LOCALIZED DEFECTS IN SEMICONDUCTORS

BY

STUART BRAND

A THEORETICAL THESIS

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SCHOOL OF PHYSICS

THE UNIVERSITY OF NEWCASTLE UPON TYNE

1977

P R E F A C E

The purpose of this thesis is to describe an investigation concerned with the problem of localized defects in semiconductors which was carried out in the Department of Theoretical Physics at the University of Newcastle upon Tyne during the period 1974-1977. The investigation was a cooperative effort involving Dr. Milan Jaros and the author.

The main results of the work have previously been put forward in the publications below:

Physical Review B14 p4496 (1976),

Proceedings of the 13th International Conference on the Physics of Semiconductors (Rome 1976) p1090,

Solid State Communications 21 p875 (1977).

In this thesis a more comprehensive discussion than was possible within the context of the published works is presented.

A C K N O W L E D G E M E N T S

I would particularly like to thank my supervisor Dr. Milan Jaros for his friendly guidance and participation during the course of this research. Gratitude is also due to the Professors of Physics for placing the facilities of the School of Physics at my disposal.

The computational labours have been made possible due to the availability of the resources of the Northumbrian Universities Multiple Access Computer installation.

Acknowledgement is made of the financial assistance of the Science Research Council who awarded me a grant for the tenure of my research.

Finally I would like to thank my parents for their generously given support during my years of studentship at the University of Newcastle upon Tyne.

CONTENTS

Page

Chapter 1 Introductory comments

1.1 Preliminary remarks 2

1.2 The localized states problem 4

1.3 The significance of localized states 6

1.4 The present work 7

Chapter 2 Background information

2.1 Omissions 10

2.2 The pure crystal Hamiltonian 10

2.3 The general form of solutions of the periodic potential problem 12

2.4 The orthogonalized plane wave pseudopotential method 15

2.5 The practical application of the pseudopotential method 19

2.6 Symmetry considerations 21

2.7 Summary 25

Chapter 3 Dealing with the localized defect problem

3.1 Opening remarks 27

3.2 General solutions of the defect problem 28

3.3 Particular solutions-the forbidden gap region 30

3.4 The choice of functions, g_N 33

3.5 Simplification of the form of the determinant 37

3.6 Labelling the defect states 43

3.7 The formation of the real space defect pseudopotential 44

3.8 Some practical aspects of the calculations 46

3.9 Summary 49

Chapter 4 Vacancies in semiconductors

4.1	Introduction	51
4.2	General review of vacancy studies	51
4.3	Vacancy-related defects in gallium arsenide-theory	58
4.4	Vacancy-related defects in gallium arsenide-experiment	76
4.5	Summary	81

Chapter 5 Displacement dependent energy levels

5.1	Introduction	83
5.2	The isolated nitrogen impurity in diamond	89
5.3	Further displacement studies : MPE capture	94
5.4	Summary	101

Chapter 6 General assessment of method

6.1	Introduction	104
6.2	Comparison with other methods	104
6.3	Concluding remarks	110

References

CHAPTER 1

INTRODUCTORY COMMENTS

Introductory Comments

Section 1.1 Preliminary remarks

In recent years there has been much interest in the general problem of point imperfections in solids, particularly in the context of the technologically important semiconducting materials. Quite apart from the possible practical implications there are also challenging academic reasons for pursuing studies of this subject. A number of features characteristic of low symmetry defect centres within an otherwise perfect crystal lattice await more full understanding by theoreticians. One of the most fundamental problems, and one which has yet to be dealt with in a totally satisfactory manner, is that of determining the response of the valence electrons to the presence of localized defects. (We shall treat this many-body problem by recourse to the perfect crystal screening function but this is obviously an approximate means.) In addition it would be valuable to better appreciate the precise numerical significance of the Jahn-Teller effect and other distortion-dependent phenomena such as the Stoke's shift and various aspects of the electron-phonon interaction. Unfortunately, much of the theory required for a completely detailed understanding of these topics, and others, has yet to be developed in a satisfactory form. Although some characteristics of defect systems have been accounted for by relatively simple approaches the necessity for large-scale computer studies is increasingly apparent if good numerical agreement is to be obtained. However, even with the aid of powerful computing facilities there are still many difficulties to be overcome. A vast amount of experimental information concerning semiconductor properties has been accumulated but much of this has still to be explained on a firm theoretical basis. This situation will almost certainly prevail for some time to come. Meanwhile, we must continually employ theoretical methods and models which may ultimately lead

to more complete understanding. The task is not an easy one, however, and even some of the more accepted results are continually in need of revision. Consequently it is not surprising that some of the less well understood properties of semiconductors, particularly those in relation to chemical impurities and intrinsic defects, are still subject to considerable discussion and indeed confusion.

The first, relatively uncomplicated impurities to be modelled theoretically were those leading to the production of "shallow" states (those with energy levels less than about 0.1 eV from the band edges) in silicon and germanium and these have been well understood for some time.(1) This understanding was achieved using the so-called Effective Mass Theory (EMT) which utilizes a straightforward coulombic impurity potential model with a dielectric constant and effective mass appropriate to the host material. This approach is directly analogous to the standard treatment of the hydrogen atom. Unfortunately, the agreement with experiment is in most cases only fair and additional central cell corrections for the particular impurity considered are necessary in order to explain the departure from the simple coulombic energy levels. Fabrication of many useful devices has followed from the deliberate modification of electrical properties of semiconductors by the controlled introduction of shallow state impurities.

In practice, it turns out that a number of observed energy levels do not fall into the shallow category. Many are found to lie well within the forbidden gap region.(2) EMT has been used in attempts to understand such levels but this theory is at its' best in dealing with states derived from a turning point or series of turning points of the upper valence band or lower conduction band with a correspondingly well defined effective mass. For a less well defined state in $\underline{k}$-space such as one composed of a spread of conduction or valence bands or both we need to formulate alternative methods. This is the situation which arises when we come to study the type

of impurity which leads to the production of levels which we classify under the heading "deep".

Section 1.2 The localized states problem

When an atom in a crystal is replaced by one of generally similar nature (in terms of atomic radius, core shells etc.) but of different valence then it is likely that predominantly coulombic type delocalized impurity states will result. The delocalization (≈50Å radius) is a direct result of the relatively long range of the impurity potential. However, this is not the only departure from perfect crystallinity which can be envisaged. If, for example, replacement by a much larger or smaller atom is attempted then distortion of some kind is likely to occur. In addition there is the classic vacancy problem which could again give rise to similar relaxation effects. Isoelectronic substituents can also produce impurity states yet these are certainly not of the aforementioned coulombic nature. They are due to short range differences in the potentials of the host and substituted atomic cores. There is also the problem of more complicated defect systems. It may be that complexes of two or more vacancies, impurities or impurity-vacancy pairs will occur. In such cases it may be expected that the strong, short range potentials will dominate in the formation of the defect states. Clearly, the long range coulomb potentials may have an influence but in general we may anticipate that these effects will be of secondary importance and would not expect energy modification to exceed the typical 0.1 eV EMT value. The energy levels produced will not necessarily be deep within the gap region, they may indeed be resonant with the valence or conduction bands. This is something which must be investigated for individual cases. However, we can expect that deep level states may result. With this in mind it is necessary to seek methods which can realistically be used in what is basically a short range potential problem.

In considering an impurity/defect system within the context of a bulk solid we are attempting to deal with what is, in effect, an infinite particle many-body problem. An exact solution is, of course, impossible and it is immediately apparent that the problem must be reduced to some sort of manageable form. In practice, the basic formulations for dealing with localized defects can be grouped into two main categories. In the molecular methods the problem is simplified by replacing the crystal with a finite size molecule which can then be treated by standard quantum chemistry techniques. Alternatively, we can formulate a representation based upon initial knowledge of the host crystal bandstructure.

The molecular orbital methods, exemplified by the early calculations of Coulson and Kearsley,[3] for example, who studied the diamond vacancy problem, rely on the treatment of the impurity/defect system as a "defect molecule". The relevant atomic orbitals or some approximation to them are required as an initial starting point and then suitable molecular orbitals are formed. Although Coulson and Kearsley included only nearest neighbours in their calculations much less restricted studies have since been attempted. In the more recent calculations of, for example, Messmer and Watkins[4] who also considered the vacancy in diamond, not only the nearest neighbours but also next nearest and possibly further removed neighbouring atoms are taken as being part of the defect molecule. It is essential that these larger molecules be studied if reliable interpretation with respect to the bulk solid is aimed for.

The Green's function Koster-Slater type methods proceed from a line of approach in which full knowledge of the host crystal eigenfunctions and energies is required. An intermediate set of localized functions is then introduced. In general, because of the need for complete knowledge of the host crystal eigenfunctions and the way in which this information appears when looking for solutions, this method can require substantial amounts of

computing time. The use of Wannier functions, as originally suggested by Koster and Slater[5,6] has been the basis of a number of calculations. However, the computer time required has usually curtailed the number of bands and/or sites used and this has inevitably reduced the value of results. It was only in 1967 that Callaway and Hughes were able to present results using for the first time a many-band calculation.[7] As will be shown later, however, localized functions other than the Wannier functions can be employed, thus leading to significant advantages.

In a given situation the symmetry of the resulting defect states is just about the only useful exact information which is available. Consequently, a feature common to all the localized defect methods is in the use of symmetry considerations to facilitate the calculations.

Section 1.3 The significance of localized states

As has already been pointed out the use of delocalized shallow donor/acceptor states in the modification of the electrical properties of semiconductors has been much exploited. There is also a great deal of interest in the practical application of the optoelectronic properties of semiconductors, particularly with regard to the production of light emitting devices. The emission of visible radiation requires that electron-hole recombination with resulting liberation of approximately 2 eV of energy occurs. Consequently, materials with band gaps, E_g, of this order are the primary source of interest with respect to this effect. Unfortunately, the most popular and hence most widely studied semiconductors, germanium and silicon, are of little use for such applications because of their relatively small E_g. This has directly led to the investigation of alternative materials because, quite apart from their intrinsic interest, they may be of useful practical importance. A number of direct band gap materials have been discovered with appropriate E_g but they have tended to suffer from difficulties of preparation or self-absorption or unsuitable electrical properties. As a result of this,

great interest has been focused upon indirect band gap materials, particularly gallium phosphide with $E_g \approx 2.3$ eV.

Due to the restrictions of momentum conservation across-the-gap recombination in indirect gap materials cannot occur unless phonon coupled and consequently emission levels are disappointingly low. However, the presence of localized states within the gap can, because of k-space spreading of the wave function and thus overlap at k=0, lead to the possibility of much higher radiative efficiencies. Unfortunately, radiative recombination must compete with other, non-radiative processes such as Auger or, perhaps, multiphonon emission (MPE) effects and so the search for satisfactory light emitting preparations is strewn with difficulties. It can be seen that a study of all localized defects, whether for their useful, or indeed adverse properties, may lead to significant knowledge.

Section 1.4 The present work

The purpose of the present work is twofold. To begin with we wish to demonstrate the practicability of a new method of dealing with the localized defect problem. The method, related to the Koster-Slater Green's function approach, is described in Chapter 3. In parallel with the feasibility study we shall attempt to gain some insight into various aspects of localized defect electron states. Although our main concern will be with the III-V semiconductors GaAs and GaP, results of studies in diamond will also be given.

In particular, we shall extend the scope of earlier calculations[8] in relation to the gallium vacancy in GaAs in order to allow for the effects of Jahn-Teller distortion. As it seems likely that many defect systems may be complexes we have generalized our formulation to include two-centre problems and results for the divacancy and vacancy-oxygen pairs in GaAs are presented. In addition, displacement studies of the nitrogen impurity in diamond and oxygen-like impurity potentials in GaAs, GaP and diamond are

carried out. Investigations of the latter type may have some relevance with regard to MPE capture processes and related phenomena.

CHAPTER 2

BACKGROUND INFORMATION

Background Information

Section 2.1 Omissions

In an account of the present kind it is not possible, nor indeed would it be appropriate, to include details of all the background ideas and theory upon which the work is based. Much of this detail has been widely discussed and documented in the literature and pressures of space preclude its' reiteration. However, in order to provide a reasonable basis of understanding for the methods used a broad outline of some of the fundamentals will be given in this chapter. Recourse to suitable references will be made both to fill in gaps and to expand upon the presented material.

Section 2.2 The pure crystal Hamiltonian

Making an immediate, explicit distinction between nuclei and electrons the Hamiltonian for a crystal can be expressed as

$$H = -\sum_I (\hbar^2/2M_I)\nabla_I^2 + (e^2/8\pi\epsilon_o) \sum_{I\neq J} Z_I Z_J/r_{IJ} - \sum_i (\hbar^2/2m_i)\nabla_i^2$$
$$+ (e^2/8\pi\epsilon_o) \sum_{i\neq j} 1/r_{ij} - (e^2/4\pi\epsilon_o) \sum_{Ii} Z_I/r_{Ii} \qquad (2.2.1)$$

where upper and lower case subscripts refer respectively to nuclear and electronic terms. Spin-orbit coupling is not included and therefore it will be assumed that all orbitals are doubly degenerate and can be occupied by two electrons of opposite spin. A simplification of the Hamiltonian can be achieved by making use of the Born-Oppenheimer approximation. The nuclei are considered as fixed at their equilibrium positions which lie at periodically positioned sites in the perfect crystal. This sets the first term of H as zero. The second term is a constant depending upon the details of the lattice system concerned and can be dismissed as we will be mainly interested in the electronic terms and not, for instance, in cohesive energies.

In order to make any reasonable progress at this point it becomes necessary to invoke the one-electron formulation. The simplest way in which we could make use of this would be to assume a total wave function of the Hartree form

$$\psi(\underline{q}_1, \underline{q}_2, \underline{q}_3, \ldots\ldots\underline{q}_N) = \prod_{i=1}^{N} \varphi_i(\underline{q}_i) \qquad (2.2.2)$$

where the $\underline{q}$ label both spatial and spin coordinates. However, ψ does not have the correct antisymmetric properties and so does not satisfy the Pauli exclusion principle. This form can, nevertheless, be useful but as we will aim towards an empirical solution eventually we can accept the additional complications which arise and utilize the full Hartree-Fock form. The determinantal function

$$\psi(\underline{q}_1, \underline{q}_2, \underline{q}_3, \ldots\ldots\underline{q}_N) = \begin{vmatrix} \varphi_1(\underline{q}_1) & \varphi_1(\underline{q}_2) & \cdots \\ \varphi_2(\underline{q}_1) & \varphi_2(\underline{q}_2) & \cdots \\ \vdots & \vdots & \\ & & \varphi_N(\underline{q}_N) \end{vmatrix} \qquad (2.2.3)$$

is automatically of the required antisymmetric form with respect to exchange of coordinates and hence satisfies the exclusion principle. Slater considered total wave functions of this type and showed that the individual one-electron wave functions are the self-consistent solutions of Equation (2.2.4).

$$\left[-\hbar^2/(2m)\nabla^2 - (e^2/4\pi\epsilon_o) \sum_I Z_I/r_{I1}\right] \varphi_k(\underline{q}_1)$$

$$+ \left[e^2/(4\pi\epsilon_o) \sum_{k'} \int \frac{|\varphi_{k'}(\underline{q}_2)|^2}{r_{12}} d\tau_2\right] \varphi_k(\underline{q}_1)$$

$$- \left[e^2/(4\pi\epsilon_o) \sum_{\substack{k' \\ \parallel\text{el spins}}} \int \frac{\varphi^*_{k'}(\underline{q}_2)\, \varphi^*_k(\underline{q}_1)\, \varphi_k(\underline{q}_2)\, \varphi_{k'}(\underline{q}_1)\, d\tau_2}{r_{12}\, \varphi^*_k(\underline{q}_1)\, \varphi_k(\underline{q}_1)}\right] \varphi_k(\underline{q}_1)$$

$$= E_k \varphi_k(\underline{q}_1) \tag{2.2.4}$$

Slater argued that the final term on the left, the so-called Fermi or exchange hole, is an insensitive function of k and can therefore be sensibly replaced by a straightforward potential function common for all k.[9] The one-electron Hamiltonian can thus be expressed as

$$H = T + V_n(\underline{r}) + V_{el}(\underline{r}) + V_{ex}(\underline{r}) \tag{2.2.5}$$

where T is the kinetic energy operator

$V_n(\underline{r})$ is the potential energy of an electron in the field of the nuclei (which is periodic for an infinite or cyclic lattice)

$V_{el}(\underline{r})$ is the potential due to all electrons, and

$V_{ex}(\underline{r})$ is the weighted average exchange potential.

Of course, it would be totally inconceivable to even consider the direct solution of such a problem for a macroscopic system. Exact analytic solutions are available for systems with one electron only. However, by using general arguments it is possible to obtain information concerning the form of the solutions as will be seen in the next section

Section 2.3 The general form of solutions of the periodic potential problem

Overlooking the precise details of the potential let us consider the general form of the solutions of

$$-\hbar^2/(2m)\nabla^2\varphi + V(\underline{r})\varphi = E\varphi \tag{2.3.1}$$

The information concerning the potential is limited to the knowledge that

$$V(\underline{r} + n_1\underline{a}_1 + n_2\underline{a}_2 + n_3\underline{a}_3) = V(\underline{r}) \tag{2.3.2}$$

where $\underline{a}_1$, $\underline{a}_2$ and $\underline{a}_3$ are the primitive translation vectors of the direct lattice and the n can take all integral values. This periodic form of potential means that we are considering a solid of infinite extent, or, more

conveniently, we can choose to employ cyclic Born-von Kármán boundary conditions. This restricts the solutions to those satisfying

$$\varphi(\underline{r} + N_1\underline{a}_1 + N_2\underline{a}_2 + N_3\underline{a}_3) = \varphi(\underline{r}) \qquad (2.3.3)$$

where N_1, N_2 and N_3 are suitable large integers. We now introduce the basic translation operators Ta_1, Ta_2 and Ta_3 which we define by their effect upon the position vector $\underline{r}$:

$$Ta_1\underline{r} = \underline{r} + \underline{a}_1, \quad Ta_2\underline{r} = \underline{r} + \underline{a}_2, \quad Ta_3\underline{r} = \underline{r} + \underline{a}_3 \qquad (2.3.4)$$

and their effect upon the wave function:

$$Ta_1\varphi(\underline{r}) = \varphi(Ta_1^{-1}\underline{r}) = \varphi(\underline{r}-\underline{a}_1) \quad \text{etc.} \qquad (2.3.5)$$

Some authors use a different notation when operating upon functions but this inverse form is employed throughout the present work. The effect of the general translation operator, T, involving any number of combinations of Ta_1, Ta_2 and Ta_3 is self-evident:

$$T\varphi(\underline{r}) = \varphi(\underline{r}-\underline{T}) \qquad (2.3.6)$$

Because of the construction of H it is apparent that

$$T(H\varphi) = H(T\varphi) = E(T\varphi) \qquad (2.3.7)$$

and this commutation relation implies that $(T\varphi)$ is also an eigenfunction of H with the same eigenvalue, E. It follows either that there are a semi-infinite number of essentially different wave functions, $(T\varphi)$, or that these are all equivelent, differing only by a phase factor such that

$$T\varphi = \lambda_T\varphi \qquad (2.3.8)$$

It is easily seen that eigenvalues given by

$$\lambda_T = e^{-i\underline{k}\cdot\underline{T}} \qquad (2.3.9)$$

are consistent with our requirements. The $\underline{k}$ are given by

$$\underline{k} = 2\pi(n_\ell/N_1)\underline{a}_1^* + 2\pi(n_m/N_2)\underline{a}_2^* + 2\pi(n_n/N_3)\underline{a}_3^* \qquad (2.3.10)$$

where $2\pi\underline{a}_1^*$, $2\pi\underline{a}_2^*$ and $2\pi\underline{a}_3^*$ are the familiar reciprocal lattice vectors which satisfy

$$\underline{a}_i^* \cdot \underline{a}_j^* = \delta_{i,j} \qquad (2.3.11)$$

Equation (2.3.8) can now be expressed as

$$\varphi_{\underline{k}}(\underline{r} + \underline{R}) = e^{i\underline{k}.\underline{R}}\,\varphi_{\underline{k}}(\underline{r}) \qquad (2.3.12)$$

where we have put $\underline{R} = -\underline{T}$ in order that the equation may take the usual form. In the limit as the $N \rightarrow \infty$, $\underline{k}$ can be treated as a continuous quantity. Equation (2.3.12) is consistent with $\varphi_{\underline{k}}(\underline{r})$ of the form

$$\varphi_{\underline{k}}(\underline{r}) = e^{i\underline{k}.\underline{r}}\,u_{\underline{k}}(\underline{r}) \qquad (2.3.13)$$

where $u_{\underline{k}}(\underline{r})$ has the same periodicity as the potential. This result was first demonstrated by Bloch for the particular case of a crystal lattice but was familiar to mathematicians from an earlier date under the title of Floquet's theorem. It can be seen that these solutions give rise to a charge density, $\varphi^*\varphi$, with the periodicity of the lattice and are thus perfectly consistent solutions of Equation (2.2.4). We have, of course, excluded in the initial formulation any terms which could have destroyed the periodicity.

Kronig and Penney applied the results which have just been presented to the case of a periodic square well potential in a one dimensional lattice and were able to show that simple results could be obtained.[10] However, the solution of a real physical problem such as that contained in Section 2.2, and which is at the root of bandstructure calculations, involves considerably greater problems and correspondingly requires the use of additional techniques. One of these techniques will be introduced in the following section.

Section 2.4 The orthogonalized plane wave pseudopotential method

We will now consider a method which will lead towards the solution of Equation (2.2.1). Equation (2.3.13) already gives an indication of the form of wave functions which may be expected. Expanding the periodic part, $u_{\underline{k}}(r)$, as a three dimensional Fourier series we obtain

$$\varphi_{\underline{k}}(\underline{r}) = e^{i\underline{k}.\underline{r}} \sum_{\underline{g}} a_{\underline{g}}(\underline{k}) e^{i\underline{g}.\underline{r}} \qquad (2.4.1)$$

where the $\underline{g}$ are the reciprocal lattice vectors which, of course, depend upon the details of the direct lattice structure. In the absence of any atomic cores we are reduced to the free electron situation in wich $a_o = 1$ and all other $a_{\underline{g}}$ are equal to zero. The solutions are then of the plane wave form, $e^{i\underline{k}.\underline{r}}$, and are normalized to unit volume. A constant potential also gives solutions of this type. When the self-consistent potential, $V(\underline{r})$, is periodic and small the use of perturbation theory might be considered in the calculation of the $a_{\underline{g}}$. The production of forbidden energy regions is then seen to be a consequence of this perturbation (in 1-d at least). This approach in which a continuous energy spectrum becomes split to form energy gaps is complementary to that of the tight-binding method in which discrete energy levels can be seen to spread out to form bands. Unfortunately, even for metallic materials perturbation theory is inadequate when applied in this way as the strength of the potential in the core regions gives rise to substantial modification of the simple plane wave solutions. The number of $a_{\underline{g}}$ with significantly greater than zero values becomes large and perturbation theory breaks down. Herring suggested that this problem could be overcome by the modification of the plane waves by the addition of atomic-like functions in the core regions.[11] These so called orthogonalized plane waves (OPW's) are given by

$$\phi_{\underline{k}+\underline{g}} = e^{i(\underline{k}+\underline{g}).\underline{r}} - \sum_{\alpha} a_{\alpha}(\underline{k}+\underline{g})\,\alpha \qquad (2.4.2)$$

where the α represent normalized atomic functions of all species present at their respective positions throughout the lattice. If the coefficients are given by

$$a_{\alpha}(\underline{k} + \underline{g}) = \int \alpha^{*} e^{i(\underline{k} + \underline{g}).\underline{r}} \, d\underline{r} \qquad (2.4.3)$$

then the ϕ are automatically orthogonal to all core states, hence giving rise to their name. This supposes, of course, that the core functions are so localized that it can be assumed that their overlap is negligible. Using (2.4.3) it is apparent that the substracted term in Equation (2.4.2) consists of a summation of Bloch-like LCAO tight-binding wave functions. For individual core states we assume that

$$\left[-\hbar^2/2m\nabla^2 + V(\underline{r})\right]|\alpha\rangle = E_{\alpha}\,|\alpha\rangle \qquad (2.4.4)$$

where the Dirac notation has now been adopted in the interests of future brevity. $V(\underline{r})$ is due predominantly to the local core potentials over the range where $\alpha(\underline{r})$ is significant and hence the deep core levels are probably little affected by the presence of the rest of the crystal. The E_{α} will be slightly different from their isolated atom values and the wave functions can reasonably be thought of as remaining unaltered. The original designation of those atomic states which can be treated in this manner must, of course, be dealt with in a sensible manner.

The solutions which we seek are expanded in terms of the OPW's:

$$|\varphi_{\underline{k}}\rangle = \sum_{\underline{g}} b_{\underline{g}}(\underline{k}) \; |\phi_{\underline{k} + \underline{g}}\rangle \qquad (2.4.5)$$

It can be seen that all eigenfunctions of this form must be orthogonal to all core states and therefore cannot converge to give these core wave functions. We have therefore produced a formulation which allows us to focus attention primarily upon the valence and conduction band characteristics, and it is

these which determine many of the properties specific to solids. Inserting Equation (2.4.5) in Equation (2.3.1) we obtain

$$\left[-\hbar^2/2m\nabla^2 + V(\underline{r})\right] \sum_{\underline{g}} b_{\underline{g}}(\underline{k}) \left[|\underline{k} + \underline{g}\rangle - \sum_{\alpha} \langle \alpha | \underline{k} + \underline{g}\rangle |\alpha\rangle \right]$$

$$= E_{\underline{k}} \sum_{\underline{g}} b_{\underline{g}}(\underline{k}) \left[|\underline{k} + \underline{g}\rangle - \sum_{\alpha} \langle \alpha | \underline{k} + \underline{g}\rangle |\alpha\rangle \right] \qquad (2.4.6)$$

which can be rearranged in the form

$$-\hbar^2/2m\nabla^2 \sum_{\underline{g}} b_{\underline{g}}(\underline{k}) |\underline{k} + \underline{g}\rangle + \left[V(\underline{r}) + \sum_{\alpha} (E_k - E_{\alpha}) |\alpha\rangle\langle\alpha| \right] \sum_{\underline{g}} b_{\underline{g}}(\underline{k}) |\underline{k} + \underline{g}\rangle$$

$$= E_k \sum_{\underline{g}} b_{\underline{g}}(\underline{k}) |\underline{k} + \underline{g}\rangle \qquad (2.4.7)$$

where use has been made of Equation (2.4.4).

Defining a new eigenfunction

$$|\varphi_{\underline{k}}^{ps}\rangle = \sum_{\underline{g}} b_{\underline{g}}(\underline{k}) |\underline{k} + \underline{g}\rangle \qquad (2.4.8)$$

which is known as the pseudo wave function, and a corresponding pseudopotential

$$V^{ps} = V(\underline{r}) + \sum (E_{\underline{k}} - E_{\alpha}) |\alpha\rangle\langle\alpha| \qquad (2.4.9)$$

Equation (2.4.7) takes the form

$$(T + V^{ps}) |\varphi_{\underline{k}}^{ps}\rangle = E_{\underline{k}} |\varphi_{\underline{k}}^{ps}\rangle. \qquad (2.4.10)$$

T is the kinetic energy operator. This new form of the equation has the same eigenvalues as, but different eigenfunctions to, those of $\varphi_{\underline{k}}$ and was originally formulated by Phillips and Kleinman.[12] It is apparent from Equations (2.4.2), (2.4.5) and (2.4.8) that the pseudo wave function is identical to the actual wave function outside the core regions and thus over most of space. Multiplying Equation (2.4.10) on the left by $|\underline{k} + \underline{g}'\rangle$ we

obtain

$$\left[\hbar^2/2m\,|\underline{k}+\underline{g}'|^2 - E_{\underline{k}}\right] b_{\underline{g}'}(\underline{k}) + \sum_{\underline{g}} b_{\underline{g}}(\underline{k})\,\langle \underline{k}+\underline{g}'\,|\,V^{ps}\,|\,\underline{k}+\underline{g}\rangle = 0 \qquad (2.4.11)$$

We therefore need to solve the secular equation

$$\left|\,(\hbar^2/2m\,|\underline{k}+\underline{g}'|^2 - E_{\underline{k}})\,\delta_{\underline{g},\underline{g}'} + \langle \underline{k}+\underline{g}'\,|\,V^{ps}\,|\,\underline{k}+\underline{g}\rangle\,\right| = 0 \qquad (2.4.12)$$

in order to obtain the required eigenvalues and eigenvectors. If the original expansion (2.4.1) had been retained the V^{ps} would simply have been replaced by V. The advantage of (2.4.12) is that the determinantal matrix is of much smaller size and consequently is easier to deal with. The reduction in size can be shown to be equivalent to a matrix partitioning process.

There are many interesting features of the pseudopotential formulation which could be described here but it is considered sufficient for our purposes that only the salient points be presented. The reader is referred to a number of alternative works in which full treatment of these various features can be found. The two properties which are of particular importance to us with regard to the use of the pseudopotential are given below.

1) The matrix elements of the pseudopotential are relatively small in comparison to the kinetic energy terms. This is because the second, orthogonalization component of V^{ps}, as was pointed out by Phillips and Kleinman originally and studied more closely by Cohen and Heine[13] and Austin, Heine and Sham,[14] is of a repulsive nature and consequently has the effect of cancelling the strong attractive potential, $V(\underline{r})$.

2) There is, in practice, considerable flexibility in the choice of pseudopotential: we need not involve ourselves in the difficulties associated with the form (2.4.9). It is basically the effect of the orthogonalization

operator upon the matrix elements and not the precise details of the operator itself which are important. Numerous models are available which take advantage of this fact.

We now briefly consider some consequences of the aforementioned points. The smallness of V^{ps} means that it may, in some circumstances, be possible to apply simple perturbation theory to the problem. This obviates the use of (2.4.12) and reduced the mathematical labour involved. Such an approach is usefully employed by Harrison in his illuminating discourse concerning the application of the pseudopotential method to the theory of metals.[15] In the case of semiconductors, in which we shall be mainly interested, the greater departure from the nearly free electron situation means that this technique would be less successful. It may be noted, although we have given no justification for it at this stage, that the replacement of the strictly non-local operator form of the orthogonalization term by a local potential (this deletes the dependence upon $\underline{k} + \underline{g}'$, $\underline{k} + \underline{g}$ and $E_{\underline{k}}$) would considerably simplify the problem.

Section 2.5 The practical application of the pseudopotential method

In the preceding sections a framework which can be usefully employed in the calculation of perfect crystal bandstructures has been developed. If the aim was to investigate the theoretical strength of this framework we could begin with an attempt to perform ab initio calculations of the electron eigenstates. The results could then be compared with experimental information. However, our main interest lies not in the testing of the veracity of the pseudopotential method but in its practical application to the study of the effects of imperfections of crystals. Consequently we desire as a starting point a reliable, but also convenient, description of the perfect crystal. If these initial conditions are not met then our final aims are unlikely to be realized. The requirements we seek are not fulfilled by the use of a completely ab initio approach and so we resort to

the adoption of an alternative technique. We choose to employ the empirical pseudopotential method. To begin with a local pseudopotential, $V^{ps}(\underline{r})$, is assumed. Consistent with earlier arguments it is taken to have the lattice periodicity:

$$V^{ps}(\underline{r}) = \sum_{\underline{g}} v^{ps}(\underline{g}) e^{i\underline{g}\cdot\underline{r}} \quad (2.5.1)$$

In the empirical pseudopotential method we determine values for $v^{ps}(\underline{g})$ (which can be complex) in order that experimental energy splittings are reproduced as nearly as possible when Equation (2.5.1) is inserted in Equation (2.4.12). As no spin dependent terms are included in the original Hamiltonian, suitably averaged energy values are used. Only the first few $v^{ps}(\underline{g})$ are fitted and all others are set to zero.

In addition to the expansion given by Equation (2.5.1) we introduce the concept of individual overlapping pseudoatom potentials such that

$$V^{ps}(\underline{r}) = \sum_{i,j} v^{ps}(|\underline{r}-\underline{R}_{i,j}|) \quad (2.5.2)$$

where i runs over all unit cells and j over atoms within the unit cell. It will later be seen that the use of individual atomic pseudopotentials is essential for our defect studies. The atomic pseudopotentials are spherically symmetric and thus consistent with linearly screened spherically symmetric core potentials (which corresponds to a system of filled atomic shells). It then follows that

$$v^{ps}(\underline{g}) = \sum_{j} \frac{e^{-i\underline{g}\cdot\underline{r}_j}}{\Omega} \int e^{-i\underline{g}\cdot\underline{r}}\, v_j^{ps}(\underline{r})\, d\underline{r} \quad (2.5.3)$$

where Ω is the unit cell volume and the $\underline{r}_j$ represent the positions of the pseudoatoms within the unit cell. Because of the sphericity of the atomic pseudopotentials the atomic form factors depend only upon the magnitude of the reciprocal lattice vectors. In calculating band structures the empirical pseudodopotentials of Cohen and Bergstresser [16] and Saslow et al.[17] will be used.

Of course, it would not be realistic to expect to achieve a completely accurate description of all crystal properties simply by the specification of a few parameters adjusted to fit observed energy separations. In any case, the experimental values themselves are not without error. However, the energy bands are well behaved, quasi-smoothly varying functions of $\underline{k}$ and if we obtain reasonable fits to the energy gaps available we can expect that the overall representation of the energy spectrum will be reasonable. In practice, agreements of fit are to about 0.1 eV in the gap regions between occupied and unoccupied states in the case of semiconductor materials. This does not in itself necessarily imply that determinations of physical properties based upon the bandstructure calculations are in similarly close agreement. It might be thought that the agreement could be improved simply by increasing the number of parameters but this does not automatically follow. This is because we have assumed a local potential for ease of calculation but we know that in principle a strictly non-local operator form should be used. We can not overcome the basic limitations which are inherent in this approximation by simply increasing the number of parameters. Ultimately, as in other cases, the value of the method must be judged on the basis of the quality of results produced.

Throughout the remainder of this work the superscript ^{ps} will be omitted for convenience and it will be implicit that all calculations are carried out within a pseudopotential scheme.

Section 2.6 Symmetry considerations

In Section 2.3 it was demonstrated that due to the translational invariance of the pure crystal Hamiltonian the eigenfunctions take a particular form and can be associated with the label $\underline{k}$. If it was so wished they could be found by resorting to the method previously prescribed for all $\underline{k}$. However, there are additional symmetries of H (i.e. operators which commute with H) and knowledge of these considerably simplifies the problem. The translations are, in fact, only a subgroup of the full crystal

space group which encompasses all geometrical symmetries of the crystal potential (and H). These extra symmetries consist of point operations (rotations etc.), possibly in association with non-primitive translations (screw rotations etc.), which leave the crystal potential unaltered. The effects of the operators defined in analogy with the so-called active viewpoint of Equations (2.3.4) and (2.3.5), are given in the Seitz notation by

$$\{\alpha | \underline{t}_\alpha\} \underline{r} = \alpha \underline{r} + \underline{t}_\alpha \tag{2.6.1}$$

and

$$\{\alpha | \underline{t}_\alpha\} \varphi(\underline{r}) = \varphi(\{\alpha | \underline{t}_\alpha\}^{-1} \underline{r}) \tag{2.6.2}$$

where

$$\{\alpha | \underline{t}_\alpha\}^{-1} = \{\alpha^{-1} | -\alpha^{-1} \underline{t}_\alpha\} \tag{2.6.3}$$

Operating upon a particular eigenfunction of the Bloch form produces another eigenfunction given by

$$\{\alpha | \underline{t}_\alpha\} \sum_{\underline{g}} a_g e^{i(\underline{k} + \underline{g}).\underline{r}} = \sum_{\underline{g}} a_{\underline{g}} e^{i(\underline{k} + \underline{g}).(\alpha^{-1}\underline{r} - \alpha^{-1}\underline{t}_\alpha)} \tag{2.6.4}$$

$$\therefore \quad \varphi_{\alpha \underline{k}}(\underline{r}) = e^{i\alpha \underline{k}.\underline{r}} \sum_{\underline{g}} a_{\underline{g}} e^{-i\alpha(\underline{k} + \underline{g}).\underline{t}_\alpha} e^{i\alpha \underline{g}.\underline{r}} \tag{2.6.5}$$

Knowledge that the α are real orthogonal matrices has been invoked. An additional phase Factor $e^{i\theta_\alpha(\underline{k})}$ can be included in the transformation in order that the Bloch functions vary smoothly in $\underline{k}$-space but this is superfluous to our arguments and so will be ignored from the outset. The reason for denoting the transformed wavefunction by $\varphi_{\alpha\underline{k}}(\underline{r})$ is apparent for in analogy with Equation (2.3.12) we have

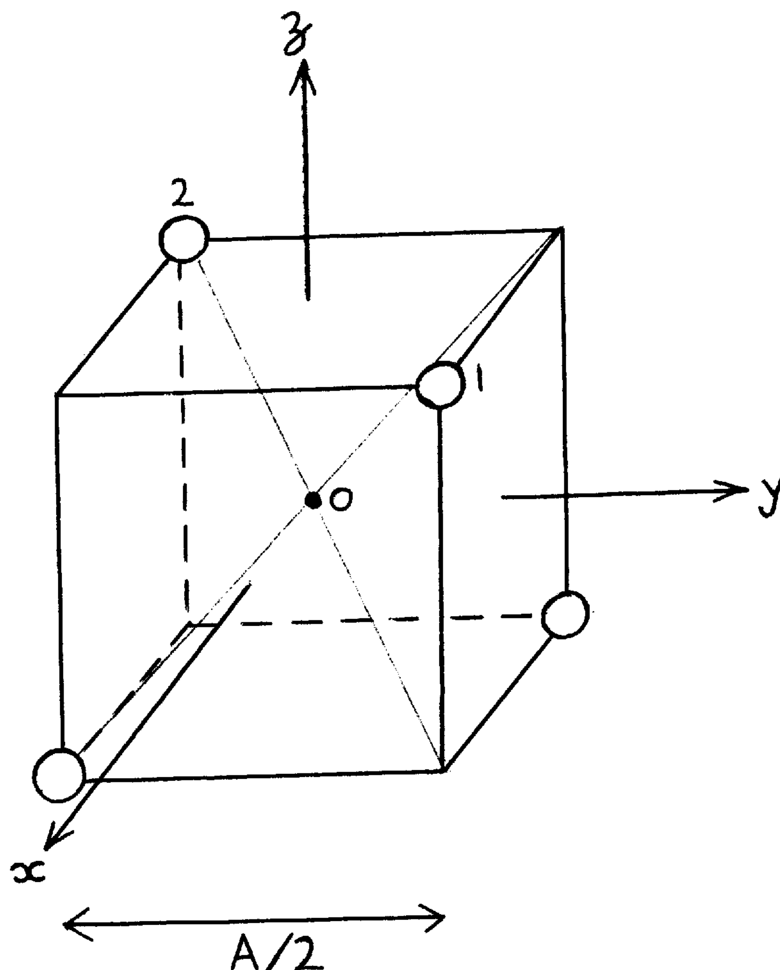
$$\varphi_{\alpha \underline{k}}(\underline{r} + \underline{R}) = e^{i\alpha \underline{k}.\underline{R}} \varphi_{\alpha \underline{k}}(\underline{r}) \tag{2.6.6}$$

The translation properties of the transformed Bloch function correspond to an eigenfunction consistent with the label $\alpha \underline{k}$. It follows that for a particular crystal system with n-fold symmetry that n-1 additional eigenfunctions associated with the same eigenvalue, $E_{\underline{k}}$, can be generated if $\varphi_{\underline{k}}$ is known. (This does not apply for $\underline{k}$ values lying at special points or lines but we shall not be concerned with such values.) In group theory terminology these degenerate $\varphi_{\underline{k}}$ are said to form an n-dimensional representation of the group of the Schrödinger equation. Only a knowledge of $\varphi_{\underline{k}}$ in 1/n of $\underline{k}$-space is initially required. This volume can be reduced still further if the geometrical inversion operator is not included in the α for then $\varphi^*_{\underline{k}}$, denoted by $\varphi_{-\underline{k}}$, must also be an additional eigenfunction (this is a consequence of time - reversal symmetry).

In the present work crystals with the face centred cubic Bravais lattice structure only will be encountered although the associated basis will vary. The reciprocal lattice, which depends only upon the Bravais lattice and not the basis, will thus be of the body centred cubic type. In the case of diamond, with the origin taken at an atom site, the full space group is denoted by $\Gamma_c^f O_h^7$ in the Schonflies notation. This contains the tetrahedral point group, T_d, as a subgroup. The total complement of 24 point group operations can be generated in terms of the direct product of the two subgroups shown overleaf. The particular factorization involving the C_{3v} subgroup will be seen to be of significance later. The space group is non-symmorphic as it includes the inversion coupled with a non-primitive translation of A/4(1,1,1) (A is the lattice constant). With the same choice of origin the symmorphic zinc-blende space group, $\Gamma_c^f T_d^2$, also has T_d symmetry. In this example the atoms of the basis are no longer identical and so the inversion related geometrical symmetry operation does not apply. However, use can be made of time-reversal symmetry and so for both structures initial knowledge of Bloch functions in only 1/48 of $\underline{k}$-space is required.

FIGURE 2·6·1

Td point group operations



24 Td operations formed from direct product

C_{3v}

$$[E, C_{2x}, C_{2y}, C_{2z}] \times [E, C_{31}^{+}, C_{31}^{-}, \sigma_{12}, \sigma_{12}C_{31}^{+}, \sigma_{12}C_{31}^{-}]$$

E represents the identity operation

C_{31}^{+-} represents $\pm$ 120° rotations about 01 axis

σ_{12} represents reflection in plane 012

$C_{2x,y,z}$ represents 180° rotations about x, y, z axes

Section 2.7 Summary

In this chapter a description has been given, albeit in a less than complete fashion, of the manner in which the pure crystal problem can be tackled. The idea of a self-consistent periodic potential is seen to be compatible with the one-electron formulation. It has been with this general result in mind, rather than with the precise details of the form of the one-electron Hamiltonian, that the solution of the pure crystal problem has been approached. Using the idea of a pseudopotential, in particular the empirical pseudopotential method of Cohen and Bergstresser, and making use of the simplifying action of symmetry considerations it has been shown how the crystal (pseudo) eigenfunctions and energies can be obtained. This initial knowledge is of crucial importance with regard to the method in which the localized defect problem shall be treated.

Much more detailed discussion of the one-electron formulation,[18] the pseudopotential method[19] and various aspects of group theory and crystal structures[20,21,22] is readily available elsewhere.

CHAPTER 3

DEALING WITH THE LOCALIZED DEFECT PROBLEM

Dealing with the localized defect problem

Section 3.1 Opening remarks

In the previous chapter attention was focused upon the calculation of the perfect crystal bandstructure (this being a pre-requisite to our further studies). Such an unblemished object is never realized in practice. Even the most carefully prepared crystal contains, although possibly small in percentage terms, a very large number of deviations from the perfect. These deviations take a number of different forms including vacancies and substituted atoms, possibly coupled with related lattice relaxation effects. For convenience, a more general definition than is standard is used and all such excursions from the perfect will be grouped together as being "defects" of the system. The nature of such defects is varied. If we consider spatial extent, for example, the effective volume of influence of simple coulombic impurities can be several hundred times greater than that of localized defects. It is not very surprising to find that no single technique is suitable for use in all cases. For example, effective mass theory is suitable for dealing with impurity states of long range coulombic origin but cannot be expected to be as successful in short range potential situations. Defects involving not only long range coulombic but also short range potentials should ideally be treated in a manner which deals equally well with both of these contributions. However, such a general method is not likely to be easy to implement as it would almost certainly lead to great, if not prohibitive, computational effort if treated comprehensively. The scheme presented here has no pretensions of dealing with totally general situations. It is specifically designed to cope with localized defect problems and any long range coulombic potential is considered as being only a modifying factor and as such of secondary importance in influencing the characteristics of the defect states. Before describing the details of the method a slight digression to clarify notation is in order.

In Chapter 2, for convenience, the pure crystal eigenfunctions and energies were distinguished by use of the label $\underline{k}$. However, scrutiny of Section 2.3 suggests that this system of labelling is not unique for any label $\underline{k} + \underline{g}$, where $\underline{g}$ is a reciprocal lattice vector, is consistent with the expected translational properties of φ. We shall choose to make use of the reduced zone representation in which $\underline{k}$ is confined to the first Brillouin zone (the unit cell of the reciprocal lattice constructed using the Wigner-Seitz method) and an additional band index, n, labels the Bloch functions in order of increasing energy.

Section 3.2 General solutions of the defect problem

Our object is to find solutions of the defect Schrödinger equation:

$$(H_o + h)\psi = E\psi. \qquad (3.2.1)$$

H_o is the pure crystal Hamiltonian for which it is assumed that (pseudo) eigenfunctions and energies given by

$$H_o \varphi_{n,\underline{k}} = E_{n,\underline{k}} \varphi_{n,\underline{k}} \qquad (3.2.2)$$

have already been found. The problem of determining the real space defect (pseudo) potential, h, is discussed in Section 3.7. ψ is expressed as an expansion in terms of the pure crystal Bloch functions, $\varphi_{n,\underline{k}}$, such that

$$\psi = \sum_n \int_{BZ} d\underline{k} \; A_{n,\underline{k}} \; \varphi_{n,\underline{k}} \qquad (3.2.3)$$

Inserting this expansion into the time-independent Schrödinger equation (3.2.1), multiplying by a particular $\varphi^*_{n',\underline{k}'}$ and integrating over all space the consistency relations

$$A_{n',\underline{k}'}(E_{n',\underline{k}'} - E) + \sum_n \int_{BZ} d\underline{k} \; A_{n,\underline{k}} \langle \varphi_{n',\underline{k}'} | h | \varphi_{n,\underline{k}} \rangle = 0 \quad (3.2.4)$$

are obtained (φ is normalized in form $\langle \varphi_{n',\underline{k}'} | \varphi_{n',\underline{k}'} \rangle = \delta(\underline{k}-\underline{k}')\delta_{n,n'}$).

Jaros and Ross have used a direct finite sampling point technique to solve this system of equations as they stand [23,24] but there are advantages in following an alternative method along the lines suggested by Bassani et al.[25,26] To begin with the potential, h, is assumed to be of factorizable form:

$$h = h^a h^b. \tag{3.2.5}$$

Making use of the complete set of orthonormal functions $g_N(\underline{r})$ the identity

$$\int d\underline{r}' \, \delta(\underline{r}-\underline{r}') = \sum_N \int d\underline{r}' \, g_N^*(\underline{r}') \, g_N(\underline{r}), \tag{3.2.6}$$

together with Equation (3.2.5) is inserted into Equation (3.2.4). The resulting equations are

$$A_{n',\underline{k}'}(E_{n'\underline{k}'} - E) + \sum_N \sum_n \int_{BZ} d\underline{k} \, A_{n,\underline{k}} \langle \varphi_{n',\underline{k}'} | h^a | g_N \rangle \langle g_N | h^b | \varphi_{n,\underline{k}} \rangle = 0 \tag{3.2.7}$$

It is apparent that the effect of the insertion of the g_N is to achieve a separation of the real space integral into a product of integrals containing only $\varphi^*_{n',\underline{k}'}$ or $\varphi_{n,\underline{k}}$ alone. Defining new coefficients A_N given by

$$A_N = \sum_n \int_{BZ} d\underline{k} \, A_{n,\underline{k}} \langle g_N | h^b | \varphi_{n,\underline{k}} \rangle \tag{3.2.8}$$

we obtain a modified version of Equation (3.2.7):

$$A_{n',\underline{k}'}(E_{n',\underline{k}'} - E) + \sum_N A_N \langle \varphi_{n'\underline{k}'} | h^a | g_N \rangle = 0 \tag{3.2.9}$$

For a particular value of E the general solution, if it exists, is given by

$$A_{n',\underline{k}'} = \frac{P\sum_N A_N \langle \varphi_{n',\underline{k}'} | h^a | g_N \rangle}{(E - E_{n',\underline{k}'})} + \beta(n',\underline{k}',E)\, \delta(E-E_{n',\underline{k}'}) \tag{3.2.10}$$

where β is determined by boundary conditions and P signifies that the principal part must be taken in any integral. As $h \rightarrow 0$ the first term in Equation (3.2.10) disappears and $\beta \rightarrow 1$ because the scattering states must tend towards the pure crystal Bloch functions as the defect potential vanishes.

the general solution of the defect problem for a state with energy E is given by

$$\psi_E = P\sum_n \int \frac{d\underline{k}}{BZ} \frac{\sum_N A_N F_N^{*a}(n,\underline{k}) \varphi_{n,\underline{k}}}{(E - E_{n,\underline{k}})}$$

$$+ \sum_n \int \frac{d\underline{k}}{BZ} \beta(n,\underline{k},E)\, \delta(E - E_{n,\underline{k}})\, \varphi_{n,\underline{k}} \tag{3.2.11}$$

where

$$F_N^a(n,\underline{k}) = \langle g_N | h^a | \varphi_{n,\underline{k}} \rangle . \tag{3.2.12}$$

Section 3.3 Particular solutions - the forbidden gap region

At this stage it does not appear that any general simplification of the defect problem has been achieved, and indeed this is true. However, if attention is directed solely towards the solutions for values of E that lie within the forbidden gap region then some progress can be made. This is, in fact, the restriction which we shall accept. The aim will be to find particular isolated solutions within the region of the pure crystal energy gap between occupied and unoccupied states. The solutions in the normal band regions will be altered to some extent but the wave functions should asymptotically tend towards the pure crystal Bloch functions as the defect potentials we shall consider will be highly localized. It will be assumed that the overall bandstructure remains essentially unaltered.

For $E \neq E_{n',\underline{k}'}$ (for any $n',\underline{k}$) we have from Equation (3.2.10)

$$A_{n',\underline{k}'} + \sum_N A_N \frac{\langle \varphi_{n',\underline{k}'} | h^a | g_N \rangle}{(E_{n',\underline{k}'} - E)} = 0 \tag{3.3.1}$$

Multiplying by $F^{b}_{N'}(n',\underline{k}')$ and integrating over all $n',\underline{k}'$ we obtain

$$A_{N'} + \sum_{N} A_N \int_{BZ} \frac{d\underline{k}' \; F^{*a}_{N}(n',\underline{k}')F^{b}_{N'}(n',\underline{k}')}{(E_{n',\underline{k}'} - E)} = 0 \qquad (3.3.2)$$

The consistency relations are now formulated completely in terms of the coefficients A_N rather than the $A_{n,\underline{k}}$. The advantage of this is that if the functions g_N are chosen sensibly then the factorized defect potential overlap integrals, F, will quickly converge to zero and the number of A_N to be considered will be small. Solutions are obtained by searching for energies, E_d, corresponding to zeros of the Fredholm determinant

$$\left| \delta_{N'N} + \sum_{n} \int_{BZ} \frac{d\underline{k} \; F^{*a}_{N}(n,\underline{k})F^{b}_{N'}(n,\underline{k})}{(E_{n,\underline{k}} - E_d)} \right| = 0 \qquad (3.3.3)$$

Sufficient trial values are used in order that the E_d be located as accurately as required. When the defect state energy eigenvalues have been found (if they exist) the related eigenvector, $\begin{pmatrix} A_1 \\ A_2 \\ \vdots \end{pmatrix}$ can be calculated.

Having done this there are, in principle, two direct methods which can be used to calculate the defect wave functions. The $\underline{k}$-space coefficients, $A_{n,k}$, can be found using Equation (3.3.1) and can then be inserted in Equation (3.2.3). Alternatively we might be able to proceed as follows.

Let us expand the defect wave function as

$$|\psi E_d\rangle = \sum_{j} a_j \, | g_j \rangle \qquad (3.3.4.)$$

Multiplying by $\langle g_i | h^b$ we obtain

$$\langle g_i | h^b | \psi_{E_d} \rangle = \sum_j a_j \langle g_i | h^b | g_j \rangle \qquad (3.3.5)$$

Inserting the $\underline{k}$-space expansion for ψ_{E_d}, then

$$\sum_n \int \frac{d\underline{k}}{BZ} A_{n,\underline{k}} \langle g_i | h^b | \varphi_{n,\underline{k}} \rangle = \sum_j a_j \langle g_i | h^b | g_j \rangle \qquad (3.3.6)$$

i.e.

$$A_i = \sum_j a_j \langle g_i | h^b | g_j \rangle \qquad (3.3.7)$$

In matrix form

$$\begin{pmatrix} \langle g_1 | h^b | g_1 \rangle & \langle g_1 | h^b | g_2 \rangle & \cdots \\ \langle g_2 | h^b | g_1 \rangle & \langle g_2 | h^b | g_2 \rangle & \cdots \\ \vdots & \vdots & \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \end{pmatrix} \qquad (3.3.8)$$

$$\therefore \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \langle g_1 | h^b | g_1 \rangle & \langle g_1 | h^b | g_2 \rangle & \cdots \\ \langle g_2 | h^b | g_1 \rangle & \langle g_2 | h^b | g_2 \rangle & \cdots \\ \vdots & & \end{pmatrix}^{-1} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \end{pmatrix} \qquad (3.3.9)$$

The real space expansion coefficients, a_i, are simply related to the A_i. It is interesting to note that if the defect potential is factorized in the form

$$h = h \times 1 \qquad (\text{i.e. } h^b = 1) \qquad (3.3.10)$$

then the particularly simple result

$$a_i = A_i \tag{3.3.11}$$

is obtained. The defect wave function is then readily calculated.

Section 3.4 The choice of functions, g_N

In section 3.2 the complete orthonormal set of functions $g_N(\underline{r})$, was introduced in order to facilitate the defect calculations. The particular choice of functions will now be revealed. We take

$$g_N(\underline{r}) = G_{n,\ell}(r)\, Y_{\ell,m}(\theta,\phi) \tag{3.4.1}$$

with $n = 0,1,2, \ldots$ $\quad \ell = 0,1,2\ldots n$ $\quad m = -\ell, -\ell+1,\ldots 0,\ldots, +\ell$

The $Y_{\ell,m}(\theta,\phi)$ are the familiar spherical harmonics which, of course satisfy

$$\int_0^{\pi}\int_0^{2\pi} Y^*_{\ell',m'}(\theta,\phi)\, Y_{\ell,m}(\theta,\phi)\sin\theta\, d\theta\, d\phi = \delta_{\ell',\ell}\,\delta_{m',m} \tag{3.4.2}$$

The $G_{n,\ell}(\underline{r})$ are given by

$$G_{n,\ell}(\underline{r}) = \frac{e^{-r/2}}{r}\, r^{\ell/2} \left(\frac{L^{\ell}_{n-\ell}(r)}{\sqrt{C_{n,\ell}}} \right) \tag{3.4.3}$$

where the $L^{\ell}_{n-\ell}(r)$ are the associated Laguerre polynomials. The normalization constant,

$$C_{n,\ell} = \frac{n!}{(n-\ell)!} \tag{3.4.4}$$

ensures that the required normalization condition for the radial functions is satisfied.

$$\int_0^{\infty} G_{n',\ell}(\underline{r})G_{n,\ell}(\underline{r})r^2 dr = \delta_{n',n}. \tag{3.4.5}$$

TABLE 3.4.1

Normalized Laguerre Coefficients $\ell = 0$ (correct to 5 sig. figs.)

Powers of r

n	r^0	r^1	r^2	r^3	r^4	r^5	r^6	r^7	r^8	r^9
0	1.0									
1	1.0	-1.0								
2	1.0	-2.0	0.5							
3	1.0	-3.0	1.5	-1.6667						
4	1.0	-4.0	3.0	-6.6667	4.1667					
5	1.0	-5.0	5.0	-16.667	20.833	-8.3333				
6	1.0	-6.0	7.5	-33.333	62.5	-50.0	1.3889			
7	1.0	-7.0	10.5	-58.333	145.83	-175.0	9.7222	-1.9841		
8	1.0	-8.0	14.0	-93.333	291.67	-466.67	38.889	-15.873	2.4802	
9	1.0	-9.0	18.0	-140.0	525.00	-1050.0	116.67	-71.429	22.321	-2.7557
	1	1	1	10^{-1}	10^{-2}	10^{-3}	10^{-3}	10^{-4}	10^{-5}	10^{-6}

Multiply coefficients by

TABLE 3.4.2

Normalized Laguerre Coefficients $\ell = 1$ (correct to 5 sig. figs.)

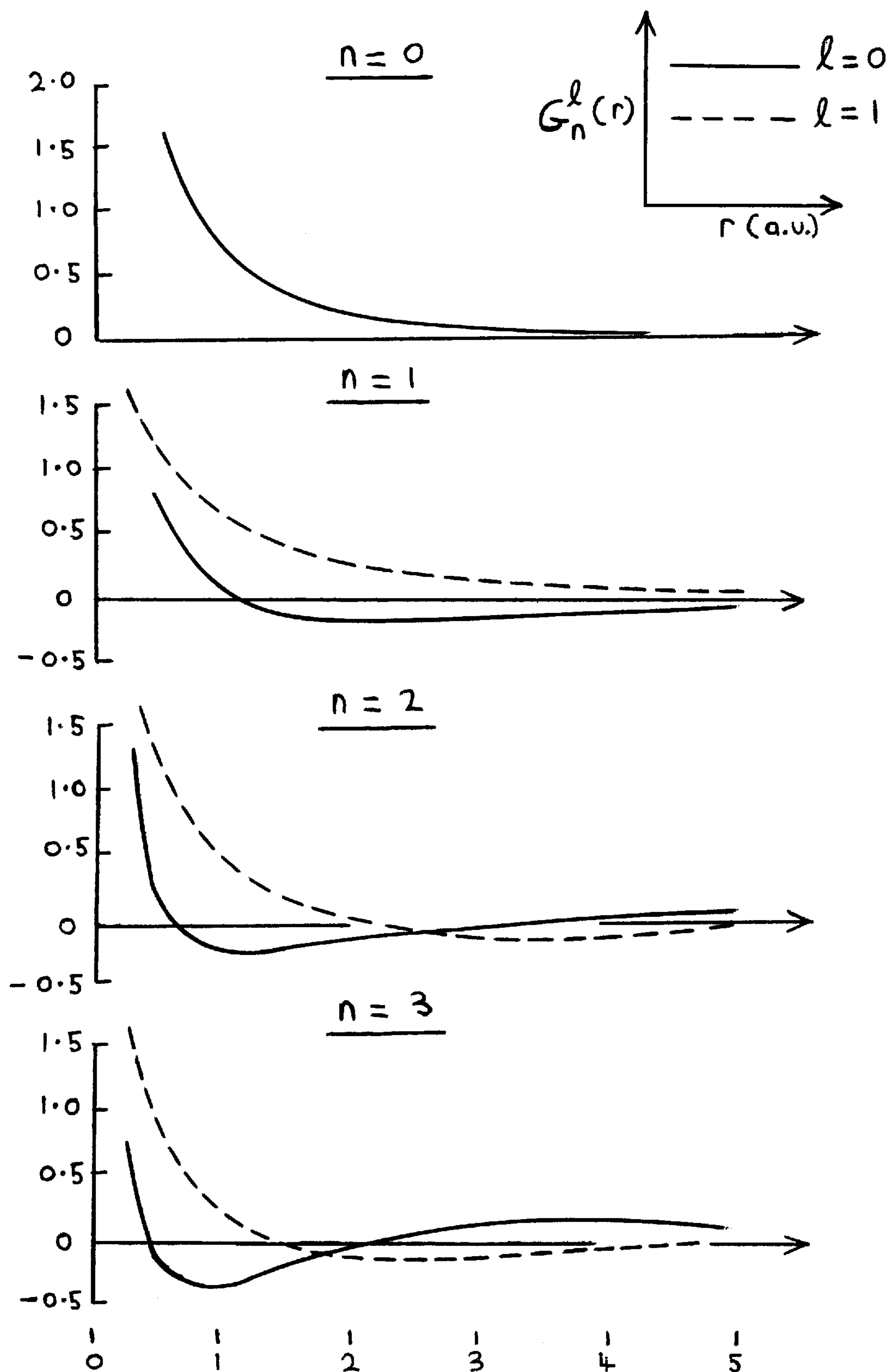
Powers of r

n	r^0	r^1	r^2	r^3	r^4	r^5	r^6	r^7	r^8	r^9
1	1.0									
2	1.4142	-0.70711								
3	1.7321	-1.7321	0.28868							
4	2.0	-3.0	1.0	-0.83333						
5	2.2361	-4.4721	2.2361	-3.7268	1.8634					
6	2.4495	-6.1237	4.0825	-10.206	10.206	-3.4021				
7	2.6458	-7.9373	6.6144	-22.048	33.072	-22.048	5.2495			
8	2.8284	-9.8995	9.8995	-41.248	82.496	-82.496	39.284	-7.0149		
9	3.0	-12.0	14.0	-70.0	175.0	-233.33	166.67	-59.524	8.2672	
10	3.1623	-14.230	18.974	-110.68	332.04	-553.40	527.05	-282.35	78.430	-8.7144
	1	1	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}

↑ Multiply coefficients by

FIGURE 3·4·1

Normalized Laguerre Functions



In practice we will use $n_{max} = 9$, $\ell = 0,1$ only. The spherical harmonics required are then:

$$
\begin{aligned}
Y_{0,0}(\theta,\phi) &= 1/(4\pi)^{\frac{1}{2}} \\
Y_{1,-1}(\theta,\phi) &= (3/8\pi)^{\frac{1}{2}}\sin\theta\, e^{-i\phi} \\
Y_{1,0}(\theta,\phi) &= (3/4\pi)^{\frac{1}{2}}\cos\theta \\
Y_{1,1}(\theta,\phi) &= -(3/8\pi)^{\frac{1}{2}}\sin\theta\, e^{i\phi}
\end{aligned}
\qquad (3.4.6)
$$

In addition it is useful to know that:

$$Y_{\ell,m}(\pi-\theta,\pi+\phi) = (-1)^{\ell}\, Y_{\ell,m}(\theta,\phi) \qquad (3.4.7)$$

$$Y^{*}_{\ell,m}(\theta,\phi) = (-1)^{m} Y_{\ell,-m}(\theta,\phi) \qquad (3.4.8)$$

$$\int_0^{\pi}\int_0^{2\pi} Y_{\ell,m}(\theta,\phi)Y_{\ell,-m}(\theta,\phi)\sin\theta\, d\theta\, d\phi = (-1)^{m} \qquad (3.4.9)$$

The appropriately normalized associated Laguerre polynomial coefficients for $\ell = 0$, $\ell = 1$ are listed in Tables 3.4.1, 3.4.2 respectively and a few of the resulting radial functions $G_{n,0}(r)$, $G_{n,1}(r)$ plotted in Figure 3.4.1.

Section 3.5 Simplification of the form of the determinant

We will now consider in some detail the form of the determinant (3.3.3) and show how the calculation of the matrix elements can be simplified. Three particular defect potential systems will be studied.

CASE 1 The defect potential is given by a sum of two individual spherically symmetric potentials at different sites i.e.

$$
\begin{aligned}
h(\underline{r}) &= h_1(|\underline{r}-\underline{R}_1|) + h_2(|\underline{r}-\underline{R}_2|) \\
&= h_1^a(|\underline{r}-\underline{R}_1|)h_1^b(|\underline{r}-\underline{R}|) \\
&\quad + h_2^a(|\underline{r}-\underline{R}_2|)h_2^b(|\underline{r}-\underline{R}_2|)
\end{aligned}
\qquad (3.5.1)
$$

Instead of the insertion of a single complete set of orthonormal functions we

insert two, one centred at each of the individual sites $\underline{R}_1, \underline{R}_2$, of the total defect potential. The quantity within the $\underline{k}$-space integral which we wish to calculate is given by

$$\frac{F^{*a}_{N,I}(n,\underline{k})F^{b}_{N',J}(n,\underline{k})}{E_{n,\underline{k}}-E} = \frac{1}{E_{n,\underline{k}}-E}\left[\int_{\substack{\text{all}\\\text{space}}} d\underline{r}\,\varphi^{*}_{n,\underline{k}}(\underline{r})h^{a}_{I}(|\underline{r}-\underline{R}_I|)g_N(\underline{r}-\underline{R}_I)\right]$$

$$\times\left[\int_{\substack{\text{all}\\\text{space}}} d\underline{r}\; g^{*}_{N'}(\underline{r}-\underline{R}_J)h^{b}_{J}(|\underline{r}-\underline{R}_J|)\;\varphi_{n,\underline{k}}(\underline{r})\right] \tag{3.5.2}$$

where I = 1 or 2, J = 1 or 2.

$$F^{*a}_{N,I}(n,\underline{k}) = \sum_{\underline{g}} a^{*}_{\underline{g}}(n,\underline{k}) \int d\underline{r}\; e^{-i(\underline{k}+\underline{g})\cdot\underline{r}}\, h^{a}_{I}(|\underline{r}-\underline{R}_I|)g_N(\underline{r}-\underline{R}_I)$$

$$= \sum_{\underline{g}} a^{*}_{\underline{g}}(n,\underline{k})e^{-i(\underline{k}+\underline{g})\cdot\underline{R}_I} \int d(\underline{r}-\underline{R}_I)e^{-i(\underline{k}+\underline{g})\cdot(\underline{r}-\underline{R}_I)}h^{a}_{I}(|\underline{r}-\underline{R}_I|)g_N(\underline{r}-\underline{R}_I)$$

$$= \sum_{\underline{g}} a^{*}_{\underline{g}}(n,\underline{k})e^{-i(\underline{k}+\underline{g})\cdot\underline{R}_I} \int_{\substack{\text{all}\\\text{space}}} d\underline{r}\; e^{-i(\underline{k}+\underline{g})\cdot\underline{r}}\, h^{a}_{I}(r)g_N(\underline{r}) \tag{3.5.3}$$

where $\underline{r}-\underline{R}_I$ has been relabelled as $\underline{r}$ for convenience. In order to calculate the integral we make use of the plane wave expansion

$$e^{i\underline{q}\cdot\underline{r}} = 4\pi\sum_{L,M} (i^L)\, j_L(qr)Y^{*}_{L,M}(\theta_{\underline{q}},\phi_{\underline{q}})Y_{L,M}(\theta,\phi) \tag{3.5.4}$$

$$\left[\; e^{-i\underline{q}\cdot\underline{r}} = 4\pi\sum_{L,M} (-i)^L j_L(qr)Y_{L,M}(\theta_{\underline{q}},\phi_{\underline{q}})Y^{*}_{L,M}(\theta,\phi)\;\right]$$

where $\theta_{\underline{q}}, \phi_{\underline{q}}$ represent the angular coordinates of $\underline{q}$ with respect to the coordinate system employed and the j_L are the spherical Bessel functions. Thus we obtain

$$F^{*a}_{N,I}(n,\underline{k}) = \sum_{\underline{g}} a^*_{\underline{g}}(n,\underline{k}) e^{-i\underline{q}.\underline{R}_I} 4\pi \sum_{L,M} (-i)^L Y_{L,M}(\theta_{\underline{q}}, \phi_{\underline{q}})$$

$$\times \int d\underline{r} \quad j_L(qr) h^a_I(r) G_{n,\ell}(r) Y^*_{L,M}(\theta, \phi) Y_{\ell,m}(\theta, \phi) \qquad (3.5.5)$$

where $\underline{q} = \underline{k} + \underline{g}$. Due to the orthogonality of the spherical harmonics only the term in which $L=\ell$ and $M=m$ is picked out of the plane wave expansion. We are left with

$$F^{*a}_{N,I}(n,\underline{k}) = 4\pi(-i)^{\ell} \sum_{\underline{g}} a^*_{\underline{g}}(n,\underline{k}) e^{-i\underline{q}.\underline{R}_I} Y_{\ell,m}(\theta_{\underline{q}}, \phi_{\underline{q}}) S^{I,a}_{n,\ell}(\underline{q}) \qquad (3.5.6)$$

where

$$S^{I,a}_{n,\ell}(q) = \int dr r^2 j_{\ell}(qr) h^a_I(r) G_{n,\ell}(r). \qquad (3.5.7)$$

Similarly we find that

$$F^b_{N',J}(n,\underline{k}) = 4\pi i^{\ell'} \sum_{\underline{g}'} a_{\underline{g}'}(n,\underline{k}) e^{i\underline{q}.\underline{R}_J} Y^*_{\ell',m'}(\theta_{\underline{q}'}, \phi_{\underline{q}'}) S^{J,b}_{n',\ell'}(q') \qquad (3.5.8)$$

We now demonstrate how the considerations of Section 2.6 can be utilized in order to simplify the problem.

In the conventional coordinate system the perfect crystal nearest neighbours are situated at $\pm$ A/8(1,1,1) with respect to the bandstructure origin. The position vectors $\underline{R}_1$, $\underline{R}_2$ are both directed along the (1,1,1) axis and hence the overall defect potential has C_{3v} symmetry. It is convenient to define a new coordinate system in which the z axis is along the (1,1,1) direction and this is the choice we shall make. Let us consider a particular $\underline{g}$, $\underline{g}'$ component, $C_{n,\underline{k}}(\underline{g},\underline{g}')$, of the product $F^{*a}_{N,I}(n,\underline{k}) F^b_{N',J}(n,\underline{k})$. We have

$$C_{n,\underline{k}}(\underline{g},\underline{g}') = 16\pi^2 (-i)^{\ell} i^{\ell'} a^*_{\underline{g}}(n,\underline{k}) e^{-i\underline{q}.\underline{R}_I} e^{i\underline{q}'.\underline{R}_J} a_{\underline{g}'}(n,\underline{k})$$

$$\times Y_{\ell,m}(\theta_{\underline{q}}, \phi_{\underline{q}}) Y^*_{\ell',m'}(\theta_{\underline{q}'}, \phi_{\underline{q}'}) S^{I,a}_{n,\ell}(q) S^{J,b}_{n',\ell'}(q')$$

$$= K(-i)^{\ell} i^{\ell'} a^*_{\underline{g}}(n,\underline{k}) a_{\underline{g}'}(n,\underline{k}) e^{-i\underline{q}.\underline{R}_I} e^{i\underline{q}'.\underline{R}_J} Y_{\ell,m}(\theta_{\underline{q}}, \phi_{\underline{q}}) Y^*_{\ell',m'}(\theta_{\underline{q}'}, \phi_{\underline{q}'})$$

(3.5.9)

where the terms which are unaltered by the symmetry operations have been combined in the constant, K. If we sum over the C_{3v} operations, β, we obtain

$$1/K \sum_{\beta} C_{n,\beta\underline{k}}(\underline{g},\underline{g}') = (-i)^{\ell} i^{\ell'} a^*_{\underline{g}}(n,\underline{k}) a_{\underline{g}'}(n,\underline{k}) \sum_{\beta} e^{-i\underline{q}.\beta^{-1}\underline{R}_I} e^{i\underline{q}'.\beta^{-1}\underline{R}_J}$$

$$\times Y_{\ell,m}(\theta_{\beta\underline{q}}, \phi_{\beta\underline{q}}) Y^*_{\ell',m'}(\theta_{\beta\underline{q}'}, \phi_{\beta\underline{q}'})$$

$$= (-i)^{\ell} i^{\ell'} a^*_{\underline{g}}(n,\underline{k}) a_{\underline{g}'}(n,\underline{k}) e^{-i\underline{q}.\underline{R}_I} e^{i\underline{q}'.\underline{R}_J} \sum_{\beta} Y_{\ell,m}(\theta_{\beta\underline{q}}, \phi_{\beta\underline{q}}) Y^*_{\ell',m'}(\theta_{\beta\underline{q}'}, \phi_{\beta\underline{q}'})$$

(3.5.10)

As $\underline{R}_I$, $\underline{R}_J$ are along the z axis they are unaffected by the C_{3v} operations, β^{-1}. Explicit insertion of the $Y_{\ell,m}(\theta, \phi)$ of Equation (3.4.6) shows that these spherical harmonics the summations in Equation (3.5.10) are in accordance with the Table 3.5.1 below.

TABLE 3.5.1

	$Y^*_{0,0}$	$Y^*_{1,0}$	$Y^*_{1,-1}$	$Y^*_{1,1}$
$Y_{0,0}$	NZ	NZ	0	0
$Y_{1,0}$	NZ	NZ	0	0
$Y_{1,-1}$	0	0	NZ	0
$Y_{1,1}$	0	0	0	NZ

This is true for all products $C_{n,\underline{k}}(\underline{g},\underline{g}')$ and hence the determinant (3.3.3) factorizes in the same block form. For the non-zero (NZ) elements we have

$$1/K \sum_{\beta} C_{n,\beta\underline{k}}(\underline{g},\underline{g}') = 6(-i)^{\ell}\, i^{\ell'}\; a^*_{\underline{g}}(n,\underline{k})a_{\underline{g}'}(n,k)e^{-i\underline{q}.\underline{R}_I}\; e^{i\underline{q}'.\underline{R}_J}$$

$$\times\, Y_{\ell,m}(\theta_{\underline{q}},\phi_{\underline{q}})Y^*_{\ell',m'}(\theta_{\underline{q}'},\phi_{\underline{q}'}) \qquad (3.5.11)$$

and hence

$$\sum_{\beta} \frac{F^{*a}_{N,I}(n,\beta\underline{k})F^{b}_{N',J}(n,\beta\underline{k})}{E_{n,\beta\underline{k}} - E} = \frac{6F^{*a}_{N,I}(n,\underline{k})F^{b}_{N',J}(n,\underline{k})}{E_{n,\underline{k}} - E}. \qquad (3.5.12)$$

The amount of labour required in order to calculate the elements of the determinant is reduced by a factor of six. Let us now consider the simplification which arises upon application of time-reversal symmetry. We have, as can easily be shown,

$$F^{*a}_{N,I}(n,-\underline{k}) = 4\pi i^{\ell} \sum_{\underline{g}} a_{\underline{g}}(n,\underline{k})e^{i\underline{q}.\underline{R}_I}\; Y_{\ell,m}(\theta_{\underline{q}},\phi_{\underline{q}})S^{I,a}_{n,\ell}(q) \qquad (3.5.13)$$

$$F^{b}_{N',J}(n,-\underline{k}) = 4\pi(-i)^{\ell'} \sum_{\underline{g}'} a^*_{\underline{g}'}(n,\underline{k})e^{-i\underline{q}.\underline{R}_J}\; Y^*_{\ell',m'}(\theta_{\underline{q}'},\phi_{\underline{q}'})S^{J,b}_{n',\ell'}(q') \qquad (3.5.14)$$

If $Y_{\ell,m}(\theta_{\underline{q}},\phi_{\underline{q}})$ and $Y_{\ell',m'}(\theta_{\underline{q}'},\phi_{\underline{q}'})$ are both real then it is immediately apparent that

$$F^{*a}_{N,I}(n,\underline{k})F^{b}_{N',J}(n,\underline{k}) + F^{*a}_{N,I}(n,-\underline{k})F^{b}_{N',J}(n,-\underline{k})$$

$$= 2\,\text{Real}\; F^{*a}_{N,I}(n,\underline{k})F^{b}_{N',J}(n,\underline{k}) \qquad (3.5.15)$$

Consequently, for the upper left-hand corner of Table 3.5.1 the length of calculation is reduced by a further factor of two. The final $\underline{k}$-space integrals are then given by

$$12 \text{ Real} \sum_{\gamma} \sum_{n} \int_{\substack{1/48 \\ BZ}} \frac{d\underline{k} F^{*a}_{N,I}(n,\gamma \underline{k}) F^{b}_{N',J}(n,\gamma \underline{k})}{E_{n,\underline{k}} - E} \tag{3.5.16}$$

where the γ represent the remaining four operations.

In the calculations we perform the sites I,J represent the nearest neighbour positions $\pm A/8(1,1,1)$.

CASE 2 This is closely related to CASE 1 as the overall Hamiltonian has the same, C_{3v}, symmetry. However, we eleminate one of the defect potentials giving a corresponding decrease in the order of the determinantal matrix. With the bandstructure centred on an atomic site the defect potential is

$$h(\underline{r}) = h(|\underline{r} - \underline{R}|) \tag{3.5.17}$$

where $\underline{R}$ is again along the (1,1,1) direction, the z axis.

CASE 3 represents a further simplification of the previous cases for we consider the situation in which we have one defect centre but in addition we put $\underline{R} = 0$. The defect potential and bandstructure are centred at an atomic site and the Hamiltonian has tetrahedral symmetry. It is a trivial matter to show that a decoupling of the $Y_{0,0}-Y_{1,0}$ products results upon application of the E,C_{2x}, C_{2y}, C_{2z} group of operations. The $Y_{0,0}$ determinant elements can then be calculated by using simply

$$48 \sum_{n} \int_{\substack{1/48 \\ BZ}} \frac{d\underline{k}\ F^{*a}_{N}(n,\underline{k}) F^{b}_{N'}(n,\underline{k})}{E_{n,\underline{k}} - E} \tag{3.5.18}$$

Section 3.6 Labelling the defect states

In Section 3.5 it was shown by practical example how the symmetry properties of the defect potential lead to a particular block factorization of the Fredholm determinant. When the potential has C_{3v} symmetry it was seen that the determinant partitions into a combined $Y_{0,0}-Y_{1,0}$ and two separate $Y_{1,-1}$ blocks. The latter give rise to two degenerate energy defect states. In analogy with the usual atomic/molecular terminology we will occasionally refer to these as corresponding to hybridized $s-p_z$ and p_x, p_y defect states. The results are not unexpected and indeed are exactly what we would anticipate on the basis of quite general group theoretical arguments. Group theory tells us that the wave functions must transform according to the irreducible representations of the particular defect Hamiltonian. In the case of a Hamiltonian with C_{3v} symmetry it is then usual for the states to be labelled a_1 $(s-p_z)$ and e (p_x, p_y). (It is customary to label 1-dimensional representations by the symbol a, 2-dimensional representations by e and 3-dimensional representations by t.) The a_1 state corresponds to the most symmetric solutions which remains unaltered upon application of the symmetry operations of the C_{3v} group i.e.

$$\{\beta \mid 0\}\, \psi_{a_1}(\underline{r}) = \psi_{a_1}(\underline{r}) \qquad (3.6.1)$$

The wave function has full C_{3v} symmetry. The doubly degenerate e states are individually transformed by the operations, β , but the resulting functions can always be expressed in terms of the two p_x, p_y basis functions.

In the case of a Hamiltonian with tetrahedral symmetry the defect wave functions are labelled $a_1(s)$ and $t_2(p_x, p_y, p_z)$. The a_1 state then has full tetrahedral symmetry and is unaltered by the 24 T_d point group operations. The triply degenerate p-states transform in accordance with the t_2 (three basis functions) representation.

In the chapters which follow we will employ both systems of terminology the usage depending upon which is most appropriate in the particular context.

Section 3.7 The formation of the real space defect potential

In Chapter 2 the idea of an atomic pseudopotential was introduced. The total crystal pseudopotential can be expanded in the form

$$V(\underline{r}) = \sum_i v_i(|\underline{r} - \underline{R}_i|) \qquad (3.7.1)$$

where the v_i and $\underline{R}_i$ represent the perfect crystal atomic pseudopotentials and positions respectively and i runs over all unit cells and atoms within the unit cell. v_i consists of the individual ionic pseudopotential together with a screening contribution due to the valence electrons. We now consider what happens when some form of defect is introduced into the system. For example, let us create a vacancy at $\underline{R}_j$. The crystal pseudopotential is then

$$V'(\underline{r}) = \sum_{i\neq j} v_i' (\underline{r} - \underline{R}_i') \qquad (3.7.2)$$

where we have excluded the atomic pseudopotential at $\underline{R}_j$ and have denoted the new relaxed equilibrium positions and pseudopotentials by $\underline{R}_i'$ and v_i' respectively. (The idea that we can deal with the defective crystal in this manner, within the pseudopotential formulation and incorporating the use of linear screening, is lengthily dealt with by Ross.[27] We shall not re-cover this ground and will assume that the method is indeed applicable.) We then have a defect potential given by

$$h(\underline{r}) = \sum_{i\neq j} v_i'(\underline{r} - \underline{R}_i') - \sum_i v_i(|\underline{r} - \underline{R}_i|) \qquad (3.7.3)$$

As it stands we would be totally unable to deal with a defect potential of this form and simplifying assumptions must be made. They are

1) The individual spherically symmetric atomic pseudopotentials are not deformed by the creation of the defect and subsequent relaxation i.e.

$$v_i' = v_i \qquad (3.7.4)$$

We will make further use of this non-deformation in Chapter 5.

2) We then further assume that the equilibrium positions of the pseudoatoms remain unchanged by creation of the defect. (An approximate method of dealing with nearest neighbour relaxation can be employed later if necessary.) This assumption is made not because in principle our method is incapable of dealing with a multiple defect centre problem-we have shown otherwise in Section 3.5 (an alternative technique using two-centre integrals could also be considered) - but because the computer time required to do so would be inordinate.

As a consequence of these assumptions the vacancy defect potential is given by

$$h(\underline{r}) = -v_j(|\underline{r} - \underline{R}_j|) \qquad (3.7.5)$$

which is the simple form which we require.

The real space vacancy potential can be found by fitting to pseudopotential coefficents, $v(q)$, such as those compiled by Cohen and Heine.[28] As these $v(q)$ include screening we have then created a vacancy by removal of what is in effect a self-consistently screened neutral pseudoatom. It is interesting to note that despite differences in screening function, $\epsilon(q)$, for different materials the $v(q)$ for a particular atomic species are little changed. This lends credence to assumption 1).

In the case of a substitutional impurity Equation (3.7.5) becomes

$$h(\underline{r}) = v_{sub}(|\underline{r} - \underline{R}_j|) - v_j(|\underline{r} - \underline{R}_j|) \qquad (3.7.6)$$

In practice the determination of the substitutional impurity potential is more involved for we should first calculate

$$h(q) = \frac{v_{sub}^{ion}(q) - v_{host}^{ion}(q)}{\epsilon(q)} \tag{3.7.7}$$

and then fit to a real space potential. (See Ross[27] for details of how this can be done using model potential parameters together with an appropriate dielectric function for the host material.)

In order that the defect potential may be satisfactorily incorporated into the defect scheme of the preceding sections we choose to fit to the factorizable form

$$h = \pm (A^2r^2-B^2)e^{-\alpha r} \tag{3.7.8}$$

$$h^a h^b = (\pm (Ar + B)e^{-\alpha r/2}) \times ((Ar-B)e^{-\alpha r/2}) \tag{3.7.9}$$

It is appreciated that the method outlined for forming the real space defect potential leads to a function which is neither unique nor error free. As a consequence it is one of the primary sources of inaccuracies in our calculations. Further discussion of the significance of these errors will be presented in the following chapters.

Section 3.8 Some practical aspects of the calculations

In the previous sections a description of the main theoretical details relevant to the method of dealing with the defect problem has been given. Some of the more practical aspects will now be mentioned. To begin with, we need to calculate the real space integrals, S. Inserting the explicit forms for h, G and the spherical Bessel functions, j_L, given by

$$j_L(x) = x^L \left(-1/x \frac{d}{dx}\right)^L \left(\frac{\sin x}{x}\right) \tag{3.8.1}$$

and in particular

$$J_0(x) = \frac{\sin x}{x} \tag{3.8.2}$$

TABLE 3.8.1

k-space sampling points

units 2 π /A

kx	ky	kz
0.14	0.09	0.04
0.26	0.09	0.04
0.50	0.09	0.04
0.74	0.09	0.04
0.94	0.09	0.04
0.38	0.09	0.04
0.86	0.09	0.04
0.62	0.09	0.04
0.26	0.21	0.04
0.38	0.21	0.04
0.50	0.21	0.04
0.62	0.21	0.04
0.74	0.21	0.04
0.89	0.21	0.04
0.94	0.21	0.04
0.38	0.33	0.04
0.50	0.33	0.04
0.62	0.33	0.04
0.74	0.33	0.04
0.86	0.33	0.04
0.50	0.45	0.04
0.62	0.45	0.04
0.74	0.45	0.04
0.62	0.57	0.04
0.26	0.21	0.15
0.38	0.21	0.15
0.50	0.21	0.15
0.62	0.21	0.15
0.74	0.21	0.15
0.86	0.21	0.15
0.38	0.33	0.15
0.50	0.33	0.15
0.62	0.33	0.15
0.74	0.33	0.15
0.50	0.45	0.15
0.62	0.45	0.15
0.38	0.33	0.27
0.50	0.33	0.27
0.62	0.33	0.27
0.50	0.45	0.27
0.50	0.45	0.39

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \qquad (3.8.3)$$

it can be seen that only component integrals of form

$$\int_0^\infty dr \; e^{-ar} \, r^y \begin{pmatrix} \sin \\ \cos \end{pmatrix} (qr) \qquad (3.8.4)$$

result (y is integral or half integral). These are standard integrals and so there is no great difficulty encountered in the calculation of the S by the formation of the appropriate combinations. In calculating the $\underline{k}$-space integrals it is, however, necessary to resort to approximate means because the Bloch functions, $\varphi_{n,\underline{k}}$, are known only at chosen points and not as a continuous function of $n,\underline{k}$. We use a point sampling technique and take

$$\sum_n \int \frac{d\underline{k} \; F_N^{*a}(n,\underline{k}) F_{N'}^{b}(n,\underline{k})}{E_{n,\underline{k}} - E} = \frac{\Omega}{N_s} \sum_{i=1}^{N_s} \sum_n \frac{F_N^{*a}(n,\underline{k}_i) F_{N'}^{b}(n,\underline{k}_i)}{E_{n,\underline{k}_i} - E} \qquad (3.8.5)$$

where Ω is the Brillouin zone volume and N_s is the number of sampling points. In practice 41 general sampling points within the irreducible 1/48 BZ volume will be employed. These points (Table 3.8.1) form a grid within the BZ section Γ - LUXWKL illustrated in Figure (3.8.1) below.

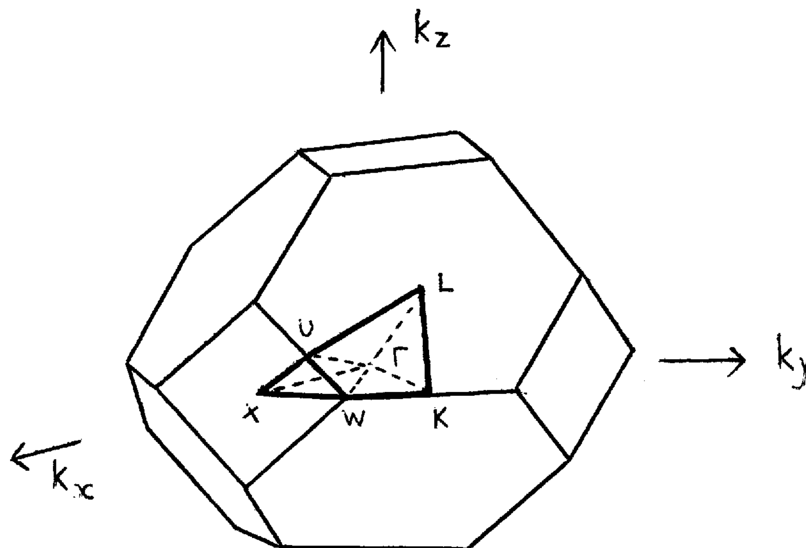


FIGURE 3.8.1

We sum over 10 bands and so the numerical integration involves a total of 19,680 points. In expanding the Bloch functions as

$$\varphi_{n,\underline{k}_i}(\underline{r}) = \sum_{\underline{g}} a_{\underline{g}}(n,\underline{k}_i)e^{i(\underline{k}_i + \underline{g}).\underline{r}} \tag{3.8.6}$$

the first 65 plane wave coefficients, $a_{\underline{g}}$, associated with type (0,0,0), (1,1,1), (2,0,0), (2,2,0), (3,1,1), (2,2,2), and (4,0,0) reciprocal lattice vectors are taken into account. The errors arising from the plane wave truncation and sampling technique lead to uncertainties of about 0.1 eV in the final results (see the works of Vinsome[29] and Ross[27] for details of convergence of energy bands and sampling). Although this is a non-negligible error it is in line with the magnitude of other errors which may be present and consequently there is no great advantage to be gained in reducing it at the expense of much increased computational effort.

Section 3.9 Summary

In this chapter a method of dealing with the localized defect problem has been outlined. The introduction of suitable sets of localized functions and the restriction of attention to the forbidden gap region has been suggested. A real space defect potential of suitable form is constructed and then a point sampling technique is used to carry out the calculations for a number of trial defect energy values. The actual defect levels, E_d, can then be located if they exist within or near the band gap.

During the course of the following chapters we shall embark upon the application of the method to a number of particular defect problems. The method will be assessed in the light of experimental information and a comparison with other theoretical approaches will be made.

CHAPTER 4

VACANCIES IN SEMICONDUCTORS

Vacancies in semiconductors

Section 4.1 Introduction

In the previous chapters an attempt has been made to develop a reasonable theoretical foundation from which we may build in order to study some aspects of localized defect problems. In this chapter some of the results obtained when the method is applied to the problem of vacancies and vacancy complexes in semiconductors, and in particular to the III-V compound gallium arsenide, will be presented. Previously most of the theoretical studies have been confined to the elemental materials silicon and diamond. It is for this reason that we choose to investigate alternative materials in the hope that useful insight of these less frequently considered compounds will be gained. Before doing so, however, a historically oriented review will be given of some of the significant theoretical and experimental studies of vacancies in covalently bonded materials.

Section 4.2 General review of vacancy studies

By the beginning of the nineteen fifties a number of experiments had been performed on irradiated germanium and silicon and it was apparent that some form of theoretical model was required in order to account for the observed results. The defects produced were clearly of an intrinsic-related nature and thus James and Lark-Horovitz[30] tried to explain the p-type compensation of n-type germanium by considering the possible consequences of the introduction of isolated vacancies and interstitials. Using simple, qualitative arguments derived from the hydrogenic model they attempted to deduce the approximate positions of the resulting defect electronic energy levels and then to relate them to the alteration in conductivity. Some years later Blount[31] proposed an alternative model with a hybridized sp^3 tight-binding scheme as a starting point. The ordering of the energy levels implied different charge states for the defects in the two approaches. Neither model

was adequate for they did not really predict the actual energy levels but rather were designed to position these levels with respect to the band edges in order to account for observed results. In addition the models were not suitable for dealing with more general situations. The authors of both formulations were, of course, aware of these limitations and also were careful to point out the possibility of more complex defect systems which could cloud the issue. In the light of later knowledge this caution was well justified.

The studies of Coulson and Kearsley[3] and Yamaguchi[32] were probably the first to theoretically consider the valence crystal vacancy problem in a really quantitative manner. Using molecular orbital techniques they attempted to directly compare the total electronic energies for different total wave function configurations in the case of a vacancy in diamond. By this means Coulson and Kearsley (CK) hoped to be able to explain the absorption spectrum of diamond and in particular the 2 eV GR1 band. Being a material of low polarizability the influence of the bulk of the crystal upon the local defect characteristics was considered to be minimal and this made diamond a particularly suitable subject for such studies. This property was one of the factors leading to the further considerable interest in this defect problem. In their investigation it was CK who put forward the concept of the "defect molecule" which, as we shall see, has continued to be frequently employed. They formed one-electron wave functions transforming as the irreducible representations of the T_d point group from the sp^3 "dangling bonds" on the four nearest neighbour atoms to the vacancy. The influence of the remainder of the crystal was largely, but not entirely, neglected. They then formed appropriate many-electron wave functions and calculated the associated energies. They were able to put forward a number of proposals including the suggestion that the neutral vacancy was the most energetically favourable configuration. CK were aware of the possible consequences of the Jahn-Teller effect[33,34] (the removal of electronic

degeneracy in response to distortion of the system) although they did not incorporate this into their calculations. This effect has since been a topic of much discussion. Yamaguchi used a similar approach to that of CK but came to the conclusion that the singly negative charged vacancy was the more energetically stable state.

Probably the most significant contribution to our knowledge of vacancy-related defects in silicon is due to the important experimental work of Corbett and Watkins and co-workers. Consequently we shall consider the findings of this work in some detail. In a series of reports beginning in 1959 they recorded the indentification of a number of defects in 1.5 MeV electron irradiated silicon, primarily by use of the electron spin resonance technique but also in conjunction with ENDOR, infra-red absorption, illumination and strain response experiments. The actual positions of the energy levels cannot be revealed by EPR work and so the complete determination of the defect characteristics must be inferred by association with the results of complementary procedures. (This process is not always simple, however, and there are still considerable information gaps.) The existence of two particularly interesting EPR spectra (later to be denoted Si-B1, Si-G8) due to the so-called A and E centres was apparent. These were initially tentatively suggested as comprising of a negatively charged vacancy-oxygen interstitial pair and a neutral vacancy-phosphorus pair respectively.(35) Subsequent work(36,37,38) confirmed these suppositions. In general, the experiments were carried out well below room temperature and so it was evident, contrary to previous belief, that the isolated vacancy was mobile at low temperatures. The positively charged isolated vacancy was detected in the form of an EPR spectrum (Si-G1) in p-type silicon below 20°K.(39) The results indicated that no nearby silicon interstitial was present and so it appeared that the interstitial was an exceedingly mobile species even at very low temperatures. Later, the analysis of another spectrum (Si-G2) in n-type

material led to the identification of the negatively charged vacancy.[40] In addition the doubly negatively charged vacancy, which is unobservable by EPR, was inferred from the generation by light of the Si-G2 spectrum. An initial report[41] also indicated the existence of several charge states for the divacancy and this was subsequently confirmed.[42] The EPR spectra Si-G6, Si-G7 were associated with the singly positively and singly negatively charged states respectively. In their defect analyses Corbett and Watkins found their results to be consistent with a simple Jahn-Teller distorted one-electron model in which only the nearest neighbour atomic orbitals need be included. This seemed to be an adequate sized defect molecule for it appeared that approximately 60-70% of the wave function was localized within this region. For the various charge states of the isolated vacancy there are two alternative Jahn-Teller distortion routes which could be postulated and these are shown in Figure 4.2.1

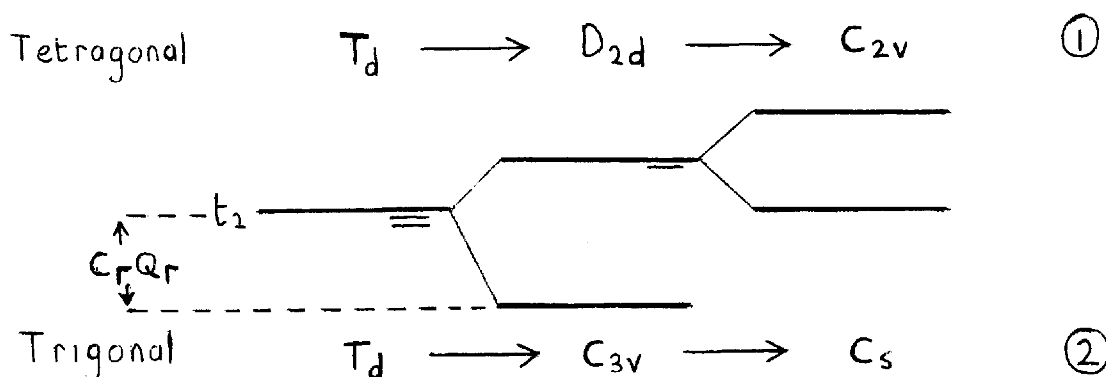


FIGURE 4.2.1

The first of these was found to be in agreement with the experimental results. A small activation energy associated with the rate of reorientation following stress was initially thought to be a direct measure of the Jahn-Teller stabilization energy. If this had indeed been so then a quite modest stabilization energy would have been implied. However, much to the detriment of simple theoretical analysis, this line of reasoning was rejected and when

Elkin and Watkins came to recalculate this quantity a disconcertingly large value of the order of 1 eV was obtained.[43] In addition there is also the complication due to possible symmetric relaxation effects to be considered. It is usual to assume a linear relationship for the lowering of the electronic energy in response to a Jahn-Teller distortion amplitude, Q_Γ. The total energy relative to the undistorted state is then

$$E = -C_\Gamma Q_\Gamma + 1/2\, K_\Gamma Q_\Gamma^2 \qquad (4.2.1)$$

where C_Γ is the Jahn-Teller coupling coefficient and K_Γ is an effective force constant for a particular distortion mode, Γ. $C_\Gamma Q_\Gamma$ might then, for example, represent the magnitude of the one-electron energy reduction when one electron is placed in the degenerate t_2 level of Figure 4.2.1. and subsequent distortion takes place. In order to calculate the minimum energy,

$$E_{JT} = -C_\Gamma^2 / (2K_\Gamma) \qquad (4.2.2)$$

it is first necessary to obtain values for C_Γ and K_Γ. A number of calculations giving values for the Jahn-Teller stabilization energy have been made such as those of Larkins for diamond[44] (but see also reference (45)) and silicon[46] vacancies. Values for silicon are also quoted by Watkins[47] and Stoneham.[48] (These references also contain useful reviews of the vacancy situation.) There is, however, much scope for error in these calculations and agreement upon the dominant mode of distortion is not always obtained. For the moment we shall take leave from such complications and will return to the progress of basic theoretical calculations of vacancy energy levels.

The Koster-Slater (K-S) Wannier function approach[5,6] was suggested as early as 1954 but the tremendous computational effort required in performing such a calculation precluded any realistic application of the method to the semiconductor defect problem for a number of years. Then in 1967, working

within the pseudopotential formulation, Callaway and Hughes published results for the silicon vacancy which demonstrated that the method was indeed feasible.[7] Later the silicon divacancy was also considered.[49] Although the achievement of Callaway and Hughes was creditable the final results were not completely satisfactory because bound states within the band gap could only be produced by an arbitrary scaling of the vacancy pseudopotential (in the case of the single vacancy a factor of 1.6 was required). In addition, the results showed an alarming sensitivity to scaling (scaling by a further 25% pushed the level right through the gap) and no definite conclusion concerning convergence was reached. Despite this the method has certain advantages for in contrast to the molecular orbital methods the energy levels obtained are well defined with respect to the band edges. The K-S technique has since been applied to the problem of vacancies in the IV-VI compound PbTe by Parada and Pratt[50,51] and the results are in broad agreement with experiment. The non-freeze-out of conductivity and resulting n and p-type properties were attributed to the existence of tellurium and lead vacancies respectively.

A number of calculations have continued to be based upon the original C-K many-electron molecular orbital method. The works of Coulson and Larkins on the diamond vacancy[52] and divacancy[53] and the additional work by Larkins on the isolated silicon[45,46] and diamond[45] vacancies are probably representative. It must be stressed that due to the difficulty of relating the results of this type of calculation to the band edges their main purpose lies in the prediction of the ordering and symmetry of the ground and excited states and the difference in their energies. Attempts are then made to explain optical asborption spectra in terms of transitions between the localized states.

In addition to the K-S and CK approaches a number of alternative methods of dealing with localized defects have been proposed. Two of these, the one-electron extended Huckel theory (EHT)[54] and $X\alpha$ scattered wave methods,[55] have

attempted to bridge the gap between the K-S and CK approaches. Both consider clusters of atoms much larger than the number used in the CK method and in doing so hope to give a reasonable description of the bulk material as well as the properties of the local defect. The measure of success in describing the bulk material is usually judged in terms of the ability to represent the band gap in an adequate manner. Messmer and Watkins, who first applied the EHT method to the problem of the sustitutional nitrogen impurity in diamond,[56] also investigated the diamond vacancy and found a localized state with energy close to the valence band edge (see, for example references (4) and (57)). Lee and McGill applied the EHT technique to the calculation of the electronic energy levels of the silicon divacancy and found two localized states with energies in the gap region.[58] They modelled a distortion mode consistent with the divacancy interpretation of Watkins and Corbett[42] and reached qualitative agreement with experiment. The levels were associated with the charge states leading to the Si-G6 and Si-G7 electron paramagnetic resonance spectra. In using the self-consistent $X\alpha$ scattered wave method (which involves the first principles numerical solution of the Schrödinger equation) Hemstreet has considered vacancies in the IV-VI compounds PbS and PbTe[59,60] and has obtained results in general agreement with the K-S calculations of Parada and Pratt although some of the details differ. We note that self-consistent calculations based on alternative schemes have been performed by Louie et al.[61] for silicon and by Surratt and Goddard[62] for both diamond and silicon vacancies.

We conclude this section by saying that the given list of vacancy calculations is by no means exhaustive but a representative sample has been included. Results specific to calculations on III-V compounds will be referred to in the following section. Brief outlines of mathematical details of some of the theoretical methods which have been mentioned can be found in Chapter 6 where comparison with the method of Chapter 3 is made.

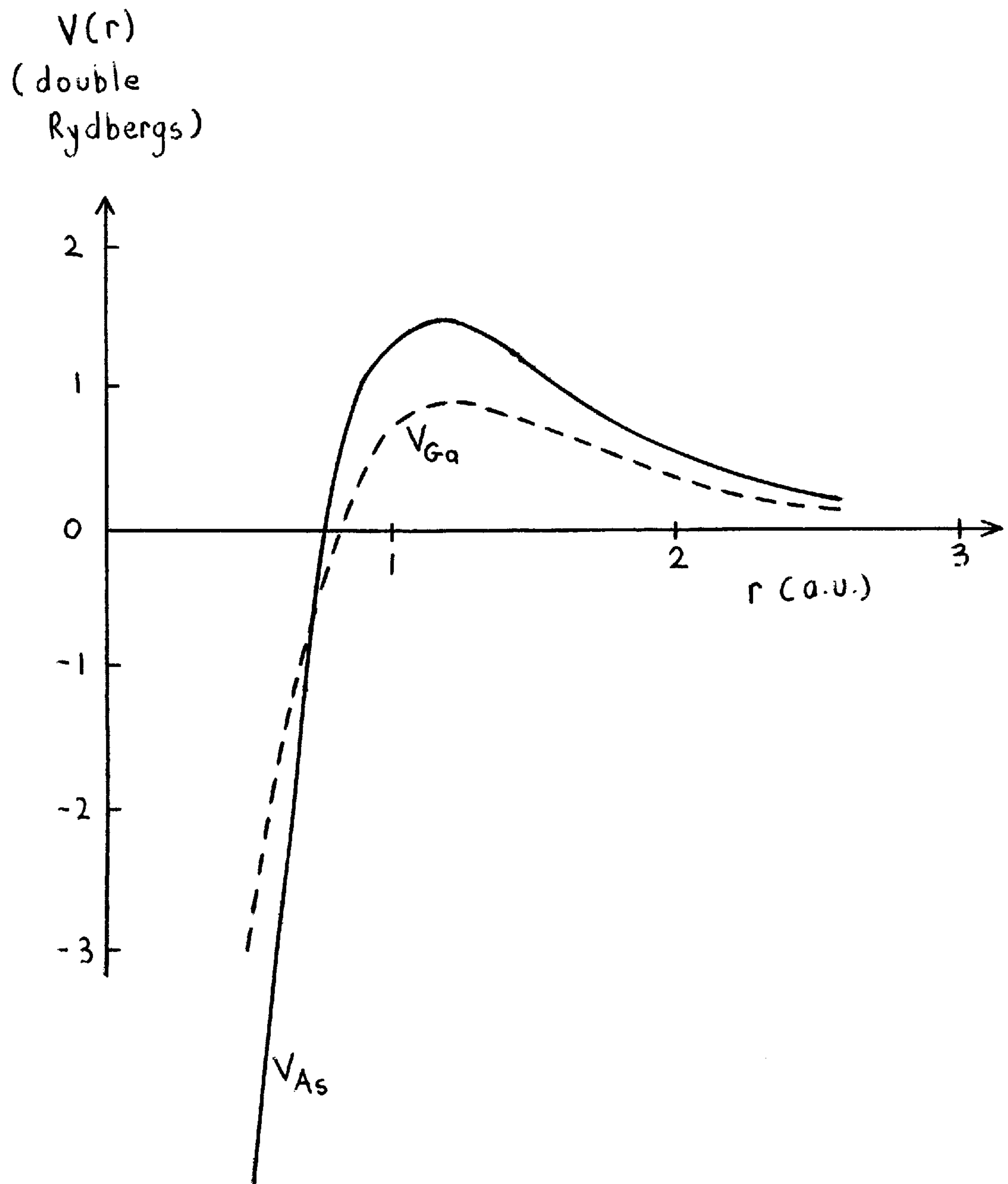
Section 4.3 Vacancy-related defects in gallium arsenide-theory

In the previous chapter a theoretical scheme which we believe to be suitable for dealing with the problem of calculating the electronic energy levels associated with localized defects was developed. The results obtained when the method is applied to the problems of single vacancies, the divacancy and the vacancy-oxygen pair in GaAs will be described. A number of general conclusions which may be useful within the overall context of vacancies in similar materials are put forward. This presentation is based upon the initial short account by Jaros[8] and the later expansion by Jaros and Brand.[63]

A natural starting point is the study of isolated gallium and arsenic vacancies in the unrelaxed GaAs lattice. We recognize that this may not well represent the real physical situation for some form of local distortion is likely to occur but we must be able to deal with this basic problem before any complications are introduced. These complications will be considered later. Applying the methods already described we obtain self-consistently screened vacancy pseudopotentials. These are illustrated in Figure 4.3.1. (At this stage the validity of these potentials could be questioned for it may be argued that the introduction of the vacancy will be such as to change the local screening properties of the system. It will be seen later why such effects are not thought to be greatly significant.) Using these potentials we try to locate s and p-type (a_1 and t_2) electronic states by searching for energies within the forbidden gap region which give zeros for the Fredholm determinant (3.3.3). In practice only one out of four states with possible energies within this region is detected. This corresponds to a threefold degenerate t_2 level associated with the gallium vacancy approximately 0.2 eV above the valence band edge. Although our method is, strictly speaking, only valid within the forbidden gap region we can infer the existence of resonance levels near the band edges if these can be "pulled" into the gap region by slightly scaling the potential. When this

FIGURE 4·3·1

Real space vacancy pseudopotentials



scaling procedure is carried out it is found that the additional three states are detectable. Both gallium and arsenic a_1 levels appear to be about 0.1 eV below the valence band edges whereas the arsenic t_2 level is about the same amount above the conduction band edge. In the light of the inherent errors in the calculations (sampling, band structure inaccuracies and potential strength) it is not possible to say for certain that these one-electron levels lie outside the gap region. However, if any of them do lie within this region then the a_1 states are the most likely candidates for, as we shall see when the results for the defect pairs are presented, the higher symmetry states are more sensitive to scaling and thus the error is likely to be greater.

In comparison with the location of the energy eigenvalues, E_d, the calculation of the associated wave functions is a much less straightforward task. Unfortunately, the pseudopotential factorization (3.3.10) which would lead to very easily obtainable wave functions fails to produce a convergent eigenvalue. When the factorization (3.7.9) is employed (this factorization is used throughout the present work) a perfectly satisfactory energy convergence is achieved (see for example Figs. 4.3.5 and 4.3.7). However, the A_N obtained from the eigenvector of (3.3.3) do not converge to any significant degree. As a consequence it is not possible to find a suitable unique wave function in terms of the g_m that we use. It may be that an alternative basis set would give adequate convergence of both energy and wave function but this is not known at present. The non-convergence of wave function does not imply a failure of our overall method but only of this particular means of obtaining it. If the defect wave function is obtained via the Bloch function expansion (3.2.3) then this convergence problem is not encountered. For the present purposes a wave function of this form is not very convenient for our only usage of it will be in the calculation of the electron-electron interaction and a simple analytic real space function is

much more useful in this context. Such a function can be obtained by fitting to the results of the Bloch function expansion or by various trial function methods. For example, if we have an analytic trial function, ψ_t, which is a good representation of the Bloch function expansion, ψ, then the Schrödinger equation can be written in the form

$$\left(H_0 + \frac{h|\psi_t\rangle\langle\psi_t|h)}{\langle\psi_t|h|\psi_t\rangle}\right)|\psi\rangle = E_d|\psi\rangle \qquad (4.3.1)$$

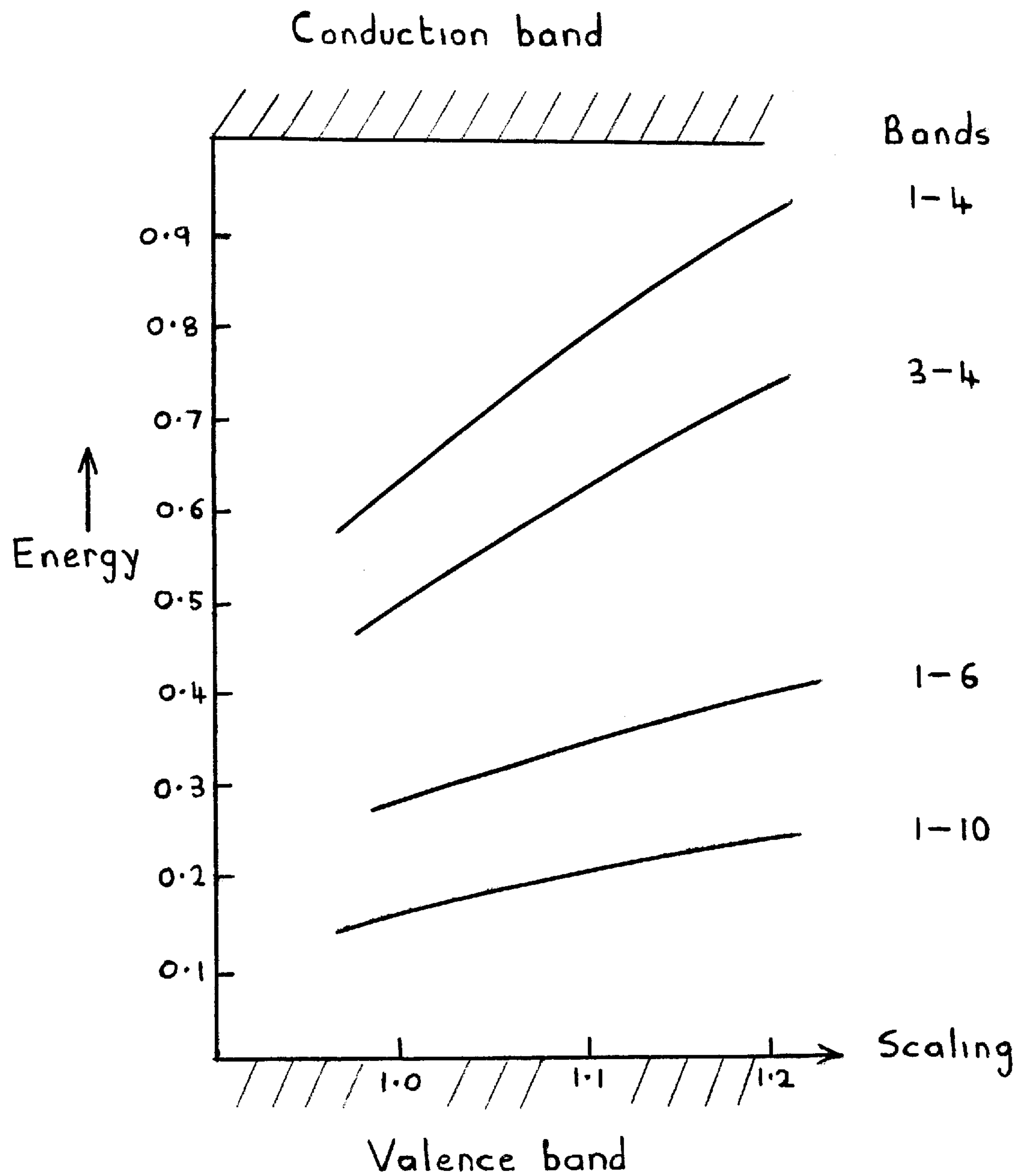
By suitable manipulation[64] we can obtain the consistency relation

$$1 + \frac{1}{\langle\psi_t|h|\psi_t\rangle}\sum_{n,\underline{k}}\frac{|\langle\varphi_{n,\underline{k}}|h|\psi_t\rangle|^2}{E_{n,\underline{k}} - E_d} = 0 \qquad (4.3.2)$$

which may be satisfied by adjusting the parameters of ψ_t accordingly. Alternative trial function methods such as that described by Jaros[65] may also be possible. All of these methods give broadly similar results but none can be considered as ideal. The s-type wave functions tend to be highly localized ($r^2 |\psi|^2$ having a maximum within a few atomic units) and the p-type rather less so. Because of the less reliable nature of the wave functions little will be said of their specific details. However, there is still useful general information concerning the construction of the defect states which can be obtained by other means. For example, consider Figure 4.3.2 in which the significance of various bands in the determination of the energy of the gallium vacancy t_2 state is displayed. It can be seen that ten bands are necessary in order to obtain reasonable convergence. This immediately makes us suspicious of arguments that are based upon a truncated band expansion. It is apparent that both the upper valence and lower conduction bands are particularly significant in determining

FIGURE 4·3·2

Gallium vacancy t_2-state energy as a function of bands and scaling



the final energy of the vacancy state. Not only are the conduction bands important with respect to the value of the energy, they also have the effect of greatly reducing the sensitivity to scaling of the final result. This lack of sensitivity to scaling leads us to believe that any initial error in the strength of the potential will not be of crucial importance. However, it is the upper valence bands which really dominate the state for it is these bands which ensure that the vacancy level lies in the gap region to begin with. The conduction bands, although important in determining the precise position of the level, are essentially a modifying influence.

On the basis of the results for the gallium vacancy, V_{Ga}, in GaAs we can attempt to anticipate the approximate positions of the t_2 level in other III-V compounds. In GaP the valence bands, which will again be expected to dominate the state, are very similar to those of GaAs. The V_{Ga} energy would thus be expected to be similar, quite close to the valence band edge. The larger band gap would be expected to result in a smaller conduction band contribution and hence a larger t_2-state energy but the difference in shape of the conduction bands is a complicating factor. A full calculation gives a level at about 0.1 eV from the valence band edge. This is a little less than might be anticipated but the overall position within the gap is still what we would expect. The valence bands of InSb are not as wide as those of GaAs but are of similar shape so we would expect the valence band contribution to lead to a higher V_{In} t_2 level. On the other hand the smaller band gap should lead to a compensating reduction in energy in response to the increased conduction band contribution. A full calculation yields a level in the middle of the gap at about 0.25 eV from the valence band edge (the large spin-orbit splittings means that the InSb band gap is not well represented in the CB scheme). Although not particularly startling the overall results are in line with what we would expect.

The arsenic vacancy potential, as can be seen from Figure 4.3.1 is significantly stronger than that of V_{Ga} and this explains why the corresponding t_2 level is pushed into the conduction band region. General arguments concerning the group V vacancies are less likely to be reliable but we would expect to find the t_2 levels in the region of the conduction band edge as was the case for V_{As}. Calculations of the position of the t_2 level associated with V_P in GaP give support to this conclusion for we locate this level about 0.4 eV into the conduction band.

In addition to modelling single vacancies or substitutional impurities we are also able to deal with simple complexes such as divacancies or vacancy-impurity pairs. We shall begin by considering the V_{Ga}-O_{As} pair. As has already been shown the isolated gallium vacancy gives rise to a triply degenerate one-electron level within the forbidden gap region. The substitutional oxygen impurity at the arsenic site produces an s-type state with energy in the middle of the gap. Any p-type state which may result has an energy well into the conduction band region and is thus not of interest. The oxygen impurity potential is shown in Figure 4.3.3. A similar strength potential representing O_P in GaP gives an energy in good agreement with experiment.(66) Figure 4.3.4 shows the energy levels associated with V_{Ga}-O_{As} which appear within the band gap.

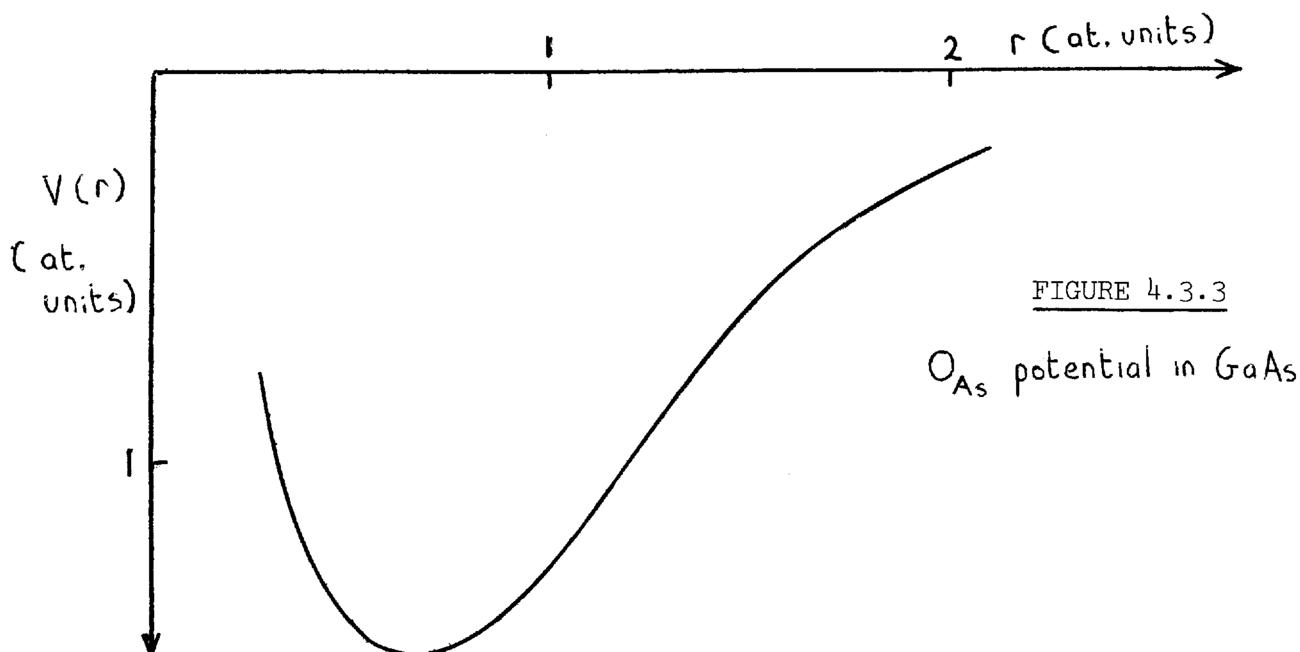
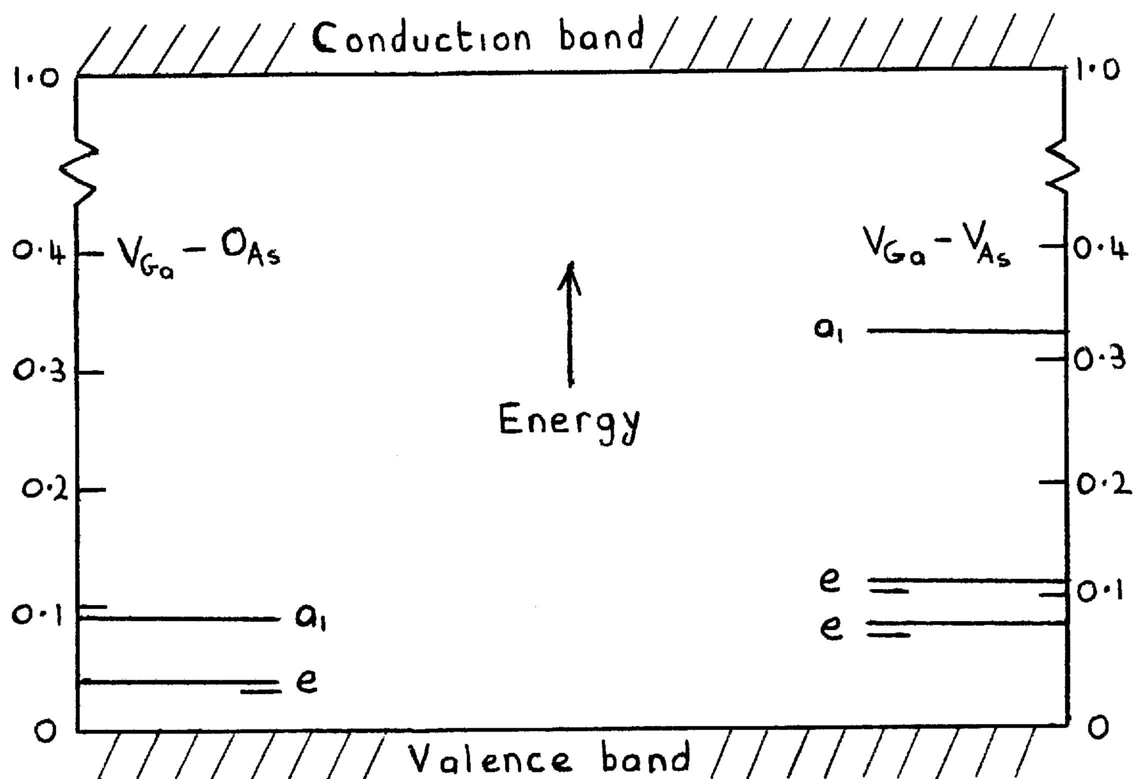


FIGURE 4.3.3

O_{As} potential in GaAs

FIGURE 4·3·4

Energy levels associated with the divacancy and vacancy-oxygen pair in GaAs



Also shown are the results for the divacancy. The pairing of the attractive oxygen impurity potential with the gallium vacancy leads to lowering and splitting of the $V_{Ga}t_2$ level to produce a non-degenerate a_1 level and a doubly degenerate e level. The V_{Ga} potential has the effect of pushing the oxygen impurity a_1 level up into the conduction band region. The interaction between the two vacancies in the V_{Ga}-V_{As} pair again leads to splitting and shifts of the t_2 levels. A precise analysis of the positions of the levels is difficult due to the dual attractive/repulsive nature of the vacancy potentials, uncertainties in the initial wave functions and the exact role of the host crystal.

In Figure 4.3.5 the convergence of the defect pair singlet state energies as a function of n(see Section 3.4) is shown. It can be seen that good convergence is achieved for a value of $n = 9$. The sensitivity to scaling of the same two levels is shown in Figure 4.3.6. Fortunately, the energy is much less sensitive to scaling than the early results of Callaway and Hughes for silicon. Even for the worst of the two curves a 25% increase in potential strength only causes approximately 0.25 eV change in energy. We can be assured that an initial small error in potential strength will not prove too disastrous. Figures 4.3.7 and 4.3.8 show the corresponding results for the doubly degenerate levels. Again there is good convergence as a function of n. The e-state energy is even less sensitive to changes in strength of the potential (note the change in scale). It seems that the role of the host crystal in determining the nature of the state is much more important in the case of the lower symmetry, more delocalized states. Both the convergence with n and sensitivity to potential strength of the defect pair energy levels are similar to the results for O_P in gallium phosphide and single vacancies. It now seems well established that the method of Chapter 3 is capable of giving truly convergent one-electron energy levels for a variety of localized defect potentials. Also it is clear that the inherent difficulties of producing a good real space defect potential do not

FIGURE 4·3·5

Energy of divacancy and vacancy-oxygen pair singlet states as a function of n (n=9 implies a total of 38 Laguerre functions, 19 per site)

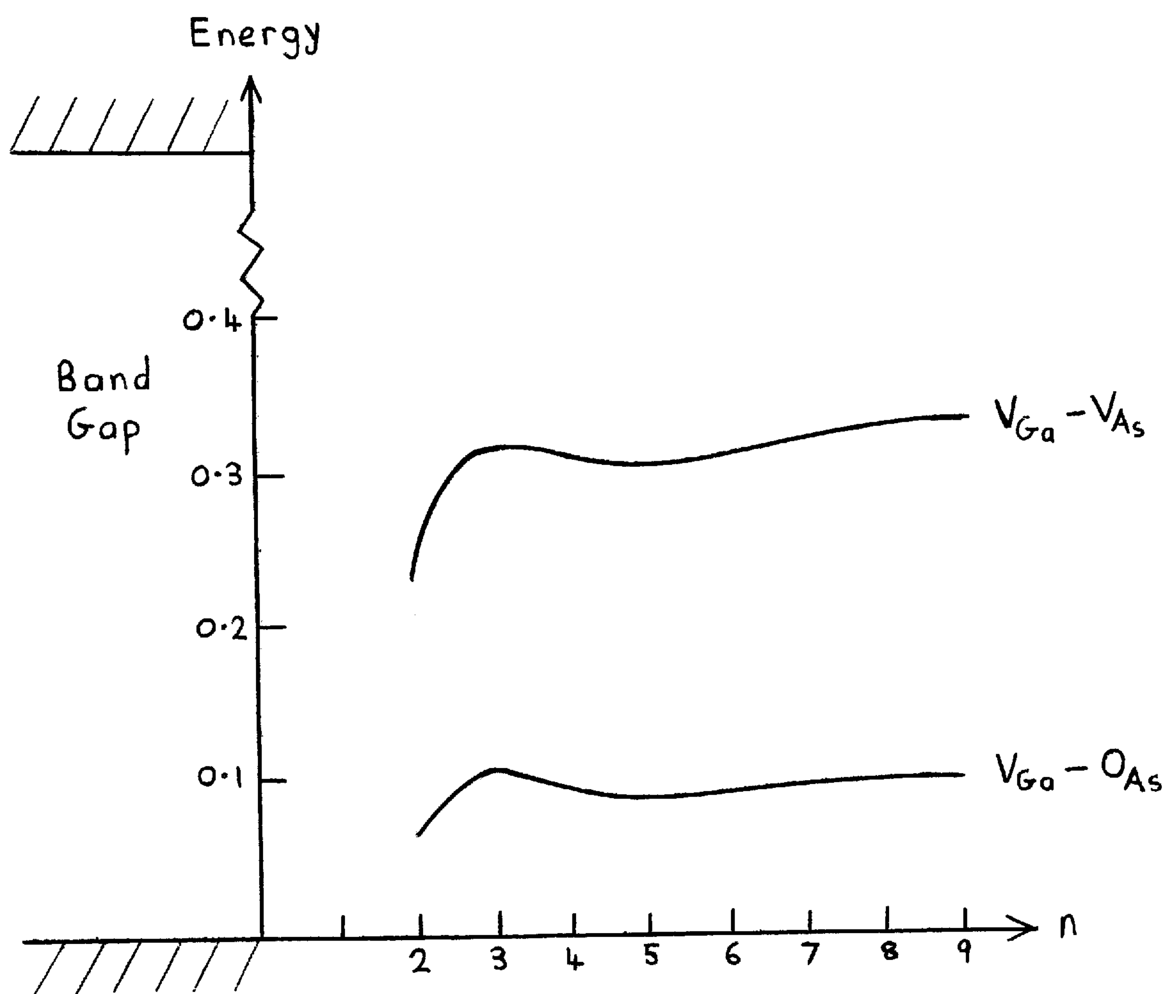


FIGURE 4·3·6

Energy of divacancy and vacancy-oxygen pair singlet states as a function of scaling (both potentials scaled)

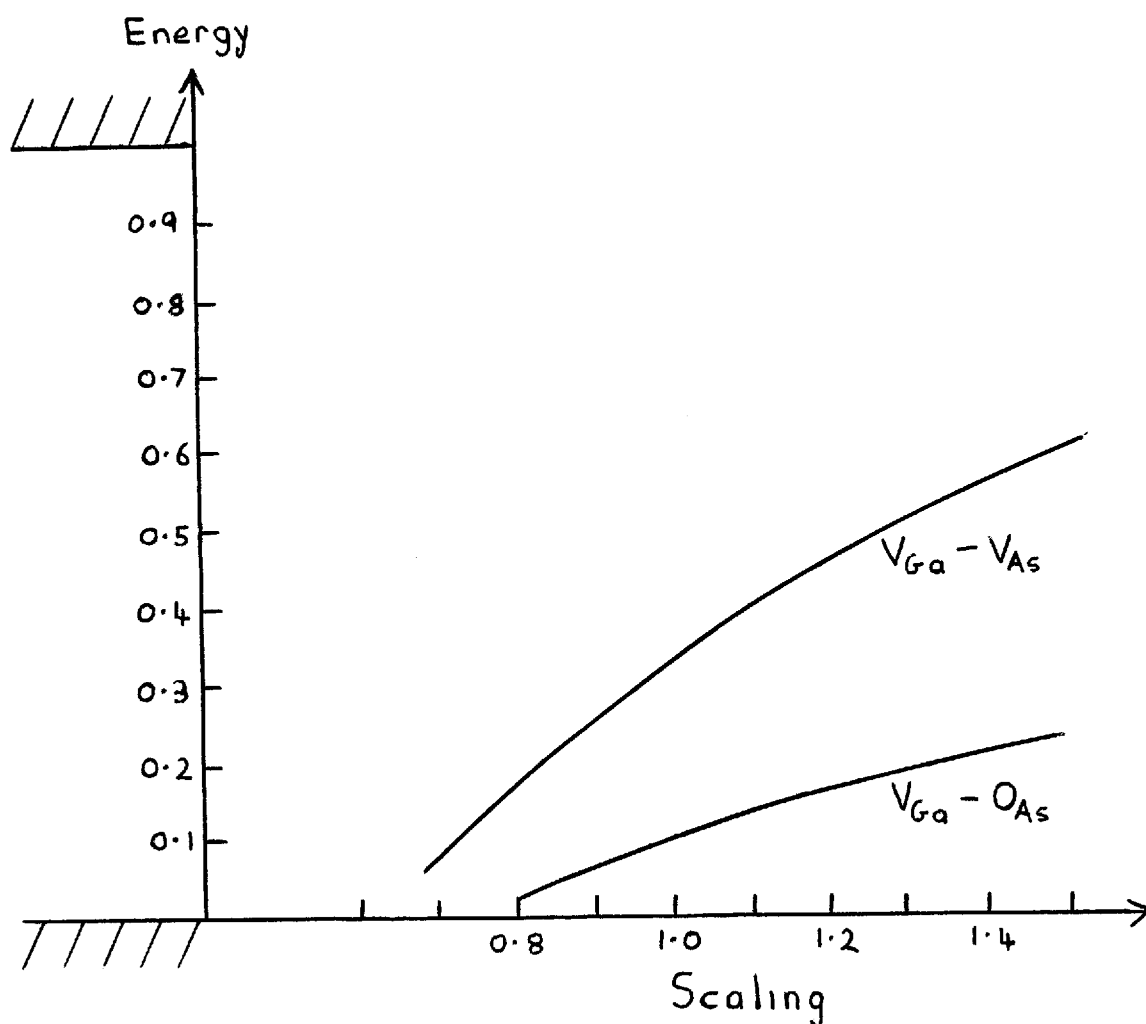


FIGURE 4·3·7

Energy of divacancy and vacancy-oxygen pair doubly degenerate states as a function of n (n=9 implies a total of 18 Laguerre functions, 9 per site)

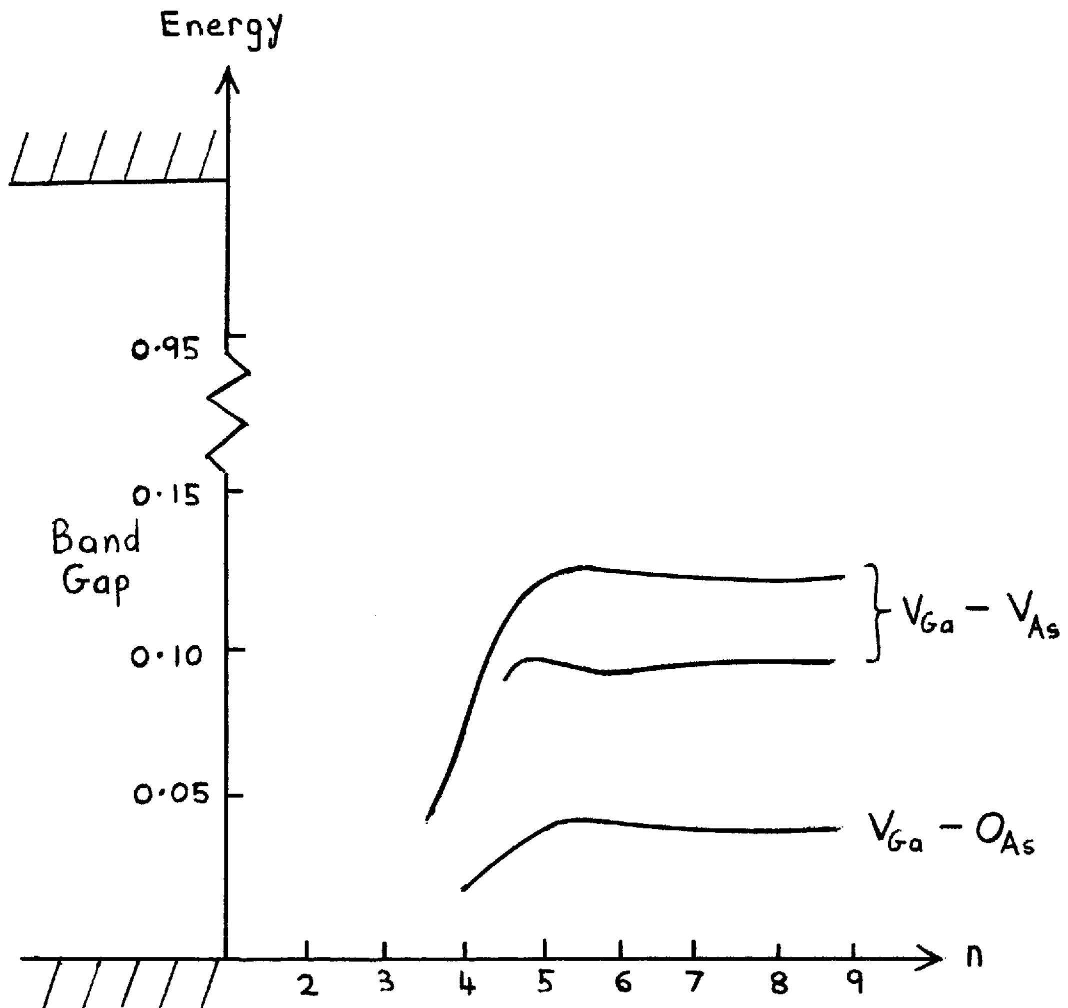
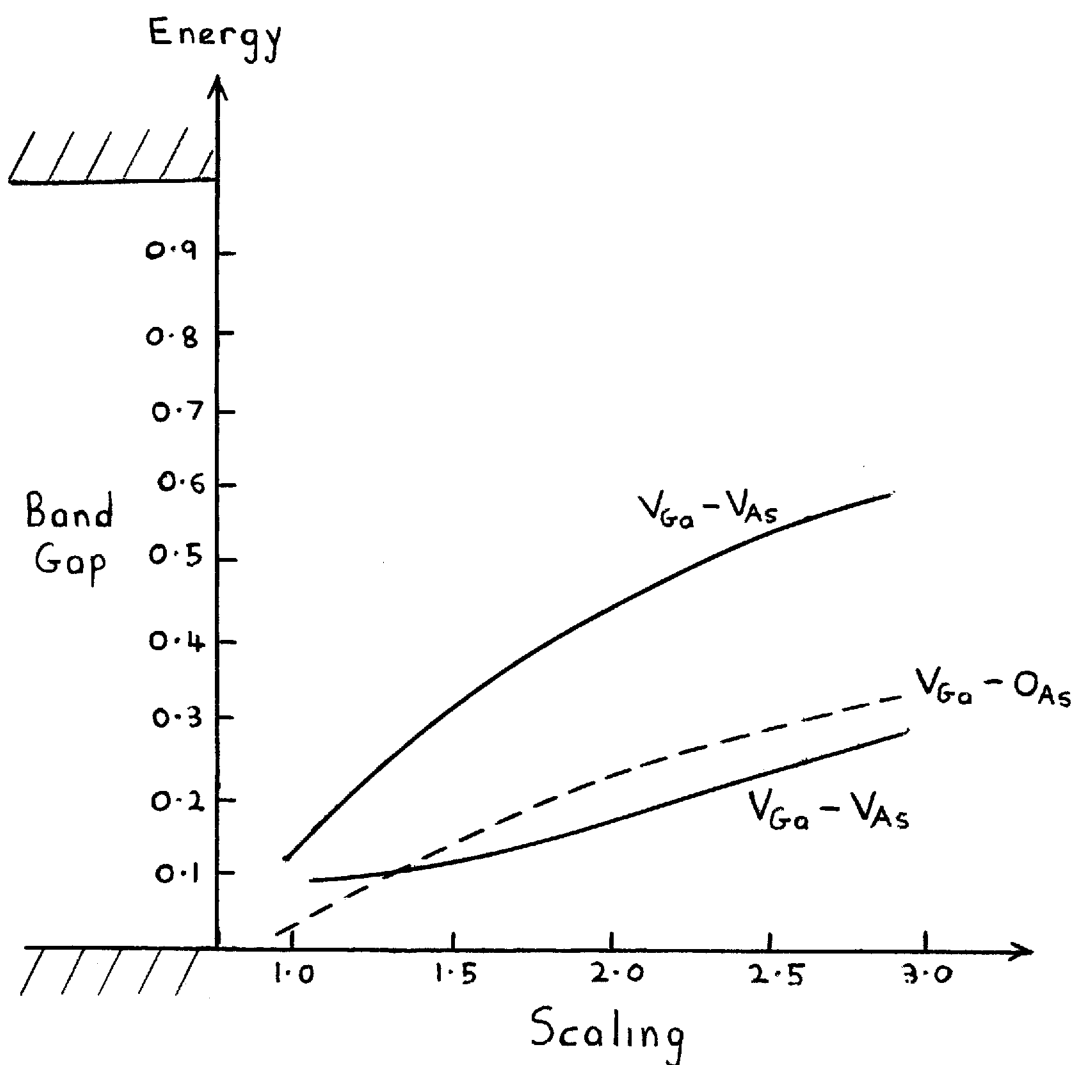


FIGURE 4·3·8

Energy of divacancy and vacancy-oxygen pair doubly degenerate states as a function of scaling (both potentials scaled)



constitute a serious threat to the accuracy of the final results.

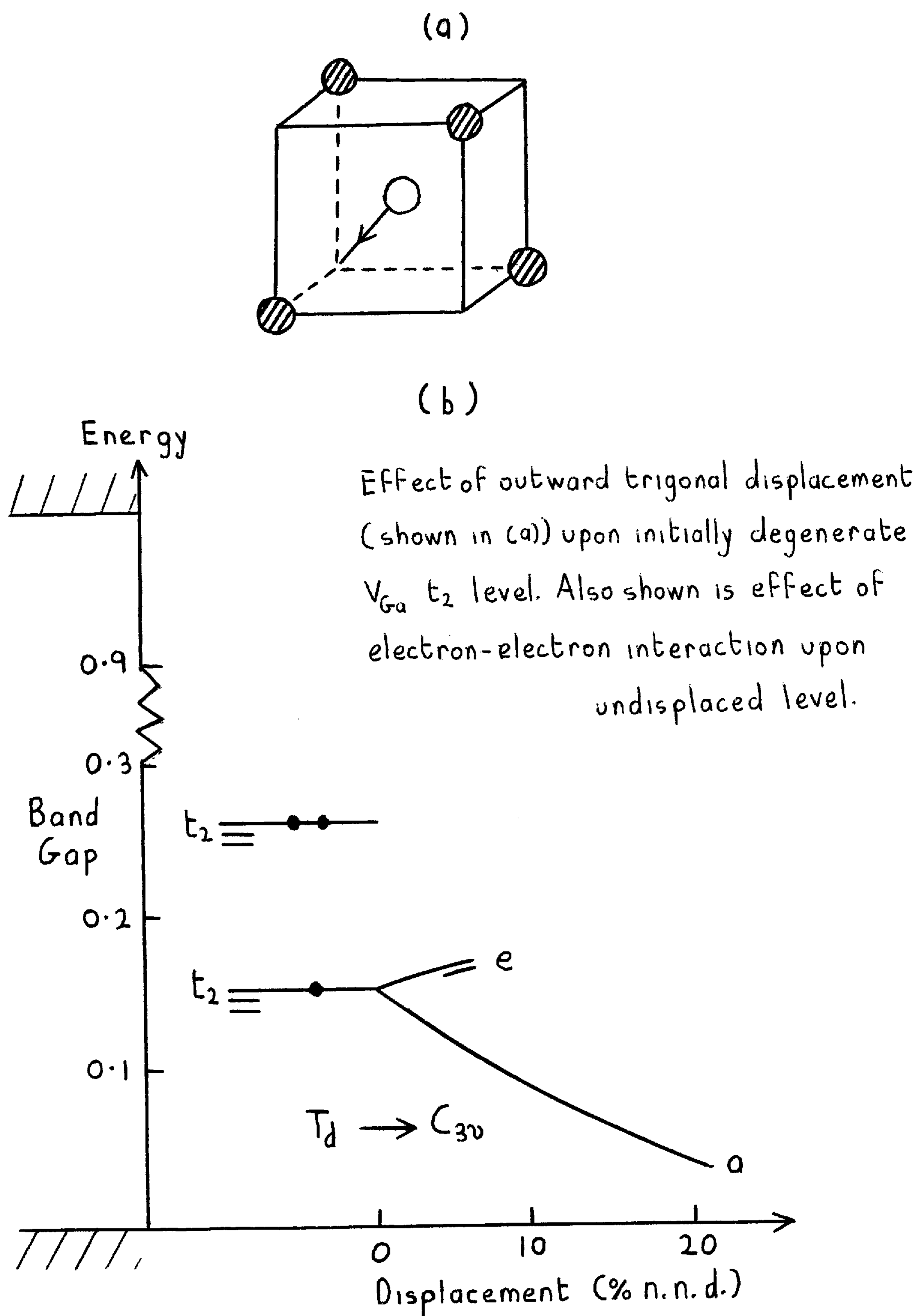
We will now return to the question of the basic validity of using the screened atomic pseudopotentials as we have done. It was pointed out that it could be argued that the introduction of the defect potential could alter the local screening properties thus leading to an error in the initial potential. However, the potentials we use are of short range and hence the most important k-space contributions are at large k i.e. short wavelength. For such values of k the crystal screening effects do not dominate. For example, if we consider $\epsilon(\underline{k})$, the dielectric screening function, at k_{Fermi} (which corresponds to the sort of wavelength we are interested in) then we find $\epsilon(k_{Fermi}) \approx 1.5$. Thus the polarizablility contribution, the part which can be altered by valence charge redistribution, is only 0.5. It would require a very drastic change in order to give rise to any meaningful alteration in final electronic energy given the insensitivity to potential strength which we have observed. Therefore it would appear that the defect potentials we employ are quite adequate for our purposes.

The discussion so far has been centred upon the location of the one-electron levels as determined by the self-consistent screened defect potential in an unrelaxed lattice. However, we cannot in general expect to directly relate these levels to experiment. To explain why this is so we will suggest a simple model for the neutral vacancy. The defect molecule approach of Coulson and Kearsley will be followed. In creating a neutral vacancy it is first necessary to break the bonds connecting the atom which is to be removed from the nearest neighbours. Normally there are eight valence electrons forming bonding pairs with these nearest neighbours. If we remove, say, a gallium core plus three valence electrons in order to create a neutral gallium vacancy then there will be five arsenic nearest neighbour electrons to be accomodated in the resulting dangling bonds. If we create an arsenic vacancy then three electrons remain. In our model these remaining electrons

occupy the localized orbitals generated by our defect formalism and which we associate with the dangling bonds. However, if we do attempt to occupy these orbitals in such a manner then we can expect a certain degree of energy modification to occur as a result of electron-electron interaction. In addition, when we come to occupy the threefold degenerate t_2 levels it can be anticipated that a Jahn-Teller distortion will occur in an effort to reduce the total energy of the system. Symmetric relaxation effects must also be expected. Ideally all such effects should be included in a self-consistent manner from the outset but this is at present beyond our capabilities. However, we can attempt to gain some insight into the consequences of these complicating factors.

We begin by considering the neutral gallium vacancy in which case we must be prepared to occupy the localized states with a total of five electrons. The first two, with opposite spin, will inhabit the lowest energy a_1 level. Electron-electron interaction will cause an increase in energy of the level but a compensating symmetric relaxation will almost certainly ensure that the level remains within the valence band region. The next electron to be added must occupy the triply degenerate t_2 level and we expect Jahn-Teller distortion to cause splitting into a non-degenerate (apart from spin) level and a higher energy doubly degenerate level. We can attempt to estimate the size of the splitting by searching for zeros of the determinant (3.3.3) for a trigonally displaced gallium vacancy potential (see figure 4.3.9(a)). This corresponds to distortion route ② of Figure 4.2.1. Figure 4.3.9 (b) shows the resultant splitting as a function of displacement. With the present programs it is not possible to simulate the alternative tetragonal distortion route. Because of the different localization of the a_1 and t_2 states we do not expect significant electron-electron interaction between electrons in these states. Consequently it will be assumed that we can populate a_1 states within the valence band region without giving rise to noticeable energy

FIGURE 4·3·9



Effect of outward trigonal displacement (shown in (a)) upon initially degenerate V_{Ga} t_2 level. Also shown is effect of electron-electron interaction upon undisplaced level.

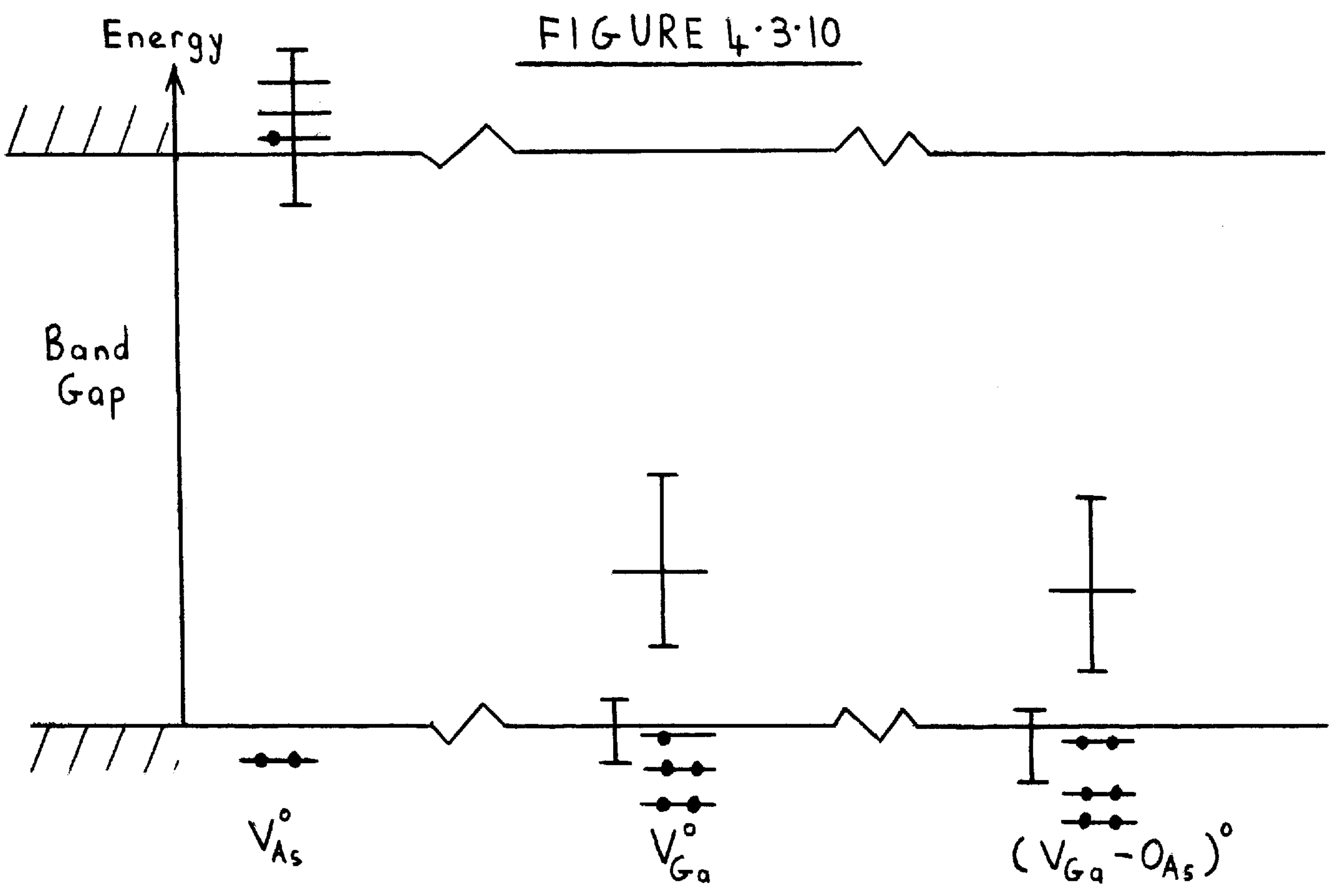
alteration of the t_2 levels. When we come to place a second spin paired electron into the lower t_2 derived state a non-negligible electron-electron interaction is to be expected. Within the Hartree scheme and using a wave function obtained by trial methods the estimation of energy increase shown in Figure 4.3.9(b) is obtained. A previous study by Jaros[(66)] indicated that symmetric relaxation in response to localization of charge can lead to an energy reduction greater than the Hartree scheme electron-electron interaction increase and we assume a similar result in this case. It is then likely that the doubly occupied t_2 derived level will "relax" into the valence band. At this stage our model still only represents the V^+ charge state. In order to obtain V^O we need to add a further electron which will lead to a Jahn-Teller splitting of the remaining doubly degenerate e states. We expect to find the fifth electron occupying a level just within or in the region of the valence band edge. The final situation is illustrated in Figure 4.3.10.

For the arsenic vacancy the position is much simpler for there are only three electrons to deal with. Two are placed in the valence band a_1 level and the third occupies a t_2 level resonant with the conduction band. Jahn-Teller distortion may bring this level down into the band gap region

The situation for the V_{Ga}-O_{As} pair may be viewed in analogy with the V_{Ga} arguments. In this case we have six electrons to deal with however. The additional electron occupies the V_{Ga} hole and the remaining unoccupied level within the gap is slightly lower due to the initially lower energy levels. Given the gallium vacancy domination of this complex it seems reasonable to anticipate similar results for other V_{Ga}-impurity systems. We are then led to predict the existence of a number of V_{Ga} related defect states with similar properties. In particular, we would expect to find a number of V_{Ga} related energy levels close to the isolated vacancy level.

There are again six electrons to be accomodated when considering the divacancy but the additional states within the gap complicate the issue.

FIGURE 4·3·10



Final assessment of positions of levels associated with neutral defects

In this case it is probably simplest and safest to put forward the suggestion that there are likely to be a number of unoccupied levels introduced into the lower part of the band gap.

The immediately preceding arguments concerning the population of the defect states are at best of a semi-quantitative nature only. We cannot at this time include the effects of valence charge redistribution, electron-electron interaction and distortion in a truly self-consistent manner. It is not possible to base the calculations upon a minimum energy configuration, Despite this it is hoped that some measure of insight has been gained. In particular it is indicated that isolated gallium and arsenic vacancies behave as single acceptors and donors repectively and we expect corresponding results for group III and group V vacancies in similar materials.

There have been few other reported theoretical studies of the electronic energy levels associated with vacancies in III-V semiconductors but a few do exist. Using a tight-binding cluster method Lowther obtained results similar to those of Jaros concerning the ordering and general positions of the Ga and As vacancy one-electron energy levels in GaAs.(67) Unpublished results by Bernholc and Pantelides are also in very good agreement. Calculations for vacancies in both GaAs and GaP have been performed by Il'in and Masterov.(68) The results for V_{Ga} in GaAs and GaP and also those for the t_2 level for V_{As} in GaAs are similar to those obtained using the present method. The V_{As} a_1 level and the results for V_P do not agree however. These calculations were based upon a two band model. The general agreement between theoretical calculations for vacancies in GaAs seems encouraging.

Section 4.4 Vacancy-related defects in gallium arsenide-experiment

Compared with our knowledge of vacancy-related defects in silicon, which is still far from complete, the situation for gallium arsenide is considerably more vague. This state of affairs is not only due to the more recent birth of interest in this material. There are a number of inherent experimental

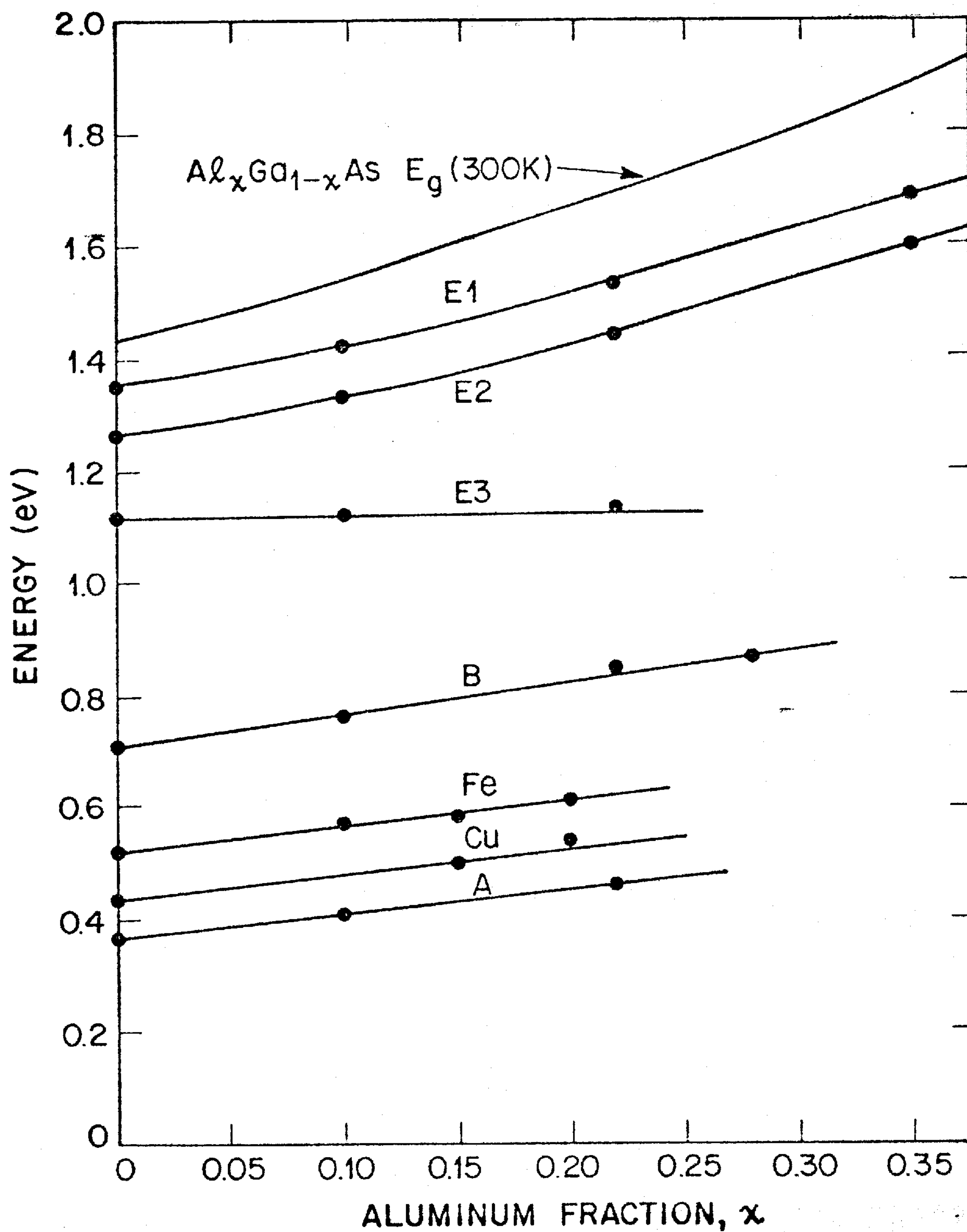
difficulties which have impeded progress. The electron paramagnetic resonance technique, which has proved such an invaluable tool in the study of silicon, is much less well suited to dealing with defects in GaAs and there is little information available from this source. Before giving the reasons for this it is, perhaps, appropriate to point out why the method is so successful with silicon, The silicon isotope ${}_{14}Si^{28}$ which comprises 92% of the naturally occurring element has no nuclear magnetic moment. ${}_{14}Si^{30}$, also with zero magnetic moment, comprises a further 3%. Only ${}_{14}Si^{29}$ with about 5% natural abundance has a resultant nuclear spin. Thus the EPR spectra of unpaired defect electrons in silicon exhibit only a small hyperfine splitting effect due to the presence of only a small percentage of magnetically active nuclei. This interaction is useful for it does give information concerning the defect centre. However, it does not drown out the effects of nearby impurity atoms. The situation is similar for diamond. In gallium arsenide the situation is not favourable for all the nuclei of the two constituent elements have magnetic moments and there is a total of three naturally occurring isotopes, all in significant quantities. The task of resolving the multiply split and broadened EPR defect spectra is thus a considerable one. In addition the problem of interpretation is more difficult to begin with. Not only is there the possibility of two alternative types of vacancies but due to there being two inequivalent interpenetrating f.c.c. lattices there are also more interstitial configurations. The two anti-site possibilities must also be taken into account. There are, of course, all the other complications which have already been encountered in silicon. Defect complexes, multiple charge states, conductivity type and other existing defect centres can all affect the results. Add to this the extra preparation difficulties and the need to eliminate trace impurities and it can be appreciated that the task of experimental interpretation is not an easy one. It thus comes as no surprise to find that there are no absolutely unambiguously indentified native defects in GaAs, although proposals regarding possible identifications

are legion. For instance, it has been suggested that gallium vacancies are acceptors, that arsenic vacancies are acceptors, that both gallium and arsenic vacancies are acceptors or, as proposed by Chiang and Pearson, that arsenic vacancies are donors and gallium vacancies acceptors.[69] (Chiang and Pearson give a number of references where evidence supporting the other suggestions can be found.) Hasegawa and Majerfield have reported an electron trap 0.83 eV from the conduction band and a hole trap 0.64 eV from the valence band of n-type GaAs which seem to be related to gallium and arsenic vacancies respectively.[70] Given the multitude of conflicting alternatives it can be fairly said that it is possible to find some experimental evidence to support almost any theoretical prediction. In fact, there can be some argument as to whether isolated vacancies in GaAs are likely to exist in observable quantities at all near room temperature. Certainly on the basis of our knowledge of vacancies in silicon, germanium and other materials this question would seem very reasonable. As we have already seen, gallium and arsenic vacancies may behave as acceptors and donors respectively. They are then likely to be in opposite charge states and be attracted to form divacancies. It is known that in silicon vacancies and interstitial oxygen have an affinity for each other and combine to form the so-called A centre. Oxygen, a ubiquitous impurity, is also present in GaAs although in this case is more likely to be in a substitutional position. The two might then combine to form the neutral $V_{Ga}-O_{As}$ pair with the oxygen atom providing the electron to fill the gallium hole as suggested in the last section. The concentration of such defect complexes is crucially dependent upon vacancy mobility, however, and it can be argued that this is much lower in GaAs than in silicon. In GaAs vacancy migration requires a next nearest neighbour jump and this process must be considered less likely than the straightforward nearest neighbour exchange in silicon. Therefore we might well expect to find stable isolated vacancies in GaAs even at room temperature.

Recently considerable information concerning the positions of defect energy levels in GaAs has become available. This is due to the Deep Level Transient Spectroscopy (DLTS) method of Lang.[71] Although there had been previous interest in pn or Schottky junction capacitance measurements it was Lang who extended the basic technique to give it the status of a truly spectroscopic method. The basic method involves the observation of capacitance transients resulting from the non-equilibrium occupation of energy levels within the band gap produced by the use of bias voltage pulses. Langs' DLTS expansion allows the location of individual electronic energy levels in a convenient manner. Although much information can be obtained using this method it cannot directly identify the defect centres responsible for a particular level. Probably the most convincing possible identification of a level associated with a vacancy in GaAs has been made in conjunction with the use of DLTS. On the basis of considerable circumstantial evidence Lang, Logan and Kimerling have associated the DLTS E3 level shown in Figure 4.4.1 with the gallium vacancy.[72] The peculiarly constant position of the electron irradiation induced E3 level relative to the valence band edge of $Al_xGa_{1-x}As$ implies a defect wave function strongly related to the valence band. As pointed out in the previous section and in line with the theoretical results for silicon of Louie et al.[61] and others, and indeed the early work of Callaway and Hughes,[7] such a property is characteristic of vacancies. On the basis of the results presented in the last section the postion of the E3 level seems consistent with it being due to the arsenic vacancy. However, the experimental results seem to exclude this possibility and Lang et al. suspect that V_{Ga} is the culprit. At first sight this conclusion seems totally incompatible with the theoretical results. However it may be that a negatively charged defect centre is involved. Also, the measured E3 level cannot be directly equated with the theoretical levels for the electron-electron interaction energy is given up to the electron on emission to the

FIGURE 4·4·1

Composition dependent deep levels in $Al_x Ga_{1-x} As$



conduction band. In our theoretical scheme the true level would then lie near the middle of the gap. Given this and assuming that a V_{Ga}^{-} charge state energy level is involved then it may be possible to reconcile the experimental and theoretical results.

Section 4.5 Summary

The results of a number of calculations giving the electronic energy levels associated with vacancy-related defects in gallium arsenide have been presented. It is apparent that Jahn-Teller and symmetric distortion together with electron-electron interaction are all of equal importance in determining the final energy levels. An attempt has been made to incorporate these factors into a semi-quantitative analysis of the gallium and arsenic vacancy and vacancy-oxygen pair problems. There seems to be reasonable agreement between alternative theoretical schemes, at least at an elementary level, but the extent to which this may merely represent a mutual inability to cope with what may be the real problems remains to be seen. It is clear that much more "hard" experimental information is required before an objective assessment of the theoretical results can be made.

CHAPTER 5

DISPLACEMENT DEPENDENT ENERGY LEVELS

Displacement dependent energy levels

Section 5.1 Introduction

We shall now concentrate attention upon the topic of displacement dependent energy level behaviour. Previously we have used the available computer programs to calculate the size of the Jahn-Teller splitting in response to a particular distortion mode simulated by axial movement of the vacancy potential. However, this is not the only context in which such studies can be of importance. For example, we might wish to know if the substitutional site is likely to be the stable configuration for an impurity atom when an electron is bound into a localized state. The energy level may be a sensitive function of displacement and indeed may be lowered when these movements take place. In such cases we would expect lattice readjustment to occur in order to minimize the total energy of the system when the level is occupied by an electron. Theoretical investigations of the type to be presented could be of value in predicting whether such readjustments are likely to be energetically favourable. As a consequence useful information concerning the symmetry of the defect states may be obtained.

Another of the features of interest with respect to the properties of defect centres is the manner in which carriers are captured. By what mechanism is the energy lost by an electron falling into a deep level carried away? It is now quite apparent that in many instances non-radiative capture is occurring. There are three processes which can be invoked to explain such non-radiative transitions. These are the cascade, Auger and multiphonon emission processes.

In the cascade model of Lax[(73)] energy is dissipated by independent emission of single phonons as an electron drops through a series of closely spaced energy levels with separation equal to a typical phonon energy. We would only expect to find such a series of levels associated with a charged, attractive centre. The final transition to the ground state may involve the

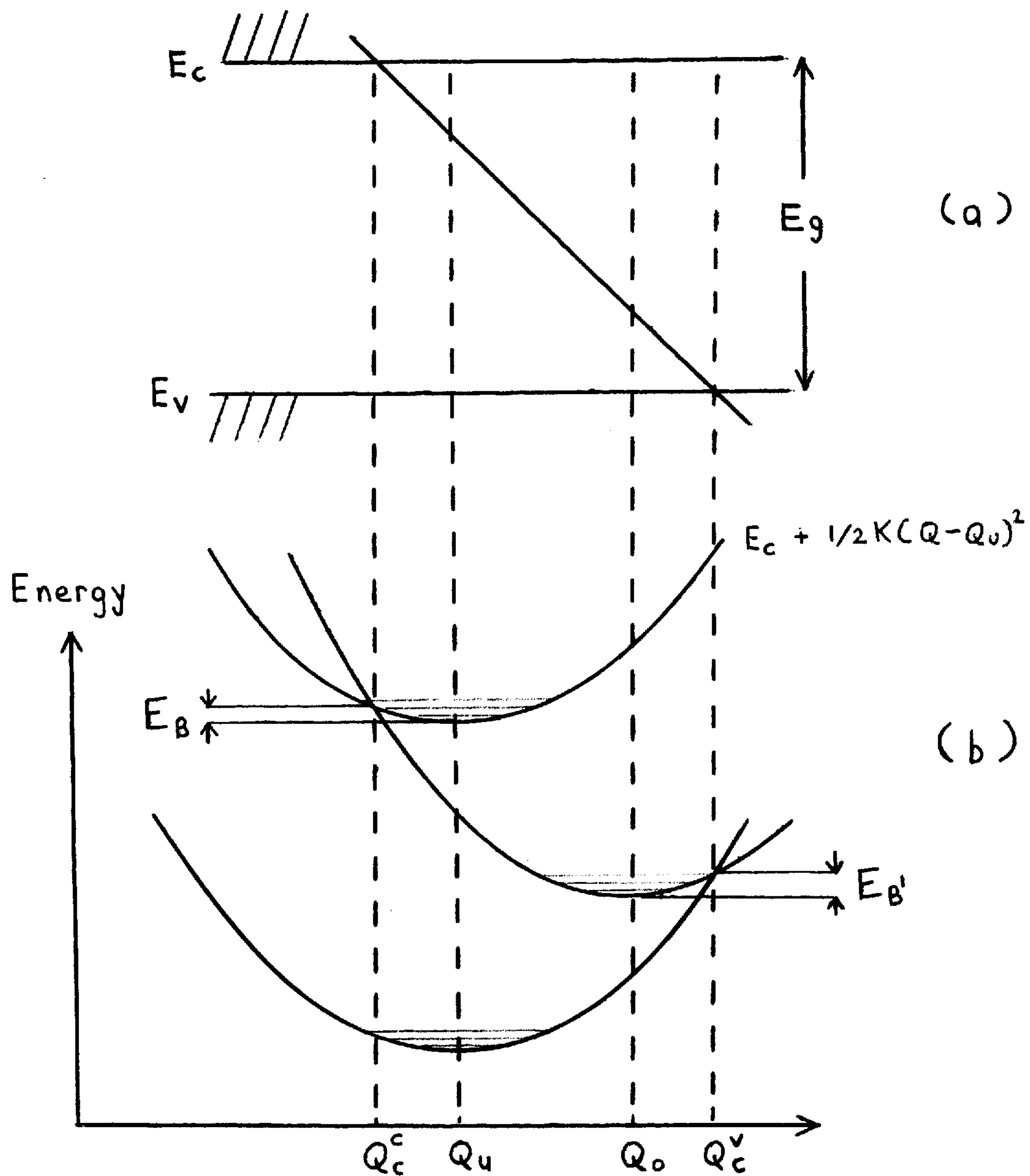
dissipation of a considerable amount of energy. However, provided that thermal emission is not much more likely to occur then this last step does not constitute a limiting factor for the overall process. The capture cross sections of attractive defects are large, typically 10^{-15} to 10^{-12} cm^2. Lax originally suggested that polarization could give rise to a similar series of levels associated with neutral defects which could explain a number of rather smaller observed cross sections. Later, Bonch-Bruevitz and Glasko showed that improbably high polarizabilities would be required for this to be possible.[74] In principle, as we have recently demonstrated,[63] realistic short range potentials can produce excited states within the gap which would be likely to enhance the probablility of capture but we shall not consider the consequences of this result. In general we cannot expect the cascade mechanism to be of great importance when considering capture by neutral centres.

We will not consider the Auger effect in detail but note that it has been identified as occurring in several instances of bound exciton recombination. Auger effect energy losses via free carriers can be easily recognized for the capture cross sections are then linearly related to carrier concentration. There is little dependence of cross section upon temperature and so there is not much probability of confusion with the MPE process which shall be described next.

The multiphonon emission process is probably best understood in terms of the configuration coordinate diagram. Figure 5.1.1 (a) shows a hypothetical unoccupied defect energy level with a linear dependence upon coordinate which we denote by a single variable, Q, for simplicity. Although in our theoretical calculations no such simple linear dependency has been observed this linear form is appropriate for illustrative purposes for it serves to conveniently demonstrate the basic MPE capture process. The defect energy level is a sensitive function of displacement and at Q_c^c and Q_c^v is equal to the respective conduction and valence band electronic energies. The elastic

FIGURE 5·1·1

Configuration coordinate diagram



Q_u — unoccupied equilibrium coordinate

Q_o — occupied equilibrium coordinate

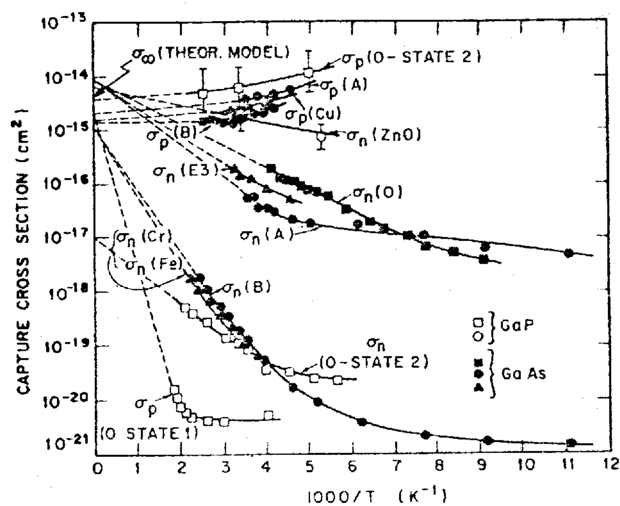
energy of the lattice is $1/2K(Q-Q_u)^2$ relative to the unoccupied equilibrium minimum at Q_u where K is a suitable force constant. In Figure 5.1.1(b) the total of elastic + electronic energy is plotted as a function of Q for the defect and band edge states. If sufficient thermal lattice vibration energy $E_B(=1/2K(Q_c^c - Q_u)^2)$ is available then the unoccupied defect level crosses into the conduction band and electron capture can occur. Upon capture the system is far from the new equilibrium position at Q_o and the excess elastic energy is liberated by means of violent local vibrations thus justifying the multiphonon emission nomenclature. Hole capture can occur in a similar manner if the electronic energy of the occupied defect state crosses the valence band at Q_c^v. Relative to the occupied equilibrium positon at Q_o thermal energy $E_{B'}(=1/2K(Q_c^v-Q_o)^2)$ is required for this to occur. The value of our defect scheme with respect to the MPE problem is in the investigation of whether the band edge crossings are likely for we can calculate the energy of the localized state as a function of Q. We shall not in this account involve ourselves in the theoretical calculation of the capture cross sections. Theoretical treatment of non-radiative transitions by MPE was first attempted by Huang and Rhys[75] and later by various other workers. The most recent study is that of Henry and Lang[76] (references to earlier works are given by these authors). Henry and Lang found that the experimental temperature dependent behaviour of a number of cross sections was consistent with their theoretical MPE model. As can be seen from the extrapolations in Figure 5.1.2 the cross sections near room temperature can be described by an expression of the form

$$\sigma_{NR}(T) = \sigma_{\infty} e^{-E_{\infty}/kT} \tag{5.1.1}$$

where σ_{∞} is usually of value 10^{-15}-10^{-14} cm^2. and

$$E_{\infty} = E_B - kT_R . \tag{5.1.2}$$

FIGURE 5·1·2



Carrier-capture cross sections vs inverse temperature for various deep levels in GaAs and GaP. An *n* subscript denotes electron capture while a *p* denotes hole capture. Levels shown for GaAs are due to Cu, Fe, Cr, O, the *E*3 radiation damage defect, and two unidentified but commonly occurring levels *A* and *B*. For GaP the cross sections are associated with the Zn-O center and the two states of the oxygen center. The dashed lines show the temperature dependence of the cross sections extrapolated to $T \to \infty$. The value of σ_∞ predicted by the theoretical model is indicated on the ordinate.

In addition to the academic interest engendered by the possibility of MPE processes there are a number of very good practical reasons for considering the consequences of this phenomenon. To begin with, it is quite clear that any competing non-radiative carrier capture process is not likely to be looked upon favourably by producers of L.E.D.'s. The quantum efficiency of these devices is notoriously low and understanding of one of the possible reasons for this is obviously of importance. Apart from the problem of carrier removal there has also been some interest in the topic of recombination-enhanced defect reactions (REDR's). In the complete across-the-gap MPE recombination process a total energy E_g is released in the region of the defect and this energy may promote REDR. For example, let us consider an interstitial atom with a barrier height E to be overcome in order that the interstitial may jump to a new site. If E is much greater than the average available thermal energy then we would expect to see little migration under normal circumstances. However, if the interstitial acts as an MPE recombination centre then the violent lattice vibrations resulting from the local release of energy during the electron/hole capture processes are likely to enhance the probability of interstitial movement. This could lead to the eventual disappearance of the isolated interstitial due to joint interstitial-vacancy annihilation or to combination with other defects. Alternatively, a more complex defect aggregate could be dissociated under the action of the energy release. The experimental work of Lang and Kimerling indicates that such REDR effects do occur in semiconductors. As shown in Figure 5.1.3 they found the annealing rate of the electron irradiation induced E3 level in GaAs (the same level encountered in the previous chapter) to be related to forward diode current and thus to recombination rate.(77,78) Lang and Kimerling found that a number of irradiation induced defects in GaP also exhibit similar enhanced annealing rates.(79) There has also been some speculation that the degradation of the Zn_{Ga}-O_p red luminescence could be

related to MPE carrier capture energy release. Recombination enhanced defect reactions have been discussed in a recent review article by Dean and Choyke.[80]

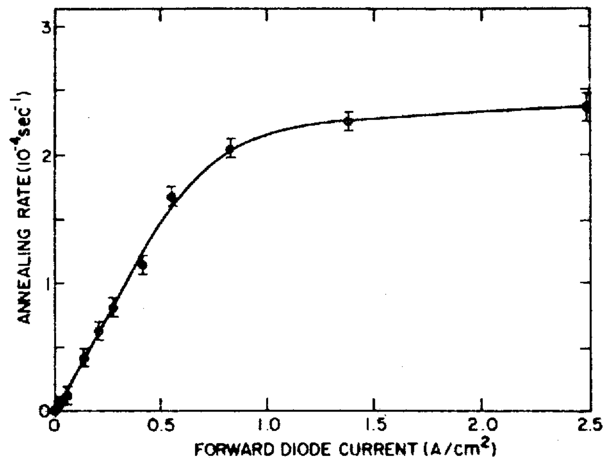


FIGURE 5.1.3

Annealing rate of GaAs E3 electron trap as a function of diode current.

The results in the following sections demonstrate the response to static and vibrational modifications to lattice coordinate, Q, of various impurity energy levels. In particular, we shall be interested in the trigonal displacement mode. General studies of this sort may be of some interest with regard to the prediction of symmetry type and the possible carrier capture mechanisms occurring at defect centres.

Section 5.2 The isolated nitrogen impurity in diamond

One of the features of natural diamonds is the large concentration of impurities which they are known to contain. Nitrogen is of particular interest for in some cases the abundance can be as much as 0.1%. We might attempt to theoretically treat such a group V impurity as a simple donor within the effective mass theory approach. However, the low dielectric constant and high effective mass for diamond would then lead to a small orbital radius from which we would infer that the use of EMT was inappropriate. We must expect a well localized defect state dominated by the short range characteristics of

the impurity potential. Such a defect is suitable for treatment by our theoretical scheme and is also a candidate for experimental investigation by electron spin resonance methods (as has already been stated diamond is a good subject material for such experimental studies). Indeed, it was with the use of the EPR technique that Smith and co-workers[81] were first able to analyse the characteristics of this impurity. Interpretation of the hyperfine structure made it quite evident that the impurity state was axially symmetric with the fifth nitrogen electron located primarily within the region of the nitrogen sustituent and a single carbon atom. It was also suggested that there was an outward displacement of the nitrogen atom leading to a lengthening of the C-N bond. Later, analysis of more detailed results by Loubser and Du Preez[82] confirmed the general conclusions of Smith et al. although some modification concerning the precise details of the charge localization was proposed.

Before beginning our theoretical calculations we first need to decide upon the choice of a suitable nitrogen pseudopotential. There is some difficulty in this respect for more than one potential is available. We could contemplate the use of the model (Heine-Aberenkov) potentials produced by Jones and Lettington[83] or Appapillai and Heine.[84] To avoid what would in effect have to be an arbitrary choice between these alternatives we perform calculations utilizing both. The difference in the final energy levels will give an indication of the error which is to be expected as a result of uncertainties in the initial potential.

We now consider the total defect potential which corresponds to a static outward axial displacement of the substitutional nitrogen impurity as shown in Figure 4.3.9(a). In real space we have

$$h(\underline{r}) = v_N(|\underline{r}|) - v_C(|\underline{r} + \underline{R}|) \qquad (5.2.1)$$

where v_N, v_C are respectively the nitrogen and carbon screened ion potentials.

Relative to the perfect crystal substitutional site the nitrogen impurity is situated at $\underline{R}$ which is taken to be the origin for the purpose of the calculation. We could treat the system as a two centre problem as was done for $V_{Ga}-V_{As}$ (CASE 1 of Section 3.5) but we will elect to follow a different line of approach. In reciprocal space the total defect potential is given by

$$h(\underline{q}) = v_N(\underline{q}) - e^{i\underline{q}\cdot\underline{R}} v_C(\underline{q}) \qquad (5.2.2)$$

Although the individual $v(\underline{q})$ are taken to be dependent upon $|\underline{q}|$ only, $h(\underline{q})$ has an angular dependence due to the $e^{i\underline{q}\cdot\underline{R}}$ factor arising from the different positions of the two potentials. Making use of (3.5.4), Equation (5.2.2) can be expressed as

$$h(\underline{q}) = v_N(\underline{q}) - v_C(\underline{q}) 4\pi \sum_{L,M} i^L j_L(qR) Y^*_{L,M}(\theta_{\underline{R}}, \phi_{\underline{R}}) Y_{L,M}(\theta, \phi) \qquad (5.2.3)$$

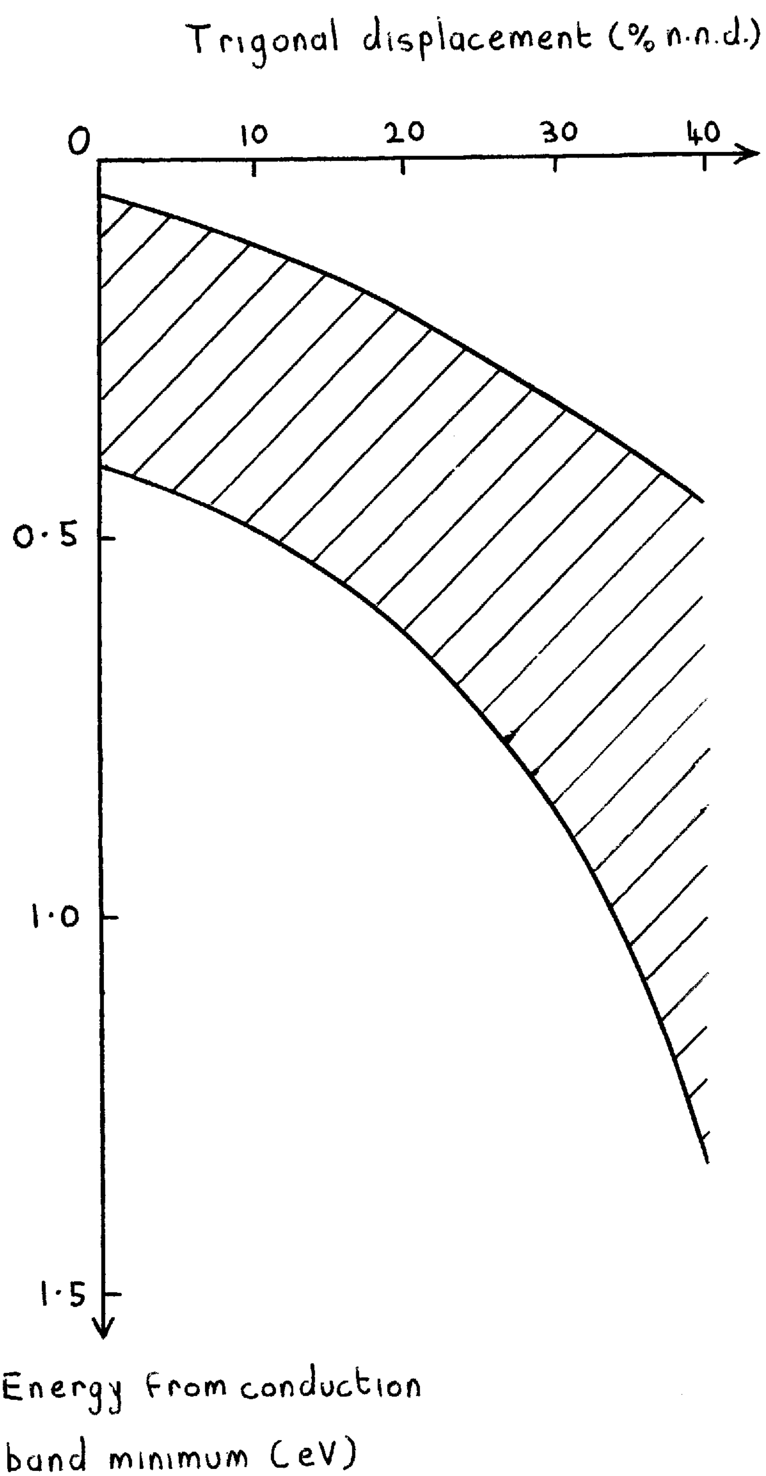
We include only the first term in the expansion of $e^{i\underline{q}\cdot\underline{R}}$ and this gives

$$h(q) = v_N(q) - v_C(q) \frac{\sin qR}{qR} \qquad (5.2.4)$$

This is equivalent to an angular averaging process. As $R \rightarrow 0$ the simple substitutional form is obtained. After fitting to a suitable real space potential(3.7.8) we are ready to procede with the calculation using CASE 2 of Section 3.5. We include g_m of s and p_z type. The use of CASE 2 rather than CASE 1 leads to a considerable saving in computer time which is particularly significant as we need to repeat the calculation for various $\underline{R}$ and also for the alternative potentials. Figure 5.2.1 shows the resultant energy level as a function of displacement. As is illustrated by the shaded area, there is a large uncertainty in the results but it is quite evident that the overall position of the energy level is in the upper part of the forbidden gap region. Also, there can be no doubt that the energy

FIGURE 5·2·1

Energy level of substitutional nitrogen impurity in diamond as a function of displacement



is lowered in response to outward movements of the nitrogen impurity. Therefore, when the impurity state is occupied by an electron we expect a trigonal distortion to occur. Unfortunately, we cannot calculate the precise position of the occupied level for we are unable to adequately account for the changes in the valance band which would allow us to predict the minimum energy configuration. Any such attempt would also be frustrated by the errors arising from the lack of self-consistency, the increasing failure of the approximation (5.2.4) and, of course, the initial uncertainties in the potential. Although we have sacrificed some degree of accuracy by using the truncated form (5.2.4) this is well justified considering the time-saving which is gained. On investigation of the importance of the individual contributions in determining the position of the level we find that the purely s-type matrix elements of (3.3.3) dominate. However, the energy level would rise by up to 1/2 eV if we did not include p_z functions in the g_m. The convergence of the energy with number of functions and also the response to scaling the potential are similar to the vacancy results of Chapter 4 and the results presented in the following section.

It may be of interest to compare our findings with those of an earlier thoeretical study of the same problem. We refer to the EHT work of Messmer and Watkins.[56,57] These authors found the simple substitutional nitrogen impurity level to be in the region below the conduction band minimum in broad agreement with our result. They also found the outward trigonal displacement to lead to an increase in binding energy. However, in these calculations the energy was found to be very sensitive to displacement: the binding energy was increased by about 6 eV in response to a 25% displacement (the EHT cluster model gave a band gap of 9.5 eV rather than the observed value of 5.5 eV). In comparison the figure obtained from our calculation is only $\approx$ 0.5 eV. Messmer and Watkins attempted to locate the equilibrium position of the impurity level by minimizing the total one-electron energy. Using this method they concluded

that the energy level was to be found at $\approx E_v + 2$ eV. However, there is some doubt concerning the validity of this energy minimization scheme for electron-electron and nuclear interaction terms are not dealt with in a totally satisfactory manner. To quote Larkins, who has also made use of the EHT method "it is improbable that reliable distortion studies can be undertaken within this framework."[85] We do not encounter such problems using our method for any realistic displacement still leads to the same general result i.e. the nitrogen donor level is to be found in the upper part of the band gap. This conclusion seems to be supported by the practical results which have been accumulated for type lb diamonds (type lb diamonds are believed to contain nitrogen impurities in isolated form as opposed to the associated la type). The optical and thermal excitation experiments of Farrer[86] suggest that the nitrogen donor level may lie 1.7 eV below the conduction band edge. A recent analysis by Davies[87] of available experimental information lends support to this claim.

In conclusion we may say that our results for the nitrogen impurity in diamond are in broad agreement with experiment. The theoretical studies support the proposition that this impurity centre has C_{3v} symmetry and also that the energy level lies in the upper part of the band gap.

Section 5.3 Further displacement studies : MPE capture

In this section we shall continue to be concerned with the displacement dependent behaviour of energy levels but our interpretation will mainly be within a vibrating lattice, MPE context. It will be assumed that the adiabatic approximation is applicable.

We begin by considering the deep level associated with a single electron bound by the oxygen impurity in gallium phosphide. From experiment this familiar level is known to be $\approx$ 0.9 eV from the conduction band and it is believed that the state is of a_1 symmetry. As has been previously demonstrated

by Jaros[66] the energy obtained using the present theoretical scheme is in very good agreement with experiment. The success in dealing with this particular problem leads us to think that it will be amenable to further theoretical investigation. Consequently, we have decided to calculate the position of the oxygen impurity state-1 energy level as a function of trigonal displacement. The results obtained are displayed in Figure 5.3.1. (Obviously, the reliability of the large displacement results is rather suspect and, in any event, normal thermal vibrations are not likely to give rise to such large excursions. However, our comments will be of a fairly general nature and we think the form of the results to be of some interest so these large displacements have been retained.) In Figures 5.3.2 and 5.3.3 the convergence and response to scaling is shown. It is apparent that there is no basic difference between the substitutional and displaced results. To simplify the calculations we have made no attempt to include the energy contribution arising from the long range component of the impurity potential. The neglect of this term means that the level is a little higher than otherwise but the form of the results will be little affected. It is clear from Figure 5.3.1 that the occupied state will be of high symmetry for the energy is at a minimum at the substitutional site. It can be seen that if we include only s-type functions in (3.3.3) we would expect MPE capture to be likely because the energy level fairly rapidly crosses the conduction band edge. However, the full s-p_z results contradict this conclusion for we find that crossing does not occur for any displacement, no matter how large. In addition we already know that the level is not sensitive to symmetric distortion. Consequently, we are led to believe that electron capture via MPE processes cannot occur in this case. This conclusion is substantiated by experiment for the temperature dependent cross section does not exhibit the behaviour we would expect for MPE capture.

FIGURE 5·3·1

Response to displacement of O and Z_{Ga}–O levels in GaP

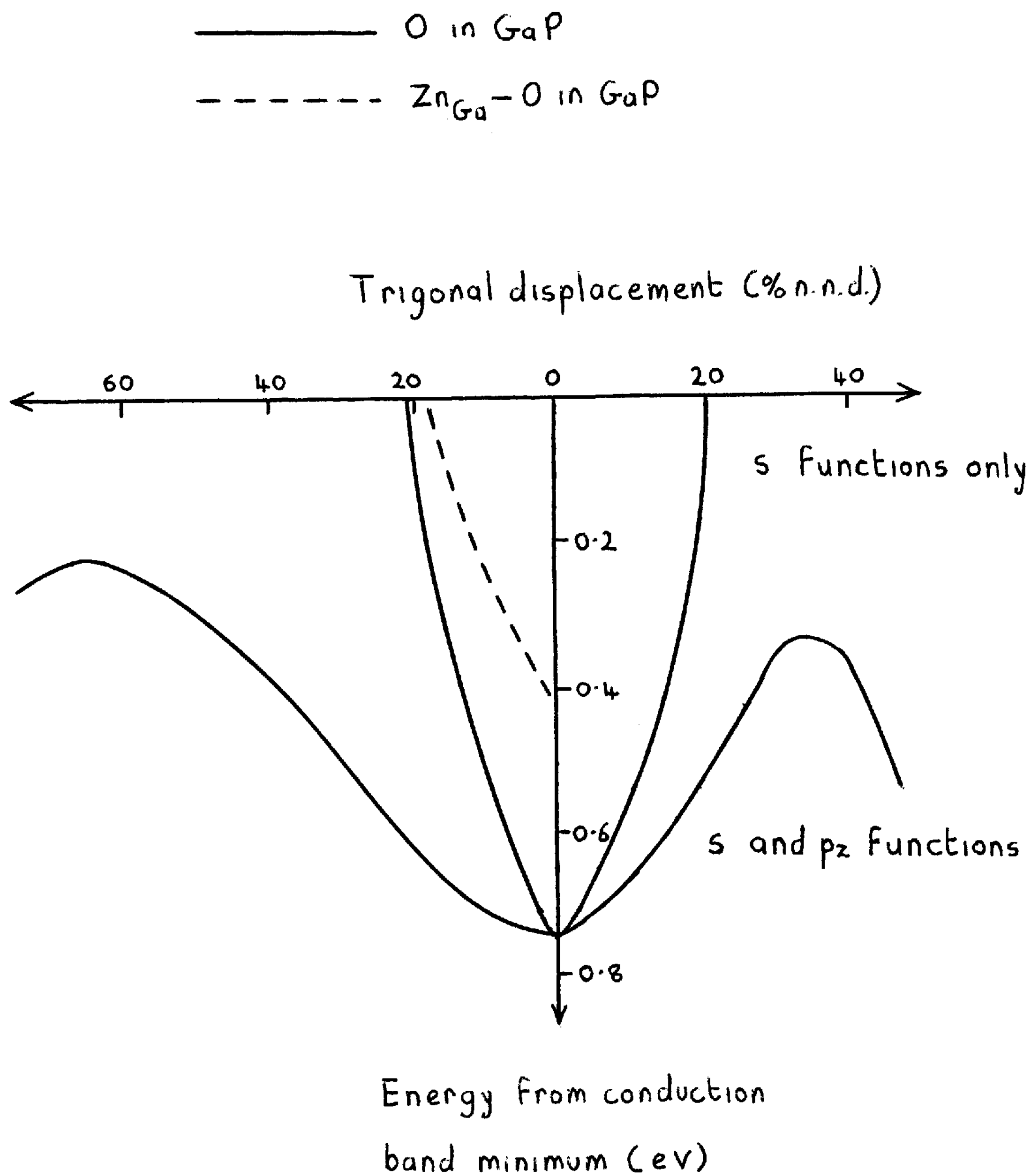
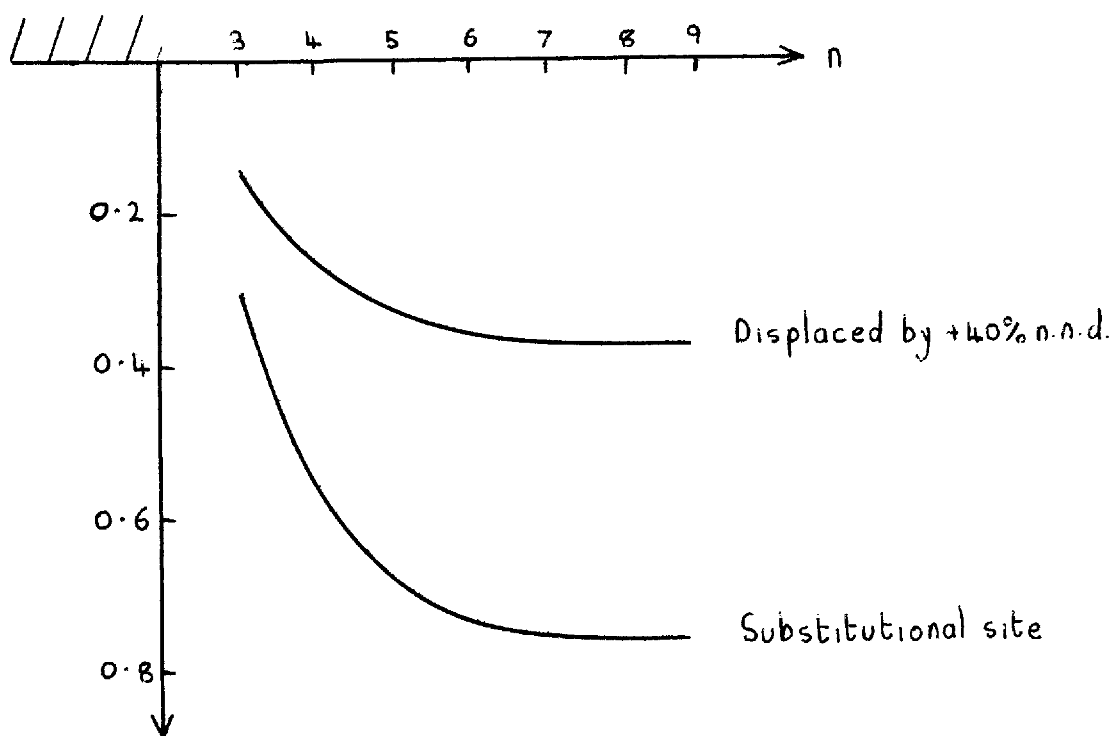


FIGURE 5·3·2

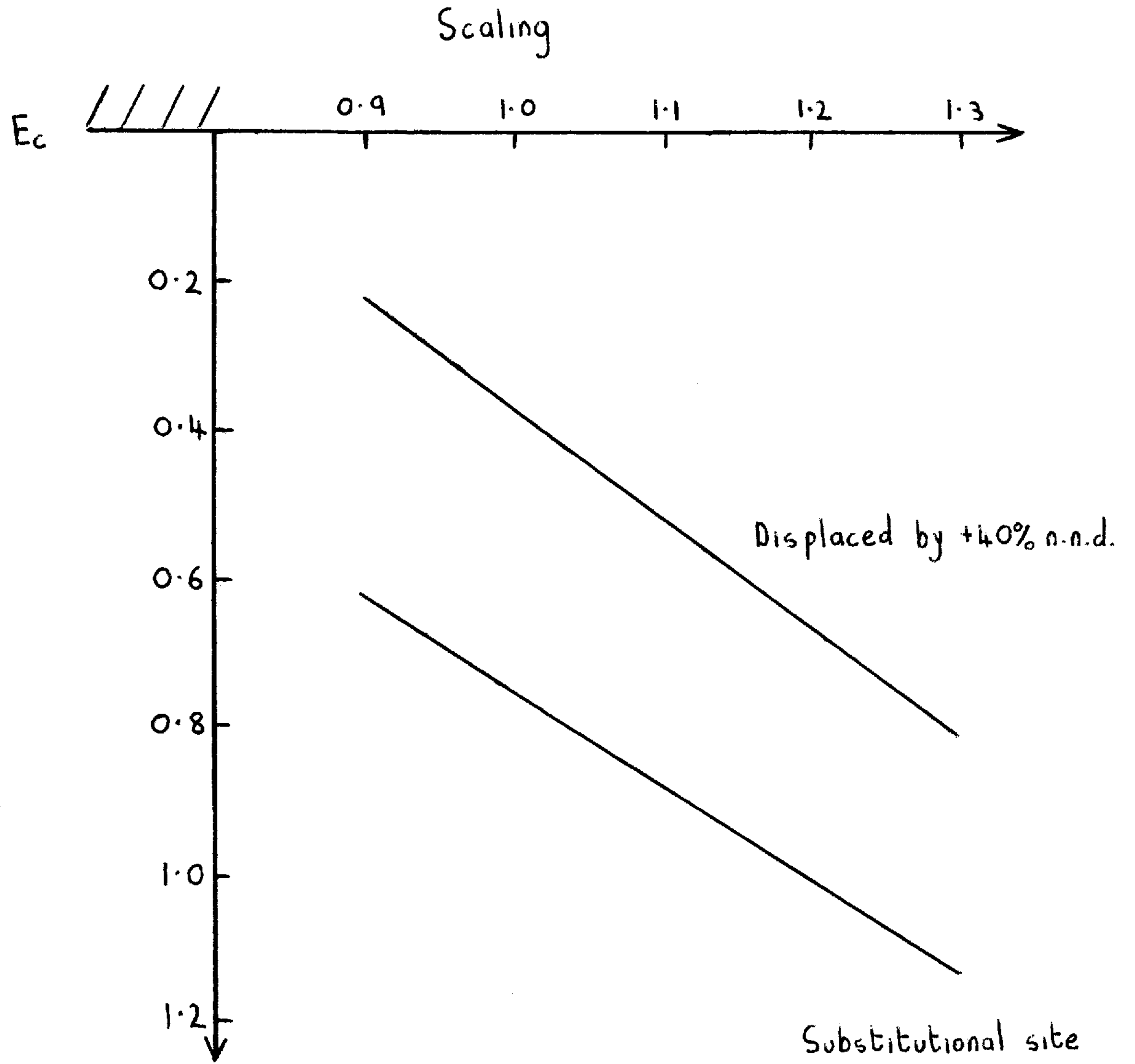
Convergence of oxygen impurity energy level with n



Energy from conduction
band minimum (eV)

FIGURE 5·3·3

Response to scaling of the oxygen impurity in GaP



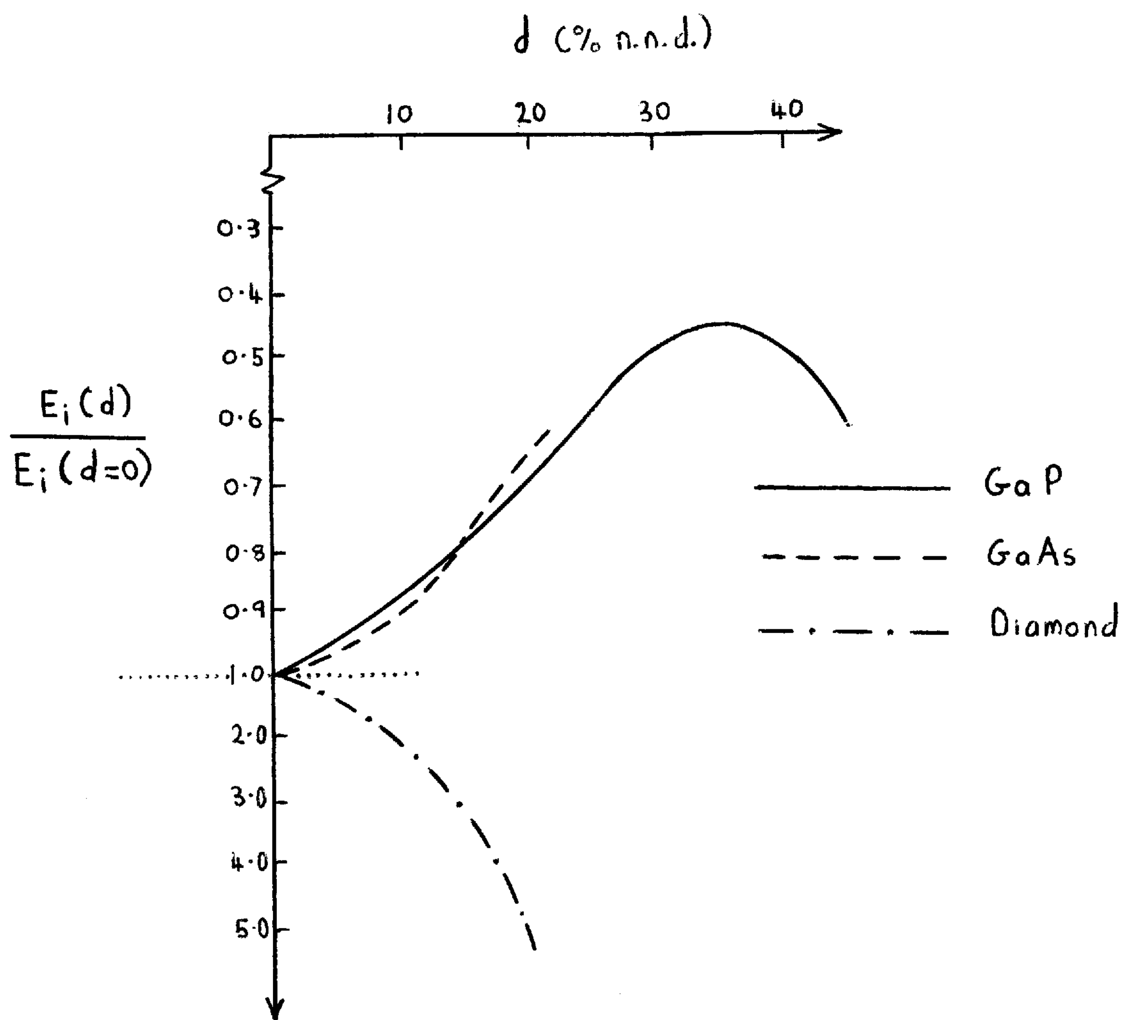
Energy from conduction band minimum (eV)

In contrast to our results for the one electron oxygen state we find that the neutral Zn_{Ga}-O acceptor-donor pair should be capable of MPE capture. It can be seen from Figure 5.3.1 that the electron binding energy for this centre is a sensitive function of displacement. Additionally, the level is fairly close the the conduction band to begin with. Therefore, a relatively small displacement is necessary in order that the conduction band edge crossing may occur and thus we expect a large MPE capture cross section. Consultation of Figure 5.1.2 shows that this prediction is confirmed by the experimental data. Theoretical calculations indicate that MPE capture to form the two electron oxygen state is also possible. In this example, however, the symmetric rather than the trigonal vibration mode is responsible for the energy crossing. Our efforts to obtain numerical agreement with the observed cross sections for the Zn_{Ga}-O and oxygen state-2 impurity systems can be found elsewhere.[88] The agreement is reasonably encouraging in both cases but additional work will be necessary before really reliable quantitative theoretical results can be expected.

In addition to our efforts in relation to the one electron impurity level in GaP a comparitive study utilizing the same potential in alternative materials has been made. In particular, we have considered gallium arsenide and diamond, similar and dissimilar materials to GaP respectively. The potential characteristic of the oxygen impurity in GaP has been used as an input for impurity level calculations in GaAs and diamond. The purpose of the study is not necessarily to relate the results obtained to any particular impurity (i.e. oxygen). Our aim is rather to investigate the role of the host crystal in determining the general properties, symmetry, capture processes etc. associated with typical deep donors in these materials. The substitutional site impurity level is consistently found in the upper half of the band gap. Not unexpectedly, the results for GaAs are quite similar to those for GaP as can been seen in Figure 5.3.4. The O_{As} and O_{P} impurity potentials are not

FIGURE 5·3·4

Relative impurity energy as a function of displacement
(energy is measured from conduction band of host material)



very different and so in this case we may fairly reliably relate our results to what we would actually expect for substitutional oxygen in GaAs (this does not apply for diamond). Again, we find that the energy is a minimum at the substitutional site thus implying high symmetry for the impurity centre. Although the GaAs E_c - 0.75 eV level (associated with the curve labelled $\sigma_n(0)$ in Figure 5.1.2) which has been observed by a number of experimentalists has previously been thought to be due to isolated oxygen, the high non-radiative capture cross section cannot be reconciled with our theoretical findings. The theoretical results indicate that this one electron state should not exhibit a large MPE capture cross section. There have also been some doubts regarding the precise indentification of this level from experimental sources. See for example, the report by Tyler et al.[89]

For diamond we obtain results considerably different from those for GaP and GaAs. In this case the depth of the impurity level is a rapidly increasing function of trigonal displacement as indeed it was for the nitrogen impurity in the previous section. It seems that this purely covalent material responds in a fundamentally different manner to the presence of deep donors. The substitutional site appears to be an inherently unstable configuration and we therefore predict substantial lattice relaxation.

Section 5.4 Summary

In this chapter we have used our localized defect scheme to model the displacement dependent behaviour of various impurity energy levels. Unfortunately, at this stage we cannot produce any reliable quantitative results for the necessary theory which would allow us to do so is not available. Nevertheless, even at the present "computational experiment" level valuable general information can be obtained. The results concerning the symmetry and overall position of the nitrogen impurity level in diamond and the MPE predictions are examples of the possible applications of the method.

In addition, studies of the temperature broadening of impurity levels may be considered. Our results indicate that any convincing theory of MPE capture will need to take into account the symmetry of the localized state for this factor can be of crucial importance in determining the likelyhood of the non-radiative transition.

CHAPTER 6

GENERAL ASSESSMENT OF METHOD

General assessment of method

Section 6.1 Introduction

In the previous chapters we have shown that our localized defect scheme can be usefully applied to a number of problems. The specific results of the calculations have already been discussed and we shall not here be concerned with their reiteration. However, little has been said with regard to the practical advantages of our method. There are, of course, numerous alternative methods which are available and it is instructive to make a direct comparison with some of these. By this means we shall demonstrate the particular areas of superiority and also the weak points of the present scheme. Finally, we shall bring together some of the general results which have been obtained while using this method with may be of some interest.

Section 6.2 Comparison with other methods

At the outset it should be said that it is not intended to comprehensively review the various defect energy level methods. The aim is merely to give comparison with some of the alternatives. No attempt shall be made to compare our method with EMT type approaches which are certainly better suited to dealing with defect states dominated by long range potentials. Nor shall we discuss the many-electron CK type schemes which do not try to relate the resulting energy levels to the band edges. We shall consider two approaches similar to our own in that they require initial knowledge of the host crystal bandstructure, and also two of the defect molecule cluster methods. We begin with brief descriptions of these four formulations.

a) The Koster-Slater Wannier function method

In the Wannier function (W.F.) representation we express the defect wave function as

$$|\psi\rangle = \sum_{n,\underline{R}} W_n(\underline{R}) \; | \; w_n(\underline{r}-\underline{R}) \rangle \qquad (6.2.1)$$

where n and $\underline{R}$ refer to band index and lattice site respectively and the localized W.F.'s are given by

$$|w_n(\underline{r}-\underline{R})\rangle = (\Omega/8\pi^3)^{1/2} \int_{BZ} d\underline{k}\ e^{i\underline{k}.\underline{R}} |\varphi_{n,\underline{k}}\rangle \qquad (6.2.2)$$

Ω is the primitive cell volume. The W.F.'s satisfy

$$\langle w_{n'}(\underline{r}-\underline{R}')|\ w_n(\underline{r}-\underline{R})\rangle = \delta_{n',n}\ \delta_{\underline{R}',\underline{R}} \qquad (6.2.3)$$

Inserting (6.2.1) into the Schrödinger Equation (3.2.1) and multiplying by $\langle w_{n'}(\underline{r}-\underline{R}')|$ we obtain

$$\sum_{n,\underline{R}} W_n(\underline{R}) h_{n',n}(\underline{R}',\underline{R}) + \sum_{\underline{R}} \left[E_{n'}(\underline{R}'-\underline{R}) - \delta_{\underline{R}',\underline{R}}\ E\right] W_{n'}(\underline{R}) = 0 \qquad (6.2.4)$$

where

$$h_{n',n}(\underline{R}'-\underline{R}) = \langle w_{n'}(\underline{r}-\underline{R}')\ |\ h\ |\ w_n(\underline{r}-\underline{R})\rangle \qquad (6.2.5)$$

and

$$E_{n'}(\underline{R}'-\underline{R}) = (\Omega/8\pi^3)^{1/2} \int_{BZ} d\underline{k}\ \ E_{n',\underline{k}} e^{i\underline{k}.(\underline{R}'-\underline{R})} \qquad (6.2.6)$$

The basic problem is to find the energy, E, such that the system of equations (6.2.4) is satisfied. If the defect potential h, is well localized then $h_{n',n}(\underline{R}',\underline{R}) \rightarrow 0$ for sites increasingly removed from the defect centre. Thus W.F.'s on only a few sites need to be taken into consideration. The complexity of the associated secular equation is correspondingly reduced. In the most extreme approximation only one W.F. derived from a single site and band is taken into account. The method is not really suitable for dealing with problems involving long range defect potentials for W.F.'s on many sites would then be necessary. The main problem with the W.F. representation is in the calculation of the matrix elements (6.2.5) which requires a double $\underline{k}$-space integration. The integrals must be calculated numerically and as the sampling

density increases the computational labour becomes very substantial.

Various Wannier function calcutions have been reported but none have conclusively demonstrated convergence in terms of number of lattice sites, bands and sampling. In such cases there must remain some doubt concerning the reliability of the final result.

b) Direct solution of the Bloch function formulation.

The present method as described in Chapter 3 and the Wannier function formulation, a), both rely upon the introduction of sets of localized functions in order to hasten convergence by limiting the size of the associated matrices. However, the method employed by Jaros and Ross [23,24] avoids such intermediate stages and the systems of equations (3.2.4) is solved directly by numerical means. The advantages of this method are that it can more conveniently deal with long range potentials and the scattered (as opposed to bound) states. When the defect energy level, E, is found, the associated eigenvector immediately yields the Bloch function coefficients, $A_{n,k}$. The dimension of the matrices involved in such calculations depends upon the product of sampling point number and number of energy bands. It is apparent that the size of the matrix grows alarmingly as additional bands and sampling points are included. As a consequence, it is the availability of computer resources which ultimately restricts the application of this approach.

c) The extended Hückel theory method

In the EHT method the host material is represented by a relatively small cluster of atoms (usually less that 50). The one-electron wave functions are expanded in the form

$$|\psi_i\rangle = \sum_{\beta} a_{i\beta} |\beta\rangle \qquad (6.2.7)$$

where the summation is over all valence atomic orbitals for each atom in the cluster. The atomic orbitals can be taken as simple Slater-type functions. We then solve the secular determinantal equation

$$\left| H_{\alpha\beta} - E_i \langle\alpha|\beta\rangle \right| = 0 \qquad (6.2.8)$$

where

$$H_{\alpha\beta} = 1/2K(I_\alpha + I_\beta)\langle\alpha|\beta\rangle. \qquad (6.2.9)$$

The I are the free atom ionization potentials and K is a constant, usually taken to be 1.75 (except for $\alpha = \beta$ in which case K = 1). It is hoped that as the cluster is enlarged an increasingly better representation of the bulk solid will be achieved. Identification of the defect states is made by comparison of the results for the "perfect crystal" with those obtained when the appropriate defect atom(s) is substituted (omitted).

The basic appeal of the EMT method lies in it's simplicity for the calculations are fairly straightforward to perform. However, although it may be possible to obtain qualitatively correct results in some cases, the EHT method must be regarded with some suspicion. There is no theoretical justification for the assignment of the value of K and in addition there are theoretical inconsistencies in the formulation. It is known that unreliable predictions are likely to be obtained for ionic systems. In practice, the results are sensitive to cluster size, boundary conditions and choice of orbital exponent:the positions and ordering of the energy levels can be significantly altered by changes in these factors. Extensions of the basic method such as the molecular unit cell superlattice approach (MUCA) have been attempted but it is difficult to imagine the development of a truly quantitative EHT defect scheme. The reader is directed to the works of Larkins,[85] Coulson[90] and Lidiard[91] for more detailed discussion of the limitations of EHT.

d) The $X\alpha$ scattered wave method[55]

A more mathematically satisfying cluster approach is that of the so-called $X\alpha$ scattered wave method. In this approach the perfect and defect crystals

are modelled by limited size molecules partitioned into atomic, interatomic and extramolecular regions. The initial wave functions are taken to be those associated with the free atoms. Within each region of the system a suitable spherically averaged or volume averaged potential is calculated and the one-electron Schrödinger equation is solved by numerical means. Matching of the resulting wave functions is enforced at the boundaries of the various regions. The new molecular wave functions are then used as a starting point for further calculation and iteration is continued until self-consistency is achieved. To account for the effects of exchange the $X\alpha$ approximation, in which the exchange potential is dependent upon local charge density only, is employed.

In practice the $X\alpha$ SW method fails to adequately represent the host crystal for the observed band gap is not reproduced. Therefore, although interesting general information may be obtained, it cannot be expected that this method will give reliable quantitative results.

Traditionally, methods requiring full knowledge of the host crystal such as a) and b) have involved considerable utilization of computer resources. As a consequence, the application and flexibility of such approaches have been somewhat limited. In addition, none of the calculations based upon these methods have previously demonstrated the capability of producing a truly convergent result. It is thus not surprising that much effort has been redirected towards the cluster-type methods. However, as we have tried to indicate, methods of this kind such as c) and d) have their own associated problems. The finite cluster size and the significance of the related surface effects simply cannot be overlooked. The advantages of convenience and possible incorporation of self-consistency do not outweigh the resulting disadvantages. It would seem that none of the methods a)-d) is a generally viable means of dealing with the problem of localized defect energy levels in

semiconductors. They are too prohibitive in the use of computer time to implement or alternatively provide an inadequate representation of reality.

We have argued that the other methods are far from being ideal in treating the localized defect problem but what, it may be asked, have we to gain by employing the method we have used? It is apparent that many of the details of our approach are similar to those of a) and b). It is obvious that the input, the bandstructure and real space defect potential, is an identical requirement for all three approaches and is thus subject to the same degree of error. The fairly inconvenient form of the defect wave function and the lack of a truly self-consistent procedure are features common to all. The crucial difference is in the computational labour. The method we use is much less demanding in this respect and thus many advantages result. As an example if we were to calculate the energy of the a_1 level associated with O_p in GaP we would require about 20 seconds of central processing unit time on an IBM 370/168 computer using our method. (This is for a calculation involving 41 sampling points in the 1/48 BZ, 10 energy bands, Bloch functions consisting of a 65 plane wave expansion, 10 basis functions, g_m and 10 trial energies.) In comparison, literally hours of CPU time would be required to perform the equivalent calculation using the direct Bloch function method, b). The lower symmetry calculations such as those involved in the displacement studies of Chapter 5 require approximately 3 mins. of CPU time. Such calculations would be all but impossible using method b). The Koster-Slater approach is similarly profligate in it's demands. Due to the speed of the calculation it is a trivial matter to demonstrate that our final result is convergent in terms of number of basis functions and also number of bands. Additional sampling points, bands and basis functions can easily be incorporated if desired. Convergence has never been satisfactorily demonstrated by users of methods a) and b) and usually much more restrictive conditions that we have employed have been necessary in order to make the calculations feasible.

It is thus quite clear that our method is superior to a) and b). We can see little reason to favour the cluster approaches such as c) and d) for, as we have already pointed out they do not at present appear to promise quantitative results. (It should be said, however, that the cluster methods do provide more convenient real space defect wave functions.) The previous support for calculations of this type has primarily been due to their low CPU demands but, as we have seen for the simpler problems at least, this is also a characteristic of our present method. With this in mind there seems little justification for pursuing the more time-consuming MUCA calculations.

Section 6.3 Concluding remarks

In this account an attempt has been made to study various aspects of the localized defect problem while simultaneously demonstrating the applicability of our basic method. Although little in the way of quantitative results has been presented this is to some extent a consequence of the particular investigations which have been undertaken. It can be expected that much additional work will be needed before the vacancy problem is satisfactorily resolved. While not as yet being able to completely adequately tackle problems such as this useful information can, nevertheless, be obtained. In a more well defined situation our method can be expected to be more successful. For instance, in the case of a given short range defect potential we can in principle accurately compute the associated one-electron energy level provided that reliable bandstructure is provided as an input.

Our calculations tend to reinforce the belief that both valence and conduction bands have an important role in determining the characteristics of localized states. Therefore, it is unreasonable to think, generally speaking, that formulations based on EMT will give realistic results. The precise position of the energy level associated with a localized state depends upon a delicate balance between the contributions of the various bands. If this balance is changed then significant alteration of the binding energy can

occur. However, there does not seem to be a correspondingly great alteration in the details of the localization of the defect wave function (it would seem that the angular and nodal characteristics of the defect wave functions are mainly responsible for determining the observable properties of localized states). Thus the practice of relating the depth of the energy level to the localization, as can be done with states of coulombic origin, is not appropriate in the case of "deep" levels. If we employ the coulombic model then the localized state wave function collapses and merges into the delocalized band edge Bloch function as the binding energy (measured from the appropriate band edge) decreases. However, if we merely change the balance between the valence and conduction bands then no such drastic alteration of wave function occurs and the existence of well localized states resonant with the conduction band is possible. Defect centres giving rise to electronic states of this type may exhibit interesting properties The persistent photoconductivity of the so-called DX centres in $Al_xGa_{1-x}As$ may be an example. This effect has been explained by Lang and Logan[92] on the basis of a large-lattice-relaxation model involving a localized state with an unoccupied energy level resonant with the conduction band and an occupied energy level within the band gap (detailed discussion of the DX centre has been given by Lang, Logan and Jaros[93]).

In conclusion, it can be said that we believe our scheme to be a promising means of investigating the localized defect problem. At present it is not possible to obtain really good defect wave functions but it is hoped that increased effort will to some extent improve this situation. One of the main difficulties we encounter in our formulation is in the production of a reliable defect potential although, as we have seen, this does not always greatly affect the results. The simple process of screening the bare potential with the perfect crystal dielectric function probably does not represent an adequate self-consistent precedure. Attempts to produce a more realistic defect

potential are proceeding. It is hoped that with additional work means whereby the more comprehensive treatment of the general problem of localized defects in semiconductors will be developed.

REFERENCES

1. Kohn W. Solid State Physics (Eds. F. Seitz, D. Turnbull) Vol. 5 p257 Academic Press (1957)

2. Milnes A.G. Deep Impurities in Semiconductors Wiley (1973)

3. Coulson C.A., Kearsley M.J. Proc. Royal Soc. A241 p433 (1957)

4. Messmer R.P., Watkins G.D. Radiation Damage and Defects in Semiconductors p255 IOP (1973)

5. Koster G.F., Slater J.C. Phys.Rev. 95 p1167 (1954)

6. Koster G.F., Slater J.C. Phys.Rev. 96 p1208 (1954)

7. Callaway J., Hughes A.J. Phys.Rev. 156 p860 (1967)

8. Jaros M. J.Phys. C8 pL550 (1975)

9. Slater J.C. Phys.Rev. 81 p385 (1951)

10. Kronig R., Penney W.G. Proc.Royal Soc. A130 p499 (1931)

11. Herring C. Phys.Rev. 57 p1169 (1940)

12. Phillips J.C., Kleinman L. Phys.Rev. 116 p287, p880 (1959)

13. Cohen M.H., Heine V. Phys.Rev. 122 p1821 (1961)

14. Austin B.J., Heine V., Sham L.J. Phys.Rev. 127 p276 (1962)

15. Harrison W.A. Pseudopotentials in the Theory of Metals, Benjamin (1966)

16. Cohen M.L., Bergstresser T.K. Phys.Rev. 141 p789 (1966)

17. Saslow W., Bergstresser T.K., Cohen M.L. Phys.Rev.Letts. 16 p354 (1966)

18. Reitz J.R. Solid State Physics (Eds. F. Seitz, D. Turnbull) Vol.1 p1 (1955)

19. Heine V. et alia Solid State Physics (Eds. H. Ehrenreich, F.Seitz, D. Turnbull) Vol.24 (1970)

20. Heine V. Group Theory in Quantum Mechanics Pergamon Press (1960)

21. Tinkham M. Group Theory and Quantum Mechanics McGraw-Hill (1964)

22. Bradley C.J., Cracknell A.P. The Mathematical Theory of Symmetry in Solids Clarendon Press (1972)

23. Jaros M., Ross S.F. J.Phys. C6 p1753 (1973)

24. Jaros M., Ross S.F. J.Phys. C6 p3451 (1973)

25. Bassani F., Iadonisi G., Preziosi B. Phys.Rev. 186 p735 (1969)

26. Bassani F., Iadonisi G., Preziosi B. Reps. on Prog. in Phys. 37 p1099 (1974)

27. Ross S.F. Ph.D. thesis unpublished (1975)

28. Cohen M.L., Heine V. Solid State Physics (Eds. H. Ehrenreich, F. Seitz, D. Turnbull) Vol. 24 p37 (1970)

29. Vinsome P.K.W. Ph.D. Thesis unpublished (1971)

30. James H.M., Lark-Horovitz K. Z.Physik Chem. 198 p107 (1951)

31. Blount E.I. Phys.Rev. 113 p995 (1959)

32. Yamaguchi T. J.Phys.Soc. Japan 17 p1359 (1962)

33. Jahn H.A., Teller E. Proc.Royal Soc. A161 p220 (1937)

34. Engleman R. The Jahn-Teller Effect in Molecules and Crystals Wiley (1972)

35. Watkins G.D., Corbett J.W., Walker R.M. J. Appl.Phys. 30 p1198 (1959)

36. Watkins G.D., Corbett J.W. Phys.Rev. 121 p1001 (1961)

37. Corbett J.W., Watkins G.D., Schrenko R.M., McDonald R.S. Phys.Rev. 121 p1015 (1961)

38. Watkins G.D., Corbett J.W. Phys.Rev. 134 p1359 (1964)

39. Watkins G.D. J.Phys.Soc. Japan 18 Suppl.II p22 (1961)

40. Watkins G.D. Radiation Damage in Semiconductors p97 DUNOD (1964)

41. Corbett J.W., Watkins G.D. Phys.Rev.Letts. 7 p314 (1961)

42. Watkins G.D., Corbett J.W. Phys.Rev. 138 p543 (1965)

43. Elkin E.L., Watkins G.D. Phys.Rev. 174 p881 (1968)

44. Larkins F.P. J.Phys.Chem.Solids 32 p2123 (1971)

45. Larkins F.P. J.Phys. C6 pL345 (1973)

46. Larkins F.P. J.Phys.Chem.Solids 32 p965 (1971)

47. Watkins G.D. Lattice Defects in Semiconductors p1 IOP (1975)

48. Stoneham A.M. Theory of Defects in Solids OXFORD (1975)

49. Callaway J., Hughes A.J. Phys.Rev. 164 p1043 (1967)

50. Parada N.J., Pratt G.W. Phys.Rev.Letts 22 p180 (1969)

51. Parada N.J. Phys.Rev. B3 p2042 (1971)

52. Coulson C.A., Larkins F.P. J.Phys.Chem.Solids 32 p2245 (1971)

53. Coulson C.A., Larkins F.P. J.Phys.Chem.Solids 30 p1963 (1969)

54. Hoffman R. J. Chem.Phys 39 p1397 (1963)

55. Johnson K.H. Adv.Quantum Chem. 7 p143 (1973)

56. Messmer R.P., Watkins G.D. Phys.Rev.Letts 25 p656 (1970)

57. Messmer R.P., Watkins G.D. Phys.Rev. B7 p2568 (1973)

58. Lee T.F., McGill T.C. J.Phys. C6 p3438 (1973)

59. Hemstreet L.A. Phys.Rev. B11 p2260 (1975)

60. Hemstreet L.A. Phys.Rev. B12 p1212 (1975)

61. Louie S.G., Schluter M., Chelikowski J.R., Cohen, M.L. Phys.Rev. B13 p1654 (1976)

62. Surratt G.T., Goddard W.A. S.S. Comms. 22 p413 (1977)

63. Jaros M., Brand S. Phys.Rev. B14 p4494 (1976)

64. Jaros M., Srivastava G.P. J.Phys.Chem.Solids 38 p1399 (1977)

65. Jaros M. Phys.Rev. B16 p3694 (1977)

66. Jaros M. J.Phys. C8 p2455 (1975)

67. Lowther J.E. J.Phys. C9 p2519 (1976)

68. Il'in N.P., Masterov F. Sov.Phys.Semic. 10 p496 (1976)

69. Chiang S.Y., Pearson G.L. J.Appl.Phys. 46 p2986 (1975)

70. Hasegawa F., Majerfeld A. Elect.Letts. 12 p52 (1976)

71. Lang D.V. J.Appl.Phys. 45 p3023 (1974)

72. Lang D.V., Logan R.A., Kimerling L.C. Phys.Rev. B15 p4874 (1977)

73. Lax M. Phys.Rev. 119 p1502 (1960)

74. Bonch-Bruevitz V.L., Glasko V.B. Sov.Phys. - Sol.State 4 p371 (1962)

75. Huang K., Rhys A. Proc.Royal Soc. A204 p406 (1950)

76. Henry C.H., Lang D.V. Phys.Rev. B15 p989 (1977)

77. Lang D.V., Kimerling L.C. Phys.Rev.Letts. 33 p489 (1974)

78. Kimerling L.C., Lang D.V. Lattice Defects in Semiconductors p589 IOP (1975)

79. Lang D.V., Kimerling L.C. Appl.Phys.Letts. 28 p248 (1976)

80. Dean P.J., Choyke W.J. Adv. in Phys. 26 p1 (1977)

81. Smith W.V., Sorokin P.P., Gelles I.L., Lasher G.L. Phys.Rev. 115 p1546 (1959)

82. Loubser J.H.N., Du Preez L. Brit.J.Appl.Phys. 16 p457 (1965)

83. Jones D., Lettington A.H. S.S. Comms. 11 p701 (1972)

84. Appapillai M., Heine V. Tech.Rep. No.5 Cavendish Lab. Cambridge unpublished (1972)

85. Larkins F.P. J.Phys. C4 p3065 (1971)

86. Farrer R.G S.S. Comms. 7 p685 (1969)

87. Davies G. Chemistry and Physics of Carbon Vol.12

88. Jaros M., Brand S. Proc. of 13th Int.Conf. of Phys.Semic's p1090 (1976)

89. Tyler E.H., Jaros M., Penchina C.M. Appl.Phys.Letts p208 (1977)

90. Coulson C.A. Radiation Damage and Defects in Semiconductors p249 IOP (1973)

91. Lidiard A.B. Radiation Damage and Defects in Semiconductors p238 IOP (1973)

92. Lang D.V., Logan R.A. Phys.Rev.Letts. 34 p635 (1977)

93. Lang D.V., Logan R.A., Jaros M. To be published in Phys.Rev.B.