SECURITY CLASSIFICATION OF THIS PAGE

DISTRIBUTION STATEMENT A
Approved for public release; Distribution Unlimited

REPORT DOCUMENTATION PAGE

DTIC AD-A248 758
ELECTE APR 17 1992 S B

1a. REPORT SECURITY CLASSIFICATION
Unclassified

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE
Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
University of Notre Dame
Department of Physics

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
Office of Naval Research
Resident Representative

6c. ADDRESS (City, State, and ZIP Code)
Notre Dame, Indiana 46556

7b. ADDRESS (City, State, and ZIP Code)
536 S. Clark Street Room 286
Chicago, Illinois 60605-1588

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
Office of Naval Research

8b. OFFICE SYMBOL (If applicable)
1114 SC/ONR

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
N 00014-89-J-1136

8c. ADDRESS (City, State, and ZIP Code)
800 N. Quincy Street
Arlington, Virginia 22217-5000

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Theory of Semiconducting Superlattices and Microstructures

12. PERSONAL AUTHOR(S)
Dow, John D

13a. TYPE OF REPORT
Final

13b. TIME COVERED
FROM 88-10-15 TO 91-10-14

14. DATE OF REPORT (Year, Month, Day)
92-03-01

15. PAGE COUNT
75

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Superlattices, Hjalmarson-Frenkel Theory, Deep level theory, Mobius transforms, Amorphus silicon, ZnSe, Gap, Theory of dopants, Ti Tc Superconductors

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This final report of work summarizes the theoretical development of Semiconducting Superlattices and Microstructures. These include:

- The first theory of Hjalmarson-Frenkel core excitons in superlattices, and applied it to strained-layer systems. This theory will be useful in characterizing phenomena in such superlattices, using synchrotron radiation.

- A unified theory of dopants in II-VI materials, with emphasis on ZnSe and candidates for blue-green lasers. This explains why Ga is no longer an effective dopant in $Zn_{1-x}Mn_xSe$ for $x > 0.1$, with a theory that also provides a natural explanation of why ZnTe is almost unique among the II-VI semiconductors in that it is naturally p-type. This work provides, a theoretical framework for viewing and understanding the recent successes at 3M by Haas et al. in obtaining blue emission from ZnSe. It also provides strong evidence against the thirty-year-old

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
George B. Wright

22b. TELEPHONE (Include Area Code)
(703)696-4202

22c. OFFICE SYMBOL
1114SS/ONR

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

U.S. Government Printing Office: 1985-507-047

#19 Abstract continued.

explanation that self-compensation is responsible for the inability to make a p-n junction in II-VI semiconductors.

- A general theory of substitutional impurities in superlattices that predicts anomalous changes of dopant character with layer thicknesses, and establishes an intellectual framework for understanding the properties of impurities in superlattices.

- The mathematical theory of Möbius transforms its applications to physical inverse problems. This is an important new area of theoretical physics, especially for the Navy: with such transforms, one should be able to detect the thermal radiation of a distant object and, from the radiation, reconstruct the shape and temperature profile of the object.

- A new molecular dynamics of III-V semiconductors, based on Sankey's ideas for Si. This method solves the quantum local density equations very rapidly for the forces on atoms, and then moves the atomic nuclei according to Newton's equations. The result is a program that a computer can digest that can compute realistic trajectories of atoms in condensed matter. This will allow us to simulate many interesting microscopic processes, such as semiconductor growth, impurity diffusion, chemical reactions, etc.

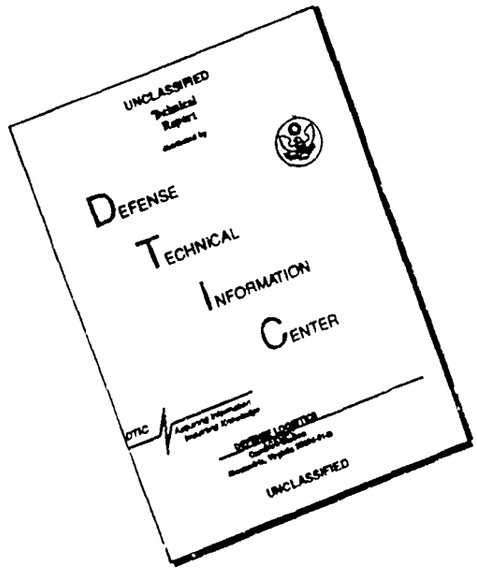
- A new, relaxed-lattice model of isoelectronic traps in GaP, explaining some old mysteries of the data.

- A theory of the localized states in the HOMO-LUMO gaps of macromolecules. The theory has the property that it can often be evaluated approximately for very large molecule, whose electronic structure is poorly known. The localized states associated with frontier orbitals often are responsible for the critical chemical reactions in which macromolecules partake.

- A new picture of doping amorphous Si, together with an explanation of its band-tail states. This picture appears to agree with recent data.

- An addition to the field of high-critical-temperature superconductors by computing the electronic structures of La_2CuO_4 and other high-T_c materials.

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

OFFICE OF NAVAL RESEARCH

FINAL REPORT

for

Contract N00014-89-J-1136

Task No. NR 616-027

Theory of semiconducting superlattices and microstructures

John D. Dow
Freimann Professor of Physics
Department of Physics
University of Notre Dame
Notre Dame, Indiana 46556

March 1, 1992

Reproduction in whole, or in part, is permitted for any purpose of the United States Government.

* This document has been approved for public release and sale; its distribution is unlimited.

92-09820

92 4 16 041

THEORY OF SEMICONDUCTING SUPERLATTICES AND MICROSTRUCTURES

John D. Dow
Department of Physics
University of Notre Dame
Notre Dame, Indiana 46556

A. Core excitons in superlattices

We have developed the first theory of Hjalmarson-Frenkel core excitons in superlattices, and applied it to strained-layer systems. We expect that this theory will be useful in characterizing phenomena in such superlattices, using synchrotron radiation.

B. Theory of dopants in ZnSe and related materials

We have provided a unified theory of dopants in II-VI materials, with emphasis on ZnSe and candidates for blue-green lasers. We have explained why Ga is no longer an effective dopant in $Zn_{1-x}Mn_xSe$ for $x>0.1$, with a theory that also provides a natural explanation of why ZnTe is almost unique among the II-VI semiconductors in that it is naturally p-type. This work provides, we believe, a theoretical framework for viewing and understanding the recent successes at 3M by Haase et al. in obtaining blue emission from ZnSe. It also provides strong evidence against the thirty-year-old explanation that self-compensation is responsible for the inability to make a p-n junction in II-VI semiconductors.

C. Theory of deep levels in superlattices

We have developed a general theory of substitutional impurities in superlattices that predicts anomalous changes of dopant character with layer thicknesses, and we have established an intellectual framework for understanding the properties of impurities in superlattices.

D. Möbius transforms

We have developed the mathematical theory of Möbius transforms and applied it to physical inverse problems. This is an important new area of theoretical physics, especially for the Navy: with such transforms, one should be able to detect the thermal radiation of a distant object and, from the radiation, reconstruct the shape and temperature profile of the object.

E. Molecular dynamics of semiconductors

We developed a new molecular dynamics of III-V semiconductors, based on Sankey's ideas for Si. This method solves the quantum local density equations very rapidly for the forces on atoms, and then moves the atomic nuclei according to Newton's equations. The result is a program that a computer can digest that can compute realistic trajectories of atoms in condensed matter. This will allow us to simulate many interesting microscopic processes, such as semiconductor growth, impurity diffusion, chemical reactions, etc.

F. Isoelectronic traps in GaP

We have developed a new, relaxed-lattice model of isoelectronic traps in GaP, explaining some old mysteries of the data.

G. Macromolecules

We have developed a theory of the localized states in the HOMO-LUMO gaps of macromolecules. The theory has the property that it can often be evaluated approximately for very large molecules, whose electronic structure is poorly know. The localized states associated with frontier orbitals often are responsible for the critical chemical reactions in which macromolecules partake.

Doping of a-Si

We have provided a new picture of doping of amorphous Si, together with an explanation of its band-tail states. This picture appears to agree with recent data.

High-T_c superconductors

We have added to the field of high-critical-temperature superconductors by computing the electronic structures of La_2CuO_4 and other high-T_c materials.

Publications

243. J. D. Dow, J. Shen, and S. Y. Ren. Core excitons in strained-layer superlattices. "Progress in Electronic Properties of Solids," Physics and Chemistry of Materials with Low-dimensional Structure, Festschrift in honor of Professor Franco Bassani, ed. by E. Doni, R. Girlanda, G. Pastori Parravicini, and A. Quattropani, (Kluwer, Dordrecht, 1989), pp. 439-449.

253. S. Y. Ren, J. Shen, R.-D. Hong, S. Klemm, M.-H. Tsai, and J. D. Dow. Deep levels and shallow-deep transitions in $ZnSe/Zn_{1-x}Mn_xSe$ superlattices. Surf. Sci. 228, 49-52 (1990).

254. J. D. Dow, S. Y. Ren, J. Shen, R.-D. Hong, and R.-P. Wang. Deep levels in superlattices. J. Electron. Mater. 19, 829-835 (1990).

255. R. V. Kasowski, M.-H. Tsai, J. D. Dow, and M. T. Czyzyk. Self-consistent antiferromagnetic ground state for La_2CuO_4 and CuO via energy band theory. Physica C 162, 1349-1350 (1989).

258. P. A. Fedders, J. D. Dow, O. F. Sankey, and D. A. Drabold. The role of topology and bond-distortions in determining the character of dopants and band-tailing in a-Si. Phys. Rev. B submitted

259. J. Shen, S. Y. Ren, and J. D. Dow. Relaxed-lattice model of isolated and paired isoelectronic traps in GaP. Phys. Rev. B 42, 9119-9126 (1990).

261. J. D. Dow, S. Y. Ren, J. Shen, and M.-H. Tsai. Deep levels in superlattices. Mater. Res. Soc. Symp. Proc. Vol. 163, 349-360 (1990).

262. O. F. Sankey, D. J. Niklewski, D. A. Drabold, and J. D. Dow. Molecular dynamics determination of electronic and vibrational spectra, and equilibrium structures of small Si clusters. Phys. Rev. B 41, 12750-12759 (1990).

263. S. Y. Ren and J. D. Dow. Change of the frontier electronic orbitals due to substitutional impurities in large chemical or biological molecules. Intl. J Quantum Chem., Quant. Biol. Symp. 17, 73-80 (1990).

270. S. Y. Ren and J. D. Dow. Generalized Möbius transforms for inverse problems. Phys. Lett. A 154, 215-216 (1991).

273. D. A. Drabold, J. D. Dow, P. A. Fedders, A. E. Carlsson, and O. F. Sankey. Convergence of force calculations for noncrystalline Si. Phys. Rev. B 42, 5345-5348 (1990).

These publications are presented in the Appendix.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1 | 23

QUALITY INSPECTED 1

APPENDIX

Publications generated
as a result of this
sponsored project.

Paper referenced #243 in the list of publications.

CORE EXCITONS IN STRAINED-LAYER SUPERLATTICES

JOHN D. DOW AND JUN SHEN

Department of Physics
University of Notre Dame
Notre Dame, Indiana 46556 USA

SHANG YUAN REN

Department of Physics
University of Notre Dame, Indiana, USA
and
University of Science and Technology of China
Hefei, Anhui, People's Republic of China

The physics of core excitons in semiconductors is reviewed, with emphasis on the fact that Hjalmarson-Frenkel 'deep' core excitons are observed, and co-exist with Wannier-Mott 'shallow' excitons which are not normally resolved experimentally. The theory of Hjalmarson-Frenkel excitons is extended to excitons in superlattices, and the $Ga3d$ core exciton in $GaAs_{1-x}P_x/GaP$ strained-layer superlattices is predicted to change from a resonance in the conduction band (with apparent negative binding energy) to a bound state in the gap (positive binding energy), as the $GaAs_{1-x}P_x$ layer thickness decreases.

1. Introduction

Franco Bassani pioneered the theory of core excitons and was one of the first theorists to call attention to the fact that the $Si\,2p$ core exciton exhibits a binding energy significantly different from the shallow donor binding energy in Si [1]. Since he was also one of the first theorists interested in artificial superlattices, we shall honor him by discussing the physics of core excitons in superlattices.

When a soft x-ray excites an atom in a semiconductor, it creates a core hole of very small radius and an electron with the same wave vector (see Figures 1 and 2). Because the core orbital has such a small radius, it does not overlap the corresponding orbital on an adjacent site, and so the core band is flat, with infinite mass. As a result, the hole is immobile, and is a fixed point charge around which the electron can revolve. One expects the electron's orbit to be hydrogenic, with an envelope wave function ψ obeying the effective-mass Schrödinger equation

$$[(-\hbar^2/2m^*)\nabla^2 - e^2/\epsilon r]\psi(\mathbf{r}) = E\psi(\mathbf{r}).$$

Here E is the energy of the orbiting electron with respect to the conduction band edge, ϵ is the dielectric constant of the semiconductor, and m^* is the conduction band's effective mass [2] (we have assumed an isotropic, non-degenerate effective mass, for simplicity). The ground state of such an electron in the presence of its hole should be the hydrogenic $1s$ state, with a binding energy relative to the conduction band minimum of

$$E_B = (13.6\,eV)(m^*/m_0\epsilon^2),$$

R. Girlanda et al. (eds.), Progress on Electron Properties of Solids, 439–449.
© 1989 by Kluwer Academic Publishers.

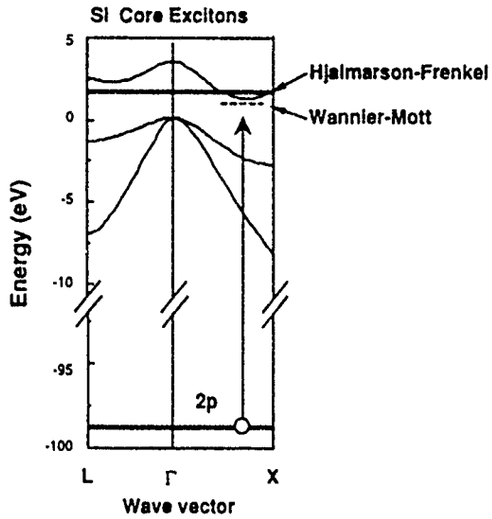


Figure 1. Energy versus wave vector: illustration of a *Si 2p* core excitation in *Si*. The core hole band is flat (lower heavy line), and the excited electron is either (i) associated with the conduction band minima near *X* for the shallow Wannier-Mott exciton (dashed line), or (ii) in the antibonding *s*-like A_1 deep level resonant with the conduction band for the Hjalmarson-Frenkel core exciton (upper heavy line).

typically tens of *meV*. Such a correlated electron-hole pair is a Wannier-Mott exciton [3].

Analogous impurity levels to the Wannier-Mott core excitons have been thoroughly studied and were thought to be completely understood until recently [4]: the shallow hydrogenic donor states [2]. The electrons of these states obey the same effective-mass Schrödinger equation as the electrons of the core excitons, and exhibit the predicted binding energies. In the case of *P* substituting for a *Si* atom in bulk *Si*, the physics of shallow donor levels has been well-established for decades. Therefore one expected that the *Si 2p* core exciton would exhibit the same binding energy as the *P* donor. It did not, being in many experiments two orders of magnitude larger: 0.1 to 0.9 *eV* [5-9].

The failure of the *Si 2p* core exciton binding energy to match the *P* donor binding energy was particularly perplexing because, to an adequate approximation (the $Z+1$ rule [1,10] or the optical alchemy approximation [11]), the core hole has zero radius and the same charge distribution as a proton. In the case of core-excited *Si*, the excited atom with its 'proton' and electron is '*P*' (similarly core-excited *Ga* and *In* are '*Ge*' and '*Sn*', respectively).

Hence in the case of the *Si 2p* core exciton, one expected an absorption spectrum that reflected the spectrum of the '*P* donor' impurity on *Si*, with its small hydrogenic effective-mass binding energy: 13.6 *eV* $(m^*/m_0\epsilon^2)$. The observation of much larger *Si 2p* core exciton binding energies ranging from 0.1 to 0.9 *eV* [5] became known as the '*Si* core exciton problem' - a problem whose explanation was complicated by the fact that the experimental results apparently did not agree. Indeed, the experiments calling for a binding energy almost the size of the *Si* band gap were particularly difficult to accept in the light of theories existing at that time. Bassani [1] was one of the first theorists to recognize that these observations called for a new theory of core excitons.

The differences between *P* in *Si* and a core-excited *Si* atom, Si^*, are small: (i) because

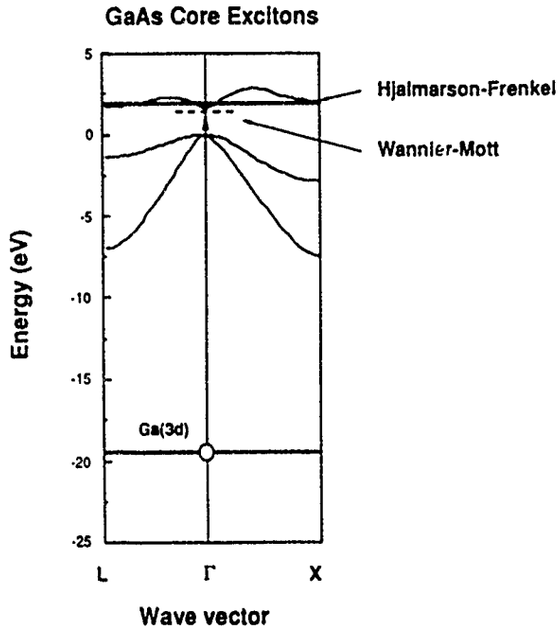


Figure 2. Illustration of the *Ga* 3*d* core excitons of *GaAs*. The shallow Wannier-Mott state lies only 2 *meV* below the conduction band edge. The deep Hjalmarson-Frenkel state lies 0.4 *eV* above the edge and is a T_2 resonance [13].

of the Franck-Condon principle [12], there is no lattice relaxation surrounding the core hole, whereas the lattice certainly relaxes around a *P* impurity - an effect that has always been thought to be negligible, and (ii) the differences in the core hole and proton charge distributions have always been believed to be negligible. Hence a successful theory of the *Si* core exciton problem would have either relied on dynamical effects such as time-dependent screening or explicitly embraced the sameness of Si^+ and *P* - and ascribed the apparent differences in binding energies to the fact that experiments probe different aspects of the same defect [1]. The latter approach now provides a simple and natural solution to the *Si* core exciton problem, but had been rejected by many theorists in the 1970's because it implied that the then-current understanding of *P* in *Si* had been incomplete.

2. Co-Existence of 'Deep' and 'Shallow' States

The many apparently contradictory facts about the *Si* 2*p* exciton can be understood once one recognizes that impurities such as *P* in *Si* have *both* 'deep' and 'shallow' states [4] - and that there are corresponding deep or 'Hjalmarson-Frenkel' [13] and shallow or 'Wannier-Mott' excitons [3]. Infrared and transport experiments are sensitive to the shallow impurity levels while core exciton experiments resolve the 'deep' Hjalmarson-Frenkel excitons rather than the shallow Wannier-Mott states. For *P* in bulk *Si* the deep level lies slighly *above* the conduction band edge, so that the corresponding exciton has an apparent *negative* binding energy [4,14] (see Figure 1). However, if the Hjalmarson-

Frenkel $Si\ 2p$ exciton is near a surface, its energy is rather dramatically perturbed, so that its apparent binding energy can be $0.8\,eV$ [15] (see Figure 3).

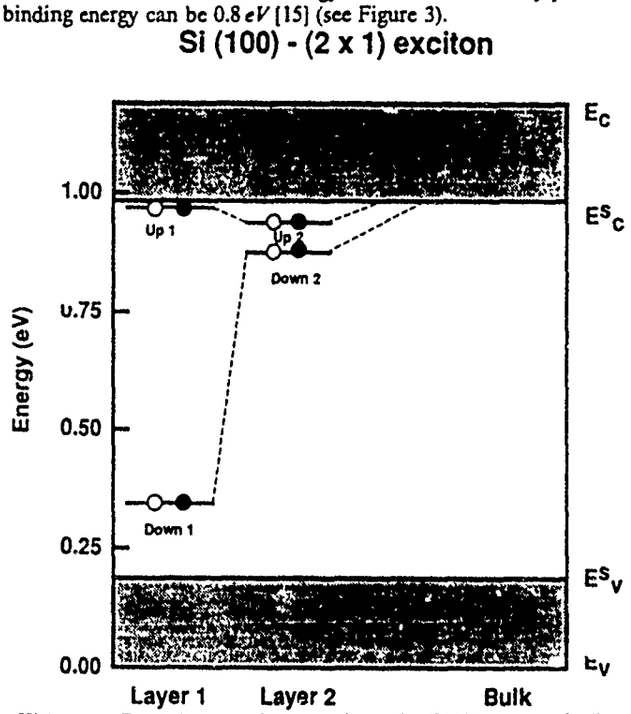


Figure 3. $Si\ 2p$ Hjalmarson-Frenkel core exciton energies at the (2×1) asymmetric-dimer (100) surface of Si, after Ref. [15]. 'Up' ('down') denotes the upper (lower) surface site (assuming the Si occupies the lower half plane), and results are given for the surface layer (Layer 1) and the uppermost sub-surface layer (Layer 2). These levels coalesce into a resonance in the conduction band for excitons in the bulk (layer ∞).

Thus the explanation of the wide range of observed exciton energies is that the different measurements had different surface sensitivities, and probed core exciton binding energies in Si, ranging from slightly negative values characteristic of the bulk to large positive values (nearly the size of the band gap) at the surface.

A central element of this picture for the $Si\ 2p$ core exciton was the prediction that the Hjalmarson-Frenkel exciton lies slightly *above* the conduction band edge in Si, with an apparent negative binding energy - a prediction that could be tested by measuring its energy in Si_xGe_{1-x} alloys as a function of alloy composition x. The theory [14] predicted that the Hjalmarson-Frenkel state would descend into the gap and assume a positive binding energy for $x\approx0.27$, and then re-enter the conduction band for larger x. The composition $x\approx0.27$ [16] corresponds to a cross-over in the conduction band minima of Si_xGe_{1-x} from being near the (100) X-point of the Brillouin zone of Si for $x = 1$ to the (111) L-point of Ge. Bunker et al. [16] performed an experiment to test these ideas, and concluded that the picture is correct for the $Si\ 2p$ core exciton, but that the predicted energies of the Hjalmarson-Frenkel exciton were slightly lower than observed (see Figure 4). The resulting $Si\ 2p$ core exciton binding energy is about $-0.06\,eV$.

The Hjalmarson-Frenkel core exciton level slightly above the band gap in bulk Si

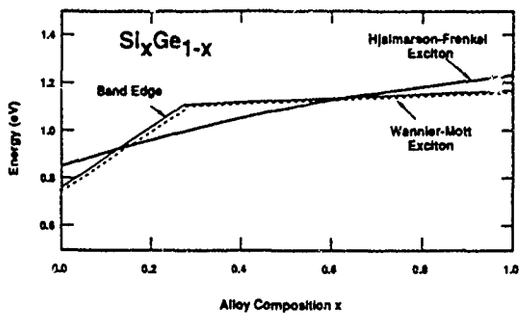


Figure 4. Energies (in eV) versus alloy composition x in Si_xGe_{1-x}, after Ref. [16]. conduction band edges (light solid line), (2) Wannier-Mott exciton levels (dashed line), and experimental conclusions [16] (heavy solid line), obtained by analyzing data with an adjusted version of the theory of Ref. [14].

corresponds to a 'deep' level of P in Si that is resonant with the conduction band. This deep level had not been a part of the theory of the P impurity, and yet was demanded by a fact that was well-known, but not fully appreciated: The central-cell defect potential of a P substitutional impurity in Si must be greater than $4\,eV$ deep, because the difference between P and Si s-state atomic energies is $4\,eV$. In a solid, it is impossible to have a $4\,eV$ perturbation potential that does not alter the electronic structure on a $4\,eV$ energy scale - and the old shallow donor theory indicated alterations of the electronic structure only on the tens of meV scale of the donor binding energy. Thus, in retrospect [4], P in Si must have, in addition to the shallow donor levels, four deep levels that lie above the conduction band edge and are resonances: one s-like level that corresponds to the Hjalmarson-Frenkel core exciton and a triply degenerate p-like level at higher energy. These four levels are localized in space and arise from the four perturbed bonds formed by the P impurity when it replaces Si.

The same theoretical picture of Hjalmarson-Frenkel core excitons in Si described Ga $3d$ core excitons in the bulk [13,17,18] (Figure 5) and Ga $3d$ and In $4d$ excitons at surfaces (Figures 6 and 7) [17,19,20] in $GaAs$, $GaSb$, $InAs$, $InSb$, and InP - a dramatic success because the surface exciton energies differ markedly from those in the bulk, yet generally agree with the theory.

The major difference between the Si $2p$ core exciton on the one hand and the Ga $3d$ and In $4d$ excitons on the other is symmetry. Dipole selection rules imply that the core-excited Ga and In generate electrons in odd-parity states and therefore have the energies of the p-like T_2 deep levels of Ge and Sn, respectively. In Si, the core-excited electron occupies an s-like A_1 level.

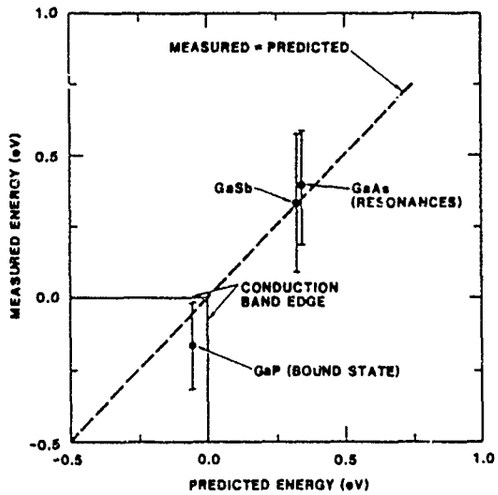


Figure 5. Comparison of measured [18] and predicted energies (in eV) of Hjalmarson-Frenkel $Ga\,3d$ core excitons in GaP, $GaSb$, and $GaAs$, after Ref. [13]. The zero of energy is the conduction band edge.

3. Core Excitons in Superlattices

Superlattices are particularly interesting materials for studying core excitons because the superlattice band edges are very sensitive to the layer thicknesses, while deep impurity levels are not. Therefore a $Ga\,3d$ core excitation in a $GaAs$ quantum well, for example, would have its core excitation energy (i.e., the Ge impurity T_2 level) remain almost independent of the width of the $GaAs$ well while the conduction band edge would dramatically increase in energy as the well became thinner (due to quantum confinement). As a result, the apparent core exciton binding energy (i.e., the energy of the Ge impurity level with respect to the conduction band edge) increases markedly as the well-width decreases. The physics is even more complex when the superlattice is not lattice-matched, and strain is important.

In this paper we apply these ideas to $Ga\,3d$ core excitons in $GaAs_{0.6}P_{0.4}/GaP$ strained-layer superlattices. We believe that these are the first theoretical results for this type of Hjalmarson-Frenkel core exciton in III-V superlattices. Our goal is to show that a Hjalmarson-Frenkel $Ga\,3d$ core exciton at a site near the center of a $GaAs_{0.6}P_{0.4}$ layer can emerge from the conduction band into the fundamental band gap of the superlattice as the layer thickness decreases. That is, its apparent binding energy with respect to the conduction band edge changes from negative to positive with decreasing layer thickness (see Figure 8).

The theory is identical to a theory for a Ge impurity at a central Ga site of a $GaAs_{0.6}P_{0.4}/GaP$ superlattice. The theoretical formalism has been published for such a defect in strain-free $GaAs/Al_xGa_{1-x}As$ superlattices [21-24], and can be modified in a straight-forward manner to treat strained-layer superlattices. The strain energy is

$$E_s = (a^3/64)(\Sigma_{ij}[6c_{11}+12c_{12}]([d_{ij}-d_0]/d_0)^2+\Sigma_{ijk}[c_{11}-c_{12}](\delta\theta_{ijk})^2)$$

where the sums are over the bonds (of length d_{ij}) connecting atoms i and j and over the

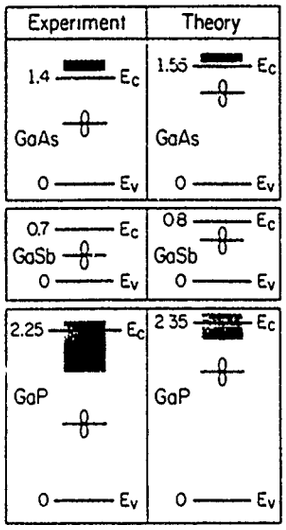


Figure 6. Experimental and theoretical *Ga 3d* Hjalmarson-Frenkel core exciton energies (propellers) at the (110) surfaces of *GaAs*, *GaSb*, and *GaP*, after Ref. [19]. The valence and conduction band edges are denoted E_v and E_c. The horizontal lines denote the bottom of the surface state band.

distortions θ_{ijk} of bond angles defined by atoms i,j, and k. Here c_{11}, c_{12}, d_0, and a are the elastic constants, perfect-crystal bond lengths, and perfect-crystal lattice constant of the relevant material. There are three types of contributions to the strain energy: (i) from *GaP*, (ii) from $GaAs_{1-x}P_x$, and (iii) from the interfaces. Three parameters define the geometry of the strained superlattice: (i) a_{perp}, the lattice constant in the direction perpendicular to the growth or z direction (this is the same in both $GaAs_{1-x}P_x$ and *GaP* layers), (ii) $c(GaP)$, the strained lattice constant in the z or growth direction in a *GaP* layer, and (iii) $c(GaAs_{1-x}P_x)$, the corresponding lattice constant in $GaAs_{1-x}P_x$. The strain energy is minimized with respect to these three parameters, and a_{perp}, $c(GaP)$, and $c(GaAs_{1-x}P_x)$ are determined. Then the off-diagonal (only) matrix elements of the Hamiltonian are scaled according to the rules of Harrison (i.e., as d^{-2} [25]) and Slater and Koster (i.e., the appropriate angle dependences [25,26]). Finally the superlattice electronic structure and Green's functions are calculated, in the usual fashion [21-24].

The predictions of the theory are contained in the p-like T_2 '*Ge* impurity' states, because the $3d$ core hole is associated with T_2 electron states as a consequence of the dipole selection rules for optical absorption. The predicted band gaps for [001] $GaAs_{0.6}P_{0.4}/GaP$ superlattices, are expected to lie slightly below [27,28] the experimental gap, but within about $\approx 0.1\,eV$ of it. The band edge of the bulk materials, both unstrained and strained as in the superlattice, are compared with the superlattice band edges in Figure 9 (we have assumed a valence band offset for a strained $GaAs_{0.6}P_{0.4}/GaP$ heterojunction of 60% of $0.33\,eV$ [29]). The corresponding 'binding

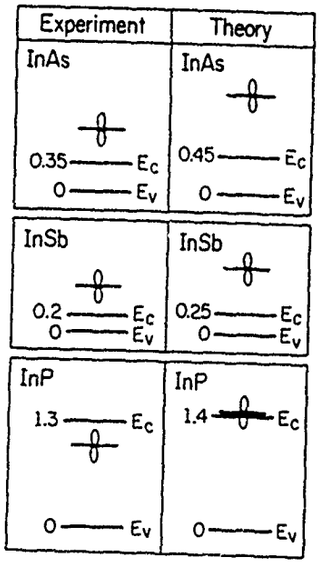


Figure 7. Experimental and theoretical *In* 4*d* Hjalmarson-Frenkel core exciton energies (propellers) at the (110) surfaces of *InAs*, *InSb*, and *InP*, after Ref. [19].

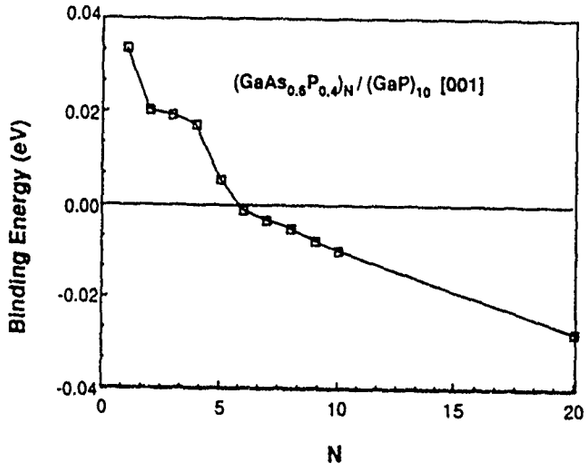


Figure 8. Predicted apparent binding energy of a *Ga* 3*d* core exciton p_z-like level at the center of a $GaAs_{0.6}P_{0.4}$ layer in an $N \times 10$ $GaAs_{0.6}P_{0.4}/GaP$ superlattice as a function of N. Note that for small (large) N the binding energy is positive (negative).

energies' of the Hjalmarson-Frenkel excitons are displayed in Figure 8 [28]. As the $GaAs_{0.6}P_{0.4}$ layer becomes thinner, the conduction band edge moves up in energy with

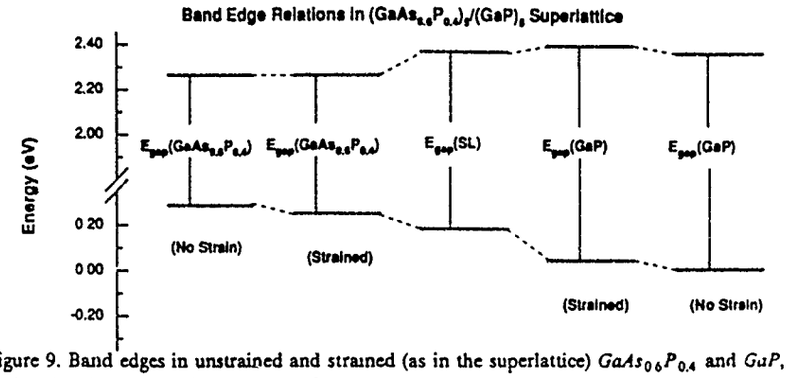


Figure 9. Band edges in unstrained and strained (as in the superlattice) $GaAs_{0.6}P_{0.4}$ and GaP, and in the 5×5 $GaAs_{0.6}P_{0.4}/GaP$ superlattice.

respect to the core exciton, until the core-exciton level descends into the gap. For simplicity of presentation in Figure 8, we have plotted only the p_z-like T_2 core exciton level for an exciton far from an interface. In the superlattice, and especially near interfaces, the T_2 levels (which are degenerate in the bulk) split, with the splittings being largest near the interfaces.

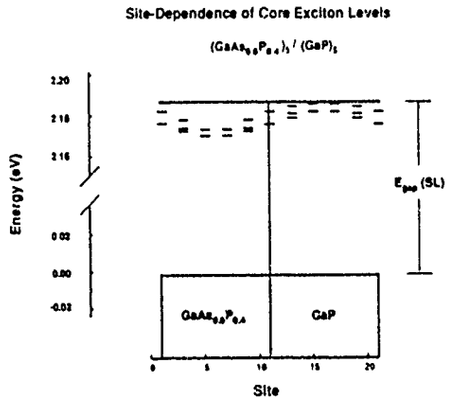


Figure 10. Predicted site dependence of the $Ga3d$ core exciton levels (in eV) in a 5×5 $GaAs_{0.6}P_{0.4}/GaP$ superlattice. The band gap of the superlattice is denoted $E_{gap}(SL)$ The Ga atoms are at odd-numbered sites. When only two lines are plotted at a site, two of the exciton levels are nearly degenerate.

Figure 10 illustrates how the core exciton levels should vary in energy for a 5×5 $GaAs_{0.6}P_{0.4}/GaP$ superlattice, as the position of the core-excited Ga atom is changed. Of course, experiments measuring such excitons should detect a broadened line whose shape reflects the distribution of sites at which the excitons are created and the splittings at each site.

4. Summary

We hope that this work will stimulate efforts to measure core excitons in superlattices and to show that suitable manipulation of band edges in superlattices by controlling layer thicknesses will cause Hjalmarson-Frenkel excitons to move into and out of the band gap.

Acknowledgements

We are grateful to the US Air Force Office of Scientific Research and the US Office of Naval Research for their generous support (Contract Nos. AF-AFOSR-89-0063-DEF and N00014-89-J-1136). We have benefitted from conversations with B. Bunker and K. Newman.

References

1. F. Bassani, *Appl. Optics* **19**, 4093 (1980), and reference therein.
2. W. Kohn, in *Solid State Physics* (edited by F. Seitz and D. Turnbull, Academic Press, New York, 1957), Vol. 5, pp. 258-321; J.M. Luttinger and W. Kohn, *Phys. Rev.* **97**, 969 (1955).
3. R.S. Knox, *Theory of Excitons*, Academic, New York, 1963.
4. H.P. Hjalmarson, P. Vogl, D.J. Wolford and J.D. Dow, *Phys. Rev. Lett.* **44**, 810 (1980).
5. A. Quattropani, F. Bassani, G. Margaritondo and G. Tinivella, *Nuovo Cimento* **51B**, 335 (1979), and reference therein.
6. F.C. Brown, in *Solid State Physics* edited by H. Ehrenreich, F. Seitz and D. Turnbull, Academic, New York, 1974, Vol. 29, p. 1 and references therein.
7. F.C. Brown and O.P. Rustgi, *Phys. Rev. Lett.* **28**, 497 (1972).
8. G. Margaritondo and J.E. Rowe, *Phys. Lett.* **59A**, 464 (1977).
9. R.S. Bauer, R.Z. Bachrach, D.E. Aspnes and J.C. McMenamin, *Nuovo Cimento* **B39**, 409 (1977).
10. M. Altarelli, *J. Phys. C* **4**, 95 (1978).
11. J.D. Dow, D.R. Franceschetti, P.C. Gibbons and S.E. Schnatterly, *J. Phys. F* **5**, L211 (1975).
12. M. Lax, *J. Chem. Phys.* **20**, 1752 (1952).
13. H.P. Hjalmarson, H. Büttner and J.D. Dow, *Phys. Rev. B* **24**, 6010 (1981); *Phys. Lett.* **85A**, 293 (1981).
14. K.E. Newman and J.D. Dow, *Solid State Commun.* **50**, 587 (1984).
15. M.A. Bowen, R.E. Allen and J.D. Dow, *Phys. Rev.* **B30**, 4617 (1984).
16. B.A. Bunker, S.L. Hulbert, J.P. Stott and F.C. Brown, *Phys. Rev. Lett.* **53**, 2157 (1984). The cross-over of conduction band edges occurs between $x = 0.1$ and $x = 0.3$, depending on temperature. Figure 4 corresponds to a low temperature.
17. M. Skibowski, G. Sprüssel and V. Saile, in *Proceedings of the Fourteenth International Conference on the Physics of Semiconductors*, Edinburgh, 1978, edited by B.L.H. Wilson, Institute of Physics, Bristol, 1979, p. 1359.
18. D.E. Aspnes, C.G. Olson and D.W. Lynch, in *Proc. XIII-th Intern. Conf. Phys. Semiconductors* (Rome, 1976), edited by F.G. Fumi (North-Holland, Amsterdam, 1976) pp. 1000; *Phys.*

Rev. **B12**, 2527 (1975); **B14**, 2534, 4450 (1976); D.E. Aspnes, M. Cardona, V. Saile and G. Sprüssel, *Solid State Commun.* **31**, 99 (1979).

19. R.E. Allen and J.D. Dow, *Phys. Rev.* **B24**, 911 (1981).
20. D.E. Eastman and J.L. Freeouf, *Phys. Rev. Lett.* **33**, 1601 (1974); **34**, 1624 (1975); W. Gudat and D.E. Eastman, *J. Vac. Sci. Technol.* **13**, 831 (1976); D.E. Eastman, T.-C. Chiang, P. Heimann and F.J. Himpsel, *Phys. Rev. Lett.* **45**, 6546 (1980).
21. S.Y. Ren, J.D. Dow and J. Shen, *Phys. Rev.* **B38**, 10677 (1988).
22. S.Y. Ren and J.D. Dow, *J. Appl. Phys.* **65** 1987 (1989).
23. R.-D. Hong, D.W. Jenkins, S.Y. Ren and J.D. Dow, *Mater. Res. Soc. Symp. Proc.* **77**, 545 (1987), in *Interfaces, Superlattices, and Thin Films*, ed. J.D. Dow and I.K. Schuller.
24. J.D. Dow, S.Y. Ren and J. Shen, NATO Advanced Science Institutes Series **B183**: *Properties of Impurity States in Superlattice Semiconductors*, ed. by C.Y. Fong, I.P. Batra and S. Ciraci (Plenum Press, New York, 1988), p. 175.
25. J.C. Slater and G.F. Koster, *Phys. Rev.* **94**, 1498 (1954).
26. W.A. Harrison, *Electronic Structure and the Properties of Solids*, W.H. Freeman, San Francisco, 1980, p. 481.
27. P.Vogl, H.P. Hjalmarson and J.D. Dow, *J. Phys. Chem. Solids* **44**, 365 (1983).
28. The electronic structure model employed [27] is an sp^3s^* empirical tight-binding model whose matrix elements exhibit chemical trends depending on atomic energies and bond lengths. The matrix elements were obtained by simultaneously fitting the band structures at the Γ and X points of sixteen semiconductors. In the case of $GaAs_{1-x}P_x$, the L conduction minima are slightly lower than they should be, especially for $x \approx 0.5$, where they can lie slightly below the Γ and X minima. Thus the quantitative conclusions of the present work for the band gap and the exciton binding energy have theoretical uncertainties of order $0.2\,eV$, and the core exciton binding energies are, if anything, *underestimated* by about this amount. Hence the binding energies we predict are likely to be even larger in Nature.
29. C. Delerue, M. Lannoo and J.M. Langer, *Phys. Rev. Lett.* **61**, 199 (1988).

Surface Science 228 (1990) 49–52
North-Holland

Paper referenced #253 in the list of publications.

DEEP LEVELS AND SHALLOW–DEEP TRANSITIONS IN $ZnSe/Zn_{1-x}Mn_xSe$ SUPERLATTICES

Shang Yuan REN, Jun SHEN, Run-Di HONG, Stefan KLEMM, M.-H. TSAI and John D. DOW
Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

Received 2 June 1989; accepted for publication 18 July 1989

A theory of deep impurity levels in superlattices is outlined, and applied to $ZnSe/Zn_{0.5}Mn_{0.5}Se$ with the substitutional dopant Ga_{Zn}. Ga is predicted to undergo a shallow–deep transition as a function of x in bulk $Zn_{1-x}Mn_xSe$, and so Ga dopes ZnSe n-type but prevents $Zn_{0.5}Mn_{0.5}Se$ from being doped n-type even by modulation doping. In $ZnSe/Zn_{0.5}Mn_{0.5}Se$ superlattices, the band edges are quite sensitive to changes in the layer thicknesses, but the deep levels are not. As a result, shallow–deep transitions as functions of layer thickness are predicted to occur. The physics of shallow–deep transitions in superlattices is elucidated, and its relevance to the II–VI doping problem is discussed.

1. Introduction

Every s- and p-bonded substitutional impurity in a semiconductor produces four "deep" levels that lie near or in the fundamental band gap of the host. These levels are due to the central-cell defect potential, and may all lie resonant with the host bands, in which case the impurity is termed "shallow." Or at least one of these levels may lie within the gap, in which case the impurity is "deep" [1]. Normally one such deep level is A_1-symmetric or s-like and three are p-like (and possibly degenerate, depending on the site symmetry).

2. Shallow–deep transitions

In a very crude (but instructive) approximation [2], the deep levels are insensitive to changes of the host composition, atomic ordering (e.g., superlattice versus random alloy), or pressure, and retain their absolute energies. In contrast, the conduction and valence band edges are sensitive to such changes, and so it is rather common that a band edge passes through a deep level, changing the character of the impurity from shallow to deep. This is believed to be the case [3] for Si_{cation} (Si on a cation site) in $Al_xGa_{1-x}As$: for $x < 0.2$ the A_1-symmetric deep level lies in the conduction band, making Si a shallow donor; but for $x > 0.3$ the deep level is in the fundamental band gap, allowing the Si atom to trap an extra electron rather than donate one to the conduction band, rendering the material semi-insulating rather than n-type [3]. This shallow–deep transition is particularly interesting in superlattices, where the band edges are sensistive to the choice of layer thickness, but the deep levels are not [2].

3. Shallow–deep transitions in superlattices

One example of such a transition is the Ga_{Zn} impurity near the center of a ZnSe layer in a $ZnSe/Zn_{0.5}Mn_{0.5}Se$ [001] superlattice. This impurity is a shallow donor, with its A_1-symmetric deep level in the conduction band, for thick ZnSe layers. Fig. 1 illustrates how the conduction band edge of a $N \times 10$ superlattice passes through the Ga deep level as N decreases from $N = 10$ to $N = 1$ (a single layer of ZnSe). In the thin superlattices (for $N < 3$, according to the theory [2,4]), the superlattice's conduction band edge lies above the deep level; and Ga becomes a deep impurity: the extra electron (relative to Zn) of neutral Ga is trapped in the deep level, which can also trap an additional electron of opposite spin. For thick ZnSe layers ($N \geq 3$), the extra electron of neutral

0039-6028/90/$03.50 © Elsevier Science Publishers B.V.
(North-Holland)

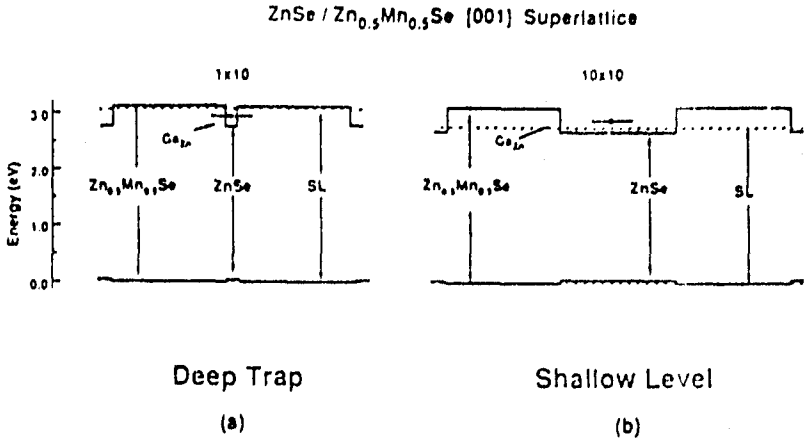


Fig. 1. Well-center Ga_{Zn} A_1 deep levels in (a) thin (1 × 10) and in (b) thick (10 × 10) quantum well ZnSe/$Zn_{0.5}Mn_{0.5}Se$ [001] superlattices. The thick solid lines are the band edges of ZnSe and $Zn_{0.5}Mn_{0.5}Se$ bulk semiconductors respectively. The dashed lines are the superlattice band edges. The short solid lines are the Ga_{Zn} deep levels. The Ga_{Zn} deep level is above the superlattice conduction band edge in a 10 × 10 superlattice and is in the gap of a 1 × 10 superlattice. The extra valence electron in the Ga_{Zn} resonant deep level will fall to the conduction band edge in the 10 × 10 superlattice, while in the 1 × 10 superlattice the extra electron will occupy the deep level which can also trap another electron of opposite spin.

Ga spills out of the deep level (which lies above the superlattice's conduction band edge) and the Ga is autoionized, creating a long-ranged Coulomb potential which binds the electron at zero temperature in a shallow donor level.

Fig. 2 illustrates how the Ga deep level, the conduction band minimum (CBM), the valence band maximum (VBM), and the shallow level are predicted to vary with ZnSe layer thickness N in an $N \times 10$ ZnSe/$Zn_{0.5}Mn_{0.5}Se$ superlattice. The predictions use an empirical tight-binding Hamiltonian [5,6] together with the Green's function method [1].

This behavior of the Ga deep level as a function of layer thickness N is similar to that found as a function of alloy composition x in $Zn_{1-x}Mn_xSe$: for $x > 0.1$ the Ga deep level lies in the band gap, not in the conduction band and traps electrons rather than donating them. This means that doping of $Zn_{1-x}Mn_xSe$ for $x > 0.1$ with Ga should produce semi-insulating rather than n-type material, which appears to be the case experimentally [7]. Even modulation doping of $Zn_{1-x}Mn_xSe$ with Ga will not produce n-type material for $x > 0.1$, because Ga is a deep trap in both layers of a $Zn_{1-x}Mn_xSe/Zn_{1-y}Mn_ySe$ superlattice for $y > x > 0.1$.

Fig. 3 illustrates the predicted dependence on alloy composition x of the levels of a Ga_{Zn} impurity in the ZnSe layer of a 1 × 10 ZnSe/$Zn_{1-x}Mn_xSe$ superlattice. For $x = 0$, the superlattice reduces to bulk ZnSe, and Ga has a shallow hydrogenic ground state donor level slightly below the conduction band minimum, which provides n-type doping. The Ga deep level lies above the conduction band minimum. As the alloy composition x of the ZnSe/$Zn_{1-x}Mn_xSe$ superlattice increases, the band gap opens up and the conduction band edge (measured with respect to the valence band maximum) moves to higher energy until, near $x = 0.4$, the band edge passes through the deep level. For $x > 0.4$, the stable ground state of the neutral Ga impurity in the 1 × 10 ZnSe/$Zn_{1-x}Mn_xSe$ superlattice has the deep level occupied by one electron. This deep level can trap a second electron of opposite spin, and so it removes electrons from the conduction band, making the material semi-insulating rather than n-type.

Shallow–deep transitions can occur when the valence band edge passes through a deep level, much the same as when a conduction band edge does. The valence-band shallow–deep transitions normally have a much more dramatic effect on the

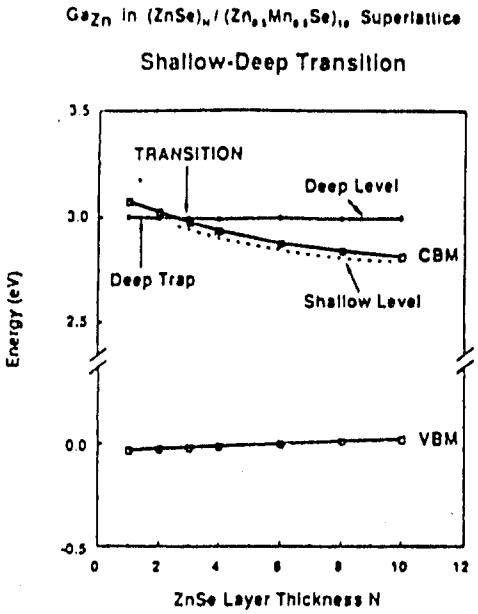


Fig. 2. Dependence of deep levels and superlattice band edges on ZnSe layer thickness N (number of ZnSe molecules thick) for $N \times 10$ ZnSe/$Zn_{0.5}Mn_{0.5}Se$ superlattices. The thick lines are the superlattice conduction band (CBM) and valence band (VBM) edges respectively. The top of the ZnSe valence band (without strain) is taken to be the zero of energy. The thin solid line is the predicted well-center Ga_{Zn} A_1 deep level. The shallow level, which follows the conduction band edge, is depicted by a dashed line. A shallow–deep transition is predicted to occur around $N = 3$.

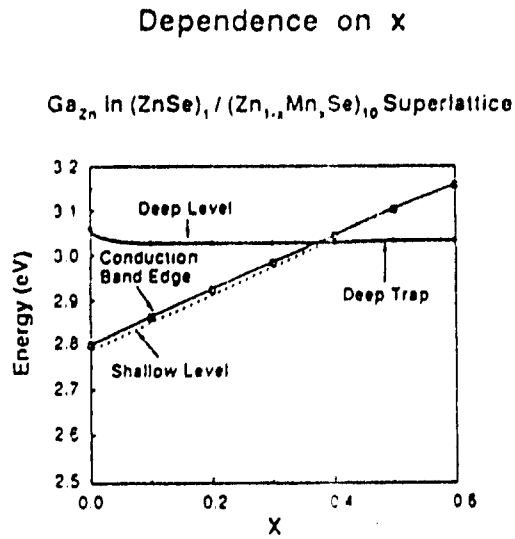


Fig. 3. Predicted dependence on Mn concentration x of the Ga_{Zn} deep level, and the shallow donor level in 1×10 ZnSe/$Zn_{1-x}Mn_xSe$ [001] superlattices. The Ga_{Zn} A_1 deep level is resonant with the conduction band when $x < 0.4$ (making Ga a shallow donor impurity), and is a deep trap occupied by the extra electron for $x > 0.4$.

doping character of a material, however, because they invariably involve p-like deep levels capable of containing six electrons (whereas the conduction-band shallow–deep transitions normally involve A_1-symmetric levels capable of trapping only two electrons).

We believe that such shallow–deep transitions are responsible for the different doping characters [8,9] of ZnSe (which can be easily doped n-type but not p-type) and ZnTe (which can be doped p-type): deep levels that lie in the gap of ZnSe and trap holes instead lie below the valence band maximum in ZnTe and donate holes. Clearly one way to enhance the p-type dopability of a II–VI semiconductor is to manipulate the semiconductor's valence band maximum, moving it up in energy until it covers the deep hole traps. For example, the p-dopability of CdTe can be improved [10] by fabricating a CdTe/ZnTe strained-layer superlattice. In this case the strain splits the

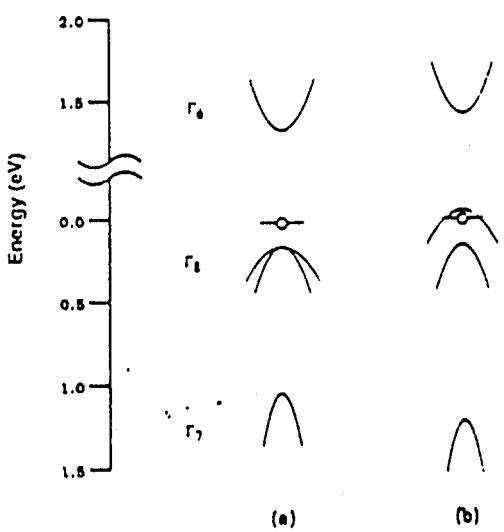


Fig. 4. Schematic energy band structure (energy in eV versus wave vector) of CdTe, illustrating how strain qualitatively changes the valence band level structure with respect to the deep level energy and covers the deep level. (a) The bulk semiconductor with a deep hole trap (that also contains at least one hole) within 0.2 eV of the valence band edge. The Γ_7 and Γ_8 bands are p-like bands that are split due to the large spin–orbit interaction in CdTe. (b) A 2×4 superlattice has an internal strain that further splits the valence band and covers up the deep level, autoionizing the hole.

valence band maximum of the CdTe and covers up deep hole traps in the gap slightly above the valence band maximum (fig. 4). A more complete discussion of this p-doping problem will be published elsewhere [11].

4. Summary

The physics of shallow–deep transitions plays a major role in determining the doping properties of II–VI semiconductors. Band edges pass through deep levels and change the doping character of the impurity from n-type (donor) to semi-insulating (trap) or from p-type (acceptor) to semi-insulating. We believe that by better understanding and using these shallow–deep transitions, it will be possible to circumvent many of the doping problems that currently plague II–VI semiconductors.

Acknowledgments

We are grateful to the Office of Naval Research, the Air Force Office of Scientific Research, the Army Research Office, and the Defense Advanced Research Projects Agency for their generous support (Contract Nos. N00014-89-J-1136, AF-AFOSR-89-0063, DAAL03-87-K-0112, and N0530-0716-05).

References

[1] H.P. Hjalmarson, P. Vogl, D.J. Wolford and J.D. Dow, Phys. Rev. Lett. 44 (1980) 810;
for the concepts that form the foundation of this work, see also W.Y. Hsu, J.D. Dow, D.J. Wolford and B.G. Streetman, Phys. Rev. B 14 (1977) 1597.
[2] S.Y. Ren, J.D. Dow and J. Shen, Phys. Rev. B 38 (1988) 10677.
[3] H.P. Hjalmarson and T.J. Drummond, Appl. Phys. Lett. 48 (1986) 657; see also, ref. [1].
[4] J. Shen, S.Y. Ren and J.D. Dow, to be published.
[5] P. Vogl, H.P. Hjalmarson and J.D. Dow, J. Phys. Chem. Solids 44 (1983) 365.
[6] A. Kobayashi, O.F. Sankey and J.D. Dow, Phys. Rev. B 28 (1983) 946.
[7] R. Gunshor and G. Prinz, private communications.
[8] R.N. Bhargava, J. Cryst. Growth 59 (1982) 15, and references therein.
[9] See, for example, C.H. Henry, K. Nassau and J.W. Shiever, Phys. Rev. B 4 (1971) 2453, and references therein.
[10] S.Y. Ren, J.D. Dow and S. Klemm, J. Appl. Phys., in press.
[11] J.D. Dow, Run-Di Hong, S. Klemm, S.Y. Ren and O.F. Sankey, to be published.

Paper, referenced #254 in the list of publications.

Journal of Electronic Materials, Vol. 19, No. 8, 1990

Deep Levels in Superlattices

John D. Dow*, Shang Yuan Ren*, Jun Shen, Run-Di Hong and Ruo-Ping Wang*

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556 U.S.A.
Department of Physics and Astronomy, Arizona State University* Tempe, Arizona 85287 U.S.A.

The physics governing deep levels in superlattices and quantum wells is elucidated, with emphasis on the importance of shallow-deep transitions caused by a band edge passing through a deep level, and the accompanying change in doping character of the impurity.

Key words: Deep impurity levels, superlattices

I. INTRODUCTION

The modern definition of a deep energy level is a level that originates from the central-cell defect potential of the impurity that produces it.[1] By this definition, which supplants the older energy criterion of a level within the fundamental band gap by at least 0.1 eV, many deep levels in semiconductors lie resonant with the host energy bands and so do not fall within the band gap. Indeed, for *s*- and *p*-bonded substitutional impurities in semiconductors, one expects four deep levels near or within the fundamental gap, associated with the four perturbed impurity bonds: one of *s*-like symmetry and three of *p*-like symmetry. An impurity is termed a deep impurity if one or more of these deep levels lies within the fundamental band gap, and so can trap an electron or a hole—removing carriers rather than donating them.

In this paper we consider how deep levels behave differently in superlattices and in quantum wells from in bulk semiconductors. We find that the effects of a superlattice on the absolute energy of a deep level tend to be (i) small, shifting or splitting the level by an amount of order 0.1 eV or less, and (ii) in most cases localized to within about three biatomic layers of an interface between the superlattice's constituent materials. Therefore, at first glance, the physics of deep levels in superlattices would appear to be uninteresting. However, the effect of the superlattice on the host band edges is quite dramatic for small-period superlattices, producing quantum confinement effects and typically shifting the superlattice band edges away from the large-period superlattice edges by several tenths of an eV. As a result of the confinement-induced movement of band edges in superlattices, these edges can often pass through a deep level, converting a shallow impurity into a deep impurity. (Note the distinction between a deep *impurity*, which has at least one of its deep levels in the fundamental gap, and a deep *level* which may or may not be in the gap.) Such shallow-deep transitions as a function of decreasing superlattice layer thickness are the most dramatic superlattice effects on deep levels.

(Received July 17, 1989)

0361-3235/1990/1401—455 00/$0 AIME

II. SHALLOW-DEEP TRANSITIONS IN BULK ALLOYS

The theory of doping in semiconductors[2] has to be revised to account for the deep levels, which play a major role in determining the doping character of an impurity and its stable ground state. The previous theory of doping, effective-mass theory,[2] simply assumed that substitutional impurities from Columns to the right (left) in the Periodic Table of the host atom would be donors (acceptors). This is not always the case. Fig. 1 shows how the relative energies of the deep levels and the band edges determine the character of a dopant that by the old rules should be a single donor, *e.g.*, P in Si. In most cases considered, the *s*- and *p*-like bonding orbitals of the impurity lie well below the valence band maximum and are fully occupied by electrons; the antibonding *p*-like orbitals are empty and high in the conduction band. The case of P in Si corresponds to the conventional shallow donor limit: its *s*-like deep level lies above the conduction band minimum (CBM), and the extra electron of neutral P, which would occupy that level, spills out and falls to the conduction band edge, ionizing the P. The extra Coulomb potential of P^+ then binds the electron (at zero temperature) in a large-radius hydrogenic shallow impurity level, which is easily ionized thermally: in this case P is a shallow donor impurity (Fig. 1a). Si on a Ga site in GaAs and Ga on a Zn site in ZnSe are similarly shallow impurities.

If the conduction band edge of Si were above the deep level (Fig. 1b), then the extra P electron would occupy the deep level, and the P would be neutral instead of ionized. The deep level would be capable of trapping another electron of opposite spin (converting P^0 to P^-) or another hole (converting P^0 to P^+). In this hypothetical case of a deep level in the gap, P would be a deep impurity. Of course, P in Si is not a deep impurity but Bunker *et al.*[3] have shown that in Si_xGe_{1-x} there is a range of alloy compositions x, such that the conduction band edge is above the P *s*-like deep level and P is indeed a deep trap. Other similar deep traps are oxygen on a P site in GaP, Si on a Ga site in $Al_{0.7}Ga_{0.3}As$, and Ga on a cation site in $Zn_{0.5}Mn_{0.5}Se$.

A third possibility exists, namely that the *s*-like

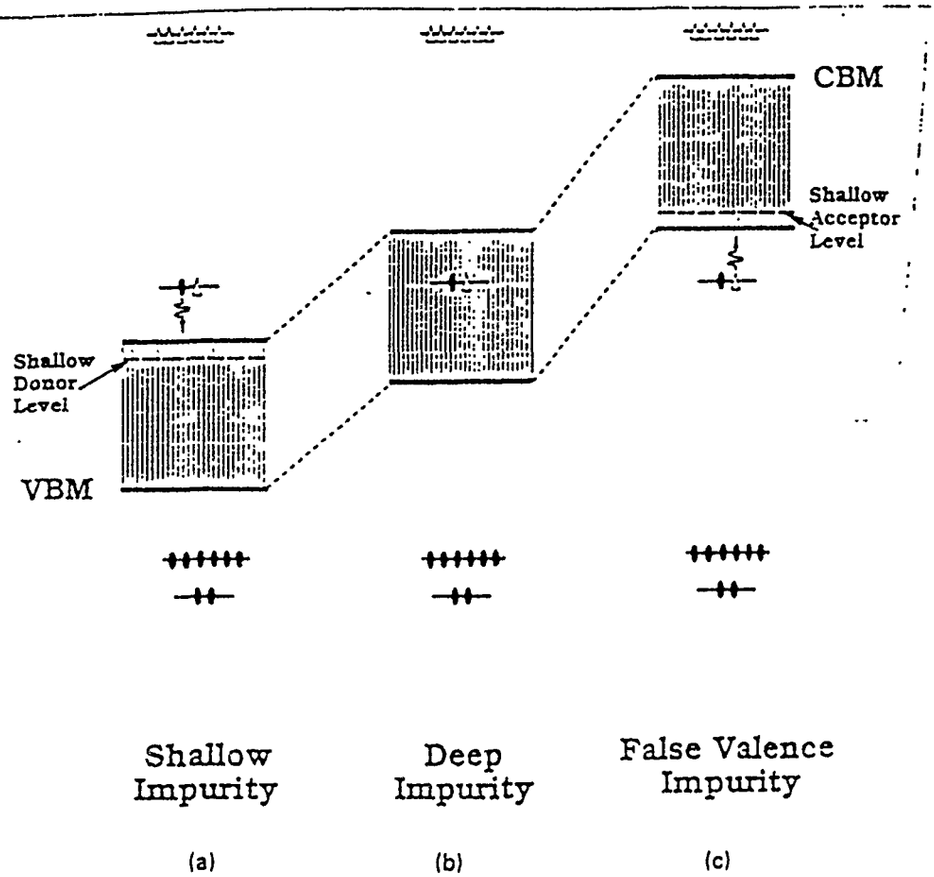


Fig. 1 — Illustration of how the deep level structure relative to the band edges determines whether an impurity is (a) shallow, (b) deep, or (c) false valence. Electrons (holes) are denoted by filled circles (open triangles). It is assumed that the impurity is from one Column to the right in the Periodic Table of the atom it replaces, such as P in Si. Deep (shallow) levels are solid (dashed) lines and the band gap is striped. CBM and VBM denote the conduction band minimum and the valence band maximum, respectively.

P deep level might lie *below* the valence band edge of Si (Fig. 1c). In this case the hole that would naturally occupy that deep level were it in the gap would instead bubble up to the valence band maximum: the P impurity would be negatively charged, and the charged impurity would bind a hole in the resulting shallow acceptor level. In this hypothetical case, we would say that the P impurity has a *false valence* of minus two relative to its normal valence, in that it would act as though it came from Column III of the Periodic Table instead of Column V, having the character of a shallow acceptor. Of course, the resultant P deep level does not actually lie below the valence band maximum, and so this case—which appears to be quite preposterous to persons who think in terms of the old effective-mass ideas—does not occur for P in Si. But it does occur for In on a Te site in $Pb_{1-x}Sn_xTe$: In is well known to be a donor in PbTe and an acceptor in SnTe—a behavior that is now known to be a consequence of the small band gap passing through a *p*-like deep level as a function of alloy composition[4] (rather than In changing its site). As a rule, false valence impurities are only to be expected in small band-gap semiconductors because false valence is associated with a level moving across the gap and is most easily achieved if the gap is small.

Shallow-deep transitions were first demonstrated as a function of alloy composition by Wolford and Streetman for the N isoelectronic impurity substituting for P and As in $GaAs_{1-x}P_x$ alloys (Fig. 2).[5] Their systematic studies on carefully ion-implanted alloys showed that the Column VI dopants S and Se produced hydrogenic effective-mass donor levels that followed the band edges as a function of alloy composition x, but that oxygen produced a deep level that was unattached to any band edge and varied linearly with x. Column-V nitrogen behaved like Column-VI oxygen, as a function of x, although its energy level in GaP is shallower energetically than S's or Se's, and for $x < 0.2$ the N level disappears into the conduction band. These data showed that the N level is actually a deep level, similar to oxygen's and distinct from the shallow levels, and that the deep level passes into the conduction band rather easily. (For simplicity, we call this a shallow-deep transition even though the N, being isoelectronic to As and P, does not produce a shallow donor level.) They also indicated that S and Se must also have deep levels similar to those of N and oxygen, lying *above* the conduction band minimum. Finally, since N is isoelectronic to P and As, the defect potential responsible for producing the deep level is, to a good approximation, confined to the central cell—and so

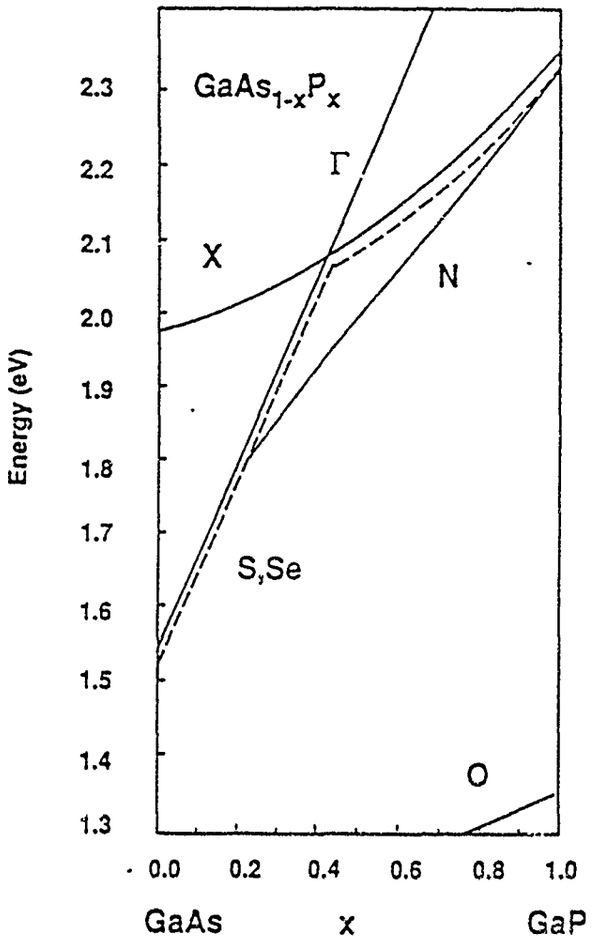


Fig. 2 — Schematic summary of the energy levels of $GaAs_{1-x}P_x$ with N, O, S, and Se anion-substitutional impurities, versus alloy composition, as determined by Wolford *et al.*[8] The zero of energy is the valence band maximum. Γ and X label the conduction band minima at $\vec{k} = 0$ and $\vec{k} = (2\pi/a_L)(1,0,0)$ of Brillouin zone, respectively. The thin solid lines labeled N are the measured nitrogen and oxygen deep levels, and the ned lines are the S and Se shallow levels.

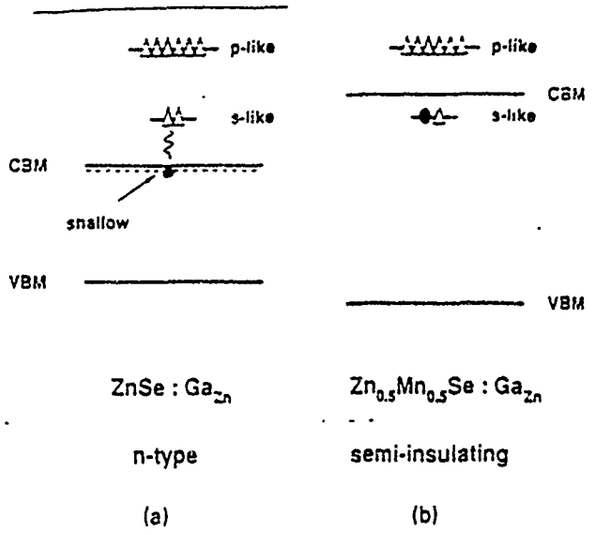


Fig. 3 — Schematic energy level diagram of Ga_{Zn} in (a) ZnSe and (b) $Zn_{0.5}Mn_{0.5}Se$, illustrating the *s*-like and *p*-like deep levels. Holes (electrons) are denoted by open triangles (closed circles). The lowest shallow level is dashed. In ZnSe the *s*-like deep level of Ga is above the conduction band minimum (CBM), causing its electron to be autoionized and trapped in the resulting shallow level (at zero temperature). In this case, Ga is a shallow donor impurity, making ZnSe *n*-type. The *s*-like level is in the gap for $Zn_{0.5}Mn_{0.5}Se$, is a deep trap (for either an electron or a hole), and makes the material semi-insulating.

the data demonstrated that deep levels are due to the central-cell potential, giving rise to the modern definition of a deep level. All of these ideas were revolutionary and controversial at the time they were first proposed, but are now well-established. Some other impurities that exhibit shallow-deep transitions as functions of alloy composition are Si on a Ga site in $A\ell_xGa_{1-x}As$[1,6,7] and Ga on a cation site in $Zn_{1-x}Mn_xSe$[4] (See Fig. 3). For small x, these impurities are donors, causing the host semiconductor to be *n*-type. For larger x they become deep traps, inhibiting conductivity, and producing semi-insulating behavior.

III. SHALLOW-DEEP TRANSITIONS VERSUS LAYER THICKNESS IN SUPERLATTICES

In the alloy hosts, the band edges vary with alloy composition and pass through the deep levels, leading to shallow-deep transitions, which are now well-documented.

Analogous transitions occur in superlattices as functions of the layer thicknesses.[7,9–13] To understand this, consider Si *in the middle of a GaAs layer* in a $GaAs/A\ell_{0.7}Ga_{0.3}As$ superlattice, and recognize that a Si on a Ga site in GaAs is a shallow donor impurity, but that Si on an $A\ell$ site in $A\ell As$ is a deep trap. As an initial approximation, think of the absolute energy of the Si deep level as being independent of the GaAs layer thickness.

If the GaAs layer is *thick*, then the band gap of the superlattice is approximately equal to the band gap of GaAs; if the layer is *thin*, then the superlattice band gap is almost the $A\ell As$ gap.[14] Hence, by reducing the thickness of the GaAs layers, it is possible to move the superlattice conduction band edge from the energy of the bulk GaAs conduction edge to that of $A\ell As$. This is a quantum confinement effect. Since the *s*-like deep level of Si is relatively constant in energy, the superlattice band edge *moves up through the Si deep level* as the GaAs layer-width decreases—and Si is predicted to undergo a shallow-deep transition. This is illustrated in Fig. 4 for the thick GaAs wells in an 18 × 18 $GaAs/A\ell_{0.7}Ga_{0.3}As$ superlattice and for thin GaAs wells in a 2×34 superlattice. In thin GaAs quantum wells, Si on a Ga site is predicted to be a deep trap.

Figures 5 and 6 show how relatively insensitive the deep levels in a superlattice are to the superlattice ordering, in comparison with the band edges. Note that the Si *s*-like level shifts only slightly near an interface (<0.1 eV) and assumes its bulk value near the center of a layer. The superlattice effects

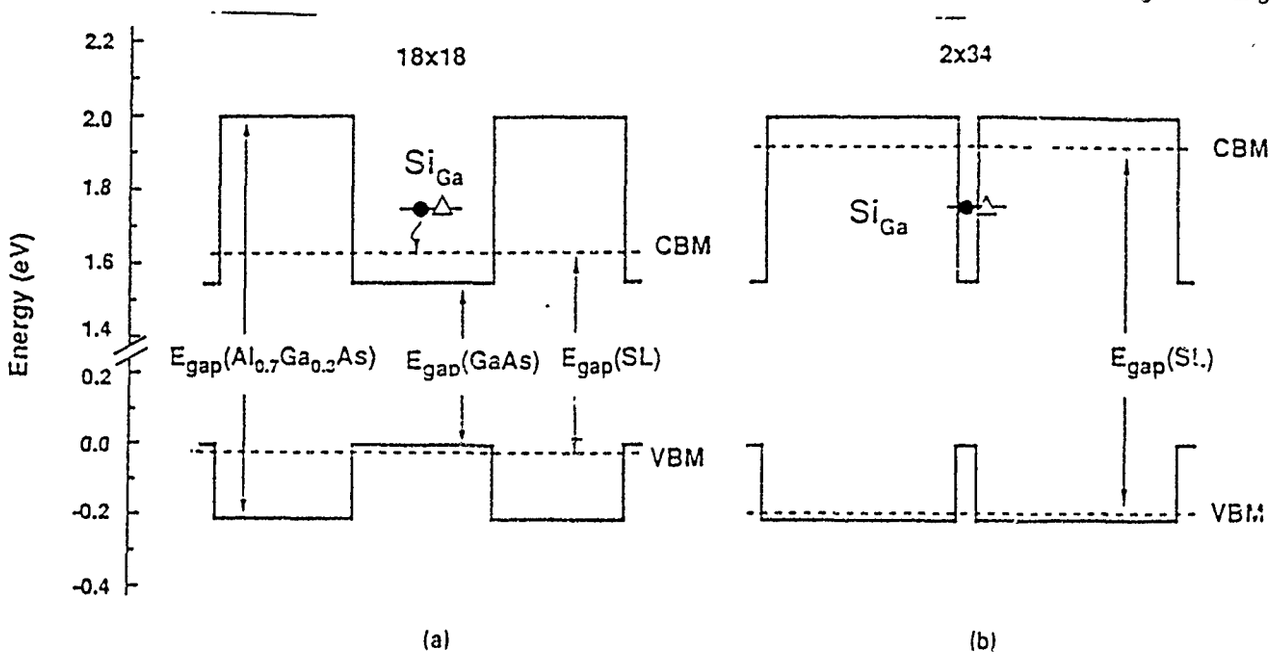


Fig. 4 — Schematic illustration of the shallow-deep transition as a function of layer thickness for a Si_{Ga} impurity near the center of a GaAs quantum well in a [001] $GaAs/Al_{0.7}Ga_{0.3}As$ superlattice. The bulk energy band edges vs position in the [001] direction are denoted by thick solid lines. The superlattice band edges are denoted by dashed lines. For the 18 × 18 superlattice, the well is 18 bi-atomic layers thick, and the Si deep level lies above the superlattice's conduction band minimum (CBM; dashed line)—causing Si to be a shallow donor. In the 2 × 34 superlattice, the *s*-like deep level is below the CBM of the superlattice, and Si is a deep trap. Note that the CBM of the superlattice is the first confinement level in a Kronig-Penney model.

have virtually disappeared for impurities more than three bi-atomic layers from an interface. (See Fig. 5.)

The *p*-like levels have similar behavior, splitting near an interface (of order 0.1 eV) with the highest-energy superlattice state of a cation vacancy in 10 × 10 $GaAs/Al_{0.7}Ga_{0.3}As$ corresponding to the *p*-like orbital oriented toward the GaAs and the lowest being directed toward the $Al_{0.7}Ga_{0.3}As$. (See Fig. 6.) The variations in the deep level energies that do occur are caused by the greater electropositivity of the Al and the valence band offset. For deep levels near the valence band maximum of the superlattice, the superlattice-induced splittings of the va-

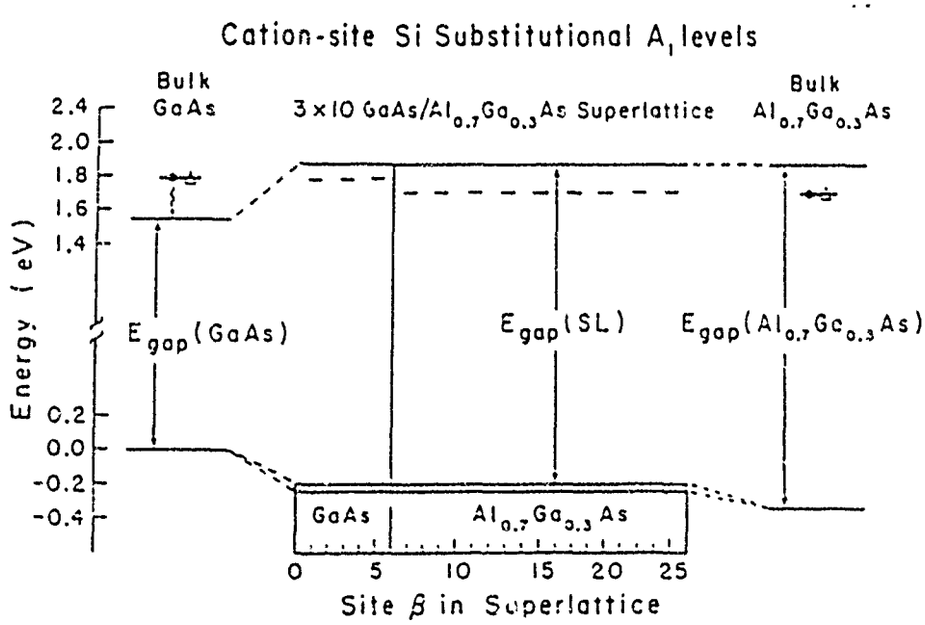


Fig. 5 — Predicted energy levels of a Si_{Ga} impurity in bulk GaAs, at various sites β in a 3 × 10 [001] $GaAs/Al_{0.7}Ga_{0.3}As$ superlattice and in bulk $Al_{0.7}Ga_{0.3}As$, after Ref. 7. Note that Si is a shallow donor in GaAs, and a deep trap in the superlattice and in $Al_{0.7}Ga_{0.3}As$. The cation sites of the superlattice correspond to β being an odd integer.

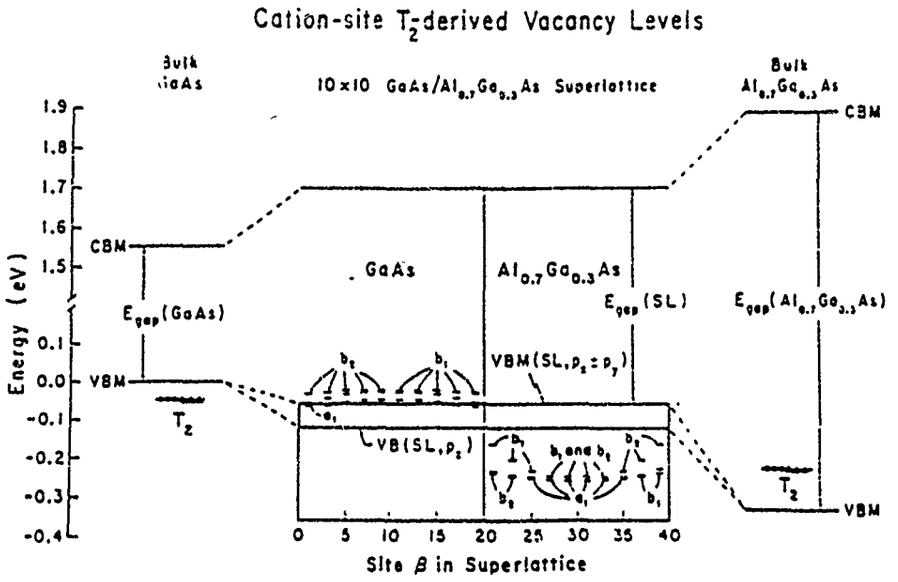


Fig. 6 — Predicted energy levels of a cation vacancy in bulk GaAs, at various sites β in a 10 × 10 [001] GaAs/$Al_{0.7}Ga_{0.3}As$ superlattice, and in bulk $Al_{0.7}Ga_{0.3}As$, after Ref. 7. Note the significant splitting of the valence band maximum in the superlattice, and the corresponding splitting of the p-like T_2-derived deep levels, even when the vacancy is distant from the interface (because the host spectral density of the superlattice is likewise split).

lence band and the host spectral density are reflected in the deep levels, even for levels distant from an interface.

In contrast to the deep levels, the band edges shift and split considerably from their bulk energies in the superlattice.

IV. APPLICATION TO THE II–VI DOPING PROBLEM

These ideas about deep levels can be used to circumvent the II–VI doping problem. Stated simply, the II–VI semiconductors are prime candidates for optical semiconductors operating in the green, blue, and ultraviolet portions of the spectrum, but are limited by the difficulty of doping them both n- and p-type—which is required for fabricating light-emitting diodes. The p-type doping is especially problematic in II–VTs. A notable exception is ZnTe, which is rather easily doped p-type, but not n-type—in contrast to ZnSe and most other II–VI semiconductors which are n-type dopable but not p-type. Thus a major mystery has been *why ZnTe's doping properties are so different from ZnSe's* and those of other II–VI's.

We proposed the following explanation of ZnTe's unusual proclivity toward p-type doping:[15,16] All but the most carefully grown ZnSe and ZnTe have defects which produce p-like deep levels at energies slight *above* the valence band maximum of ZnSe, but *below* the valence band edge of ZnTe. As a result, the deep level traps holes in ZnSe (making the material semi-insulating), but provides free holes in ZnTe (making ZnTe p-type). Namely, if random $ZnSe_{1-x}Te_x$ alloys could be grown, these deep levels would cause the defect to undergo a deep to shallow transition with increasing x.

We speculate that the defect responsible for the relevant deep level is antisite Zn, namely Zn_{Se} and Zn_{Te}, or, in the case of Li doping, an antisite dopant Li_{Se} or Li_{Te}.[17] Our analysis, however, does not depend on either of these identifications, and the picture we propose depends only on there being a defect with a deep level slightly above the valence band maximum capable of trapping holes in ZnSe; the corresponding deep level in ZnTe must lie below the valence band maximum and donate holes to the valence band. (Fig. 7.)

Our goal is to construct a superlattice with a band gap near that of ZnSe, but with a superlattice valence band maximum that lies above the ZnSe hole trap and hence will convert it into a shallow acceptor. One way to do that is to embed thin layers of GaAs in ZnSe. The valence band edge of GaAs is at higher energy than that of ZnSe, while the conduction band minima of the two materials are at almost the same energy.[14] Therefore by choosing the thickness of the GaAs we can "tune" the superlattice's valence band edge so that it lies above the ZnSe hole trap, without greatly altering the energy of the conduction band edge. (See Fig. 8.) The resulting predicted (low temperature) band gaps are given in Fig. 9. Band gaps even further toward the blue are possible with $(GaAs)_{1-x}(ZnSe)_x$/ZnSe superlattices.[13]

V. SUMMARY

In summary we have shown that in typical superlattices the deep level energies do not change much: typically of order 0.1 eV for an impurity at an interface. The s-like levels shift and the p-like levels split with the p-orbitals directed toward one material having absolute energies close to the bulk

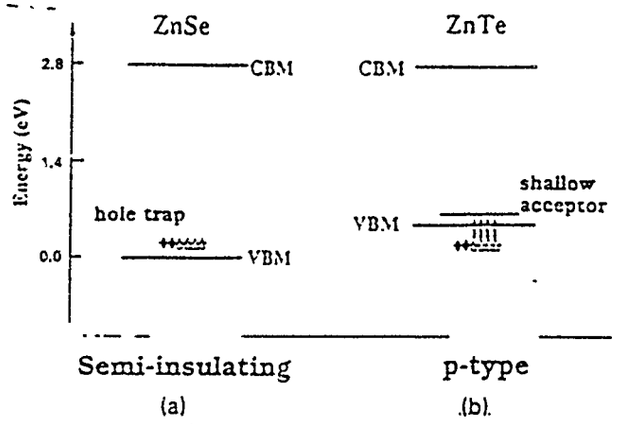


Fig. 7 — Illustration of the proposed shallow-deep transition explanation[17] of why ZnTe is easily doped *p*-type but ZnSe is not: *p*-like hole traps (a) in ZnSe above the valence band maximum instead (b) lie *below* the valence band maximum of ZnTe—and hence are shallow acceptors.

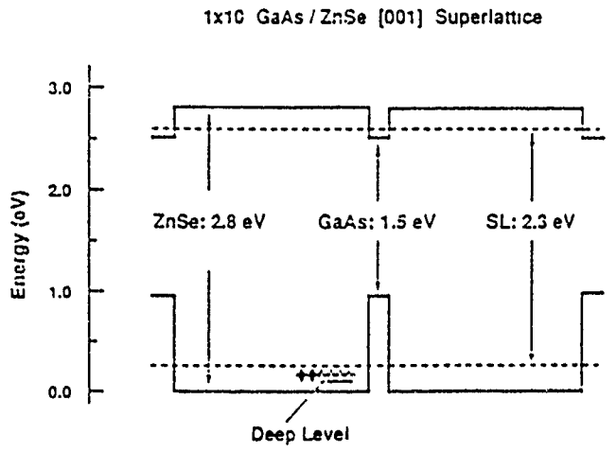


Fig. 8 — Schematic illustration of how thin-layers of GaAs in a GaAs/ZnSe superlattice can help cover up a deep level, which is a hole trap in ZnSe, causing it to become a shallow acceptor. The holes bubble up to the valence band maximum of the superlattice, after Ref. 18.

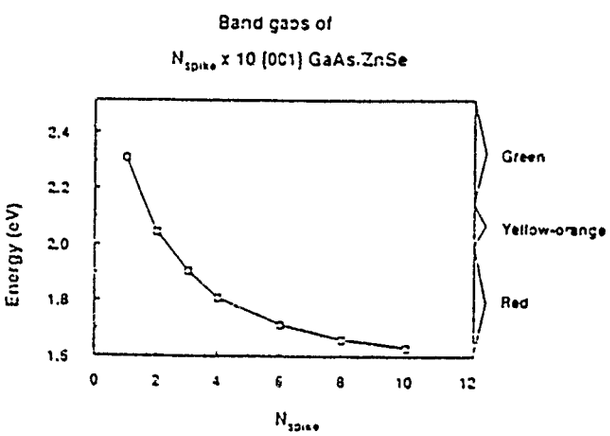


Fig. 9 — Predicted band gaps of N_{spike} × 10 [001] GaAs/ZnSe superlattices, showing how the band gap can be in the green for spike superlattices (small N_{spike}), after Ref. 18. N_{spike} is the number of GaAs layers in a period of the superlattice.

defect energy for that material (*e.g.* GaAs or $Al_{0.7}Ga_{0.3}As$). The interface splitting and shifts die out rather rapidly with distance from the interface, virtually disappearing within three atomic bi-layers unless the level is *p*-like with nearby *p*-like host states, in which case its splitting reflects the splitting of the host superlattice's spectral density.

In contrast to the deep levels, which are relatively unaffected by the superlattice order, the band edges of small-period superlattices exhibit the effects of quantum confinement, and so can lie anywhere between the lower and upper conduction band energies of the two bulk materials which constitute the superlattice—with the exact superlattice band edge energy depending on the layer thicknesses, especially the small-gap layer (*e.g.* GaAs). Therefore, the *interplay* between the confinement effects on the band edges and the insensitivity of the deep levels to the superlattice ordering can lead to shallow-deep transitions which change the character of a dopant from, say, an *n*-type donor to a semi-insulating deep trap.

The concept of shallow-deep transitions for *p*-like deep hole traps lying slightly above the valence band maximum of ZnSe has been proposed as a possible explanation of why ZnTe is relatively easily doped *p*-type, in contrast to other II–VI semiconductors. Finally this concept has been used to design GaAs/ZnSe small-period superlattices which have almost the band gap of ZnSe and almost the doping properties of GaAs.[18, 19]

ACKNOWLEDGMENTS

We are grateful to the Office of Naval Research and the Defense Advanced Research Projects Agency for their generous support (Contract Nos. N00014-89-J-1136 and N0530-0716-05).

REFERENCES

*Present address.
1. H. P. Hjalmarson and P. Vogl, Phys. Rev. Lett. *44*, 810 (1980). See also W. Y. Hsu, J. D. Dow, D. J. Wolford and B. G. Streetman, Phys. Rev. *B16*, 1597 (1977) for a discussion of the concepts on which this work is based.
2. W. Kohn, Solid State Phys. *5*, 257 (1957).
3. B. A. Bunker, S. L. Hulbert, J. P. Stott and F. C. Brown, Phys. Rev. Lett. *53*, 2157 (1984).
4. Craig S. Lent, M. A. Bowen, R. S. Allgaier, J. D. Dow, O. F. Sankey and E. S. Ho, Solid State Commun. *61*, 83 (1987).
5. D. J. Wolford, W. Y. Hsu, J. D. Dow and B. G. Streetman, J. Lumin. *18/19*, 863 (1979), and private communication.
6. H. P. Hjalmarson, private communication.
7. S. Y. Ren, J. D. Dow and J. Shen, Phys. Rev. *B38*, 10677 (1988).
8. R.-D. Hong and J. D. Dow, "Doping $Zn_{1-x}Mn_xSe$ n-type," Appl. Phys. Lett., in press.
9. R.-D. Hong, D. W. Jenkins, S. Y. Ren, and J. D. Dow, MRS Symp. Proc. *77*, 545 (1987), Interfaces, Superlattices, and Thin Films, eds. J. D. Dow and I. K. Schuller.
10. S. Y. Ren and J. D. Dow, J. Appl. Phys. *65*, 1987 (1989).
11. J. D. Dow, S. Y. Ren and J. Shen, NATO Advanced Science Institutes Series *B183*: Properties of Impurity States in Superlattice Semiconductors, (Plenum Press, New York, 1988).

pp. 175–187, eds. C. Y. Fong, Inder P. Batra and S. Ciraci. See also the other papers in this book for an overview of the field.

12. S. Y. Ren and J. D. Dow, J. Appl. Phys. 65, 1987 (1989).
13. J. D. Dow, J. Shen, and S. Y. Ren, in "Progress in Electronic Properties of Solids," Physics and Chemistry of Materials with Low-dimensional Structure, Festschrift in honor of Professor Franco Bassani, eds. E. Doni, G. Pastori Parravicini, A. Quatropani and R. Girlanda, (Kluwer, Dordrecht, 1989).
14. This limit on the band gap of the superlattice may not be obvious in Kronig-Penney models, in which the conduction band edge of the superlattice is the first confined quantum well state.
15. J. D. Dow, R.-D. Hong, S. Klemm and S. Y. Ren, "A proposed explanation of the *p*-doping proclivity of ZnTe," to be published.
16. S. Y. Ren, J. D. Dow and S. Klemm, J. Appl. Phys. 66, 2065 (1989).
17. S. Y. Ren, J. Shen, R.-D. Hong, S. Klemm, M.-H. Tsai and J. D. Dow, Surf. Sci., to be published.
18. J. Shen, J. D. Dow and S. Y. Ren, J. Appl. Phys., 67, 3761 (1990).

PD 505
58
228, 4p-521 (1990)
Surf. Sci

Bio Blank

Physica C 162-164 (1989) 1349-1350
North-Holland

Paper referenced #255 in the list of publications.

Self-Consistent Antiferromagnetic Ground State for La_2CuO_4 and CuO Via Energy Band Theory

Robert V. Kasowski[1], M.-H. Tsai[2], John D. Dow[2], M. T. Czyzyk[3]

[1]E. I. du Pont de Nemours & Company, Central Research & Development Department, Experimental Station, Wilmington, De. 19880-0356
[2]Physics Department, University of Notre Dame, Notre Dame, Indiana 46556
[3]Catholic University, Nijmegen, Netherlands

We have used the pseudofunction (PSF) method to compute self-consistent spin- polarized energy bands for La_2CuO_4 and CuO. The ground state is found to be semiconducting and anti-ferromagetic (AF) for both La_2CuO_4 and CuO. For La_2CuO_4 the Cu moment is $0.35\mu_B$. For CuO, the moment is 0.68 μ_B on the Cu site and 0.19 μ_B on the O site. The moments are in agreement with neutron diffraction data.

The pseudofunction (PSF) method[1] with a local density potential[2] has been used to compute spin-polarized energy bands for La_2CuO_4 (214) and CuO which give properties in good agreement with experiment. The PSF method uses a local orbital basis set employing the full potential including the core states and non-spherical corrections throughout the unit cell. Nine s, p, d basis functions were used on the metal atoms with s, p sets on the O atoms. The basis functions are continually changed during iteration to self-consistency so as to optimize the description of the charge density. The pseudofuntions and the non-spherical part of the potential are expanded in plane waves.

In Figure 1, the spin polarized bands near E_f are plotted for 214. The band gap between filled and empty states varies from approximately 2 eV at Γ to 0.06 eV on the hexagonal face. The bands are relatively flat in the Z direction because of the layered structure and are not shown. For the spin-polarized bands, the gap varies from 0.35 eV at X and M to 0.06 along the M to N and N to S directions. The necessity that bands be narrow on this hexagonal face thus becomes obvious since the spin effect which gives the semiconducting gap varies from 0.35 eV to 0.06 eV. This band is 0.17 eV wide for the PSF method[3] and approximately 0.5 eV for other methods[4].

The spin polarized energy bands for CuO give a larger gap of 0.2 eV. The recent crystal structure determination with neutrons was used[3]. There are 16 atoms in a monoclinic cell. The non-spherical potential was expanded in 18513 plane waves in order to allow very accurate description of the potential.

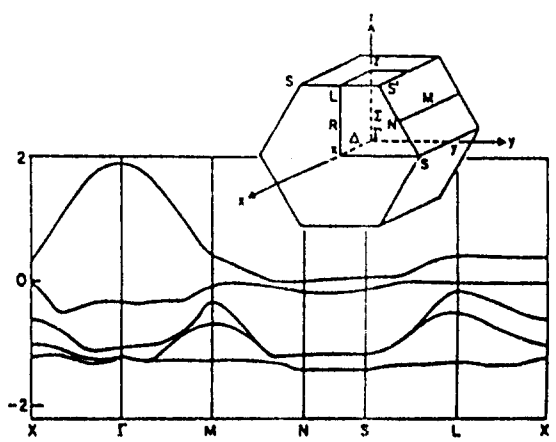


Figure 1) Spin polarized energy bands near the Fermi energy for La_2CuO_4

0921-4534/89/$03.50 © Elsevier Science Publishers B.V. (North-Holland)

In Table I, the moments on the Cu and O sites are shown with comparison to recent neutron diffraction data. The agreement is very close indicating that the local density energy bands are giving an accurate description of the ground state.

Table I. Comparison of calculated magnetic moments with experiment.

	Calculated	Experiment
	La_2CuO_4	
Cu	$0.35\mu_B$	$0.35\mu_B$
O	0.0	small
	CuO	
Cu	$0.68\mu_B$	$0.65\mu_B$
O	$0.19\mu_B$	$0.14\mu_B$

The importance of the description of the potential is illustrated in Figure 2 where only the band at the Fermi energy is shown. These calcualtions are non spin polarized as we wish to illustrate how the band width is sensitive to description. The band at E_f is very broad in the top panel because the potential was limited to only 4913 plane waves. The band in the bottom panel is very narrow because 15,625 plane waves were used. The wave functions and their plane wave description were identical for both calculations. This result apppears to be a convincing means of relating band widths and accuracy of description.

Finally, the CuO bands were also calculated with spin polarized augmented spherical wave(ASW) method. Identical postions, unit cells and k point sampling were used in the ASW and PSF calculations. The ASW method found the ground state to be metallic with no magnetic moment on the Cu or O sites contrary to the results obtained with the PSF method. Thus, the ASW and PSF methods give different solutions.

In summary, the itinerant energy band model yields a band structure which has a moment of approximately the correct magnitude on the Cu site and O sites for both 214 and CuO.

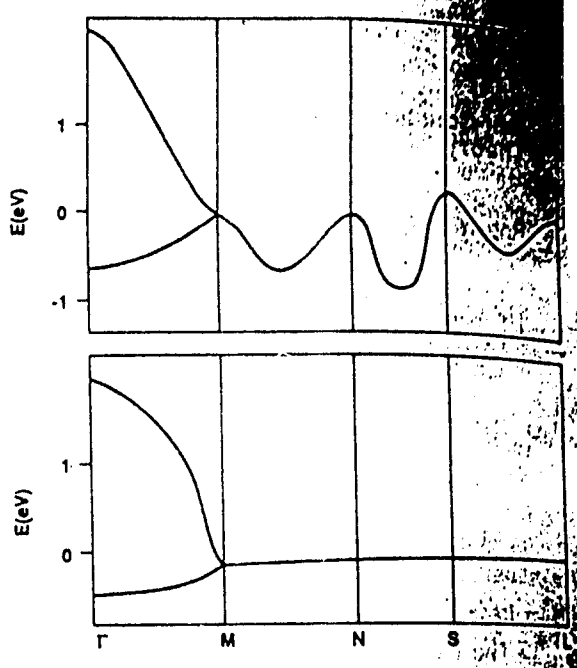


Figure 2) Non-spin polarized energy band at E_f for 4913 and 15356 plane waves in potential.

References

1. R. V. Kasowski, M.-H. Tsai, T. N. Rhodin, and D. D. Chambliss, Phys. Rev. B34, 2856 (1986).

2. U. von Barth and L. Hedin, J. Phys. C: Solid St. Phys. 5, 1629 (1982).

3. J.B. Forsyth, P.J. Brown, and B.M. Wanklyn, J. Phys. C: Solid State Phys. 21, 1988(2917).

4. T. C. Leung, X. W. Wang, and B. N. Harmon, Phys. Rev. B37, 384 (1988).
P. A. Sterne, C. S. Wang, and W. M. Temmerman, Phys. Rev. B (May 1, 1988).

Best Available Copy

Paper referenced #259 in the list of publications.

PHYSICAL REVIEW B VOLUME 42, NUMBER 14 15 NOVEMBER 1990-I

Relaxed-lattice model of isolated and paired isoelectronic traps in GaP

Jun Shen, Shang Yuan Ren, and John D. Dow
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556
(Received 25 October 1989)

A simple model of isolated and paired isoelectronic substitutional traps in GaP is presented that successfully explains the anomalous monotonic-in-n ordering of the nth nearest-neighbor $(NN)_n$-pair levels, with pair energies that vary as the inverse cube of the N-N separation. The model is a multiband deep-level theory that includes the effects of lattice relaxation around the impurities. It is predicted that $(Bi,Bi)_n$ pairs will not exhibit the monotonic-in-n ordering of NN_n pairs, that $(Bi,Bi)_1$ and $(Bi,Bi)_2$ will be resonant with the valence band, and that hydrostatic pressure can drive the Bi trap into the valence band.

I. INTRODUCTION

Although luminescence associated with isoelectronic P-substitutional defects such as Bi and N in GaP has been studied for over a quarter of a century,[1-3] there still is no theory which can explain all of the following facts: (i) isolated Bi_P produces a deep hole trap $\simeq 40$ meV above the valence-band maximum,[4,5] (ii) isolated N_P produces an electron trap 11 meV below the conduction-band minimum,[2] (iii) luminescence associated with $(Bi,Bi)_n$ pairs is not observed, and (iv) discrete luminescence lines associated with $(N,N)_n$ pairs are observed with the energy of the nth nearest-neighbor pair being simply related to the separation R_n between the two N atoms in the pair:

$$E(NN_n) \simeq E(N) - \beta R_n^{-3} \quad \text{for } n \geq 2 .$$

Here NN_n refers to the nth nearest-neighbor pair (with both N impurities on anion sites), $E(N)$ is the energy of isolated N, and β is a constant. Indeed, existing theories have been incapable of satisfactorily describing the observed ordering of the NN_n-pair levels, and do not even predict the observed monotonic variation with the separation R_n (or the neighbor number n).

In this paper we shall assume the accepted viewpoint that the isoelectronic trap levels are deep levels[6-8] due to the central-cell potentials of N and Bi, and we shall demonstrate that (i) a strain-free model cannot explain the observed monotonic-in-n behavior of the NN_n-pair lines, namely, if the impurities and host atoms are assumed to occupy sites on the undistorted lattice (i.e., if the effects of strain due to the size mismatch of impurity and host are neglected), then the energies of the NN_n-pair lines are not monotonic functions of n or R_n, and (ii) if strain is incorporated in a simple model, then all of the above unexplained facts can be easily understood.

The successful theories of the prototypical isoelectronic trap, isolated substitutional N in GaP, have all assumed that virtually the entire defect potential is localized in the central cell; the theories differed, however, in the number of host bands significantly mixed into the chemical bonds of this defect. The earlier theories, notably by Faulkner[3] and by Hsu *et al.*,[9] employed only one band until Swarts *et al.*[10] demonstrated that the natural defect potential had a strength so large ($\simeq 7$ eV) that a multiband model is necessary to renormalize the defect potential. It is now generally accepted that the minimum number of bands required to produce the sp^3 chemical bonding is eight: The N trap level is a "deep" level originating from the central-cell potential, as described by the Green's-function theory of Hjalmarson *et al.*[7]

II. INADEQUACY OF STRAIN-FREE THEORIES

Efforts to extend the isolated-N theory to NN_n pairs[11] have not produced the observed monotonic dependence on n (or R_n); see Fig. 1. While this failure might appear, at first glance, to be a property of the specific theoretical models, closer examination of the theories reveals that it is a general property of models which omit the effects of lattice strain and which instead attempt to obtain the NN_n-pair energies from electronic coupling alone, while constraining the N impurities and the host atoms to sites of a perfect zinc-blende lattice. To see this, consider the Schrödinger equation for an isolated-N impurity at site A ($A=0$ or $A=R_n$):

$$(H_0 + V_A)\phi_A = E(N)\phi_A .$$

Here ϕ_A is the deep-level wave function of an isolated-N impurity at site A, and V_A is the defect potential. For an NN_n pair, with one N impurity at the origin and the other at R_n, the Schrödinger equation is

$$(H_0 + V_0 + V_{R_n})\Phi = E(NN_n)\Phi ,$$

with approximate bonding and antibonding orbitals

$$\Phi \simeq 2^{-1/2}(\phi_0 \pm \phi_{R_n}) .$$

 ©1990 The American Physical Society

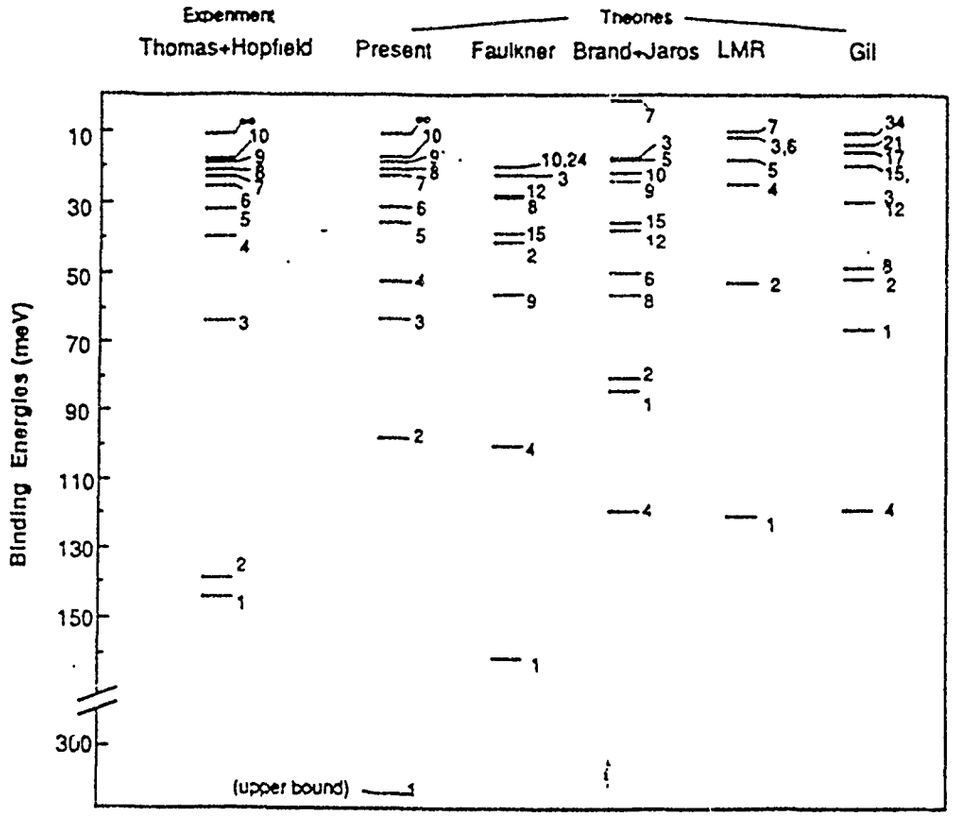


FIG. 1. Binding energies of NN_n pairs observed by Thomas and Hopfield (Ref. 2), and calculated with the present model, by Faulkner (Ref. 3), by Brand and Jaros (Ref. 16), by Li *et al.* (LMR) (Ref. 15), and by Gil *et al.* (Ref. 17). The zero of energy is the free-exciton luminescence line. Note the ordering of levels with n. The NN_1 binding energy for the present theory is an upper bound estimated by assuming $d'=d_0$; see text.

The corresponding energy is

$$E(NN_n)=E(N)+(\pm\langle\phi_0|V_0|\phi_{R_n}\rangle+\langle\phi_0|V_{R_n}|\phi_0\rangle)\times(1\pm\langle\phi_0|\phi_{R_n}\rangle)^{-1}$$

or

$$E(NN_n)\simeq E(N)\pm\phi_0(0)\phi_{R_n}(0)v_0$$

for large R_n.[12] (This approximate result agrees well with more exact calculations.) Here the overlap integral is $\langle\phi_0|\phi_{R_n}\rangle$; we have assumed that R_n is moderately large, and we have taken the defect potential to have strength v_0 and to be localized within the impurity's cell. Hence the NN-pair energy $E(NN_n)$ has the same dependence on R_n as the isolated-N deep-level wave function centered at R_n: $\phi_{R_n}(0)$. Since this wave function is known to oscillate[13] as a function of R_n and is not monotonic, we conclude that the observed NN_n-pair energies, which are monotonic, are not determined predominantly by the electronic coupling between the two impurities.

III. ROLE OF STRAIN

Since the electronic coupling between the two N impurities in the NN_n pairs does not produce the observed monotonic-in-n ordering of the pair levels, there must be some other effect which is larger in magnitude than the electronic coupling and which produces monotonic level ordering. We shall show that elastic strain due to deformation of the host lattice in the vicinity of the defect is responsible for the observed ordering and also produces the correct magnitudes for the energies of the NN_n pairs.

We are not the first to suggest strain as the mechanism primarily responsible for the NN_n-pair energy levels in GaP. Twenty years ago Allen[14] proposed that strain, and consequentially deformation potentials, determined the NN_n-pair binding energies in GaP. His theory fell into disfavor, however, because his N-related levels were rigidly attached to the conduction band edges and would have preserved their energies, with respect to the conduction minimum in $GaAs_{1-x}P_x$ as the alloy composition x varied—contrary to the observations.[9] The important demonstration of Allen is that strain produces an effect *of the correct order of magnitude* to explain the NN_n-pair data. This raises the question of how the strain-free theories were able to obtain NN_n-pair energies of the correct order of magnitude, albeit incorrectly ordered: If both the electronic coupling and strain effects are of the required order of magnitude to explain the data, then a correct theory must account for the competition between these effects. However, in the two multiband theories, namely those of Li *et al.*[15] and Jaros and Brand,[16] the

coupling between N atoms or the defect potential strength is an adjustable parameter—and this parameter leads to the large apparent electronic coupling with incorrect ordering. The theories of Faulkner[3] and Gil *et al.*[17] are missing the important effects of the valence band[7] on the impurity levels. Thus the previous theories that attempted to explain the NN_n-pair energies in terms of a strain-free electronic-coupling model considerably overestimated the effect.

IV. ISOLATED IMPURITIES

We incorporate the effects of lattice deformation and strain around an isolated impurity using the widely accepted sp^3s^* tight-binding model of Hjalmarson *et al.*[7] This model successfully described the isolated-N trap in $GaAs_{1-x}P_x$ and many other deep levels in a wide variety of semiconductors. Most of the published calculations based on this model did not incorporate the effects of strain or lattice relaxation, however, although lattice distortions are easily incorporated into the nearest-neighbor tight-binding model Hamiltonian, because the various matrix elements T between orbitals centered on adjacent sites obey Harrison's scaling rule and so are approximately inversely proportional to the square of the bond length d: $T=T_0(d_0/d)^2$. Therefore, the defect-potential matrix V of, say a N_P defect in GaP, is localized to the impurity site and its four neighbors and to the basis orbitals s, p_x, p_y, and p_z (Ref. 18) centered on those sites. The impurity-site diagonal elements are the same as given by Hjalmarson *et al.* for the strain-free theory[19]

$$V_s=\beta_s[w_s(\mathrm{N})-w_s(\mathrm{P})]\ ,$$

and

$$V_p=\beta_p[w_p(\mathrm{N})-w_p(\mathrm{P})]\ .$$

Here the w's are the atomic energies in the solid, as given by Vogl *et al.*[20]

The wave function of the N deep level is dominantly hostlike and Ga-dangling-bond-like in character. Hence, the effects of host d states are well simulated by the sp^3s^* model, which is known to treat the host electronic structure well, and the relativistic effects (which are important for Bi levels) can be neglected. The nonzero nearest-neighbor off-diagonal elements of the defect potential V are obtained from the host-crystal matrix elements, scaled for the altered (inwardly relaxed) bond length of the impurity with its four neighbors:

$$V_{\text{off diagonal}}=D=T_0[(d_0/d)^2-1]\ ,$$

where d is the relaxed bond length and has a value between $d_0(=2.36\ \text{Å})$ and the sum of the covalent radii of N and Ga, 2.01 Å. Details of the calculational procedure for obtaining the N deep-level energy are given in Appendix A.

V. PHYSICS OF THE RELAXATION EFFECT

The qualitative physics governing the strain-relaxation effect is displayed in Fig. 2, where we use a defect-molecule model. When a Ga atom (energy ε_{Ga}) and a P atom (energy ε_P) are brought together, a bonding and an

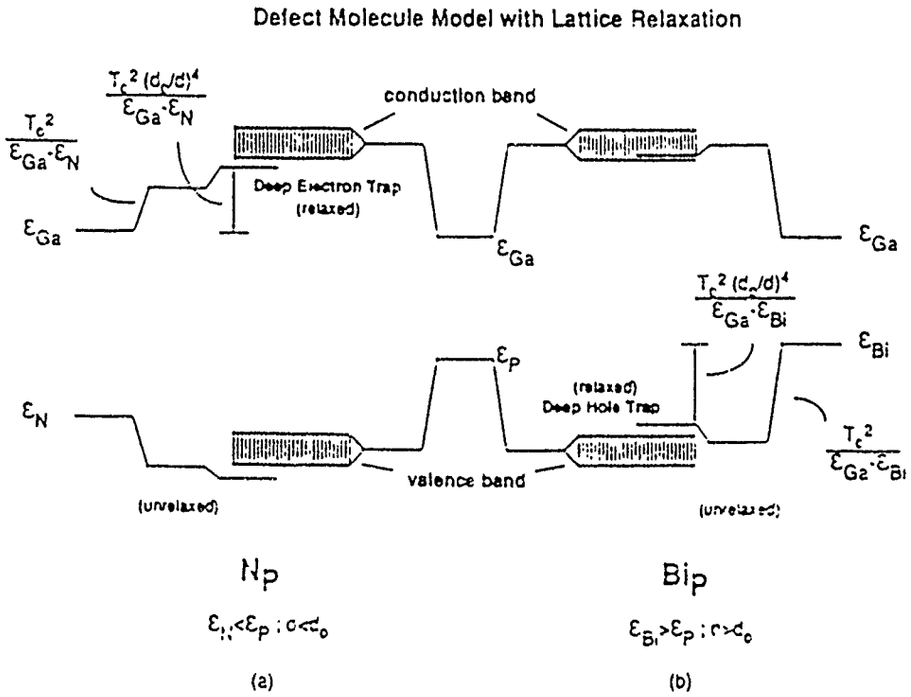


FIG. 2. Schematic illustration of the effects of the central-cell defect potential and lattice relaxation on the deep impurity levels of (a) N_P; (b) Bi_P in GaP. See text for discussion.

antibonding state will be formed. The bonding state is broadened in the crystal and forms the valence band, while the antibonding state produces the conduction band. When a N atom replaces a P atom ($\varepsilon_{Ga} > \varepsilon_P > \varepsilon_N$), its central-cell defect potential produces a bonding hyperdeep level in or below the valence band and an antibonding deep level in the band gap. The depth of the defect potential is related to the difference in atomic energies of P and N (Ref. 7), and is $\simeq -7$ eV. Because the nearest-neighbor transfer-matrix element T is almost the same for N—Ga and P—Ga bonds, it is the larger (by $\simeq 7$ eV) energy denominator $\varepsilon_{Ga} - \varepsilon_N$ which causes the defect's bonding-antibonding splitting [of order $T_0^2(\varepsilon_{Ga} - \varepsilon_P)^{-1}$ in the extreme tight-binding limit] to be smaller than the host's—leading to a deep level in the gap. Because N has a smaller covalent radius than P, the Ga collapses inward and the transfer-matrix element T increases, owing to Harrison's universal rule $T = T_0 d_0^2/d^2$ (Ref. 21). The antibonding deep level is pushed up due to such a lattice relaxation by an amount of order $T_0^2[(d_0/d)^4 - 1](\varepsilon_{Ga} - \varepsilon_N)^{-1}$. When the impurity atom is larger than the P host atom it replaces, as in the case of Bi, the bonding deep level is pulled up in energy by approximately $T_0^2[(d_0/d)^4 - 1](\varepsilon_{Ga} - \varepsilon_{Bi})^{-1}$ as a result of relaxation.

We assume that the relaxed bond length d is a linear function of Δr, the difference between the covalent radii of the impurity and the host:[22]

$$d = d_0 + \lambda \Delta r \ .$$

The parameter λ is necessarily positive and less than unity, since the relaxation around a vacancy (the smallest-radius "impurity") does not annihilate the vacancy. We determine the precise value of λ by requiring the theory for isolated N_P to produce *exactly* the observed energy level, namely 11 meV below the conduction-band edge.[23,24] Note that without the lattice-relaxation effect the N level would lie 270 meV lower in energy. (See Fig. 3.) It should be emphasized, however, that λ is a phenomenological parameter whose precise value should not be overinterpreted. We have performed calculations for a variety of values of λ, using various models of the defect potential,[19] and find the physics insensitive to the detailed parametric choice.[25]

At this point the theory is completely determined for isolated impurities, and we can compute the energy level of Bi_P—a defect that, without the lattice-relaxation effect, would give a T_2-symmetric deep-level resonant with the valence band. But Bi is larger than P and causes an outward relaxation of its neighboring Ga atoms—and this relaxation pushes the deep level up into the gap (see Fig. 3), where it lies gratifyingly close to the observed Bi energy.

According to the theory, neither As_P nor Sb_P produces a deep level in the fundamental band gap of GaP—in agreement with experiment.[26]

When hydrostatic pressure is applied to GaP, the Bi level should move down into the valence band[27] (see Fig. 4).

In GaAs, we predict that the N level descends into the fundamental band gap with increased pressure—as ob-

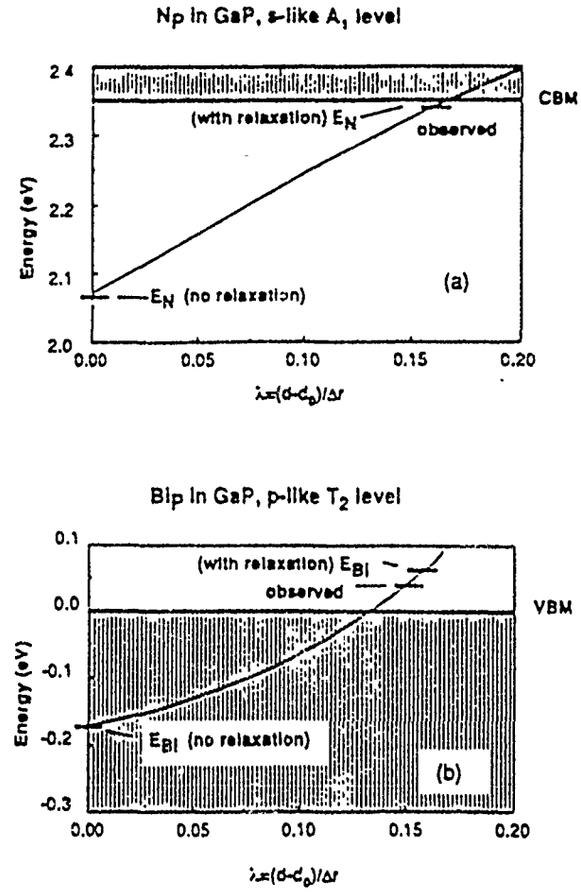


FIG. 3. Predicted results of lattice relaxation vs the parameter λ on (a) the s-like A_1 level of N_P and (b) the p-like T_2 level Bi_P in GaP. The conduction-band minimum is denoted CBM and the valence-band maximum is VBM. The shaded area corresponds to the energies of the host bands.

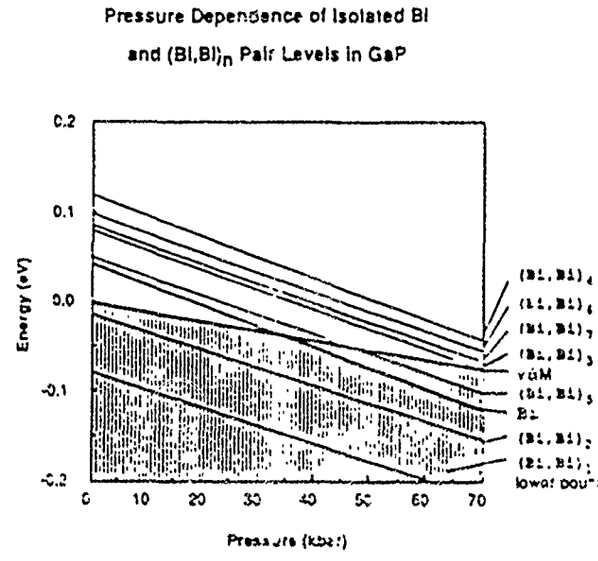


FIG. 4. Predicted dependences on pressure of the GaP valence-band maximum (VBM; dark solid line) and the isolated-Bi (dark solid line) and $(Bi,Bi)_n$-paired isoelectronic traps (light solid lines).

served[28] (see Fig. 5).

This same theory, applied to Bi on an anion site in GaAs or InP, produces a level above the valence-band maximum in InP that descends into the valence band with hydrostatic pressure and a level in the valence band of GaAs that remains there—in agreement with the observations.[29] Clearly, this simple model convincingly accounts for the principal experimental facts concerning isolated isoelectronic traps in GaP and other III-V semiconductors.

VI. NN_n pairs

The model Hamiltonian for a pair of impurities, such as NN_n, can be constructed from the isolated-impurity Hamiltonian as discussed in Appendix B. One new feature arises in the case of pairs: one must (in principle) determine the relaxed bond length d' for each bond using elasticity theory.

An important point is that the strain field created by one N impurity at the origin influences the nearest-neighbor bond lengths of the second impurity at R_n. One N, being smaller than P, causes the second N of an NN_n pair to move toward it by an amount approximately given by elasticity theory. The second N's neighboring Ga atom that is closest to the first N also moves toward the origin, causing a net elongation of the second N—Ga bond length, given by the gradient of the strain field:

$$d' = d(1 + A/R_n^3) \quad \text{for } n \geq 2 ,$$

where we have $A \simeq 2a^2\delta a$, and $a^2\delta a \simeq \pm 0.7904$ Å^3.[30] This dependence of the strained bond length on R_n^{-3} is reflected in the NN_n-pair energies. In the case of NN_1 pairs this expression for the bond length, which is based on linear-elasticity theory and the assumption that the strain fields of the two interacting N atoms do not interfere, breaks down. In this case, both N atoms will relax toward their common Ga neighbor, assuming a bond length less than d_0, GaP's bond length. Thus, for $n = 1$, we can only say $d' < d_0$. To simplify the calculation of the pair energies, we assume that a single stretched bond length d', that of the N—Ga bond closest to the first N atom, characterizes all of the N—Ga nearest-neighbor bonds of the second N. (In fact there is often a slight angular dependence to these bond lengths.) Then we obtain the energies of the NN_n pairs using the Green's-function method, as outlined in Appendix B.

The NN_n-pair energies are in excellent agreement with the data, as shown in Figs. 1 and 6. The relaxation of the strain is responsible for the R_n^{-3} and monotonic-in-n ordering of the NN_n-pair levels as observed first by Thomas and Hopfield.[2] Note that the other theories attempted unsuccessfully to explain the observed ordering in terms of electronic coupling between the two N atoms rather than in terms of strain.

It is noteworthy that the theory does *not* predict that $(Bi,Bi)_n$ pairs order monotonically with n or R_n. In this case the isolated-Bi states are p-like or T_2 symmetric and near the valence-band maximum. We find that isolated Bi lies in the gap, as observed, and that some $(Bi,Bi)_n$ pairs, for $n \geq 3$, lie in the gap as well. However, $(Bi,Bi)_1$ and $(Bi,Bi)_2$ lie in the valence band, due to the strain associated with its larger radius than the radius of P.

No $(Bi,Bi)_n$ pairs have been reported in GaP, to our knowledge. It stands to reason that for very large n these pairs will lie close to the isolated-Bi level—and so may not be resolved. For smaller n the formation of enough

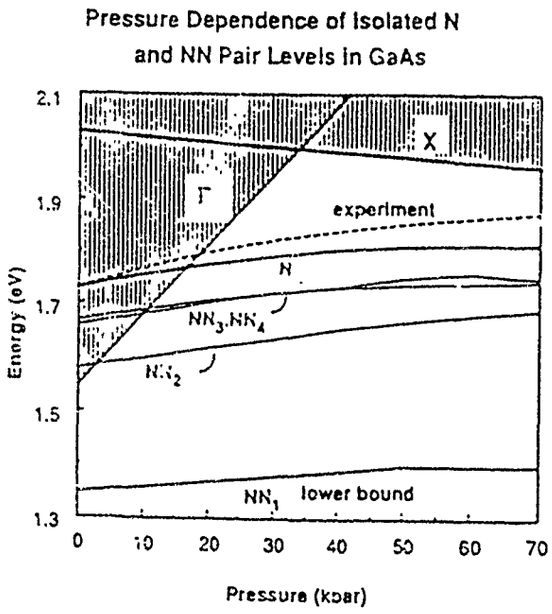


FIG. 5. Predicted dependences on pressure of the GaAs band edges Γ and X (heavy solid lines) and the isoelectronic trap N, in comparison with the data of Ref. 28. Some NN_n pairs are also shown, and have similar pressure dependences to the isolated-N level.

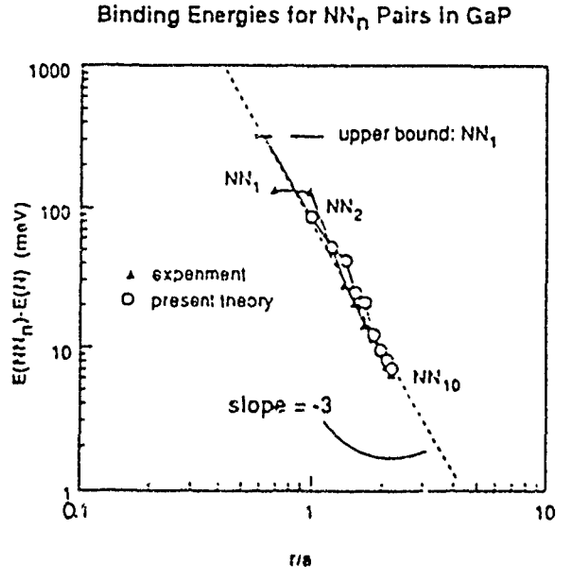


FIG. 6. Energies of NN_n-pair luminescence lines in GaP (with respect to the isolated N_P luminescence) vs pair separation R_n divided by the lattice constant of GaP, a. Open circles are the present theory for $n \geq 2$; data are denoted by solid triangles. Note the R_n^{-3} dependence caused by strain. The upper bound on the NN_1-pair energy obtained by assuming $d' = d_0$ is displayed as a bar.

pairs to observe may be limited by the Bi solubility. The interesting feature of the present theory is that it explains why $(Bi,Bi)_1$ and $(Bi,Bi)_2$ should not be observed, even if these pairs form in sufficient concentration, namely they lie outside the fundamental band gap. To a good approximation, the $(Bi,Bi)_n$ levels have the same dependence on hydrostatic pressure as isolated Bi, according to the theory. It would be interesting if these missing levels predicted by the theory, and their pressure dependences, were measured.

ACKNOWLEDGMENTS

We would like to thank D. J. Wolford for stimulating discussions of the data and the Office of Naval Research for their generous support (Contract No. N00014-89-J-1136).

APPENDIX A: ISOLATED IMPURITY

We consider an impurity (e.g., Bi_P) in GaP at the origin with nearest neighbors at the sites

$$\mathbf{v}_1=(1,1,1)a/4 ,$$

$$\mathbf{v}_2=(1,-1,-1)a/4 ,$$

$$\mathbf{v}_3=(-1,1,-1)a/4 ,$$

and

$$\mathbf{v}_4=(-1,-1,1)a/4 ,$$

where a is the lattice constant. On the impurity (anion or "a") site we have five orbitals $|0,a,s\rangle$, $|0,a,p_x\rangle$, $|0,a,p_y\rangle$, $|0,a,p_z\rangle$, and $|0,a,s^*\rangle$. Since the defect-potential matrix does not involve the excited state s^*, we ignore that orbital. The s orbital is a basis for the A_1 or s-like irreducible representation of the tetrahedral group, and the three p orbitals transform according to the T_2 representation. Another set of basis orbitals can be formed from the inward-directed sp^3 hybrids centered on the first shell of adjacent (cation or "c") sites to the impurity. They are

$$|1,A_1,s\rangle=(+|h_1\rangle+|h_2\rangle+|h_3\rangle+|h_4\rangle)/2 ,$$

$$|1,T_2,x\rangle=(-|h_1\rangle-|h_2\rangle+|h_3\rangle+|h_4\rangle)/2 ,$$

$$|1,T_2,y\rangle=(-|h_1\rangle+|h_2\rangle-|h_3\rangle+|h_4\rangle)/2 ,$$

and

$$|1,T_2,z\rangle=(-|h_1\rangle+|h_2\rangle+|h_3\rangle-|h_4\rangle)/2 ,$$

where the hybrids are

$$|h_1\rangle=(|\mathbf{v}_1,c,s\rangle-|\mathbf{v}_1,c,x\rangle-|\mathbf{v}_1,c,y\rangle-|\mathbf{v}_1,c,z\rangle)/2 ,$$

$$|h_2\rangle=(|\mathbf{v}_2,c,s\rangle-|\mathbf{v}_2,c,x\rangle+|\mathbf{v}_2,c,y\rangle+|\mathbf{v}_2,c,z\rangle)/2 ,$$

$$|h_3\rangle=(|\mathbf{v}_3,c,s\rangle+|\mathbf{v}_3,c,x\rangle-|\mathbf{v}_3,c,y\rangle+|\mathbf{v}_3,c,z\rangle)/2 ,$$

and

$$|h_4\rangle=(|\mathbf{v}_4,c,s\rangle+|\mathbf{v}_4,c,x\rangle+|\mathbf{v}_4,c,y\rangle-|\mathbf{v}_4,c,z\rangle)/2 .$$

If the defect potential matrix V associated with the impurity is confined to the central cell and its coupling is limited to first neighbors (the latter being a strain effect), then we have the nonzero matrix elements of V

$$V_i=\begin{array}{c} \begin{matrix} |0,a,s\rangle & |1,A_1,s\rangle & |0,a,x\rangle & |1,T_2,x\rangle & |0,a,y\rangle & |1,T_2,y\rangle & |0,a,z\rangle & |1,T_2,z\rangle \end{matrix} \\ \begin{pmatrix} V_s & D_s & 0 & 0 & 0 & 0 & 0 & 0 \\ D_s & V_s & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & V_p & D_p & 0 & 0 & 0 & 0 \\ 0 & 0 & D_p & V_p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & V_p & D_p & 0 & 0 \\ 0 & 0 & 0 & 0 & D_p & V_p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & V_p & D_p \\ 0 & 0 & 0 & 0 & 0 & 0 & D_p & V_p \end{pmatrix} \end{array} ,$$

where according to Harrison's rule,[22] the interatomic perturbation matrix elements are

$$D_s=U_s[(d_0/d)^2-1] ,$$

$$D_p=U_p[(d_0/d)^2-1] ,$$

and we have

$$U_s=T(s,s)-3T(sa,pc) ,$$

and

$$U_p=T(sc,pa)+T(x,x)+2T(x,y) .$$

Here d_0 and d are the perfect and relaxed bond lengths, respectively, and the matrix elements are the same as in Refs. 20 and 7. The deep impurity levels associated with defect are obtained by solving the secular equation

$$\det[1-G(E)V]=0 ,$$

where $G(E)=(E-H_0)^{-1}$ is the host Green's function, with matrix elements

$$G(\mathrm{R},b,i;\mathrm{R}',b',i')=\sum_{\nu,\mathbf{k}} \langle \mathrm{R},b,i|\nu,\mathbf{k}\rangle[E-E(\nu,\mathbf{k})]^{-1}\langle \nu,\mathbf{k}|\mathrm{R}',b',i'\rangle .$$

Here the sum is over the host bands ν and ten special points[24] in the Brillouin zone. Because of the simple form of the defect matrix, the eigenvalue equations for A_1 and T_2 levels become

$$[1-G(0,a,s;0,a,s)V_s]+[G(0,a,s;1,A_1,s)^2-G(0,a,s;0,a,s)G(1,A_1,s;1,A_1,s)]D_s^2-2G(0,a,s;1,A_1,s)D_s=0 ,$$

$$[1-G(0,a,x;0,a,x)V_p]+[G(0,a,x;1,T_2,x)^2-G(0,a,x;0,a,x)G(1,T_2,x;1,T_2,x)]D_p^2-2G(0,a,x;1,T_2,x)D_p=0 ,$$

threefold degenerate).

This calculation has not included the effects of bond-length changes between the first- and second-nearest neighbors for two reasons: (i) we wanted to present the simplest possible model containing the essential physics of (N,N) pairs, and (ii) these effects are very small, because they affect in first-order only the hybrid orbitals centered on the nearest-neighbor sites and directed *outward* from the impurity, whereas the level positions are determined primarily by the hybridization of the impurity orbitals with the *inward*-directed hybrids.

APPENDIX B: NN_n-PAIRED IMPURITIES

For paired impurities the defect potential V_2 is a direct sum matrix

$$V_2=V(0)+V(\mathrm{R}_n) ,$$

where each of $V(0)$ and $V(\mathrm{R}_n)$ is an 8×8 matrix, and the bond lengths in these matrices are the appropriate values d' for the pair (as opposed to d). The resulting secular equation is a 16×16 matrix:

$$\det[1-G(E)V_2]=0 ,$$

which is solved for the pair energies E. The ten special points were extended to 240 points by the 24 T_d symmetry operations[15] and then were used to evaluate the Green's-function matrix elements.

M. Gershenzon, D. G. Thomas, and R. E. Dietz, in *Proceedings of the International Conference on the Physics of Semiconductors, Exeter, 1962*, edited by A. C. Stickland (The Institute of Physics and Physical Society, London, 1962), p. 752.

D. G. Thomas, J. J. Hopfield, and C. J. Frosch, Phys. Rev. Lett. 15, 857 (1965); D. G. Thomas and J. J. Hopfield, Phys. Rev. 150, 680 (1966).

R. A. Faulkner, Phys. Rev. 175, 991 (1968).

F. A. Trumbore, M. Gershenzon, and D. G. Thomas, Appl. Phys. Lett. 9, 4 (1966).

P. J. Dean, J. D. Cuthbert, and R. T. Lynch, Phys. Rev. 179, 754 (1969); P. J. Dean and R. A. Faulkner, *ibid.* 185, 1064 (1969).

J. J. Hopfield, D. G. Thomas, and R. T. Lynch, Phys. Rev. Lett. 17, 312 (1966).

H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. 44, 810 (1980).

For a review, see J. D. Dow, in *Highlights of Condensed Matter Theory* (Proceedings of the International School of Physics "Enrico Fermi," Course 89, Varenna, 1983), edited by F. Bassani, F. Fumi, and M. P. Tosi (North Holland, Amsterdam, 1985), pp. 465.

A. Y. Hsu, J. D. Dow, D. J. Wolford, and B. G. Streetman, Phys. Rev. B 16, 1597 (1977).

C. A. Swarts, D. L. Miller, D. R. Franceschetti, H. P. Hjalmarson, P. Vogl, J. D. Dow, D. J. Wolford, and B. G. Streetman, Phys. Rev. B 21, 1708 (1980).

[11]O. F. Sankey, H. P. Hjalmarson, J. D. Dow, D. J. Wolford, and B. G. Streetman, Phys. Rev. Lett. 45, 1656 (1980).

[12]The approximation $E(\mathrm{NN}_n)\simeq E(\mathrm{N})=\phi_0(0)\phi_{\mathrm{R}_n}(0)w_0$ is remarkably close (within 10%) to the energies we calculate exactly.

[13]S. Y. Ren, Sci. Sin. A 27, 443 (1984); and S. Y. Ren, W. M. Hu, O. F. Sankey, and J. D. Dow, Phys. Rev. B 26, 951 (1982). The same oscillatory behavior of the total wave function occurs for N in GaP. (Note that ϕ is not just an envelope function.

[14]J. W. Allen, J. Phys. C 1, 1136 (1968).

[15]M.-F. Li, D.-Q. Mao, and S. Y. Ren, Phys. Rev. B 32, 6907 (1985).

[16]S. Brand and M. Jaros, J. Phys. C 12, 2789 (1979); see also M. Jaros and S. Brand, *ibid.* 12, 525 (1979).

[17]B. Gil, J. P. Albert, J. Camassel, H. Mathieu, and C. Benoit a la Guillaume, Phys. Rev. B 33, 2701 (1986). See also B. Gil, Physica B+C (Amsterdam) 146B, 84 (1987).

[18]The s^* orbital, by convention (Ref. 7), does not contribute to the defect potential.

[19]Ref. 20 proposes $\beta_s=0.8$ and $\beta_p=0.6$, as a phenomenological way to simulate screening and relaxation in the solid. Since we incorporate relaxation explicitly, we use $\beta_s=\beta_p=1$ here.

[20]P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983). To avoid confusion with the defect po-

tential, we use the notation T for the transfer-matrix elements where Vogl *et al.* used V.

[21]W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980).

[22]The covalent radii of N, P, As, Sb, and Bi are 0.75, 1.06, 1.20, 1.40, and 1.46 Å, respectively. The approximate proportionality of impurity bond-length change and covalent radius difference has been demonstrated by J. L. Martins and A. Zunger, Phys. Rev. B 30, 6217 (1984). For an alternative viewpoint, see S. T. Pantelides, *ibid.* 12, 2543 (1975).

[23]The exciton bound to N lies 11 meV lower than the free exciton (Ref. 2). We do not consider the electron-hole interaction here, and so take this to be the effective energy of the N one-electron level below the conduction-band minimum.

[24]The resulting value of λ is 0.1592.

[25]The resulting energy levels from the various different phenomenologies differed by almost a constant and by typically <0.1 eV.

[26]D. J. Wolford (private communication).

[27]We use $\delta p = -3B\delta d/d$, where B is the bulk modulus, to compute the corresponding bond length changes under pressure. See R.-D. Hong, D. W. Jenkins, S. Y. Ren, and J. D. Dow, Phys. Rev. B 38, 12549 (1988), and D. W. Jenkins, S. Y. Ren, and J. D. Dow, *ibid.* 39, 7881 (1989).

[28]D. J. Wolford, J. A. Bradley, K. Fry, and J. Thompson, *Proceedings of the International Conference on the Physics of Semiconductors, San Francisco, 1984*, edited by J. D. Chadi and W. A. Harrison (Springer, New York, 1985), p. 627.

[29]P. J. Dean, A. M. White, E. W. Williams, and M. G. Astles, Solid State Commun. 9, 1555 (1971).

[30]The sign is positive (negative) if the impurity atom is smaller (larger) than the host atom.

[31]D. J. Chadi and M. L. Cohen, Phys. Rev. B 8, 5747 (1973).

Paper referenced #243 in the list of publications.

DEEP LEVELS IN SUPERLATTICES

JOHN D. DOW, SHANG YUAN REN, JUN SHEN, AND MIN-HSIUNG TSAI
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556 U.S.A.

ABSTRACT

The physics of deep levels in semiconductors is reviewed, with emphasis on the fact that all substitutional impurities produce deep levels - some of which may not lie within the fundamental band gap. The character of a dopant changes when one of the deep levels moves into or out of the fundamental gap in response to a perturbation such as pressure or change of host composition. For example, Si on a Ga site in GaAs is a shallow donor, but becomes a deep trap for x>0.3 in $Al_xGa_{1-x}As$. Such shallow-deep transitions can be induced in superlattices by changing the period-widths and quantum confinement. A good rule of thumb for deep levels in superlattices is that the energy levels with respect to vacuum are relatively insensitive (on a >0.1 eV scale) to superlattice period-widths, but that the band edges of the superlattices are sensitive to changes of period. Hence the deep level positions *relative to the band edges* are sensitive to the period-widths, and shallow-deep transitions can be induced by band-gap engineering the superlattice periods.

DOPING AND DEEP LEVELS IN BULK SEMICONDUCTORS

In recent years, the theory of doping has been revised from the old effective-mass theory [1] which explained the behavior of substitutional P in Si and Se on an As site (Se_{As}) in GaAs so well. In the effective-mass model, one simply assumed that an impurity such as Se substituting for As would have a singly-ionized stable ground state Se^+ were it not for the long-ranged screened Coulomb potential $-e^2/\epsilon r$ which binds the "extra" electron of Se in a hydrogenic shallow-donor orbital with radius $a \approx \hbar^2\epsilon/m^*e^2$ much larger than a typical lattice constant. (Here m^* is the effective mass of the electron, ϵ is the dielectric constant of the host, and we have, for simplicity of presentation, overlooked the anisotropy of m^*.) Hydrogenic effective-mass theory has a significant weakness, however: it does not *predict* that a particular impurity such as Se in GaAs should be a shallow donor, but rather *assumes* that all atoms coming from Columns to the right of As (Column-V) in the Periodic Table are donors, when they substitute for As. Therefore effective-mass theory offers no criterion for determining when the extra electron of an impurity such as Se in GaAs should occupy a delocalized "shallow" hydrogenic orbital rather than a localized "deep" orbital, but rather assumes that all impurity levels are shallow.

Because there was no accepted theory of deep impurity levels in semiconductors several years ago, an unfortunate semantic problem has arisen in the literature. Shallow energy levels were defined as levels within ≈0.1 eV of a band edge and hence thermally ionizable with significant probability at room temperature. Deep levels were levels *in the fundamental band gap* by more than 0.1 eV. With the advent of the theory of deep levels [2], this energy definition has been supplanted by one that relies more on the localization of the state than its energy with respect to a nearby band edge. Shallow levels arise from the long-ranged Coulomb potential of the defect, and have extended, hydrogenic wavefunctions described by effective-mass theory. Deep levels originate from the central-cell potential of the defect. As a result, all levels that were shallow by the previous definition are shallow by the new one, except those few deep levels (by the new definition) that accidentally lie

within 0.1 eV of a band edge. All levels that were deep by the old definition are also deep by the new one, but now there are previously unanticipated deep levels that may lie in the bands or in the gap within 0.1 eV of a band edge. In this paper, we always adopt the following modern definitions: A shallow donor (acceptor) is a defect whose only energy levels in the fundamental band gap are shallow effective-mass theory levels. A deep trap is a defect that has one or more deep levels in the gap. An isoelectronic trap is a deep trap associated with an isoelectronic substitutional impurity. An isoelectronic resonant scattering center is an isoelectronic impurity whose deep levels lie energetically within the host bands and hence resonantly scatter carriers.

The character of a substitutional impurity, namely whether it is a shallow donor, a shallow acceptor, a deep trap, an isoelectronic trap, or an isoelectronic resonant scattering center is now known to be determined not solely by the position of an impurity in the Periodic Table relative to the host atom it replaces, but also by the strength of the central-cell defect potential and the energies of the "deep" levels produced by that potential relative to the band edges of the fundamental gap. Missing from the effective-mass model are the local bonds that are perturbed by the defect (and their bonding and antibonding states), and a symptom of this short-coming is the fact that the s-atomic energies of Se and As differ by ≈4 eV, indicating that the central-cell potential which models the difference must be larger yet. Clearly one cannot introduce a ≈4 eV attractive potential into a solid and have its *sole* effect be the formation of shallow donor states with binding energies of less than 10 meV; something must also happen to the electronic structure on the 4 eV scale. What happens is the formation of the hyperdeep and deep levels.

Fig. {1} [3.4] illustrates the case of N substituting for P in GaP, but (for simplicity) considers only a single directed orbital of each atom. The Ga and P atoms, when brought together into a molecule, form bonding and antibonding orbitals, with a bonding-antibonding splitting inversely proportional to the atomic energy difference $\epsilon_{Ga}-\epsilon_P$. In the solid, these molecular levels broaden into bands. When N substitutes for P, its energy ϵ_N is ≈7 eV lower than that of P; hence the energy denominator $\epsilon_{Ga}-\epsilon_N$ is ≈7 eV larger than $\epsilon_{Ga}-\epsilon_P$ and the impurity bonding-antibonding splitting is correspondingly smaller, causing the antibonding deep level to lie in the gap. The N-like impurity level is the unobserved hyperdeep level below the valence band maximum, whereas the deep level that lies in the gap has a wavefunction that is Ga-dangling-bond-like (almost independent of the N impurity). If the conduction band were somewhat broader in energy, the deep level would be covered up (as happens for N in GaAs).

N, being isoelectronic to P, simply produces a deep level in GaP that lies in the fundamental band gap, slightly below the conduction band edge. For Se in GaAs and P in Si, the bonding-antibonding physics is the same, but there is an extra electron (the "donor" electron) that will occupy this deep state if it is within the gap, or alternatively, if the deep level lies above the conduction band edge, the extra electron will fall to the conduction band edge, produce Se^{+} or P^{+}, and be bound to the ionized impurity in the hydrogenic effective-mass donor state.

For Se in GaAs (or for P in Si) the hyperdeep or isoelectronic levels have s-like and p-like symmetry, are fully occupied by electrons, and lie well below the valence band maximum and are unobserved and normally uninteresting. The antibonding or deep levels of Se in GaAs are also s-like and p-like, and lie above the conduction band minimum. These deep levels in particular are missing from the effective-mass model, and *their energies with respect to the conduction band edge determine the doping character of the impurity.*

In the effective-mass model, the anion substitutional dopants oxygen, sulfur, selenium, and tellurium should all be shallow donors in GaAs, GaP, and $GaAs_{1-x}P_x$. Yet oxygen has been observed to be a deep trap in GaP, and is thought to be deep in GaAs as well

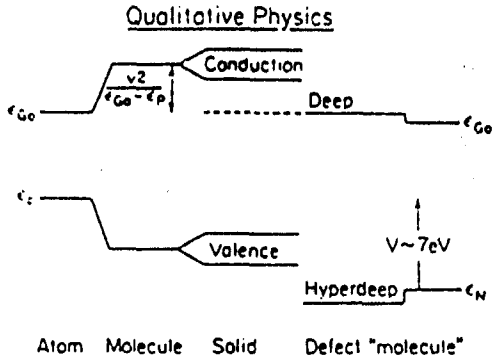


Fig. {1}. Schematic energy levels of Ga, P, molecular GaP, and solid GaP, together with the energy levels of a N anion-substitutional impurity "defect molecule" in GaP, after Refs. [3] and [4].

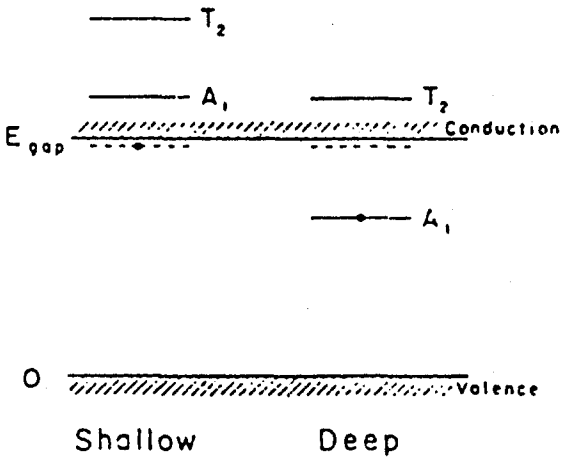


Fig. {2}. Energy level diagram illustrating the difference between shallow impurities such as S and Se in GaAs, which have their deep levels (of A_1 and T_2 symmetry for substitutional defects in zincblende hosts) outside the fundamental band gap and deep impurities such as O in GaAs or GaP, which have at least one deep level in the gap.

[5]. In contrast to oxygen, S and Se are shallow donors. The reason for the different doping character in GaAs of S and Se on the one hand and oxygen on the other is that the s-like (A_1-symmetric) antibonding deep level of oxygen lies *below* the conduction band minimum, but the corresponding deep levels of S and Se lie *above* the conduction band edge. (See Fig. {2}.) As a result, the ground state of the neutral oxygen impurity in GaAs has the deep level in the gap occupied by one electron. This level lies about 0.4 eV [5] below the conduction band edge, and so is not thermally ionizable – and does not contribute to the semiconductivity of GaAs. Moreover, the neutral oxygen impurity can trap another electron of opposite spin, and so rather than donating an electron and making GaAs n-type, it can remove one and make the host semi-insulating.

In contrast to oxygen, S and Se are shallow donor impurities that make GaAs n-type, because their A_1symmetric antibonding deep levels lie *above* the conduction band minimum. As a result, their extra electrons are autoionized, fall to the conduction band edge, and are then bound (at zero temperature) in the hydrogenic shallow donor levels produced by the ionized S^+ or Se^+. These shallow levels lie within 0.1 eV of the conduction band edge and so can be thermally ionized rather easily at room temperatures. Since the A_1 deep level lies well above the Fermi energy and the Coulombic potential of S^+ or Se^+ is screened by the donor electron, each of these impurities is incapable of binding a second electron [6]. Thus the shallow donor impurities S and Se are thermally ionized and contribute to the n-type semiconductivity of GaAs. In contrast, the deep impurity oxygen traps electrons and detracts from conductivity.

A central notion of the modern theory of doping is that *the character of a dopant is determined by the relative positions of its deep levels and the band edges (and Fermi energy).*

Fig. {3} shows how the doping character of an impurity such as Se_{As} in GaAs could be changed from a shallow donor to a deep trap to a false valence shallow acceptor, provided one could perturb the band edges of the host GaAs relative to the deep level so that first the conduction band edge and then the valence band maximum passed through the A_1-symmetric deep level. In the cases considered in Fig. {3}, the bonding s-like and p-like levels are fully occupied by electrons, lie well below the valence band maximum, and are electrically inactive and uninteresting. The anti-bonding p-like levels lie well above the conduction band minimum, and are unoccupied and also uninteresting. Thus, for Se_{As} in GaAs, the interesting deep level is the s-like or A_1-symmetric deep level and it lies above the conduction band edge. But we could imagine applying a very large pressure, which would cause the GaAs conduction band edge to pass above the Se deep level – changing the doping character of this impurity from a shallow donor to a deep-trap, and causing the conductivity to drop precipitously. One could also imagine applying a uniaxial stress so strong that the Se deep level would pass into the valence band. (The magnitude of the stress would actually exceed the breaking point of GaAs.) In this case Se_{As} would become a shallow acceptor because the hole in the deep level would bubble up to the valence band maximum or Fermi energy, leaving the stable ion of Se to be Se^-. A hole could then orbit the negative ion in a hydrogenic effective-mass state. We call this case the case of false valence. It does not occur often, because it generally requires a small band-gap host, whose vacancy has no deep level in the gap. Nevertheless when the conditions for false valence are met, an impurity such as Se_{As} from Column-VI can appear to act as if it were from Column-IV (from the viewpoint of someone familiar only with effective-mass theory). In general the false valence will differ from the true valence by two or six, depending on whether the deep level that has crossed the fundamental band gap is an s-state or a p-state.

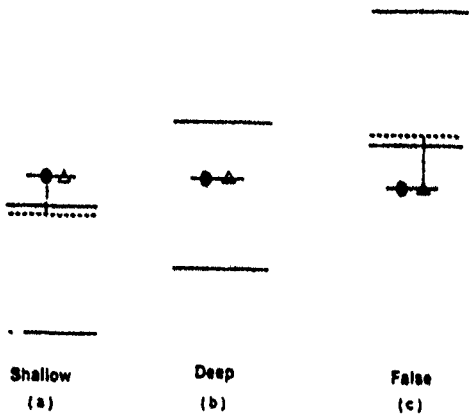


Fig. {3}. Energy level diagram illustrating how the relative positions of a deep level and the band edges could affect the character of a dopant. Band edges are long solid lines; shallow levels are dashed, and the deep level is a short solid line. The relevant deep level is assumed to be s-like and capable of holding two electrons of opposite spin. This impurity is assumed to come from one Column to the right in the Periodic Table of the atom it replaces (such as Se_{As} in GaAs). If the relevant deep level is in the gap (b), it is occupied by an electron (solid circle) and a hole (open triangle), and the neutral impurity is a deep trap for an electron or a hole. If the level is in the conduction band (a), the electron is autoionized, falls to the conduction band edge, and is trapped in the shallow donor level (dashed). If the level is in the valence band (c), the hole will bubble up into the shallow acceptor level.

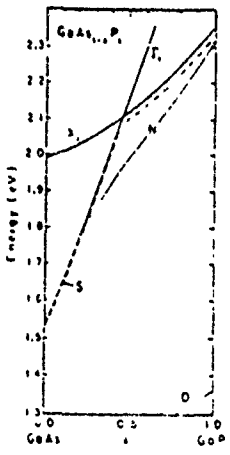


Fig. {4}. The band edges at the X-point and the Γ-point of the Brillouin zone, the N deep level, the O deep level, and the S and Se shallow levels in $GaAs_{1-x}P_x$ versus alloy composition x, after Refs. {7} and {8}.

ALTERING THE CHARACTER OF A DOPANT

Interesting physics occurs when a band edge passes through a deep level, because the character of a dopant changes. This effect was first explicitly recognized by Wolford et al. [5], who studied anion-site N, O, S, and Se in $GaAs_{1-x}P_x$. These impurity-host combinations were especially interesting because O, S, and Se are all Column-VI impurities and were naively expected to have the same doping character as shallow donors, but N is a Column-V isoelectronic defect and should not have produced any level according to the then-conventional thinking. Instead, only S and Se produced shallow donors. N produced an "isoelectronic trap" in GaP that actually lies closer to the conduction band edge than the shallow donor states. We now know that this isoelectronic trap state is in fact a deep level that by accident lies only slightly below the conduction band edge, but when it was first discovered in GaP, the isoelectronic trap concept was novel and no one considered the possibility that the trap level was in fact a deep level (with "deep" defined in a localization sense) although it was energetically shallower than the shallow levels. Oxygen in GaP produced a level energetically deep in the gap, ≈1 eV below the conduction band edge.

The key to the difference between S and Se on the one hand and oxygen on the other is provided by the data for N in the alloy $GaAs_{1-x}P_x$ [5]. In the alloy the N and O energy levels behave similarly, being roughly linear functions of alloy composition x, while the S and Se levels are rather obviously "attached" to the band edges and so vary quadratically (Fig. {4}). This fact indicated that the physics responsible for the N and O levels is the same, and, since N is isoelectronic to As and P and hence has no long-ranged defect potential, that the long-ranged or Coulombic parts of the defect potential are not responsible for the O level. This was an important fact because it showed that the O level was not a shallow level that somehow appeared energetically lower in the gap – lower due to some extra physics that is unimportant in most shallow donor problems. The data demonstrated that the O deep level originated instead from the central-cell potential of the defect. This meant that there were two separate limits to the impurity problem: (1) ordinary shallow-donor, one-band, effective-mass theory in which the long-ranged Coulomb potential of the impurity created hydrogenic states localized in $\vec{k}$-space near a band extremum and delocalized in $\vec{r}$-space, and (2) deep-level theory which relied upon the localized central-cell potential, created bonding and antibonding states, and had sp^3-character wavefunctions involving at least eight energy bands – the wavefunctions of such levels are delocalized in $\vec{k}$-space.

A direct consequence of this way of thinking about impurities was the notion that every s- and p-bonded impurity produces four localized bonding and four antibonding deep levels associated with the impurity's perturbed bonds. Typically the antibonding levels of an electronegative defect are near the band gap, while the bonding levels are in or below the valence band, are termed hyperdeep levels [2], and normally are neither electrically active nor interesting. Therefore, there must be both an s-like and a three-fold degenerate p-like deep level of S (or Se) in GaP that lies *above* the conduction band edge and varies approximately linearly with x in $GaAs_{1-x}P_x$ – as the N and O (s-like) levels do.

The data for N in $GaAs_{1-x}P_x$ indicate that a deep level, namely an antibonding (s-like, in this case) level can be either in the gap or in the band (for x<0.2). Indeed the level passes rather readily into the conduction band near x≈0.2, without a large Fano resonance effect, indicating that its wavefunction has rather different character from the shallow levels or the $GaAs_{1-x}P_x$ conduction band minimum. (Recall that GaP has an indirect band gap at the X point of the Brillouin zone, in the [001] direction, while GaAs has a direct gap at Γ or $\vec{k} = \vec{0}$.)

DEEP LEVELS IN SUPERLATTICES

Shallow-deep transitions

Since the relative positions of the substitutional impurity's deep levels and the band edges of the host determine the character of a dopant, the physics of deep levels in superlattices is particularly interesting [7-14]. The energy of a deep level with respect to vacuum is relatively insensitive (on a scale of >0.1 eV) to the structure of the superlattice, because the spatial extent of the deep level wavefunction is only of order ≈10 Å and only infrequently comes into contact with the interfaces of all but the smallest period superlattices. Therefore *the wavefunction and energies of a deep impurity in a superlattice do not change much when the period of the superlattice changes.* In contrast to the deep levels themselves, *the superlattice band edges are readily perturbed* by changes of the superlattice periods, exhibiting the well-known effects of quantum confinement: for example, reducing the width of the GaAs layers in a GaAs/$Al_xGa_{1-x}As$ superlattice causes a quantum-well effect, which results in the conduction band edge of the superlattice moving up in energy while the valence band maximum descends to lower energy. (See Fig. {5}.) Hence it is possible to "band-gap engineer" the period sizes of a superlattice so that, for example, the small-period superlattice's conduction band maximum lies *above* the deep level of a dopant – although the corresponding large-period superlattice's conduction band edge lies *below* the deep level. This means that the doping character of the impurity changes with period size [7].

To be specific, consider $N_1 \times N_2$ GaAs/$Al_{0.7}Ga_{0.3}As$ superlattices grown in the [001] direction and doped with Ga-site Si near the center of the GaAs quantum wells (Fig. {6}). Here N_1 is the number of GaAs layers in the superlattice period, and N_2 is the number of $Al_{0.7}Ga_{0.3}$ layers. For the 18×18 large-period superlattice, the Si deep level lies above the conduction band edge, resonant with the conduction band, so that the "extra" Si electron is autoionized, making Si^+ – and the Coulombic potential of Si^+ binds the electron (at low temperature) in a shallow donor level. However, for the small-period 2×34 superlattice, the conduction band minimum lies at higher energy that the Si deep level (according to theory [7,8,15]) and so the Si deep level lies in the gap, where it can trap an additional electron of opposite spin to its own electron, becoming Si^-. In the thick-period superlattice, Si_{Ga} near the center of a GaAs well is a shallow donor, making the quantum well n-type, but in the thin-period 2×34 structure, Si_{Ga} is a deep trap which tends to make the GaAs well semi-insulating. This transition from shallow donor to deep trap behavior of the Si substitutional impurity, as a function of GaAs layer thickness (Fig. {7}), is one of many example of shallow-to-deep transitions of substitutional impurities in semiconductor superlattices. The fact that the most common donor in the most widely studied superlattice exhibits this behavior should be disturbing to people who have assumed that impurities do not change their doping character in superlattices.

Indeed, one can show theoretically that in *two-dimensional semiconductor hosts* every substitutional impurity has at least one deep level in the fundamental band gap. That is, in two-dimensions all impurities are deep traps with levels in the gap. This behavior is to be contrasted with the situation for three dimensions: most impurities do not produce deep levels in the gap. Thus *with increasing quantum confinement, many shallow donors or acceptors can be expected to become deep traps and to cease providing free carriers.* In other words, thin quantum well structures will have a greater tendency to exhibit semi-insulating behavior rather than doping behavior.

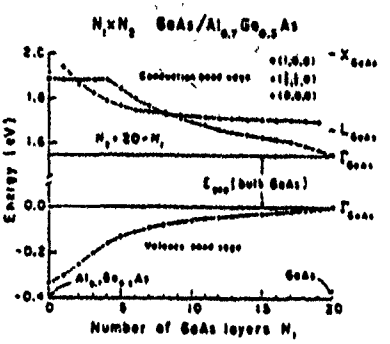


Fig. {5}. Predicted energies (in eV) of the superlattice conduction band minimum and valence band maximum with respect to the valence band maximum of bulk GaAs for a GaAs/$Al_xGa_{1-x}As$ [001] superlattice versus reduced layer thicknesses N_1 and N_2 for various $N_1 \times N_2$ [001] GaAs/$Al_xGa_{1-x}As$ superlattices, with x=0.7 and N_1+N_2 fixed to be 20, after Ref. [7]. Note the broken scale on the ordinate. The positions of the band extrema of bulk GaAs at Γ, L, and X are shown on the right of the figure, at $N_1=20$.

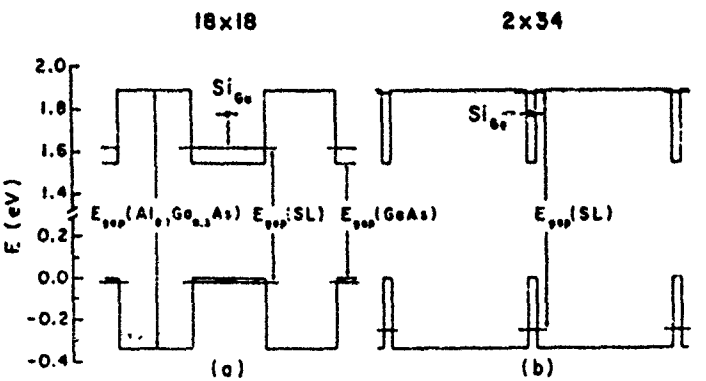


Fig. {6}. Illustrating the quantum-well effect on the band gap $E_{gap}(SL)$ of a $N_1 \times N_2$ GaAs/$Al_{0.7}Ga_{0.3}As$ superlattice, after Ref. [7]: (a) $N_1=N_2=18$; and (b) $N_1=2$, $N_2=34$. The band edges of the superlattice are denoted by chained lines. For this alloy composition the superlattice gap is indirect for case (b), with the conduction band edge at $\vec{k}=(2\pi/a_L)(1/2,1/2,0)$. Note the broken energy scale. The zero of energy is the valence band maximum of GaAs.

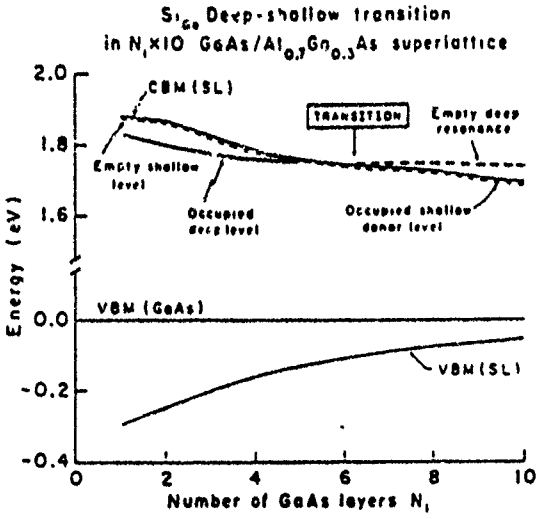


Fig. {7}. Illustrating the deep-to-shallow transition as a function of GaAs layer thickness N_1 in a GaAs/$Al_xGa_{1-x}As$ N_1×10 superlattice (SL) with x=0.7 for a Si impurity on a Column-III site in the center of a GaAs layer of the superlattice host, after Ref. [7]. The conduction band edge (CBM) and valence band maximum (VBM) are indicated by light solid lines. The Si deep level is denoted by a heavy line, which is solid when the level is in the gap but dashed when the level is resonant with the conduction band. The deep level in the band gap for $N_1<6$ is covered up by the conduction band as a result of changes in the host for $N_1>6$. The impurity's deep level lies in the gap for $N_1<6$ and is occupied by the extra Si electron; the Si, in this case, is thus a "deep impurity." For $N_1>6$, the deep level lies above the conduction band edge as a resonance. The daughter electron from the Si impurity which was destined for this deep level is autoionized, spills out of the deep resonance level, and falls to the conduction band edge (light solid line) where it is subsequently bound (at low temperature) in a shallow level associated with the long-ranged Coulomb potential of the donor (indicated by the short dashed line). It is important to realize that *both* the deep level and the shallow levels coexist and are distinct levels with qualitatively different wavefunctions. The issue of whether an *impurity* is "deep" or "shallow" is determined by whether or not a deep level associated with the impurity lies in the band gap. The computed deep-shallow transition occurs for $N_1\approx6$ layers. While the qualitative physics is completely reliable, it is possible that the transition layer thickness may differ somewhat from $N_1=6$ in real superlattices. The predicted fundamental band gap of the superlattice is indirect for $1<N_1\leq8$. All energies are with respect to the valence band maximum of GaAs.

Dependence of deep levels on location

Although shallow-deep transitions that occur when a quantum-confined band edge passes through a deep level are the most dramatic superlattice effect, the interfaces between layers in superlattices have the effect of shifting and splitting deep levels on a scale of order 0.1 eV. Fig. {8} shows the predicted [8] dependence on site of an As vacancy's deep levels in a 10×10 GaAs/$A\ell_{0.7}Ga_{0.3}As$ superlattice grown in the [111] direction. Note that the valence band edges of GaAs, $A\ell_{0.7}Ga_{0.3}As$, and the superlattice are aligned according to the measured band offset [16], and that once this is done, the deep levels are relatively independent of site, except for a variation on the scale of 0.1 eV. The bulk p-like valence band maximum is split into an upper π-like and lower σ-like band edge due to the reduced symmetry of the superlattice. This splitting is reflected in different energies for π-like and σ-like deep levels, even if the parent impurity is distant from an interface. The T_2 or p-like deep levels split the most at an interface into one or two levels corresponding to hybrid orbitals directed primarily into GaAs (which orbitals have energies near the bulk GaAs vacancy energy) and hybrids directed primarily into $A\ell_{0.7}Ga_{0.3}As$ with corresponding energies [7,8]. In addition to this effect, the band offset between the two materials is reflected in both the p-like deep level and the s-like level.

The s-like deep levels of cation-site Si in a 3×10 GaAs/$A\ell_{0.7}Ga_{0.3}As$ superlattice are given in Fig. {9}, and show (i) how undramatic the effects of the superlattice are on the level itself and (ii) how an impurity can produce a shallow donor and n-type doping in bulk GaAs but yield a deep trap and semi-insulating behavior in the superlattice. Si (or a defect involving Si) is thought to be the DX center in $A\ell_xGa_{1-x}As$ [15].

These shifts and splittings of the deep levels as a function of position in the superlattice will manifest themselves experimentally as inhomogeneously broadened deep level energies.

Formalism

The formalism employed here is rather simple, and has been discussed in Refs. [7] and [8] in detail. Basically, one solves the secular equation

$$\det \{ 1 - GV \} = 0,$$

where $G=(E-H)^{-1}$ is the Green's function operator associated with the host Hamiltonian H and E is the deep level energy. The defect potential operator is V, and we have obtained solutions to this equation using the tight-binding model of Vogl et al. [17]. The basic approach is the same as that used by Hjalmarson et al. [2] for substitutional impurities in bulk semiconductors.

SUMMARY

Deep impurity levels in semiconductor superlattices will, on an absolute energy scale, be shifted only slightly and split (or inhomogeneously broadened) from their bulk semiconductor energies. This relative insensitivity of the levels to superlattice geometry, combined with the superlattice's property of being sensitive to quantum confinement effects on its band edges should lead to shallow-deep transitions and novel doping properties of impurities in thin quantum wells. The inhomogeneous broadening of deep levels will be a less dramatic effect on the electronic structure, but will nonetheless have important consequences in areas such as exciton transport and resonant energy transfer of excitons bound to deep impurities.

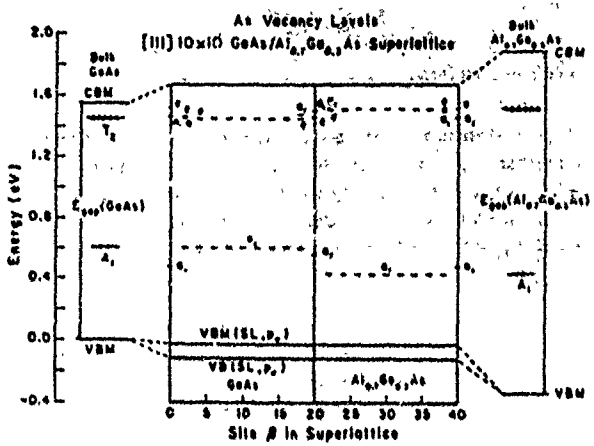


Fig. {8}. Predicted energy levels of an As-vacancy in a $(GaAs)_{10}(A\ell_{0.7}Ga_{0.3}As)_{10}$ [111] superlattice, as a function of β, the position of the vacancy (even values of β correspond to As-sites), after Ref. [8]. Note the splitting of the T_2 levels at and near the interfaces (β = 0, 20, and 40), and that the T_2-derived vacancy levels lie at higher energy in an $A\ell_{0.7}Ga_{0.3}As$ layer than in a GaAs layer. The a_1 and e ordering changes at successive interfaces. The zero of energy is the valence band maximum of bulk GaAs, and the corresponding valence band (VBM) and conduction band (CBM) edges and deep levels in bulk GaAs and bulk $A\ell_{0.7}Ga_{0.3}As$ are given to the left and right of the central figure, respectively. The top of the central figure is the conduction band edge of the superlattice, and the bottom corresponds to the split valence band in the superlattice – the valence band maximum of the superlattice being of e symmetry (p_π) and the split-off a_1 (p_σ) band maximum lying 0.087 eV lower in energy. The A_1 level in the $A\ell_xGa_{1-x}As$ layer of the superlattice is lower than the corresponding level in the GaAs layer because of the band offset of 0.334 eV. The electron (hole) occupancies of the deep levels in bulk GaAs and bulk $A\ell_{0.7}Ga_{0.3}As$ are denoted by solid circles (open triangles).

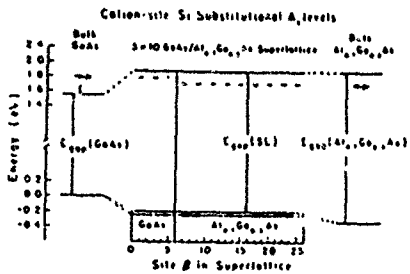


Fig. {9}. Predicted A_1-derived deep levels of Si in GaAs, in a 3×10 $GaAs/A\ell_{0.7}Ga_{0.3}As$ superlattice, after J. D. Dow, S. Y. Ren, and J. Shen, Ref. [10] as a function of the position β of the Si in the superlattice, and in the bulk materials. Interfaces correspond to sites β= 0, 6, and 26. Note that in bulk GaAs Si is a shallow donor, but that it is a deep trap in this superlattice, according to the theory.

ACKNOWLEDGMENT

We are grateful to the U. S. Office of Naval Research and the Air Force Office of Scientific Research (Contract Nos. N00014-89-J-1136 and AF-AFOSR-89-0063) for their generous support.

REFERENCES

[1] W. Kohn, in *Solid State Physics*, edited by F. Seitz and D. Turnbull, (Academic Press, New York, 1957), Vol. 5, P. 258-321.

[2] H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Letters 44, 810 (1980); H. P. Hjalmarson, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1979, unpublished. See also W. Y. Hsu, J. D. Dow, D. J. Wolford, and B. G. Streetman, Phys. Rev. B 16, 1597 (1977) for the concepts that form the foundation of this work.

[3] J. D. Dow, Mater. Res. Soc. Symp. Proc. 46, 71 (1985).

[4] J. D. Dow, in *Highlights of Condensed Matter Theory* (Proc. Intl. School of Phys. "Enrico Fermi" Course 89, Varenna, 1983) ed. by F. Bassani, F. Fumi, and M. P. Tosi (Societa Italiana di Fisica, Bologna, Italy, and North Holland, Amsterdam, 1985), pp. 465 et seq.

[5] D. J. Wolford, W. Y. Hsu, J. D. Dow, and B. G. Streetman, J. Lumin. 18/19, 863 (1979).

[6] The hydrogenic potential, strictly speaking, can bind a second electron, since the ion H^- is stable. However, this second-electron binding energy is so small compared with typical thermal energies that we neglect it.

[7] S. Y. Ren, J. D. Dow, and J. Shen, Phys. Rev. B 38, 10677 (1988).

[8] S. Y. Ren and J. D. Dow, J. Appl. Phys. 65, 1987 (1989).

[9] R.-D. Hong, D. W. Jenkins, S. Y. Ren, and J. D. Dow, Proc. Materials Research Soc. 77, 545 (1987), *Interfaces, Superlattices, and Thin Films*, ed. J. D. Dow and I. K. Schuller.

[10] For a review of the physics of impurities in superlattices, see *Properties of impurity states in superlattice semiconductors*, ed. by C. Y. Fong, I. P. Batra, and S. Ciraci (Plenum Publishing Corporation, 1988). See also Refs. [7], [8], [11], [12], [13], and [14].

[11] S. Das Sarma and A. Madhukar, J. Vac. Sci. Technol. 19, 447 (1981).

[12] D. K. Maude, J. C. Portal, L. Dmowski, T. Foster, L. Eaves, M. Nathan, M. Heiblum, J. J. Harris, and R. B. Beall, Phys. Rev. Letters 59, 815 (1987).

[13] J. S. Nelson, C. Y. Fong, I. P. Batra, W. E. Pickett, and B. M. Klein, unpublished

[14] M. Lannoo and J. Bourgoin, to be published.

[15] H. P. Hjalmarson and T. J. Drummond, Appl. Phys. Letters 48, 657 (1986). A preliminary account was given by H. P. Hjalmarson, Bull. Amer. Phys. Soc. 32, 814 (1987).

[16] D. J. Wolford, T. F. Kuech, J. A. Bradley, M. A. Grell, D. Ninno, and M. Jaros, J. Vac. Sci. Technol. B4, 1043 (1986).

[17] P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983).

Change of the Frontier Electronic Orbitals due to Substitutional Impurities in Large Chemical or Biological Molecules

SHANG YUAN REN AND JOHN D. DOW

Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287, USA

Abstract

A molecule whose electronic structures are poorly understood, in some cases, can be thought of as another perhaps well-understood "host chemical molecule" with substitutional "impurities." From this point of view, the Green's function technique in solid state physics can be extended to investigate the *change* due to substitutional impurities in large chemical or biological molecules of the frontier electronic orbitals—the highest occupied molecular orbitals (HOMO) and the lowest unoccupied molecular orbitals (LUMO). These orbitals are the ones most active in chemical reactions, and are conceptually very similar to the valence band and conduction band states of semiconductors. Substitutional impurities will often introduce new "deep" electronic states into the gap between the HOMO and LUMO states of the original molecule and these states will have the following properties. (*i*) The likelihood that an impurity will produce a state in the gap depends on the site of the impurity. (*ii*) For impurities on a particular site, the wavefunction of the gap state will be relatively independent of the impurity. (*iii*) Only a small fraction of the deep level wavefunction (typically 10%) will lie within the impurity's cell and the largest part (about 40–50%) will be on the impurity's nearest-neighbor atoms. (*iv*) The derivative of the energy of the impurity state with respect to the central-cell impurity potential dE/dV is approximately equal to the probability of the gap-state electron being found on the impurity's site.

The frontier orbitals of chemical or biological molecules, namely the highest occupied molecular orbitals (HOMO) and the lowest unoccupied molecular orbitals (LUMO), are the most active electronic orbitals in chemical reactions and biological processes [1]. A theory of how these orbitals change when the chemical composition of the macromolecule is altered would be particularly useful. For reasonably small molecules, with fewer than 100 atoms, the standard quantum-chemistry calculations can provide the needed information about electronic structure. But for very large macromolecules, such as DNA or proteins, such calculations are currently impractical, and so one must find an alternative scheme for predicting how the frontier orbitals of large molecules, their energy levels, and their electronic properties change with chemical composition. Such a scheme is suggested by the fact that many large chemical or biological molecules are electrically semiconductive [2], with HOMO and LUMO states that are conceptually similar to the valence bands and conduction bands in semiconductors. Therefore replacement of one atom in a macromolecule is analogous to introduction of a substitutional impurity into a semiconductor, and so the qualitative concepts of the theory of deep impurity levels in semiconductors

International Journal of Quantum Chemistry, Quantum Biology Symposium 17, 73–80 (1990)
© 1990 John Wiley & Sons, Inc.
CCC 0360-8832/90/010073-08$04.00

Diazabicyclooctane can be considered as 2.2.2-bicyclooctane with "two substitutional impurity nitrogen atoms" at carbon sites 1 and 4 (with two H atoms, each of them bonded to each of the two C atoms, removed).

The Green's function technique, which so successfully described the wavefunctions [6] and the electronic structure of deep impurity levels in semiconductors can be readily extended to investigate the change of the electronic structure of the frontier orbitals due to one or more substitutional impurity atoms in a large chemical or biological molecule. That is to say, for very large chemical or biological molecules, although it may be very difficult to calculate the complete electronic structures of the frontier orbitals themselves, the changes in electronic structure due to an "impurity" can be predicted rather easily. Moreover, in some molecules so large that calculations of even the electronic structure of a single "perfect host" is impractical, qualitative predictions are nevertheless still feasible, based on the general structure of the theory of deep levels. For example, many extremely large biological molecules have several P atoms. The theory of deep levels (without any calculation) would predict that molecules with one P atom replaced by an As atom should have similar electronic structures to the "host," but that N substitution might yield a deep level in the HOMO–LUMO gap.

A substitutional impurity produces a defect potential [9], the short-ranged part of which is normally associated with the "deep" impurity levels in the gap. The states associated with s- and p-orbitals on each site will be especially perturbed by this potential (but the number of states near the HOMO–LUMO gap will not be altered) and one or more of the states lying in the bands of the "host" may move into the gap as a result of the perturbation. The gap states or "deep levels" interest us in this paper. In particular, we are especially interested in determining whether a particular "impurity" produces a "deep level" in the HOMO–LUMO gap, and if it does, what would be the properties of such a level—a problem similar to the deep impurity problem in semiconductors.

The approach we use to investigate the wavefunctions of deep impurity levels is a semiempirical Green's function approach [3–6] with a nearest-neighbor tight-binding Hamiltonian describing the electronic band structures of the "host" crystalline semiconductor or macromolecule. In the molecular case the major difference is that chemical or biological molecules generally do not have periodic symmetry [10,11]. Here, for illustrative purposes, we treat a rather small molecule, 2.2.2-bicyclooctane, as our "host." It has 8 carbon atoms and 14 hydrogen atoms. We use a nearest-neighbor tight-binding Hamiltonian to describe the electronic structure of the "host molecule," 2.2.2-bicyclooctane. We include one s orbital and three p orbitals for each carbon atom and one s orbital for each hydrogen atom. The resulting tight-binding Hamiltonian is an $8 \times 4 + 14 = 46$-dimensional matrix. To obtain the tight-binding parameters, we fit the *ab initio* SCF–LCAO–MO calculation of Lehn and Wipff [12], obtaining the results of Figure 2.

This Hamiltonian has 46 energy eigenvalues, each of which can contain two electrons of opposite spin. The 46 valence electrons of this "host molecule" will occupy the 23 lower energy levels, and the 23 higher energy levels are unoccupied at zero temperature. We find a 4.35 eV energy gap between the highest occupied

[3–6] should have useful analogues in macromolecular physics. In particular, it should be possible to predict which "impurities" or chemical changes are likely to introduce a deep level into the HOMO–LUMO gap.

In crystalline semiconductors such as Si or GaAs, an impurity produces "deep" levels associated with the parts of the "central-cell" impurity potential at and near the impurity site [5]. Normally four such levels should appear in the energetic vicinity of the fundamental band gap, if the impurity is s- and p-bonded: one s-like level and three p-like levels associated with the perturbed impurity bonds. The theory of deep impurities is concerned with predicting whether one or more of the deep levels lies within the fundamental band gap, in which case the level often alters the electron dynamics, by trapping carriers, by serving as a center for non-radiative recombination of electrons and holes, or by providing an initiation site for a chemical reaction. Similar phenomena should occur in macromolecules.

Two cases amenable to treatment by the theory of deep impurities in macromolecules are:

(*i*) Several macromolecules with the same crystal structure are so large that it is practical to compute the complete electronic structure of only one prototype, in which case the other molecules are considered as the prototype "host" with "impurities."

(*ii*) The macromolecules are so large that it is impractical to compute the electronic structure of even a prototype.

In this latter case one can often make a crude approximation to the Green's function of the "host" and then estimate how the frontier orbitals of the host are affected by "impurities." For example, if we take the rather small 2,2,2-bicyclooctane [7] molecule as the "host molecule" (see Fig. 1), then quinuclidine can be considered as 2,2,2-bicyclooctane with "a substitutional impurity nitrogen atom" at carbon site 1 (with the H atom, which is bonded to the C atom, also taken out [8]).

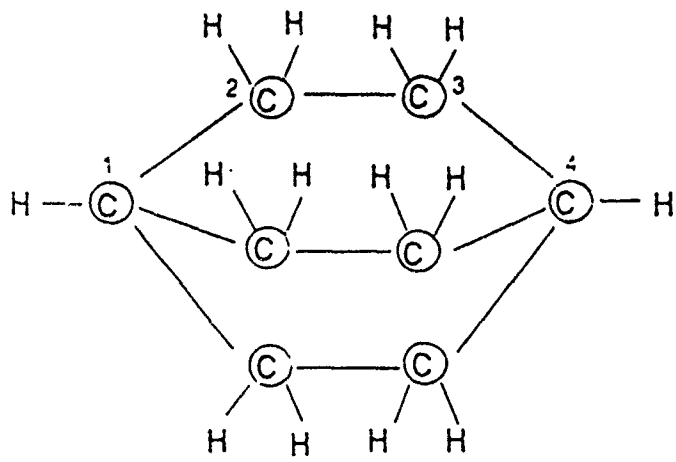


Figure 1. The geometrical structure of the 2,2,2-bicyclooctane molecule after Reference [7], indicating the site labeling

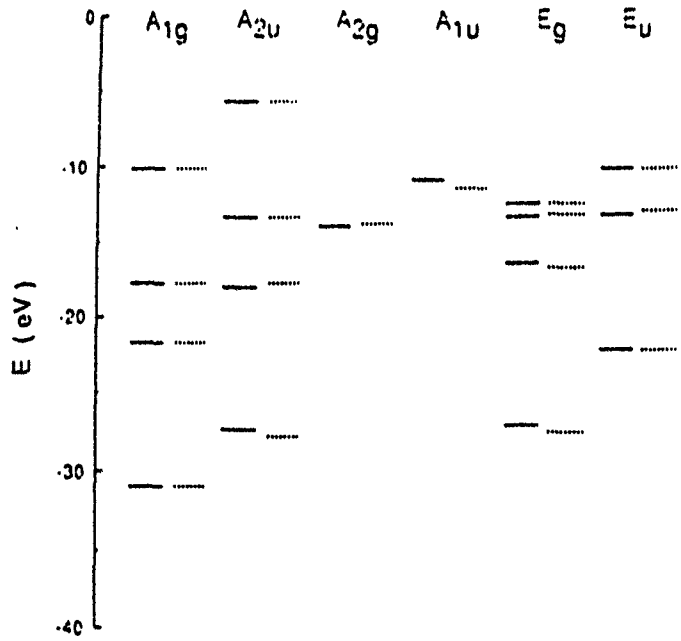


Figure 2. Comparison of the valence electronic structure of 2,2,2-bicyclooctane obtained with *ab initio* SCF theory [12]—and the tight-binding parametrization of that theory (- - -). The LUMO state is the highest energy A_{2u} state and the HOMO state is the highest E_u state in this figure.

energy level—the 23rd eigenvalue—and the lowest unoccupied energy level—the 24th eigenvalue. This gap is analogous to the band gap of semiconductors, and the lower 23 energy levels are very similar to the valence bands of semiconductors, while the higher 23 unoccupied energy levels correspond to the conduction bands.

If a substitutional impurity atom is introduced into this host molecule, the difference in Hamiltonians will be the "impurity potential" of the system [4–6]. Here we consider only the effect of the short-ranged potential, in analogy with the case of crystalline semiconductors, for which the short-ranged potential rather than the long-ranged Coulomb potential is dominant in producing states in the gap more than 0.1 eV from a band edge. We model the defect potential in a tight-binding basis of s and p orbitals centered on each atomic site. Following the theory of deep impurity levels in semiconductors [4–6], we take the defect potential to be short-ranged in such a basis (assuming either that the basis states on different sites do not overlap, or that the basis states are symmetrically orthogonalized): The effective short-ranged potential may be expressed as an "on-site" potential on the "impurity site":

$$\begin{aligned} \langle ia|V|jb\rangle &= V_s \quad \text{if} \quad i = j = 0, \quad \text{and} \quad a = b = s, \\ &= V_p \quad \text{if} \quad i = j = 0, \quad \text{and} \quad a = b = x \text{ or } y \text{ or } z, \\ &= 0 \text{ otherwise} \end{aligned} \tag{1}$$

Here i and j denote atomic sites in the molecule, and a and b are symbols of orbitals s, p_x, p_y, and p_z. With this approximation, the energy level associated with the impurity in the host molecule will be determined by the following secular equation [4,5]

$$\det|1 - GV| = 0 \tag{2}$$

Here G is the Green's function of the host molecule:

$$\langle ia|G|jb\rangle = \sum_n \langle ia|n\rangle\langle n|jb\rangle/(E - E_n) \tag{3}$$

and we are interested in the levels E in the gap. Here n represents one of the solutions of the tight-binding Hamiltonian of the perfect host molecule. For many cases, especially when the impurity has an attractive short-ranged potential, V_s plays a more important role, and the effect of V_p can be neglected. In this case, the energy level of the gap state is determined by,

$$\langle 0s|G|0s\rangle V_s = 1 \tag{4}$$

The associated wavefunction is [6]

$$|\psi\rangle = \sum_{ia} C_{ia}|ia\rangle \tag{5}$$

where we have

$$C_{0s} = -\langle 0s|G|0s\rangle\langle 0s|G|0s\rangle/(d\langle 0s|G|0s\rangle/dE) \tag{6}$$

and

$$C_{ia} = C_{0s}\langle ia|G|0s\rangle/\langle 0s|G|0s\rangle\,. \tag{7}$$

We take 2,2,2-bicyclooctane as an illustrative example of a "host" although it is not a very large molecule. It has eight carbon atoms, but only two of the carbon sites are inequivalent, namely those at site 1 and site 2 (See Fig. 1). In Figure 3 we show the energy levels of the gap state for site 1 and site 2 as functions of the impurity potential V_s [see Eq. (4)].

For an impurity at site 1, an impurity potential of about −0.12 eV is enough to produce an impurity state in the HOMO–LUMO gap, while for an impurity at site 2, a much stronger impurity potential, −2.0 eV is needed to produce an impurity state in the gap. The following atoms, substituting for C at either site, should produce a deep level in the gap: the rare gases, the halogens, and N, O, S, and Se. In addition At is the only s- and p-bonded impurity with an impurity potential between −0.12 and −2.0 eV when it substitutes for C [8], and so At produces a "deep level" in the HOMO–LUMO gap when it occupies site 1, but not when it occupies site 2 [13].

In Figure 4 we show that the on-site coefficient C_{0s} for the wavefunction at site 1 depends very little on V_s, as determined by Eq. (6). This is an example of deep-level "pinning" [5]. It means that the deep level in the HOMO–LUMO gap is associated with a wavefunction and an energy that are almost independent of the impurity: different impurities produce almost the same wavefunctions. This happens because

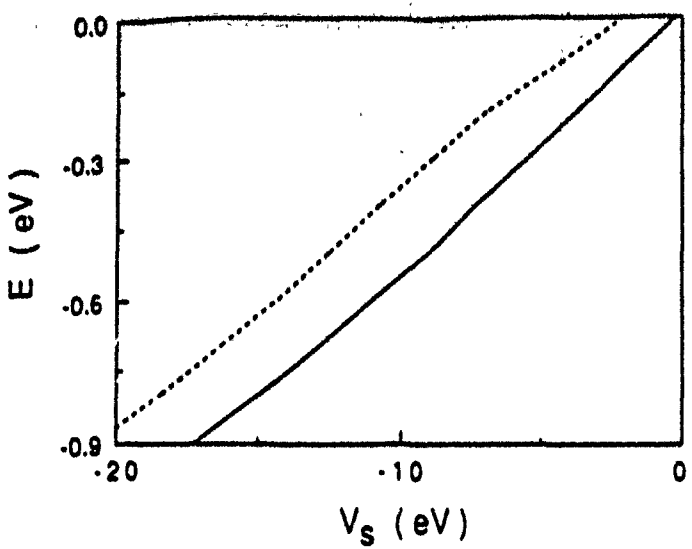


Figure 3. The energy level of the gap state for the site 1 (—) and site 2 (- - -) as functions of the impurity short-range potential V_s, determined by Eq. (3). We chose the lowest energy level of the LUMO states of the host molecule as the energy zero.

the deep level in the gap associated with a strongly attractive defect is an antibonding state and has a predominantly host-like character.

The charge density associated with the deep level has typically 40–50% of its charge on the atomic sites of the nearest-neighbors to the impurity, and only about

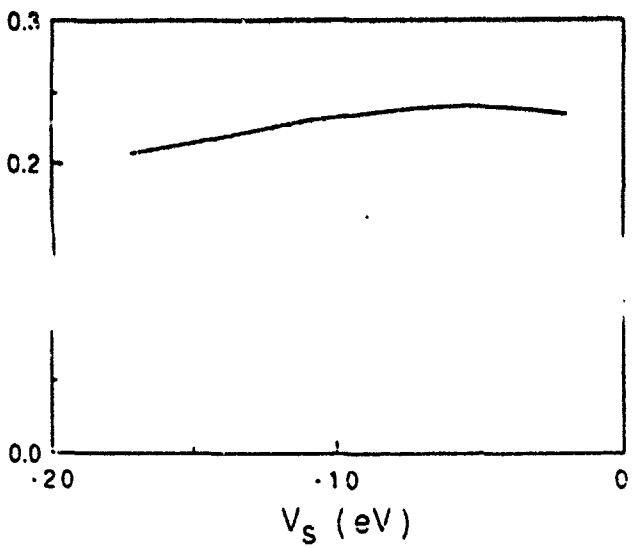


Figure 4. The on-site wavefunction coefficient C_{s1} of the gap state for the site 1 as a function of the impurity short-range potential V_s, determined by Eq. (6)

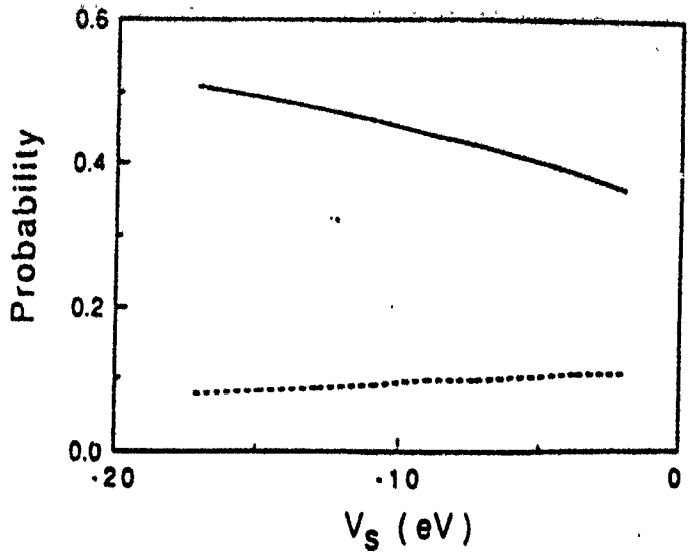


Figure 5. The on-site probability (- - -) $\Sigma_i |C_{0i}|^2$, where $i = s, p_x, p_y$, and p_z, and the corresponding nearest-neighbor probability (—) of the gap state for the site 1 as a function of the impurity short-range potential V_s, determined by Eqs. (6) and (7).

10% on the impurity's site (see Fig. 5). The 10% on the impurity site for this molecule is similar to the corresponding charge densities for deep levels in elementary crystalline Si. Hence, the basic physics of deep levels in semiconductors carries over to macromolecules: the charge density in the impurity cell is rather smaller than an uninformed person might have guessed, and is related to the derivative.

$$dE/dV_s = (C_{0i})^2 \tag{8}$$

by the Hellmann–Feynman theorem [13].

In summary, the calculation for the rather moderate-sized 2,2,2-bicyclooctane molecule reveals that deep impurity levels can be produced in the HOMO–LUMO gap with similar properties to those found for deep impurity levels in crystalline semiconductors. Since this physics holds for both the rather small 2,2,2-bicyclooctane molecule and for very large semiconductor crystals, it appears to be very general, and hence can be exploited for elucidating the electronic structures of wide classes of macromolecules that correspond to a "host" plus one or more substitutional "impurities."

Acknowledgment

We are grateful to the U.S. Office of Naval Research and the Air Force Office of Scientific Research (Contract Nos. N00014-89-J-1136 and AFOSR-89-0063) for their generous support.

Bibliography

[1] For example, see K. Fukui, *Theory of Orientation and Stereoselection* (Springer-Verlag, Berlin, 1975).

[2] For example, see A. Szent Gyorgyi, *Introduction to a Submolecular Biology* (Academic, New York, 1960).

[3] G. F. Koster and J. C. Slater, Phys. Rev. 94, 1392 (1954); G. F. Koster and J. C. Slater, Phys. Rev. 95, 1167 (1954).

[4] For a review, see J. D. Dow, in *Highlights of Condensed Matter Theory (Proc. Intl. School of Phys. "Enrico Fermi" Course 89, Varenna, 1983)*, F. Bassani, F. Fumi, and M. P. Tosi, Eds. (Societa Italiana di Fisica, Bologna, Italy, and North Holland, Amsterdam, 1985), pp. 465 et seq.

[5] H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. 44, 810 (1980); see also W. Y. Hsu, J. D. Dow, D. J. Wolford, and B. G. Streetman, Phys. Rev. B16, 1597 (1977) for the concepts that form the foundation of this work.

[6] S. Y. Ren, W. M. Hu, O. F. Sankey, and J. D. Dow, Phys. Rev. B 26, 951 (1982); S. Y. Ren, Scientia Sinica XXVII, 443 (1984); S. Y. Ren, "Electronic structure of deep impurities in semiconductors," in *Lattice Dynamics and Semiconductor Physics*, Festschrift for Professor Kun Huang, (World Scientific Publishing Co., Singapore, 1990), pp. 536–547.

[7] 2,2,2-Bicyclooctane is a chemical molecule with 8 carbon atoms and 14 hydrogen atoms. The two carbon atoms, each of which is bonded to three carbon atoms and one hydrogen atom, are labeled as site 1 and site 4. The other six carbon atom sites, each of which is bonded to two carbon atoms and two hydrogen atoms, are labeled as 2, 3, 5, 6, 7, and 8. See Figure 1 and, e.g., *Handbook of Chemistry and Physics*, 57th ed. (CRC Press, Boca Raton, FL, 1976–1977), pp. c-208, c-362, and c-628.

[8] For simplicity of presentation, we execute the calculations assuming that the C atom is replaced, but making the approximation that the H atoms bonded to it are not removed.

[9] The short-ranged impurity potential V_s can be approximated as $V_s = \beta_s(E_{si} - E_{sh})$, where E_{si} is the s state atomic energy of the impurity atom and E_{sh} is the s state atom energy of the replaced host atom, which may be found, e.g., from P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983). β_s is a constant, 0.8. See also Reference [5].

[10] P.-O. Löwdin, J. Mol. Spectros. 14, 119 (1964). Some of the formulas in this early paper are very similar to those in the present paper. The authors thank Prof. Löwdin for calling their attention to this paper.

[11] M. A. Ali and R. F. Wood, J. Mol. Spectros. 23, 53 (1967) uses the formalism of Reference [10] to study a few heteroatomic conjugated systems.

[12] J. M. Lehn and G. Wipff, Theoret. Chim. Acta (Berlin) 33, 43 (1974).

[13] H. Helimann, *Einfuehrung in Quanten Chemie* (Franz Deutsch, Leipzig, 1937); R. P. Feynman, Phys. Rev. 56, 340 (1939).

Received June 29, 1990

Generalized Möbius transforms for inverse problems

Shang Yuan Ren [1] and John D. Dow

Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287, USA [2]
and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

Received 19 July 1990; revised manuscript received 4 February 1991; accepted for publication 7 February 1991
Communicated by J.P. Vigier

An inverse formula is given for a new, generalized Möbius transform $G(x, y, \ldots) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \ldots F(n^{\alpha}x, m^{\beta}y, \ldots)$, namely $F(x, y, \ldots) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \ldots \mu(n)\mu(m) G(n^{\alpha}x, m^{\beta}y, \ldots)$. In the case of one dimension and $\alpha = -1$, this reduces to the Chen–Möbius transform which has been applied to the inverse black-body radiation problem. For $\alpha = +1$ this becomes the Hardy–Wright–Möbius transform.

Recently Chen [1] modified Möbius' number-theory transform [2] and showed how it could be used to solve a class of inverse problems in physics, including the problem of determining the temperature distribution of a black body from its radiation spectrum: given a quantity $F(x)$ that is related to a function $G(x)$ by

$$G(x) = \sum_{n=1}^{\infty} F(x/n) , \tag{1a}$$

the inverse function F is given by

$$F(x) = \sum_{n=1}^{\infty} \mu(n)\, G(x/n) , \tag{1b}$$

where $\mu(n)$ is the Möbius function: $\mu(n) = 1$ if $n = 1$, $\mu(n) = (-1)^r$ if n is the product of r distinct prime factors, and $\mu(n) = 0$ otherwise [2]. For example, in the physics of black bodies, $G(\nu)$ may be related to the radiation spectrum and $F(T)$ may be related to the areal thermal distribution of the body [1]. As shown by Chen, this theorem is particularly useful for inverting data, for example, obtaining the unknown areal distribution of temperatures on a black body from its radiation spectrum.

Here (i) we combine Chen's result with another inverse Möbius transform from number theory [2] [#1] to obtain a more general inverse formula, and (ii) we extend this generalized formula to two and higher dimensions. The new and more general inverse Möbius transform formula for one dimension is:

If we have

$$G(x) = \sum_{n=1}^{\infty} F(n^{\alpha}x) \tag{2a}$$

then it is the case that

$$F(x) = \sum_{n=1}^{\infty} \mu(n) G(n^{\alpha}x) . \tag{2b}$$

Hardy and Wright proved this theorem for $\alpha = 1$ ("theorem 270" [2] [#1]) and Chen's Möbius theorem corresponds to $\alpha = -1$. We show that the theorem is valid for any real α, and hence there are entire classes of Möbius transforms, each corresponding to a different choice of α. Our proof uses "theorem 263" [2] for the Möbius function,

$$\sum_{n|k} \mu(n) = 1, \quad \text{for } k = 1,$$
$$= 0, \quad \text{otherwise.} \tag{3}$$

Here the sum is over all the different divisors n of the integer k, including 1 and k. The proof of the gen-

[1] On leave from the University of Science and Technology of China, Hefei 230026, China.
[2] Present address.

[#1] In fact, due to eq. (31) of ref. [1], eq. (32) in ref. [1] is a direct result of theorem 270 of ref. [2].

eralized formula is obtained by evaluating

$$\sum_{n=1}^{\infty} \mu(n) G(n^{\alpha} x) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mu(n) F(m^{\alpha} n^{\alpha} x)$$

$$= \sum_{k=1}^{\infty} F(k^{\alpha} x) \sum_{n|k} \mu(n) = F(x) . \qquad (4)$$

In this derivation, we performed the sum over $k=mn$ and n, instead of over m and n; this reordering will converge if

$$\sum_{m=1} \sum_{n=1} |F(m^{\alpha} n^{\alpha} x)| = \sum_{k} d(k) |F(k^{\alpha} x)| \qquad (5)$$

converges, where $d(k)$ is the number of divisors of k [2]. Note that for $\alpha=0$ this convergence criterion is normally *not* met, and hence $\alpha=0$ is to be avoided. Furthermore the converse of eq. (2) is true if

$$\sum_{n=1} \sum_{m=1} |G(m^{\alpha} n^{\alpha} x)| = \sum_{k} d(k) \, |G(k^{\alpha} x)| \qquad (6)$$

converges.

Generalization of formula (2) to higher dimensions is straightforward; for example, in two dimensions for $F(x, y)$ of the form

$$G(x, y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} F(m^{\alpha} x, n^{\beta} y) , \qquad (7a)$$

we have

$$F(x, y) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \mu(k) \mu(l) \, G(k^{\alpha} x, l^{\beta} y) . \qquad (7b)$$

A sufficient condition for this inverse formula to be valid is the convergence of

$$\sum_{k} \sum_{m} \sum_{l} \sum_{n} |F(m^{\alpha} k^{\alpha} x, n^{\beta} l^{\beta} y)|$$

$$= \sum_{i} \sum_{j} d(i) d(j) |F(i^{\alpha} x, j^{\beta} y)| . \qquad (8)$$

The new inverse Möbius formula (2) can be applied to problems in mathematics and physics, and a convenient choice of the indices α, β, ... can be selected to best suit each problem.

The one-dimensional version leads to the new mathematical identity

$$\frac{1}{\zeta(p)} = \sum_{n=1}^{\infty} \mu(n) n^{-p} \quad \text{for } p>1 , \qquad (9)$$

where $\zeta(p) = \sum_{n=1}^{\infty} n^{-p}$ is the Riemann zeta function. It also leads to the identities

$$\frac{x^2}{x^2 \pm 1} = \sum_{n=1}^{\infty} \tfrac{1}{2} \mu(n) [(\pi x/n) f_{\pm}(\pi x/n) - 1] \qquad (10)$$

and

$$J_0(x) = \sum_{n=1}^{\infty} \mu(n) \Big(-\frac{1}{2} + \frac{1}{nx}$$

$$+ 2 \sum_{m=1}^{N} (n^2 x^2 - 4m^2 \pi^2)^{-1/2} \Big) , \qquad (11)$$

where we have $2N\pi < nx < (2N+2)\pi$, J_0 is the cylindrical Bessel function, $f_+(x) = \coth(x)$, and $f_-(x) = \cot(x)$, and we have used the identities [3]

$$f_{\pm}(\pi x) = \frac{1}{\pi x} + \frac{2x}{\pi} \sum_{n=1}^{\infty} (x^2 \pm n^2)^{-1} ,$$

and

$$\sum_{n=1}^{\infty} J_0(nx) = -\frac{1}{2} + \frac{1}{x} + 2 \sum_{m=1}^{N} (x^2 - 4m^2 \pi^2)^{-1/2}$$

for $2N\pi < x < (2N+2)\pi$.

There are undoubtedly many more uses for this inverse formalism in both mathematics and theoretical physics that will be uncovered as it is used more.

We are grateful to the U.S. Office of Naval Research and the Air Force Office of Scientific Research (Contract Nos. N00014-89-J-1136 and AFOSR-89-0063) for their generous support.

References

[1] N.X. Chen, Phys. Rev. Lett. 64 (1990) 1193.
[2] G.H. Hardy and E.M. Wright, An introduction to the theory of numbers, 5th Ed. (Oxford Univ. Press, Oxford, 1979) pp. 235–237.
[3] I.S. Gradshteyn and I.M. Ryzhik, Table of integrals, series and products (Academic Press, New York, 1980) pp. 46, 976.

Convergence of force calculations for noncrystalline Si

D. A. Drabold and J. D. Dow
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

P. A. Fedders and A. E. Carlsson
Department of Physics, Washington University, Saint Louis, Missouri 63130

Otto F. Sankey
Department of Physics, Arizona State University, Tempe, Arizona 85287

(Received 16 January 1990; revised manuscript received 1 May 1990)

We compare the forces generated by various angle-dependent potentials and by *ab initio* band-structure calculations with a limited number of k points. In cells with 32 and 54 atoms we find substantial errors for all of the angle-dependent-force models, as well as for the band-structure forces with very few k points.

Recently there have been a number of computer simulations of amorphous solid Si (*a*-Si) and of liquid Si (*l*-Si) using two distinct methods: (i) empirically fitted angle-dependent forces (ADF) (Refs. 1–8), and (ii) first-principles band-structure-derived forces (BSF).[9,10] Finite Si clusters and reconstructed Si surfaces have also been studied. Similar studies have been performed or are being performed for other materials.[11] Both of these methods, ADF and BSF, have inaccuracies that have not been systematically evaluated for liquid and amorphous geometries. The ADF models are fit to a fairly small database so that one might question how well they predict the forces in an amorphous solid of liquid. This is especially true since relatively small errors in total energies can lead to rather large errors in forces. The BSF methods, which use first-principles molecular-dynamics simulations within a supercell geometry are not fit. However, because of the computational complexity in actually implementing them, they have a different set of difficulties. One can ask how important cell size is and/or what the dependence is on the band curvature or number of k points in the Brillouin zone. In this communication, we briefly investigate these questions. For the ADF models we have limited our investigations to the potentials of Stillinger and Weber (SW),[1] the two potentials of Biswas and Hamann [BH-1 (Ref. 5) and BH-2 (Ref. 6)], and the model of Carlsson, Fedders, and Myles (CFM).[7] Concerning the BSF method, the cell size and k-point sampling were investigated using the *ab initio* total-energy–molecular-dynamics scheme of Sankey and Niklewski.[12]

For the BSF model, we find that the use of only the Γ (k=0) point in samples of from 32 to 54 atoms leads to typical errors of 0.3–0.4 eV/Å. These are small in comparison with typical force scales in Si that are of order several eV/Å. However, errors of this magnitude caused one sample to remain amorphous upon annealing, when a more accurate calculation led to crystallization. Further, we have generally found fewer defects in samples annealed with a more accurate procedure. This suggests that the energy barriers (or necessary forces) for crystallization and defect healing are much lower than the natural scale in Si. Increasing the number of points to eight can still lead to non-negligible error. The ADF models all yield errors several times larger than the Γ-point BSF method.

We first examine the accuracy of the BSF method. *Ab initio* molecular-dynamics simulations usually involve the use of "supercells," because it is computationally impossible to simulate a cluster large enough to ignore surface states. In addition, supercells are convenient since it is quite straightforward to use the apparatus of crystalline solids to simulate approximately an infinite system. One chooses a particular Bravais lattice with a large unit cell, and periodically repeats the cell through all space. The aim of this construction is to eliminate spurious surface effects, while providing a cell large enough to include a representative sample of local microstructures. The properties of this infinite system are then easily handled with k-space methods. A crucial aspect of this procedure is the choice of geometry of the cell.

If the supercell is large enough so that intercell interactions are negligible, we are still left with the requirement of handling the "supercell bands" properly. Because the supercell lattice constant is much larger than typical interatomic spacings, the supercell bands are much narrower than for a system with a smaller unit cell. In this paper we examine the effects of curvature and dispersion of the supercell bands on the atomic forces, in some commonly used supercell geometries. The systems we treat are disordered and have a finite electronic density of states at the Fermi level. The Fermi level's crossing the supercell bands is expected to enhance the effects of the band dispersion on the total energy. If all bands were completely filled or empty, then only the center of gravity of a band, and not its width, would affect the total energy. This is not true for partially filled bands. Typical ADF methods yield 15% or more defects, which yields a sizable density of states at the Fermi level.

It has become a common practice in *ab initio* molecular-dynamics simulations to use one k point in the Brillouin zone to calculate energies and forces. In effect this assumes that the supercell bands are flat. For computational reasons, the Γ point (k=0) is usually selected for this purpose. To evaluate this procedure, we have calculated the supercell bands for a 54-atom fcc structure that is highly disordered. There is significant dispersion in the bands for the 54-atom cell, with widths of order 0.1 eV. We performed a similar calculation for the 54-atom

 © 1990 The American Physical Society

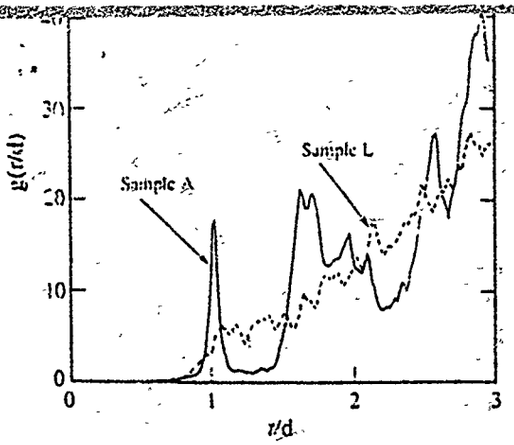


FIG. 1. Radial distribution function of samples A and L in reduced units of $d=2.35$ Å.

ever, starting with the same near equilibrium sample, when it was annealed further with four k points, it became a crystal. Thus, the structures that emerge are very sensitive to the interatomic forces. Although (as mentioned above) the annealing sequence that was used is certainly far removed from the actual physical growth process for a-Si, we see no reason why the physical growth process should be less sensitive to the forces. Sample L was obtained in the same fashion as sample A, but the procedure was stopped when the kinetic energy was still several thousand degrees per atom. Its radial distribution function is shown in Fig. 1, and is basically liquid-like. However, the radial distribution function is quite different from the experimental one.[17] This is not a serious problem, since these samples were created in order to test the potentials and not to simulate real l-Si. Thus our two samples include a sample that is close to being a-Si and a sample that is still quite liquid-like. Our results on the forces are summarized in Table I (for the 32-atom sample A) and Table II (for the 54-atom sample L). We have defined ΔF as a root-mean-square deviation

$$\Delta F=\left[N^{-1}\sum_i[\mathrm{F}(i)-\mathrm{F}_e(i)]^2\right]^{1/2},$$

where $\mathrm{F}(i)$ is the computed force on atom i, $\mathrm{F}_e(i)$ is the exact (16-point BSF method) force on atom i, and N is the number of atoms (32 or 54). The quantities $\delta_i F$ in the tables are the errors in the forces for the 3 or 5 worst atoms. The quantity $\Delta F'$ is the same as ΔF except that the worst 10% of the atoms were taken out of the sum and thus N is 29 or 49. The rms force on the atoms in

Since sample A had been extensively annealed, the exact forces are small and thus the relative error in the forces is rather large compared to the rms total force. It is interesting to note the atoms on which the errors were the greatest are highly correlated not only among the various ADF methods but also between the ADF methods and the Γ-point BSF method. For this reason atom numbers of the worst atom are included in parentheses in Tables I and II. Atoms no. 12, 24, and 29 all have one bond angle of less than 90°, atom no. 31 has bond angles between 96° and 127°, and atom no. 9 has a dangling bond. Evidently almost all of the methods have more trouble with badly strained bonds than with dangling bonds. Further, from the tables, one can see that a substantial part of the rms error comes from 10% of the atoms.

For the 54-atom more-liquid-like sample L, we note that all of the ADF methods predict forces that are in error by about the magnitude of the exact forces, while the BSF method is much better, even with one k point. Since we have found no correlations between the methods on which atoms were treated worst, we have not included the atom numbers in this case. We have also noted some dynamical differences between the methods. For example, if one plots $r^2(t)$ versus t with the exact BSF method, one obtains a reasonably straight line indicating the expected diffusive behavior. However, with only one k point, r^2 is supralinear, suggesting a partially ballistic behavior.

Some further general trends are evident as one goes from the least-accurate methods (the ADF models) to Γ-point BSF calculations and finally to exact BSF calculations. One trend that is evident among all of the methods is that more accurate methods tend to favor crystals and methods with larger errors tend to favor disorder. As noted earlier, we found samples that had converged to an amorphous sample using the Γ-point BSF method but that became crystalline when annealed further with more k points. Further, it has become clear from other investigations[18–21] that the ADF methods invariably give very large (or order 15–20%) concentrations of defects. Another trend is evident on a shorter time scale. This is the observation that distortions from perfect crystal or from tetrahedral coordination are relaxed faster and more completely with the better methods.

In order to understand better the origins of the errors caused by using a finite number of k points, we have performed some additional calculations. It is well known

TABLE I. Comparison of the errors in the 32-atom cluster, sample A. The root-mean-square error ΔF is defined in the text and $\delta_i F$ are the errors in the forces for the worst three atoms. The numbers in parentheses refer to the one atom number in the cluster and all units are eV/Å.

Method	ΔF	$\Delta F'$	$\delta_1 F$	$\delta_2 F$	$\delta_3 F$
BS (k=0)	0.34	0.23	0.74(12)	0.62(21)	0.54(29)
BH-1	1.54	1.41	2.80(9)	2.27(29)	2.25(19)
BH-2	0.74	0.57	2.06(12)	1.49(31)	1.28(29)
SW	1.09	0.66	3.13(31)	3.12(18)	2.48(12)
CFM	1.01	0.81	2.43(12)	2.07(31)	1.84(24)

TABLE II. Errors in the forces for the 54-atom cluster, sample L. The root-mean-square error ΔF is defined in the text and $\delta_i F$ are the errors in the forces for the worst five atoms. All units are eV/Å.

Method	ΔF	$\Delta F'$	$\delta_1 F$	$\delta_2 F$	$\delta_3 F$	$\delta_4 F$	$\delta_5 F$
BS (4 k points)	0.15	0.14	0.24	0.23	0.20	0.20	0.20
BS(k=0)	0.43	0.39	0.83	0.80	0.74	0.72	0.68
BH-1	15.54	7.48	65.65	61.41	31.38	30.63	17.11
BH-2	2.45	2.19	4.40	4.40	4.39	4.09	3.85
SW	4.01	3.52	8.45	7.95	6.91	6.40	6.34
CFM	6.98	4.54	25.69	23.99	12.96	10.97	9.99

that in order to obtain accurate Brillouin-zone (BZ) integrations, one needs more k points for materials with a finite density of states at the Fermi energy than for materials with a gap at the Fermi energy. To establish the importance of this effect, we performed calculations for the energies and forces on crystals that were allowed to disorder slightly by small random motions. In case I we allowed rms excursions of roughly 0.1 Å. This simple retained a gap of about 1 eV. In case II we allowed rms excursions of roughly 0.2 Å, and the gap barely closed. Both samples contained 54 atoms. In both cases the error obtained by using only one k points was 0.26 eV per atom. However, the rms error in the force in case I was 0.13 eV/Å, while the error was 0.53 eV/Å in case II. This suggests that the majority of the error is in the bands crossing the Fermi surface. Further, we note that the energy differences between different crystal structures are often less than 0.26 eV per atom.

One can understand this result in the following way. We view the use of different numbers of k points in the BZ as corresponding to different interpolations for $E(\mathbf{k})$ versus k between the chosen k points. The approximate $E(\mathbf{k})$ then oscillates about the exact $E(\mathbf{k})$ over a distance $\Delta\mathbf{k}$ in k space and with an amplitude proportional to $\Delta\mathbf{k}$. For bands crossing the Fermi surface the integral is cut short, leading to errors of order $(\Delta\mathbf{k})^2$ (a length times a height). The first factor of $\Delta\mathbf{k}$ comes from the number of wave functions that are affected, and the second one comes from the distance of these wave functions from the Fermi level. On the other hand, the error in the charge density, which determines the electric field and thus the forces, is of order $\Delta\mathbf{k}$. Thus the total energy is in error by terms of order $(\Delta\mathbf{k})^2$, while the forces are in error by terms that are linear in $\Delta\mathbf{k}$.

In conclusion, we have shown that the ADF methods studied yield forces for badly disordered samples far from equilibrium that are in error by about the magnitude of the forces. In terms of absolute error, these methods are much better for samples that are nearer to equilibrium but are still only qualitatively correct. Whether they can yield reliable defect or surface structures is open to question. The Γ-point BSF method for small clusters is worst when far from equilibrium. For such clusters the forces are in error by about 0.4 eV/Å and in error by about 0.1–0.2 eV/Å when using four k points. The absolute error in the forces when using only one k point is rather small for nearly converged samples, but this small error can make the difference between an amorphous and crystalline sample.

This work was supported in part by the Office of Naval Research (ONR), U.S. Department of Defense, under Grants No. N00014-90-J-1304 and N-00014-84-K-03552, by the National Science Foundation (NSF) under Grant No. DMR-88-01260, the U.S. Department of Energy under Grant No. DE-FG02-84ER45130, and by the Solar Energy Research Institute (SERI) under Subcontract No. XB-7-06055-1 of Contract No. DE-AC02-83CH10093. We wish to acknowledge considerable help from Stefan Klemm.

[1]F. H. Stillinger and T. A. Weber, Phys. Rev. B 31, 5202 (1985).
[2]J. Tersoff, Phys. Rev. Lett. 56, 632 (1986); Phys. Rev. B 37, 6991 (1988).
[3]M. I. Baskes, Phys. Rev. Lett. 59, 2666 (1988).
[4]M. I. Baskes, J. S. Nelson, and A. F. Wright, Phys. Rev. B 40, 6085 (1989).
[5]R. Biswas and D. R. Hamann, Phys. Rev. Lett. 55, 2001 (1985).
[6]R. Biswas and D. R. Hamann, Phys. Rev. B 36, 6434 (1987).
[7]A. E. Carlsson, P. A. Fedders, and C. W. Myles, Phys. Rev. B 41, 1247 (1990).
[8]A. E. Carlsson, in *Solid State Physics: Advances in Research and Applications*, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1990), Vol. 43, p. 1.
[9]I. Štich, R. Car, and M. Parrinello, Phys. Rev. Lett. 63, 2240 (1989).
[10]R. Car and M. Parrinello, Phys. Rev. Lett. 60, 204 (1988).
[11]G. Galli, R. Martin, R. Car, and M. Parrinello, Phys. Rev. Lett. 62, 555 (1989).
[12]O. F. Sankey and D. J. Niklewski, Phys. Rev. B 40, 3979 (1989).
[13]H. J. Monkhorst and J. D. Pack, Phys. Rev. B 13, 5188 (1976).
[14]Coordinates and forces for these samples are available upon request.
[15]F. Fortner and J. S. Lanin, Phys. Rev. B 39, 5527 (1989).
[16]J. H. Etherington, A. C. Wright, J. T. Wenzel, J. C. Dore, J. H. Clarke, and R. N. Sinclair, J. Non-Cryst. Solids 48, 265 (1982).
[17]J. P. Gabathuler and S. Steeb, Z. Naturforsch. 34A, 1314 (1979).
[18]R. Biswas, I. Kwon, A. M. Bouchard, C. M. Soukoulis, and G. S. Grest, Phys. Rev. B 39, 5101 (1989); R. Biswas, Gary S. Grest, and C. M. Soukoulis, *ibid.* 38, 8154 (1988).
[19]J. Q. Broughton and X. P. Li, Phys. Rev. B 35, 9120 (1987).
[20]M. D. Kluge, J. R. Ray, and A. Rahman, Phys. Rev. B 36, 4234 (1987).
[21]W. D. Luedtke and U. Landman, Phys. Rev. B 37, 4656 (1988).

function, but rather perform the integrals numerically from a wave-function tabulation.

The total energy in Eq. (1) can be conveniently rewritten as a sum over eigenvalues of the single-particle Hamiltonian,

$$E_{\rm tot}(n)=E_{\rm BS}(n)+[U_{ii}-U_{ee}(n,n)]+\delta U_{\rm xc}(n)\ , \quad (3a)$$

where $E_{\rm BS}(n)$ is the "band-structure" energy and is given by a sum of one-electron eigenvalues over occupied states i,

$$E_{\rm BS}(n)=2\sum_i \varepsilon_i\ . \quad (3b)$$

The quantities ε_i are the single-particle Hamiltonian eigenvalues that satisfy the Schrödinger equation in matrix form

$$\sum_{\nu,\alpha'} h^{\alpha,\alpha'}_{\mu,\nu} a_i(\nu,\alpha')=\varepsilon_i \sum_{\nu,\alpha'} S^{\alpha,\alpha'}_{\mu,\nu} a_i(\nu,\alpha')\ , \quad (4a)$$

where the single-particle Hamiltonian matrix elements are

$$h^{\alpha,\alpha'}_{\mu,\nu}=\langle \phi^{\rm PAO}_\mu(\mathbf{r}-\mathbf{r}_\alpha)|h(n)|\phi^{\rm PAO}_\nu(\mathbf{r}-\mathbf{r}_{\alpha'})\rangle\ , \quad (4b)$$

and the overlap matrix is

$$S^{\alpha,\alpha'}_{\mu,\nu}=\langle \phi^{\rm PAO}_\mu(\mathbf{r}-\mathbf{r}_\alpha)|\phi^{\rm PAO}_\nu(\mathbf{r}-\mathbf{r}_{\alpha'})\rangle\ . \quad (4c)$$

The standard single-particle LDA Hamiltonian operator is

$$h(n)=\frac{p^2}{2m}+\sum_\alpha [V_{\rm ion}(\mathbf{r}-\mathbf{r}_\alpha)+V_{\rm nl}(\mathbf{r}-\mathbf{r}_\alpha)] +e^2\int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r'+\mu_{\rm xc}(n)\ , \quad (5)$$

where $\mu_{\rm xc}(n)=d[n\varepsilon_{\rm xc}(n)]/dn$ is the exchange-correlation potential. The other terms in Eq. (3a) are the "short-ranged" repulsive potential

$$U_{\rm SR}=U_{ii}-U_{ee}(n,n) =\frac{e^2}{2}\left[\sum_{\alpha,\alpha'}{}' \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{r}_\alpha-\mathbf{r}_{\alpha'}|}-\int\int d^3r\, n(\mathbf{r})\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r'\right] \quad (6)$$

and an exchange-correlation correction, $\delta U_{\rm xc}$,

$$\delta U_{\rm xc}=\int n(\mathbf{r})[\varepsilon_{\rm xc}(n)-\mu_{\rm xc}(n)]d^3r\ . \quad (7)$$

The Schrödinger-like Eq. (4a) follows from a variational principle for the total energy of Eq. (1). The single-particle Hamiltonian, Eq. (5), defines a problem requiring a self-consistent solution. To avoid the difficulties associated with iterating to self-consistency, we adopt an approximation suggested by Harris.[27] Thus our *third* major approximation is to consider a sum of neutral-atom spherical charge densities as a zero-order approximation to the self-consistent density of the cluster, and to keep only first-order changes from this density in the energy functional. Thus we write

$$n(\mathbf{r})=n^0(\mathbf{r})+\delta n(\mathbf{r})\ , \quad (8)$$

where

$$n^0(\mathbf{r})=\sum_\alpha n_{\rm atom}(\mathbf{r}-\mathbf{r}_\alpha)\ . \quad (9)$$

The neutral-atom density for Si is taken to be a spherically symmetric s^2p^2 configuration. To first order in δn, the energy functional of Eq. (3a) simplifies to

$$E^{(1)}_{\rm tot}=E^{(1)}_{\rm BS}+[U_{ii}-U_{ee}(n^0,n^0)]+\delta U_{\rm xc}(n^0)\ , \quad (10)$$

where $E^{(1)}_{\rm BS}$ is the "band-structure" energy determined from the single-particle Hamiltonian of Eq. (5) with n replaced by n^0, viz., $h(n^0)$. This functional neglects terms of order δn^2.

Our approach is to use this approximate total-energy functional rather than the fully self-consistent total-energy functional. The advantages of this method are (1) the electronic eigenvalue equation only needs to be solved once for each atomic configuration instead of ~10 times as in a fully self-consistent calculation and (2) four-center Coulomb integrals do not need to be evaluated since they appear only in the δn^2 terms, which are neglected. This non-self-consistent method has been tested by Harris,[27] Polatoglou *et al.*,[28] and Read *et al.*[29] on metals, semiconductors, and an ionic compound (NaCl) and it is found to give surprisingly good agreement with self-consistent calculations.

Our *fourth* major approximation involves the evaluation of the matrix elements of the various terms that make up $h(n^0)$. The kinetic energy, overlap, and nonlocal part of the pseudopotential are easily evaluated exactly (numerically) for each geometry. A table of results on a fine grid of separations is constructed so that the matrix elements for any separation between the atoms is accurately interpolated. The three center matrix elements of the neutral-atom potential given by

$$V_{\rm NA}(\mathbf{r}-\mathbf{r}_\alpha)=V_{\rm ion}(\mathbf{r}-\mathbf{r}_\alpha)+e^2\int \frac{n(\mathbf{r}'-\mathbf{r}_\alpha)}{|\mathbf{r}-\mathbf{r}'|}d^3r' \quad (11)$$

are accurately approximated by an r-dependent expansion in multipole moments. The nonlinear exchange-correlation matrix elements are calculated within the average density approximation described in Ref. 23. All matrix elements are calculated in real space.

The forces are determined by differentiating the total energy $E^{(1)}_{\rm tot}$ given by Eq. (10), $\mathbf{F}_\gamma=-\partial E^{(1)}_{\rm tot}/\partial \mathbf{r}_\gamma$. The most difficult term comes from the band structure. Its derivative is

$$\frac{-\partial E^{(1)}_{\rm BS}}{\partial \mathbf{r}_\gamma}=-2\sum_{i\,({\rm occ})}\frac{\partial}{\partial \mathbf{r}_\gamma}\langle\psi_i|h(n^0)|\psi_i\rangle =-\frac{\partial}{\partial \mathbf{r}_\gamma}\sum_{\substack{\mu,\nu\\ \alpha,\alpha'}}\rho^{\alpha\alpha'}_{\mu\nu}(h^0)^{\alpha\alpha'}_{\mu\nu} =-\sum_{\substack{\mu,\nu\\ \alpha,\alpha'}}\left[\rho^{\alpha\alpha'}_{\mu\nu}\frac{\partial(h^0)^{\alpha\alpha'}_{\mu\nu}}{\partial \mathbf{r}_\gamma}-E^{\alpha\alpha'}_{\mu\nu}\frac{\partial S^{\alpha\alpha'}_{\mu\nu}}{\partial \mathbf{r}_\gamma}\right]\ , \quad (12)$$

where $\rho^{\alpha\alpha'}_{\mu\nu}$ and $E^{\alpha\alpha'}_{\mu\nu}$ are the density and energy-density matrices, respectively,

similar to that recently proposed by Foulkes *et al.*[24]

In this paper, we use this method to determine the equilibrium structures and vibrational spectra of small Si clusters (Si_n with $n=2-7$). The technique of molecular dynamics is employed, in which the many-body classical equations of motion, $\mathbf{F}=m\,d^2\mathbf{r}/dt^2$, is solved in time, and the subsequent motion of the atoms is determined from some set of initial conditions. Equilibrium structures were found by simulated annealing and dynamical quenching (in which the system is periodically quenched by removing kinetic energy from its atoms). In finding the vibrational spectra, the velocity autocorrelation functions are Fourier transformed. The theory and techniques described in this paper are immediately transferable to condensed-matter systems, which will be dealt with in future publications.

II. ELECTRONIC-STRUCTURE THEORY

In this section, we briefly describe the theory used to determine the electronic structure and to obtain the total energy of these systems and forces between atoms. We will describe four major approximations that simplify the electronic structure tremendously, so that medium scale (~100 atoms or more) molecular-dynamics simulations can easily be performed. Only a brief description of the theory will be given, as a more complete description can be found in Ref. 23.

The theoretical foundation used is density-function theory. Within this rigorous ground-state theory, we make our *first* major approximations. The approximations are the use of the Hohenberg-Kohn-Sham local-density approximation (LDA) and the pseudopotential approximation. The LDA replaces the exchange-correlation energy functional by a local function of the density. We use the local exchange-correlation functional of Ceperley and Alder as parametrized by Perdew and Zunger.[25] The pseudopotential approximation replaces the core electrons by an effective potential that acts on the valence electrons. Accurate nonlocal (angular-momentum-dependent) pseudopotentials of the norm-conserving Hamann-Schlüter-Chiang type are used.[26]

In the LDA and nonlocal pseudopotential approximation, the total-energy functional is given by

$$E_{\rm tot}=T_K+\int\left\{\sum_a[V_{\rm ion}(\mathbf{r}-\mathbf{r}_a)+V_{\rm nl}(\mathbf{r}-\mathbf{r}_a)]+\frac{e^2}{2}\int\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r'+\varepsilon_{\rm xc}(n)\right\}n(\mathbf{r})\,d^3r\ , \tag{1}$$

where the individual terms are, respectively, the kinetic energy, the ionic local pseudopotential of the ion at r_a, the nonlocal pseudopotential, the electron-electron Hartree repulsion, and the exchange-correlation energy, which is a functional of the electron density n.

The electronic energy eigenstates ψ_i are the expanded in a tight-binding-like linear combination of pseudo-atomic-orbitals (PAO's):

$$|\psi_i(r)\rangle=\sum_{\mu,a}a_i(\mu,a)|\phi_\mu^{\rm PAO}(\mathbf{r}-\mathbf{r}_a)\rangle\ . \tag{2}$$

The PAO's are self-consistently determined eigenfunctions of the valence electrons of the free atom in the nonlocal pseudopotential approximation and are nodeless. The pseudoatom contains only the valence electrons, so that the orbital types μ for Si are s, p_x, p_y, and p_z.

Our *second* major approximation is motivated by the need to reduce the range of interaction between atomic orbitals, and hence, greatly reduce the number of neighbors each atom (or pair of atoms) interacts with. This reduction leads to a corresponding reduction in computer time required for the calculations. We do this by slightly exciting the PAO's by imposing the boundary condition that they vanish at and outside a predetermined radius r_c. (As $r_c\to\infty$, the atom approaches the ground state.) The value of r_c that we use here and in previous work is $5.0a_B$. The motivation for this value is that it rigorously yields a third-neighbor model for matrix elements of the single-particle Hamiltonian in crystalline Si. We have found that our results are not critically dependent on the precise value of r_c as long as r_c is not too small. The kinetic energy associated with the confinement of the electron in the atom begins a sharp increase at values of r_c slightly less than $5a_B$. A plot of the Si s-orbital wave function in the pseudopotential approximation for various values of r_c is shown in Fig. 1. The curve for $r_c=8a_B$ is virtually identical to the ground-state wave function. Notice that for $r_c=5a_B$, the wave function in the bonding region is well represented, but the long-ranged tail is eliminated. In the multicenter integrals needed in this work, we do not fit the wave function to any analytic

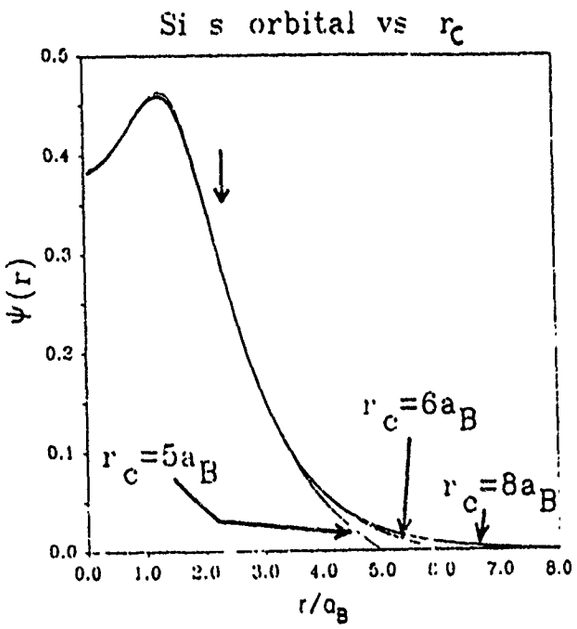


FIG. 1. The s pseudo-atomic-orbital of Si using various values of r_c. The wave function for $r_c=8a_B$ is very close to the pseudoatom ground state. The bonding region (defined to be half the nearest-neighbor distance in bulk Si) is shown by the vertical arrow.

function, but rather perform the integrals numerically from a wave-function tabulation.

The total energy in Eq. (1) can be conveniently rewritten as a sum over eigenvalues of the single-particle Hamiltonian,

$$E_{\rm tot}(n)=E_{\rm BS}(n)+[U_{ii}-U_{ee}(n,n)]+\delta U_{\rm xc}(n)\ , \quad (3a)$$

where $E_{\rm BS}(n)$ is the "band-structure" energy and is given by a sum of one-electron eigenvalues over occupied states i,

$$E_{\rm BS}(n)=2\sum_i \varepsilon_i\ . \quad (3b)$$

The quantities ε_i are the single-particle Hamiltonian eigenvalues that satisfy the Schrödinger equation in matrix form

$$\sum_{\nu,\alpha'} h_{\mu,\nu}^{\alpha,\alpha'} a_i(\nu,\alpha')=\varepsilon_i \sum_{\nu,\alpha'} S_{\mu,\nu}^{\alpha,\alpha'} a_i(\nu,\alpha')\ , \quad (4a)$$

where the single-particle Hamiltonian matrix elements are

$$h_{\mu,\nu}^{\alpha,\alpha'}=\langle \phi_\mu^{\rm PAO}(\mathbf{r}-\mathbf{r}_\alpha)|h(n)|\phi_\nu^{\rm PAO}(\mathbf{r}-\mathbf{r}_{\alpha'})\rangle\ , \quad (4b)$$

and the overlap matrix is

$$S_{\mu,\nu}^{\alpha,\alpha'}=\langle \phi_\mu^{\rm PAO}(\mathbf{r}-\mathbf{r}_\alpha)|\phi_\nu^{\rm PAO}(\mathbf{r}-\mathbf{r}_{\alpha'})\rangle\ . \quad (4c)$$

The standard single-particle LDA Hamiltonian operator is

$$h(n)=\frac{p^2}{2m}+\sum_\alpha [V_{\rm ion}(\mathbf{r}-\mathbf{r}_\alpha)+V_{\rm nl}(\mathbf{r}-\mathbf{r}_\alpha)] +e^2\int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r'+\mu_{\rm xc}(n)\ , \quad (5)$$

where $\mu_{\rm xc}(n)=d[n\varepsilon_{\rm xc}(n)]/dn$ is the exchange-correlation potential. The other terms in Eq. (3a) are the "short-ranged" repulsive potential

$$U_{\rm SR}=U_{ii}-U_{ee}(n,n) =\frac{e^2}{2}\left[\sum_{\alpha,\alpha'}{}' \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{r}_\alpha-\mathbf{r}_{\alpha'}|}-\int\int d^3r\, n(\mathbf{r})\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r'\right] \quad (6)$$

and an exchange-correlation correction, $\delta U_{\rm xc}$,

$$\delta U_{\rm xc}=\int n(\mathbf{r})[\varepsilon_{\rm xc}(n)-\mu_{\rm xc}(n)]d^3r\ . \quad (7)$$

The Schrödinger-like Eq. (4a) follows from a variational principle for the total energy of Eq. (1). The single-particle Hamiltonian, Eq. (5), defines a problem requiring a self-consistent solution. To avoid the difficulties associated with iterating to self-consistency, we adopt an approximation suggested by Harris.[27] Thus our *third* major approximation is to consider a sum of neutral-atom spherical charge densities as a zero-order approximation to the self-consistent density of the cluster, and to keep only first-order changes from this density in the energy functional. Thus we write

$$n(\mathbf{r})=n^0(\mathbf{r})+\delta n(\mathbf{r})\ , \quad (8)$$

where

$$n^0(\mathbf{r})=\sum_\alpha n_{\rm atom}(\mathbf{r}-\mathbf{r}_\alpha)\ , \quad (9)$$

The neutral-atom density for Si is taken to be a spherically symmetric s^2p^2 configuration. To first order in δn, the energy functional of Eq. (3a) simplifies to

$$E_{\rm tot}^{(1)}=E_{\rm BS}^{(1)}+[U_{ii}-U_{ee}(n^0,n^0)]+\delta U_{\rm xc}(n^0)\ , \quad (10)$$

where $E_{\rm BS}^{(1)}$ is the "band-structure" energy determined from the single-particle Hamiltonian of Eq. (5) with n replaced by n^0, viz., $h(n^0)$. This functional neglects terms of order δn^2.

Our approach is to use this approximate total-energy functional rather than the fully self-consistent total-energy functional. The advantages of this method are (1) the electronic eigenvalue equation only needs to be solved once for each atomic configuration instead of ~10 times as in a fully self-consistent calculation and (2) four-center Coulomb integrals do not need to be evaluated since they appear only in the δn^2 terms, which are neglected. This non-self-consistent method has been tested by Harris,[27] Polatoglou *et al.*,[28] and Read *et al.*[29] on metals, semiconductors, and an ionic compound (NaCl) and it is found to give surprisingly good agreement with self-consistent calculations.

Our *fourth* major approximation involves the evaluation of the matrix elements of the various terms that make up $h(n^0)$. The kinetic energy, overlap, and nonlocal part of the pseudopotential are easily evaluated exactly (numerically) for each geometry. A table of results on a fine grid of separations is constructed so that the matrix elements for any separation between the atoms is accurately interpolated. The three center matrix elements of the neutral-atom potential given by

$$V_{\rm NA}(\mathbf{r}-\mathbf{r}_\alpha)=V_{\rm ion}(\mathbf{r}-\mathbf{r}_\alpha)+e^2\int \frac{n(\mathbf{r}'-\mathbf{r}_\alpha)}{|\mathbf{r}-\mathbf{r}'|}d^3r' \quad (11)$$

are accurately approximated by an r-dependent expansion in multipole moments. The nonlinear exchange-correlation matrix elements are calculated within the average density approximation described in Ref. 23. All matrix elements are calculated in real space.

The forces are determined by differentiating the total energy $E_{\rm tot}^{(1)}$ given by Eq. (10), $\mathbf{F}_\gamma=-\partial E_{\rm tot}^{(1)}/\partial \mathbf{r}_\gamma$. The most difficult term comes from the band structure. Its derivative is

$$\frac{-\partial E_{\rm BS}^{(1)}}{\partial \mathbf{r}_\gamma}=-2\sum_{i\,({\rm occ})}\frac{\partial}{\partial \mathbf{r}_\gamma}\langle \psi_i|h(n^0)|\psi_i\rangle =-\frac{\partial}{\partial \mathbf{r}_\gamma}\sum_{\substack{\mu,\nu\\ \alpha,\alpha'}}\rho_{\mu\nu}^{\alpha\alpha'}(h^0)_{\mu\nu}^{\alpha\alpha'} =-\sum_{\substack{\mu,\nu\\ \alpha,\alpha'}}\left[\rho_{\mu\nu}^{\alpha\alpha'}\frac{\partial (h^0)_{\mu\nu}^{\alpha\alpha'}}{\partial \mathbf{r}_\gamma}-E_{\mu\nu}^{\alpha\alpha'}\frac{\partial S_{\mu\nu}^{\alpha\alpha'}}{\partial \mathbf{r}_\gamma}\right]\ , \quad (12)$$

where $\rho_{\mu\nu}^{\alpha\alpha'}$ and $E_{\mu\nu}^{\alpha\alpha'}$ are the density and energy-density matrices, respectively,

$$\rho_{\mu\nu}^{\alpha\alpha'} = \sum_{\substack{i \\ (\mathrm{occ})}} a_i^*(\mu,\alpha) a_i(\nu,\alpha') ,$$

$$E_{\mu\nu}^{\alpha\alpha'} = \sum_{\substack{i \\ (\mathrm{occ})}} \varepsilon_i a_i^*(\mu,\alpha) a_i(\nu,\alpha') ,$$

where $a_i(\mu,\alpha)$ are the expansion coefficients of the wave function in Eq. (2).

After the forces are evaluated, the equations of motion can be solved and the positions and velocities of the atoms updated. The energy functional has been tested in bulk Si and found to give bulk static properties that are in good agreement with experiment.[23] It is the purpose of this paper to extend these calculations to the static and dynamic properties of small Si clusters using an *ab initio* molecular-dynamics technique.

III. MOLECULAR DYNAMICS

The equilibrium structures and vibrational spectra of Si clusters have been determined using the technique of molecular dynamics. The forces are calculated quantum mechanically from the total energy [Eq. (10)] and its derivatives [e.g., Eq. (12)]. The nuclear coordinates are moved in time according to the classical equations of motion, $\mathbf{F}_n = m_n d^2\mathbf{r}/dt^2$. These equations are integrated using the Gear predictor-corrector algorithm,[30] with a time step of ~ 1.5 fs.

The ground-state equilibrium structures are found by one of two techniques: dynamical quenching or simulated annealing. In either of these two techniques, the cluster is started in some trial structure, generally with very little symmetry.

In the technique of dynamical quenching, the atoms in the cluster are allowed to respond to internal forces and are accelerated. A kinetic temperature T_K is defined as the average classical kinetic energy of the atoms, $3/2kT_K = (1/N)\sum_i \frac{1}{2} m_i v_i^2$, where $i = 1,2,\ldots,n$ is the atom index. As the atoms accelerate, the kinetic temperature increases until a maximum is reached. The system is then quenched by setting all the velocities to zero, thus removing all the kinetic energy. The atoms again are allowed to accelerate and the quenching process is repeated. After several dynamical quenching cycles, a minimum energy configuration is obtained. The minimum may be only a local minimum, so that the procedure needs to be repeated several times with different initial configurations.

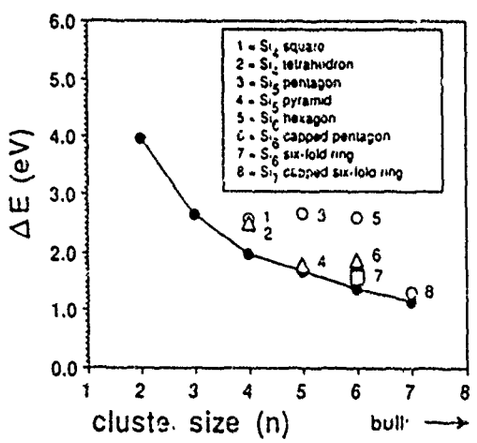


FIG. 2. The energy difference per atom between Si clusters of size $n = 2$–7 and bulk Si. The solid dots correspond to equilibrium structures explained in the text, and other metastable configurations are also shown for $n = 4$–7.

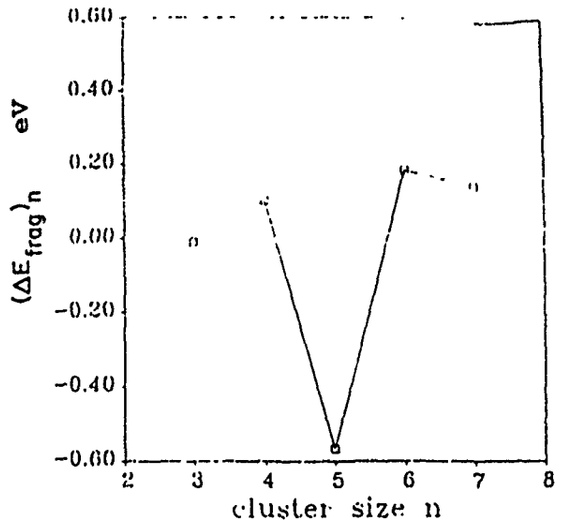


FIG. 3. The n-dependent fragmentation energy $(\Delta E_{\mathrm{frag}})_n$ as a function of n. The dip at $n = 5$ indicates that fragmentation is more likely to occur for this cluster.

In the simulated annealing technique, the energy is removed more gradually so that a global minimum is more likely to be obtained. The technique has been described by Kirkpatrick *et al.*[31] and uses the Monte Carlo algorithm of Metropolis *et al.*[32] Monte Carlo steps are taken where the coordinates of each particle are changed as $\mathbf{r}_i \to \mathbf{r}_i + \delta\mathbf{r}_i$. Here $\delta\mathbf{r}_i = r_0\,\delta\mathbf{R}$, where $\delta\mathbf{R}$ is a triplet of random numbers between -1 and $+1$ for each atom i, and r_0 is the maximum step size. An energy difference $\delta E = -\mathbf{F}_i \cdot \delta\mathbf{r}_i$ is computed and the step is taken if $\delta E < 0$, while if $\delta E > 0$ the step is taken with probability $e^{-\delta E/kT_A}$, where T_A is an annealing temperature. This ensures that after many moves, the ensemble tends to the Boltzmann distribution. The temperature T_A is gradually reduced so that the system settles into the ground state. We have found that the dynamical quenching finds the ground state much more readily than simulated annealing for these small clusters. However, simulated annealing is more likely to find the true ground state than dynamical quenching. In either case, however, a variety of initial configurations must be tried before one can believe that the true ground state has been found.

The vibrational spectrum is found directly from the molecular-dynamics simulation without annealing or quenching. We start in a ground-state configuration and have constrained the motion so that the center of mass of the system of atoms is stationary and there is no angular momentum about the center of mass. To accomplish this, we first give the atoms random velocities commensu-

rate with a given initial kinetic temperature. In general, however, the angular momentum (L) is nonzero. To keep the system from rotating, we adjust the velocities by subtracting the quantity $\mathbf{v}_i = (\mathbf{I}^{-1}\cdot\mathbf{L})\times\mathbf{r}_i$ from the velocity vector $\mathbf{v}_i$ ($i=1,2,\ldots,n$), where $\mathbf{I}$ is the inertia tensor. This procedure removes the angular momentum while still allowing for randomness in the velocities. Finally the velocities are uniformly shifted so that there is no center-of-mass motion.

The frequency spectrum is determined by Fourier transforming the velocity autocorrelation function,

$$g(t)=\sum_{n=1}^{N}\frac{\langle \mathbf{v}_n(t)\cdot\mathbf{v}_n(0)\rangle}{\langle \mathbf{v}_n(0)\cdot\mathbf{v}_n(0)\rangle}\ .$$

Here n is the atom index and the brackets $\langle\cdots\rangle$ indi-

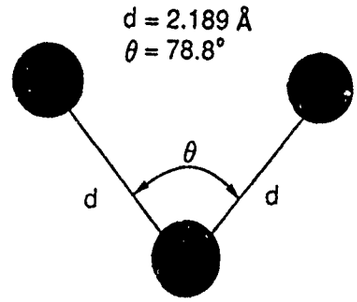


FIG. 5. The ground-state configuration of Si_3.

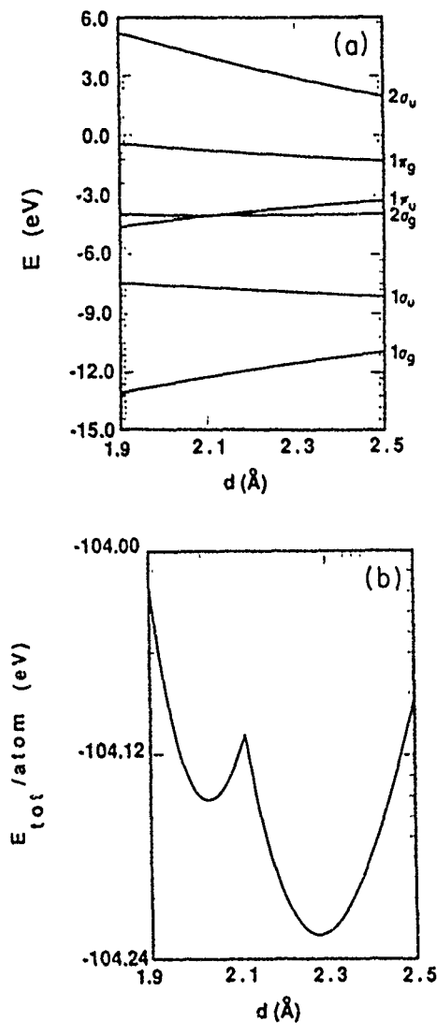


FIG. 4. (a) The one-electron LDA energy eigenvalues vs separation distance d for Si_2. The doubly degenerate (neglecting spin) $1\pi_u$ level crosses the singly degenerate $2\sigma_g$ level at about $d=2.12$ Å. (b) Total energy per atom vs separation distance d for Si_2. A secondary minimum occurs because of the level crossing shown in (a).

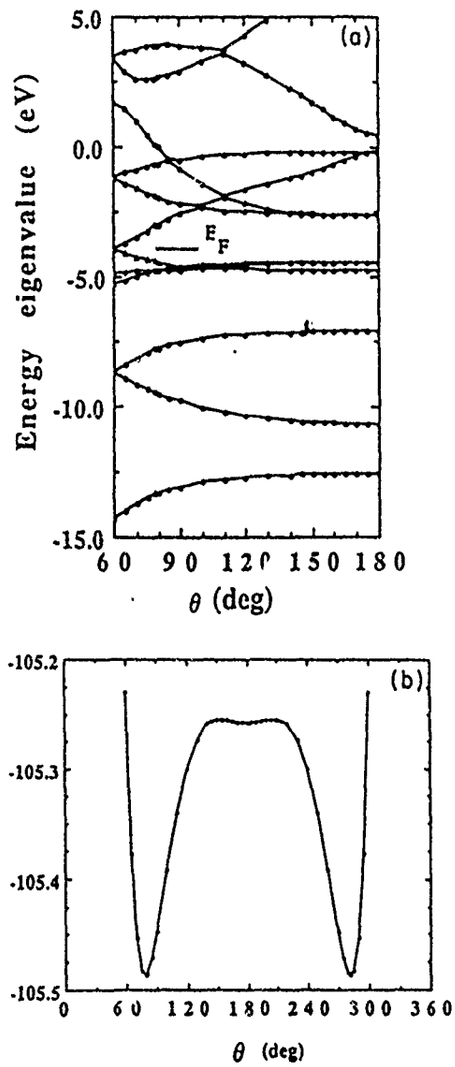


FIG. 6. (a) The one-electron LDA energy eigenvalues versus opening angle θ for Si_3, where the bond lengths d are kept constant at their equilibrium value (2.189 Å). Note the level degeneracies which occur at $\theta=60°$, which is the highly symmetric case of an equilateral triangle. The Fermi level separating the occupied from the unoccupied levels is schematically shown. (b) Total energy per atom vs opening angle θ for Si_3 illustrating the bond bending forces of the Si_3 molecule.

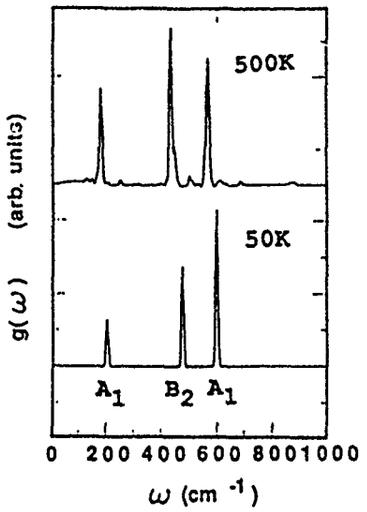
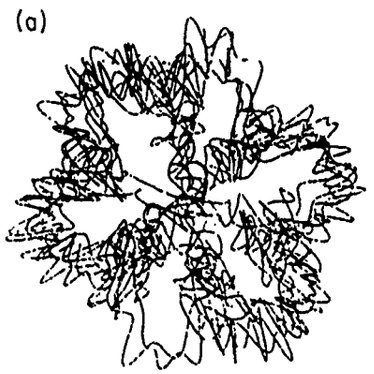


FIG. 7. Spectral density $g(\omega)$ of Si_3 in its equilibrium configuration at a low (lower figure) and a high (upper figure) kinetic temperature. The kinetic temperatures are 50 and 500 K, respectively. The modes soften at higher excitation levels, and there is added structure due to anharmonicity.



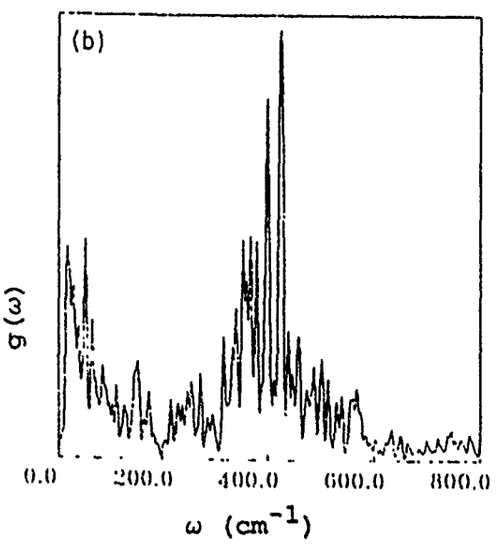


FIG. 8. (a) A snapshot of the planar motion of an Si_3 molecule at high excitation corresponding to a kinetic temperature of ~2500 K. The chaotic motion is very much anharmonic. (b) The spectral density function $g(\omega)$ for the motion in (a). The three peak structure is lost and the spectrum becomes continuous.

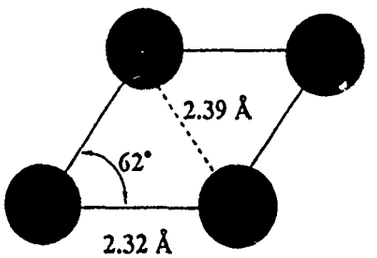


FIG. 9. The ground-state configuration of Si_4.

cate that an ensemble average over all atoms is taken. This ensemble average is defined by

$$\langle f(t_i)f(0)\rangle = \frac{1}{M-i+1}\sum_{m=1}^{M-1} f(t_{i+m})f(t_m)\,,$$

where t_j is the time at the jth step and $j=0,1,2,\ldots,M$, with M being the total number of time steps in the simu-

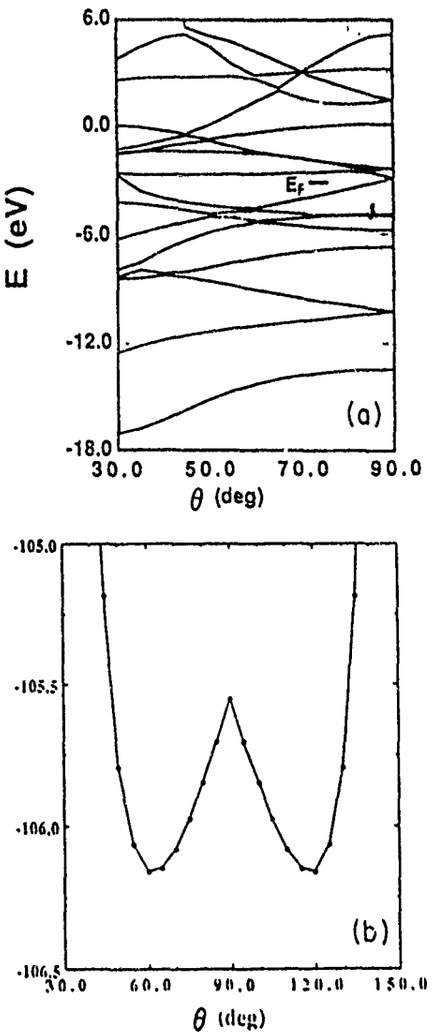


FIG. 10. (a) The one-electron LDA energy eigenvalues vs interior angle θ for Si_4. In the ground state, $\theta=62°$. The side length was kept constant at its ground-state value. Degeneracies and Jahn-Teller instabilities occur at 90°. (b) Total energy per atom vs interior angle θ for Si_4.

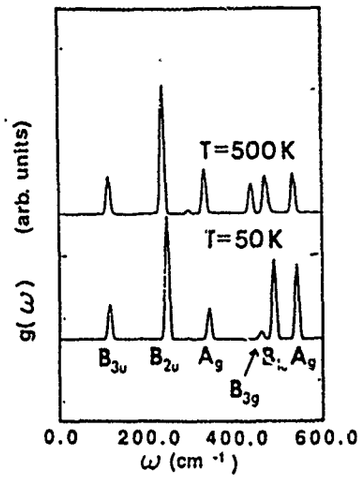


FIG. 11. Spectral density $g(\omega)$ of Si_4 in its equilibrium configuration for high and low kinetic temperatures. The peaks in the spectrum are labeled according to the symmetry of the displacements.

lation. The spectral density $g(\omega)$ is the Fourier cosine transform of the velocity autocorrelation function $g(t)$,

$$g(\omega)=(1/T)\int_0^T g(t)W(t)\cos(\omega t)\,dt\ ,$$

where T is the total time of the simulation. $W(t)$ is the Blackman window function[33] used to reduce oscillations due to finite time sampling.

IV. RESULTS FOR Si CLUSTERS

In this section we individually discuss the results of clusters from Si_n with $n=2$–7. We begin by summarizing the energetics of the various geometries found by quenching and annealing. In Fig. 2 we show the energy difference per atom, $\Delta E_n=(E_{tot}/\text{atom})_{cluster}-(E_{tot}/\text{atom})_{bulk}$, between bulk (diamond) Si and the Si clusters in their ground state and metastable equilibrium configurations.

The solid circles indicate the ground-state configurations (to be described individually), while the open circles, squares, and triangles represent local-minimum (metastable) structures. The energy per atom of the cluster approaches the bulk binding energy monotonically, but there is a marked decrease in slope in going

TABLE I. Comparison of the harmonic frequencies (in cm^{-1}) of the Si_4 molecule in this work with quantum-chemistry SCF results.

Mode symmetry	This work	6-31 G* (Ref. 43)	6-31 G (Ref. 43)
B_{3u}	116	130	137
B_{2u}	248	223	157
A_g	346	380	328
B_{3g}	464	453	371
B_{1u}	495	543	472
A_g	546	503	425

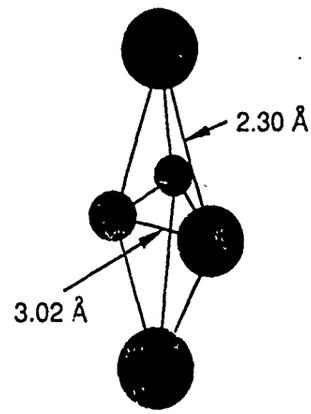


FIG. 12. Ground-state configuration of Si_5. This structure is a trigonal bipyramid, consisting of an equilateral triangle with one atom above the plane of the triangle and one atom below.

from $n=4$ to $n=5$. This can be seen more clearly by considering the fragmentation energy as defined by Raghavachari and Logovinsky[34] according to the reaction $Si_n \rightarrow Si_{n-1}+Si$. The energy of this reaction is

$$(E_{frag})_n=(\Delta E_{frag})_n+[E_1-(E_{tot}/\text{atom})_{bulk}]\ , \qquad (13)$$

where $(\Delta E_{frag})_n=(n-1)\Delta E_{n-1}+n\Delta E_n$. The last term in Eq. (13) is a constant (independent of n). This term involves the energy of an isolated atom, which is not well described in our method because of spin effects and the use of compact orbitals.[35] The trends with n of $(E_{frag})_n$ are entirely contained in the first term $(\Delta E_{frag})_n$ that is plotted in Fig. 3. The lesser stability of Si_5 compared to neighboring clusters is clearly evident. The results are in very close agreement with the fragmentation trends found by Raghavachari.[36]

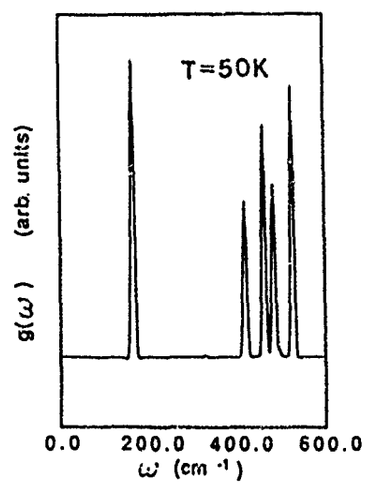


FIG. 13. Spectral density $g(\omega)$ of Si_5 in its equilibrium configuration at low kinetic temperature ($T=50$ K).

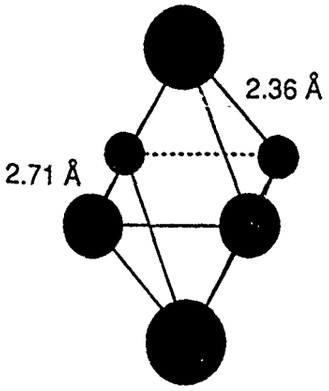


FIG. 14. Ground-state configuration of Si_6. This structure is a square bipyramid, consisting of a square with one atom above the plane of the square and one atom below.

A. Si_2

Dynamic quenching of Si_2 gives the equilibrium bond distance at $d=2.27$ Å, in good agreement with the experimental value of 2.24 Å.[37] The bulk bond length of crystalline Si is 2.35 Å, while the present theory gives 2.38 Å. Thus we find especially good agreement between theory and experiment for the large reduction in nearest-neighbor distance in going from bulk Si to Si_2 (0.11 Å experiment and 0.11 Å theory).

The single harmonic-vibrational stretch mode is determined to be $\omega=531$ cm^{-1}, which compares well to the experimental result of 511 cm^{-1}.[37] In this case at least, our results are comparable to the considerably more sophisticated quantum chemistry self-consistent-field (SCF) results of Ref. 36 who find 568 cm^{-1}.

In Fig. 4(a) are shown the eigenvalues of the one-electron LDA Hamiltonian for Si_2 as a function of Si-Si distance. The $1\pi_u$ electronic level is fourfold degenerate (including spin) and contains two electrons in equilibrium ($d=2.27$ Å). The $2\sigma_g$ level crosses the $1\pi_u$ level as the distance is decreased. This is in excellent agreement with the spin-polarized results of Northrup *et al.*,[38] who find that the $1\pi_u$ level crosses the doubly degenerate $2\sigma_g$ level at about $4.0a_B$ (2.12 Å), resulting in a new ground-state configuration at smaller separations. We also find the $1\pi_u$ level crossing the $2\sigma_g$ level at $d=2.12$ Å. This results in a potential-energy surface with two minima as shown in Fig. 4(b). A metastable configuration exists at $d\simeq 2.03$ Å.

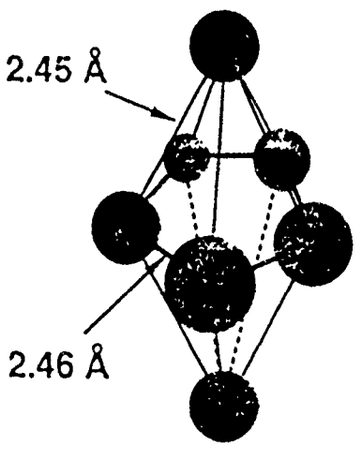


FIG. 15. Ground-state configuration of Si_7. This structure is a pentagonal bipyramid, consisting of a pentagon with one atom above the plane of the pentagon and one atom below.

B. Si_3

The Si_3 ground-state configuration we find by dynamical quenching is an isosceles triangle with the two equal sides of length $d=2.189$ Å and an opening angle of $\theta=78.8°$ (see Fig. 5). These results are in good agreement with the more rigorous quantum-chemistry SCF calculations of Diercksen *et al.*,[39] who find $d=2.196$ Å and $\theta=80.6°$. Grev *et al.*[40] who find $d=2.160$ Å and $\theta=78.1°$, and Raghavachari *et al.*[34] who find $d=2.17°$ and $\theta=77.8°$.

The energy eigenvalues as a function of the angle θ (with the distance d kept constant at its value at equilibrium) are shown in Fig. 6(a). The Fermi level lies between the sixth and seventh levels. These two levels become degenerate at $\theta=60°$ (an equilateral triangle), which make this geometry Jahn-Teller unstable. The total energy as a function of θ (again with d fixed) is shown in Fig. 6(b). Two equivalent minimum are seen corresponding to $\theta=78.8°$ and $\theta=281.2°$. The system is clearly unstable at 60°, and a relatively large barrier separates the two equivalent minimum as the molecule passes through 180°, where the molecule forms a linear chain as occurs for C_3. Curiously, we find a very shallow local minimum at the linear chain. This is to be contrasted to the SCF results of Raghavachari,[14] which find the linear chain to be unstable. The imaginary frequency found in that work (82i cm^{-1}) is very small, indicating a relatively flat potential surface similar to ours.

The spectral density function $g(\omega)$ in the harmonic ($T_K=50$ K) and the near-harmonic ($T_K=500$ K) regions for Si_3 vibrations are shown in Fig. 7. Three peaks are apparent corresponding to the $3n-6$ normal modes of vibration. The width of the peaks is due to the finite time of the simulation. The position of the peaks indicates the frequency of the normal mode, while the relative heights reflect the (randomly chosen) relative amplitudes in each of the modes. The spectrum at low temperature is pure, indicating very little anharmonicity, while at higher temperatures, small contributions to anharmonic affects (sum and difference frequencies) are evident. Also notice the overall shift toward lower frequencies. The "harmonic" frequencies at 50 K are 202, 474, and 596 cm^{-1} for the A_1, B_2, and A_1 symmetric modes, respectively. These can be compared with those obtained by Grev *et al.*[40] of 157, 570, and 574 and Raghavachari[41] of 206, 560, and 582. We note that both of our A_1 modes are in good agreement with these quantum-chemistry SCF calculations, while the B_2 mode is only in fair agreement, with our mode lying ~16% lower than theirs. This is not unreasonable as typically calculated frequencies of molecules are accurate to ~10–15 %.[42]

The molecular-dynamics technique makes no assumptions about harmonic potentials and does not construct a dynamical matrix. Very harmonic systems are treated in exactly the same way as an anharmonic system. We show in Fig. 8 the results of a simulation of Si_3 at a very high temperature of ~2500 K. The simulation was of 3600 time steps and there is no indication that the molecule is about to disassociate. The motion is two dimensional and a snapshot of the motion is shown in Fig. 8(a). Within the chaotic motion, one sees a nearly sixfold pattern coming from a molecule containing only three atoms. This occurs because the atom that acts as the apex of the ground-state triangular molecule is being changed from one atom to the next. Also there is considerable classical tunnelinglike behavior of the apex atom through the center line connecting the other two atoms. The spectral density function $g(\omega)$ [Fig. 8(b)] has lost the three-peak structure of the harmonic system, and takes on a continuous spectrum. The frequencies depend on the amplitude of each of the "modes," and the amplitudes are continuously changing.

C. Si_4

The equilibrium configuration determined for Si_4 is a rhombus with side length 2.32 Å and a minor diagonal length of 2.39 Å (see Fig. 9). This compares very well with the results of Tománek and Schlüter,[14] who find a rhombus of side length 2.3 Å with a diagonal of 2.4 Å, and the results of Raghavachari and Logovinsky,[34] who also find a rhombus of side 2.30 Å and diagonal 2.40 Å. Other possible geometries include the square and tetrahedron. The square forms a metastable minimum-energy configuration with a side of length 2.32 Å, and the tetrahedron is metastable with minimum energy of side length 2.46 Å. The energies of these two structures are shown in Fig. 2 and are significantly higher in energy ($\simeq$0.61 eV/atom and $\simeq$0.53 eV/atom, respectively) than the rhombus.

The one-electron LDA energy eigenvalues are plotted as a function of the angle θ in Fig. 10(a). As in the case of Si_3, there is again a crossing of levels at the configuration of highest symmetry (in this case, $\theta=90°$). This causes a Jahn-Teller unstable system that shows up as a cusp in the total energy per atom also as a function of the angle θ in Fig. 10(b).

The vibrational spectrum of Si_4 is shown in Fig. 11 for a high and low kinetic temperature. There are $3n-6=6$ normal modes for Si_4 and the high-temperature spectrum shows the mode-softening characteristic of anharmonicity at higher-excitation levels. The normal coordinates that give rise to the various peaks in the spectrum are labeled according to symmetry. The lowest frequency mode is B_{3u}, which is the "bond-bending" mode in which alternate atoms move out of and into the plane of the paper (see Fig. 9). The other bond-bending mode is B_{2u} in which the atoms move in the plane of the paper. The frequencies found from an examination of Fig. 11 are compared with those of Raghavichari and Rohlfing[43] in Table I. Overall agreement with these far more sophisticated calculations is excellent.

D. Si_5

We find the equilibrium structure of Si_5 based on annealing and quenching to be a trigonal bipyramid, consisting of an equilateral triangle of side 3.02 Å capped on top and bottom, with the distance between the corners of the triangle and a cap being 2.30 Å (see Fig. 12). This structure agrees with those found by Tománek and Schlüter,[14] who find a trigonal bipyramid with the relevant lengths of 3.1 and 2.4 Å, and Raghavachari and Logovinsky,[34] who find the same structure with distances of 2.34 and 3.26 Å. Metastable structures for Si_5 include the pentagon and pyramid. The energetics of these structures are shown in Fig. 2, where the pentagon is $\simeq$1.00 eV/atom higher and the pyramid is $\simeq$0.13 eV/atom higher.

The vibrational spectral density function of Si_5 is shown in Fig. 13 at low temperature ($T\simeq50$ K). The five peaks in the figure come about because three frequencies are doubly degenerate, while an additional frequency is "accidentally" degenerate, and unresolvable.

E. Si_6

The equilibrium structure for Si_6 is that of a bipyramid, consisting of a square with side length 2.71 Å capped on top and bottom, with the distance between a vertex of the square and a cap being 2.36 Å (see Fig. 14). This structure agrees with that found by Tománek and Schlüter,[14] who also find such a "distorted octahedron" with the relevant distances of 2.6 and 2.3 Å. Raghavachari,[36] however, finds the ground-state structure to be an edge-capped trigonal bipyramid.

The energies of higher-energy metastable states for $n=6$ are shown in Fig. 2. Three structures investigated are structures higher in energy than the equilibrium structure: (1) a hexagon ($\simeq$1.22 eV/atom higher in energy), (2) the singly capped pentagon ($\simeq$0.50 eV/atom higher), and (3) a sixfold buckled ring which is a bulk fragment ($\simeq$0.22 eV/atom higher).

F. Si_7

The ground-state structure found for Si_7 is a bicapped pentagon, consisting of a pentagon of side length 2.46 Å capped on top and bottom, with the distance between a vertex of the pentagon and a cap being 2.45 Å (see Fig. 15). This is in agreement with Ballone *et al.*,[44] who also find a pentagonal bipyramid as the equilibrium structure. In addition, when just dynamical quenching is performed, the system invariably settles into a capped and strongly reconstructed version of the sixfold buckled ring. This metastable structure is relatively close in energy ($\simeq$0.16 eV/atom higher) to this equilibrium structure.

V. CONCLUSIONS

We have used an approximate electronic-structure method to determine the electronic and vibrational spectrum and the equilibrium structures of small Si clusters. The method used is a simplified *ab initio* tight-binding model based on local-density theory, and has no adjust-

able parameters. We have compared a number of our results with those of others and find substantial agreement with more rigorous calculations. Based on these results for small molecules, we are optimistic that the molecular-dynamics technique and energy functional used here will be applicable to bulk crystalline and amorphous covalent semiconductors and to semiconductor surfaces as well. The method is computationally fast and efficient, and should permit us to study covalent systems of previously incalculable complexity.

ACKNOWLEDGMENTS

Two of us (O.F.S. and D.J.N.) wish to thank the Office of Naval Research, U.S. Department of Defense, for their encouragement and support under Contract No. ONR-N00014-90-J-1304. Two others (D.A.D. and J.D.D.) wish to acknowledge support under Contracts No. AF-AFOSR-89-0063 and No. ONR-N00014-89-J-1136. We wish to thank Gary Adams and Stefan Klemm for their insights, helpful comments, and suggestions.

[1]E. Pearson, T. Takai, T. Halicioglu, and W. A. Tiller, J. Cryst. Growth 70, 33 (1984); T. Takai, T. Halicioglu, and W. A. Tiller, Scr. Metall. 19, 709 (1985).

[2]F. H. Stillinger and T. A. Weber, Phys. Rev. B 31, 5262 (1985).

[3]M. I. Biswas and D. R. Hamann, Phys. Rev. Lett. 55, 2001 (1985); R. Biswas and D. R. Hamann, Phys. Rev. B 36, 6434 (1987).

[4]D. W. Brenner and B. J. Garrison, Phys. Rev. B 34, 1304 (1986).

[5]J. Tersoff, Phys. Rev. Lett. 56, 632 (1986); Phys. Rev. B 37, 6991 (1988); 38, 9902 (1988).

[6]M. I. Baskes, Phys. Rev. Lett. 59, 2666 (1987); B. W. Dodson, Phys. Rev. B 35, 2795 (1987).

[7]J. Chelikowsky, J. C. Phillips, M. Kamal, and M. Strauss, Phys. Rev. Lett. 62, 292 (1989).

[8]G. Ackland, Phys. Rev. B 40, 10 351 (1989).

[9]A. Carlsson, P. A. Fedders, and C. W. Myles, Phys. Rev. B 41, 1247 (1990).

[10]K. E. Khor and S. Das Sarma, Phys. Rev. B 40, 1319 (1989).

[11]B. C. Bolding and H. C. Andersen, Phys. Rev. B (to be published).

[12]J. R. Chelikowsky, J. C. Phillips, M. Kamal, and M. Strauss, Phys. Rev. Lett. 62, 292 (1989); J. R. Chelikowsky and J. C. Phillips, *ibid.* 63, 1653 (1989).

[13]D. J. Chadi, Phys. Rev. B 29, 785 (1985); Phys. Rev. Lett. 59, 1691 (1987); G. X. Xian and D. J. Chadi, Phys. Rev. B 35, 1288 (1987).

[14]D. Tománek and M. A. Schlüter, Phys. Rev. B 36, 1208 (1987).

[15]D. G. Pettifor, A. J. Skinner, and R. A. Davies, in *Atomistic Simulation of Materials—Beyond Pair Potentials,* edited by V. Vitek and D. J. Srolovitz (Plenum, New York, 1989), p. 317; A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, J. Phys. C 21, 35 (1988).

[16]A. T. Paxton and A. P. Sutton, J. Phys. C 21, L481 (1988); A. T. Paxton, A. P. Sutton, C. M. M. Nex, *ibid.* 20, L263 (1987); A. T. Paxton, *Atomistic Simulation of Materials—Beyond Pair Potentials,* edited by V. Vitek and D. J. Srolovitz (Plenum, New York, 1989), p. 265.

[17]J. A. Majewski and P. Vogl, Phys. Rev. Lett. 57, 1366 (1986).

[18]J. R. Chelikowsky and R. Redwing, Solid State Commun. 64, 843 (1987).

[19]O. F. Sankey and R. E. Allen, Phys. Rev. B 33, 7164 (1986).

[20]R. E. Allen and M. Menon, Phys. Rev. B 33, 5611 (1986); M. Menon and R. E. Allen, *ibid.* 33, 7099 (1986); M. Menon and R. E. Allen, Superlatt. Microstruct. 3, 295 (1987); M. Menon and R. E. Allen, Solid State Commun. 64, 53 (1987).

[21]C. Z. Wang, C. T. Chan, and K. M. Ho, Phys. Rev. B 39, 8586 (1989).

[22]F. S. Khan and J. Q. Broughton, Phys. Rev. B 39, 3688 (1989).

[23]O. F. Sankey and D. J. Niklewski, Phys. Rev. B 40, 3979 (1989); in *Atomistic Simulation of Materials—Beyond Pair Potentials,* edited by V. Vitek and D. J. Srolovitz (Plenum, New York, 1989), p. 95.

[24]W. M. Foulkes and R. Haydock, Phys. Rev. B 39, 12 520 (1989).

[25]D. M. Ceperley and G. J. Adler, Phys. Rev. Lett. 45, 566 (1980). These numerical results have been parametrized by J. Perdew and A. Zunger, Phys. Rev. B 23, 5048 (1981).

[26]D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. 43, 1494 (1979).

[27]J. Harris, Phys. Rev. B 31, 1770 (1985).

[28]H. M. Polatoglou and M. Methfessel, Phys. Rev. B 37, 10 403 (1988).

[29]A. J. Read and R. J. Needs, J. Phys. Condens. Matter 1, 7565 (1989).

[30]C. W. Gear, Argonne National Laboratory Report No. ANL-7126, 1966 (unpublished); Y. H. Lee, Ph.D. thesis, Kent State University, 1986.

[31]S. Kirkpatrick, C. D. Gellatt, Jr., and M. P. Vecchi, Science 220, 671 (1983).

[32]M. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. 21, 1087 (1953).

[33]F. J. Harris, Proc. IEEE 66, 51 (1978).

[34]K. Raghavachari and V. Logovinsky, Phys. Rev. Lett. 55, 2853 (1985).

[35]The increase of kinetic energy due to the compact orbitals is much smaller for clusters and solids, due to hopping amongst the atoms, than it is for the isolated atom.

[36]K. Raghavachari, J. Chem. Phys. 84, 5672 (1986).

[37]D. E. Milligan and M. E. Jacox, J. Chem. Phys. 52, 2594 (1970), and references therein.

[38]J. E. Northrup, M. T. Yin, and M. L. Cohen, Phys. Rev. A 28, 1945 (1983).

[39]G. H. F. Diercksen, N. E. Gruner, J. Oddershele, and J. R. Sabin, Chem. Phys. Lett. 117, 29 (1985).

[40]R. S. Grev and H. F. S. Schaefer, Chem. Phys. Lett. 119, 111 (1985).

[41]K. Raghavachari, J. Chem. Phys. 83, 3520 (1985).

[42]J. A. Pople, H. B. Schlegel, R. Krishnan, D. J. Defrees, J. S. Brinkley, M. J. Frisch, and R. A. Whiteside, Int. J. Quantum Chem. Symp. 15, 269 (1981).

[43]K. Raghavachari and C. M. Rohlfing, J. Chem. Phys. 89, 2219 (1988).

[44]P. Ballone, W. Andreoni, R. Car, and M. Parrinello, Phys. Rev. Lett. 60, 271 (1988).

Generalized Möbius transforms for inverse problems

Shang Yuan Ren [1] and John D. Dow

Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287, USA [2]
and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

Received 19 July 1990; revised manuscript received 4 February 1991; accepted for publication 7 February 1991
Communicated by J.P. Vigier

An inverse formula is given for a new, generalized Möbius transform $G(x,y,\ldots)=\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\ldots F(n^{\alpha}x,m^{\beta}y,\ldots)$, namely $F(x,y,\ldots)=\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\ldots\mu(n)\mu(m)G(n^{\alpha}x,m^{\beta}y,\ldots)$. In the case of one dimension and $\alpha=-1$, this reduces to the Chen–Möbius transform which has been applied to the inverse black-body radiation problem. For $\alpha=+1$ this becomes the Hardy–Wright–Möbius transform.

Recently Chen [1] modified Möbius' number-theory transform [2] and showed how it could be used to solve a class of inverse problems in physics, including the problem of determining the temperature distribution of a black body from its radiation spectrum: given a quantity $F(x)$ that is related to a function $G(x)$ by

$$G(x)=\sum_{n=1}^{\infty}F(x/n)\,, \tag{1a}$$

the inverse function F is given by

$$F(x)=\sum_{n=1}^{\infty}\mu(n)\,G(x/n)\,, \tag{1b}$$

where $\mu(n)$ is the Möbius function: $\mu(n)=1$ if $n=1$, $\mu(n)=(-1)^r$ if n is the product of r distinct prime factors, and $\mu(n)=0$ otherwise [2]. For example, in the physics of black bodies, $G(\nu)$ may be related to the radiation spectrum and $F(T)$ may be related to the areal thermal distribution of the body [1]. As shown by Chen, this theorem is particularly useful for inverting data, for example, obtaining the unknown areal distribution of temperatures on a black body from its radiation spectrum.

Here (i) we combine Chen's result with another inverse Möbius transform from number theory [2][#1] to obtain a more general inverse formula, and (ii) we extend this generalized formula to two and higher dimensions. The new and more general inverse Möbius transform formula for one dimension is:

If we have

$$G(x)=\sum_{n=1}^{\infty}F(n^{\alpha}x) \tag{2a}$$

then it is the case that

$$F(x)=\sum_{n=1}^{\infty}\mu(n)G(n^{\alpha}x)\,. \tag{2b}$$

Hardy and Wright proved this theorem for $\alpha=1$ ("theorem 270" [2] [#1]) and Chen's Möbius theorem corresponds to $\alpha=-1$. We show that the theorem is valid for any real α, and hence there are entire classes of Möbius transforms, each corresponding to a different choice of α. Our proof uses "theorem 263" [2] for the Möbius function,

$$\sum_{n|k}\mu(n)=1,\quad \text{for } k=1,$$
$$=0,\quad \text{otherwise.} \tag{3}$$

Here the sum is over all the different divisors n of the integer k, including 1 and k. The proof of the gen-

[1] On leave from the University of Science and Technology of China, Hefei 230026, China.
[2] Present address.

[#1] In fact, due to eq. (31) of ref. [1], eq. (32) in ref. [1] is a direct result of theorem 270 of ref. [2].

0375-9601/91/$ 03.50 © 1991 – Elsevier Science Publishers B.V. (North-Holland)

eralized formula is obtained by evaluating

$$\sum_{n=1}^{\infty} \mu(n) G(n^{\alpha}x) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mu(n) F(m^{\alpha}n^{\alpha}x)$$

$$= \sum_{k=1}^{\infty} F(k^{\alpha}x) \sum_{n|k} \mu(n) = F(x) . \qquad (4)$$

In this derivation, we performed the sum over $k=mn$ and n, instead of over m and n; this reordering will converge if

$$\sum_{m=1} \sum_{n=1} |F(m^{\alpha}n^{\alpha}x)| = \sum_{k} d(k) |F(k^{\alpha}x)| \qquad (5)$$

converges, where $d(k)$ is the number of divisors of k [2]. Note that for $\alpha=0$ this convergence criterion is normally *not* met, and hence $\alpha=0$ is to be avoided. Furthermore the converse of eq. (2) is true if

$$\sum_{n=1} \sum_{m=1} |G(m^{\alpha}n^{\alpha}x)| = \sum_{k} d(k) |G(k^{\alpha}x)| \qquad (6)$$

converges.

Generalization of formula (2) to higher dimensions is straightforward; for example, in two dimensions for $F(x, y)$ of the form

$$G(x, y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} F(m^{\alpha}x, n^{\beta}y) , \qquad (7a)$$

we have

$$F(x, y) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \mu(k)\mu(l) G(k^{\alpha}x, l^{\beta}y) . \qquad (7b)$$

A sufficient condition for this inverse formula to be valid is the convergence of

$$\sum_{k} \sum_{m} \sum_{l} \sum_{n} |F(m^{\alpha}k^{\alpha}x, n^{\beta}l^{\beta}y)|$$

$$= \sum_{i} \sum_{j} d(i)d(j) |F(i^{\alpha}x, j^{\beta}y)| . \qquad (8)$$

The new inverse Möbius formula (2) can be applied to problems in mathematics and physics, and a convenient choice of the indices α, β, ... can be selected to best suit each problem.

The one-dimensional version leads to the new mathematical identity

$$\frac{1}{\zeta(p)} = \sum_{n=1}^{\infty} \mu(n) n^{-p} \quad \text{for } p>1 . \qquad (9)$$

where $\zeta(p)=\sum_{n=1}^{\infty} n^{-p}$ is the Riemann zeta function. It also leads to the identities

$$\frac{x^2}{x^2 \pm 1} = \sum_{n=1}^{\infty} \tfrac{1}{2}\mu(n) [(\pi x/n) f_{\pm}(\pi x/n) - 1] \qquad (10)$$

and

$$J_0(x) = \sum_{n=1}^{\infty} \mu(n) \left(-\frac{1}{2} + \frac{1}{nx} \right.$$

$$\left. + 2 \sum_{m=1}^{N} (n^2x^2 - 4m^2\pi^2)^{-1/2} \right) , \qquad (11)$$

where we have $2N\pi < nx < (2N+2)\pi$, J_0 is the cylindrical Bessel function, $f_+(x)=\coth(x)$, and $f_-(x)=\cot(x)$, and we have used the identities [3]

$$f_{\pm}(\pi x) = \frac{1}{\pi x} + \frac{2x}{\pi} \sum_{n=1}^{\infty} (x^2 \pm n^2)^{-1} ,$$

and

$$\sum_{n=1}^{\infty} J_0(nx) = -\frac{1}{2} + \frac{1}{x} + 2 \sum_{m=1}^{N} (x^2 - 4m^2\pi^2)^{-1/2}$$

for $2N\pi < x < (2N+2)\pi$.

There are undoubtedly many more uses for this inverse formalism in both mathematics and theoretical physics that will be uncovered as it is used more.

We are grateful to the U.S. Office of Naval Research and the Air Force Office of Scientific Research (Contract Nos. N00014-89-J-1136 and AFOSR-89-0063) for their generous support.

References

[1] N.Y. Chen, Phys. Rev. Lett. 64 (1990) 1193.
[2] G.H. Hardy and E.M. Wright, An introduction to the theory of numbers, 5th Ed. (Oxford Univ. Press, Oxford, 1979) pp. 235–237.
[3] I.S. Gradshtein and I.M. Ryzhik, Table of integrals, series and products (Academic Press, New York, 1980) pp. 46, 976.