Quantum mechanical modelling of defects in semiconductors

2004 J. Phys.: Condens. Matter 16 S2643

(http://iopscience.iop.org/0953-8984/16/27/004)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **16** (2004) S2643–S2657 PII: S0953-8984(04)70024-7

Quantum mechanical modelling of defects in semiconductors

R Jones[1], **B J Coomer**[1] **and P R Briddon**[2]

[1] School of Physics, University of Exeter, Exeter EX4 4QL, UK
[2] School of Natural Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

Received 6 October 2003
Published 25 June 2004
Online at stacks.iop.org/JPhysCM/16/S2643
doi:10.1088/0953-8984/16/27/004

Abstract

The theoretical basis of density functional pseudopotential methods for determining the properties of defects in semiconductors is given. The formalism is applied to a single illustrative defect, the carbon interstitial in silicon, which has been identified by electron paramagnetic resonance, vibrational mode spectroscopy, photoluminescence and deep-level transient spectroscopy. The theory is shown to largely account for the properties of the defect.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Bill Norgett was associated with the HADES code used to determine the properties of defects in ionic systems. In these materials, the main interaction between a defect and the host is determined by the crystal field. The electrostatic interactions between the constituents of the defect and its surroundings are then crucial and the success of the code is due to the accuracy by which these are modelled. This approach will not work for covalent or metallic systems where the kinetic energies of the electrons and their modification by the surrounding atoms are a crucial ingredient in bonding. Methods based on local density functional theory (DFT) for determining the structure and electronic properties of covalent and metallic bonded materials have increased in importance in the 20 years or so since the pioneering studies of Yin and Cohen [1]. Most of these calculations are carried out in a supercell or a cluster containing a defect. In the former, periodic boundary conditions are imposed across the boundaries of the supercell which typically contains of the order of 100 atoms. In the latter, the surface has to be passivated—usually by hydrogen—to prevent overlap of the wavefunctions of surface dangling bonds with those of the defect at the centre of the cluster. The wavefunctions corresponding to either crystalline or defect related states are expressed in a basis of either plane waves or localized orbitals. Of course, these approaches are ideal for the treatment of a high concentration of defects, of the order of a few atomic per cent as in an ordered alloy,

but the interactions between defects in different cells, when the supercell method is used, may limit the usefulness of the approach when one is interested in isolated *point defects*.

There are a large number of computer codes based on the density functional formalism available, e.g. FINGER, AIMPRO, CASTEPS, SIESTA and VASP. It is not the intention of this review to detail the differences in these codes but to explain the background theory and illustrate the type of information which can be extracted, using as an example AIMPRO. This is an acronym for *ab initio modelling program* and is used by groups based at Exeter, Newcastle, Sussex, Luleå and Aveiro.

We shall show that the modelling is able to give a good account of the properties of a single defect—the carbon interstitial (C_i) in silicon. This defect is particularly interesting because it has been characterized by the four main experimental techniques applied to point defects: electron paramagnetic resonance (EPR), deep-level transient spectroscopy (DLTS), infra-red absorption (IR) and photoluminescence (PL). It has also been well investigated by theory and the theoretical results can then be compared with experiment.

2. Quantum mechanical modelling

The structure and electronic properties of a defect can only be found if the Schrödinger equation for the defect embedded in a solid is solved. This cannot be done exactly and there are several layers of approximations before a tractable solution can be found. The Schrödinger equation involves both the electrons and the ions and contains contributions from the kinetic energy of the ions and electrons as well as the electron–electron and electron–ion interaction.

The first approximation is due to Born and Oppenheimer, who argue that, since the nuclear masses are so much larger than the electron mass, then we can treat the electrons as moving in a field of static nuclei. The Schrödinger equation for the electrons then involves a potential due to fixed nuclear sites collectively denoted by R. The solution of this equation is the structural potential energy $E(R)$ and the electronic wavefunction. To determine dynamical quantities such as vibrational frequencies of the ions and diffusion barriers, we need then to solve the classical problem of the nuclei moving in the force field $-\nabla_R E(R)$. The Born–Oppenheimer approximation is invariably used but it must be remembered that it does break down and the motions of the electrons and ions are then independent. This happens in a Jahn–Teller distortion or during an electronic transition involving the emission of phonons. Details of these are found in [2].

Let us now consider the Schrödinger equation of the electrons in a field of frozen nuclei. This equation still cannot be solved exactly because of the inter-electronic Coulomb terms, $\sum_{i,j} V_{\mathrm{e-e}}(\mathbf{r}_i - \mathbf{r}_j)$. If these terms were neglected, then the Schrödinger equation is separable in the degrees of freedom of each electron and the exact wavefunction can be written down as a product of orbitals $\Pi_\lambda \psi_\lambda(\mathbf{r}_\lambda, s_\lambda)$ where $\psi_\lambda(\mathbf{r}, s)$ is the λth wavefunction of the one-electron Hamiltonian. Of course, we require that the total wavefunction is antisymmetric and this can be achieved by a suitable sum over the product orbitals. For example, for the H_2 molecule, the antisymmetric electronic wavefunction is

$$\psi_1(\mathbf{r}_1, s_1)\psi_2(\mathbf{r}_2, s_2) - \psi_1(\mathbf{r}_2, s_2)\psi_2(\mathbf{r}_1, s_1).$$

Here s_1, s_2 are spins (up or down) of the electrons. This type of wavefunction is called a Slater determinant and is specified by the electronic configuration, i.e. a listing of the orbitals λ occupied by the N electrons: for example, the ground state configuration, $1s^{\uparrow\downarrow}$ or $1s^2$. The true wavefunction is then a linear combination of Slater determinants.

The neglect of electron–electron interaction is too extreme to make this theory useful and various attempts have been made to take this into account. There are two principal schemes:

Hartree–Fock (HF) and DFT. Both of these replace the electrostatic potential acting on electron i by the other electrons, i.e.

$$\sum_j V_{\mathrm{e-e}}(\mathbf{r}_i - \mathbf{r}_j)$$

by an average over the positions of the other electrons. Thus this potential becomes $V_{\mathrm{eff}}(\mathbf{r}_i, s_i)$. Clearly the Schrödinger equation is then separable and the problem simplified. In both theories, the effective potential is composed of two terms: the Hartree potential $V_{\mathrm{H}}(\mathbf{r})$ and the exchange–correlation potential $V_{\mathrm{XC}}(\mathbf{r}, s)$. The former is the electrostatic potential due to the electron density $n(\mathbf{r})$ at point $\mathbf{r}$. The electron density is derived from the spin-density, $\sum_\lambda |\psi_\lambda(\mathbf{r}, s)|^2$, by summing over the two spin states. This is simply the probability of finding an electron at the point $\mathbf{r}$ with either spin. The exchange–correlation potential differs in the two theories. In HF, it is a complicated function determined by all the orbitals $\psi_\lambda(\mathbf{r}, s)$ and differs for different orbitals. In DFT, it is rigorously determined only by the electron spin density. This is an important result but the precise dependence on density is not known except for a particular problem: the homogeneous electron gas. In this case, the equations have been solved numerically and the dependence of V_{XC} on n is known exactly. For any other problem where the electron density varies throughout space, we assume that $V_{\mathrm{XC}}(\mathbf{r})$ at point $\mathbf{r}$ is given by the homogeneous electron gas value involving the spin-density $n(\mathbf{r}, s)$ at the same point $\mathbf{r}$. This is called the local spin-density approximation (LDA) and the theory is then called local density functional (LDF) theory. For high electron densities, it can be shown that V_{XC} is proportional to $n^{1/3}$ and thus in LDF theory $V_{\mathrm{XC}}(\mathbf{r})$ is also proportional to $n^{1/3}(\mathbf{r}, s)$. For smaller densities, V_{XC} can still be written down but is given by a more complicated expression. In both approaches the effective potential that enters the Schrödinger equation for the orbitals $\psi_\lambda(\mathbf{r}, s)$ depends on the occupied orbitals. The equation can only be solved through a self-consistent scheme where a guess is made for the input spin-density. The Schrödinger equation is then solved, and the output orbitals used to update the effective potential. The whole process is then iterated until the output wavefunctions are close to the input ones. This self-consistent method varies the spin-density until the energy is minimized which can be regarded as allowing charge to flow between atoms until the energy $E(R)$ is least.

The force acting on each atom can then be found from $-\nabla E(R)$ and the atoms moved until equilibrium is found. Of course the bond lengths and angles as well as cell parameters should then be related to those found experimentally by e.g. x-ray diffraction. The curvature of the energy about the equilibrium positions then determines the vibrational frequencies which can be related to modes found in IR experiments. The barriers to atomic movement can also be found and related to diffusion studies. Finally, the electronic energy levels can be found from the total energy of neutral and charged cells as described below. The optical transitions can also be investigated although the results for the absolute transition energies have so far been disappointing and will be discussed further below.

Although the self-consistent Schrödinger equation mentioned above seems simple to solve, there are still many steps required before a solution is at hand. Not all the electrons need be considered. They are divided into two groups: core and valence electrons. The former are bound close to the ions and do not play a part in bonding. It is highly advantageous to remove these from the theory. This can be done through the use of the pseudopotential. For example, in Si the 1s, 2s, and 2p electrons are regarded as part of the core while the 3s and 3p electrons are part of the valence shell. The four bonds associated with each Si atom in the bulk arise from these four valence electrons. The pseudopotential is constructed by insisting that it has exactly the same valence energy levels as the true atomic potential, i.e. the 3s and 3p and even the 3d levels are the same (the 3d levels are required because these orbitals are occupied to some

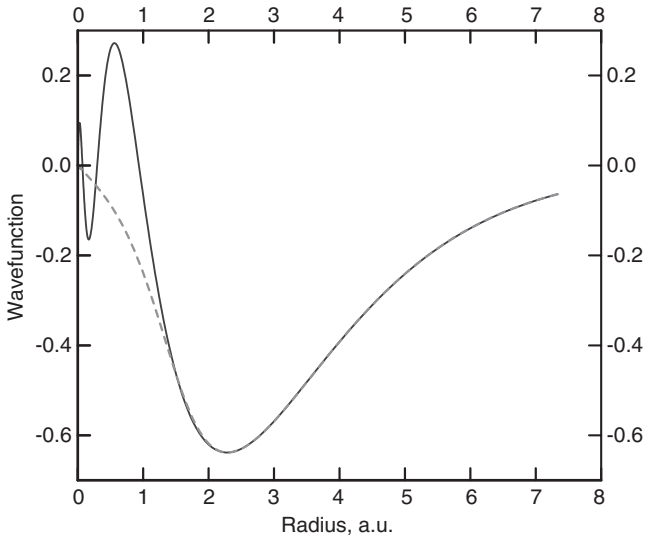

Figure 1. The 4s radial wavefunction (full curve) for the Ni atom where all electrons have been included and the pseudo-radial wavefunction (dashed).

extent in the solid). Moreover, the corresponding pseudo-wavefunctions are exactly equal to the true wavefunctions outside a core whose size depends on the identity of the atom.

Inside the core, the pseudo-wavefunctions are not equal to the true valence ones but are very smooth nodeless functions which are easy to represent mathematically. The true 3s and 3p wavefunctions of Si oscillate inside the core and are rather difficult to represent mathematically. Figure 1 shows these functions for Ni where the core is about 1.8 au (0.953 Å).

The crucial point is that pseudopotentials have no core states at all! The lowest bound state solutions are the valence energy eigenvalues and, as stated above, these are the same as the true valence energy levels. Whereas the superposition principle tells us that the electrostatic potential acting on the valence electron is simply the sum of the potentials due to the core and the valence electrons, this is not true for the exchange–correlation potential which is a non-linear function of the spin-density. Nevertheless, for many materials this potential can be linearized in the valence spin-density. For others however, such as GaN, it is necessary to include a non-linear core correction.

We have reached the point where the eigenvalue problem,

$$H\psi_\lambda(\mathbf{r}, s) = E_\lambda\psi_\lambda(\mathbf{r}, s),$$

has to be solved. Here, H includes the kinetic energy of a single electron, the sum of the pseudopotentials arising from all of the ions, and the effective potential V_{eff} describing the interaction between electrons and which depends on the spin-density of all the electrons $n(\mathbf{r}, s)$. There is an additional energy due to the interaction between the ions but this does not involve the coordinates of the electrons. To proceed, we write the wavefunction as the product of a space dependent term and a spin-function $\chi(s)$ and expand the spatial part in terms of a basis $\phi_i(\mathbf{r})$. Thus

$$\psi_\lambda(\mathbf{r}, s) = \chi_\lambda(s)\sum_i c_i^\lambda\phi_i(\mathbf{r}).$$

Here $\chi_\lambda(s)$ is simply the spin-up wavefunction if $s = 1/2$ and the spin-down counterpart otherwise. Two choices of ϕ are often made: plane waves and Cartesian Gaussian orbitals.

Using the first is equivalent to making a Fourier transform of the wavefunction. This has problems with certain elements, e.g. first row, transition and rare-earth ones, where the wavefunction varies rapidly and a great number of plane waves are required. AIMPRO uses Cartesian Gaussian orbitals of the form

$$x^{n_1} y^{n_2} z^{n_3} \mathrm{e}^{-ar^2},$$

when centred at the origin. In practice, Gaussian functions which are centred at the nuclei, and sometimes at bond-centred sites between the nuclei, are used. If $n_i = 0$ then the orbitals are simply spherically symmetric Gaussian functions sometimes called s Gaussian orbitals. If one of the n_i is unity and the others zero then the functions are p Gaussian orbitals, while if the sum of the n_i is two the set of six orbitals generates five d orbitals and one s orbital. The great advantage of Gaussian orbitals is that analytical expressions can be given for all the integrals that appear.

We now have to solve the equations

$$H \sum_j \chi_\lambda(s) c_j^\lambda \phi_j(r) = E_\lambda \chi_\lambda(s) \sum_j c_j^\lambda \phi_j(r).$$

The next step is to multiply this equation through by $\chi_\lambda^*(s)\phi_i^*(r)$ and integrate over r (and sum over s). This reduces the differential equation to a set of matrix equations:

$$\sum_j (H_{ij} - E_\lambda S_{ij}) c_j^\lambda = 0$$

which can be solved by matrix methods. H_{ij} and S_{ij} are matrix elements of the Hamiltonian and the overlap respectively.

Once this equation has been solved we can generate the output spin density

$$n_\mathrm{o}(\mathbf{r}, s) = \sum_\lambda |\chi_\lambda(s)\psi_\lambda(\mathbf{r})|^2$$

and the process iterated as explained above to achieve self-consistency.

$$n(\mathbf{r}, s) = \sum_k c_k(s) g_k(\mathbf{r}).$$

Once the self-consistent density has been found, the structural energy $E(R)$ can be evaluated. We then compute the forces on each atom from $-\nabla E(R)$ and move the atoms until this energy is a minimum and these forces vanish. This is called relaxing the assembly of atoms.

3. Applications to defects: the carbon interstitial

Carbon exists as an impurity in Si with a concentration rarely below 10^{14} cm^{-3}. Although as an isolated point defect it is a benign impurity, without electrical, magnetic or optical effects, it has the ability to trap other defects like hydrogen, oxygen and especially silicon interstitials. The resulting defects have important effects on the electronic properties of the material and these have ensured that carbon-related defects have received much attention from both experimentalists and theorists.

Substitutional carbon, C_s, was first identified in silicon in the 1960s [3, 4] when it was shown to be responsible for a local vibrational mode (LVM) at 607 cm^{-1}. Such a mode lies above the Raman frequency of Si and is easily detected in absorption experiments. Irradiation of silicon produces self-interstitials (Si_i) which migrate rapidly and bind to substitutional carbon to form the C_i complex. The first identification of C_i was made using infra-red spectroscopy [5]. This technique measures the absorption of infra-red light as it traverses a thick Si crystal

containing the defects. Absorption occurs by the excitation of a vibrational state of the defect, e.g. the stretch in the carbon–silicon bond lying along the C_2 axis. Two LVMs are observed at 960 and 936 cm^{-1} which appear just as the 607 cm^{-1} mode, due to C_s, disappears. Later work [6] correlated an EPR spectrum observed in irradiated p-type Si labelled G12 with the LVMs at 960 and 936 cm^{-1}. In EPR, a magnetic field splits the energy levels in $C_i^{\pm}$ whose spin is $1/2$ into two states (Zeeman splitting). The actual splitting depends on the g-tensor, which can then be found and whose symmetry is related to that of the defect. In addition, satellite transitions accompanying the main lines are due to a hyperfine interaction with nuclei having an overlap with the highest occupied wavefunction of the defect. G12 has C_{2v} symmetry with the principal axis along $\langle 001 \rangle$. The hyperfine interaction of the unpaired spin showed that it possessed p-like character with 36% localization on a unique carbon atom. The spectrum was identified with the positive charge state of the carbon interstitial, C_i^+. This shows that the defect has a donor level in the band gap as it is positively charged in boron doped Si.

Another important experimental quantity is the stress-tensor or piezospectroscopic tensor B. If a stress is imposed along the [100] axis, then the energy of C_i defects aligned along this direction increases (as the defect is compressive along this axis) while the energy of those defects aligned along [010] and [001] decreases. At a temperature T, the defect may reorientate and assume lower energy configurations. The population of those defects lying along [010] and [001] then increases at the expense of the [100] population. If the strain tensor caused by external stress is ϵ, then we can write the change in energy $\Delta E = \mathrm{tr}\, B \cdot \epsilon$ where the B-tensor is a measure of the tensile nature of the defect along its principal axis [7].

Stress alignment studies were performed to extract the B tensor and the reorientation barrier, found to be 0.88 eV. Isochronal annealing of G12 showed that heating the sample to 65 °C for 30 min removes the spectrum. This anneal is correlated with the loss of a hole trap at $E_v + 0.28$ eV in DLTS studies [8]. DLTS is a technique which measures the emission rates of majority carriers trapped in defects in the depletion region of a p–n junction when a voltage pulse is applied. From the emission rates, the depths of the traps from the band edges can be measured. These experiments placed the donor or (0/+) level of the defect to lie at $E_v + 0.28$ eV.

More recently, studies [9] of irradiated n-type material correlated the L6 EPR spectrum (C_{2v} symmetry) with an acceptor level measured by DLTS at $E_c - 0.10$ eV. The L6 centre has a similar B tensor, annealing range and reorientation barrier (0.77 eV) to G12. In contrast to the G12 EPR centre, no hyperfine splitting due to the presence of C or Si isotopes was observed. The lack of observable hyperfine interaction could be explained if the wavefunction is relatively delocalized and has a small overlap with the carbon atom.

PL experiments also detect a signal arising from C_i [10, 11]. The 856 meV line also has C_{2v} symmetry and corresponds to the neutral charge state of the defect. The activation energy [11] to migration was found to be 0.87 eV, similar to that found for the charged species.

Through consideration of the EPR symmetry and unpaired wavefunction the structure of C_i has been shown to exist in a $\langle 001 \rangle$ split configuration with a C and Si atom sharing a lattice site. Unambiguous confirmation that the neutral species also possesses this structure came from uniaxial stress measurements of the PL line [10] and the 902 and 931 cm^{-1} LVMs [12].

4. Theoretical density functional studies of C_i

The C_i defect has been investigated using both cluster-based and supercell methods. Two clusters were employed with configuration $C_1Si_{70}H_{60}$ and $C_1Si_{181}H_{116}$ and one supercell, configuration C_1Si_{64}. The vibrational modes are calculated using the supercell and are compared with previous AIMPRO calculations which used a $C_1Si_{35}H_{36}$ cluster [13].

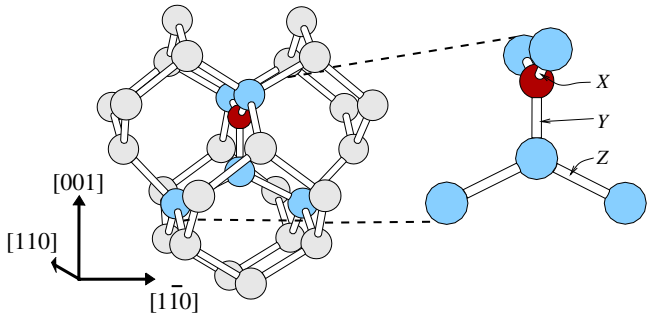

Figure 2. Structure of the carbon interstitial in silicon. Two atoms: one carbon and one silicon share a lattice site and are aligned along [100]. The lengths of the labelled bonds are given in table 1.

Table 1. Bond lengths (Å) of C_i evaluated using cluster and supercell approaches. Also shown for the neutral charge state are the results of *ab initio* calculations of Capaz *et al* [14]. C denotes cluster calculations and S denotes supercell.

		Method (No of Si atoms)				
Charge state	Bond	C (35) [13]	C (70)	C (181)	S (32) [14]	S (64)
	X	—	1.83	1.79	—	1.78
−	*Y*	—	1.75	1.77	—	1.74
	Z	—	2.26	2.24	—	2.26
	X	1.82	1.88	1.82	1.80	1.80
0	*Y*	1.73	1.70	1.71	1.77	1.73
	Z	2.30	2.29	2.25	2.22	2.26
	X	—	1.93	1.86	—	1.78
+	*Y*	—	1.74	1.77	—	1.74
	Z	—	2.28	2.26	—	2.25

In the clusters, the wavefunction fitting basis consists of independent s and p Gaussian orbitals with four different exponents, placed at each Si site. A fixed linear combination of two Gaussian orbitals are sited on the terminating H atoms. In addition, s and p orbitals are placed at each Si–Si bond centre to ensure a good description of the Kohn–Sham states. The charge density is fitted with four independent Gaussian functions with different widths on each Si atom and three on the terminating H atoms. One extra Gaussian function is placed at each Si–Si bond centre. In the supercell, the wavefunction fitting method is the same as that for the cluster. The charge density is fitted with a plane-wave basis using a cut-off energy of 40 au.

5. Structure

The relaxed structure of the neutral defect is shown in figure 2. C_i was optimized in all three charge states and the bond lengths at the defect core are given in table 1. There is a small dependence of the structure of C_i on charge state. Most of the calculated values from different methods agree to within ∼3%. The small differences between the bond lengths calculated in the small and large clusters/supercells demonstrate that the calculated structures are reasonably converged with respect to the system size.

Table 2. LVMs (cm^{-1}) of C_i in silicon calculated using the supercell and cluster method compared with experiment. The isotopic shifts of the LVMs with ^{13}C and ^{14}C are also shown.

	B_1			A_1		
	^{12}C	^{13}C	^{14}C	^{12}C	^{13}C	^{14}C
Experiment	930	26	49	921	29	54
Supercell	960	31	58	936	27	50
Cluster	867	28	52	922	28	52

6. Vibrational modes

Vibrational modes can provide quite accurate and valuable information on a defect, allowing immediate comparison with experiment. Such calculations give considerable help to an experimentalist in identifying various absorption peaks in the infra-red.

The vibrational modes for C_i have been evaluated here using the supercell method [15] and the results of previous cluster calculations [13] are also given. The first step is to optimize the defect cluster or supercell until the forces are zero. The energy double derivatives of each atom in the system must now be found. This involves moving each atom along an axis l by a distance ϵ (usually chosen to be 0.025 au), recalculating the self-consistent charge density, and finding the new forces. Say atom A is displaced along l. Atom B now feels a force along direction m of $f^+_{m,\mathrm{B}}(l, \mathrm{A})$. Atom A is then moved a distance $-\epsilon$ along l from its equilibrium position and a new force $f^-_{m,\mathrm{B}}(l, \mathrm{A})$ is found. The three points are used to determine the energy second derivatives to second order in ϵ:

$$D_{l,\mathrm{A},m,\mathrm{B}} = \frac{f^+_{m,\mathrm{B}}(l, \mathrm{A}) - f^-_{m,\mathrm{B}}(l, \mathrm{A})}{2\epsilon}. \tag{1}$$

Diagonalization of the complete dynamical matrix allows the eigenvalues of the vibrational modes to be determined. Since the displacement, ϵ, is finite, this method includes some contribution from all even powers of ϵ, not just the quadratic term and so some anharmonicity is included. For this reason the calculated frequencies are termed *quasi-harmonic*.

In practice it is not necessary to calculate the energy double derivatives for all the atoms within the system. Only those atoms in the defect core lie in a potential well which is significantly different from that of a bulk atom. For bulklike atoms away from the core then, the Musgrave–Pople valence-force potential [16] can be used.

From the optimized C_i in the 64-atom supercell, the energy double derivatives were calculated explicitly for the six atoms in the defect core (see figure 2). Using the Musgrave–Pople potential to simulate the potential energy surface for the remaining bulklike atoms of the supercell the vibrational modes are calculated. Three modes are calculated to lie above the Raman frequency with frequencies 554 and 936, 960 cm^{-1}.

The two higher modes have symmetries B_1 and A_1. The B_1 mode involves a swinging motion of the carbon atom between the two equivalent Si atoms to which it is bonded. The A_1 mode consists of stretching motion of bond Y (see figure 2). The mode symmetries confirm the experimental assignments and the frequencies are within 3% of the experimental values (table 2). The isotopic shifts of both modes upon replacement of the carbon atom with ^{13}C and ^{14}C have also been evaluated by experiment. The calculated values are also found to agree well with experiment.

Calculations employing a cluster with configuration $C_1Si_{35}H_{36}$ give reasonable results. The A_1 mode is calculated to have a frequency and isotopic shifts very close to experiment.

The calculated frequency of the B_1 mode however is about 7% lower than that observed so that it lies below the A_1 mode. The isotopic shifts of this mode are found to agree well.

These calculated values are quite typical of AIMPRO results. Vibrational mode frequencies calculated using the cluster method may differ by as much as 10% from experiment but the isotopic shifts are usually found to agree well. The supercell approach generally results in slightly more accurate determination of vibrational frequencies.

7. Migration paths and diffusion barriers

7.1. Methods

In order to calculate a migration path, several techniques can be used to constrain the system in order to force the defect away from its ground state configuration. The best technique depends on the complexity of the application.

For a simple diffusion process where the saddle point position for the diffusion step is known, one can take advantage of point group symmetry. Take for example VOH in silicon. In this case there are four Si atoms surrounding a vacant site with an oxygen atom pairing two Si dangling bonds. One other dangling bond is saturated by the H atom leaving a single dangling bond. At some finite temperature the H atom gains enough thermal energy to hop over to the dangling bond. This diffusion step has a saddle point corresponding to the configuration where the H atom is sited intermediate between two dangling bond atoms. The symmetry of this configuration is C_{2v}. The diffusion barrier may be calculated by taking the difference between the total energies of the optimized ground state and the optimized saddle point. Optimization of the saddle point is made possible by constraining the symmetry to C_{2v}.

A more sophisticated way to find the migration barrier is necessary for those cases where the saddle point is unknown. The first of these is the *orthogonal relaxation* method. The vector displacement of all equivalent atoms from the initial atomic configuration, $\mathbf{R}_i^1$ to the final structures $\mathbf{R}_i^2$ of the diffusion path is found, $\mathbf{v}_i = \mathbf{R}_i^1 - \mathbf{R}_i^2$. All atoms are shifted by a fraction α along vector $\mathbf{v}_i$. The atoms are then allowed to relax from these intermediate positions *only in directions orthogonal to* $\mathbf{v}_i$. Stepping α from zero to unity allows the diffusion path and energy to be mapped.

The third method uses a constraint on the relative bond lengths between specified atoms. For each constraint imposed, three atoms are included, say atoms A, B and C. The relative bond lengths between atoms A and B ($|\mathbf{R}_A - \mathbf{R}_B|$) and atoms A and C ($|\mathbf{R}_A - \mathbf{R}_C|$) are constrained such that $x = |\mathbf{R}_A - \mathbf{R}_B|^2 - |\mathbf{R}_A - \mathbf{R}_C|^2 =$ constant. This method is typically used when atom A is the diffusing atom, breaking a bond with atom B and subsequently making a bond with atom C during the diffusion. For more complex diffusion paths, a number of bond length constraints may be applied.

7.2. Diffusion of C_i

The migration barrier has been found experimentally to be 0.87–0.88 eV irrespective of charge state. The determination of a reorientation barrier for the L6 and G12 defect with a similar energy (0.78 eV) suggests that the migration path and reorientation occur simultaneously.

Previous calculations [14] have determined the activation energies for two diffusion paths indicated in figure 3. At the midpoint of diffusion path 1, the diffusing C atom lies at a bond-centred position. The complete diffusion step results in a 180° reorientation of the defect. The energy barrier to diffusion along this path has been calculated to be 1.1 eV using a $C_1Si_{35}H_{36}$ cluster [13]. Path 2 is markedly different from path 1 with a midpoint configuration which has

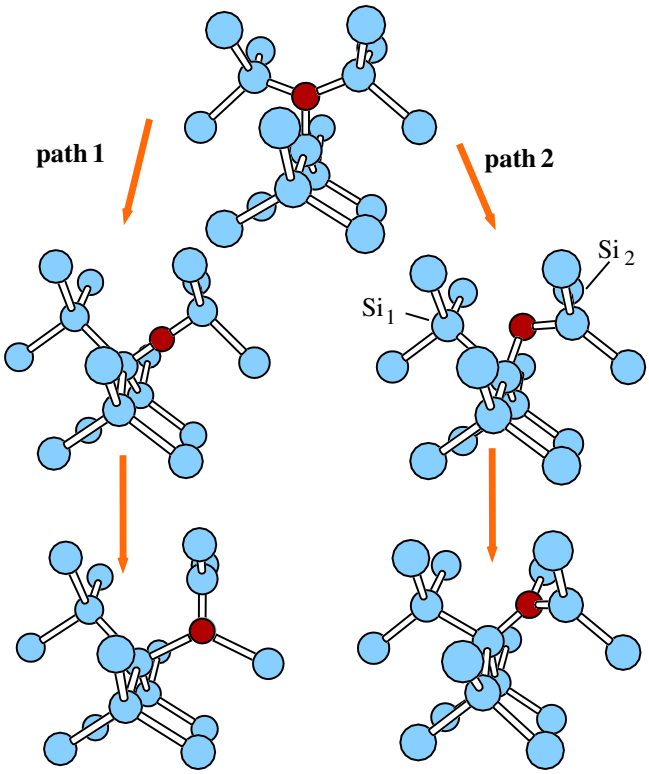

Figure 3. Possible migration paths of C_i. Path 1 has a saddle point for diffusion at the bond-centred site. The saddle point for path 2 has C_2 symmetry.

C_2 symmetry. This path explains the small difference between the observed reorientation and diffusion barrier as the defect principal axis reorients during each diffusion step. Capaz *et al* calculated the barrier to migration along this path to be 0.51 eV, in reasonable agreement with experiment, but the cluster results gave a value of 1.1 eV.

The different calculations give quite different results. It is often the case that the cluster method overestimates diffusion barriers whilst the supercell approach tends to underestimate barriers. Calculations of oxygen diffusion clearly follow this trend [17]. Recent results [18] for the diffusion of isolated oxygen in silicon give values of 2.2 and 2.8 eV using supercell and cluster AIMPRO respectively (cf experiment: 2.53 eV).

8. Stress response

The energy–stress tensor (B) is calculated using the supercell approach in the following way. The defect is first optimized in a number of supercells with different lattice parameters, typically within 0.2–0.4% of the ideal lattice parameter. The total energies of the supercell are fitted to a parabola to find the equilibrium lattice constant shown in figure 4 for the C_i defect. In this case we find the lattice parameter to expand by 0.33%. This expansion could in principle be measured by x-ray methods. It is comparable to that found for interstitial oxygen [18] and can be expected to arise for C_i as the extra atom is incorporated into the lattice and forms strong bonds with its neighbours. From this equilibrium volume, a uniaxial strain or shear is applied.

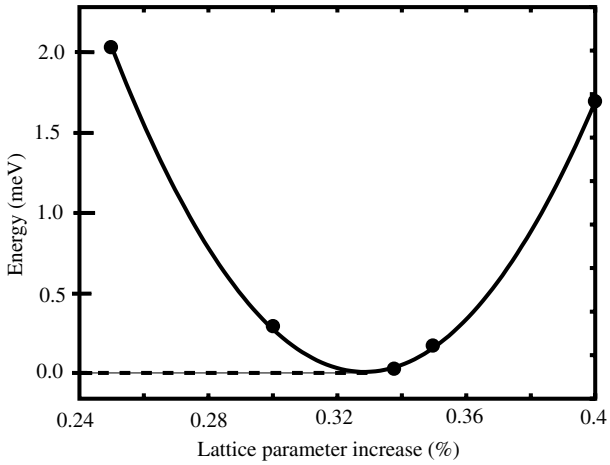

Figure 4. Total energy differences of the C_i defect in the supercell for various values of the lattice parameter, a_0.

Table 3. Calculated and experimentally determined B-tensor elements (eV) of the carbon interstitial.

Element	Calc.(0)	L6(−)	G12(+)
B_1	9.0	7.4	7.3
B_2	0.9	−0.2	0.3
B_3	−9.6	−7.2	−7.6

The number of different stresses or shears which need to be applied in order to evaluate all the tensor elements depends upon the defect symmetry. In the case of the C_i with C_{2v} symmetry, the strain must be applied along two separate directions.

The B-tensor was calculated for the neutral charge state of C_i. It is compared with the experimental values extracted from the stress response of the G12 and L6 EPR centres in table 3. In general we would expect the stress response of a defect to depend upon the charge state. The experimental values of the positive and negative charge states of C_i, however, show that the stress response is quite insensitive to charge state. The directions of the tensor elements are shown in figure 5. The principal element of the B-tensor, B_3, has a negative sign reflecting the compressive nature of the defect along its principal axis. The B_1 tensor element is much larger than B_2. Since B_1 lies along the axis containing a large component of the two equivalent C–Si bonds, it is probable that this reflects the sensitivity of these bonds to compression. The calculated values of the tensor elements shown in table 3 are accurate to around 20%. This is typical of the accuracy of these calculations.

9. Electronic structure

The Kohn–Sham eigenvalue spectrum for the three charge states of C_i is shown in figure 6. The neutral defect possesses two deep one-electron levels which arise from the two p orbitals in the defect, one associated with the threefold coordinated C atom, the other with the threefold coordinated Si atom. The lower of the two levels arises from the carbon orbital, reflecting the higher electronegativity of C than Si. The addition or extraction of one electron from the neutral defect results in a shift of the Kohn–Sham levels. The electronic wavefunction

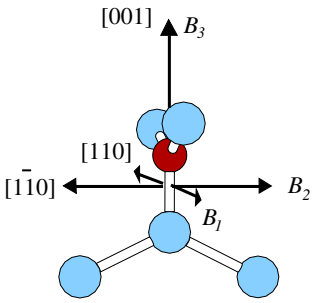

Figure 5. Directions of B-tensor elements of the carbon interstitial in silicon.

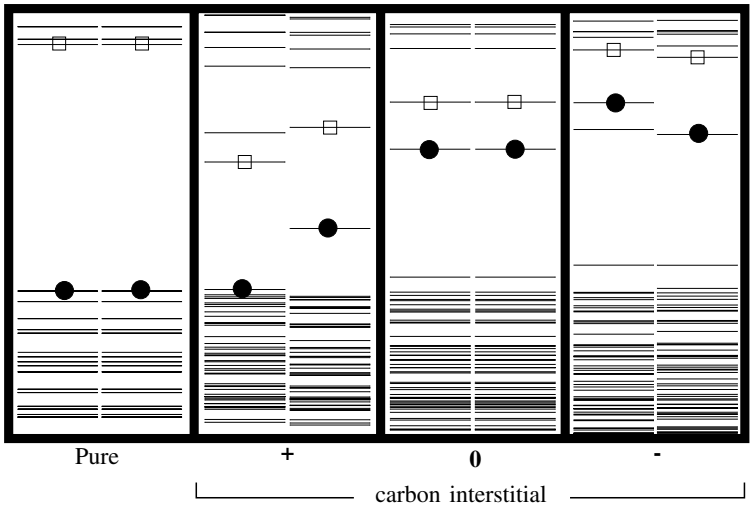

Figure 6. Kohn–Sham levels of the carbon interstitial in silicon. Left: eigenspectrum of a pure 131-atom cluster. Right: spin-up and down eigenspectra of the carbon interstitial in three charge states. The valence band edges have been aligned. The highest occupied and lowest unoccupied levels are indicated by filled circles and empty squares respectively.

corresponding to the unpaired spins of C_i^- and C_i^+ can be compared with the data extracted from the hyperfine satellites of the EPR centres L6 and G12.

The electronic distribution of the unpaired electron in both the positive and negative charge states is shown in figure 7. Clearly, the large spin density on carbon for the positive defect is in agreement with experiment. The two HOMO orbitals shown in figure 7 possess B_1 and B_2 symmetry. Performing a Mulliken analysis calculation allows the calculated Kohn–Sham wavefunction to be compared with the results of EPR experiments. This approach characterizes the wavefunction by estimating the amount of a wavefunction accounted for by each Gaussian fitting function. In this way the localization on each atom (and bond-centred function, if applicable) is estimated and also the s:p character of the wavefunction can be estimated. For the case of C_i^+ around 30% of the spin is localized on the C atom in good agreement with experiment (36%). This wavefunction at the C atom is purely p-like. The unresolved hyperfine interaction of the unpaired electron of C_i^+ (G12) with six to eight Si atoms is consistent with the calculations which show up to ∼4% amplitude on second shell atom sites. In the negative charge state the calculation shows 30% of the unpaired wavefunction is localized in a p orbital

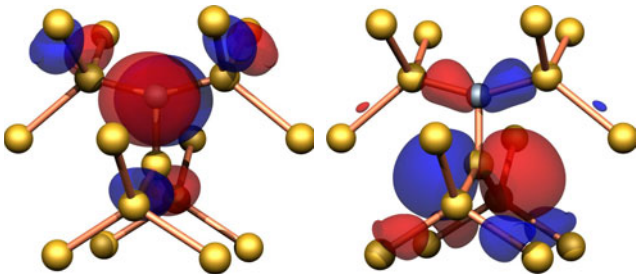

Figure 7. Isosurface plot of the unpaired wavefunction of C_i^+ (left) and C_i^- (right).

on the unique Si atom. The Mulliken analysis shows that the remaining portion of the unpaired spin is slightly more delocalized than for the positive charge state. The more distributed nature of the unpaired spin for the negative charge state may explain the lack of observed hyperfine splitting with ^{13}C or ^{29}Si in outer shells in the L6 EPR spectrum. However, from the calculated localization we would expect to find a hyperfine interaction with the unique Si nucleus in the negative charge state, and it is unclear why this is not observed.

9.1. Calculation of electrical levels

The method used to estimate the positions of the donor/acceptor levels includes an empirical correction. This technique involves a cluster calculation of the energy required to add electrons to (electron affinity) or extract electrons from (ionization energy) a defect. This energy is then compared with that for a defect with *known* electrical levels. The method has been used to characterize the electronic structures of chalcogen and chalcogen–hydrogen defects in Si [19], resulting in errors of around 0.2 eV.

The electrical levels of a defect correspond to the positions of the Fermi energy at which a charge state change occurs. The trap type is characterized in terms of the defect charge state when the trap is filled and when it is emptied. Hence, a simple donor level is denoted (0/+) and the acceptor levels is denoted (−/0).

The donor level is defined as the energy difference of the defect when a hole is promoted to the defect level from the top of the valence band, E_v

$$E_{tot}[E_d(0), E_v(h)] - E_{tot}[E_d(h), E_v(o)]. \tag{2}$$

The acceptor level is defined similarly. It is the difference in energy of the defect (a) when an additional electron is inserted into the bottom of the conduction band at E_c, and (b) when this electron occupies the lowest empty defect level:

$$E_{tot}[E_d(e), E_c(0)] - E_{tot}[E_d(0), E_c(e)]. \tag{3}$$

Alternatively, the donor level is the difference between the ionization energy of the defect and that of bulk Si. In cases where the defect wavefunction is enclosed by the cluster and does not interact with the cluster surface, the ionization energy can, in principle, be calculated. The cluster method, however, cannot be used to compute the ionization energy of bulk Si because the valence electron wavefunctions are delocalized and interact with the surface. The method employed circumvents this problem by including an empirical correction. The ionization energy, I of two defects, say defect A and defect B, is calculated. A is the defect for which we wish to evaluate the donor level and defect B is a defect with a known donor level position, $E(0/+)_B$. The position of the donor level of defect A can be easily evaluated,

$$E(0/+)_B = E(0/+)_A + I_B - I_A. \tag{4}$$

Table 4. Electrical levels of substitutional gold and the carbon interstitial. The calculated C_i levels are also shown.

	(0/+)	(−/0)
Au_s	Exp: $E_v + 0.35$	Exp: $E_c - 0.56$
C_i	Exp: $E_v + 0.28$	Exp: $E_c - 0.1$
C_i	Calc: $E_v + 0.42$	Calc: $E - 0.01$

The acceptor level is found in a similar fashion, this time using the difference in electron affinities to determine the difference in acceptor level positions.

Defect B should ideally be a well characterized defect, with well characterized levels. Additionally, the defect structure should be relatively insensitive to calculation details, making routine calculation unproblematic. We usually take the carbon interstitial as the marker defect as it possesses both an acceptor and a donor level and fulfils the above criteria. For the purposes of demonstration we calculate here the C_i donor and acceptor levels using substitutional gold as the marker defect. The Au_s defect possesses a donor and an acceptor level at $E_v + 0.35$ and $E_c - 0.56$ eV respectively [20–22]. The calculated and experimental electrical levels are within 0.2 eV of each other (see table 4).

10. Discussion

We have set out the main ingredients in the modelling of defects in semiconductors by quantum mechanical methods and illustrated the theory by concentrating on the carbon interstitial defect in Si. Both cluster and supercell calculations can be used to predict and explain a number of observations. The structural, electronic and vibrational characteristics of the defect have been calculated and compared with experiment. Good accuracy is achieved for the calculation of vibrational modes. The stress response tensor is calculated to be within ∼20% of the values determined from EPR stress studies. A semi-empirical method is employed to find the electrical levels of defects. This has been demonstrated to give an accuracy of around ±0.2 eV [19]. The Kohn–Sham wavefunction can often be used, together with Mulliken analysis of the wavefunction localization and s:p ratio in order to compare the calculation results with hyperfine data associated with paramagnetic defects measured by EPR.

One area where modelling has clear limitations is the prediction of optical transition energies as experimentally found in PL. The errors in the theory caused by deficiencies in the bandgap and a treatment of excited states are not the main reasons for the lack of agreement. As is often the case in silicon, Zeeman studies demonstrate that the radiative transition at 0.856 eV in C_i is due to the recombination of an exciton where the hole is a strongly bound to C_i in its donor level at $E_v + 0.28$ eV, and its electron partner is weakly bound to the Coulombic centre. The transition energy is then related to the donor level reduced by the weak binding energy of the exciton to the Coulombic centre. The precision by which the donor level can be calculated is, however, very much less than can be observed, and this limits the usefulness of the theory in modelling optical properties in detail.

References

[1] Yin M T and Cohen M L 1980 *Phys. Rev. Lett.* **45** 1004
[2] Stoneham A M 1975 *Theory of Defects in Semiconductors* (Oxford: Oxford University Press)
[3] Newman R C and Willis J B 1965 *J. Phys. Chem. Solids* **26** 373
[4] Newman R C and Willis J B 1969 *J. Phys. Chem. Solids* **30** 1943

[5] Bean A R and Newmann R C 1970 *Semicond. Sci. Technol.* **8** 175
[6] Watkins G D and Brower K L 1976 *Phys. Rev. Lett.* **36** 1329
[7] Watkins G D 1996 *Early Stages of Oxygen Precipitation in Silicon* (*NATO ASI Series* vol 17) ed R Jones (Dordrecht: Kluwer) p 1
[8] Kimerling L C 1977 *(Inst. Phys. Conf. Ser. 31)* p 221
[9] Song L W and Watkins G D 1990 *Phys. Rev. Lett.* **42** 5759
[10] Wooley R, Lightowlers E C, Tipping A K and Newman R C 1986 *Mater. Sci. Forum* **10–12** 929
[11] Thonke K, Teschner A and Sauer R 1987 *Phys. Rev.* B **61** 241
[12] Zheng J F, Stavola M and Watkins G D 1994 *The Physics of Semiconductors* ed D J Lockwood (Singapore: World Scientific) chapter 5, p 2363
[13] Leary P W 1997 *PhD Thesis* University of Exeter
[14] Capaz R B, Dal Pino A J and Joannopoulos J D 1994 *Phys. Rev.* B **50** 7439
[15] Coutinho J P 2001 private communication, unpublished
[16] Musgrave M J P and Pople J A 1962 *Proc. R. Soc.* **1268** 474
[17] Newmann R C and Jones R 1994 *Oxygen in Silicon* vol 42, ed F Shimura (New York: Academic) chapter 8
[18] Coutinho J, Jones R, Briddon P R and Öberg S 2000 *Phys. Rev.* B **62** 10824
[19] Coutinho J, Torres V J B, Jones R and Briddon P R 2003 *Phys. Rev.* **67** 035205
[20] Collins C B, Carlson R D and Gallagher C J 1957 *Phys. Rev.* **105** 1168
[21] Tasch A F and Sah C T 1970 *Phys. Rev.* B **1** 800
[22] Mesli A, Courcelle E, Zundel T and Siffert P 1987 *Phys. Rev.* B **36** 8049