure is relaxed according to the modified forces.

Method 2 uses a different constraint. Consider three atoms labelled 1, 2 and 3. The constraint in this case is defined as $C = R_{12}^2 - R_{13}^2$, where R_{12} is the distance between atoms 1 and 2 and R_{13} is the distance between atoms 1 and 3. If when going from initial to final configuration R_{12} increases while R_{13} instead decreases then the constraint C can be increased in steps to take the structure from its initial to its final configuration. The structure is relaxed (subject to the constraint C) at each step along the migration path.

Both Method 2 and Method 4 give the information required to plot a diffusion barrier: a plot of formation or total energy against configuration. The height of the barrier is the migration energy. It is important to choose constraints carefully to ensure that the system is not over constrained and of course it is necessary to have enough steps between initial and final configuration for the migration energy to be accurate. In analogy to the global minima problem faced during relaxation, a number of different paths must be studied to ensure that the path with the lowest barrier has been found. A structural reorientation can be treated in exactly the same way as a migration process.

Thermal stability

Many experiments are able to observe the temperature at which a defect disappears and thus it is desirable to be able to estimate the thermal stability of a defect. Assuming that the complex anneals by dissociation and not by reacting with another species that becomes mobile at the annealing temperature making a simple estimate of the thermal stability is rather simple. The rate at which a complex dissociates is given by,

$$R = \nu \exp\left(\frac{-E_A}{kT}\right) \tag{1.78}$$

where ν is the attempt frequency, k is Boltzmann's constant and T is the temperature. E_A is the activation energy for the defect to dissociate. This is expected to be approximately the sum of the binding energy of the complex and the migration energy of the species that will diffuse away. The calculation of both of these energies is described above. The attempt frequency is usually taken to be the Debye frequency. This is the characteristic maximum vibration frequency of the atoms in the crystal, given by (Hill, 1986)

$$\nu_D = \left(\frac{3}{4\pi}\frac{N}{V}\right)^{1/3} v_s \tag{1.79}$$

where N/V is the number of atoms per unit volume in the crystal and v_s is the speed of sound in the material. For silicon the Debye frequency is around 10^{13} s^{-1}.

The temperature at which a complex anneals can be estimated by solving equation 1.78 for T.

Electrical levels

Calculating the formation energy of charged defects is described above. The resulting formation energy is a function of electron chemical-potential or equivalently Fermi-level.

Therefore, the formation energy of a defect or defect complex can be calculated as a function of E_F for different charge states and for a given value of E_F one of the charge states will be lowest in energy and hence most stable. The value of E_F for which two charge states are equal in energy corresponds to an occupancy level. In other words the value of E_F for which $E_f(q) = E_f(q+1)$ corresponds to the $(q/q+1)$ level of the defect. When E_F is greater than this critical value the defect will be in the charge state q and when E_F is below this level the defect will be in the charge state $(q+1)$. This method for calculating electrical levels is referred to as the Formation Energy Method.

Although the Formation Energy Method often reproduces the observed levels of defects and it is satisfying since it is entirely *ab initio*, it is often found that the technique known as the Marker Method is able to calculate levels with a greater degree of accuracy. The Marker Method calculates levels by solving,

$$E_d(q) - E_d(q+1) + E_d(q/q+1) = E_m(q) - E_m(q+1) + E_m(q/q+1) \tag{1.80}$$

where $E_d(q)$ and $E_d(q+1)$ are the energies of the defect of interest in the q and $(q+1)$ charge states and $E_d(q/q+1)$ is the position of the $(q/q+1)$ level of the defect. $E_m(q)$ and $E_m(q+1)$ are the energies of a marker defect in the specified charge states. The marker defect should ideally be a defect having similar electrical properties to the defect of interest and must have an $(q/q+1)$ level at a position $E_m(q/q+1)$ that is known empirically. Using the empirical value of $E_m(q/q+1)$ and the calculated $E_m(q)$, $E_m(q+1)$ and $E_d(q)$, $E_d(q+1)$ equation 1.80 is solved for the defect level $E_d(q/q+1)$. The basic principle of this method is that the difference in ionisation energy and corresponding occupancy level is similar for similar defects. The Marker Method is the method used throughout this thesis and, as a rule of thumb, the error associated with calculated levels is around 0.2 eV. Even this relatively large error is remarkable given the ground-state nature and the large underestimation of the band-gap by DFT. The Marker Method is thought to be more reliable due to the cancellation of errors arising from strain and charge interactions between supercells. Calculated electrical levels are invaluable tools for the identification of electrically active defects observed for example, by DLTS (section 2.2).

Vibrational modes

The vibrational modes of a crystal or defect are given by the system's dynamical matrix $\mathbf{D}$ (Born and Huang, 1954) by solving the eigenvalue problem

$$\mathbf{D} \cdot \mathbf{U} = \omega^2 \mathbf{U} \tag{1.81}$$

where the $3N$ eigenvalues ω^2 are the square frequencies associated with the $3N$ normal modes $\mathbf{U}$. Each normal mode is a $3N$ dimensional vector describing the motion of all N atoms for that mode. The $3N \times 3N$ dynamical matrix $\mathbf{D}$ has elements,

$$D_{ab}(i,j) = \frac{1}{\sqrt{M_i M_j}} \frac{\partial^2 E}{\partial u_{ia} \partial u_{jb}} \tag{1.82}$$

where a and b are any of the three Cartesian coordinates and u_{ia} and u_{jb} are the displacements of atoms having mass M_i and M_j respectively, in those directions.

To calculate the second derivative of the energy the supercell is first relaxed so that all forces are effectively zero. Then atom i is moved a small amount ϵ (~ 0.025 a.u.) along the Cartesian direction a. The charge density of this new configuration must then be calculated and from the new self-consistent charge density the new forces must be derived. The forces will no longer be zero since the structure has been perturbed. We label the component of the new force on atom j acting in the b Cartesian direction $f^+_{bj}(a,i)$. Atom i is then moved the same distance in the opposite direction $-\epsilon a$ which gives rise to a force $f^-_{bj}(a,i)$. The second derivative of the energy is then to second order in ϵ,

$$\frac{\partial^2 E}{\partial u_{ia} \partial u_{jb}} = \frac{f^+_{bj}(a,i) - f^-_{bj}(a,i)}{2\epsilon}. \tag{1.83}$$

It is important to note that this method of obtaining the energy second derivatives includes some anharmonic contributions, in contrast to the infinitesimal derivative. For this reason the frequencies obtained by the method described here are sometimes referred to as quasi-harmonic frequencies (Jones *et al.*, 1994).

In practice due to the time consuming nature of the procedure, usually only the defect atoms and their nearest neighbours will be treated in the way described above. The energy second derivative of the bulk-like atoms further from the defect are obtained from an empirical potential such as that proposed by Musgrave and Pople (1962). The potential for atom i is

$$\begin{aligned} V_i &= \frac{1}{4}\sum_j k_r^i(\Delta r_{ij})^2 + \frac{r_0^2}{2}\sum_{j>k} k_\theta^{(i)}(\Delta\theta_{jik})^2 + r_0\sum_{k>j} k_{r\theta}^{(i)}(\Delta r_{ij}+\Delta r_{ik})\Delta\theta_{jik} \\ &+ \sum_{k>j} k_{rr}^{(i)}\Delta r_{ij}\Delta r_{ik} + r_0^2\sum_{l>k>j} k_{\theta\theta}^{(i)}\Delta\theta_{jik}\Delta\theta_{kil} \end{aligned} \tag{1.84}$$

where Δr_{ij} is the change in the length of the i-j bond and $\Delta\theta_{jik}$ is the change in angle between the i-j and i-k bonds. The sums are over the nearest neighbour atoms to i. The coefficients k^i are fitted so that experimental properties of the species of atom i are reproduced.

The vibration modes of a defect can be calculated with great accuracy and compared to the results of infra-red (section 2.3) or photoluminescence (section 2.4) experiments possibly providing confirmation of a defect structure. Since the frequencies of vibrational modes are highly sensitive to defect structure, an agreement between calculated and observed modes is a strong indication that the modelled and observed structures are very similar.

Mulliken bond populations

Mulliken analysis is a valuable tool in the analysis of the detailed nature of a Kohn-Sham level. This is especially useful when the Kohn-Sham level corresponds to a defect's band-gap level because in this case the calculated properties can be compared to those obtained from Electron Paramagnetic Resonance (EPR) experiments (section 2.5).

The Mulliken bond population is a measure of the contribution to a Kohn-Sham level λ coming from the basis function ψ_i. This is given by (Pople and Beverage, 1970),

$$p_\lambda(i) = \frac{1}{N_L}\sum_{j,\mathbf{k}} c^{\mathbf{k}}_{\lambda i} S^{\mathbf{k}}_{ij} (c^{\mathbf{k}}_{\lambda j})^* \tag{1.85}$$

where $S^{\mathbf{k}}_{ij}$ are the overlap matrix elements given by,

$$S^{\mathbf{k}}_{ij} = \int B^*_{\mathbf{k}i} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} B_{\mathbf{k}j} d\mathbf{r} \tag{1.86}$$

where $\sum_i p_\lambda(i) = 1$ and as in the previous equations, $\mathbf{k}$ are the special $\mathbf{k}$-points (section 1.8). If a state λ has a strong contribution from the orbital ψ_i then the coefficients $c^{\mathbf{k}}_{\lambda i}$ will be large and hence the bond-population $p_\lambda(i)$ will also be large. By comparing contributions to $p_\lambda(i)$ coming from s, p and d-like basis functions it is possible to obtain the hybridisation of the state. Both the hybridisation and the bond-population can be compared to the results of EPR experiments.

1.13 Summary

This chapter has described the method used to calculate the properties of defects namely their structure and total energy and the properties that may be derived from these. Despite the approximations and simplifications that are necessary to allow systems of appreciable size to be treated, experimentally observable properties of defects can be calculated with remarkable accuracy. It is also noteworthy that with the exception of the local vibrational modes (and the Marker Method for calculating electrical levels), all experimental observables are calculated entirely from first principles. Even in the calculation of local vibrational modes the only empirical input is the mass of the atoms (if a Musgrave-Pople potential is not used). Where theory really comes into its own is in the explanation of experimental results and so in the next chapter many of the relevant experimental techniques will be described with the remainder of this thesis being devoted to the explanation of results obtained from them.

Chapter 2

Experimental Techniques

2.1 Introduction

As described in the previous chapter, density functional theory is capable of calculating a large number of experimental observables. These include defect structures and hence symmetries, electrical energy levels and local vibrational modes to name but a few. Comparing the calculated properties of a defect to experimental observations allows the assignment of a precise structure to an observed defect on the atomic and electronic scale. Such an assignment and the knowledge that comes with it has many benefits, the greatest being the possibility of avoiding "bad" defects by for example, defect engineering. This may involve deliberately creating certain "good" defects at the expense of the bad defects that need to be avoided. The focus of this chapter is the experimental methods most commonly relied upon to ascertain information on an observed defect. The observables that these experiments measure provide the knowledge required to begin the modelling process.

2.2 Deep Level Transient Spectroscopy

Deep level transient spectroscopy (DLTS) is a highly sensitive method for the study of deep levels. The technique has been reviewed in several places. For a complete review of DLTS and deep levels see Mooney (1999, and references therein). There are many extensions to the DLTS technique such as Laplace DLTS (Dobaczewski *et al.*, 1994) but these will not be discussed here. Instead an overview of the techniques will be given to enable the reader to understand results discussed in proceeding chapters.

2.2.1 Deep levels

Impurities or intrinsic defects in semiconductors usually have the effect of introducing levels in the semiconductor band-gap. The levels comes from the fact that a defect will capture and emit electrons and holes at different rates depending upon background doping and temperature, two things that define the Fermi-level (E_F). Thus the charge state of a defect depends upon E_F and the defect level is the energy at which the stable charge state of the defect would change if E_F were swept across the band-gap.

Defect levels located far from the band edges are known as deep levels. Since a large activation energy is required to ionise a deep level, it must be that the trapped electrons or holes are held by a strong potential, localising their wavefunction in the vicinity of the defect. The localisation of the trapped carriers wavefunction in real space means that in k-space the wavefunction will be highly delocalised and hence able to couple strongly with a wide variety of phonons. This means that deep levels often act as non-radiative recombination centres.

The capture rates of electrons (equation 2.1) and holes (equation 2.2) at a defect are given respectively by

$$c_n = \sigma_n \langle v_n \rangle n \tag{2.1}$$

$$c_p = \sigma_p \langle v_p \rangle p \tag{2.2}$$

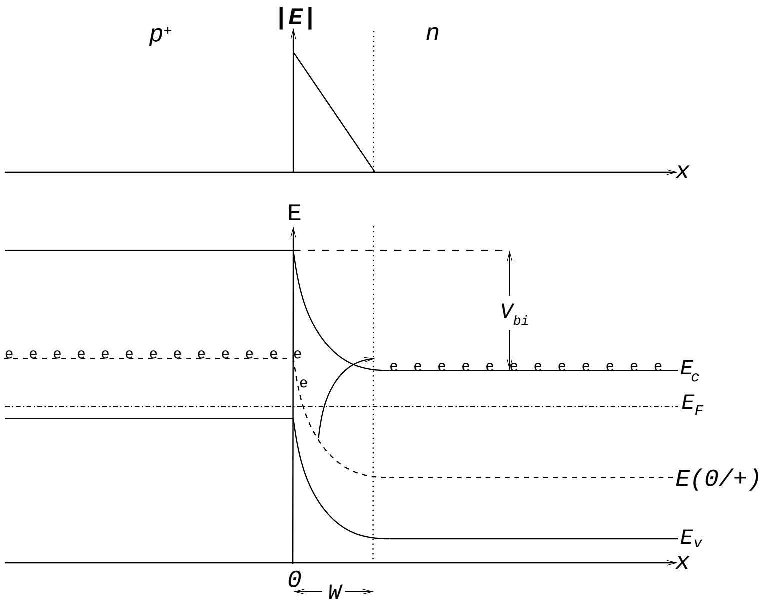

Figure 2.1: A p^+n Schottky barrier diode showing the associated band bending. Above the band structure is a plot of the electric field caused by the band bending on the same x-axis. The n-type region has $x > 0$ while $x < 0$ is p^+-type.

where σ_n and σ_p are the electron and hole capture cross sections respectively and n and p are the electron and hole concentrations respectively. The thermal velocities (or mean Fermi-velocities) of electrons and holes are $\langle v_n \rangle$ and $\langle v_p \rangle$ respectively. For electrons the thermal velocity is given by,

$$\frac{3}{2}kT = \frac{1}{2}m_n^*\langle v_n^2 \rangle \tag{2.3}$$

with

$$\langle v_n \rangle \approx \sqrt{\langle v_n^2 \rangle} \tag{2.4}$$

where k is the Boltzmann's constant, T temperature and m_n^* the effective electron mass. The hole thermal velocity is given by analogous equations.

A defect with an electron-capture rate greater than its hole-capture rate is known as an electron trap. Conversely a defect with $c_p > c_n$ is a hole trap. A defect will act as a recombination centre if both c_n and c_p are large.

2.2.2 Schottky diodes

DLTS experiments probe the space charge region of a Schottky diode or p-n junction. The inherent band bending at the junction has two effects. Due to the constant position of E_F, as the bands bend the Fermi-level effectively shifts across the band-gap. Impurities within the space charge region will then become ionised. The second effect of the band bending is the introduction of an electric field. These two points are illustrated in figure 2.1.

An important property of a Schottky diode is that the width of the space charge region, W, can be varied by applying an external voltage. The width is given by,

$$W = \sqrt{\frac{2\epsilon(V_{bi} + V)}{qN}} \tag{2.5}$$

where V_{bi} is the built in voltage of the junction as shown in figure 2.1, ϵ is the permittivity of the depleted semiconductor, q is the charge of an electron, N is the density of ionised impurities in the space charge region and V is the applied reverse bias. Since the electric field in the space charge region sweeps away any free carriers in around 10^{-12}-10^{-10} seconds a Schottky diode is effectively a variable width insulator. The capacitance of such a device is given by

$$C = \frac{\epsilon A}{W} \tag{2.6}$$

where A is the area of the junction. Since W is dependent upon the density of ionised impurities in the space charge region (equation 2.5), measuring the capacitance of the device provides information on the carrier traps within the depletion region. Specifically, a change in the concentration of carriers trapped at deep levels within the depletion region, caused by the thermal emission of carriers to the appropriate band edge, can be studied by monitoring the change in capacitance with time at a constant voltage.

2.2.3 The filling and emptying of traps

The thermal emission rate of carriers from a trap to the appropriate band edge is defined by a Boltzmann distribution. For electrons, emission to the conduction band occurs at a rate,

$$e_n = A_n \exp(-\Delta E/kT) \tag{2.7}$$

where ΔE is the Gibbs free energy required for the emission of the electron from the trap level (E_T) to the conduction band (E_c). ΔE is given by

$$\Delta E = E_c - E_T = \Delta H - T\Delta S \tag{2.8}$$

where ΔH and ΔS are the change in enthalpy and entropy resulting from the emission of the carrier.

The pre-exponential factor A_n is given by

$$A_n = \frac{\sigma_n \langle v_n \rangle N_c}{g} \tag{2.9}$$

where N_c is the effective density of states of the conduction band and g is the degeneracy of the defect level.

Hence

$$e_n = \frac{\sigma_n \langle v_n \rangle N_c}{g} \exp(\Delta S/k) \exp(-\Delta H/kT). \tag{2.10}$$

Since $\langle v_n \rangle \propto T^{1/2}$ and $N_c \propto T^{3/2}$, if $\ln(e_n/T^2)$ is plotted against $1/T$ the gradient will give ΔH and the y-axis intercept provides information on σ_n. Any graph, such as this, where $\ln x$ is plotted against $1/T$ is known as an Arrhenius plot. Since σ_n is typically temperature dependent the value of σ_n derived from an Arrhenius plot will often differ from that found by a more direct measurement. The Arrhenius plot is often referred to as the signature of a deep level since it should be entirely independent of experimental setup. Clearly similar equations to those above may be derived for the emission of a trapped hole to the valence band.

To measure the transient thermal emission of carriers from the deep levels the traps must first be filled. This is accomplished by a filling pulse which narrows the width of the depletion region, allowing carriers to enter and be trapped at deep levels. After the pulse the depletion region returns to its original width as traps are thermally ionised at a rate

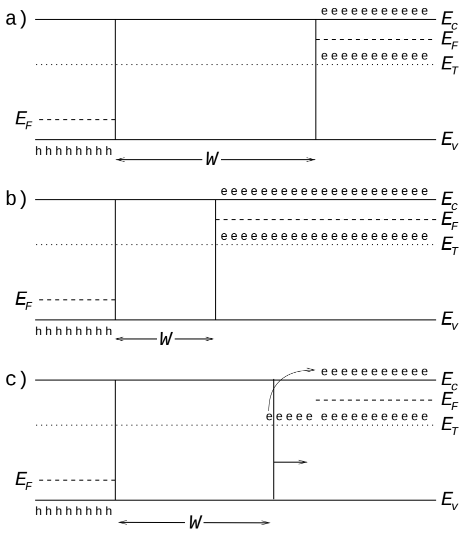

Figure 2.2: Diagram a shows the depletion region under steady-state reverse bias when all traps are empty. During the voltage pulse the size of the depletion region (W) is reduced and traps are filled as shown in diagram b. After the pulse, W returns to its steady-state value as the traps thermally empty as shown in diagram c. Carriers are removed by the electric field within the depletion region. Note that for clarity band bending (see figure 2.1) has not been shown in this figure.

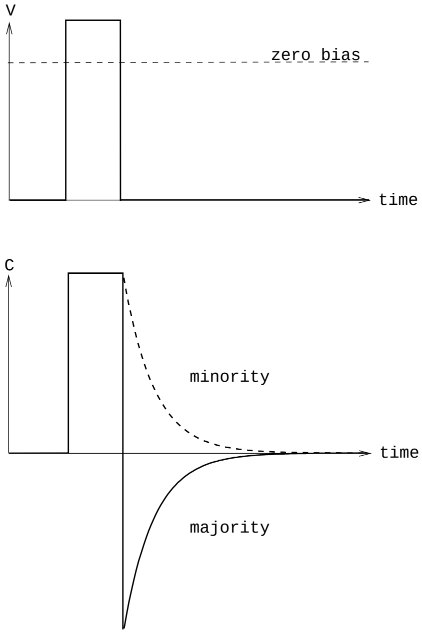

Figure 2.3: The different capacitance transients of minority and majority carriers. The filling pulse is shown above on the same time scale.

given by equation 2.7 returning the system to equilibrium. This is shown schematically in figure 2.2. The electric field within the space charge region means that ejected carriers are quickly swept away and re-trapping can be neglected. The thermal emission of carriers from deep levels results in the capacitance transient, the direction of which reveals whether the trap is a minority or majority carrier trap as shown in figure 2.3.

The Poole Frenkel effect

The Poole Frenkel effect is a manifestation of the fact that the emission of a negative carrier to leave a positive trap (or vice versa) will be assisted by the electric field within the depletion region (see figure 2.1), enhancing the thermal emission (Hartke, 1968). This is observable as a dependence of ΔH upon field strength and equivalently V. The observation of a Poole Frenkel effect gives valuable information on the nature of the trap: It implies that the trap is neutral when occupied and charged when empty (*i.e.* a $(0/+)$ like electron trap or a $(-/0)$ like hole trap).

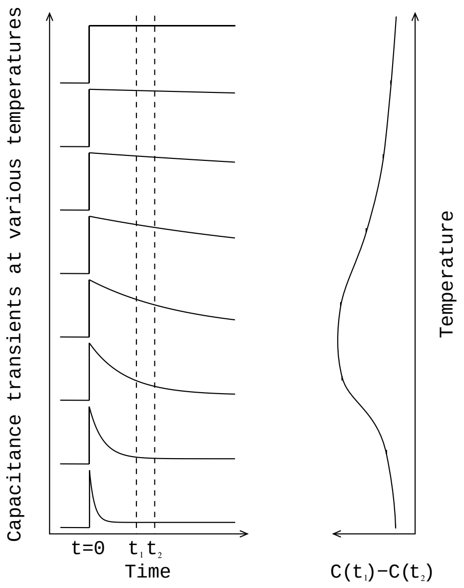

Figure 2.4: The double boxcar technique for analysing capacitance transients, introduced by Lang. The rate window is defined by t_1 and t_2, the times at which the capacitance is measured to give ΔC, plotted on the right hand side.

2.2.4 The double boxcar technique

The DLTS technique was pioneered by Lang who introduced the concept of using rate windows (Lang, 1974), turning the analysis of capacitance transients into a powerful spectroscopic technique. The technique involves repeatedly pulsing the diode to produce a series of capacitance transients at different temperatures. The value of C is measured at two specific times, t_1 and t_2 which define the rate window. If we consider a pulse at $t = 0$ then $C(t_1)$ is the capacitance after a time t_1 and $C(t_2)$ the capacitance after a time t_2. The value of interest is $\Delta C = C(t_1) - C(t_2)$ [1]. Scanning over temperature varies the time constant of the thermal emission of carriers and hence at some temperature the thermal emission rate and the instrument rate window will produce a maximum in ΔC. This is shown schematically in figure 2.4. This set-up is called a dual gated signal averager, more

[1] Note that the capacitance transient is expected to be exponential since it depends upon the thermal emission of carriers from a single level at a given temperature. Hence $C(t)$ can be defined by a measurement of C at just two points in time.

commonly known as the double boxcar technique. The dependence of the emission rate upon temperature is required to produce an Arrhenius plot. This dependence can be found by varying the rate window, which shifts the DLTS peak to a different temperature and hence, in a standard DLTS experiment, a series of scans is required to find the signature of a given deep level.

2.3 Infra-red Spectroscopy

2.3.1 The origin of vibrational modes

A point defect placed in an otherwise perfect crystal will break the translational symmetry and hence modify the material's modes of vibrations. It is possible that these modified modes will still lie within the band of lattice modes and hence the vibrational energy will be transmitted throughout the crystal. Alternatively the modified modes may be above the maximum lattice frequency ω_{max} (also known as the Debye frequency which for silicon coincides with the Raman frequency) or within the gap in frequencies of the lattice optical and acoustic modes if this gap exists in the given material. Both of these types of modified modes will be highly localised spatially at the defect. A third alternative is that the modified modes are located at a frequency where the phonon density of states is a minimum giving what is know as a pseudo-localised mode or in band resonance. In silicon there is no gap between the acoustic and optical bands but there is a minimum in the density of states in the region of 250 cm^{-1}. Hence gap modes will not exist but in band resonances may be present in addition to true localised modes. These modes may be observed via the interaction of applied infra-red radiation with the dipole moment of the defect. This technique is called infra-red spectroscopy and good reviews on the topic have been given by Newman (1973) and Stavola (1999) among others.

2.3.2 The removal of free carriers

Free carriers will be present in a semiconductor at a finite temperature. It is crucial that their concentration is kept low enough that the infra-red radiation is absorbed at defects rather than by the free carriers. A successful technique for reducing free carrier absorption is that of counter doping. In the case of the acceptor, co-doping with a donor or vice versa will greatly reduce the number of free carriers. Alternatively an electron or neutron irradiation may be performed to create lattice damage which produces deep levels where remaining free carriers will be trapped (Smith and Angress, 1963). It is important to remember in these cases that the impurity of interest may interact with the co-dopant and/or radiation damage (Newman and Smith, 1967; Newman, 1973).

2.3.3 Local vibrational modes

Impurity atoms having a mass less than that of the host crystal atoms are likely to exhibit local vibrational modes, *i.e.* modes with angular frequency greater than ω_{max}. By considering a simple one-dimensional linear chain it can be shown that the angular frequency of the mode is proportional to $1/\sqrt{m'}$ where m' is the mass of the light impurity (Newman, 1973), a finding that holds true for a three-dimensional crystal. An extension of the simplistic one-dimensional linear chain to three-dimensions has been performed by several authors (Dawber and Elliot, 1963*a*,*b*; Elliot and Pfeuty, 1967) and is reviewed by Newman (1973). Only the results will be summarised here. One important result of this analysis is that integrated absorption due to an impurity of mass m' in a host crystal where the atoms have mass m ($m' < m$) can be shown to be given by,

$$\int \alpha(\omega)d\omega = \frac{2\pi^2 D\eta^2}{nc}|\chi(0)|^2 \tag{2.11}$$

where D is the concentration of impurities of mass m', c is the velocity of light in vacuum, n is the refractive index of the crystal and η is the apparent electric charge of the impurity. In equation 2.11 the integrated absorption can be considered as the absorption due to an atom of mass m' vibrating with frequency ω in a fixed potential well, reduced by a factor

$m'|\chi(0)|^2$ which reflects the non-zero amplitude of the vibration on neighbouring atoms. $|\chi(0)|^2$ is given by,

$$|\chi(0)|^2 = \frac{1}{m}\left(\epsilon^2 z'^2 \int \frac{\nu(\mu)}{(z'-\mu)^2} d\mu - \epsilon\right)^{-1} \tag{2.12}$$

where ϵ is the mass defect $(m - m')/m$, z' is ω_L^2 (ω_L is the angular frequency of the local mode), $\nu(\mu)$ is the density of unperturbed states and μ is the electric dipole moment induced by the oscillation.

By considering the effect of a small displacement of the atoms from their equilibrium position Q on the dipole moment, the dipole moment may be written as,

$$\mu = \mu_e + \left(\frac{d\mu}{dQ}\right)_e Q + \left(\frac{d^2\mu}{dQ^2}\right)_e \frac{Q^2}{2} + ... \tag{2.13}$$

The probability of an optical transition from a quantum state i to j is given by $\langle i|\mu|j\rangle$. For a static moment there will be no coupling to vibrational states and the dipole moment will contain only a rotational component μ_e. The strength of the fundamental vibrational transitions will be proportional to $[(d\mu/dQ)_e]^2$ implying that the magnitude but not the sign of $(d\mu/dQ)_e$ may be determined experimentally. The quantity $(d\mu/dQ)_e$ is the same as the apparent charge η appearing in equation 2.11. The strength of the second harmonic transitions is proportion to $[(d^2\mu/dQ^2)_e]^2$ with higher order harmonics given by the higher order terms of equation 2.13.

2.3.4 Symmetry determination

A light, substitutional impurity having cubic symmetry will give rise to a triply degenerate local vibrational mode due to the degeneracy of the x, y and z directions. If the symmetry is lowered, the equivalence of these directions is removed and hence the degeneracy of the states is reduced. For example, if the symmetry is lowered to tetragonal or trigonal the triply degenerate state will split into a doublet and a singlet. If the symmetry is further lowered, for orthorhombic symmetry or lower, the states will be split into three non-degenerate states.

Axial symmetries cannot be determined within the harmonic approximation as all give rise to two fundamental modes. However, anharmonic terms allow the admixing of wavefunctions which permits transitions between the ground and higher excited states. The number of allowed fundamental, second harmonic and higher harmonic transitions depends upon the point group of the oscillator and hence the distinction between different axial symmetries is possible.

2.3.5 Isotope effects

Isotope effects are observed strongly by infra-red spectroscopy since the frequency of a local mode depends upon the atoms mass, as described in section 2.3.3. By studying material that is isotopically enriched, from the observed shift in a local mode with isotope it is often possible to demonstrate the presence of a species in the defect to which that mode belongs. Further benefits of isotope studies are discussed in section 2.4.5 but also apply to infra-red spectroscopy. The power of infra-red spectroscopy in collaboration with isotopic studies is demonstrated nicely in section 4.2.

2.4 Photoluminescence Spectroscopy

2.4.1 Optical absorption and emission

Photoluminescence (PL) is an extremely sensitive technique which probes the excited states of defects in semiconductors, often allowing unambiguous determination of the atomic species and giving information on the defect structure. Before discussing photoluminescence in any detail it is useful to review the processes of optical absorption and emission. For a complete review of optical techniques and photoluminescence the reader should see Davies (1999).

For absorption, light is passed through a crystal and transmitted light is examined spectroscopically. A dark line in the transmitted spectrum has a frequency that corresponds to the energy of an optical transition between two electronic states within the crystal. The process of optical emission is essentially the same apart from that the light is emitted at the given energy rather than absorbed. However emission is complicated by processes where some of the energy may be emitted non-radiatively, for example in the form of phonons.

If we consider the simple case where no light is externally or internally reflected, the intensity of transmitted light will be given by,

$$I_t = I_0 \exp(-\mu t) \tag{2.14}$$

where I_0 is the intensity of the incident light, t the sample thickness and μ the absorption coefficient.

It can be shown, by considering the detailed balance of absorption and luminescence processes, that the concentration of absorbing centres N is given by,

$$N = \frac{g_f}{g_i} \frac{q}{\pi^2} \frac{E^2 \tau}{\hbar^3 c^2} \frac{n^2}{(n^2+2)^2} \int \mu(E) dE \tag{2.15}$$

where g_i and g_f are the degeneracies of the initial and final electronic states respectively, E is the energy of the transmitted light and τ the radiative lifetime of the transition. An important implication of this equation is that

$$N \propto \int \mu(E) dE. \tag{2.16}$$

Thus absorption can give the relative number of defects although the absolute number will usually be unknown.

2.4.2 Photoluminescence

Above band-gap photoluminescence

Crystals typically have a high absorption coefficient for above band-gap light and hence the energy is deposited close to the surface, typically within a few micrometres. Since this energy is larger than the crystal's band-gap the deposited energy creates electron hole pairs (excitons). In Si the binding energy of the electron and hole is small ($\sim 14.3 \pm 0.5$ meV (Shaklee and Nahory, 1970)) making it probable that many of the excitons will become ionised rather than immediately recombining. Furthermore, since Si has an indirect band-gap, electrons and holes have different wave-vectors and so their recombination requires the participation of a phonon which reduces the probability of recombination occurring. Thus excitons in Si can have a considerable lifetime ($\sim 60\ \mu$s Merle *et al.* (1978)) giving them time to diffuse to defect centres where they may be trapped. Trapping at defects reduces the lifetime of excitons via a few possible mechanisms, but the most simple comes from the fact that a localisation of the trapped electron and hole in real space implies a delocalisation of electron and hole in k-space making recombination possible without the necessity of a phonon. The recombination energy will place the centre in an excited electronic state and the centre may then relax to its ground-state via the emission of a photon of specific energy; thus luminescence characteristic of the centre is observed.

It is important to note that since the input energy may be captured by a variety of different types of defect centres the amount of energy dedicated to the excitation of the optical centre of interest is unknown. The implication of this is that the luminescence intensity of a centre in different samples is generally not directly proportional to its concentration. This is in contrast to absorption where the energy is directly absorbed by the optical centre at its characteristic frequency. Furthermore the luminescence intensity may be reduced by de-excitation via non-radiative processes.

Below band-gap photoluminescence

A perfect crystal will not absorb below band-gap photons; however, such photons may be absorbed by defects directly. Since there is no intermediate process between absorption of the photon and excitation of the centre the absorption as a function of the polarisation of incident light may be studied. This is discussed in more detail in section 2.4.6.

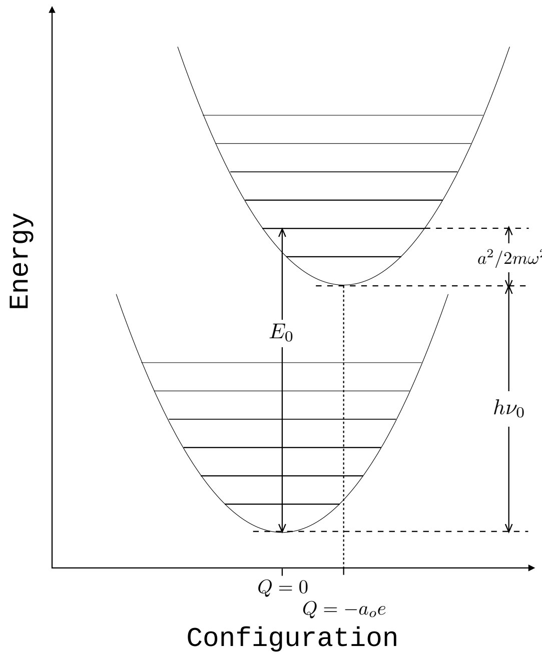

Figure 2.5: A configuration coordinate diagram showing the ground and excited states of a PL transition. The zero phonon line has energy $h\nu_0$.

2.4.3 The photoluminescence spectrum

Zero-phonon line

The zero phonon line is the main signature of a defect in photoluminescence and corresponds to the purely electrical transition with no phonon involvement. The zero phonon energy is shown schematically in figure 2.5 and is discussed in more detail below.

2.4.4 Local vibrational modes

The energy of an electronic state, E, that is sufficiently separated in energy not to interact with other electron states, varies in a linear way with applied strain e for small strains, that is $\Delta E = Ae$. The position of the atoms is labelled Q where $Q = 0$ corresponds to the equilibrium positions of the atoms in their electronic ground state. Now if the system is excited to a state at energy E_0 the total energy can be minimised by moving the atoms from $Q = 0$ to $Q = -a_0 e$ where a_0 is the inter-atomic spacing. This lowers the total energy by $\Delta E = -Ae = aQ$ where $a = A/a_0$, shown in figure 2.5. The amount by which the energy can be lowered is constrained by the elastic energy, kQ^2 where k is a constant in this simple model. The energy of the excited state is then given by,

$$V_e = E_0 + kQ^2 - aQ = E_0 + k(Q - a/2k)^2 - a^2/4k. \tag{2.17}$$

It is now seen that the excitation leads to simple harmonic motion of the atoms around an equilibrium point at $Q_0 = a/2k = a/m\omega^2$ where m is the mass of the vibrational mode. Equation 2.17 shows that different excited states will have vibrational modes centred at different equilibrium positions. This is known as electron-phonon coupling. Note that anharmonicity and vibronic coupling (the process of vibrational modes causing the interaction of electronic states) will modify equation 2.17 and have been neglected in this simple discussion.

According to the Born-Oppenheimer approximation, the wavefunction of a quantum state ψ, can be separated into an electronic and vibrational part, ϕ and χ respectively. Hence we can write the ground and excited states of a system as,

$$\psi_{g,n} = \phi_g(r)\chi_n(Q), \tag{2.18}$$

$$\psi_{e,N} = \phi_e(r)\chi_N(Q - Q_0) \tag{2.19}$$

where the subscripts g and e denote ground and excited states and n and N are their respective quantum numbers. The probability of an electric-dipole transition between states is given by,

$$\begin{aligned} I_{e,N,g,n} &= \iint \psi^*_{e,N} \mathbf{r} \psi_{g,n} dr dQ \\ &= \int \phi^*_e(r) \mathbf{r} \phi_g(r) dr \int \chi_N(Q - Q_0)\chi_n(Q) dQ \\ &= C \int \chi^*_N(Q - Q_0)\chi_n(Q) dQ \end{aligned} \tag{2.20}$$

where we have considered the electronic contribution as a constant C. From the orthogonality of states of different energies, in the case where $Q_0 = 0$ (*i.e.* the equilibrium atomic positions are unchanged by the excitation),

$$\int \chi_n^*(Q)\chi_N(Q)dQ = \delta_{N,n} \tag{2.21}$$

where $\delta_{N,n}$ is the Kronecker delta.

If we now consider absorption at low temperature so that $n = 0$ the squared overlap integral is,

$$\left| \int \chi_N^*(Q - Q_0)\chi_0(Q)dQ \right|^2 = \frac{S^N \exp(-S)}{N!} \tag{2.22}$$

where S is the known as the Huang-Rhys factor which has the physical significance of being the relaxation energy, $a^2/2m\omega^2$ in units of $\hbar\omega$.

It can be seen in figure 2.5 that the zero phonon line (the transition from n to N) corresponds to an energy $h\nu_0 = E_0 - a^2/2m\omega^2$. From equation 2.22 the intensity of this transition is $I_0 = I_t \exp(-S)$ where I_t is the total intensity of the absorption band. Similarly the intensity of the one-phonon sideband which occurs $\hbar\omega$ higher in energy ($n = 0$, $N = 1$) is $I_1 = I_t S \exp(-S)$ from equation 2.22.

2.4.5 Isotope effects

Isotopes play a key role in the identification of defects using photoluminescence. It is often possible to correlate the relative intensities of a series of lines with the abundances (natural or artificially enriched) of stable isotopes of a given atomic species. Thus the possible presence of that species within the optical centre can be demonstrated. Furthermore the number of components to the splitting of the spectral lines, when more than one isotope of a species is present, reveals the lower limit to the number of atoms of that species within the centre and may reveal whether they are at equivalent or inequivalent sites.

2.4.6 Selection rules

The probability of a dipole transition from a state ψ_1 to ψ_2 is given by,

$$P = \langle \psi_1 | \mathbf{r} | \psi_2 \rangle \tag{2.23}$$

where $\mathbf{r}$ is the direction of the electric field of the exciting light. For example, for the simple case of excitation from an s state to a p state by light of electric field polarised in the x direction, $\langle s|\mathbf{x}|p_x\rangle \neq 0$ whereas $\langle s|\mathbf{x}|p_y\rangle = 0$ and $\langle s|\mathbf{x}|p_z\rangle = 0$. It is important to note that group theory can state when the probability of a transition is zero but it cannot give the magnitude of a non-zero probability which may be so small that it is effectively zero.

These selection rules may lead to a case where a centre will absorb photons with an electric field component only along a specific crystallographic direction. In this situation centres may be excited with polarised light and the polarisation of the luminescence analysed. Let us consider a defect which absorbs light only with an electric field polarised in $\langle 100\rangle$. Then the light having electric field along [100] will be absorbed only by centres aligned along that direction. Assuming that there is enough direct absorption at the centres and that they luminesce by the emission of a single photon so that the polarisation is conserved, the intensity of the luminescence will vary as $\cos^2\theta$ where θ is the angle between the [100] axis and the electric field of the emitted light.

2.4.7 Symmetry determination

Photoluminescence in conjunction with the application of uniaxial stress to a material can provide a large amount of reliable information on the symmetry of the optical centre of interest. The stress applied to a cubic crystal can be defined by six elements of the stress tensor s_{ij} where i and j run over the cubic axis directions x, y and z. Applying a stress to the crystal along a $\langle 100\rangle$ direction results in a s_{xx} type stress and applying stress in a $\langle 111\rangle$ direction gives shear s_{xy} type stresses.

For small stresses the ground and excited electron states will not interact with other electronic states. If such a stress is applied in a direction which reduces the symmetry of the centre the energy level for the transition will split by an energy proportional to the stress (for non-interacting electronic states). Hence the symmetry group of the centre may be determined but since there is a requirement for the stress to be applied in equal and opposite directions (for the crystal to be static), centres with inversion symmetry cannot be separated from those without. This means from stress perturbations alone it may be possible to determine that a defect has trigonal symmetry but it will not be possible to determine whether the point group is, for example, C_{3v} or D_{3d}. The differentiation of point groups with and without inversion symmetry may be made using an electric field in addition to the stress. A good discussion of this technique has been given by Davies (1999)

2.5 Electron Paramagnetic Resonance

Electron Paramagnetic Resonance or EPR is an extremely powerful tool in the study of defects in semiconductors due to the rich information it is able to provide, even more so when the extended form of EPR known as Electron Nuclear Double Resonance (ENDOR) is applied. EPR gives not only the electrical properties of a defect but also much information on the defect structure, chemical composition and even the electron wave function. The main limit to EPR is that it relies on the interaction of the defect with a magnetic field and hence the defect being studied must be paramagnetic. Since the main defects of interest in semiconductors are those that are electrically active it is often the case that in at least one of a defects stable charge states it will have an unpaired electron and hence be paramagnetic and so in principle visible to EPR and ENDOR. A good review of the EPR and ENDOR techniques as well as examples of their application can be found in Watkins (1998, and reference therein).

2.5.1 Experimental Setup

EPR studies the transitions between Zeeman levels induced at a defect by an external field. These transitions usually lie within the microwave region of the electromagnetic spectrum and so the sample is placed within a microwave cavity where there is an intense and constant microwave field. A variable magnetic field is then applied to the sample. When the magnitude of the magnetic field $|\mathbf{B}|$, is such that a Zeeman transition coincides with the energy of the fixed microwave radiation, a resonance will occur. This resonance is detected either by a decrease in the quality factor Q of the cavity (absorption) or by a small shift in the cavity's resonant frequency (dispersion).

2.5.2 The g-tensor

The energy of the Zeeman interaction is given by the Hamiltonian,

$$\hat{\mathcal{H}} = -\mu.\mathbf{B} \tag{2.24}$$

where μ is the electronic magnetic dipole moment of the defect, induced by electron spin $\mathbf{S}$ and electron orbital angular momentum $\mathbf{L}$. That dipole moment μ is given by,

$$\mu = -g_s \mu_B \mathbf{S} - g_L \mu_B \mathbf{L} \tag{2.25}$$

where μ_B is the Bohr magneton $(e\hbar/2mc)$, $g_s = 2.0023$ and $g_L = 1$.

In semiconductors and solids in general, $\mathbf{L}$ is often suppressed and hence μ predominantly originates from $\mathbf{S}$ with the additional spin-orbit term $\lambda\mathbf{L}.\mathbf{S}$ which can be considered as being effectively due to the spin alone. Hence μ may be written as

$$\mu = -\mu_B \mathbf{S}.\mathbf{g} \tag{2.26}$$

where $\mathbf{g}$ is the symmetric tensor $\mathbf{g} = 2.0023 + \Delta\mathbf{g}$. The small orbital component of the dipole moment due to the spin-orbit interaction has been introduced via the term $\Delta\mathbf{g}$. The Hamiltonian then becomes,

$$\hat{\mathcal{H}} = \mu_B \mathbf{S}.\mathbf{g}.\mathbf{B} \tag{2.27}$$

and the energy levels are split due to the Zeeman effect, giving levels shifted by

$$E(M) = g\mu_B BM \tag{2.28}$$

where M is the azimuthal spin quantum number which takes the values $-S \leq M \leq +S$ giving $(2S+1)$ equally spaced levels. From equation 2.28 it is seen that for a defect with $S = 1/2$ a resonance will occur when $B = h\nu_0/g\mu_B$. This is shown schematically in figure 2.6.

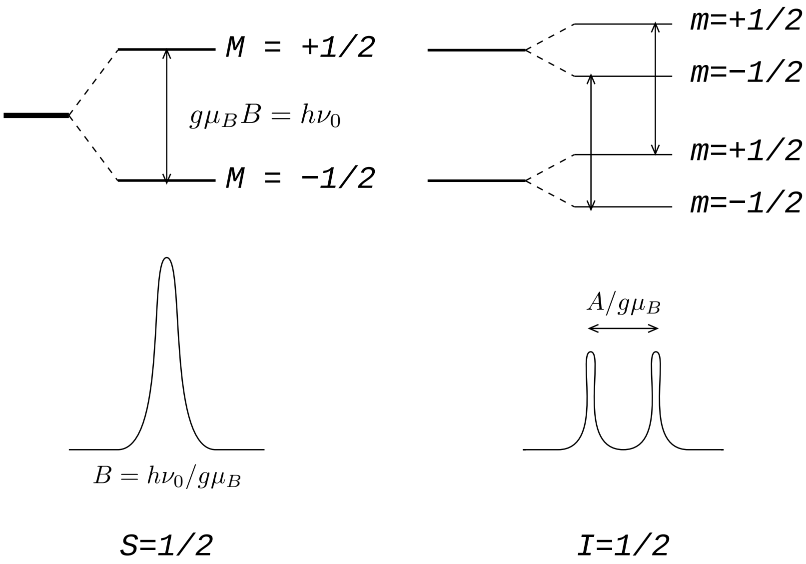

Figure 2.6: The allowed EPR transitions for a system with electron spin $S = 1/2$ (left) and with nuclear spin $I = 1/2$ (right). The curves show the resonance as a function of magnetic field while the lines above show the transitions between levels that have been split due to the Zeeman effect. For ENDOR there are also transitions allowed for $\Delta M = 0$, $\Delta m = 1$.

The g-tensor is the principal quantity studied by EPR and its angular dependence is given by,

$$g = \left(\sum_i g_i^2 n_i^2\right)^{\frac{1}{2}} \tag{2.29}$$

where g_i are the principle values of g and n_i are the directional cosines of $\mathbf{B}$ with respect to the principle axes of the g-tensor ($i = 1, 2, 3$). Thus by varying the magnetic field orientation the angular dependence of g may be determined. The resulting splitting of levels as the magnetic field is directed away from directions of high crystal symmetry provides information on the symmetry of the defect. EPR spectra characteristic of several common defect point group symmetries are shown in figure 2.7 for a cubic crystal.

The electronic structure of the defect can be determined from Δg using second order perturbation theory which gives,

$$\Delta g = -2\lambda \sum_n \frac{\langle 0|\mathbf{L}|n\rangle\langle n|\mathbf{L}|0\rangle}{E_n - E_0} \tag{2.30}$$

where n denotes the excited states.

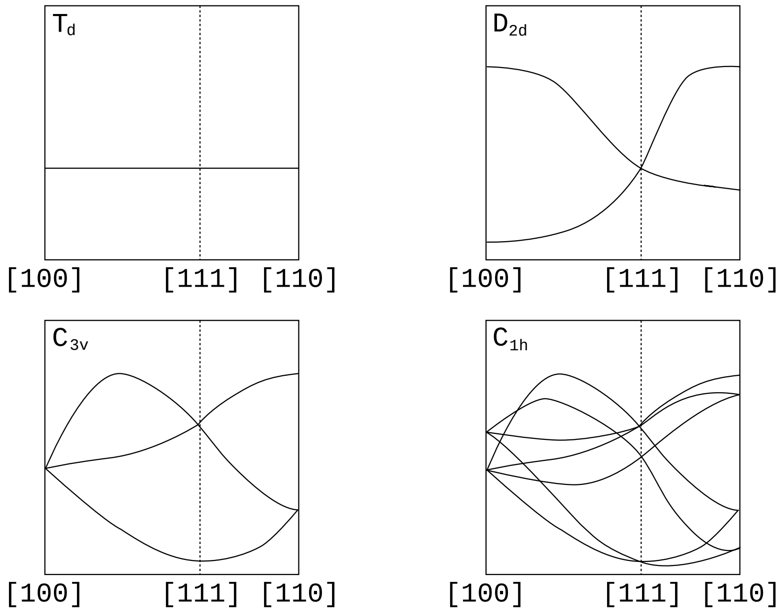

Figure 2.7: The characteristic angular dependencies of EPR signals for some common point group symmetries in a cubic semiconductor.

2.5.3 Hyperfine Interaction

So far within this section the only magnetic field considered has been the applied magnetic field B_{app}. There is however, likely to be an additional magnetic field experienced by the paramagnetic electron (B_{loc}) due to the magnetic moment induced by the nuclear spins ($\mathbf{I}$) of nearby atomic nuclei. Hence strictly we should write,

$$\mathbf{B} = \mathbf{B}_{app} + \mathbf{B}_{loc}. \tag{2.31}$$

This will introduce an additional term in the Hamiltonian given in equation 2.27 giving,

$$\hat{\mathcal{H}} = \mu_B \mathbf{S}.\mathbf{g}.\mathbf{B} + \sum_j \mathbf{S}.\mathbf{A}_j.\mathbf{I}_j \tag{2.32}$$

where the sum is over the nearby nuclei. The tensor $\mathbf{A}_j$ is the magnitude and angular dependence of the nuclear dipole for each nucleus.

The $(2I_j + 1)$ allowed energy levels are found by first order perturbation theory to lie at,

$$E(M, m_j) \simeq (g\mu_B B + \sum_j A_j m_j)M \tag{2.33}$$

where m_j is the nuclear azimuthal quantum number. The condition for EPR resonance is now $\Delta M = \pm 1$ and $\Delta m_j = 0$ and hence a resonance will occur when

$$B = \frac{1}{g\mu_B}(h\nu - \sum_j A_j m_j). \tag{2.34}$$

The corresponding transitions are shown on figure 2.6.

2.5.4 Wavefunction analysis

Further information on the electron wavefunction can be inferred from the angular dependence of A_j. In analogy to equation 2.29 A_j can be expressed as,

$$A_j = \left(\sum_\alpha A_{j\alpha}^2 n_{j\alpha}^2\right)^{\frac{1}{2}} \tag{2.35}$$

where $A_{j\alpha}$ and $n_{j\alpha}$ are the principle values of $A_{j\alpha}$ and the directional cosines with respect to $\mathbf{g}.\mathbf{B}$ respectively for each axis α and each nucleus j. By substituting terms for the electronic and nuclear dipole moments and adding a Fermi contact term it is possible to write $\mathbf{A}$ as a function of the electron wavefunction, ψ, at the nucleus ($r = 0$),

$$A_{ij} = g g_N \mu_B \mu_N \left\{ \left\langle \frac{3x_i x_j}{r^5} - \frac{\delta_{ij}}{r^3} \right\rangle + \frac{8\pi}{3}|\psi(0)|^2 \delta_{ij} \right\}. \tag{2.36}$$

Often the wavefunction is considered as a linear combination of atomic orbitals,

$$\psi = \sum_j \eta_j \psi_j \tag{2.37}$$

where the sum is over each atomic site j and η_j^2 represents the fraction of the charge density localised at that site. Then for each site,

$$\psi_j = \alpha_j(\psi_s)_j + \beta_j(\psi_p)_j + \ldots \tag{2.38}$$

where $(\psi_s)_j$ and $(\psi_p)_j$ are the s and p components of ψ at atomic site j and α_j^2 and β_j^2 give the fraction of ψ forming the s or p orbital. The value of η_j^2 is given by the hyperfine interaction at site j while α_j^2 and β_j^2 are given by the isotropic and anisotropic components of the hyperfine interaction respectively.

2.5.5 Fine structure terms

When $S > 1/2$ higher order fine structure may appear. This gives rise to an additional term within the Hamiltonian of the form

$$\mathbf{S}.\mathbf{D}.\mathbf{S} \tag{2.39}$$

which originates from the spin-spin and spin-orbit interactions between electrons.

In the case where $D \ll g\mu_B B$ a term in M^2 is added to equation 2.33 which now becomes,

$$E(M, m_j) \simeq (g\mu_B B + \sum_j A_j m_j)M + \frac{3}{2}M^2 \sum_\gamma D_\gamma n_\gamma^2 \tag{2.40}$$

where D_α are the principle values of D while n_α are the directional cosines of the principle axes with respect to $\mathbf{g}.\mathbf{B}$. The condition for resonance now becomes,

$$B(M \rightarrow M-1) = \frac{1}{g\mu_B}(h\nu_0 - \sum_\alpha A_\alpha^2 n_\alpha^2 - \frac{3}{2}(2M-1)\sum_\gamma D_\gamma^2 n_\gamma^2). \tag{2.41}$$

2.5.6 ENDOR

Electron nuclear double resonance spectroscopy is a relatively simple extension to the EPR technique. The experimental setup is as described in section 2.5.1 but with the addition of a coil either wrapped around or placed within the cavity. Once an EPR resonance is found as described above, the magnetic field **B** is held constant and a radio frequency is induced by the coil. The radio frequency is used to induce a resonance with the nuclei of atomic species close to the defect. Thus resolution of hyperfine interactions (described in section 2.5.3) is greatly increased.

The added resolution has several advantages but perhaps the most significant is that the nuclear interaction with the external field may be studied. This often makes it possible to unambiguously identify the chemical species of atoms at the defect.

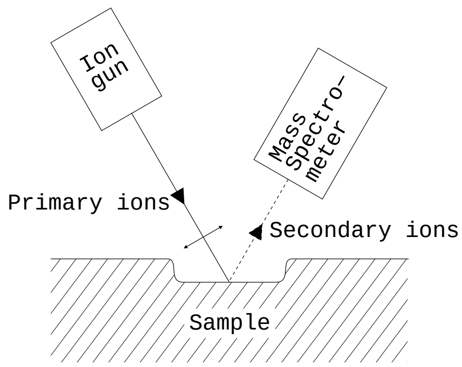

Figure 2.8: A schematic diagram of a dynamic SIMS experimental setup where the primary ion beam is rastered over the sample surface to create a flat bottomed crater of well controlled depth. The crater sizes produced in SIMS range from around 5×5 μm to around 500×500 μm.

2.6 Secondary Ion Mass Spectrometry

Secondary Ion Mass Spectrometry (SIMS) is a sensitive characterisation technique, especially useful for analysing dopant depth profiles. The basic principles of the technique presented here have been discussed in more detail by Clegg (1990) as well as Benninghoven *et al.* (1987).

2.6.1 Experimental setup

The basic principle of SIMS is rather simple. The sample of interest is bombarded with ions having energies of 1-20 keV. Upon collision with the sample, these primary ions will sputter secondary ions from the surface. The secondary ions are then extracted to a mass spectrometer which determines their mass to charge ratio. The experiment is shown schematically in figure 2.8. Since the surface will be sputtered away as a function of time it may be possible to use the evolution of the mass peak of an element of interest to evaluate the depth profile of that element. In order to ensure that the surface is reduced in a well controlled manner, the ion beam is usually rastered over the surface. Furthermore, to prevent errors due to the counting of particles sputtered from the crater edge, electronic or optical gating is employed. This either physically restricts the flow of these ions into the mass spectrometer or electronically removes their signal ensuring that only secondary ions from the flat crater centre are included in the SIMS profile.

2.6.2 Projectile-target interaction

Collision cascades and intermixing

The processes by which primary ions interact with the sample and secondary ions are produced are more complicated than they may first seem and hence deserve further discussion. The deposition of the projectiles energy within the target can be described by linear cascade theory (Sigmund, 1969) where the incident particle undergoes a series of binary collisions with atoms within the target. The target atoms, which initially are at rest, receive large amounts of recoil energy and these atoms then undergo their own binary collisions resulting in a collision cascade continuing until the energy imparted by each collision becomes less that the atomic displacement energy ($\sim$ 10 eV). If a collision cascade intersects the sample surface and the recoiling particle (which may be an atom or a group of atoms) has been given enough energy to overcome the surface binding energy it will be sputtered from the sample. Sputtered particles typically have energies of a few eV and originate from within the first two or three atomic layers from the surface.

This complicated interaction of projectile and target has several consequences, the main one being that atoms within the sample will generally be moved deeper into the target by the collisions. If the SIMS analysis is being performed to study the depth profile of a given impurity, that impurity will appear to have been present deeper within the sample that it actually was. In other words the depth profile of the impurity will be broadened. This effect is known as intermixing and can be minimised by lowering the beam energy and increasing the bombardment angle with respect to the surface normal. A second consequence of the projectile target interaction is that a majority of the primary ions will remain in the target, changing the composition of the sample surface. Crystal orientation plays an important role in these phenomena since the penetration depth of implanted ions is drastically increased by channelling when the ions are implanted into a crystal along a low-index crystallographic axis.

Secondary ion formation

The process by which sputtered particles acquire charge is complex and only qualitative trends shall be discussed here. It is known that an electronegative element in the surface region of the sample will increase the fraction of particles sputtered in the form of positively charged ions. Conversely the presence of an electropositive species in the surface region increases the fraction sputtered in the negative charge state. This effect is likely to be due to the strong dependence of a sample's work function on the concentration of electronegative or electropositive species in the near surface region (Williams and Evans, 1978). It is for precisely this reason that highly electronegative oxygen and electropositive caesium are the species most often used as primary ions.

2.6.3 Depth profiling

As mentioned in section 2.6.1, SIMS is a powerful tool for investigating the depth profiles of dopants or impurities. This is done by converting the raw SIMS output, counts of an atomic species i as a function of time, into the concentration c_i of element i as a function of depth z. The count rate for a species i is given by,

$$I_i = \dot{z}\beta_i\eta_i A N c_i \tag{2.42}$$

where $\dot{z}$ is the sample erosion rate, β_i is the fractional amount of i sputtered in the form of an ion, η_i is the number of sputtered ions originating from A, the analysis or gated area and N is the atomic density of the sample. The value of $\dot{z}$ is obtained empirically. After the original sample surface has been sputtered away (when $z >\sim 10$ nm) the rate of change of crater depth with time (t) is practically constant. Hence $\dot{z} = z/t$. The product $\beta_i\eta_i$ is known as the useful ion yield. β_i cannot be calculated and it is difficult to measure η_i; hence a reference sample, with known c_i is used to determine the products value. Since A, N and I_i are known the concentration of i can be obtained from equation 2.42 and the raw SIMS data can then be converted into c_i as a function of z. This has been proved an invaluable technique in the study of the transient enhanced diffusion of boron by authors such as Pelaz *et al.* (1999) and Mannino *et al.* (2000).

2.7 Injection dependent lifetime spectroscopy

A critical characteristic of solar cell material is the carrier lifetime. The carrier lifetime is likely to be reduced from that of perfect bulk silicon due to the trapping of carriers at defects and/or impurities. A reduction in carrier lifetime results in a serious reduction in efficiency of the solar cell since a reduction in carrier lifetime implies that fewer carriers can reach the external circuit. One such lifetime degrading defect is discussed in more detail in chapter 6. A very valuable technique in the study of solar cells is Injection Dependent Lifetime Spectroscopy, reviewed below. A good discussion of this technique has been provided by Macdonald and coworkers (Macdonald and Cuevas, 1999; Macdonald *et al.*, 2001) while the more fundamental examination of carrier lifetimes which forms the foundations the above discussions is provided by Hornbeck and Haynes (1955). A good discussion on the Quasi-Steady-State Photo-Conductance (QSSPC) technique, summarised below (Section 2.7.1), has been given by Sinton and Cuevas (1996).

2.7.1 Quasi-steady-state photoconductance

There are several available techniques for the measurement of effective carrier lifetimes in semiconductors, most of which are based on the measurement of photoconductance. One common technique is to study the decay of the photoconductance induced by a flash of light where the duration of the flash is short with respect to the effective carrier lifetime (Hornbeck and Haynes, 1955). The effective lifetime may be inferred from the decay transient. However, the measurement of very low lifetimes is made difficult by complications arising from surface recombination transients and minority carrier spreading.

Quasi-steady-state photoconductance avoids these problems by applying a light pulse that is long with respect to the effective carrier lifetime and measuring the photoconductance

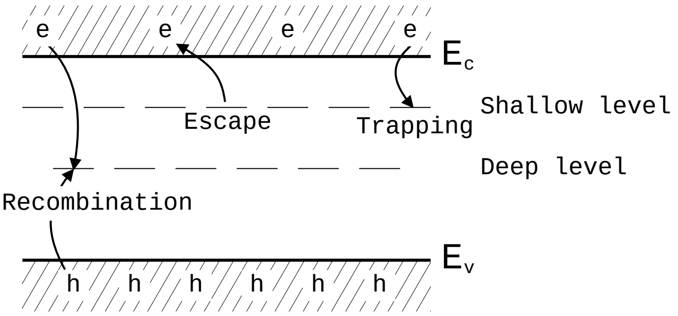

Figure 2.9: The trapping model of Hornbeck and Haynes (1955) for a p-type semiconductor. Recombination is considered to occur at a single deep level. A second, shallow level is able to trap carriers which may later be released.

under the quasi-steady-state illumination. The photoconductance may be measured via a number of methods, many of which have the advantage of being contactless. For example, the photoconductance may be measured by microwave reflectance or via a coil in a bridge circuit that is inductively coupled to the wafer conductivity (Sinton and Cuevas, 1996). Sample preparation may need to include the getter-removal of metallic impurities (which are strong recombination centres). To avoid surface recombination the surface is usually etched and subsequently passivated (*e.g.* by SiN deposition).

2.7.2 Determination of the effective carrier lifetime

The trapping model originally implemented by Hornbeck and Haynes (1955) in the study of photoconductance decay transients can be adapted to the quasi-steady-state regime (Macdonald and Cuevas, 1999). The model for the case of p-type material described here, is shown in figure 2.9. Recombination occurs at a deep recombination level at a rate $1/\tau_r$. The effect of illumination is to free electrons from the shallow level at a rate $1/\tau_g$ where, under steady-state conditions, τ_r and τ_g are equal. The shallow traps are filled at a rate $1/\tau_t$ and the mean time spent by a carrier in the trap is τ_g. The density of normally empty shallow traps is N_t which is dependent upon the Fermi-level E_F and temperature T. Note that in the limit where $E_F \sim E_v$ and the shallow trap is close to the conduction band, N_t is simply the total trap density.

Consider the regime where the excess carrier concentration is much greater than N_t which can be realised by intense illumination. Since the number of excess electrons is much larger than N_t, regardless of the number of electrons in traps, the excess electron density Δn will be approximately equal to the excess hole density Δp. Recombination will occur at a rate proportional to the photogeneration rate, reflecting the bulk lifetime of the material. An alternative regime is that where the excess carrier concentration is closer to or less than N_t, realised by less intense illumination. In this case the density of trapped electrons n_t, will be large. To maintain charge neutrality $\Delta p = \Delta n + n_t$ must always hold which implies that if the density of trapped electrons is large Δp will also become large resulting in an increased photoconductivity. The excess conductivity is given by,

$$\Delta\sigma = q\Delta n(\mu_n + \mu_p) + qn_t\mu_p \tag{2.43}$$

where μ_n and μ_p are the electron and hole mobilities. It can now be seen that the term *Injection Dependent* Lifetime Spectroscopy follows from the dependence of Δn upon illumination intensity.

Hornbeck and Haynes (1955) described the trapping process in p-type material by the following equations,

$$\frac{d\Delta n}{dt} = g_e - \frac{\Delta n}{\tau_r} + \frac{n_t}{\tau_g} - \frac{\Delta n(1 - n_t/N_t)}{\tau_t} \tag{2.44}$$

$$\frac{dn_t}{dt} = -\frac{n_t}{\tau_g} + \frac{\Delta n(1 - n_t/N_t)}{\tau_t} \tag{2.45}$$

where g_e is the photogeneration rate of electron-hole pairs. Since steady-state conditions mean that the left hand sides of equations 2.44 and 2.45 must equal zero,

$$\Delta n = g_e\tau_r \tag{2.46}$$

which in turn implies that

$$n_t = \frac{N_t\Delta n}{\Delta n + N_t\tau_t/\tau_g} \tag{2.47}$$

although in general QSSPC actually obtains $\Delta n_{eff} = g_e\tau_{eff}$ rather than Δn given by equation 2.46. It is important to note that Δn is always equal to that expected in the absence of traps for a given g_e. Trapping increases the number of excess holes which leads to the excess conductivity described by equation 2.43.

From the measured excess conductivity (equation 2.43) and the photogeneration rate which is generally known, it is possible to derive

$$\tau_{eff} = \frac{\Delta\sigma}{g_e(\mu_n + \mu_p)} \tag{2.48}$$

where it has been assumed that $\Delta n = \Delta p$.

2.7.3 Determination of electrical levels

By fitting experimental data to the trapping model described above (specifically equations 2.43,2.46,2.47,2.48), the values of τ_r, τ_t/τ_g and N_t may be derived. Hornbeck and Haynes (1955) have shown that the energy level E_T of the trap is given by,

$$E_c - E_T = kT \ln\left(\frac{N_c\tau_g}{N_t\tau_t}\right) \tag{2.49}$$

where kT has its usual meaning and N_c is the effective density of states of the conduction band edge. It is important to note that if there is more than one trap the calculated level will be a weighted average of all those present.

2.7.4 Determination of capture cross sections

The use of QSSPC in the determination of another important fundamental property of a carrier trap, namely the capture cross section, has been described by Schmidt and Cuevas (1999) and later by Macdonald *et al.* (2001). Since the injection-level dependence of the effective carrier lifetime is highly sensitive to the resistivity of the sample being studied, fitting Shockley-Read-Hall (SRH) recombination curves to many samples of different dopant concentration allows unique and accurate values of the electron and hole capture cross sections, σ_n and σ_p to be obtained. This method is claimed to be able to determine capture cross sections more accurately than by extrapolation of a DLTS Arrhenius plot (section 2.2)[2].

[2]Note that although capture cross sections obtained by extrapolation of a DLTS Arrhenius plot may have relatively large errors, changing the filling pulse in a DLTS experiment can give values of the capture

Shockley-Read-Hall statistics

Shockley-Read-Hall statistics (Shockley and Read Jr., 1952; Hall, 1952) are used to fit plots of excess carrier concentration Δn against the effective carrier lifetime τ_{eff}. These plots are known as recombination curves. The SRH carrier lifetime is given by,

$$\frac{1}{\tau_{SRH}} = \frac{N_A + \Delta n}{\tau_{p0}(n_1 + \Delta n) + \tau_{n0}(N_A + p_1 + \Delta n)} \tag{2.50}$$

where N_A is the dopant density and it is assumed that $\Delta n = \Delta p$ as in section 2.7.2. We will define n_1 and p_1 later (equations 2.53 and 2.54). The electron and hole lifetimes are τ_{n0} and τ_{p0} respectively. These are given by,

$$\tau_{n0} = 1/\big((v_{th})_n \sigma_n N_{SRH}\big) \tag{2.51}$$

and

$$\tau_{p0} = 1/\big((v_{th})_p \sigma_p N_{SRH}\big) \tag{2.52}$$

where $(v_{th})_n$ and $(v_{th})_p$ are the respective carrier's thermal velocity. Equations 2.51 and 2.52 thus define the electron and hole capture cross sections, σ_n and σ_p. The electron and hole densities when the Fermi-level is coincident with the recombination centre level are

$$n_1 = N_c \exp\left(\frac{E_T - E_c}{kT}\right) \tag{2.53}$$

and

$$p_1 = N_v \exp\left(\frac{E_c - E_{gap} - E_T}{kT}\right) \tag{2.54}$$

respectively, where E_c is the energy level of the conduction band, E_{gap} is the band-gap and E_T is the energy level of the trap. N_c and N_v are the effective densities of states of the conduction band and valence band edges respectively and again kT has its usual meaning.

The assumption that $\Delta n = \Delta p$ and the application of the standard SRH statistics described above is valid when trapping effects are insignificant and while the recombination centre density is much less than the injected carrier density ($N_{SRH} \ll \Delta n$).

cross sections with an accuracy comparable to the method described above.

It is possible to simplify equation 2.50 for low or high injection conditions ($\Delta n \ll N_A$ and $\Delta n \gg N_A$ respectively). Since for a deep level n_1 and p_1 are likely to be very much less than N_A, if $\Delta n \ll N_A$, equation 2.50 will become,

$$\frac{1}{\tau_{SRH}} = \frac{1}{\tau_{n0}} \tag{2.55}$$

while if $\Delta n \gg N_A$ it becomes

$$\frac{1}{\tau_{SRH}} = \frac{1}{\tau_{p0}}. \tag{2.56}$$

Fitting Shockley-Read-Hall curves

The measured effective lifetime τ_{eff} is plotted against the excess carrier concentration Δn for samples of various N_A. This data is then fitted with curves obtained from the above equations. By adjusting parameters to give a good fit to all curves, unique values of σ_n and σ_p can be found.

2.8 Summary

The preceding sections of this chapter have outlined the basic principles behind the experimental techniques most relevant to the rest of the work herein. Of course there are many other useful techniques not discussed here. Even the techniques that are discussed here have many extensions and refinements that have not all been mentioned. Its is important to note that many ideas and techniques are transferable. For example the application of stress to determine symmetry, discussed in the section on PL (section 2.4) can also be used by DLTS (section 2.2) and EPR (section 2.5). Furthermore the method of determining activation energies from an Arrhenius plot described in section 2.2 can be used to determine the activation energy of any reaction (*e.g.* defect dissociation or migration) and not just for the activation energy for the removal of a carrier from a deep level.

Several techniques have been described that are capable of providing, for example the position of an electrical level or the frequency of a local vibrational mode of a defect.

The fact that many techniques attempt to measure the same property in different ways is rather useful for two main reasons. It gives confidence to the results, but also the additional information obtained by each alternative technique complements the information available previously. Similarly this is the main strength and most important use of theory: to model the properties observed experimentally thus confirming or disproving a given model. Hence the strong collaboration of theory and experiment is extremely valuable, as it is hoped will be made clear during the proceeding chapters. The proceeding chapters all follow a format similar to this: A review of previous (mainly experimental) work will be given, followed by a presentation of the calculations preformed by this author on that subject, followed by the conclusions that can be drawn in the light of the theoretical investigation.

Chapter 3

Substitutional and Interstitial Boron

3.1 Introduction

The fundamental elements of any semiconductor device are p-n junctions: adjoining n and p-type regions. Since boron is a group III element, substitutional boron acts as a shallow acceptor in silicon. Boron is the acceptor most commonly used in silicon device manufacture today and so an understanding of its properties is critical. As discussed in the Introduction, due to the relentless reduction in junction size, extremely small and extremely highly boron doped regions are being demanded. Small, highly doped regions are achieved by ion-implantation. Boron ions or complexes containing boron (*e.g.* BF_2) are fired into the crystal. Such an approach has the advantage of providing the ability to dope crystals at concentrations greater than the dopants solid solubility limit as well as offering the ability to accurately choose the doped region. However, as the ions enter the crystal, although most of their energy is dissipated via interactions with electrons, some of the ions will collide with host nuclei and displace them. Following implantation there will therefore be much damage in the crystal and a large fraction of implanted atoms will not be located at substitutional sites as required for boron to act as a shallow

	^{11}B	^{10}B	shift
$B^- \ T_d$			
T_2	626 ($^{11}B_s$ 623)	650 ($^{10}B_s$ 646)	24 (23)
$BI^+ \ C_{3v}$			
A_1	697 (^{11}R 730)	726 (^{10}R 757)	29 (27)

Table 3.1: A comparison of the calculated and observed local vibrational modes of substitutional boron and the boron interstitial defect. Experimental data are in parenthesis and were obtained by infrared spectroscopy (Bean *et al.*, 1972; Tipping and Newman, 1987). All modes are given in wavenumbers (cm^{-1}). The symmetry of the calculated modes and defects are also given. For an explanation of these symmetries the reader is referred to Tinkham (1964).

acceptor. Therefore, to activate the boron (to make it substitutional) and to remove damage to the host crystal, the material must be annealed. This can have the side effect of triggering transient-enhanced diffusion, recently reviewed by Jain *et al.* (2002), during which the implanted boron may diffuse over large distances in a very short time, placing a limit on the minimum length scales of devices. Furthermore, in the regions with the highest dopant concentration boron-interstitial clusters (BICs) may be formed. These BICs though immobile, are likely to be effectively electrically inactive (Stolk *et al.*, 1995*a*; Pelaz *et al.*, 1997; Cowern *et al.*, 1990) therefore reducing the effectiveness of the doping. Before considering BICs it is instructive to study substitutional and interstitial boron, both of which have been investigated experimentally for many years.

3.2 Substitutional boron

The acceptor level of substitutional boron B_s is around 0.045 eV above the valence band (Morin and Maita, 1954). In material counter-doped with either arsenic or phosphorus, using infra-red spectroscopy the local vibrational modes of B_s at 646 and 623 cm^{-1} for $^{10}B_s$ and $^{11}B_s$ respectively have been observed (Smith and Angress, 1963; Angress *et al.*, 1965; Balkanski and Nazarewicz, 1966; Bean *et al.*, 1972).

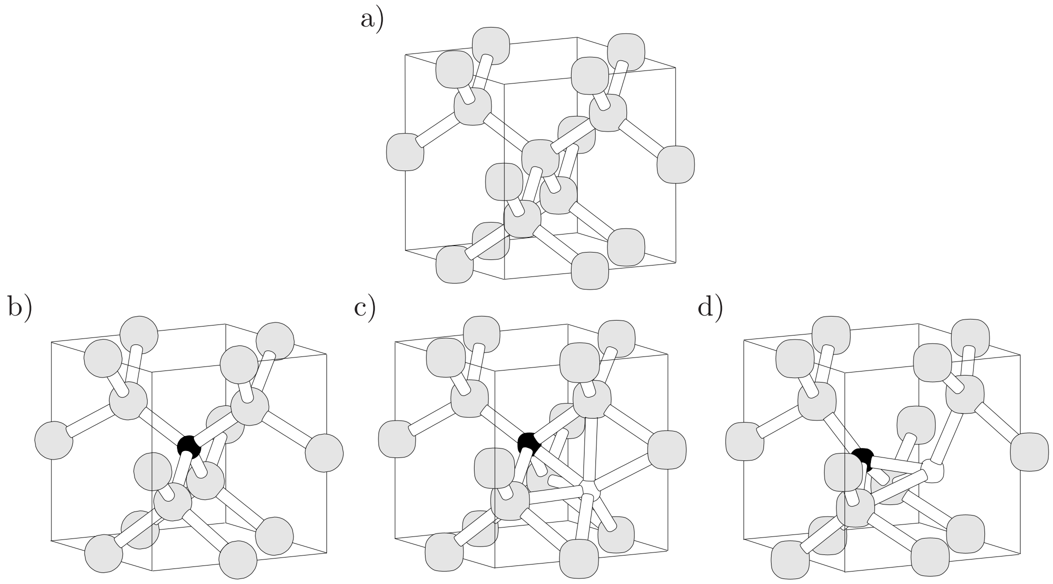

Figure 3.1: The structures of substitutional boron and the boron interstitial defect. Structure *b* is substitutional boron. The boron interstitial defect in the C_{3v} and C_{1h} configurations, referred to as $\mathrm{B}_s\mathrm{Si}_i^T$ and B_i^X, are represented by structures *c* and *d* respectively. Also shown is a piece of equivalent bulk material to aid the reader (structure *a*). Small black balls represent boron, large grey balls represent substitutional silicon atoms and small white atoms represent interstitial silicon atoms.

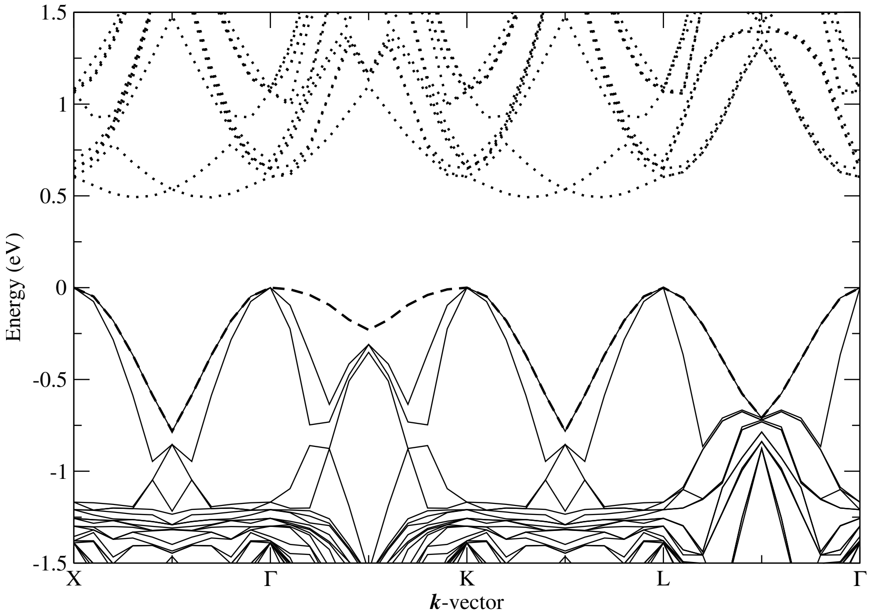

Figure 3.2: The band-structure of substitutional boron. In the neutral charge state the solid lines represent filled levels, the dotted lines represent empty levels and the dashed line is a half occupied level. The dashed line represents the acceptor level of boron. In this and all following band-structures the labels X, Γ, K and L correspond to the **k**-points having coordinates (1/2 0 0), (0 0 0), (1/2 1/2 0) and (1/2 1/2 1/2) respectively in the Brillouin zone of the conventional unit supercell.

	^{11}B	^{10}B	shift	Mixed isotope mode
B_2^{2-} D_{3d} (C_{3v})				
A_{1g} (A_1)	635	664	29	650
E_g (E)	602	627	25	617
E_u (E)	551 (P_1 553)	568 (P_3 570)	17 (17)	558 (P_2 560)

Table 3.2: The calculated and observed (Newman and Smith, 1967) local vibrational modes of the substitutional boron pair. Observed modes are given in parenthesis. All modes are given in wavenumbers (cm^{-1}). The symmetries of the calculated structures and modes are also given.

The structure of substitutional boron, shown in figure 3.1, involves an inward relaxation of the surrounding silicon atoms due to the smaller radius of the boron atom. The shallow acceptor level of boron is shown as a dashed line in the calculated band-structure presented in figure 3.2. Using the marker method with bulk silicon as the marker, the calculated $(-/0)$ level is placed at $E_v + 0.2$ eV. This is in reasonable agreement with the experimentally observed position with the disagreement indicative of the size of the possible errors associated with the calculation of defect levels using DFT. However it is worth noting that bulk silicon is not a particularly good marker in this case since the $(-/0)$ level of bulk silicon can be considered to be the conduction band and hence the position and furthermore the character of the marker's $(-/0)$ level is significantly different from that of substitutional boron. The calculated local vibration modes of B_s are presented in table 3.1 and show excellent agreement with experiment. The fact that the local vibrational modes of B_s agree so well with experiment while the calculated position of the acceptor level is in disagreement with experiment by around 0.2 eV highlights both the accuracy with which DFT is able to calculate LVMs and the relatively large errors present in the calculation of electrical levels.

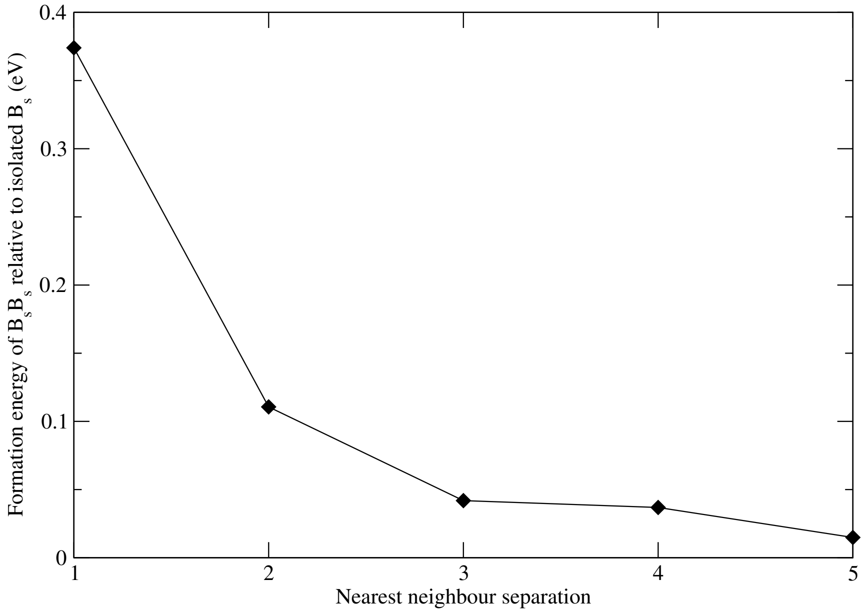

Figure 3.3: The calculated formation energy of the substitutional boron pair, relative to isolated substitutional boron, as a function of separation. A separation of 1 corresponds to nearest neighbour B_sB_s, a separation of 2 corresponds to next nearest neighbour B_sB_s and so on. The formation energy of isolated substitutional boron has been set to zero. The plot shows that two substitutional boron atoms in a nearest neighbour configuration is unstable with respect to two isolated substitutional boron atoms by 0.37 eV.

3.3 Substitutional Boron Pairs

As well as isolated substitutional boron, local vibrational modes have been observed and assigned to pairs of substitutional boron atoms (Newman and Smith, 1967). As shown in table 3.2 the calculated local vibrational modes of a nearest neighbour substitutional boron pair agree well with the observed modes which are labelled the P lines. It is interesting to note that this nearest-neighbour substitutional boron pair is metastable with respect to two isolated substitutional boron atoms. This is clearly seen when the formation energy of B_2 is plotted as a function of distance between the two boron atoms (figure 3.3). The energy clearly tends toward the formation energy of isolated substitutional boron as the pair are separated. Hence the assignment of the nearest neighbour substitutional boron pair to the P centre provides evidence for the existence of metastable defects in silicon. In this case it is believed that the nearest neighbour substitutional boron pair is grown in or formed by some reaction and that the barrier for the boron atoms to move apart is too large for a more stable arrangement to be achieved (Tipping and Newman, 1987).

3.4 The Boron Interstitial Defect

Electron irradiation induces lattice damage in the form of vacancy-interstitial (Frenkel) pairs. If boron and phosphorus co-doped material is electron irradiated at low temperature new vibrational modes labelled the R lines are observed at 757 and 730 cm^{-1} for ^{10}B and ^{11}B respectively. These lines have been assigned to interstitial boron B_i, formed when self-interstitials react with B_s (Tipping and Newman, 1987). The vacancies generated by the irradiation are trapped by the counter dopant creating E centres. The R lines are stable up to 230 K.

The Si-G28 EPR (and ENDOR) centre is also observed following low temperature electron irradiation of boron doped silicon and has also been assigned to B_i (Watkins, 1975). The production rate of B_i is similar to that of the single vacancy implying a remarkably efficient capturing of the irradiation-induced self-interstitials by B_s. Near band-gap light is required to observe the Si-$G28$ spectrum implying that the paramagnetic charge state that is visible to EPR (neutral in this case) is metastable. An analysis of the hyperfine interactions estimates that $\sim 33\%$ of the wavefunction lies in a p-orbital of the single boron atom while two nearby silicon atoms have $\sim 17\%$ each, accounting for a total of $\sim 67\%$ of the wavefunction. The angular dependece of the EPR spectrum is not fully resolved but the symmetry is observed to be C_{1h}, lower than that of the more typical tetrahedral (T_d) or hexagonal (D_{3d}) interstitial sites. Biaxial stress experiments demonstrate that an electron or hole capture process induces a change in the centre's structure from one of low monoclinic to trigonal symmetry. The Si-$G28$ centre anneals with an activation energy of 0.60 ± 0.05 eV. The prefactor for the annealing rate is a factor of 10^6 smaller than the prefactor for reorientation obtained from the biaxial stress study. Hence Watkins (1975) proposed that B_i anneals by diffusing with a migration energy of 0.6 eV until, after making $\sim 10^6$ single jumps, it is trapped by some other defect.

The fact that B_i is observed by EPR in neither n nor p-type material in the absence of illumination implies that B_i has both a donor and acceptor level allowing it to be non-paramagnetic in its stable charge state in both n and p-type material. Since the Si-$G28$ EPR centre has not been observed in n-type silicon, only the donor (0/+) level could be located by EPR which placed it around $E_c - 0.15$ eV. Later DLTS placed the donor (0/+) level at $E_c - 0.13$ eV. The acceptor ($-$/0) level was also observed by DLTS experiments (Watkins and Troxell, 1980; Troxell and Watkins, 1980) and placed at $E_c - 0.45$ eV later corrected by Harris *et al.* (1987) who showed that a better estimate of the level position was $E_c - 0.37$ eV. Thus B_i is a negative-U system, meaning that when either an electron is trapped at the positive defect or a hole is trapped at the negative one, a second charge carrier will be spontaneously trapped changing the charge state directly from $+1$ to -1 or vice versa. The structural reorientation associated with the change in charge state lowers the total energy by more than enough to compensate for the Coulombic repulsion between the pair of electrons or holes. Negative-U was a phenomenon first proposed by Anderson (1975) to explain the failure to detect paramagnetism for intrinsic defects in chalcogenide glasses.

In agreement with other theoretical studies (Hakala *et al.*, 2000; Jeong and Oshiyama, 2001), the present work finds that the most stable configuration formed when a self interstitial is trapped by substitutional boron in the positive and neutral charge state has C_{3v} symmetry and is essentially *substitutional* boron with a self interstitial trapped at an adjacent tetrahedral interstitial site. In the negative charge state the C_{3v} symmetry is broken forming a complex similar to a [110] split-interstitial (a complex where a dumbbell of two atoms, aligned along [110], replaces a single host atom). This configuration has C_{1h} symmetry and is, according to theory, metastable in the neutral charge state being 0.18 eV higher in energy than the C_{3v} form. The C_{3v} and C_{1h} configurations are labelled $\mathrm{B}_s\mathrm{Si}_i^T$ and B_i^X respectively and are shown in figure 3.1. Given that for the case of $\mathrm{B}_s\mathrm{Si}_i^T$ the boron atom is not interstitial and for B_i^X it is not well defined whether the boron is interstitial or not, interstitial boron is a rather misleading name for this complex. Hence, from this point it will be referred to as the boron interstitial defect or just BI where $\mathrm{B}_n\mathrm{I}_m$ is a complex formed between n boron atoms and m interstitial atoms of either species. The latter notation conveniently removes the requirement of defining which atoms are and which are not interstitial, a task that is often non-trivial and unnecessary. Note that in agreement with the EPR data (Watkins, 1975) the symmetry of the positively charged defect is trigonal. However, theory finds that the symmetry of the most stable form in the neutral charge state is also trigonal, the C_{1h} form, as observed by EPR, is found to be metastable. Nevertheless, this already provides evidence that the structures shown in figure 3.1 are indeed the correct form of the boron interstitial defect. The donor and acceptor levels of BI, calculated using bulk silicon as a marker, are $E_c - 0.05$ and $E_c - 0.33$ eV, in good agreement with EPR and DLTS and reproducing the observed negative-U. Boron located at the more usual tetrahedral or hexagonal interstitial site, B_i^T or B_i^H respectively, is significantly less stable than $\mathrm{B}_s\mathrm{Si}_i^T$. In p-type material the position of the electrical levels of $\mathrm{B}_s\mathrm{Si}_i^T$, B_i^T and B_i^H dictates that they will all be positively charged and in this charge state both B_i^T and B_i^H are around 1.0 eV less stable than $\mathrm{B}_s\mathrm{Si}_i^T$.

Further evidence that the correct structure of BI is that shown in figure 3.1 comes from a comparison of the observed and calculated local vibrational modes. The position of the

	EPR			DFT		
Atom	α_j^2	β_j^2	η_j^2	α_j^2	β_j^2	η_j^2
B	0.03	0.97	0.33	0.005	0.99	0.17
Si_1	0.14	0.86	0.15	0.01	0.77	0.08
Si_2	0.02	0.98	0.19	0.05	0.83	0.11

Table 3.3: The calculated and observed (Watkins, 1975) wavefunction parameters for the boron interstitial defect. α_j^2 is the s-like fraction, β_j^2 the p-like fraction and η_j^2 is the total fraction of the paramagnetic electron localised on the given atom.

donor and acceptor level of BI implies that in the compensated material in which local vibrational modes were measured, the boron interstitial defect will be positively charged. Hence the local vibrational modes were calculated for $B_sSi_i^{T+}$, the stable form of BI under such conditions. The calculated and observed modes are again given in table 3.1 and agree well. A comparison of the calculated and observed shift in wavenumber with boron isotope shows an agreement to within 2 cm^{-1}.

Finally, a comparison of the calculated wavefunction parameters for the form of BI that corresponds to the observed paramagnetic state, B_i^X to those measured by Watkins (1975) for the boron interstitial defect, reveals several similar trends, shown in table 3.3. Although the quantitative agreement is not superb it is clear that a significant fraction, around 36%, of the paramagnetic electron of B_i^X lies on the boron and just two other silicon atoms, in agreement with the EPR study. Furthermore, in all cases the wavefunctions is largely accommodated in a p-orbital. Thus a study of the nature of the wavefunction of B_i^X provides the final piece of evidence that BI is none other than B_i^X and $B_sSi_i^T$ shown in figure 3.1.

3.5 Summary

Self-interstitials are produced by irradiation or ion-implantation and the reaction of self-interstitials with substitutional boron generates boron interstitial defects (BI). It has been demonstrated above that the generation of BI is not a kick-out mechanism but is rather a pairing between a self-interstitial and substitutional boron atom. BI has been shown to be a negative-U centre which in the positive charge state has the form of substitutional boron next to a tetrahedral self-interstitial. In the negative charge state the defects symmetry is reduced. BI becomes mobile below room temperature as evidenced by the annealing of the R lines at 230 K. It is likely that when BI becomes mobile it will be trapped at other centres to generate larger, more stable complexes. The complexes formed in the high-boron concentration regime are the topic of the next chapter, while chapter 5 will deal with the complexes formed when the boron concentration is comparable to that of other impurities common to silicon.

Chapter 4

Boron Interstitial Clusters

4.1 Introduction

There is much interest in the stable, electrically inactive boron interstitial clusters (BICs) formed during transient enhanced diffusion (TED) (Stolk *et al.*, 1995*a*; Pelaz *et al.*, 1997; Cowern *et al.*, 1990). At present rather little is known about them since they are too small to be resolved by transmission electron microscopy (TEM), are electrically inert and hence not observable by electrical measurements such as DLTS and at present there is no information on local vibrational modes of these clusters of which there are likely to be many different forms. One way to gain an insight into the nature of these clusters is to employ theory but since the experimental knowledge of these clusters is sparse, attempting to model the clusters stable at the temperatures associated with TED is difficult with no means of gauging success. This is especially apparent when it is realised that some BICs may exist in metastable forms, just like the nearest neighbour substitutional pair discussed in section 3.3. Here, the small clusters formed when BI anneals have been studied since in this case there are experimental observations of the defects formed. It is hoped that by developing an understanding of the evolution of boron in implanted or irradiated material

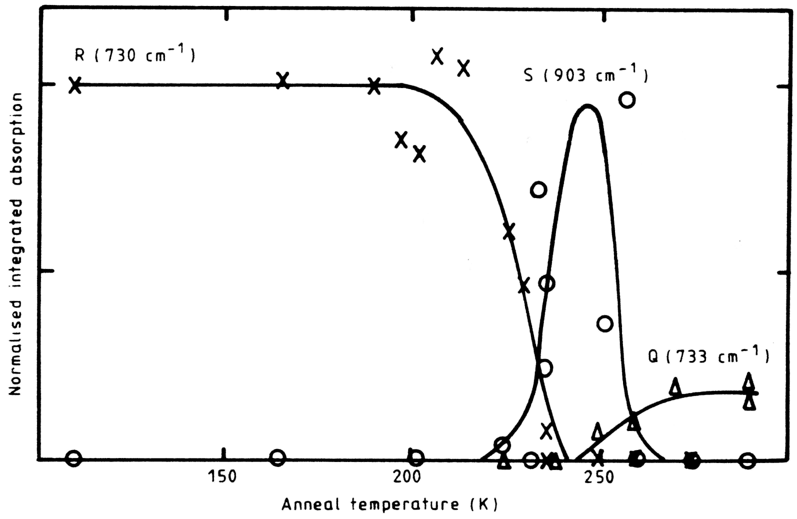

Figure 4.1: The normalised integrated absorption of the 730 cm^{-1} R line, 903 cm^{-1} S line and 733 cm^{-1} Q line for various anneal temperatures. This figure is reproduced with kind permission from Tipping and Newman (1987).

this approach will provide a clearer insight into the properties of the BICs observed during TED.

4.2 The S and Q centres

Isochronal anneals in association with infrared absorption measurements (Tipping and Newman, 1987) reveal much of what is known about small BICs. As described in the previous chapter, local vibrational modes assigned to BI are observed in compensated, boron doped silicon that has been irradiated with electrons at low temperature. It is critical that the electron irradiation is performed at low temperature since the R lines (assigned to BI) begin to anneal at 200 K and are removed completely by a 10 minute anneal at 250 K. This disappearance of the R lines coincides with the growth of set of lines labelled the S lines. The defect responsible for these lines, the S centre, is stable only within a narrow temperature window, annealing at 270 K. A new set of lines labelled the Q lines then appear. The centre responsible for these lines is stable to 220 °C. The Q centre can also be formed directly by the same dose of room temperature irradiation (Laithwaite *et al.*, 1975). Figure 4.1 shows the variation of the normalised integrated absorption of the R, S and Q lines with annealing temperature, clearly demonstrating the evolution from the R centre to the Q centre via the S centre.

The frequencies of the S and Q lines are shown in table 4.1. Five local vibrational modes have been correlated with the S centre. Tipping and Newman (1987) reasoned that the S centre could be assigned to a boron pair having axial symmetry. The lines S_1, S_2 and S_3 would then be the non-degenerate modes, parallel to the symmetry axis, while the lines S_4, S_5 and S_6 would be the doubly degenerate modes perpendicular to the axis. Note that S_6 is an unobserved line predicted by Tipping and Newman (1987) to lie at 624 cm^{-1}. If it were present it would be masked by the much stronger absorption due to $^{11}B_s$. Thus, infra-red spectroscopy suggests that the S lines originate from a complex of two boron atoms having axial symmetry. The Q centre on the other hand gives rise to only one local mode per isotope. Bean *et al.* (1972) assigned these modes to interstitial boron. In material compensated with phosphorus the E centre (PV) is observed. Since B_i is expected to react with PV to generate B_s-P_s complexes and no formation of B_s-P_s was observed during irradiation once compensation had been achieved, Bean *et al.* (1972) reasoned that B_i must be immobile at 300 K and since it is immobile, clusters of interstitial boron must be unable to form. Furthermore, since BV is known to be unstable at 300 K (Watkins, 1976) and almost all of the vacancies generated by the irradiation are thought to be contained in E centres, the most likely possibility is that the Q lines are due to B_i. However, as discussed in chapter 3, Tipping and Newman (1987) later assigned the R lines to interstitial boron and since the S centre has been assigned to a pair of boron atoms they concluded that the Q centre, formed when the S centre anneals, must be a complex of three or more interstitial boron atoms.

When BI becomes mobile it is likely that, in the absence of other impurities it will be trapped by either substitutional boron or by a vacancy. In the infrared experiment described above the vacancies are trapped in the form of stable E centres so it is probable that the anneal of BI will result in its trapping by B_s and the formation of B_2I (B_sB_i using

	^{11}B	^{10}B	shift	Mixed isotope mode
B_2I^S D_{3d} (C_{3v})				
A_{2u} (A_1)	919 (S_1 903)	946 (S_3 928)	27 (25)	934 (S_2 917)
A_{1g} (A_1)	702	736	34	718
E_u (E)	611 (S_4 599)	634	23	628 (S_5 603)
E_g (E)	595	616	21	600
B_2I^Q D_{2d} (C_{2v})				
A_1 (A_1)	1132	1186	54	1159
E (B_1)	756 (^{11}Q 733)	785 (^{10}Q 760)	29 (27)	785, 756
$B_2I_3^Y$ C_{1h} (C_1)				
A' (A)	1043	1093	50	1068
A'' (A)	808	841	33	836
A' (A)	801 ($I2$ 843)	832 ($I2$ 882)	31 (39)	804
A'' (A)	580	591	11	584
A' (A)	578	581	2	580
A' (A)	563	567	5	565
A' (A)	552	556	3	554
A'' (A)	232 ($I2$ 242)	234 ($I2$ 242)	2 (<1)	233

Table 4.1: The local vibrational modes of small boron interstitial clusters as calculated in this work, observed modes being given in parenthesis. All modes are given in wavenumbers (cm^{-1}). The S and Q centres observed by IR spectroscopy (Newman and Smith, 1967; Laithwaite *et al.*, 1975; Tipping and Newman, 1987) are assigned to a metastable and stable form of B_2I respectively. The vibrational modes of the $I2$ centre, observed by PL (Thonke *et al.*, 1984) are assigned to a metastable form of B_2I_3. Symmetries of defects and modes are given for the isotopically pure defects with mixed isotope symmetries in parenthesis.

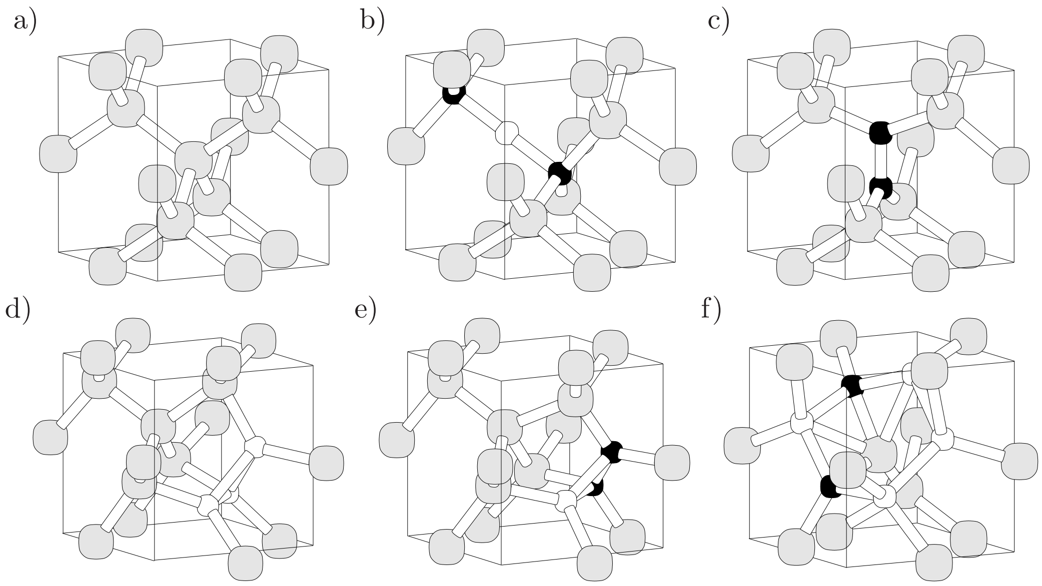

Figure 4.2: The structures of interstitial and boron interstitial clusters as calculated in this work with the exception of structure *d* which was calculated by Jones *et al.* (2002). B_2I^S (structure *b*) is assigned to the *S* centre, B_2I^Q (structure *c*) to the *Q* centre, I_3^W (structure *d*) to the *W* or *I*1 centre and $B_2I_3^Y$ (structure *e*) is assigned to the *Y* or *I*2 centre. $B_2I_3^{6r}$ is the more stable form of B_2I_3 but it has not been observed in any experiment. Small black balls represent boron, large grey balls represent substitutional silicon atoms and small white atoms represent interstitial silicon atoms.

more conventional notation). Of the different configurations of B_2I studied it was found that B_2I^Q shown in figure 4.2 is the most stable, in agreement with previous theoretical investigations (Tarnow, 1992; Zhu *et al.*, 1996; Liu *et al.*, 2000). B_2I^Q consists of two boron atoms sharing a single lattice site in a split interstitial configuration aligned along [100] with D_{2d} symmetry. IR-active local modes with E-symmetry occur at 756 cm^{-1} (^{11}B-^{11}B) and 785 cm^{-1} (^{10}B-^{10}B) and within 30 cm^{-1} of the Q lines, shown in table 4.1. The displacements of the boron atoms in the E mode are along $\langle 110 \rangle$ directions and perpendicular to the boron-boron bond. Changing the mass of one of the boron atoms has negligible effect on the frequency of the second boron atom and for ^{10}B-^{11}B the two E modes occur within 0.01 cm^{-1} of the isotopically pure ones. The position of these modes and their isotopic shifts are in very good agreement with the Q lines at 733 cm^{-1} (^{11}B) and 760 cm^{-1} (^{10}B) which are formed after BI has annealed. It is seen that the absence of mixed modes does not require there to be only one boron atom in the defect. The boron-boron bond length of 1.54 Å implies a displacement of boron from its lattice site of 0.77 Å along $\langle 100 \rangle$, which is consistent with the channelling studies of Smulders *et al.* (1990). These studies were carried out with low temperature (35 K) proton irradiations and showed that substantially more B_s atoms are displaced from substitutional sites following a 300 K anneal, consistent with the formation of B_2I^Q. B_2I^Q is electrically inactive as expected from simple one electron arguments (all bonds are fully satisfied and close to the ideal bond-angle). The binding energy of B_s^- and B_i^+ in B_2I^Q is 1.3 eV and if it is assumed that the defect dissociates at a rate $\nu \exp(-(W + E_B)/kT)$ where ν is the Debye frequency and W is the migration barrier of BI^+, known to be 0.6 eV (Troxell and Watkins, 1980), then B_2I^Q would dissociate at around 400°C. Its disappearance at 220°C is in accordance with the suggestion of Bean *et al.* (1972) that B_2I traps a divacancy that becomes mobile at this temperature. The resulting B_2V complex breaks up forming substitutional boron and resulting in a partial recovery to the electrical activity of boron observed at this temperature (Bean *et al.*, 1972).

The assignment of B_2I^Q to the Q centre suggests that the S centre, which is stable only over the narrow temperature range illustrated by figure 4.1, is likely to be some intermediate centre formed during the conversion of BI to B_2I^Q. Indeed calculations suggest that this is the case. It is found that the local vibrational modes of a complex labelled B_2I^S, shown in figure 4.2 correspond closely to the modes of the S centre. The calculated and observed modes are shown for comparison in table 4.1. The calculated (observed) modes of A_{2u} symmetry lie at 919 (903) and 946 (928) cm^{-1} for ^{11}B-^{11}B and ^{10}B-^{10}B respectively and shift to 934 (917) cm^{-1} for ^{11}B-^{10}B. E_u modes lie at 611 (599) and 634 (-) cm^{-1} for ^{11}B-^{11}B and ^{10}B-^{10}B respectively and shift to 628 (603) cm^{-1} for ^{11}B-^{10}B. The unattributed E_u mode, calculated to lie at 634 cm^{-1} would probably be masked by absorption of $^{11}B_s$ and is in agreement with the line S_6 predicted by Tipping and Newman (1987) to lie at 624 cm^{-1}. The S defect is 0.4 eV less stable than B_2I^Q and hence is a second example of a metastable species that has been frozen in, in this case during the capture of BI by B_s. The assignment of the S lines to a metastable defect complements that of the metastable boron substitutional pair discussed in section 3.3. This work finds that B_2I^S is electrically active with a $(-/0)$ level placed at $E_c - 0.2$ eV when bulk silicon is used as a marker. In the negative charge state its symmetry drops to C_2 or C_{1h} when the Si interstitial moves out of the bond centre. B_2I^S is similar in structure to the neutral di-carbon interstitial centre (Song *et al.*, 1990; Leary *et al.*, 1998).

4.3 The $I2$ centre

To our knowledge, only one other point-like defect thought to be composed of purely boron and self-interstitial atoms has been observed experimentally. It is known as the $I2$ or Y centre. The $I2$ centre has a zero phonon line at 1.080 eV and is observed by photoluminescence in irradiated p-type material that has been annealed at around 350 °C (Thonke *et al.*, 1984; Sauer and Weber, 1983; Terashima *et al.*, 1997). The $I2$ line is composed of four peaks and from the observed isotopic shift of the zero phonon line it has been deduced that $I2$ possess two optically equivalent boron atoms. The centre is found to have a local mode satellite shifted from the zero phonon line by -30.2 meV (labelled $L1$) (Thonke *et al.*, 1984; Sauer and Weber, 1983). Later a second local mode satellite shifted from the zero phonon line by 104.6 meV (^{11}B) and 109.4 meV (^{10}B) was observed (labelled $L2$) (Thonke *et al.*, 1984). $L1$ is reported to display only a small isotope effect (^{10}B 30.18 meV and ^{11}B 30.04 meV). These modes are given in units of wavenumbers in table 4.1. It is noted in the later work that a slight shift between the deconvoluted positions of the modes of the fourfold structure for the natural isotopic abundances of boron and those observed in the isotopically pure cases implies that the two boron are only very nearly equivalent. Zeeman measurements indicate trigonal symmetry although uniaxial stress measurements reveal the true symmetry to be C_{1h} (Thonke *et al.*, 1983). The $I2$ centre is stable to around 400 °C (Sauer and Weber, 1983) or 500 °C (Terashima *et al.*, 1997). Along with $I2$ an intrinsic centre labelled $I1$ (or W) is observed (Thonke *et al.*, 1984; Sauer and Weber, 1983; Terashima *et al.*, 1997). The trigonal $I1$ defect has a zero phonon line at 1.0182 eV and is believed to be composed of three self-interstitials, shown in figure 4.2 (Jones *et al.*, 2002; Estreicher *et al.*, 2001).

In irradiated or implanted material annealed to high temperature, if there are sufficient self-interstitials present that B_2I^Q is not removed by the divacancy as described above, it is possible that B_2I^Q may trap a di-interstitial which is considerable more mobile than the single self-interstitial (Estreicher *et al.*, 2001; Eberlein *et al.*, 2001). Thus the likely result of further annealing of B_2I^Q is the formation of a B_2I_3 complex. In agreement with Liu *et al.* (2000) the most stable form of B_2I_3 is found to be a configuration constructed by arranging three split-interstitials (two B-Si and one Si-Si dumbbells) in a triangle so as to form a six member ring lying approximately in a $(2\bar{3}\bar{3})$ plane. Here we call this complex, which is shown in shown in figure 4.2, $B_2I_3^{6r}$. $B_2I_3^{6r}$ has C_1 symmetry but a similar configuration having the same C_{1h} symmetry as $I2$ can be easily constructed and is only marginally higher in energy ($\sim$ 0.05 eV in the relevant positive charge state). However, both forms of $B_2I_3^{6r}$ give rise to a mid-gap donor level and are thus likely to act as nonradiative recombination centres and certainly not shallow exciton traps as expected for $I2$ which has a zero phonon line energy slightly smaller than the silicon band-gap.

A complex that may be linked to the $I2$ PL centre is a metastable configuration of B_2I_3 similar in form to the structure of $I1$. We call this complex $B_2I_3^{Y}$ and it is shown in figure 4.2. The $I1$ centre ($\mathrm{I}_3^{\mathrm{W}}$) consists of three bond-centred self-interstitials placed in a $[1\bar{1}0]$ chain and can be modified to form $\mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$ by replacing the self-interstitial at each end of the chain with a boron atom. This reaction can be written $\mathrm{I}_3^{\mathrm{W}} + 2\mathrm{B}_s \rightarrow \mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$ and is exothermic, releasing ~ 3 eV. $\mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$ is 0.7 eV higher in energy than $\mathrm{B}_2\mathrm{I}_3^{\mathrm{6r}}$ but it is possible that like $I1$, it is formed through the aggregation of interstitials and like $I1$ once formed cannot dissociate below about 500°C (Estreicher *et al.*, 2001).

The local vibrational modes of $\mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$ have been calculated in both 96 and 216 atom supercells. The true local modes were found to be fully converged in the smaller cell but it was found that an in-band resonance around 230 cm^{-1} became more localised as the supercell size was increased. Local vibrational modes having A'-symmetry lie at 801 cm^{-1} (^{11}B-^{11}B) and 832 cm^{-1} (^{10}B-^{10}B) as shown in table 4.1. These are within 50 cm^{-1} of the phonon-replica modes of $I2$ observed at ~ 843 cm^{-1} and 882 cm^{-1} which probably possess A'-symmetry (Thonke *et al.*, 1984). The calculated A'' mode around 800 cm^{-1} however, is in slightly better agreement with experiment. A low frequency resonant mode with a high degree of localisation on the boron atoms and their silicon neighbours is present at 232 cm^{-1} (^{11}B-^{11}B) and 234 cm^{-1} (^{10}B-^{10}B) where there is a minimum in the bulk phonon density of states. These are in agreement with the phonon replica of $I2$ observed at ~ 242 cm^{-1} with an isotopic shift ~ 1 cm^{-1}. These modes transform as A''-symmetry and there is no selection rule excluding their participation in a one-phonon replica.

A Kohn-Sham band-structure of $\mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$ is presented in figure 4.3. The band-structure demonstrates the lack of deep levels associated with $\mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$. Only shallow levels are present, close to both the valence and conduction bands. If an exciton were trapped at such shallow levels its recombination would result in a zero phonon line with an energy slightly smaller than the band-gap, as is the case for $I2$. Thus the electronic structure, along with the symmetry and vibrational modes, suggests that $\mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$ is a strong candidate for the defect responsible for the $I2$ PL centre. It is a peculiar feature of the $I2$ luminescence that Zeeman studies indicate a trigonal centre in contrast with uniaxial stress measurements. We note that, like the $I1$ line, $\mathrm{B}_2\mathrm{I}_3^{\mathrm{Y}}$ is compressive along [111]. A Zeeman splitting, exhibiting trigonal symmetry due to the strain field, might then arise from a weakly bound carrier

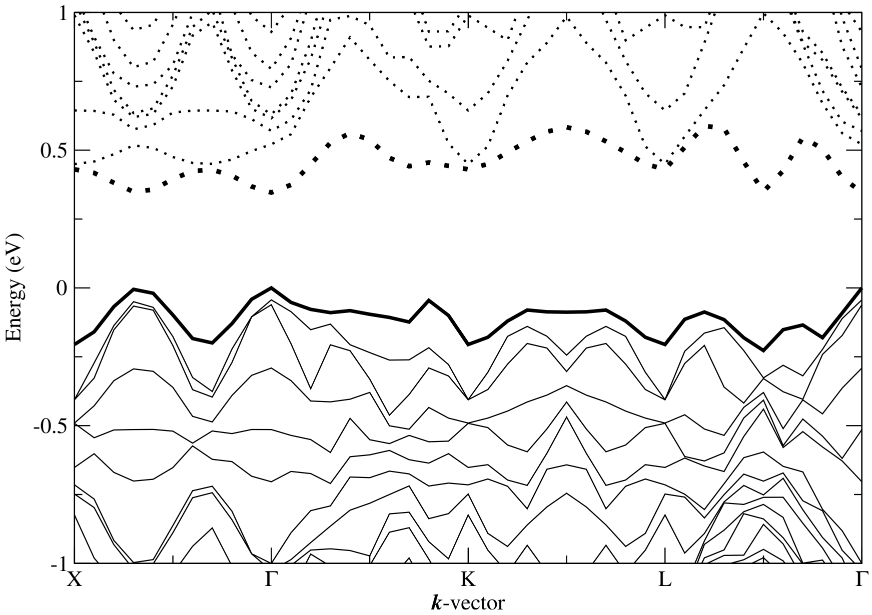

Figure 4.3: The calculated Kohn-Sham band-structure of $B_2I_3^Y$ in a 96 atom supercell. Dotted lines indicate levels which are empty while solid lines indicate levels which are filled in the neutral charge state. Note the shallow level present at each band edge indicated by heavy lines.

Coulombically trapped to the ionised centre.

4.4 Summary

In the previous chapter BI, formed when a self-interstitial is trapped by substitutional boron, was discussed. Above it has been shown that when BI becomes mobile at 230 K it is trapped by B_s forming a [100] boron-boron split interstitial that has been labelled B_2I^Q after its correlation with the Q lines. It has been shown that B_2I^Q is not formed directly from this reaction. Instead a complex that has been labelled B_2I^S, after its assignment to the S lines, is formed. This metastable defect is then converted into B_2I^Q via a small energy barrier. B_2I^Q is likely to be destroyed by reaction with a divacancy which becomes mobile at around 220°C. However, in interstitial-rich regions it is possible that B_2I^Q reacts with a highly mobile di-self-interstitial to form a metastable defect labelled $B_2I_3^Y$ which can be assigned to the $I2$ PL centre. This centre is then stable to around 500°C. Above this temperature there is, to our knowledge, no experimental evidence to facilitate the further development of this proposed reaction path. Recent evidence has however shown that boron may eventually form extended defects that are large enough to be observable by TEM (Cristiano *et al.*, 2003). So far the possibility of the interaction of boron with common impurities found in silicon has been neglected. This important aspect to the reaction paths of boron in silicon will be addressed in the next chapter.

Chapter 5

Boron-impurity complexes

5.1 Introduction

As discussed previously, the properties of the boron related defects formed following ion-implantation or irradiation and subsequent annealing are of great interest. It is hoped that understanding the reactions undertaken by boron will increase the ability to control Transient-Enhanced Diffusion (TED) and facilitate the pursuit of smaller device scales. Chapter 3 discussed substitutional boron and the boron interstitial defect and chapter 4 considered the complexes of boron and self-interstitials formed when BI anneals. In this chapter the role played by impurities common to silicon is discussed.

Oxygen is found in Czochralski (Cz) grown silicon in concentrations of around 10^{17} cm^{-3} or higher and in concentrations around two orders of magnitude smaller in float-zone FZ silicon. The large amount found in Cz silicon is due to incorporation of oxygen from the silica crucible used during wafer growth. Oxygen is a desirable impurity since the interaction of oxygen with dislocations greatly increases the energy required for the dislocations to move. Thus dislocation locking by oxygen makes a silicon wafer much harder and more resistant to warpage. Furthermore, oxygen precipitates getter transition metals which are highly undesirable impurities. The concentration of carbon in silicon is usually around 10^{15}-10^{16} cm^{-3} and it is incorporated during growth by a number of methods including from gaseous contaminants and from contact with hot graphitic growth equipment. As for oxygen, the incorporation of carbon has some positive effects. Substitutional carbon traps self-interstitials, reducing the supersaturation of self-interstitials and thus retarding the diffusion of boron (for example Rücker *et al.* (1998)). Hence carbon has a positive role to play in the reduction of TED. Hydrogen is used in many sample preparation processes such as etching and is thus easily and often introduced. The hydrogen concentration depends very much upon the previous treatment of the wafer but it is almost always present. It is the interaction with boron in the presence of self-interstitials that is discussed below for these three impurities, oxygen, carbon and hydrogen.

5.2 Oxygen

In Cz silicon the annealing of BI coincides with the growth of a new level at $E_c - 0.23$ eV observed by Troxell and Watkins (1980). Centres with near identical growth and anneal temperatures as well as energy levels have been observed at $E_c - 0.27$ eV and $E_c - 0.26$ eV by Mooney *et al.* (1977) and Drevinsky *et al.* (1988) respectively. Due to their similarities these levels are assumed to be one and the same and so to avoid confusion, from this point on, the level of interest will somewhat arbitrarily be referred to as $E_c - 0.23$ eV though there is no evidence that this estimate of activation energy is any more accurate than that of Mooney *et al.* (1977) or Drevinsky *et al.* (1988). Boron's involvement in the centre is confirmed by its effect on the introduction rate. Mooney *et al.* (1977) observed that the level's introduction rate seemed to be proportional to $[B]^{\frac{1}{2}}$.[1] Drevinsky *et al.* (1988) found that for low boron concentrations the introduction rate is directly proportional to [B] and for high concentrations it decays as $[B]^{-2}$. The observation that the introduction rate is directly proportional to [O] (Drevinsky *et al.*, 1988) points to a complex of boron and a

[1]Here and throughout the notation [X] represents the concentration of X.

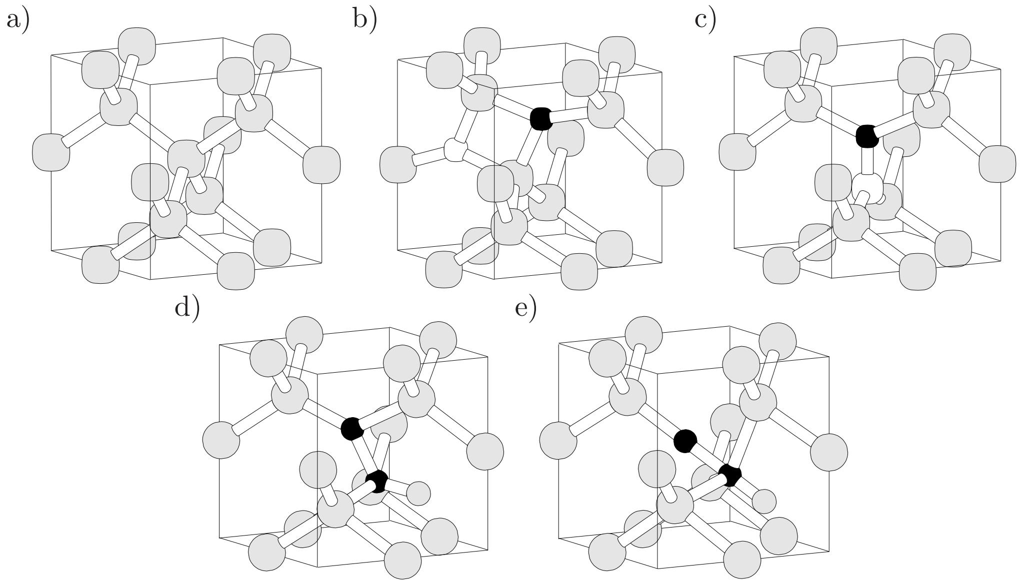

Figure 5.1: The calculated structures of the boron-impurity complexes assigned to the experimentally observed levels. (a) a piece of bulk Si to aid the reader, (b) B_iO_i, (c) B_iC_s, (d) B_iB_sH in the neutral and negative charge state, (e) B_iB_sH in the positive charge state. Large grey balls represent Si, small grey are H, small black are B, small white are O and the larger white ones are C.

single oxygen atom as a candidate for this centre.

Since the defect responsible for the $E_c - 0.23$ eV level is formed as BI anneals and since it is likely to contain just one oxygen atom B_iO_i is a probable candidate. The most stable form of B_iO_i is shown in figure 5.1. The structure is similar to that proposed for C_iO_i (Coutinho *et al.*, 2002) but in this case the oxygen atom is less strongly bound. The binding energy E_B of oxygen with BI is 0.6 eV while the binding energy of oxygen with C_i is 1.7 eV according to the calculations of Coutinho *et al.* (2002). The activation energy for dissociation of B_iO_i is approximately $W + E_B = 1.2$ eV when the migration energy W of BI measured by Troxell and Watkins (1980) (0.6 eV) is used. This is in excellent agreement with the value of 1.20 ± 0.05 eV measured by Mooney *et al.* (1977).

The single donor (0/+) level of B_iO_i is, using interstitial carbon as a marker, calculated to lie around $E_c - 0.35$ eV, near to the observed level. The second donor $(+/++)$ level is calculated to lie close to (possibly even below) the valence band top when the sulphur substitutional pair, which is a double donor is used as a marker defect. Calculations find no acceptor $(-/0)$ level in the band-gap. Thus the level observed at $E_c - 0.23$ eV can be assigned to the *single* donor level of B_iO_i and not the double donor as suggested by Mooney *et al.* (1977) on the grounds of it having a large electron capture cross section ($\sigma_n = 3 \times 10^{-13}$ cm^2) before electron capture and a small hole capture cross section ($\sigma_p \sim 10^{-20}$ cm^2) after the electron capture. Mooney *et al.* (1977) point out that these capture cross sections suggest that the defect was positively charged before and after electron capture and hence the level must be a second donor $(+/++)$ level. There is however experimental evidence in support of the *single* donor activity of B_iO_i. The loss of carriers in electron and proton-irradiated Cz p-type silicon can be correlated with the growth in concentration of B_iO_i and the other dominant radiation induced defects in that temperature range (V_2 and C_iO_i) (Yamaguchi *et al.*, 1999). To obtain good agreement between carrier and defect concentration it must be assumed that B_iO_i possesses a single and not double donor level. In conclusion B_iO_i is formed when BI becomes mobile at 230 K and diffuses to O_i [2] with an activation energy of 0.6 eV and a pre-exponential

[2] Interstitial oxygen has a migration energy of 2.5 eV and hence is immobile at 230 K (Mikkelsen, 1986)

factor of 5×10^7 s^{-1} as measured by Troxell and Watkins (1980). B_iO_i then gives rise to a single donor (0/+) level around $E_c - 0.23$ eV.

5.3 Carbon

At 150-200°C in Cz silicon the $E_c - 0.23$ eV level disappears and a new level appears at around $E_v + 0.29$ eV (Mooney *et al.*, 1977; Kimerling *et al.*, 1989). The defect responsible for this new level is formed with an activation energy and pre-exponential factor identical to that of the decay of the $E_c - 0.23$ eV level (Mooney *et al.*, 1977). The production rate of the new defect formed is independent of [O] but directly proportional to [C] and inversely proportional to [B] (Kimerling *et al.*, 1989).

If the $E_c - 0.23$ eV level is indeed B_iO_i then the evidence above suggests that when B_iO_i dissociates at ~ 150°C, BI diffuses to and is trapped by substitutional carbon generating B_iC_s with a level at $E_v + 0.29$ eV. However there are some difficulties in this interpretation (Yarykin *et al.*, 2004). In FZ silicon, one anticipates that, due to the lower oxygen concentration, B_iC_s could form at room temperature when BI or C_i becomes mobile. However, experiments show that in carbon-rich FZ silicon the $E_v + 0.29$ eV level is not formed until about 120°C (Drevinsky *et al.*, 1988; Monakhov *et al.*, 1997). One explanation for this observation could be the preferential formation of a metastable precursor $B_iC_s^*$ which is converted into B_iC_s via a barrier that can be surmounted only when the material is annealed to 120°C and which has so far evaded detection in DLTS studies. Such a metastable precursor may also explain how the concentration of B_iC_s can be several times that of B_iO_i (Yarykin *et al.*, 2004; Drevinsky *et al.*, 1988). The unexpected dependence of the production rate of B_iC_s upon boron concentration is perhaps not surprising given that for high boron concentration B_2I^Q (B_iB_s) is likely to be formed instead of B_iC_s.

The ground-state structure of B_iC_s is shown in figure 5.1 and is very similar to that of the B_iB_s defect (B_2I^Q) discussed in section 4.2. Note that B_iC_s can equivalently be described as C_iB_s. The complex has a single donor (0/+) level calculated, using interstitial carbon as a marker, to lie around $E_v + 0.26$ eV, very close to the observed level. Calculations also predict an acceptor $(-/0)$ level around $E_c - 0.3$ eV using the same marker. The complex could dissociate either by the release of C_i leaving B_s or by the release of BI leaving C_s. The former mechanisms is more likely as a self-interstitial is calculated to be ~0.7 eV more tightly bound to C_s than to B_s. The dissociation temperature is then given by $(W + E_B)/k \ln \nu$ with W, the migration energy of $C_i \sim 0.80$ eV (Watkins and Brower, 1976) and E_B, the binding energy of C_i and $B_s \sim 1.2$ eV. This temperature, around 400°C, is in very good agreement with the temperature of close to 400°C at which Mooney *et al.* (1977) and Drevinsky *et al.* (1988) observe the $E_v + 0.29$ eV level to disappear. The weight of evidence therefore points to the assignment of the $E_v + 0.29$ eV level to B_iC_s which is formed when B_iO_i dissociates and BI is trapped at C_s.

5.4 Hydrogen

In this section the effect of hydrogen on the evolution of boron interstitial clusters is discussed. A level observed at $\sim E_v + 0.51$ eV and labelled $H3$ (or $H\alpha 1$) is observed in boron doped Cz or FZ silicon that has been either room temperature electron-irradiated and then hydrogenated (Feklisova *et al.*, 2001) or room temperature electron irradiated after hydrogenation (Yarykin *et al.*, 2001) and in room temperature α-irradiated boron doped epitaxially grown Si (which contains much grown in hydrogen) (Mamor *et al.*, 1998) and in proton-irradiated boron doped FZ silicon (Mamor *et al.*, 2001). The level is not observed in material that has been made p-type by aluminium, instead of boron doping (Feklisova *et al.*, 2001). Hence the link to both hydrogen and boron is clear. The $H3$ complex forms so efficiently that its concentration can exceed that of all other DLTS centres present with levels in the lower half of the gap (Feklisova *et al.*, 2001). In hydrogenated FZ (or epi) boron doped Si, the level is observed in as-irradiated samples. In Cz silicon however, the sample must be annealed to a temperature in the range 170-370°C subsequent to the irradiation, then cooled and hydrogenated at room temperature in order to produce $H3$.

The fact that Cz silicon and only Cz silicon (which has high oxygen concentration) requires annealing at above 170°C strongly suggests that the precursor of $H3$ (unhydrogenated $H3$) is a complex other than B_iO_i. In fact no level that can be assigned to the $H3$ precursor has been detected by DLTS. The requirement of a room temperature irradiation or implantation to form $H3$ implies the involvement of an irradiation-induced defect that is stable at room temperature. This suggests two likely candidates for the $H3$ centre: B_iB_sH and B_iC_sH (Yarykin *et al.*, 2004). An assignment of $H3$ to B_iH, one of the candidates put forward by Volpi *et al.* (2004), is ruled out based on the fact that $H3$ can be formed by a room temperature irradiation followed by hydrogenation implying that the unhydrogenated precursor to $H3$ must be stable at room temperature which BI clearly is not. The assignment of the $E_v + 0.29$ eV level to B_iC_s makes the assignment of the $H3$ level to B_iC_sH unlikely, mainly due to the fact that no levels of the $H3$ precursor have been detected. It is possible that an undetected metastable precursor of B_iC_s, ($B_iC_s^*$) could lead to $H3$, but it would then be necessary to argue that not all $B_iC_s^*$ is converted to the stable form B_iC_s on heating. Moreover, the most stable structure of B_iC_sH is the same as that of B_iC_s, shown in figure 5.1 but with a hydrogen atom (not shown) attached to the C atom. This leaves carbon four-fold coordinated and B three-fold coordinated and leads to an electrically inactive defect. Hence, B_iC_sH can not be assigned to the $H3$ centre which has a level near mid-gap.

B_iB_s (B_2I^Q) on the other-hand is a much stronger candidate for the $H3$ precursor since it is known to be formed by the electron irradiation of boron-doped Si. B_iB_s is electrically inactive and hence invisible to DLTS. Furthermore, B_iB_s is not expected to form in as-irradiated Cz silicon since BI formed by the irradiation would be trapped by oxygen giving B_iO_i instead. The $H3$ precursor (B_iB_s) could then be formed only by annealing at a temperature high enough to dissociate B_iO_i but not so high as to destroy B_iB_s which is removed when V_2 becomes mobile (Bean *et al.* (1972) and section 4.2). Hence B_iB_s is an appealing candidate for the $H3$ precursor. If B_iB_s is hydrogenated it forms the structure shown in figure 5.1. In the positive charge state the symmetry changes from C_{1h} to C_{3v}. B_iB_sH has both a donor and acceptor level calculated to lie close together and around $E_v + 0.5$ eV when bulk silicon is used as a marker. It is believed to be a negative-U system as the charge driven bistability hints. Due to the generation mechanism and stability of its precursor and the position of its electrical levels, it is B_iB_sH that is assigned to the $H3$ centre. B_iB_sH is predicted to dissociate into B_iB_s and H_i around 430°C assuming that the migration energy of hydrogen W_{H_i} is 0.2 eV (Van de Walle *et al.*, 1988).

Potential problems with the assignment of B_iB_sH to $H3$ have been identified by Yarykin *et al.* (2004). The introduction rate of $H3$ is expected to be quadratic with boron concentration if the assignment of $H3$ to B_iB_sH is correct. For [B]$\sim 10^{15}$ cm^{-3} the introduction rate is observed to be linear and not quadratic and appears to saturate at around 10^{16} cm^{-3} (Yarykin *et al.*, 2004). However, it is now believed (Yarykin, 2004) that the introduction rate of $H3$ was limited by the amount of hydrogen present in the material and not by the boron concentration as believed previously. Hence the measurement of the introduction rate as a function of boron concentration must be repeated in material with a greater concentration of hydrogen.

It is concluded that B_iB_s is the precursor to $H3$ and that the complex may be activated by the trapping of a single hydrogen atom giving rise to the $H3$ centre with donor and acceptor levels around mid-gap. An alternative formation mechanism for $H3$ occurs when the material is hydrogenated prior to irradiation. In this case BI, produced by the irradiation, is trapped by B_sH possibly forming a metastable $B_iB_sH^*$ defect (MH3 observed by Yarykin *et al.* (2001)) which evolves into B_iB_sH at around 150°C. If B_iB_sH were to trap a second hydrogen, it would once again become passive although this does not appear to occur even for long hydrogenation times (Yarykin *et al.*, 2004). Again it now appears likely that the lack of observed passivation is simply due to the fact that the hydrogen concentration was overestimated (Yarykin, 2004). An alternative explanation is that since the reaction $VOH + H \rightarrow VOH_2$ [3] is about 0.9 eV more exothermic than the reaction $B_iB_sH + H \rightarrow B_iB_sH_2$, VOH_2 is preferentially formed. However, the former explanation is more appealing since it is also able to explain the linear introduction rate of $H3$ with boron concentration mentioned above.

[3] The binding energy of H to VOH had been calculated by Coutinho *et al.* (2000).

5.5 Summary

In chapter 4 it was shown that the boron interstitial defect (BI), generated by implantation or irradiation, is trapped by substitutional boron when it becomes mobile at around 230 K. The resulting B_iB_s (B_2I) complex is electrically inactive. B_iB_s (B_2I) can be formed directly by room temperature implantation or irradiation. Here it has been demonstrated that in the presence of hydrogen B_iB_s becomes electrically active with donor and acceptor level around $E_v+0.51$ eV and it is predicted to be stable up to $\sim 400°$C. An alternative reaction path occurs in Cz material or material with lower boron concentration. In this material BI is trapped in large concentrations by oxygen forming B_iO_i and giving rise to a single donor (0/+) level at around $E_c - 0.23$ eV. B_iO_i anneals by dissociation between 150 and 200°C and subsequently B_iC_s is formed. B_iC_s has a donor (0/+) level at $E_v + 0.29$ and like B_iB_sH is stable until 400°C when it dissociates.

Having understood the reactions leading to the most dominant, electrically active, interstitial defects in p-type silicon up to 400°C, the focus will now be turned to a complex of boron and oxygen that is believed to exist in in solar cells, in far lower concentrations than B_iO_i does in irradiated material, yet it causes a drastic degradation of solar cell efficiency. This complex is the focus of the next chapter.

Chapter 6

Degradation of boron doped Czochralski silicon solar cells

6.1 Introduction

There are many social and economic pressures on the development of cheap, renewable energy sources. On Earth we receive just over 1×10^3 Wm^{-2} from the Sun[1] and so solar power is a promising source of energy. Solar cells are one important device available for harnessing this energy, not only on Earth but also for example on satellites where a portable and renewable source of electricity is hugely desirable. Henry (1980) has shown that in principle it is possible to create a single band-gap solar cell with an external quantum efficiency of 31% for an illumination intensity of 1 sun. The efficiency of a single band-gap solar cell is largely limited by the fact that photons with energy less than the band-gap are transmitted through the semiconductor while those with energy greater than the band-gap generate excitons where the electron is far above the conduction band bottom, E_c and the

[1] This quantity of energy defines the value of 1 sun.

hole is far below the valence band top, E_v. Energy possessed by the exciton in excess of the band-gap energy is quickly lost as the electron and hole fall to E_c and E_v respectively.

It is desirable to manufacture solar cells from semiconductors used in devices, none more so than silicon. For a solar cell having the band-gap of silicon the maximum theoretical efficiency is around 29% (Henry, 1980). To date the highest efficiency achieved by a n^+-p silicon solar cell is 24.7% (Zhao *et al.*, 1998). This was achieved in mono-crystalline FZ material. But FZ material, though containing far less oxygen than Cz material (the importance of this will become apparent in the next section) is relatively expensive, a fact which has spurred interest in Cz silicon solar cells. In Cz material there are around two orders of magnitude more oxygen ($\sim 10^{17}$ cm^{-3} or larger) and the maximum efficiencies achieved are rather less than FZ material at around 19-20% (Bothe, 2004). Despite the significantly lower efficiency, Cz solar cells play an important role due to their lower cost and are preferred to FZ solar cells when the size of solar cell required, which is inversely proportional to cell's efficiency, is less of an issue.

A major problem faced by n^+-p Cz solar cells is that of light-induced degradation. The external efficiency of the solar cell reduces during operation due to the light-induced generation of defects. These defects act as recombination centres, trapping minority and majority carriers (electrons and holes respectively in the solar cells of interest here where the active region is p-type). For early Cz solar cells the degradation was found to be due to the light induced dissociation of Fe_iB_s pairs (Reiss *et al.*, 1996). Fe_iB_s has a donor (0/+) level at $E_v + 0.10$ eV (Feichtinger, 1979) and an acceptor level at around $E_c - 0.29$ eV (Lemke, 1981; Brotherton *et al.*, 1985). After illumination, dissociated Fe_i gives rise to a donor (0/+) level at $E_v + 0.38$ eV (Feichtinger *et al.*, 1978). Fe_i acts as a much stronger recombination than Fe_iB_s resulting in a decrease in minority carrier lifetime with illumination. Even an iron concentration as low as 10^{11}cm^{-3} can result in degradation (Reiss *et al.*, 1996) and so it is absolutely critical that iron contamination is avoided.

Even in the absence of iron, degradation is observed in Cz solar cells. The external efficiency is reduced by about one tenth following illumination (Fischer and Pschunder, 1973; Knobloch *et al.*, 1996) or minority carrier injection (Knobloch *et al.*, 1996). This degradation has been attributed to carrier recombination at a boron-oxygen centre (Schmidt *et al.*, 1997; Schmidt and Cuevas, 1999; Schmidt and Bothe, 2004) and it is this boron-oxygen defect that is the concern of this chapter.

6.2 The boron oxygen complex

Minority carrier lifetime spectroscopic studies have shown that the concentration of this recombination centre is directly proportional to boron concentration (Glunz *et al.*, 1998; Schmidt and Cuevas, 1999; Schmidt and Bothe, 2004) and to $[O]^2$ (Schmidt *et al.*, 2002; Schmidt and Bothe, 2004), suggesting a complex of a single boron atom with a pair of oxygen atoms. This dependence is contrary to earlier results which found a stronger oxygen dependence (Glunz *et al.*, 1998) but the quadratic dependence upon oxygen concentration has now been confirmed (Rein *et al.*, 2003).

Hydrogen may be introduced during processing steps such as etching. By measuring the passivation of substitutional boron by hydrogen, as a function of distance from the crystal surface it has been shown that etching introduces hydrogen only to a depth $\sim$ 10 μm (Weber *et al.*, 2003). Since the lifetime degradation has been shown to be due to a bulk property of the material (Schmidt and Cuevas, 1999), the involvement of hydrogen introduced by such processing can probably be excluded.

The generation of the defect involves two processes (Bothe *et al.*, 2004): a fast process which will not be discussed here and a slower process activated by an energy $E_A^{gen} =$ 0.37 eV (Schmidt *et al.*, 2002; Bothe *et al.*, 2004). For the slower process, its generation rate per centre is proportional to $[B]^2$ (Rein *et al.*, 2001; Bothe *et al.*, 2003) and to the light intensity up to a critical value of just 10^{-2} suns after which the generation rate saturates (Schmidt and Bothe, 2004). The defect anneals out in the dark at $\sim 200°$C with an activation energy of $E_A^{ann} = 1.3$ eV (Rein *et al.*, 2001; Bothe *et al.*, 2003; Schmidt and Bothe, 2004).

Schmidt *et al.* (Schmidt *et al.*, 2002; Bothe *et al.*, 2004; Schmidt and Bothe, 2004) have put forward a model where interstitial oxygen pairs O_{2i}, already present in the material,

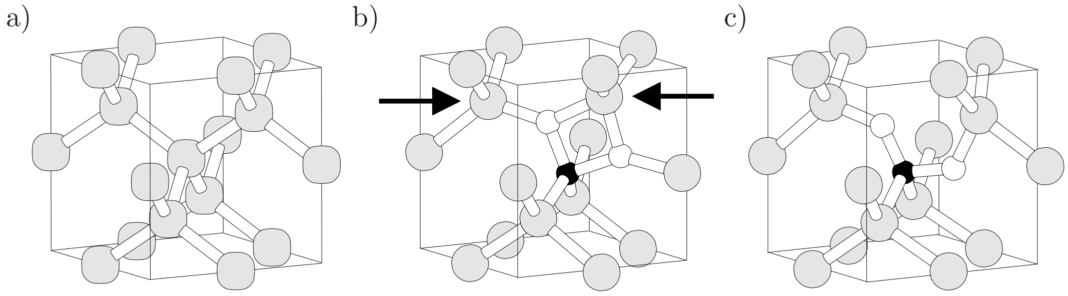

Figure 6.1: The structure of the two most stable forms of B_sO_{2i}. Large grey balls represent silicon, small white are oxygen and small black are boron. The square form of B_sO_{2i}, $B_sO_{2i}^{sq}$ (structure *b* above) is the most stable form in the single positive charge state and is metastable in the neutral charge state. The staggered form, $B_sO_{2i}^{st}$ (structure *c* above) is the most stable form in the neutral charge state. An equivalent piece of perfect bulk silicon is shown (structure *a* above) to aid the reader. The arrows indicate the Si atoms on which the largest fraction of the donor state wavefunction resides.

are converted to some fast diffusing form they call $O_{2i}^{\star}$. E_A^{gen} then corresponds to the diffusion barrier of $O_{2i}^{\star}$ which migrates quickly through the material until it is trapped by B_s forming the recombination centre B_sO_{2i}. The centre has so far evaded detection by DLTS, possibly indicating that it is present only in a very low concentration, but injection-level dependent carrier lifetime measurements place the level at which the minority carrier trapping occurs at between $E_v + 0.35$ and $E_c - 0.45$ eV(Schmidt and Cuevas, 1999). Using advanced lifetime spectroscopy the level was pinned down to $E_c - 0.41$ eV (Rein and Glunz, 2003). Although absolute carrier capture cross sections have not been determined, the ratio of electron capture cross section to hole capture cross section (σ_n/σ_p) has been determined by Rein and Glunz (2003) to be 9.3 implying that the centre is an efficient electron trap.

The present theoretical work indicates that the most stable complex formed when substitutional boron has trapped an oxygen pair is that shown in figure 6.1. The complex exhibits a charge driven bistability. In the positive charge state the oxygen atoms are over-coordinated as indeed found for other oxygen defects like C_iO_i (Coutinho *et al.*, 2002), B_iO_i (section 5.2) and thermal donors (for example Chadi (1996); Pesola *et al.* (2000);

Coutinho *et al.* (2001)). This square configuration is labelled $B_sO_{2i}^{sq}$. In the neutral or negative charge state the oxygen atoms are two fold coordinated and this staggered configuration is labelled $B_sO_{2i}^{st}$. $B_sO_{2i}^{sq}$ is 0.4 eV more stable in the single positive charge state while in the neutral charge state it is metastable to $B_sO_{2i}^{st}$ by 0.4 eV. In the negative charge state $B_sO_{2i}^{sq}$ is unstable.

The thermodynamic acceptor $(-/0)$ level is placed around $E_v + 0.3$ eV and corresponds to the $B_sO_{2i}^{st}$ configuration that is the ground-state in both the neutral and negative charge states. The thermodynamic donor $(0/+)$ level corresponds to the reaction

$$B_sO_{2i}^{+sq} + e^- \rightarrow B_sO_{2i}^{0st} \tag{6.1}$$

and is placed in the lower half of the band-gap around $E_v + 0.4$ eV. However, this reaction can be considered to proceed via two steps. First

$$B_sO_{2i}^{+sq} + e^- \rightarrow B_sO_{2i}^{0sq} \tag{6.2}$$

and then

$$B_sO_{2i}^{0sq} \rightarrow B_sO_{2i}^{0st}. \tag{6.3}$$

The rate for the reverse of reaction 6.2 (*i.e.* $B_sO_{2i}^{0sq}$ capturing a hole to return to $B_sO_{2i}^{+sq}$) is τ_{hc} where

$$\tau_{hc} = p\langle v_h \rangle \sigma_p. \tag{6.4}$$

The hole capture cross section, σ_p is estimated to be 1×10^{-15} cm^2. The mean thermal velocity of holes is given by

$$\frac{3}{2}kT = \frac{1}{2}m_p^*\langle v_p^2 \rangle \tag{6.5}$$

where m_p^*/m_0 has been assumed to be 0.16. The hole concentration p is the sum of the doping generated holes p_{dope} and the illumination generated holes p_{ill}. The p-type doping concentration is estimated to be 1×10^{15} cm^{-3} giving p_{dope}. The additional p-type doping due to illumination is given by $p_{ill} = Gt_{hl}$ where $G = I/(dE_{eh})$. The hole lifetime t_{hl} is estimated to be 1×10^{-7} s, the illumination intensity I is taken to be 1000 Wm^{-2} (1 sun), d the illumination attenuation distance is taken to be 1×10^{-4} cm and E_{eh} is the energy to create an ionised electron-hole pair in silicon, approximately 3.6 eV.

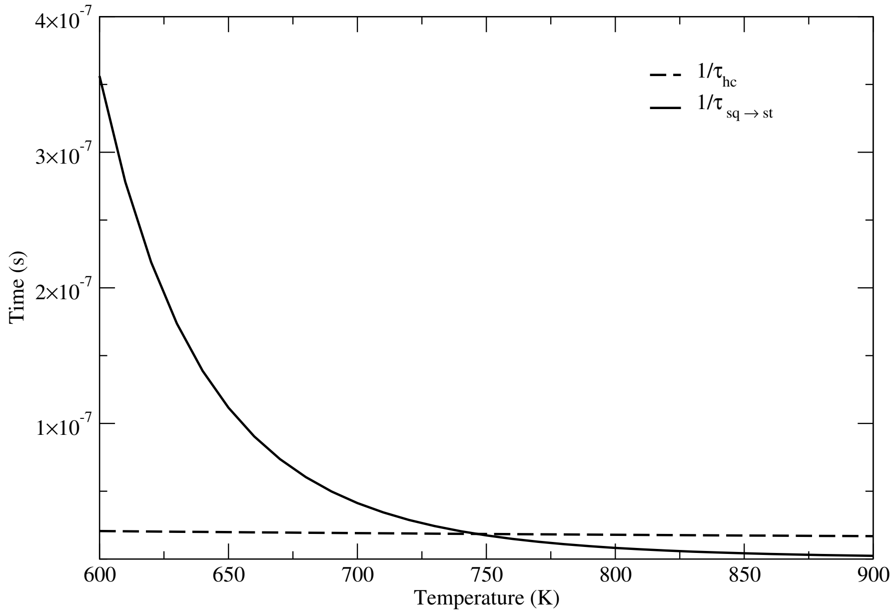

Figure 6.2: The time taken for $B_sO_{2i}^{0sq}$ to capture a hole ($1/\tau_{hc}$) and the time taken for $B_sO_{2i}^{0sq}$ to be converted to the stable form in the neutral charge state, $B_sO_{2i}^{0st}$ ($1/\tau_{\mathrm{sq}\to\mathrm{st}}$). Note that at room temperature hole capture will take place long before B_sO_{2i} is able to reconfigure.

Reaction 6.3 will proceed at a rate $\tau_{\mathrm{sq}\to\mathrm{st}}$ where,

$$\tau_{\mathrm{sq}\to\mathrm{st}} = \nu \exp\left(\frac{-E_A}{kT}\right). \tag{6.6}$$

The activation energy for reaction 6.3, E_A in the above equation, is calculated to be 0.78 eV.

Using the estimated values given above, at room temperature equations 6.4 and 6.6 predict that the reverse of reaction 6.2 will occur 8 orders of magnitude more quickly than reaction 6.3. The values of $1/\tau_{hc}$ and $1/\tau_{\mathrm{sq}\to\mathrm{st}}$ are plotted in figure 6.2 as a function of temperature. This calculation indicates that the relevant donor (0/+) level at around room temperature in the solar cells is that corresponding to reaction 6.2 and not the thermodynamic donor level which corresponds to reaction 6.1. This is calculated, using interstitial carbon as a marker, to lie in the upper part of the band-gap around $E_c - 0.3$ eV in good agreement with the level observed by Rein and Glunz (2003). The donor level can be clearly seen if the Kohn-Sham bandstructure of $B_sO_{2i}^{\mathrm{sq}}$ is plotted (figure 6.3). Note that the fact that the complex is positively charged in p-type material means that it should act as a strong

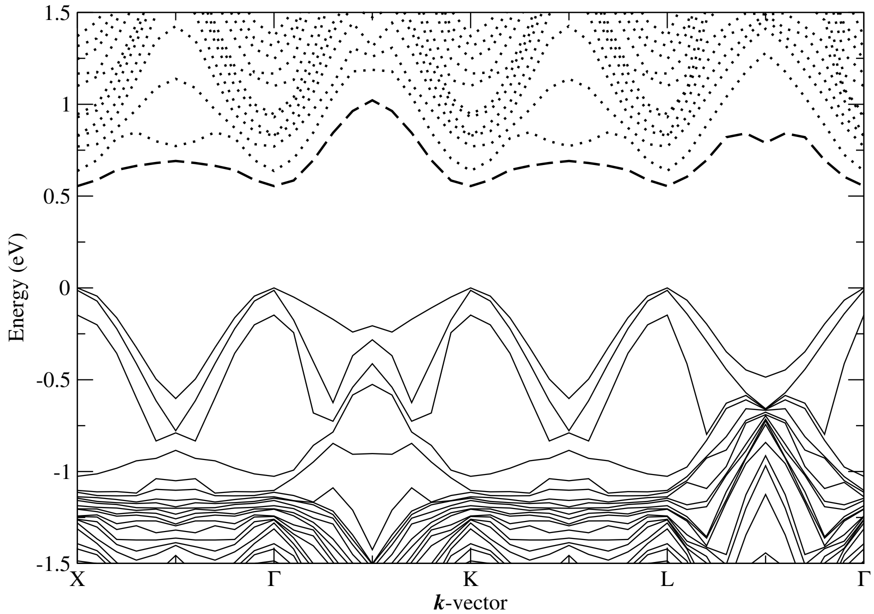

Figure 6.3: The Kohn-Sham band-structure of $B_sO_{2i}^{sq}$ calculated in a 64 atom supercell. Dotted lines represent conduction band levels, solid lines indicate valence band levels and the dashed line represents the donor (0/+) level.

electron trap, consistent with the observed ratio of capture cross sections. Thus the boron oxygen complex $B_sO_{2i}^{sq}$ is found to have very similar properties to the lifetime degrading defect observed in Cz silicon.

A Mulliken analysis and a study of the wavefunction of the highest partially filled level of neutral $B_sO_{2i}^{sq}$ have been performed. This level which corresponds to the donor (0/+) level is localised mainly on the silicon atoms indicated by arrows in figure 6.1 and is largely p-orbital like in character.

6.3 The formation mechanism of the boron oxygen complex

Having demonstrated that B_sO_{2i} has many properties in common with the observed lifetime degrading defect, in order to add weight to a possible assignment it is necessary to understand how B_sO_{2i} may be formed and destroyed. This involves understanding the physical significance of the activation energies $E_A^{gen} = 0.37$ eV and $E_A^{ann} = 1.3$ eV discussed

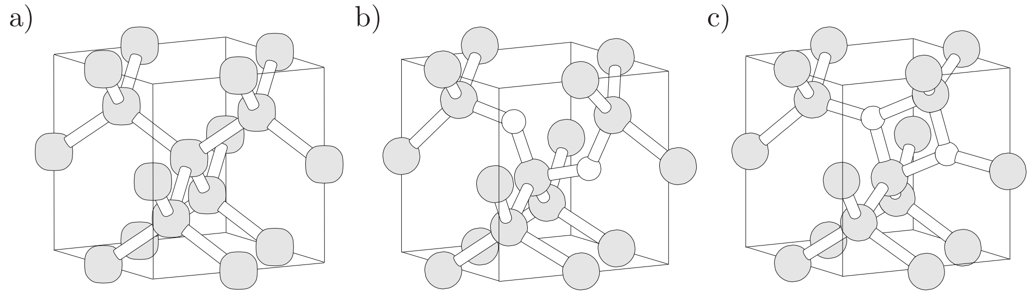

Figure 6.4: The structures of the oxygen pair. The stable form of the pair in the neutral and single positive charge states is labelled O_{2i}^{st} (structure *b* above). O_{2i}^{sq} (structure *c* above) is the stable form of the pair in the double positive charge state. A piece of equivalent bulk material (structure *c*) is also shown. As before, grey balls represent Si and small white ones represent O.

in the previous section. In order to understand the formation and annealing of B_sO_{2i} and to test the model of Schmidt *et al.* (Schmidt *et al.*, 2002; Bothe *et al.*, 2004; Schmidt and Bothe, 2004), the oxygen pair, the mobile constituent in the formation of B_sO_{2i}, has been studied. Interstitial O_i has a migration energy of 2.5 eV (Mikkelsen, 1986) and thus is immobile at room temperature. However, oxygen is bound together in the form of pairs (O_{2i}) by $E_B \sim 0.3$ eV (Murin *et al.*, 1998) which it is likely are formed and frozen into the wafers during heat treatments to remove thermal donors. The pair has been observed by infrared absorption and its concentration is found to depend upon $[O_i]^2$ (Murin *et al.*, 1998) which would in turn lead to the quadratic dependence upon oxygen concentration for the concentration of lifetime degrading complexes, observed experimentally (Schmidt *et al.*, 2002; Schmidt and Bothe, 2004; Rein *et al.*, 2003). For this reason the properties and in particular the migration energy of O_{2i} was investigated.

In the neutral and single positive charge state the staggered form of the pair, shown in figure 6.4, is the more stable by ~ 1.3 and ~ 0.6 eV respectively. In the double positive charge state the stability is reversed and the square form, also shown in figure 6.4, is more stable by ~ 0.9 eV. The pair diffuses by cycling between the square and staggered form with the addition of a 180° precession of the oxygen atoms around their bond centres. This precession costs significantly less that the reconfiguration from staggered to square forms and does not contribute to the migration barrier. A configuration coordinate diagram is

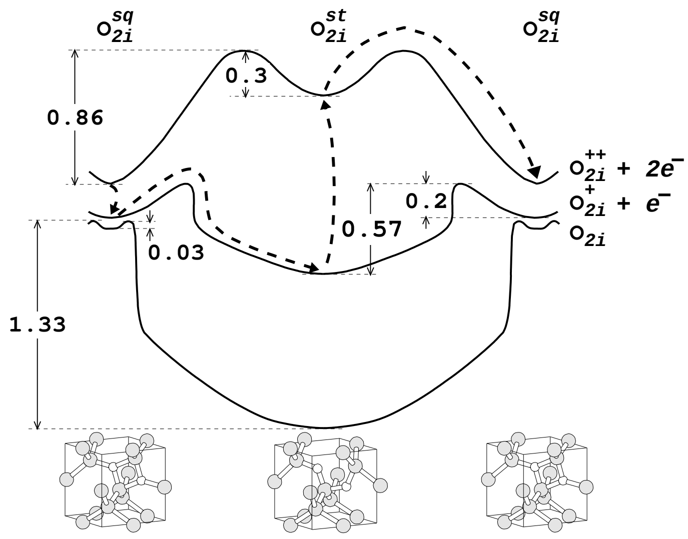

Figure 6.5: A configuration coordinate diagram for the oxygen pair. Arrows show the proposed thermally assisted Bourgoin diffusion mechanism with a thermal barrier of 0.3 eV. O_{2i}^{++sq} at A first captures a photo-generated or injected electron and, after overcoming a 0.2 eV barrier, changes its configuration to O_{2i}^{+st}. It then traps a hole becoming O_{2i}^{++st}, and executes a diffusion jump to O_{2i}^{++sq} at B after overcoming a thermal barrier of 0.3 eV.

presented in figure 6.5 showing the migration barrier in each charge state. In the neutral charge state pairs are calculated to diffuse with an energy of about 1.3 eV in agreement with previous calculations by Coutinho *et al.* (2000) and Lee *et al.* (2001). The saddle point for the migration is close to the square configuration, O_{2i}^{sq} shown in figure 6.4. In the single positive charge state the migration path is similar but the migration energy is reduced to 0.57 eV. In the double positive charge state, where the relative stabilities of the square and staggered forms are reversed, the migration energy is 0.86 eV which is again smaller than for the neutral pair.

The square configuration of the pair has a filled level high in the band-gap, shown in figure 6.6, which is empty for O_{2i}^{++} and suggests that the pair can be a bistable defect. The first and second donor levels of the pair are calculated to lie at about $E_c - 1.0$ eV and $E_c - 0.6$ eV respectively when the sulphur substitutional pair is used as a marker. Since the first donor level lies below the second, the defect has negative-U, consistent with a charge-driven bistability, with the occupancy $(0/++)$ level lying at $\sim E_v + 0.4$ eV. In p-type Si, with

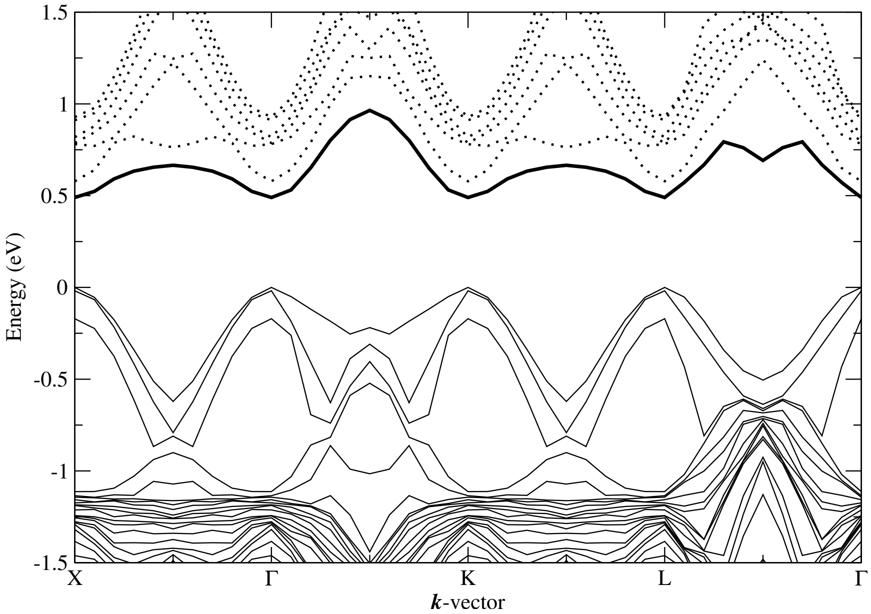

Figure 6.6: The Kohn-Sham bandstructure of O_{2i}^{sq}. When the Fermi-level lies below the heavy solid line and above 0 eV, O_{2i}^{sq} is in the double positive charge state.

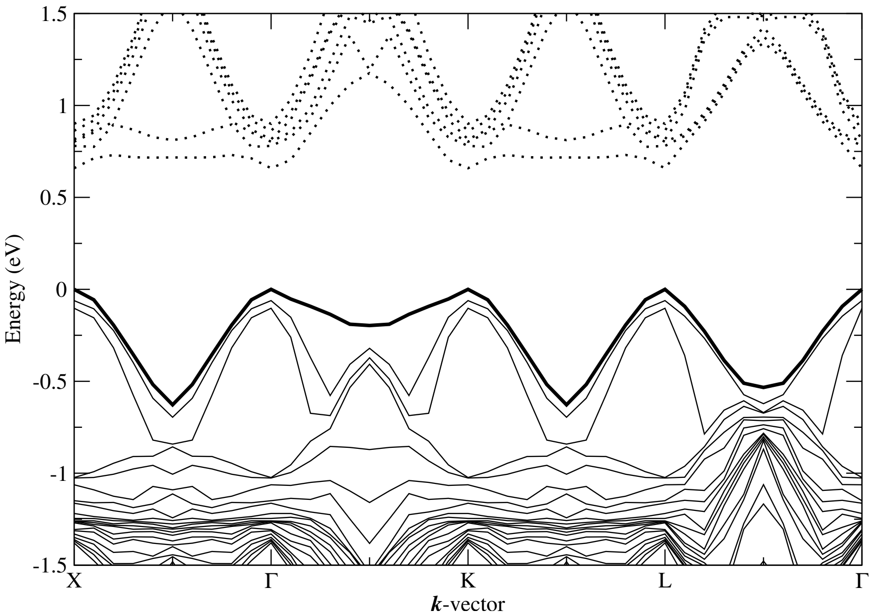

Figure 6.7: The Kohn-Sham bandstructure of O_{2i}^{st}. In the neutral charge state solid lines are filled levels and dotted lines represent unoccupied levels. It is suggested that the highest occupied level (shown as a heavy line) may be a donor (0/+) level of O_{2i}.

the Fermi level below this level, the pair would therefore be present as O_{2i}^{++sq}. However, if the first donor level, which corresponds to the reaction $O_{2i}^{+st} + e^- \rightarrow O_{2i}^{0st}$, lay very close to or beneath the valence-band top, a possible interpretation of the bandstructure of O_{2i}^{st} (figure 6.7), then hole capture would be less efficient and only a fraction of the pairs would be found in the double positive charge state.

If the pair is indeed stable in the double positive charge state in p-type material then the diffusion mechanism described by the arrows in figure 6.5 is possible. The pair, present in its stable O_{2i}^{++sq} form, can trap a minority carrier electron, generated by the illumination or alternatively by the application of a forward bias, thus becoming O_{2i}^{+sq}. It may then surmount the small 0.2 eV barrier to the stable configuration in the positive charge state, O_{2i}^{+st}. From here the pair can capture a majority carrier hole becoming O_{2i}^{++st}. To become O_{2i}^{++sq} at a new site the pair then only has to overcome a 0.3 eV barrier. Hence the migration energy for the pair in p-type material, under minority carrier injection is calculated to be just 0.3 eV, rather close to the observed activation energy for the formation of the lifetime degrading defect of 0.37 eV. Furthermore such a diffusion mechanism should result in a linear dependence of the production rate of B_sO_{2i} with illumination as observed for the lifetime degrading defect.

The proposed pair migration mechanism is summarised by,

$$\begin{aligned}
O_{2i}^{++sq} + e^- &\rightarrow O_{2i}^{+sq} && (6.7)\\
O_{2i}^{+sq} &\xrightarrow{0.2\text{ eV}} O_{2i}^{+st} && (6.8)\\
O_{2i}^{+st} + h^+ &\rightarrow O_{2i}^{++st} && (6.9)\\
O_{2i}^{++st} &\xrightarrow{0.3\text{ eV}} O_{2i}^{++sq}. && (6.10)
\end{aligned}$$

The rate that each of each of these reactions, and those reaction competing with them, have been estimated as described in section 6.2 for reactions 6.2 and 6.3. The product of reaction 6.7 will be removed by both electron and hole capture,

$$\begin{aligned}
O_{2i}^{+sq} + h^+ &\rightarrow O_{2i}^{++sq} && (6.11)\\
O_{2i}^{+sq} + e^+ &\rightarrow O_{2i}^{0sq}. && (6.12)
\end{aligned}$$

These reactions occur 3 and 2 orders of magnitude more slowly than reaction 6.8 respectively and hence $O_{2i}^{++\mathrm{sq}}$ can quickly be converted to $\mathrm{O}_{2i}^{+\mathrm{st}}$. The second part of the diffusion mechanism (reactions 6.9 and 6.10) is now considered. The rate of the reaction,

$$\mathrm{O}_{2i}^{++\mathrm{st}} + e^{-} \rightarrow \mathrm{O}_{2i}^{+\mathrm{st}} \tag{6.13}$$

is an order of magnitude greater than that of reaction 6.10 and hence will often occur before the 0.3 eV barrier can be overcome to result in $\mathrm{O}_{2i}^{++\mathrm{sq}}$ at a new site. This bottle neck suggests that the pre-exponential term defining the reaction rate is small. The prefactor though not explicitly stated in the literature, can be estimated from the Arrhenius plot presented by Hashigami *et al.* (2003) and is indeed found to be small ($\sim 2.65 \times 10^{3}$ s^{-1}) (Palmer, 2004).

Mobile oxygen pairs will be trapped both by oxygen atoms forming O_{3i}^{++} and by B_s^{-} forming $\mathrm{B}_s\mathrm{O}_{2i}^{+}$. The capture radius in the latter process will be considerably enhanced by the Coulomb attraction between B_s^{-} and O_{2i}^{++}. The binding energy between B_s^{-} and O_{2i}^{++} has been estimated from the change in energy when B_s^{-} and O_{2i}^{++} are separated in a large 144 atom supercell. Moving B_s^{-} away from $\mathrm{O}_{2i}^{++\mathrm{sq}}$ by either one or two steps along the $\langle 110 \rangle$ chain of bonds results in an increase of energy of 0.38 eV in each case. Hence B_s^{-} is bound to O_{2i}^{++} by an energy of around 0.38 eV.

If the model proposed here is correct, the production rate of lifetime degrading centres should be proportional to the concentration of the double positively charged pair given by

$$[\mathrm{O}_{2i}^{++}] = A \exp\left(\frac{-E_f^{\mathrm{O}_{2i}}(++)}{kT}\right) \tag{6.14}$$

where $E_f^{\mathrm{O}_{2i}}(++)$ is the formation energy of $\mathrm{O}_{2i}^{++\mathrm{sq}}$, given by equation 1.76. From equation 1.76 and the fact that the position of a band-gap $(q/q+1)$ level, $E(q/q+1)$ is defined by the point at which the formation energy of the defect is equal for charge states q and $(q+1)$, it is clear that for the donor $(0/+)$ level of the oxygen pair,

$$\begin{aligned} E_f^{\mathrm{O}_{2i}}(0) &= E_f^{\mathrm{O}_{2i}}(+) && (6.15)\\ E_T^{\mathrm{O}_{2i}}(0) &= E_T^{\mathrm{O}_{2i}}(+) + E_F && (6.16)\\ \Rightarrow E(0/+) &= E_T^{\mathrm{O}_{2i}}(0) - E_T^{\mathrm{O}_{2i}}(+) && (6.17) \end{aligned}$$

where $E_T^{O_{2i}}(0)$ and $E_T^{O_{2i}}(+)$ are the total energy of O_{2i}^{0st} and O_{2i}^{+st} respectively and E_F is the Fermi-energy. Similarly

$$E(+/++) = E_T^{O_{2i}}(+) - E_T^{O_{2i}}(++) \tag{6.18}$$

hence

$$\begin{aligned} E(0/++) &= \frac{E(0/+) + E(+/++)}{2} && (6.19) \\ &= \frac{E_T^{O_{2i}}(0) - E_T^{O_{2i}}(++)}{2}. && (6.20) \end{aligned}$$

Using the definition of formation energy E_f (equation 1.76), equation 6.14 can be written

$$[O_{2i}^{++}] \propto \exp\left(\frac{-[E_T^{O_{2i}}(++) + 2E_F]}{kT}\right). \tag{6.21}$$

Substituting in the value of $E_T^{O_{2i}}(++)$ from equation 6.20

$$[O_{2i}^{++}] \propto \exp\left(\frac{-[E_T^{O_{2i}}(0) - 2E(0/++) + 2E_F]}{kT}\right) \tag{6.22}$$

and since

$$\begin{aligned} [O_{2i}^{0}] &= A\exp\left(\frac{-E_f^{O_{2i}}(0)}{kT}\right) && (6.23) \\ &\propto \exp\left(\frac{-E_T^{O_{2i}}(0)}{kT}\right) && (6.24) \end{aligned}$$

the concentration of O_{2i}^{++sq} is given by,

$$[O_{2i}^{++}] \propto [O_{2i}^{0}]\exp\left(\frac{2E(0/++) - 2E_F]}{kT}\right). \tag{6.25}$$

Finally, using the fact that $E_F = kT\ln(N_v/[B])$ where N_v is the effective density of states of the valence band (Sze, 1981*a*) it is seen that

$$[O_{2i}^{++}] \propto \frac{[B]^2}{N_v^2}[O_{2i}^{0}]\exp\left(\frac{2E(0/++)]}{kT}\right) \tag{6.26}$$

and the generation rate is proportional to $[B]^2$ as observed by Rein *et al.* (2001).

The $B_sO_{2i}^{+}$ defect would anneal out in the dark, in the absence of excess minority carriers, with an activation energy, $E_A \approx E_B + E_m$ where E_B is the binding energy of O_{2i}^{++} to B_s^- and E_m is the migration energy of O_{2i}^{++}. Even in the presence of minority carriers it is predicted that should the complex become $B_sO_{2i}^{0}$, since the rate for hole capture is over an order of magnitude faster than dissociation, the complex will trap a majority carrier hole before it is able to dissociate. Therefore regardless of whether or not minority carriers are present, the defect will dissociate in the positive charge state with an activation energy $\sim$ 1.2 eV. The predicted activation for dissociation of B_sO_{2i} is in good agreement with the observed value of 1.3 eV.

An important alternative formation process that has not yet been discussed here is that of recombination-enhanced defect migration, a process considered in detail (though not for the oxygen pair) by Weeks *et al.* (1975). This process makes use of the energy released when an electron-hole pair recombine non-radiatively. For the case of the pair, the energy released by the recombination in addition to the 0.37 eV provided thermally (E_A^{gen}) would provide the energy required for the diffusion of the oxygen pair. If the pair is assumed electrically inert, a reasonable assumption given the bandstructure of O_{2i}^{st} (figure 6.7), then the recombination energy in addition to the thermal energy sum to the migration barrier of neutral O_{2i} (see figure 6.5). However it is difficult to understand how an electrically inert defect can subsequently trap an electron and hole to produce a recombination event which must occur close by if the energy is to be used for migration. If the pair is electrically active however it is possible that the recombination occurs at the double donor $(+/++)$ level, releasing around 1 eV (see figure 6.6), some of which may be dissipated by phonons while the remainder, plus the 0.37 eV thermal energy, allow the pair to overcome the 0.86 eV migration barrier in that charge state. This interesting mechanism clearly must be investigated in greater detail.

6.4 Summary and final remarks

It is proposed that the oxygen pair may diffuse with a low prefactor and low activation energy of ~ 0.37 eV in p-type material during illumination or minority carrier injection. The highly mobile pair is trapped by substitutional boron giving rise to a complex possessing a donor (0/+) level around $E_c - 0.4$ eV at which electrons and holes are efficiently and successively trapped. The defect thus leads to a reduction in the carrier lifetimes resulting in a reduction in solar cell efficiency.

Having a model for the complex responsible for the lifetime degradation of Czochralski silicon allows an understanding of preventative measures to be developed. One possible method for reducing the formation of the complex is to use Cz silicon with an increased carbon concentration. The trapping of oxygen at carbon should lead to a reduction in the formation of B_sO_{2i}, though the effectiveness of this technique is yet to be confirmed. Degradation can be avoided by using a p-type dopant other than boron. It has been demonstrated that gallium doped solar cells do not exhibit degradation (Schmidt, 2004, and references therein). A preliminary investigation of the Ga_sO_{2i} complex reveals that the staggered form is unstable due to the larger size of the dopant. The stable square form has no levels deep within the gap and does not act as a carrier trap. A similar result has been found for aluminium. Thus, *if economically viable*, doping with gallium or aluminium rather than boron is a possible solution to the problem of degradation.

Chapter 7

The solubility of boron and arsenic in biaxially strained silicon

7.1 Introduction

A reduction in device size requires an increase in doping concentration and a reduction in dopant diffusivity. Applying a strain to the host material can alter both the diffusivity and the solubility of impurities. Strained silicon is typically grown epitaxially on to SiGe which has a larger lattice constant and hence the lattice constant of silicon in the growth plane is expanded to match that of the SiGe. There is much interest in the use of SiGe with strained Si layers for devices since the reduction in symmetry of the silicon crystal results in an increased electron mobility. Hence developing an understanding of the effect of biaxial strain on the solubility and diffusivity of dopants in these systems is of great importance.

It has been demonstrated that biaxial strain introduced by growing Si epitaxially on SiGe can reduce the diffusivity of boron (Cowern *et al.*, 1994; Rajendran and Schoenmaker, 2001; Zangenberg *et al.*, 2003). For arsenic the effect of biaxial strain on diffusion is, to the knowledge of the author, yet to be investigated. While little is known about the effect of strain on the diffusion of dopants even less is known about the effect of strain upon their solubilities. The effect of biaxial strain on the solubility of both the acceptor boron and donor arsenic has not been studied experimentally but the change in equilibrium solubility of boron with biaxial strain has been studied theoretically by Sadigh *et al.* (2002). Using local density functional theory Sadigh *et al.* (2002) predicted that the equilibrium solubility of boron should be enhanced by $\sim 150\%$ for a -1% biaxial strain at $1000°$C (the convention that tensile strains are positive and compressive ones negative has been adopted). Sadigh *et al.* (2002) showed that this enhancement is due to two effects. The variation of the Fermi-energy with strain alters the stability of charged centres with respect to a neutral precipitate and the strain-induced change in lattice constant leads to increased stability and hence equilibrium concentration, of dopants that tend to induce a change in lattice constant of similar sense in unstrained material. As shall be confirmed in this chapter, both of these effects enhance the solubility of boron in compressively strained material but it is the Fermi-energy's variation with strain that dominates. The change in solubility with strain is calculated for both boron and arsenic following the method described by Sadigh *et al.* (2002, and references therein) for the case of boron. This method is described below.

7.2 Method

Inserting an impurity into an unstrained, perfect crystal leads to changes in lattice constant. The lattice parameter change is linear in dopant concentration and for a boron concentration of 4×10^{20} cm^{-3} corresponds to a strain of -0.002 (Cardona and Christensen, 1987*a*,*b*). For arsenic, the change of lattice parameter with doping is very much smaller (Cardona and Christensen, 1987*a*,*b*). Since the biaxial strain that results from growing silicon on some lattice-mismatched material such as SiGe is typically much larger than this, the calculated lattice parameter of pure bulk silicon is used as the reference

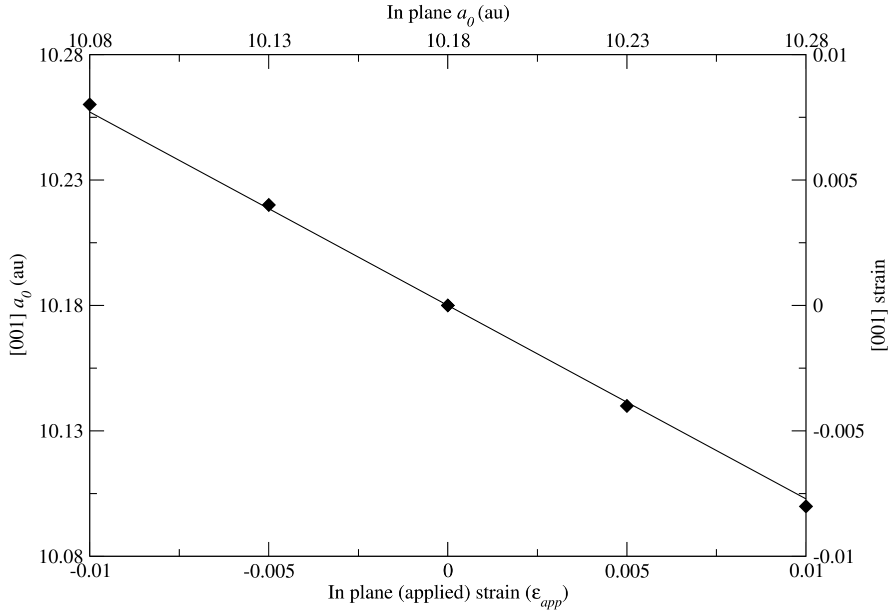

Figure 7.1: The variation of relaxed [001] lattice constant with biaxial-strain. Data points are the calculated values while the line is a plot of the equation $\epsilon_{rel} = -2(C_{12}/C_{11})\epsilon_{app}$ given by elasticity theory. The calculation was performed in supercells of 64 atoms.

lattice constant. In the present work the unstrained lattice constant of bulk silicon a_0, is calculated to be 5.39 Å which can be compared with an experimental value of 5.43 Å.

To represent the biaxial strain resulting from epitaxial growth on a lattice-mismatched material, a strain ϵ_{app} is imposed upon the [100] and [010] lattice constants, $a_{[100]}$ and $a_{[010]}$, of a supercell. The [001] lattice constant $a_{[001]}$ is then allowed to relax so as to minimise the total energy of the bulk supercell. The impurity atom is then placed in this relaxed supercell and the atomic positions are relaxed for fixed lattice vectors. Re-relaxing the [001] cell parameter has negligible effect on the [001] cell parameter ($< 0.1\%$) and total-energy of the system (< 0.5 meV). Figure 7.1 shows the calculated relaxed [001] lattice parameter as a function of biaxial strain $\epsilon_{app} = \epsilon_{[100]} = \epsilon_{[010]}$. The strain along [001] is $\epsilon_{rel} = \epsilon_{[001]} = (a_{[001]} - a_0)/a_0$ and its variation is in excellent agreement with the expression derived from elasticity theory $\epsilon_{rel} = -2(C_{12}/C_{11})\epsilon_{app}$ shown by the line in figure 7.1, where C_{ij} are components of the elastic stiffness tensor and C_{12} and C_{11} are the experimental elastic constants 6.39×10^{10} and 16.57×10^{10} N m^{-2} respectively (Hirth and Lothe, 1968).

The Gibbs free energy of a system is $G = E - TS$ where E is the enthalpy, T is the temperature and S is the entropy[1]. The solubility limit of a substitutional impurity X_s is defined as the concentration of X for which the Gibbs free energy of a system containing dissolved impurities is equal to the Gibbs free energy of a system where the impurity atoms have formed precipitates. This concentration is found by minimising the Gibbs free energy with respect to the number of substitutional impurities. It can then be shown that the solubility limit of X_s is given by,

$$[X_s] = A\exp\left(\frac{-E_f}{kT}\right) \tag{7.1}$$

where k is Boltzmann's constant and E_f is the formation energy, or enthalpy of formation, given by equation 1.76 discussed previously. The chemical potential of both boron and arsenic, μ_B and μ_{As} are related to the energy of a dopant atom in its precipitate and are assumed to be independent of strain since it is likely that the structure and hence energy of a dopant atom within the precipitate would not depend upon the precise lattice constant of the bulk material. On the other hand, the chemical potential of silicon, μ_{Si} varies with strain and is taken to be the total-energy per atom of an n atom supercell of biaxially strained pure silicon relaxed in the [001] direction ($\mu_{Si}(\epsilon_{app}) = E_T^{Si}(\epsilon_{app})/n$). The pre-exponential factor A contains terms which arise from the entropy of the system. The entropy depends upon the disorder arising from the number of possible sites available for the substitutional dopant (the entropy of mixing) and the additional disorder introduced by atomic vibrations (vibrational entropy). The entropy of mixing is taken to be constant since the change in density of substitutional sites is negligible for the small strains considered here. However, the change in vibrational entropy may not be negligible at finite temperature since the vibrational modes of X_s may be strongly affected by strain. Here it has been assumed that A is not a function of strain and therefore the results presented in this chapter are only strictly valid at $T = 0$ K when entropy does not contribute to the Gibbs free energy.

[1]The term $+PV$ where P is pressure and V is volume will be neglected here since its contribution to the change in entropy is small for solid state reactions.

The Fermi-energy at high temperature is given by (Sze, 1981*b*),

$$E_F(\epsilon_{app}) = E_{mid}(\epsilon_{app}) \pm kT \left[\ln\left|\frac{[X_s]}{n_i}\right| + \ln\left|\frac{1}{2}\left(1+\sqrt{1+\left(\frac{2n_i}{[X_s]}\right)^2}\right)\right|\right] \tag{7.2}$$

where E_{mid} is the energy mid-way between the valence-band maximum and conduction-band minimum in bulk silicon subject to biaxial strain ϵ_{app}. The second term in equation 7.2 is positive for donors and negative for acceptors. The value of E_{mid} is obtained from a Kohn-Sham band-structure which gives a gap of 0.50 eV for unstrained material. The effect of quasi-particle corrections on the change in band-gap with strain has been considered by Sadigh *et al.* (2002) where it is shown to be negligible. This implies that the infamous underestimate of the band-gap within local density functional theory is effectively independent of strain. Hence DFT is able to accurately reproduce the change in band-gap with strain despite the large error in its absolute magnitude. The density of intrinsic carriers, $n_i(\epsilon_{app})$ in equation 7.2, is given by (Sze, 1981*b*),

$$n_i(T, \epsilon_{app}) = \sqrt{N_v N_c}\exp\left(-\frac{E_{gap}(\epsilon_{app})}{2kT}\right) \tag{7.3}$$

where N_v and N_c are the effective densities of states in the valence and conduction bands respectively and E_{gap} is the size of the silicon (Kohn-Sham) band-gap for the given strain. It is stressed that it is the change in band-gap with strain and not the absolute value of E_{gap} that is important to this problem as will be explained below. N_v and N_c have been taken to be constant with strain and temperature. The variation of the product of N_v and N_c with strain and temperature has been considered by Sadigh *et al.* (2002) and is found to make no significant differences to the results.

Of interest here is the *enhancement* to the equilibrium solubility limit with strain and for this the chemical potential of the dopant as well as N_v and N_c are not required, nor are the absolute values of E_{mid} or E_{gap}. The enhancement to the equilibrium solubility limit is defined as $([X_s](\epsilon_{app}) - [X_s](0))\,/[X_s](0)$. This quantity can be more accurately calculated than the absolute value since many terms which are not easily calculated are cancelled.

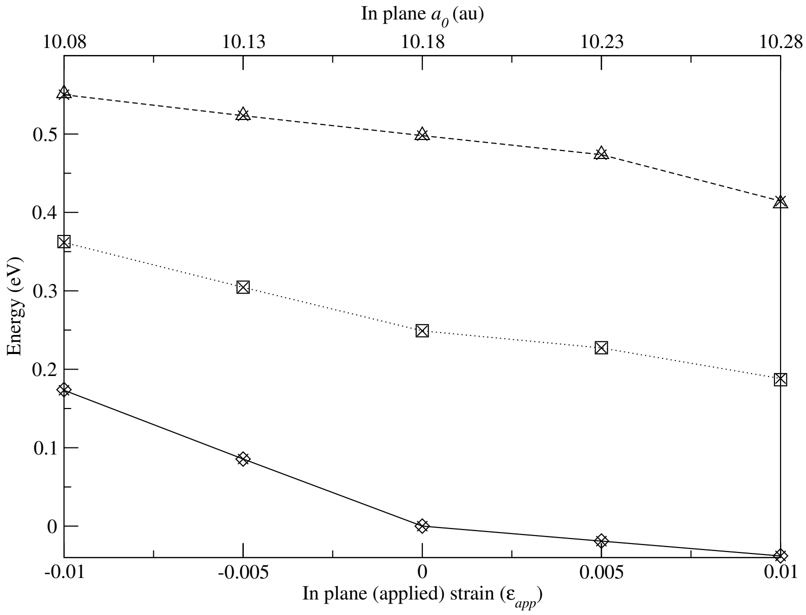

Figure 7.2: The dependence upon biaxial strain of valence band maximum, conduction band minimum and mid-gap energy. Crosses joined by a solid line show the valence-band maximum, crosses joined by a dashed line show the conduction-band minimum and crosses joined by a dotted line show the mid-gap energy level (E_{mid}) calculated using a 64 atom supercell. Using a 216 atom supercell the valence-band maximum, conduction-band minimum and E_{mid} have values shown by empty diamonds, triangles and squares respectively. Note the independence of these energies with supercell size.

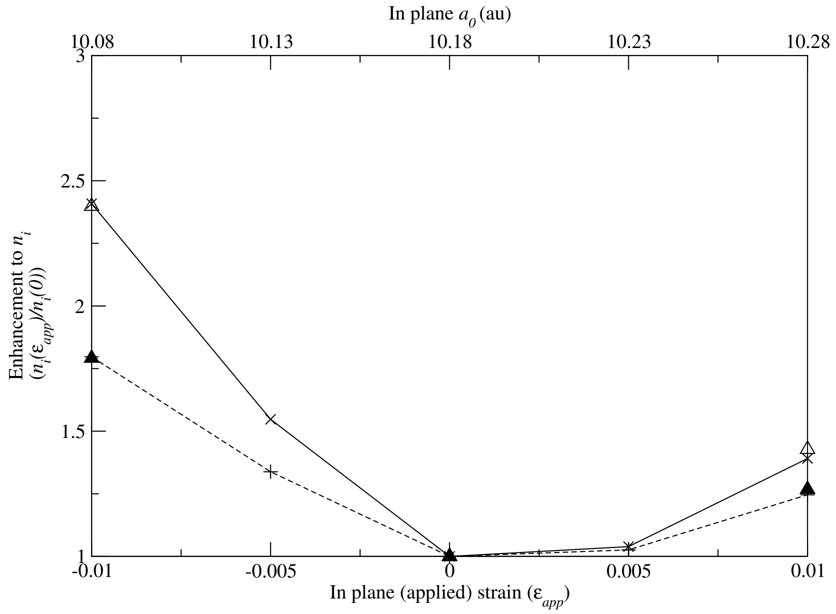

Figure 7.3: $n_i(\epsilon_{app})/n_i(0)$ calculated using 64 and 216 atom supercells for two different temperatures. The crosses joined by a solid line and the pluses joined by a dashed line are values calculated using 64 atom supercells for T = 800 K and T = 1200 K respectively. Empty and filled triangles are the values calculated in 216 atom supercells for T = 800 K and T = 1200 K respectively demonstrating the convergence with respect to cell size.

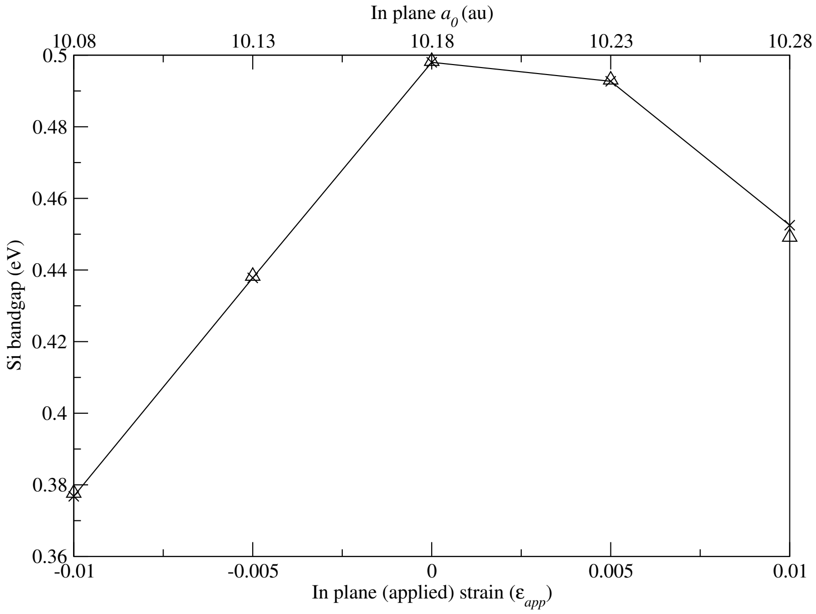

Figure 7.4: The variation in size of the Kohn-Sham band-gap of silicon with biaxial strain. The solid line with crosses shows values calculated using 64 atom supercells while triangles show the values calculated using 216 atom supercells.

7.3 The variation in Fermi-energy with biaxial strain

$E_{mid}(\epsilon_{app})$ is plotted in figure 7.2 and shows an almost linear decreases with biaxial strain ϵ_{app}. Since the Fermi-energy is approximately equal to E_{mid} (equation 7.2) this result implies that if all other terms are constant, a positively charged donor's solubility will be enhanced by tensile strain while a negatively charged acceptor's solubility will be enhanced by compressive strain. The enhancement to the equilibrium solubility is also proportional to $\sqrt{n_i(\epsilon_{app})/n_i(0)}$. The dependence of $n_i(\epsilon_{app})/n_i(0)$ upon strain is shown in figure 7.3 for two different temperatures. The number of intrinsic carriers is increased for both negative and positive strains due to the narrowing of the band-gap with strain of either sense (figure 7.4). Thus a study of biaxially strained, pure silicon alone predicts that the change in Fermi-energy with strain will enhance the solubility of positively charged defects under tensile strain and negatively charged defects under compressive strain. The effect of biaxial strain upon the total energy of substitutional boron and arsenic is considered next.

ϵ_{app}	bulk	B	As
-0.01	2.324	2.050	2.387
-0.005	2.329	2.055	2.391
0	2.333	2.059	2.396
0.005	2.338	2.064	2.401
0.01	2.343	2.069	2.406

Table 7.1: The impurity's bond lengths dependence upon biaxial strain. For comparison the bond lengths in equivalently strained bulk silicon are also shown. The unit of length is the Angström. Note that all four bond lengths are equal and the strain is accommodated by a change in bond angle.

7.4 Substitutional boron

A single substitutional boron atom in a 64 atom silicon supercell results in an equilibrium (unstrained) lattice constant $a_0(B)$ of 5.36 Å. The four silicon-boron bonds have lengths of 2.047 Å. The variation of lattice parameter is expected to be linear with boron concentration to first order. This can be written,

$$(a_0(B) - a_0)/a_0 = \beta[\mathrm{B}] \tag{7.4}$$

where for boron β, which describes the expansion to the lattice caused by the impurity, is calculated to be -5.99×10^{-24} cm^3 in good agreement with the value -5.2×10^{-24} cm^3 found experimentally (Cardona and Christensen, 1987*a*,*b*).

For the case of biaxial strain, the calculated differences between the four silicon-boron bond lengths are found to be negligible with the strain accommodated by an adjustment in bond angles. The silicon-boron bond length is shown in table 7.1 for different strains along with the equivalent distances in bulk material for comparison.

The variation of the formation energy of substitutional boron with strain has two main components: the variation in total-energy and the variation in Fermi-energy. In section 7.3 the Fermi-energy was shown to increase with compressive strain. Hence for negatively

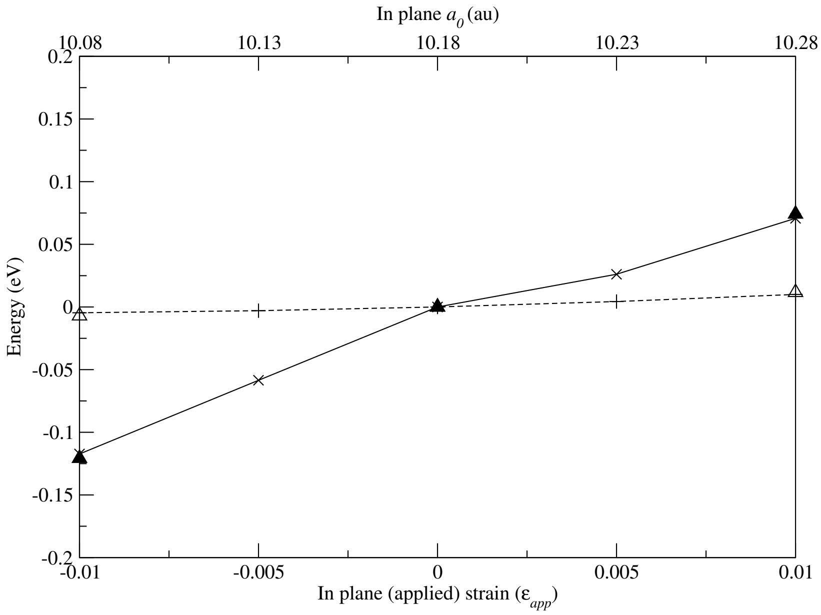

Figure 7.5: The variation of formation energy and total energy of substitutional boron with biaxial strain. Pluses joined by a dashed line and crosses joined by a solid line show values calculated using 64 atom supercells for the total energy and formation energy respectively. Empty triangles and filled triangles show the values of the total energy and formation energy respectively calculated with 216 atom supercells to demonstrate convergence.

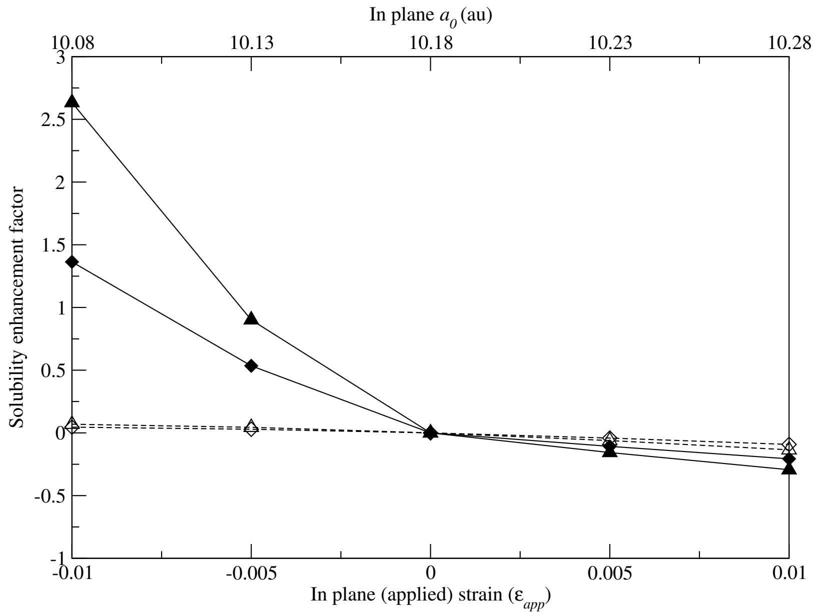

Figure 7.6: The enhancement to the equilibrium solubility of boron with biaxial strain for two different temperatures (triangles correspond to 800 K and diamonds correspond to 1200 K). To illustrate the influence of the change in Fermi-energy the enhancement is shown for a Fermi-energy set to be constant with strain (dashed lines and empty symbols) and for the value of E_F calculated as a function of strain (solid lines and filled symbols).

charged boron the effect of compressive strain on the Fermi-energy reduces the formation energy of the substitutional acceptor. Figure 7.5 shows the variation in total energy and formation energy as a function of strain, both of which are set to zero for unstrained material. This figure illustrates two points. The solubility of boron would be enhanced by compressive strain due to the change in total-energy with strain alone. Including contributions to the change in formation energy from the change in Fermi-energy with strain greatly increases this trend. This is seen in figure 7.6 which shows the enhancement to the equilibrium solubility limit for two different temperatures for the case where the Fermi-energy is fixed (so the variation in stability comes entirely from the change in total-energy) and the case where the Fermi-energy varies with strain. This clearly illustrates that the change in Fermi-energy with strain increases the predicted equilibrium solubility limit by an order of magnitude.

7.5 Substitutional arsenic

A single neutral arsenic atom in a 216 atom unstrained silicon supercell results in a relaxed lattice parameter $a_0(\mathrm{As})$ of 5.388 Å. This gives a β value of -0.06×10^{-24} cm^3 in good agreement with the small negative value -0.1×10^{-24} cm^3 found experimentally (Cardona and Christensen, 1987*a*; Cargill, III *et al.*, 1988). Using a smaller 64 atom supercell it is found that the lattice constant is identical to that of bulk material (*i.e.* $\beta = 0$). Details of the arsenic-silicon bond lengths have been studied by Koteski *et al.* (2003) using Extended X-ray Absorption Fine Structure spectroscopy (EXAFS). It is reported that, relative to the equivalent distance in pure silicon, the distance between the arsenic atom and its first nearest neighbour is $\sim 3\%$ larger, decreasing to 0.78% larger and 0.62% larger for the second and third nearest neighbours and presumably becoming negative at larger distances to result in the measured overall reduction in lattice constant. Our calculations agree well with this EXAFS data as shown in table 7.2 and we predict that the distance between arsenic and its $\sim 8^{\mathrm{th}}$ nearest-neighbour is indeed shorter than the distance between equivalent crystal sites in bulk silicon.

	EXAFS			DFT		
NN	bulk	Si:As	% increase	bulk	Si:As	% increase
1	2.351	2.43	3.36	2.333	2.397	2.71
2	3.840	3.87	0.78	3.810	3.822	0.32
3	4.502	4.53	0.62	4.468	4.479	0.25
4				5.389	5.397	0.15
5				5.872	5.875	0.05
6				6.600	6.602	0.04
7				7.000	7.003	0.05
8				7.620	7.619	-0.02

Table 7.2: A comparison of the change in first, second and third nearest neighbour (NN) distances between Si and Si:As. EXAFS (Koteski *et al.*, 2003) and the present calculations agree that there is a dilation around the As atoms which drops off sharply and in fact leads to a *smaller* lattice constant than bulk Si. The theoretical values were calculated using volume relaxed, 216 atom supercells. Lengths are given in Angströms.

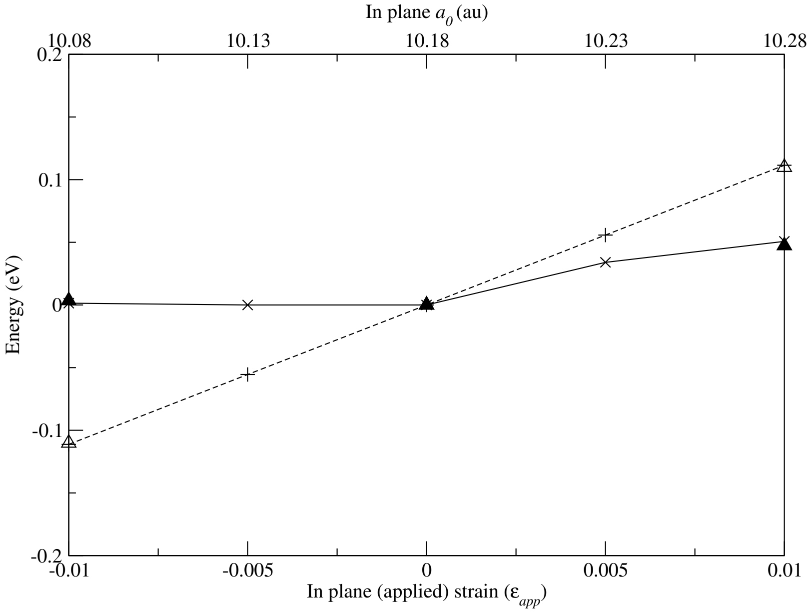

Figure 7.7: The variation of formation energy and total energy of substitutional arsenic with biaxial strain. Pluses joined by a dashed line and crosses joined by a solid line show values calculated using 64 atom supercells for the total energy and formation energy respectively. Empty triangles and filled triangles show the values of the total energy and formation energy respectively calculated with 216 atom supercells to demonstrate convergence.

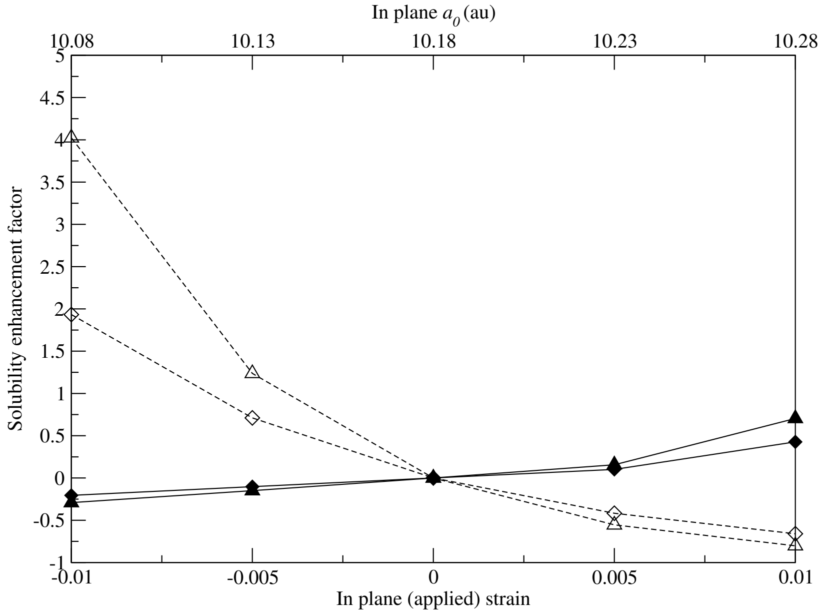

Figure 7.8: The enhancement to the equilibrium solubility of arsenic with strain for two different temperatures (triangles correspond to 800 K and diamonds correspond to 1200 K). To illustrate the influence of the change in Fermi-energy the enhancement is shown for a Fermi-energy set to be constant with strain (dashed lines and empty symbols) and for the value of E_F calculated as a function of strain (solid lines and filled symbols).

The arsenic-silicon bond-lengths in biaxially strained material are given in table 7.1. The calculated change in formation energy due to the change in total-energy and the combination of the change in total-energy and change in Fermi-energy with strain is shown for arsenic in figure 7.7. The enhancement to the equilibrium solubility limit is shown in figure 7.8. The effect of the Fermi-energy is far more critical for arsenic. When the strain dependence of the Fermi-energy is neglected the solubility of arsenic increases for tensile strains, despite the fact that the arsenic atom has a larger atomic radius than a silicon atom. This is due to the fact that arsenic actually leads to a reduction in the overall silicon lattice constant as discussed above and hence compressive strain leads to a reduction in the total-energy of the substitutional donor. It is the change in Fermi-energy with strain that results in an increase in solubility of arsenic with tensile strain.

7.6 Summary

This work shows that the Fermi-energy is decreased by tensile biaxial strain such as that imposed on a silicon epilayer grown on SiGe. This has the effect of reducing the formation energy of positively charged impurities and increasing the formation energy of negatively charged ones. Hence the stability and therefore solubility of ionised donors will be increased by tensile strain while the opposite is true for ionised acceptors. The total energy of a substitutional impurity is also dependent upon biaxial strain. Arsenic highlights that fact that the atomic size is not a good parameter in the determination of the variation in total energy with biaxial strain. Instead it is necessary to consider whether, in unstrained material, doping with that particular element will increase or decrease the lattice constant resulting in a positive and negative value of β respectively. Species with a positive value of β will have a lower total energy and hence greater stability in material subject to a tensile biaxial strain while the opposite is true of impurities with a negative β value.

In agreement with Sadigh *et al.* (2002) the solubility of boron is greatly enhanced by compressive strain and is reduced by tensile strain. This enhancement is due to both the negative β value and the fact that the dopant is negatively charged. The enhancement to the solubility of boron is largely due to the Fermi-energy effect. The solubility of arsenic is increased by tensile biaxial strains and reduced by compressive strains and hence the solubility of arsenic is expected to be enhanced in silicon grown on SiGe. The small negative value of β for arsenic tends to increase its solubility in compressively strained material but this is more that compensated by a strong increase in solubility due to the Fermi-energy effect for tensile strains. The net result is a small but significant enhancement to the equilibrium solubility limit of arsenic under positive biaxial strain. For both arsenic and boron the contribution to the solubility enhancement from the change in Fermi-energy with strain is by far the most important. Once again it is stressed that vibrational entropy, which may play an important role in determining the solubility of boron and arsenic as a function of strain has been neglected and therefore these conclusions are only strictly valid for a temperature of 0 K.

Chapter 8

Concluding remarks

The present theoretical modelling of the interaction of substitutional boron with self-interstitials and the close comparison of these calculations to experiment demonstrates the surprising structure of the boron interstitial defect. Rather than the boron atom being located at an interstitial site, it is found that it remains more or less substitutional next to a self-interstitial. Boron located at an interstitial site (what is normally meant by interstitial boron) is a metastable form of this defect. The boron interstitial defect is denoted BI since it is formed from a single boron atom with one interstitial atom (of either species). BI has been demonstrated to be a negative-U defect and as is typical of centres possessing negative-U, its structure depends upon its charge state. BI is formed extremely efficiently by the trapping of self-interstitials at substitutional boron but it becomes mobile at around 230 K and is trapped at other impurities or by other boron interstitial defects to form boron interstitial clusters.

A comparison of the local modes observed following the disappearance of the LVMs assigned to BI has allowed a far greater understanding of the clustering process undertaken by boron to be developed. A stable complex that is formed by room temperature when BI is trapped by substitutional boron is a boron-boron $\langle 100 \rangle$ split interstitial labelled B_2I^Q. This electrically inactive complex is probably removed by reaction with divacancies at around 220°C.

For lower boron concentrations the evolution of boron is modified by its more probable interaction with the common impurities oxygen and carbon. In such material, when BI becomes mobile it is trapped by interstitial oxygen giving rise to B_iO_i which has a level observed by DLTS at around $E_c - 0.23$ eV. Here it has been argued that this level is a single donor $(0/+)$ and not a double donor $(+/++)$ level as inferred from the capture cross-sections. The calculated binding energy of BI to interstitial oxygen suggests that the complex dissociates at around 150°C. At this temperature a level is observed by DLTS at $E_v + 0.29$ eV, which calculations presented here support, is due to the trapping of dissociated BI at C_s forming B_iC_s. This complex is stable until around 400°C.

Hydrogen is introduced during wafer processing steps such as etching and its interaction with interstitial boron has been considered. The reaction of hydrogen with boron and interstitial boron generates a complex B_iB_sH which is identical to B_2I^Q apart from the fact that a hydrogen atom is bound to one of the boron atoms. This complex can be formed in a number of ways, in fact by any method that introduces hydrogen, boron and self-interstitials to the material, highlighting the stability of the complex and hinting at the importance of B_2I^Q.

Within this thesis the problem of solar cell degradation has been addressed. The reduction in carrier lifetime caused by the light-induced dissociation of the iron-boron pair has been well understood for a relatively long time. On the other-hand the degradation due to the boron-oxygen complex is still rather mysterious. The model proposed by Schmidt *et al.* (Schmidt *et al.*, 2002; Bothe *et al.*, 2004; Schmidt and Bothe, 2004) has in part been confirmed by the calculations presented here where B_sO_{2i} has been shown to have many properties in common with the observed lifetime degrading defect. A mechanism for the formation of this defect via the slower of the two mechanisms observed experimentally has been proposed.

Biaxially strained silicon has been shown to have some rather interesting properties with regards to the stability of point defects. The Fermi-level in such material, at high temperature, is located approximately mid-way between the valence-band top and conduction-band bottom. This energy level was shown to increase with compressive strain and decrease with tensile strain. This results in the increased stability and hence increased solubility limit of acceptors in compressively strained material and the increased stability and hence increased solubility limit of donors in material under tensile strain.

Despite the number of experimental observations that have been successfully explained through the modelling described here, a number of questions remain unanswered and a number of further questions have been raised providing plenty of scope for future work. As mentioned in chapter 5, B_iC_s is expected to be formed by room temperature irradiation of FZ, boron-doped silicon since BI will be trapped directly by carbon rather than oxygen as is the case in Cz material. Instead B_iC_s does not form until the material has been annealed at 120°C. It was suggested in chapter 5 that this trend may be an indication of a metastable complex of boron and carbon that is not converted into B_iC_s until 120°C. Preliminary calculation suggest that there may be such a barrier in the formation of B_iC_s but the formation mechanism of B_iC_s demands further study if this trend is to be understood.

Many questions remain unanswered when it comes to the complex responsible for the degradation of Cz solar cells. Although there is strong evidence that B_sO_{2i} is the defect responsible for the degradation, the formation process must be considered in greater detail. Of the fast and slow formation processes relatively little is known about the faster mechanism although this is currently being addressed by experiment. The recombination enhanced diffusion process must be carefully considered as an alternative explanation of the slow formation process and as a candidate for the fast formation process. If recombination enhanced diffusion is unable to explain the fast process then perhaps a recombination enhanced reconfiguration of B_sO_{2i} from an inactive to active form can. Again this must be considered in detail and in the light of forthcoming experimental results. Using gallium or aluminium as the *p*-type dopant has been suggested as a possible route to the avoidance of lifetime degrading defects in Cz silicon solar cells. Calculations imply that the interaction of gallium and aluminium with the oxygen pair will not result in a lifetime degrading complex but for the case of gallium these calculations are preliminary and need to be performed in greater detail. There is a possibility for theory to play a role in the discovery of other methods that allow the formation of B_sO_{2i} to be avoided.

Having developed an understanding of the role of Fermi-energy and total-energy in the stability of defects in biaxially strained material, the contribution of vibrational entropy to their stability must now be studied. The study could then be extended to include the diffusion process of boron in such material. Understanding the effect of strain upon the diffusion of boron may lead to methods of suppressing the diffusion of boron in such material. Given the current trend for the use of silicon layers biaxially strained by growth on SiGe in high performance devices, such knowledge would be of great importance and benefit.

Of all unanswered questions, one of the most relevant relates to the BICs that have so far evaded detection by both macroscopic measurements (*e.g.* TEM) and optical and electronic measurements (*e.g.* PL, EPR). Very little is known about the BICs that are stable at the temperatures as high as those during TED. Very recently however, very large BICs have been observed in the form of boron-interstitial platelets. These platelets are so large that they can be observed via TEM (Cristiano *et al.*, 2003). The high stability of B_2I^Q is echoed by its formation in hydrogenated material and even by the high stability of the structurally similar complex B_iC_s and here it is suggested that an aggregation of B_2I^Q may be responsible for these very large BICs. This is clearly a matter to be investigated. Once these large BICs are understood it will also perhaps be possible to extrapolate from B_2I^Q to the platelet like BICs providing for the first time an experimentally grounded model for intermediately sized clusters.

Throughout this thesis the value of theory and experiment working hand-in-hand and guiding each other has been demonstrated. The true potential of each is only achieved with the aid of the other. This point I leave as my final conclusion.

Bibliography

Agarwal A, Gossmann H J, Eaglesham D J, Pelaz L, Jacobson D C, Haynes T E, and Erokhin Y E, Appl. Phys. Lett. **71**, 3141 (1997).

Anderson P W, Phys. Rev. Lett. **35**, 953 (1975).

Angress J F, Goodwin A R, and Smith S D, Proc. R. Soc. London, Ser. A **287**, 64 (1965).

Bachelet G B, Hamann D R, and Schlüter M, Phys. Rev. B **26**, 4199 (1982).

Baldereschi A, Phys. Rev. B **7**, 5212 (1973).

Balkanski M and Nazarewicz W, J. Phys. Chem. Solids **27**, 671 (1966).

von Barth U and Hedin L, J. Phys. C **5**, 1629 (1972).

Bean A R, Morrison S R, Newman R C, and Smith R S, J. Phys. C **5**, 379 (1972).

Benninghoven A, Rudenauer F G, and Werner H W, *Secondary ion mass spectrometry* (John Wiley, 1987).

Born M and Huang K, *Dynamical theory of crystal lattices* (Oxford University Press, Oxford, 1954).

Born M and Oppenheimer R, Ann. Phys. **87**, 457 (1925).

Bothe K Private communication (2004).

Bothe K, Hezel R, and Schmidt J, Solid State Phenomena **95-96**, 223 (2004).

Bothe K, Schmidt J, and Hezel R, in *3rd World Conference on Photovoltaic Energy Conversion* (2003), in press.

Brotherton S D, Bradley P, and Gill A, J. Appl. Phys. **57**, 1941 (1985).

Brust D, *Methods in Computational Physics*, volume 8, p. 33 (Academic Press, 1968).

Cardona M and Christensen N E, Phys. Rev. B **35**, 6182 (1987*a*).

Cardona M and Christensen N E, Phys. Rev. B **36**, 2906 (1987*b*).

Cargill, III G S, Angilello J, and Kavanagh K L, Phys. Rev. Lett. **61**, 1748 (1988).

Ceperley D, Phys. Rev. B **18**, 3126 (1978).

Ceperley D M and Alder B J, Phys. Rev. Lett. **45**, 566 (1980).

Chadi D J, Phys. Rev. Lett. **77**, 861 (1996).

Chadi D J and Cohen M L, Phys. Rev. Lett. **7**, 692 (1973).

Clegg J B, in *Growth and characterisation of semiconductors*, (edited by R A Stradling and P C Klipstein), chapter 5 (Adam Hilger, Bristol and New York, 1990).

Coutinho J, Jones R, Briddon P R, and Öberg S, Phys. Rev. B **62**, 10824 (2000).

Coutinho J, Jones R, Briddon P R, Öberg S, Murin L I, Markevich V P, and Lindstörm J L, Phys. Rev. B **65**, 014109 (2002).

Coutinho J, Jones R, Murin L I, Markevich V P, Lindström J L, Öberg S, and Briddon P R, Phys. Rev. Lett. **87**, 235501 (2001).

Cowern N E B, Janssen K T F, and Jos H, J. Appl. Phys. **68**, 6191 (1990).

Cowern N E B, Zalm P C, van der Sluis P, Gravesteijn D J, and de Boer W B, Phys. Rev. Lett. **72**, 2585 (1994).

Cristiano F, Hebras X, Cherkashin N, Claverie A, Lerch W, and Paul S, Appl. Phys. Lett. **83**, 5407 (2003).

Davies G, in *Identification of defects in semiconductors*, (edited by M Stavola), volume 51B of *Semiconductors and Semimetals*, chapter 1 (Academic Press, San Diego, 1999).

Dawber P G and Elliot R J, Proc. R. Soc. London, Ser. A **273**, 222 (1963*a*).

Dawber P G and Elliot R J, Proc. R. Soc. **81**, 453 (1963*b*).

Dobaczewski L, Kaczor P, Hawkins I D, and Peaker A R, J. Appl. Phys. **76**, 194 (1994).

Downey D F, Chow J W, Ishida E, and Jones K S, Appl. Phys. Lett. **73**, 1263 (1998).

Drevinsky P J, Caefer C E, Tobin S P, Mikkelsen J C J, and Kimerling L C, MRS Res. Soc. Symp. Proc. **104**, 167 (1988).

Eberlein T A G, Pinho N, Jones R, Coomer B J, Goss J P, Briddon P R, and Öberg S, Physica B **308–310**, 454 (2001).

Elliot R J and Pfeuty P, J. Phys. Chem. Solids **28**, 1789 (1967).

Estreicher S K, Gharaibeh M, Fedders P A, and Ordejón P, Phys. Rev. Lett. **86**, 1247 (2001).

Ewald P P, Ann. Phys. **64**, 253 (1921).

Feichtinger H, Inst. Phys. Conf. Ser. **46**, 528 (1979).

Feichtinger H, Waltl J, and Gschwandtner A, Solid State Commun. **27**, 867 (1978).

Feklisova O, Yarykin N, Yakimov E B, and Weber J, Physica B **308–310**, 210 (2001).

Fischer H and Pschunder W, in *Proc. 10th IEEE Photovolt. Spec. Conf.*, p. 404 (IEEE, New York, 1973).

Flynn C P, *Point defects and diffusion* (Oxford University Press, Glasgow, 1972).

Fock V, Z. Phys. **61**, 126 (1930).

Glunz S W, Rein S, Warta W, Knobloch J, and Wettling W, in *2nd World Conference on Photovoltaic Energy Conversion* (European Commision, Ispra, 1998), p. 1343.

Hakala M, Puska M J, and Nieminen R M, Phys. Rev. B **61**, 8155 (2000).

Hall R N, Phys. Rev. **87**, 387 (1952).

Harris R D, Newton J L, and Watkins G D, Phys. Rev. B **36**, 1094 (1987).

Harrison W A, *Pseudopotentials in the theory of metals* (W. A. Benjamin, New York, 1966).

Hartke J L, J. Appl. Phys. **39**, 4871 (1968).

Hartwigsen C, Goedecker S, and Hutter J, Phys. Rev. B **58**, 3641 (1998).

Hashigami H, Dhamrin M, and Saitoh T, Japan J. Appl. Phys. **42**, 2564 (2003).

Heine V, Solid State Phenomena **24**, 1 (1970).

Henry C H, J. Appl. Phys. **51**, 4494 (1980).

Hill T L, *An introduction to statistical thermodynamics* (Dover, New York, 1986).

Hirth J P and Lothe J, *Theory of dislocations*, Materials Science and Engineering (McGraw-Hill, New York, 1968).

Hohenberg P and Kohn W, Phys. Rev. **136**, B864 (1964).

Hornbeck J A and Haynes J R, Phys. Rev. **97**, 311 (1955).

Jain S C, Schoenmaker W, Lindsay R, Stolk P A, Decoutere S, Willander M, and Maes H E, J. Appl. Phys. **91**, 8919 (2002).

Jeong J W and Oshiyama A, Phys. Rev. B **64**, 235204 (2001).

Jones R and Briddon P R, in *Identification of Defects in Semiconductors*, (edited by M Stavola), volume 51A of *Semiconductors and Semimetals*, chapter 6, p. 287 (Academic Press, Boston, 1999).

Jones R, Eberlein T A G, Pinho N, Coomer B J, Goss J P, Briddon P R, and Öberg S, Nucl. Instrum. Methods B **186**, 10 (2002).

Jones R, Goss J, Ewels C, and Öberg S, Phys. Rev. B **50**, 8378 (1994).

Jones R O and Gunnarsson O, Rev. Mod. Phys. **61**, 689 (1989).

Kemble, Rev. Mod. Phys. **1**, 206 (1932).

Kimerling L C, Asom M T, Benton J L, Drevinsky P J, and Caefer C E, Mater. Sci. Forum **38-41**, 141 (1989).

Knobloch J, Glunz S W, Biro D, Warta W, Schäffer E, and Wettling W, in *Proc. 25th IEEE Photovolt. Spec. Conf.*, p. 405 (IEEE, New York, 1996).

Kohn W and Sham L J, Phys. Rev. **140**, A1133 (1965).

Koopmans T, Physica **1**, 104 (1934).

Koteski V, Ivanovic N, Haas H, Holub-Krappe E, and Mahnke H E, Nucl. Instrum. Methods B **200**, 60 (2003).

Laithwaite K, Newman R C, and Totterdell D H, J. Phys. C **8**, 236 (1975).

Lang D V, J. Appl. Phys. **45**, 3023 (1974).

Leary P, Jones R, and Öberg S, Phys. Rev. B **57**, 3887 (1998).

Lee Y J, von Boehm J, Pesola M, and Nieminen R M, Phys. Rev. Lett. **86**, 3060 (2001).

Leeuw S W D, Perram J W, and R. S E, Proc. R. Soc. London, Ser. A **373**, 27 (1980).

Lemke H, Phys. Status Solidi A **64**, 215 (1981).

Liu X, Windl W, and Masqueleir M, Appl. Phys. Lett. **77**, 2018 (2000).

Louie S G, Froyen S, and Cohen M L, Phys. Rev. B **26**, 1738 (1982).

Lundqvist S and March N H, editors, *Theory of the inhomogeneous electron gas*, Physics of solids and liquids (Plenum, New York, 1983).

Macdonald D and Cuevas A, Appl. Phys. Lett. **74**, 1710 (1999).

Macdonald D, Cuevas A, and Wong-Leung J, J. Appl. Phys. **89**, 7932 (2001).

MacDonald J K L, Phys. Rev. **43**, 830 (1933).

Mamor M, Auret F D, Goodman S A, Meyer W E, and Myburg G, Appl. Phys. Lett. **73**, 3178 (1998).

Mamor M, Willander M, Auret F D, Meyer W E, and Sveinbjörnsson, Phys. Rev. B **63**, 045201 (2001).

Mannino G, Cowern N E B, Roozeboom F, and van Berkum J G M, Appl. Phys. Lett. **76**, 855 (2000).

Mannino G, Stolk P A, Cowern N E B, de Boer W B, Dirks A G, Roozeboom F, Berkum J G M, Woerlee P H, and Toan N N, Appl. Phys. Lett. **78**, 889 (2001).

Mc Weeney R, *Methods of Molecular Quantum Mechanics* (Academic Press, New York, 1989).

Merle J C, Capizzi M, Fiorini P, and Frova A, Phys. Rev. B **17**, 4821 (1978).

Mikkelsen J C, MRS Res. Soc. Symp. Proc. **59**, 19 (1986).

Monakhov E V, Nylandsted Larsen A, and Kringhøj P, J. Appl. Phys. **81**, 1180 (1997).

Monkhorst H J and Pack J D, Phys. Rev. B **13**, 5188 (1976).

Mooney P M, in *Identification of defects in semiconductors*, (edited by M Stavola), volume 51B of *Semiconductors and Semimetals*, chapter 2 (Academic Press, San Diego, 1999).

Mooney P M, Cheng L J, Süli M, Gerson J D, and Corbett J W, Phys. Rev. B **15**, 3836 (1977).

Morin F J and Maita J P, Phys. Rev. **96**, 28 (1954).

Murin L I, Hallberg T, Markevich V P, and Lindström J L, Phys. Rev. Lett. **80**, 93 (1998).

Musgrave M J P and Pople J A, Proc. R. Soc. London, Ser. A **268**, 474 (1962).

Newman R C, *Infrared Studies of Crystal Defects* (Taylor and Francis, London, 1973).

Newman R C and Smith R S, Phys. Lett. A **24**, 671 (1967).

Pack J D and Monkhorst H J, Phys. Rev. B **16**, 1748 (1977).

Palmer D W Private communication (2004).

Parr R G and Yang W, *Density-functional theory of atoms and molecules* (Oxford University Press, Oxford, 1989).

Pelaz L, Gilmer G H, Gossmann H J, Rafferty C S, Jaraiz M, and Barbolla J, Appl. Phys. Lett. **74**, 3657 (1999).

Pelaz L, Jaraiz M, Gilmer G H, Gossman H J, Rafferty C S, Eaglesham D J, and Poate J M, Appl. Phys. Lett. **70**, 2285 (1997).

Perdew J P, in *Electronic Structure of Solids '91*, (edited by P Ziesche and H Eschrig) (Akademie Verlag, Berlin, 1991).

Perdew J P, Burke K, and Ernzerhof M, Phys. Rev. Lett. **77**, 3865 (1996*a*).

Perdew J P, Burke K, and Wang Y, Phys. Rev. B **54**, 16533 (1996*b*).

Perdew J P and Wang Y, Phys. Rev. B **45**, 13244 (1992).

Perdew J P and Zunger A, Phys. Rev. B **23**, 5048 (1981).

Pesola M, Lee Y J, von Boehm J, Kaukonen M, and Nieminen R M, Phys. Rev. Lett. **84**, 5343 (2000).

Pickett W E, Comp. Phys. Rep. **9**, 115 (1989).

Pople J A and Beverage D L, *Approximate molecular orbital theory* (McGraw-Hill, New York, 1970).

Rajendran K and Schoenmaker W, J. Appl. Phys. **89**, 980 (2001).

Reif F, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965).

Rein S and Glunz S W, Appl. Phys. Lett. **82**, 1054 (2003).

Rein S, Glunz S W, and Willeke G, in *3rd World Conference on Photovoltaic Energy Conversion*, p. 2899 (2003), in press.

Rein S, Rehrl T, Warta W, Glunz S W, and Willeke G, in *Proc. 17th European Photovolt. Solar Energy Conf.*, p. 1555 (WIP-ETA, Munich, 2001).

Reiss J H, King R R, and Mitchell K W, Appl. Phys. Lett. **68**, 3302 (1996).

Roothaan C C J, Rev. Mod. Phys. **23**, 69 (1951).

Roth E G, Holland O W, Venezia V C, and Nielsen B, J. Electr. Mat. **26**, 1349 (1997).

Rücker H, Heinemann B, Röpke W, Kurps R, Krüger D, Lippert G, and Osten H J, Appl. Phys. Lett. **73**, 1682 (1998).

Sadigh B, Lenosky T J, Caturla M J, Quong A A, Benedict L A, Diaz de la Rubia T, Giles M M, Foad M, Spataru C D, and Louie S G, Appl. Phys. Lett. **80**, 4738 (2002).

Sauer R and Weber J, Physica B & C **116**, 195 (1983).

Schmidt J, Solid State Phenomena **95-96**, 187 (2004).

Schmidt J, Aberle A G, and Hezel R, in *Proc. 26th IEEE Photovolt. Spec. Conf.*, p. 13 (IEEE, New York, 1997).

Schmidt J and Bothe K, Phys. Rev. B **69**, 024107 (2004).

Schmidt J, Bothe K, and Hezel R, in *Proc. 29th IEEE Photovolt. Spec. Conf.*, p. 178 (IEEE, New York, 2002).

Schmidt J and Cuevas A, J. Appl. Phys. **86**, 3175 (1999).

Shaklee K L and Nahory R E, Phys. Rev. Lett. **24**, 942 (1970).

Shockley W and Read Jr. W T, Phys. Rev. **87**, 835 (1952).

Sigmund P, Phys. Rev. B **184**, 383 (1969).

Sinton R A and Cuevas A, Appl. Phys. Lett. **69**, 2510 (1996).

Slater J C, Phys. Rev. **34**, 1293 (1929).

Slater J C, *Quantum Theory of Molecules and Solids*, volume 2 (McGraw Hill, New York, 1965).

Smith S D and Angress J F, Phys. Lett. **6**, 131 (1963).

Smulders P J M, Boerma D O, Bech Nielsen B, and Swanson M L, Nucl. Instrum. Methods B **45**, 438 (1990).

Song L W, Zhan X D, Benson B W, and Watkins G D, Phys. Rev. B **42**, 5765 (1990).

Stavola M, in *Identification of defects in semiconductors*, (edited by M Stavola), volume 51 of *Semiconductors and Semimetals*, chapter 3 (Academic Press, San Diego, 1999).

Stolk P A, Gossman H J, Eaglesham D J, Jacobson D C, Poate J M, and Luftman H S, Appl. Phys. Lett. **66**, 568 (1995*a*).

Stolk P A, Gossmann H J, Eaglesham D J, and Poate J M, Nucl. Instrum. Methods B **96**, 187 (1995*b*).

Stoneham A M, *Theory of Defects in Solids* (Oxford University Press, London, 1975).

Sze S M, *Physics of Semiconductor Devices* (Wiley-Interscience, New York, 1981*a*), 2nd edition.

Sze S M, *Physics of Semiconductor Devices*, pp. 16–27 (Wiley-Interscience, New York, 1981*b*), 2nd edition.

Tarnow E, J. Phys. Cond. Matter **4**, 5405 (1992).

Terashima K, Ikarashi T, Watanabe M, and Kitano T, Mater. Sci. Forum **258–263**, 587 (1997).

Thijssen J M, *Computational Physics* (Cambridge University Press, Cambridge, 1999).

Thonke K, Bürger N, Watkins G D, and Sauer R, in *Proceedings of the 13th International Conference on Defects in Semiconductors*, (edited by L C Kimerling and J M Parsey Jr.), pp. 823–831, Metallurgical Society of AIME. Electronic Materials Committee (American Institute of Mining, Metallurgical, and Petroleum Engineers, Inc., New York, 1984).

Thonke K, Weber J, Wagner J, and Sauer R, Physica B & C **116**, 252 (1983).

Tinkham M, *Group theory and quantum mechanics* (McGraw-Hill, New York, 1964).

Tipping A and Newman R C, Semicond. Sci. Technol. **2**, 389 (1987).

Troullier N and Martins J L, Phys. Rev. B **43**, 1993 (1991).

Troxell J R and Watkins G D, Phys. Rev. B **22**, 921 (1980).

Volpi F, Peaker A R, Berbezier I, and Ronda A, J. Appl. Phys. **95**, 4752 (2004).

Vosko S J, Wilk L, and Nusair M, Chinese J. Phys. **58**, 1200 (1980).

Van de Walle C G, Bar-Yam Y, and Pantelides S T, Phys. Rev. Lett. **60**, 2761 (1988).

Watkins G D, Phys. Rev. B **12**, 5824 (1975).

Watkins G D, Phys. Rev. B **13**, 2511 (1976).

Watkins G D, in *Identification of defects in semiconductors*, (edited by M Stavola), volume 51A of *Semiconductors and Semimetals*, chapter 1 (Academic Press, San Diego, 1998).

Watkins G D and Brower K L, Phys. Rev. Lett. **36**, 1329 (1976).

Watkins G D and Troxell J R, Phys. Rev. Lett. **44**, 593 (1980).

Weber J, Knack S, Feklisova O V, Yarykin N A, and Yakimov E B, Microelectronic Eng. **66**, 320 (2003).

Weeks J D, Tully J C, and Kimerling L C, Phys. Rev. B **12**, 3286 (1975).

Williams P and Evans C A, Surface Sci. **78**, 324 (1978).

Windl W, Bunea M M, Stumpf R, Dunham S T, and Masquelier M P, Phys. Rev. Lett. **83**, 4345 (1999).

Yamaguchi M, Khan A, Taylor S J, Ando K, Yamaguchi T, Matsuda S, and Aburaya T, J. Appl. Phys. **86**, 217 (1999).

Yarykin N Private communication (2004).

Yarykin N, Feklisova O, and Weber J, Physica B **308–310**, 159 (2001).

Yarykin N, Feklisova O V, and Weber J, Phys. Rev. B **69**, 045201 (2004).

Zangenberg N R, Fage-Pedersen J, Lundsgaard Hansen J, and Nylandsted Larsen A, J. Appl. Phys. **94**, 3883 (2003).

Zhao J, Wang A, Green M A, and Ferrazza F, Appl. Phys. Lett. **73**, 1991 (1998).

Zhu J, Diaz de la Rubia T, Yang L H, Mailhiot C, and Gilmer G H, Phys. Rev. B **54**, 4741 (1996).