Downloaded from rsta.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Understanding Defects in Semiconductors: Spin-Off from Technology

G. A. Baraff

Phil. Trans. R. Soc. Lond. A 1992 **341**, 195-202
doi: 10.1098/rsta.1992.0093

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. A* go to:
http://rsta.royalsocietypublishing.org/subscriptions

This journal is © 1992 The Royal Society

Downloaded from rsta.royalsocietypublishing.org

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF

Understanding defects in semiconductors: spin-off from technology

By G. A. Baraff
AT&T Bell Laboratories, Murray Hill, New Jersey 07974, U.S.A.

Defects in semiconductors play a controlling role in determining their technologically useful properties. Particularly important in this regard are defects that introduce levels far from the conduction and valence band edges. Capture or emission of electrons or holes into such levels generally results in displacement of the atoms in the immediate vicinity of the defect, and the action of the defect in determining the semiconductor properties must be understood with this complication in mind. Examples of the role of calculations in establishing some of our currently useful knowledge are presented.

1. Introduction

This paper provides an introduction to the work on deep level defects that constitutes approximately one quarter of this volume. Although the volume is about processes, there is a connection between deep level defects and processes. That connection comes about via phenomena that occur at defects and because of the presence of defects. Defects are dynamic entities that can migrate through the crystal. They can combine with each other in ways similar to the way atoms combine in ordinary chemical processes in gases. Under the influence of external stimuli such as light, or pressure, or exchange of electrons with the environment, they can alter the arrangement of atoms in their immediate vicinity. These changes give rise to some interesting phenomena, namely, large lattice relaxation, negative U, metastability and persistent photoconductivity (Pantelides 1986).

There is an important link between technologically important issues and the study of point defects. Some of this will be mentioned in §2. In §3, I review some basic ideas that underlie what is done in trying to calculate the properties of defects. This is in the nature of an elementary tutorial but it contains notions which, although understood by practitioners in the field, have not been spelled out for a general audience. Section 4 will briefly mention how progress in the ability to calculate defect properties has had an effect on the way we think of them.

2. Technologically driven problems

(a) *Diffusion*

Semiconductor devices are made by successively depositing layers of materials of varying composition and doping on a substrate, sometimes etching away parts of various layers or masking off the areas where new material is to be deposited. During processing, dopants will move around if they are mobile at processing temperatures. The dopant atoms themselves are defects in the lattice. Their diffusion may well be

Phil. Trans. R. Soc. Lond. A (1992) **341**, 195–202
Printed in Great Britain © 1992 The Royal Society

7-2

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions: Physical Sciences and Engineering.* STOR® www.jstor.org

Downloaded from rsta.royalsocietypublishing.org

mediated by other defects; vacancies, for example, with which they can trade places, or native interstitials, which eject the dopants from their lattice sites. The energetics involved in these defect reactions control the diffusion. There is therefore interest in being able to calculate the energetics so as to improve diffusion modelling, essential in the design of better manufacturing processes.

(b) *Persistent photoconductivity and the DX centre*

Aluminium gallium arsenide alloy, or AlGaAs, is an important semiconductor material in devices where high speed, low noise, low power consumption and generation or detection of light may be important. In AlGaAs, the DX centre is the lowest energy state of the donor atom when the aluminium fraction exceeds 22% (Mooney 1991), and thus this centre determines the conductivity of the material. Yet, at low temperatures, one cannot count on knowing the conductivity of the material: the DX level has large barriers to both electron capture and emission. At low temperatures, these barriers are large enough that the free electron concentration does not come into equilibrium with the donors and the conductivity of the material then depends on its thermal history. Even worse, persistent photoconductivity is observed at low temperature. When the sample is exposed to light at low temperature, the DX centres are ionized. However, when the light is turned off, the electrons remain in the conduction band for minutes, hours, or days, depending on the temperature and alloy concentration. This is clearly a disaster, both for a photodetector, and for other devices such as the MODFET (a modulation doped field effect transistor). Both of these aspects, the high capture barriers and the persistent photoconductivity, are known to result from the fact that the defect changes its geometric configuration, that is, it undergoes a large lattice relaxation, when its charge state changes. Exactly what the defect is and how it causes these peculiar effects were major intellectual mysteries and major technological annoyances until the theoretical work carried out by Chadi & Chang (1989).

In the case of the MODFET, the operating characteristic (the current versus the potential difference) is determined in large part by the amount of band bending. This is affected in turn by the spatial distribution of ionized donors. At low temperatures, these donors cannot easily recapture their electrons, and thus the operating characteristics of the MODFET changes because the donors are DX centres.

(c) *Production of semi-insulating gallium arsenide and EL2*

When devices are fabricated, material is deposited on a substrate. That substrate has to be semi-insulating to insulate properly the various parts of a device from each other. Equally important, where many transistors are on a single monolithic substrate, a semi-insulating substrate is needed to insulate the devices from each other.

Gallium arsenide, as grown, often contains shallow acceptor levels. Their energy is only slightly above the top of the valence band, and so they are easily occupied by electrons from that band, leaving behind holes that give the material p-type conductivity. However, if mid-gap donors are introduced in greater concentration than the acceptors, they compensate the acceptor levels, filling them all with electrons. Under these conditions, there can be no electrons in the conduction band (they would be captured by the empty mid-gap levels) nor holes in the valence band (they would be filled by electrons from the occupied mid-gap levels). The material becomes semi-insulating.

Downloaded from rsta.royalsocietypublishing.org

GaAs has been doped with chromium to make it semi-insulating. However, at the temperatures used in device fabrication, chromium is mobile. This is unacceptable when it is necessary to make some regions semi-insulating, and other regions n-type or p-type. Thus, chromium is unsuited for this purpose. But other mid-gap donors are known. EL2 is one such midgap donor (Martin & Makram-Ebeid 1986). It is immobile at device processing temperatures. It does happen to be metastable at low temperatures. Unlike DX, its metastability is only optically driven. It is suited for use in producing semi-insulating GaAs. Aside from its metastability, which is a fascinating topic, there was technological need to understand the microscopic structure of EL2. The need arose because technologists, although confident that they could make wafers with reproducible EL2 concentration, feared that if there were variations in the trace compositions of the source materials used in production, their recipes for producing wafers with uniform EL2 concentration might fail. Knowing the microscopic identity of EL2 would provide confidence that the recipes could be changed if the source material changed.

(*d*) *Degradation*

Since the properties of devices are controlled by the positions of the dopant atoms, it is necessary that they stay where they are first put, not only during manufacture, but also while the device is in use. The presence of free carriers can cause enhanced mobilities and can cause defect reactions to occur, even at temperatures where these processes would not normally be important. Electrons in the conduction band, for example, liberate energy when they are captured by a deep-lying defect level, and this energy can substitute for the energy that would otherwise have to be supplied thermally. Thus, a device that has a long life on the shelf can fail in use. One needs to understand these enhanced processes as well. This understanding can be aided by precise calculations of the energetics of defect reactions.

3. Basic ideas in deep level studies

(*a*) *Tools and approximations*

The generic problem is to describe the quantum mechanical state of a defect in an otherwise perfect crystal. The defect is an atom or several atoms that are wrong for the crystal, either in their chemical identity, or their location, or both. The relevant variables are the positions of all the electrons in the system and all the atomic cores in the system, but almost universally the Born–Oppenheimer approximation is brought in at this stage. Density-functional theory is most often used to describe how the electrons interact with each other, and norm-conserving pseudopotentials are used to describe how the electrons interact with the atomic cores. Finally, the region of the crystal near the effect is singled out for special attention. One way is by using Green's function. This is a technique that allows one to do a complete calculation even though the region in which the wave functions must be computed is limited to the region around the defect where the potential differs from that of the perfect crystal (Koster & Slater 1954; Baraff & Schluter 1979; Bernholc *et al.* 1980; Kelly & Car 1992).

Another way that has recently been resurrected is the large unit-cell method, where a periodic array of defects is considered. Each defect in the array is surrounded by a unit cell of perfect crystal, and these unit cells are chosen to be large enough that a defect in one unit cell is not influenced by the defect in the next cell. The shift in

Downloaded from rsta.royalsocietypublishing.org

preferred calculational method from large unit cell to Green's function back to large unit cell (either alone or used with a Green's function) was driven both by the changes in hardware and software: recent advances in calculational techniques for the large unit cell problem, such as Car–Parrinello and conjugate gradient techniques (Car & Parrinello 1985; Stitch *et al.* 1989) were used in defect calculations as well. The results of such calculations are wave functions for the various electronic defect states, and a total energy. This total energy, in the Born–Oppenheimer approximation, depends on the chemical potential μ, on the positions of the atomic cores R_i, on the number of electrons N occupying the defect levels, and on j, the state of excitation of the defect.

(b) *Levels*

There are many uses of the word 'level' in quantum mechanics. An important use is in the discussion of defects, where a level is the difference in ground state energies between two systems that are identical, except that the second contains one more electron than the first. This use is important because it defines the conditions under which a defect will be found in one or another equilibrium charge state. To see how this comes about consider the probability of finding a system with a given number of particles as given by the rules of equilibrium statistical mechanics. The system we are considering is the defect, and the occupancy of any or all of the localized states associated with it. The probability P that at temperature T and chemical potential μ the system will be found to contain N electrons is given by

$$P(N,T) = Z^{-1} g_N \exp[-(E_N - \mu N)/kT], \tag{1}$$

where
$$Z \equiv \sum_{N=0} g_N \exp[-(E_N - \mu N)/kT]$$

is the partition function. The ground state energy for the system containing N particles is E_N and the degeneracy factor is g_N. The ratio of probabilities that the system has two electrons rather than one is

$$\frac{P(2,T)}{P(1,T)} = \frac{g_2 \exp[-(E_2 - 2\mu)/kT]}{g_1 \exp[-(E_1 - \mu)/kT]}$$
$$= (g_2/g_1) \exp\{-[(E_2 - E_1) - \mu]/kT\}$$
$$= (g_2/g_1) \exp[-(\epsilon_2 - \mu)/kT], \tag{2}$$

where
$$\epsilon_2 \equiv E_2 - E_1. \tag{3}$$

(We assume here that excited states are located at appreciably higher energies. If they are not, it is trivial to include them in the argument being made.) The ratio of probabilities is thus given by an expression which involves a comparison between μ, the chemical potential, and the level ϵ_2, defined as the difference between the two ground state energies. If $\mu < \epsilon_2$, this ratio is nearly zero at low temperature and the defect will be found to have zero or one electron. Conversely, if $\mu > \epsilon_2$, the ratio will be nearly infinite and the effect will be found to be occupied by at least two electrons. This idea is illustrated in figure 1*a*, which shows the levels associated with an isolated vacancy in silicon, at least as they were understood before 1979 (Watkins *et al.* 1979).

The levels shown as lying between the conduction band and the valence band denote those values of μ for which the occupancy of the defect changes. The regions

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Downloaded from rsta.royalsocietypublishing.org

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

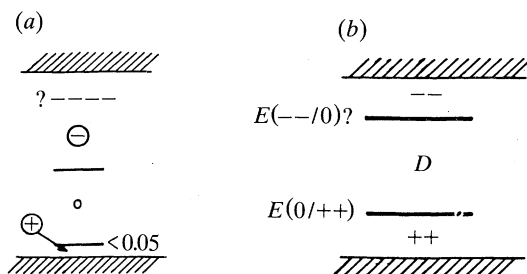

Figure 1. (*a*) Energy levels for the silicon vacancy as they were understood before 1979. (*b*) Energy levels for the silicon vacancy after 1979.

between the levels are labelled by the equilibrium charge state of the vacancy when the chemical potential is anywhere in the region.

(*c*) *Two electron levels: negative U*

Since 1980, it has become clear that the actual situation for the silicon vacancy is as shown in figure 1*b*, namely, if μ is below the lowest level, there are no electrons on the defect and its charge is $+2$, while if μ is above the lowest level, then there are two electrons on the defect and it is neutral (Newton *et al.* 1983). In this case, μ determines an equilibrium occupancy of either zero or two electrons, and this is said to be a 'two electron level'. There is no value of μ for which the system (in equilibrium with a reservoir of charge) will be found with one electron.

To understand how this situation is described by statistical mechanics, consider a system that can hold zero, one or two electrons. The levels are still defined as the differences in ground state energies:

$$\epsilon_1 \equiv E_1 - E_0, \quad \epsilon_2 \equiv E_2 - E_1. \tag{4a, b}$$

It usually takes more energy to add the second electron than the first because the second one is repelled by the one already there. So if we define

$$\epsilon_2 = \epsilon_1 + U \tag{5}$$

then U, the additional energy needed to add the second electron, is expected to be of the order of $e^2/\epsilon R$, where R is the spatial extent of the wave function for the added electron and ϵ is the dielectric constant of the crystal. The partition function in this case takes the form

$$\begin{aligned} Z &= g_0 \exp(-E_0/kT) + g_1 \exp[-(E_1-\mu)/kt] + g_2 \exp[-(E_2-2\mu)/kt] \\ &= g_0 \exp(-E_0/kT)\,[1 + (g_1/g_0)\exp[-(\epsilon_1-\mu)/kT] + (g_2/g_0)\exp\{-2[\bar{\epsilon}-\mu]/kT\}], \end{aligned} \tag{6}$$

where
$$\bar{\epsilon} \equiv \tfrac{1}{2}(\epsilon_1+\epsilon_2). \tag{7}$$

Two energies determine which of the three terms is large: ϵ_1 and $\bar{\epsilon}$. Occupancy will be decided by the placement of μ relative to ϵ_1 and $\bar{\epsilon}$. Although it most often happens that $\epsilon_2 > \epsilon_1$, if the situation is reversed, then by definition of (5), it follows that $U < 0$. This indicates that something other than electrostatic repulsion is active in determining the relative energy needed to add successive electrons. It will also follow, from (4), that

$$2E_1 > E_0 + E_2. \tag{8)}$$

Downloaded from rsta.royalsocietypublishing.org

This inequality is what Chadi & Chang (1989) calculated to be the case for the DX defect. In the case of DX, the neutral defect is the one with a single electron, while zero or two electrons correspond to a charge of plus or minus one respectively. In the case of DX, even without a reservoir of electrons to equilibrate with, a collection of neutral defects that are able to exchange electrons among themselves will have a lower energy if half of them have two electrons and the other half have zero electrons. Experiments to establish whether or not this is true for DX have occupied great attention recently and there is now general consensus that this is true.

Another consequence follows from the ordering $\epsilon_2 < \epsilon_1$. In this case, from (7), we have $\bar{\epsilon} < \epsilon_1$. Then, at low temperatures, $P(0,T) \approx 1$ for $\mu < \bar{\epsilon}$, and $P(2,T) \approx 1$ for $\mu > \bar{\epsilon}$, and $P(1,T) \approx 0$ for all μ. The system will, depending on the relative positions of μ and $\bar{\epsilon}$, contain either zero or two electrons if it is in contact with an external reservoir. This describes the two-electron level. Occupancy of the defect changes by two electrons as μ is raised. The silicon vacancy, with its two electron level, is sometimes referred to as being a 'negative U' system, because the two concepts, negative U and two-electron level, are identical.

(d) *Large lattice relaxation and negative U*

Lattice relaxation is the cause of negative U. A connection between the two can be illustrated with a very simple model. Suppose that there is only one active lattice coordinate, say Q. Suppose further that two electrons can go into the same spatial electronic wave function, one with spin up and the other with spin down. Assume that each electron exerts the same force on the coordinate, a coordinate which, in the absence of electrons, is acted upon by a harmonic oscillator potential. These conditions are what one would arrive at by making the very lowest order expansion of the energy, regarding lattice distortion as the small parameter. Then, as shown by Baraff *et al.* (1980), the energy to place the first electron on the defect is $\epsilon_L - E_{JT}$, where E_{JT} is the energy of lattice relaxation and ϵ_L is the energy which would have been needed if the lattice did not relax. The energy to put the second electron on is $\epsilon_L + U - 3E_{JT}$, where U is the added electrostatic repulsion. The significant physics here is that the amount of lattice displacement (as measured from the situation with no electrons in place) caused by adding two electrons is twice that of adding one. Therefore the energy lowering (also measured from the situation with no electrons in place) caused by two electrons is four times that of caused by one electron. The difference between energies to add the first and second electrons is then

$$\epsilon_2 - \epsilon_1 \equiv U_{\text{eff}} = U - 2E_{JT}.$$

It is this U_{eff} that is negative for the silicon vacancy.

The simple model here represents the total energy in all states of the system as being parabolas, much as virtually all configuration coordinate diagrams were constructed before 1985. Such simple diagrams can only be valid for relatively small displacements. When there is really large lattice distortion, the simple parabolas are of no use. One must instead use diagrams in which energies have been calculated in a serious manner. An early first effort in this direction was contributed by Baraff & Schluter (1985) who considered the total energy for various charge states of the gallium vacancy in GaAs. By motion of a single arsenic neighbour of the vacancy, this defect can be transformed into a nearest neighbour arsenic-vacancy–arsenic-antisite pair. The configuration coordinate on which the energy depended was taken to be the position of the displaced arsenic nearest neighbour atom. Since 1985, the

Downloaded from rsta.royalsocietypublishing.org

technique for evaluating total energies has progressed to the extent that our current understanding of both DX and EL2 rests firmly on total energies and configuration coordinate diagrams as calculated by Chadi & Chang (1989) and by Dabrowski & Scheffler (1989).

(*e*) *Configuration dependent levels*

In so far as the energies depend on the configuration coordinates, one can also consider energy differences that depend on the configuration. Unlike the other levels, which were defined as the differences between minimum energies and were of use for describing statistical occupations, these levels are useful in reasoning about forces. This comes about as follows. The total energy depends on all the lattice coordinates, and so the force on the ith coordinate is the negative of the gradient of the energy with respect to that coordinate. Consider two forces on the same coordinate, R_i, one force arising in a system with j electrons and the other arising in a system containing $j-1$ electrons. That difference in force can be regarded as the force contributed by adding the jth electron to the system. It is given by the negative gradient of an energy difference. This makes it convenient to define this configuration dependent energy difference as a level:

$$\epsilon_j(Q) \equiv E_j(Q) - E_{j-1}(Q), \quad F_i = -\partial \epsilon_j(Q)/\partial R_i. \tag{9a, b}$$

Interestingly, this type of level is the simplest type to calculate. Apart from minor corrections, it is the eigenvalue of the Schrödinger equation that is solved in the Born–Oppenheimer calculation to get the wave functions and total energies at each configuration of the system. As an example of (9*b*), consider an effective mass level where a state is pulled down a small amount below the bottom of the conduction band by the coulomb attraction to an ionized donor. Such a level is insensitive to the exact location of lattice atoms in the neighbourhood of the donor. Its energy, on a configuration coordinate diagram, is independent of the coordinate. From (9*b*), that independence means that the electron in the state does not exert appreciable force on any of the lattice atoms. But this was already expected: the effective mass state is so spread out in space that its density is low and fairly uniform, so it cannot exert a force. Conversely, the energy levels of a deep level defect are found to be profoundly sensitive to the location of the atoms nearby, as has been shown in all defect calculations.

4. Examples of changing views on defects

Two examples of how progress in evaluating total energies accurately have been important in our understanding of large-lattice relaxation defects have already been mentioned: the DX centre and EL2. An account of how the theoretical calculations for these two defects has changed our fundamental ideas about defects is already in print (Baraff 1990). As this paper must be short there is only space to mention the existence of another article (Baraff 1992). This touches on another aspect of the effect (or lack of effect) of calculations. It cites a case in which a calculation of EL2 predating that of Dabrowski & Scheffler (1989) was ignored at the time it was presented because there was no confidence in the calculational method, even though, in retrospect, it turned out to be correct in concept and result. The EL2 story is still very much open as regards the microscopic identity of the defect. A summary of the present state of understanding appears in Baraff (1992).

New methods for evaluating the forces have emerged recently, and new methods for handling bigger and bigger systems have appeared. These have made it possible

Downloaded from rsta.royalsocietypublishing.org

to determine the atomic arrangements of larger and larger numbers of atoms around the defects. The resulting accuracy has made it possible to learn more about defect properties in solids. These new methods are described in this volume.

References

Baraff, G. A. 1990 The importance of interaction between theory and experiment in understanding and identifying technologically important point defects. In *Defect control in semiconductors* (ed. K. Sumino) pp. 31–40. Amsterdam: North-Holland.

Baraff, G. A. 1992 Update: the mid-gap donor level EL2 in GaAs. In *Deep centers in semiconductors* (ed. S. T. Pantelides). New York: Gordon and Breach. (Revised version of Pantelides (1986).)

Baraff, G. A., Kane, E. O. & Schluter, M. 1980 Theory of the silicon vacancy: an Anderson negative-U system. *Phys. Rev.* B **22**, 5662–5686.

Baraff, G. A. & Schluter, M. 1979 New self-consistent approach to the electronic structure of localized defects in solids. *Phys. Rev.* B **19**, 4965–4979.

Baraff, G. A. & Schluter, M. 1985 Bistability and metastability of the gallium vacancy in GaAs: the actuator of EL2? *Phys. Rev. Lett.* **21**, 2340–2343.

Bernholc, J., Lipari, N. O. & Pantelides, S. T. 1980 Scattering-theoretic method for defects in semiconductors II: Self consistent formulation and application to the vacancy in silicon. *Phys. Rev.* B **21**, 3543–3562.

Car, P. & Parrinello, M. 1985 Unified approach for molecular dynamics and density functional theory. *Phys. Rev. Lett.* **55**, 2471–2474.

Chadi, D. J. & Chang, K. J. 1989 Energetics of DX formation in GaAs and $Al_x Ga_{1-x} As$ alloys. *Phys. Rev.* B **39**, 10063–10074.

Dabrowski, J. & Scheffler, M. 1989 Isolated arsenic–antisite defect in GaAs and the properties of EL2. *Phys. Rev.* B **40**, 10391–10401.

Kelly, P. J. & Car, R. 1992 Green's matrix calculation of total energies of defects in silicon. *Phys. Rev.* B **45**, 6543–6563.

Koster, G. F. & Slater, J. C. 1954 Simplified impurity calculation, *Phys. Rev.* **96**, 1208–1223.

Martin, G. & Makram-Ebeid, S. 1986 The mid-gap donor level EL2 in GaAs. In *Deep centers in semiconductors* (ed. S. T. Pantelides) p. 399. New York: Gordon and Breach.

Mooney, P. M. 1991 Donor related levels in GaAs and $Al_x Ga_{1-x} As$. *Semiconductor Sci. Technol.* **6** (10), B1–B8.

Newton, J. L., Chatterjee, A. P., Harris, R. D. & Watkins, G. D. 1983 Negative-U properties of the lattice vacancy in silicon. *Physica* B **116**, 219–223.

Pantelides, S. T. (ed.) 1986 *Deep centers in semiconductors*. New York: Gordon and Breach.

Stitch, I., Car, R., Parrinello, M. & Baroni, S. 1989 Conjugate gradient minimization of the energy functional: a new method for electronic structure calculations. *Phys. Rev.* B **39**, 4997–5004.

Watkins, G. D., Troxell, J. R. & Chatterjee, A. P. 1979 Vacancies and interstitials in silicon. In *Defects and radiation effects in semiconductors* (ed. J. H. Albany), pp. 16–30, Conf. Ser. N. 46. Bristol and London: Institute of Physics.