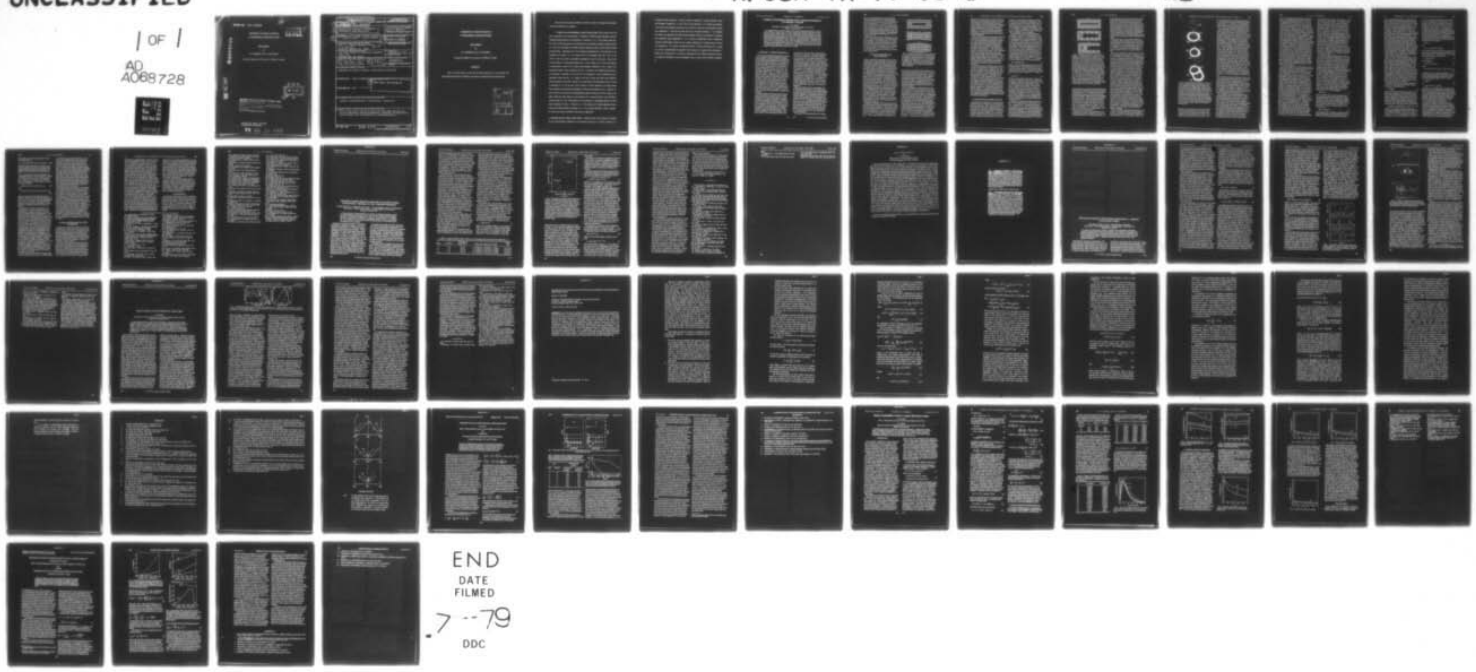
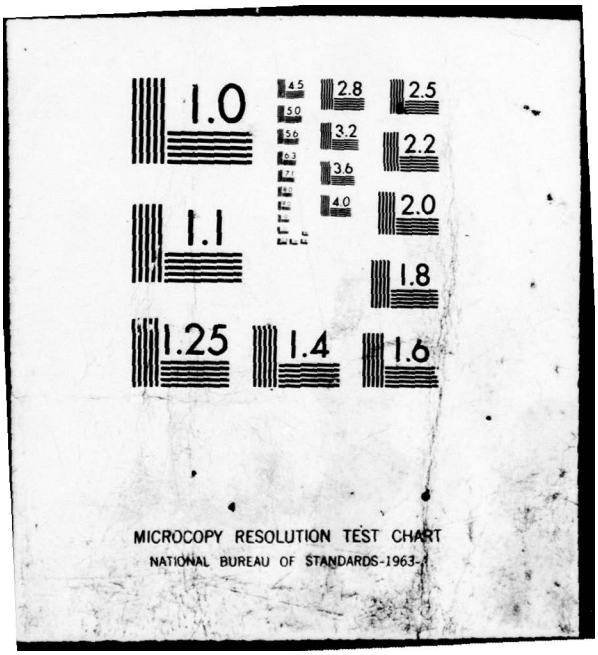
AD-A068 728 IBM THOMAS J WATSON RESEARCH CENTER YORKTOWN HEIGHTS N Y F/G 20/2
THEORETICAL STUDIES OF DEFECTS IN TETRAHEDRAL SEMICONDUCTORS.(U)
NOV 78 S T PANTELIDES, J A VAN VECHTEN F49620-77-C-0005
UNCLASSIFIED AFOSR-TR-79-0598 NL

1 OF 1
AD A068728



END
DATE FILMED
7-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 79-0598

LEVEL II

THEORETICAL STUDIES OF DEFECTS IN TETRAHEDRAL SEMICONDUCTORS

FINAL REPORT

by

S. T. Pantelides and J. A. Van Vechten

Principal Investigators for Contract No. F49620-77-C-0005

AD A068728

DDC FILE COPY

DDC
RECEIVED
MAY 18 1979
A

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

Approved for public release;
distribution unlimited.

79 05 18 047

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (18) (19) AFOSR-TR-79-0598	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) (6) Theoretical Studies of Defects in Tetrahedral Semiconductors,	5. TYPE OF REPORT & PERIOD COVERED Final Scientific Report 10/1/76 - 9/30/78
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) S. T. Pantelides, J. A. Van Vechten	8. CONTRACT OR GRANT NUMBER(s) (15) F49620-77-C-0005
9. PERFORMING ORGANIZATION NAME AND ADDRESS IBM T. J. Watson Research Center P. O. Box 218 Yorktown Heights, NY 10598 349 250	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bldg. 410 Bolling AFB, DC 20332	12. REPORT DATE November 30, 1978
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (11) 30 Nov 78 (12) 58 p.	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(9) Final rept. 1 Oct 76-30 Sep 78,

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Defects, Semiconductors, Dislocations, Impurities

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

This is the final report on work that has been carried out on the electronic and thermochemical properties of impurities, point defects and dislocations in semiconductors.

DD FORM 1 JAN 73 1473 349 250 UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

THEORETICAL STUDIES OF DEFECTS IN TETRAHEDRAL SEMICONDUCTORS

FINAL REPORT

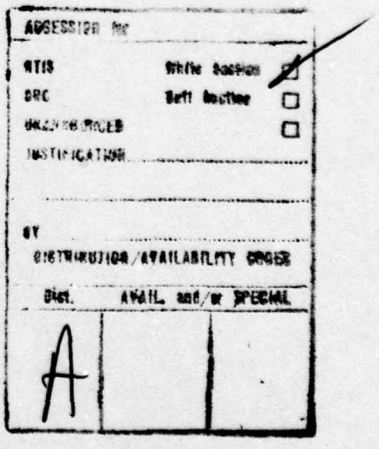
by

S. T. Pantelides and J. A. Van Vechten

Principal Investigators for Contract No. F49620-77-C-0005

ABSTRACT

This is the final report on work that has been carried out on the electronic and thermochemical properties of impurities, point defects and dislocations in semiconductors.



Work done under contract F49620-77-C-0005 on defects in tetrahedral semiconductors may be divided in two categories:

1. **Studies of the thermochemistry of point and line defects:** Work carried out in this area is described in two journal articles: "Formation of Interstitial-type Dislocation Loops in Tetrahedral Semiconductors by Precipitation of Vacancies", by J. A. Van Vechten, Phys. Rev. B 17, 3197 (1978) (also given as an invited talk at the APS March meeting in Chicago, March 21, 1979); and "Threshold for Optically Induced Dislocation Glide in GaAs-AlGaAs Double Heterostructures: Degradation via a New Cooperative Phenomenon?", by B. Monemar, R. M. Potemski, M. B. Small, J. A. Van Vechten and G. R. Woolhouse, Phys. Rev. Lett. 41, 260 (1978). These two articles are included as Appendices A and B in this report. More of this work will appear in a forthcoming chapter by J. A. Van Vechten in Vol. III of the *Handbook of Semiconductors*, edited by S. P. Keller (North Holland, Amsterdam, 1979 or 1980), in an invited talk entitled "GaAs Is Different from Si", to be given at the AIME Electronic Materials Conference, in Boulder, Co., June 24-29, 1979 (Appendix C), and in publications still in preparation which deal with: (a) variation in the rate of anion and cation atomic diffusion with stoichiometry of host III-V crystals; (b) site preference and compensation ratio of Group IV impurities (C, Si, Ge and Sn) in III-V crystals; (c) defect migration in Si during pulsed laser annealing and nonthermal mechanisms by which this annealing can occur. Subjects (a) and (b) above are treated in the aforementioned handbook chapter and will be treated in the EMC invited talk while subject (c) was treated in two contributed talks at the APS March meeting, March 30, 1979: "Mechanism for the Production of a Metastable Plasma in Pulsed Laser Annealing" by Ellen J. Yoffa and J. A. Van Vechten; and "Defect Migration During Pulsed Laser-induced Plasma Annealing" by J. A. Van Vechten and Ellen J. Yoffa. Summaries of these two talks are included in this report as Appendix D.

2. **Electronic-structure studies of point defects:** During the term of this contract we reported the first self-consistent calculation of the electronic structure of a neutral vacancy in Si

assuming no lattice distortions. The main results are contained in a paper by Bernholc, Lipari and Pantelides (Appendix E). In other work on deep impurities, T. N. Morgan has investigated the capture and excitation processes at the O center in GaP and has reinterpreted existing data (Appendix F). Work has also been carried out on shallow impurities. S. T. Pantelides has derived and discussed new effective-mass equations which are appropriate for materials with multivalley bands (Appendix G). N. O. Lipari, in collaboration with A. Baldereschi (Lausanne, Switzerland), carried out new, accurate calculations of acceptor spectra in Si and Ge (Appendix H). N. O. Lipari, in collaboration with D. L. Dexter (Rochester, N.Y.), carried out a study of polarizabilities of impurities in the presence of a magnetic field (Appendix I). Finally, and J. Pollman and N. O. Lipari, in collaboration with H. Büttner (Bayreuth, Germany), studied the quenching of exciton diamagnetic shifts in polar, layered materials (Appendix J).

Appendix A

PHYSICAL REVIEW B VOLUME 17, NUMBER 8 15 APRIL 1978

Formation of interstitial-type dislocation loops in tetrahedral semiconductors by precipitation of vacancies

J. A. Van Vechten
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 14 November 1977)

It is hypothesized that when vacancies precipitate to form voids in tetrahedral semiconductors, e.g., Si and GaAs, reconstruction reactions occur on the internal surfaces of these voids in the same manner as they are observed to occur on external surfaces of the same crystallographic orientation. Previously, it has been concluded that many of the various reconstruction reactions observed on various semiconductor surfaces produce hillocks by expelling a portion of the atoms from the unreconstructed (ideal) surface to migrate in a reaction front across the surface. From these two lemmas, it is here concluded that the corresponding waves of atoms driven by reconstruction on internal surfaces will precipitate into pillars of crystal-line material within the void and produce dislocation loops at which the lattice planes bow away from the center. Such dislocations are conventionally denoted "interstitial type." This mechanism may explain several observations of "interstitial type" dislocations in semiconductors, including Si and GaAs, for which there is much evidence for vacancies and no other evidence for self-interstitials.

I. INTRODUCTION: VACANCIES OR INTERSTITIALS?

Surely the most fundamental question to resolve when one begins the study of deep point defects and line defects in semiconductors is whether the dominant native defect involved in the formation and motion of the observed defects is the vacancy or the self-interstitial. Despite this fact and the intensive study of deep point and of line defects in several semiconductors over many years, this question is still being debated even for the most thoroughly studied cases. For example, in the case of Si, Refs. 1–6 refer to a few of the recent papers in which the respective authors interpret diffusion and formation of dislocations and stacking faults in terms of vacancies, while Refs. 7–12 refer to a few of the recent papers in which these phenomena are described in terms of Si self-interstitials. A similar situation exists in GaAs-AlGaAs, where some authors explain the formation of dark-line defects in heterostructure lasers entirely in terms of vacancies,[13,14] while others[15,16] invoke self-interstitials to explain their formation.

Those who believe that vacancies are the dominant native defects in group IV and III-V semiconductors do so for several reasons: (i) Only vacancies and vacancy complexes are identified by electron paramagnetic resonance in these materials even when they have been subject to electron irradiation and one may be certain that vacancy-interstitial (Frenkel) pairs were produced.[17] The absence of self-interstitials in irradiated samples has generally been ascribed to a rapid athermal migration of self-interstitials,[18] even at 2 K, and to a very large heat of formation of interstitials,[19] which provides a large driving force for their removal. (ii) Several impurities, such as Zn in GaAs,[20] are observed to diffuse as interstitials but to occupy predominantly substitutional sites. Therefore, there must be vacant lattice sites for them to occupy. (iii) Diffusion and related phenomena are successfully described by ascribing multiple ionization levels to the native defect which are the same as those observed for the vacancy in low-temperature irradiation experiments.[4,21,22] (iv) Theoretical estimates of the heat of formation of vacancies correspond well with those of the native defect observed in quenching and diffusion experiments.[21–25] As the threshold energies for Frenkel pairs in irradiation experiments are quite high, the energy to be ascribed to the heat of formation of the interstitial is much higher than that of the vacancy.[19] This is also in agreement with simple theoretical arguments.[19]

Those who believe that self-interstitials are the dominant native defects in these same group IV and III-V semiconductors do so mainly because the great majority of dislocation loops and stacking faults observed by transmission electron microscopy (TEM) in these materials have been determined to be of extrinsic or "interstitial" character.[7–12,15,16,26,27] Although four mechanisms by which interstitial-type dislocations may be formed without any self-interstitials being present are well established[1] (and discussed in Sec. II), these authors feel that in several cases these four alternative mechanisms do not obtain and that the only remaining mechanism is the precipitation of self-interstitials.

The purpose of the present paper is to analyze

 © 1978 The American Physical Society

(Sec. III) the mechanism by which dislocations should be expected to form if indeed vacancies are much more numerous than self-interstitials in these semiconductors. It is concluded that, because reconstruction should occur on interior surfaces as it is known to occur on exterior surfaces, the precipitation of vacancies will cause the formation of interstitial-type dislocations in most, but not all, experimental situations. The possibilities to verify this contention experimentally are discussed in Sec. IV.

The author is aware that his proposal is at odds with the conventional wisdom of the TEM-dislocation community. He maintains that it is a reasonable conclusion from what is generally accepted in high-vacuum surface science and the most attractive resolution of the dilemma of the evidence for vacancies and for self-interstitials where it is clear that they cannot both be present in significant concentrations.

II. WHAT AN "INTERSTITIAL-TYPE" DISLOCATION LOOP IS

When an electron microscopist says that a dislocation loop is "interstitial-type", he means that the host-lattice planes bow away from the center of the loop.[28] When he says the dislocation loop is "vacancy type", he means that the host-lattice planes bow toward the center of the loop (see Fig. 1). This notation was popularized by Friedel,[28] who observed that if a number of vacancies were to condense to form a planar void of sufficient size, then it would be energetically favorable for the material to remove the void by rewelding the surface on either side together. This would remove the surface energy of the void over its entire extent at the cost of deforming the lattice inward only around the perimeter. Thus, it is clear that for a sufficiently large number of vacancies in a cluster, the rewelding will reduce the total energy of the solid. It should also be clear that there will always be an activation barrier to be overcome before this rewelding can occur, because the opposite planes must be brought together. Furthermore, if foreign atoms, such as H, He, or Ar, in the presence of which crystals are often grown, precipitate into the void, they may prevent the rewelding from occuring.

It should be emphasized that the electron microscope detects the bowing of the lattice planes rather than the presence or absence of host atoms. In the absence of lattice distortion, a vacancy would have a scattering potential for electrons differing from that of an interstitial of the same atomic species only by a sign.[29] (The potential being that of the atom.) As the image intensity is proportional to

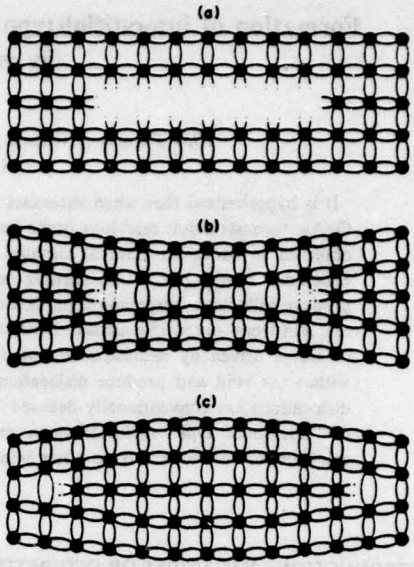


FIG. 1. Schematic representation of the conventional view of the conversion of a cluster of vacancies (a) into a "vacancy-type" dislocation loop (b) by a rewelding of the surfaces on either side of the void. (These figures are adapted from Friedel, Ref. 28). In (c) we see an "interstitial-type" dislocation loop at which the host-lattice planes bow away from the center of the loop as would be expected if, for example, self-interstitials were to precipitate.

the square of the electronic wave function, the image contrast produced by vacancies is the same as that produced by interstitial atoms. While it is true that the electron beam is less attenuated in passing through vacancies than in passing through interstitials, the difference of one atomic layer more or less in a sample typically 1 μm thick is generally not detectable and would be masked by steps and adsorbed layers on the surface of the sample.

Now, it has long been recognized that the fact that a dislocation loop is determined to be "interstitial-type" does not imply that it was formed by a precipitation of self-interstitials.[1] Indeed, four alternative processes have been firmly established in the semiconductor literature. One alternative process, which has been widely observed in Si, is the conversion of interstitially diffusing impurities (typically Cu, Au, or Li) into substitutional impurities by absorption into interstitial loops.[30-33] A second process is the emission of vacancies by an interstitial loop initially present.[14,34] The interstitial loop grows by one atom for each vacancy that is emitted from its boundary and diffuses away. In a third mechanism,[35] a precipitate particle (e.g., Cu or Fe in Si[5,6]) is formed at an edge dislocation. Compressive stress inside the pre-

cipitate is relieved by flow of vacancies from the dislocation to the precipitate particle. This flow may take place by dislocation pipe diffusion or volume diffusion and is accompanied by dislocation climb. The sense of the climb is such that an interstitial loop will grow and a vacancy loop will shrink.[5,6,35,36] A fourth mechanism is prismatic punching,[37-39] which is only rarely observed in semiconductor crystals that have not been mechanically damaged.[39]

There is a fifth alternative mechanism which has not been much discussed in the literature. Because these semiconductors expand on solidification, any inclusion of the liquid phase that may be trapped in the solid during crystal growth will tend to punch out interstitial-type dislocations when it does freeze.[40,41] One expects inclusions to occur when the crystal is growing by means of reentrant growth steps. When the growth steps are more than a few atomic layers high (on covalent crystals they are often several hundreds of layers high), one would expect[42,43] to find reentrant growth steps even if the treads of the growing surfaces of the crystal do not reconstruct. The tendency to form reentrant growth steps would be much increased if these treads do reconstruct.[41]

This fifth alternative mechanism has the attraction that it would provide a mechanism for the nucleation of dislocations in crystals of such high purity that impurity induced heterogeneous nucleation is not effective. It has been noted that the homogeneous nucleation of a dislocation from a concentration of single vacancies that was at equilibrium with the surface during crystal growth is normally an extremely unlikely event.[44,45] If one assumes the self-interstitial is the dominant native lattice defect, the same considerations as presented in Refs. 44 and 45 show that the homogeneous nucleation of an interstitial dislocation is equally unlikely. It is usually concluded that almost all observed dislocations were heterogeneously nucleated by impurity precipitates, but many semiconductors are grown so pure that this assumption should be questioned.

III. RECONSTRUCTION ON INTERNAL SURFACES

Reconstruction reactions are first-order phase transitions that are observed (in high vacuum) on most low index [e.g., (100), (110), and (111)] surfaces of group IV and III-V semiconductors by low-energy-electron diffraction (LEED), photoemission (PE), and other techniques.[46-53] It is generally agreed that these reactions occur because covalently bonded crystals can lower their free energy by removing the dangling bonds on the ideal surface (i.e., that obtained by simply truncating the bulk crystal structure) by switching to a new crystal structure for the surface layer of atoms. Whereas the atoms in the bulk of these semiconductors engage in s-p^3 hybridized bonding, those on the surface may engage in s-p^2 hybridized bonding or p bonding or some other type of bonding in order to minimize their free energy.

The first lemma of the present hypothesis is that similar reconstruction reactions will occur on internal surfaces of a crystal that are formed by the precipitation of vacancies at voids, vacancy complexes, or dislocations, however nucleated. As long as the opposite sides of such cavities remain sufficiently far apart that the interaction between them is slight, these reconstructions will surely obtain. In order to gauge how great a separation between surfaces is sufficient, we note that surface energies result from the spilling of electron density into vacuum, where it cannot be compensated by atomic cores, which occurs because the electronic wave functions are not terminated abruptly at a surface.[23,25,54-57] This spilling of the charge density will be approximately

$$\rho(x) = \rho(0)\exp(-k_s x), \tag{1}$$

where k_s is the linearized Thomas-Fermi screening wave number. The important point is that, because the valence electron density of these semiconductors is very high, k_s is large and the spilled charge density falls off rapidly.[23,25,57] Indeed, for Si

$$k_s R_w = 3.5, \tag{2}$$

where R_w is the Wigner-Seitz radius. Therefore, as long as the opposite surfaces of an internal cavity remain at least one atomic spacing apart, there is negligible interaction between them. In fact, $2R_w$ is a conservative estimate of the distance between opposite surfaces because atoms move away from the vacuum and toward the bulk immediately upon formation of the surface, before any further reconstruction can occur, and also around single vacancies. This is simply because charge is redistributed from the dangling bonds of these surfaces to the back bonds causing them to shorten.[23,25,52,58]

It should be noted that a minimum extent of surface area is required before reconstruction can occur. Empirically,[59] the minimal dimension D_c is typically 6–8 nm. The rate at which the reconstruction reaction occurs increases with increasing extent of the surface.[51,60] This variation has been shown to fit the hypothesis that once the reconstruction reaction starts on any tread of a stepped surface, it propagates rapidly across the entire tread to the risers.[60] This fact has been explained[41,60,61] by the hy-

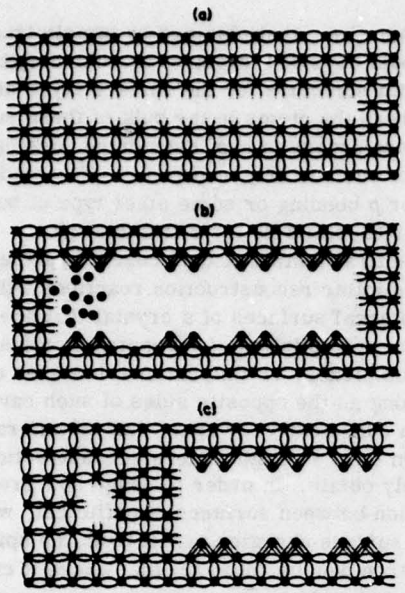


FIG. 2. Schematic representation of the author's proposal of a mechanism for the conversion of a planar precipitate of vacancies into an "interstitial-type" dislocation loop by the reconstruction of the surfaces of the cavity. In (a) we again see the cavity formed by the precipitation of vacancies, as in Fig. 1(a). In (b) both surfaces of the cavity are undergoing the surface reconstruction reaction which drives a wave of atoms, that are expelled from the surfaces by the reaction, into the extremities of the cavity. These extra atoms in the extremities are thought to be in a state similar to that of the liquid phase. The reconstructed surface is a phase of lower energy, lower density, and different bonding. In (c) the atoms that were expelled by the reconstruction have recrystallized to the bulk crystal structure. As they do so, they expand, as does the liquid phase, so that an interstitial-type dislocation loop is formed, the surfaces of the original cavity are forced apart by one lattice spacing, and a pillar of good crystalline material is left to stabilize the cavity and the dislocation.

pothesis that reconstructed surface phases generally contain fewer atoms than the ideal (unreconstructed) surface phase,[46,47,52,53] so that the reaction releases atoms and drives them in a reaction front that perturbs the metastable ideal phase and thus propagates the reaction further. When this occurs on an exterior surface of a crystal, one would expect the waves of released atoms to collide and produce hillocks. This effect has evidently been observed; the hillocks on Si being typically 40 nm high.[62] It provides an explanation for the continuous source of steps, that is required for crystal growth, in the absence of screw dislocations.[41,61]

Although the structure and coordination of the released atoms as they propagate in the wave has not been determined, it is here assumed that, as in the liquid phase, these atoms are more densely packed than in the tetrahedrally coordinated solid phase. It would seem that if the atoms were tetrahedrally coordinated in the covalent solid phase, they would not be able to migrate across the surface as required to explain the kinetics of the reconstruction on stepped surfaces.[51,60] In any case, the tetrahedral phase is the least dense of all known condensed phases of these compositions. This point is important because it implies that these atoms will expand, as does a liquid drop, when they recrystallize to the bulk solid phase.

The second lemma of the present hypothesis is the contention just introduced that reconstruction releases atoms from the ideal surface, that these will migrate in a wave across the flat surface until stopped for some reason whereupon they solidify to the bulk crystal structure and expand to produce a hillock, pillar, or similar surface defect.

Now let us consider what should happen as vacancies precipitate in the interior of a covalent crystal. It is well known that there is a binding energy of order 1 eV between vacancies[57,63,64] so that they will form multiple vacancy clusters (small voids) as the crystal cools from its growth temperature. Suppose that an approximately disk-shaped cluster (cavity) has grown to a size sufficient to have a low index surface greater than the critical extent, 6–8 nm, required for reconstruction[59] without rewelding (see Fig. 2). When the reconstruction occurs, it produces a wave of atoms which propagates through the void across the interior surface. The wave will stop when it collides with another wave, the end of the cavity, or perhaps a patch of reconstructed surface. Thereupon, the atoms solidify into the bulk solid phase. Obviously, an internal hillock or pillar produced inside the cavity of precipitated vacancies could not be as high as the 40-nm ones on exterior surfaces. One might instead suppose these atoms recrystallize as would a drop of the liquid, expanding as it solidifies and tending to drive the opposite surfaces of the cavity apart by one or a few atomic spacings. If this does occur, then an "interstitial-type" dislocation loop is produced in which the "extra plane" of lattice sites contains regions where host atoms are present and regions where they are not (see Fig. 3). Where the atoms are not present, the opposite surfaces are now reconstructed, so that there are no dangling bonds and there is a much larger activation barrier against rewelding.

Several questions regarding this hypothesis come to mind: (i) Is the proportion of pillar regions to void regions likely to be sufficient to stabilize the dislocation loop? (ii) Is the Burgers vector of the

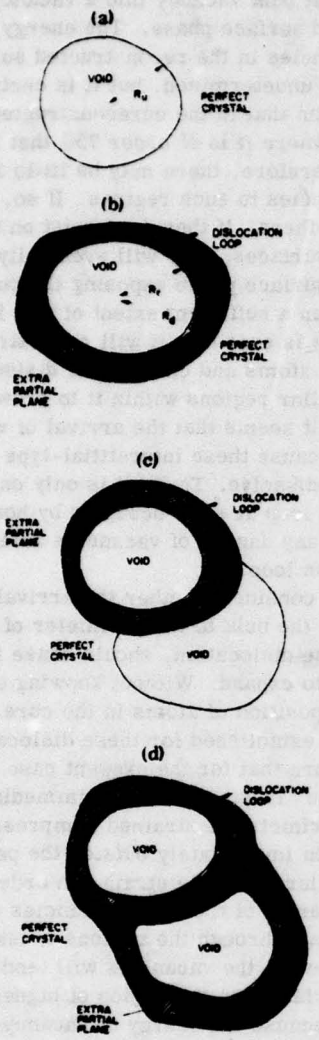


FIG. 3. Proposed process, as in Fig. 2, viewed normal to the plane of the cavity and of the loop. In (a) we see the original disk-shaped cavity produced by the vacancies prior to its reconstruction. In (b) the cavity has reconstructed and formed a pillar between the dislocation loop and the reconstructed cavity from the atoms expelled by the reconstruction reaction. In (c) more vacancies have precipitated out of the bulk to form a new unreconstructed void at the perimeter of the first dislocation loop. The vacancies are attracted to this region by the tensile strain field there. They do not annihilate the pillar because of the compressive strain field in that region. In (d) the new cavity has undergone reconstruction producing more pillar region and increasing the dislocation loop.

dislocation necessarily the same as the orientation of the cavity surface which reconstructed? (iii) How about the problem of nucleating such a dislocation in the first place? (iv) If such a dislocation were nucleated, would it grow? (v) Is the energy released by reconstructing the internal surface greater than that consumed producing the dislocation? (vi) Under what circumstances would vacancy type dislocations be expected instead? Let us consider these questions in order.

(i) The generally accepted, Lander-Morrison model for the stable (7×7) reconstruction of the (111) surfaces of Si contains 26% fewer atoms than the ideal surface.[46,47] As there are two such surfaces to the vacancy produced cavity, 52 atoms would be released for every 100 vacancies precipitating into a disk-shaped cavity one atomic spacing thick. Let us suppose for the moment that the Burgers vector of the dislocation is the same as the orientation of the reconstructed surface. If the host lattice is forced apart one atomic layer to form a dislocation loop as these atoms solidify, then 26% of the extra plane of lattice sites would be occupied, i.e., 26% of the region would be pillar and 74% void. If the initial cavity is a disk two atomic spacings thick, 26 atoms would be released for every 100 vacancies precipitating and, if the host lattice is forced apart one spacing (so there are 3 layers where there were 2) to form the dislocation, then 15.3% of the loop would be pillar. If the initial cavity is a disk three atomic spacings thick, still 13.8% of the resultant dislocation loop would be pillar. This case is typical of the stable reconstructions that occur on the low index surface of most semiconductors.[52,53] It would seem that such proportions of pillar to void would be adequate to stabilize the loop.

(ii) It seems clear that the Burgers vector of the dislocation need not be the same as the orientation of the cavity surface which reconstructs. If a liquid inclusion were to solidify and to punch out dislocations, the resultant Burgers vectors would be determined by properties of the host as well as by the shape of the inclusion. Moreover, after the dislocation is nucleated, its Burgers vector may be rotated by low-energy processes,[65] such as reaction with Shockley partial dislocations.

(iii) In addition to the dislocations that may be nucleated by the action of impurities,[30-39] some interstitial-type dislocations would be nucleated in pure semiconductor crystals grown from a melt, which is more dense,[66] by the trapping of liquid inclusions by reentrant (overhanging) growth steps that are many atoms high.[41-43] Moreover, it is here proposed that the precipitation of vacancies-interior reconstruction mechanism will itself nucleate interstitial-type dislocations due to the expansion of the released atoms as they recrystallize, and that it will do so without requiring that the vacancy cluster first assume the unlikely shape of a disk one atomic layer

thick.[44,45] Suppose that the vacancies in Si precipitate to an octahedral-shaped cavity, which is the equilibrium shape of a cavity in Si. The dimensions of the (111) surfaces of this cavity exceed the minimal dimension for reconstruction[59] (about 6 nm) when about 5.1×10^3 vacancies precipitate into it. There will then be 1.1×10^3 atoms on the 8 surfaces bounding this cavity. If all surfaces reconstruct, 2.9×10^2 atoms will be released in the wave of the reaction front. Suppose that, before recrystallization, these atoms form a single drop in one corner of the octahedron so as to minimize their surface energy. It seems clear that recrystallization will begin where the drop contacts the surfaces of the cavity and will next occur on the free surface of the drop. As with the freezing of ice cubes, the interior of the drop will be the last to recrystallize. If the drop does expand about 15% in volume when it recrystallizes, as does the liquid phase,[66] it will stress the surrounding lattice. One of the ways in which this stress can be relieved is by the formation of an interstitial dislocation loop sufficient to contain about 30 atoms in its extra partial plane. Obviously this configuration would not be a state of minimum energy, but it is well known that dislocations are not equilibrium defects; they would never exist were the crystal at equilibrium.[67]

(iv) We now suppose that an interstitial-type dislocation loop has been nucleated by one of the above mechanisms and that vacancies diffuse to it. It is commonly stated that vacancies will annihilate such a dislocation. However, it is well known that impurity atoms form precipitates on and about dislocations as the crystal cools and their concentrations become supersaturated. Vacancies will generally not annihilate such precipitates because to do so would require redissolving the impurity atom into the solid from which it had precipitated. Vacancies should not be expected to annihilate the dislocation produced by or to accommodate such precipitates either. In those central portions of the loop where the extra partial plane is composed of host atoms, the arrival of vacancies from the bulk will not remove any lattice sites from the dislocation, i.e., annihilate it, unless the vacancies nucleate a disk-shaped cavity and reweld the opposing surfaces, which, as noted above, is a very unlikely event.[44,45] Evidently, there would be little attraction of vacancies to such regions of the loop and they would diffuse on through the bulk. In those central portions of an "interstitial-type" dislocation loop formed by the vacancy precipitation-surface reconstruction mechanism where the "extra partial plane" is in fact a cavity with reconstructed surfaces, the arrival of a vacancy from the bulk would merely convert that bulk vacancy into a vacancy in the reconstructed surface phase. The energy of formation of vacancies in the reconstructed surface phase is still undetermined, but it is certain to be greater than that in the unreconstructed surface phase, where it is of order 75% that in the bulk.[41,60] Therefore, there may be little attraction for vacancies to such regions. If so, they may not stay there. If they do persist on the reconstructed surfaces, they will eventually remove the layer of surface phase exposing the bulk phase beneath. When a sufficient extent of this ideal surface phase is exposed, it will reconstruct releasing more atoms and causing the dislocation and/or the pillar regions within it to grow. In either case, it seems that the arrival of vacancies will not cause these interstitial-type dislocation loops to dissolve. Thus, it is only on the perimeter of the loop at sites occupied by host atoms that there is any danger of vacancies annihilating the dislocation loop.

Let us now consider whether the arrival of vacancies from the bulk to the perimeter of the loop, the core of the dislocation, should cause the loop to shrink or to expand. Without knowing anything about the disposition of atoms in the core, which is not firmly established for these dislocations, we may be sure that for the present case of "interstitial-type" loops, the region immediately inside the perimeter is strained compressively and that the region immediately outside the perimeter is subject to large tensile strain. In order to reach the interior of the loop, vacancies would have to migrate through the regions of tensile strain. However, the vacancies will tend to remain and to cluster in the region of highest tensile strain because the energy of vacancy formation $\Delta H_f(V)$ is a decreasing function of the interatomic spacing r, and so is less in regions of tensile strain. This may be understood by using the approximation of the single vacancy as a macroscopic cavity[23,25,57,68] and estimating

$$\Delta H_f(V) = A\sigma, \tag{3}$$

where A is the surface area of the cavity and σ is the energy per unit area of the cavity surface. (Because the valence electron density is so high, as noted at Eq. (2), this is a rather good approximation in semiconductors,[23,25,57] although it is not very accurate for many metals.[68]) As A is proportional to r^2 while σ is proportional[56] to $r^{-5/2}$, Eq. (3) implies

$$\Delta H_f(V) = \text{const} r^{-1/2}. \tag{4}$$

One may also wish to consider the effect of breaking the covalent bonds. Several authors[69-71] have analyzed the bond bending and bond stretching

force constants between atoms. From these analyses, and from the observed signs of the third-order elastic constants, it is clear that the interatomic forces are decreasing functions of r. (One might also consider the chemical trends in the commonly accepted experimental values[23,25,57,72] for $\Delta H_f(V)$, which are greater for C than for Si and greater for Si than for Ge, etc.) Of course, some of the vacancies will pass through the regions of tensile strain to the interior perimeter of the loop despite this repulsion due to the strain field there. But before the compressive stress can be removed and the energy of the crystal reduced (as assumed by those who assert that vacancies will annihilate interstitial-type dislocations), it would be necessary to cluster a sufficient number of vacancies in the interior perimeter so that part of the extra partial plane there could be removed by the mechanism of rewelding of the surfaces of a disk-shaped cavity. Evidently, the probability of such an event in this case is even less than in regions without compressive strain.

Thus, we come to the conclusion that as vacancies arrive in the vicinity of the dislocation loop they will cluster predominately in the regions of highest tensile strain outside the loop. As they do so, they will produce extended cavities. The shape of these cavities will be influenced by the strain field outside the dislocation. While they will probably not be disks of thickness one atomic layer, they will probably also not have the equilibrium shape of cavities in the host lattice either. Consequently, when these cavities reach the critical extent, they will reconstruct releasing more atoms that will cause the dislocation to grow so that the process may be repeated until the supersaturation of vacancies is depleted.

(v) Referring to Fig. 3(a), we may estimate the energy of an initial approximately disk-shaped cavity before reconstruction as

$$E_1 = 2\pi R_u^2 \sigma_u, \tag{5}$$

where R_u is the radius of this cavity and σ_u is the energy per unit area of the unreconstructed surface. When reconstruction occurs, the surface energy is reduced to σ_r and the radius of the void region is reduced to R_r because a fraction of the atoms released from the surface fill up part of the cavity. The energy of the cavity plus dislocation loop defect in Fig. 3(b) may be estimated[73] as

$$E_2 = 2\pi R_r^2 \sigma_r + \pi \mu R_d b^2 [\ln(R_d/b) - 1]/2(1-\nu), \tag{6}$$

where R_d is the radius of the dislocation loop, μ is the shear modulus of the host crystal (in terms of the elastic constants,[74] $\mu = C_{44} - H/5$, where $H = 2C_{44} + C_{12} - C_{11}$), b is the Burgers vector of the dislocation, and ν is Poisson's ratio. Clearly, $E_1 > E_2$ for R_u greater than some critical value, R_{uc} or for μ less than a critical value, μ_c. For the case of Si at room temperature, the shear modulus is[70] $\mu(\mathrm{RT}) = 6.81 \times 10^{11}$ erg/cm^3, while[75] $b = 0.384$ nm, and $\nu = 0.215$. Also, for Si(111) surfaces,[25] $\sigma_u = 904$ erg/cm^2. The value of σ_r is not known empirically but can be estimated[25] by noting that the reconstructed surface is about 26% less dense than the unreconstructed surface and is free of dangling bonds. This estimate is $\sigma_r = 613$ erg/cm^2. If we further suppose that the pillar regions discussed in (i) lie entirely within the original cavity, we estimate

$$R_d = R_u, \tag{7}$$

and

$$R_r = (1.0 - 0.26)^{1/2} R_u = 0.86 R_u. \tag{8}$$

Let us further assume that the cavity reconstructs as soon as it attains the minimum extent,[59] $D_c = 6$–8 nm, necessary for reconstruction. Then, taking the minimum empirical value for D_c,

$$R_c = \tfrac{1}{2} D_c = 3 \text{ nm}, \tag{9}$$

one calculates,

$$E_1 = 5.11 \times 10^{-10} \text{ erg}, \tag{10}$$

while

$$E_2 = 2.56 \times 10^{-10} + 2.02 \times 10^{-10} = 4.58 \times 10^{-10} \text{ erg}. \tag{11}$$

Therefore, with Si at room temperature, any cavity that is large enough to reconstruct can do so with sufficient energy to produce the dislocation loop indicated in Figs. 2 and 3. Thus,

$$R_{uc}(\mathrm{Si}) = 0,$$

and

$$\mu_c(\mathrm{Si}) = 8.6 \times 10^{11} \text{ erg/cm}^3. \tag{12}$$

However, as this minimal excess energy is only about 10% of E_1, the formation of the dislocation may be an improbable event in such cases. If the cavity reconstructs without producing a dislocation, the defect will produce little lattice strain and probably would not be detected by TEM unless it becomes decorated with impurities, in which case it would appear as a precipitate.

As one goes to higher temperatures, the temperature variation of μ must be considered. The author has found empirical values of $d\mu/dT$ only for room temperature and below.[76] However, it can be calculated from theories relating the elastic constants to the electronic structure[71] and a knowledge of the temperature dependence of the electronic structure.[77,78] (This theory fits the room-temperature temperature dependence to

about 20%.) This leads to the estimate for Si at its melting point T_m,

$$\mu(\mathrm{Si}, T_m) = 0.8\mu(\mathrm{Si}, \mathrm{RT}) = 6.45 \times 10^{11} \text{ erg/cm}^3. \quad (13)$$

Neglecting thermal expansion of b and any temperature variation of σ_u and σ_r, this leads to the estimate that E_1 exceeds E_2 by about 20% for the minimum sized cavity that could reconstruct at T_m.

In the case of GaAs at room temperature, the shear modulus is[70] $\mu = 4.85 \times 10^{11}$ erg/cm^3, while[79] $\nu = 0.29$ and $b = 0.399$ nm, and[25] $\sigma_u = 812$ erg/cm^3 and $\sigma_r = 495$ erg/cm^3. With $R_u = 3$ nm, the minimum value possible for reconstruction, a calculation as above leads to

$$E_1(\mathrm{GaAs}, \mathrm{RT}, 3 \text{ nm}) = 4.59 \times 10^{-10} \text{ erg}, \quad (14)$$

while

$$E_2 = 2.07 \times 10^{-10} + 1.66 \times 10^{-10} = 3.73 \times 10^{-10} \text{ erg}. \quad (15)$$

Again we conclude that any cavity which can reconstruct would do so with sufficient energy to produce the sort of dislocation loop here hypothesized (Figs. 2 and 3). However, the minimum energy release is only about 19% so that again it is probable that the dislocation loop will not always be formed.

(vi) Finally, we consider the question of the conditions under which one ought to expect the precipitation of vacancies to produce "vacancy-type" dislocations loops in semiconductors. This question is relevant because there are a few observations[80,81] of vacancy-type dislocations in these materials where most dislocations are observed to show interstitial-type lattice bowing in the electron microscope. The important point here would seem to be that reconstruction is a thermally activated process.[41,60,61] It would seem that rewelding should occur without an activation barrier if unreconstructed surfaces (with dangling bonds in compatible orientation) are mechanically forced together by any stress field that happens to be present. If the dislocation is nucleated at too low a temperature, the reconstruction mechanism will not be competitive against rewelding. Although one would have to consider the details of a particular case to make an accurate estimate, a crude assumption, that the critical temperature T_c, dividing the regimes of nucleation of vacancy-type and of interstitial-type dislocation loops is that for which the mean time to reconstruction is 600 sec, leads[41,60] to the estimate that $T_c(\mathrm{Si}) = 390\,^\circ\mathrm{C}$ and that $T_c(\mathrm{GaAs}) = 380\,^\circ\mathrm{C}$.

Once the loop has been nucleated and has started to grow as one type or the other, one should expect it to continue to grow as that same type. In the case of an "interstitial-type" loop, this follows because of the tensile strain field outside the perimeter, where the vacancies would tend to cluster and permit the lattice planes on either side of the extra partial plane to separate further apart. In the case of a "vacancy-type" dislocation, this would follow because the vacancies would tend to cluster on the outside perimeter of the loop, where the atoms are not well bonded (Fig. 1) so that $\Delta H_f(V)$ is small, and would permit the host-lattice planes on each side of the missing partial plane to come together and reweld. Note that in both cases vacancies will be approaching the core of the dislocation predominantly from the outside of the loop because the dislocation is gettering them from the bulk of the crystal.

The distinction just made may serve to resolve the discrepancy between the reports by Kimerling *et al.*[15] and by O'Hara *et al.*[16] that dark line defects (DLD's) in GaAlAs lasers are a convolution of interstitial-type dislocations, and the report by Woolhouse *et al.*[80] that the dislocations, at least in optically degraded material, are in fact of vacancy type. The former two groups studied DLD's which formed about dislocations that were produced during crystal growth and thus were nucleated while the crystal was hot enough for reconstruction to occur in times of order 10 msec. Woolhouse *et al.* studied DLD's which formed about dislocations that were introduced into the sample by scratching its surface at room temperature and causing it to glide by optical excitation. The mean time to reconstruction over the activation barrier at room temperature is calculated to be[41] or order 10^{12} years.

IV. PROPOSED EXPERIMENTS THAT MAY VERIFY THE MECHANISM

Although the mechanism proposed here has the attraction that it would serve to resolve the fundamental question of the nature of deep point and line defects by explaining the general observation of extrinsic stacking faults and "interstitial-type" dislocations without invoking any self-interstitials, and thus contradicting the general conclusion that vacancies are far more numerous, it should be subjected to further scrutiny.

One fairly clear distinction between this mechanism and the alternative that the dislocations are formed by the precipitation of self-interstitials that were incorporated during crystal growth is that the present mechanism implies the presence of voids in the crystal taking up somewhat fewer lattice sites than the initial number of vacancies, i.e., of order 10^{15}–10^{16} per cm^3, while the self-

interstitial mechanism implies there will be many orders of magnitude less void. As was noted in Sec. II, the presence of small (or order 6 nm) voids is difficult to detect[29] by TEM. One could not distinguish the difference in absorbing power from surface effects and noise. The scattering intensity from vacancies and interstitials is the same in the absence of lattice distortions. In either case there would be lattice distortion due to impurities because both types of dislocation would tend to attract impurities[82] by their strain fields. (In fact, the dislocations are almost always observed to be decorated with impurities.[83]) However, the presence of such voids should be detected if one diffuses a radioactive gas, such as ^{3}H or ^{37}Ar or ^{41}Ar or ^{39}Ar, through the sample[84] and measures the radioactivity remaining in the sample. As these gases diffuse rapidly and interstitially, there would be a negligible concentration of these on substitutional lattice sites, but they would tend to be trapped within any voids in the sample.

Other interstitially diffusing impurities, such as Cu or Li, would also be expected to precipitate into any cavities that are present in the interior of the loop as well as around the perimeter. With the self-interstitial mechanism, one would expect to find such precipitates only about the core of the dislocation and not in the body of the loop, which should be perfect material according to that model. Therefore, by determining the number of atoms of Li or of Cu which precipitate at low temperatures on dislocation loops that were previously grown into the sample, such as the swirl defects[85] in Si, and comparing this to the number of sites on the perimeter of the loops and to the number of sites within the loops, one should be able to determine if the impurity is restricted to the dislocation core. It may also be possible to detect the presence of such impurities in the body of the loop by high-resolution electron energy-loss analysis[86] or by x-ray microanalysis.[87]

It should be noted that the absence of Moire fringes in the body of many of these loops is not conclusive evidence that the material in the loop is perfect. It may just as well be that the material is highly disordered, essentially amorphous in that region, or that it consists of pillars of perfect material dispersed between regions of void or of disordered material.

ACKNOWLEDGMENTS

This work was encouraged by John W. Matthews, now deceased, as fine a scientist and a man as the author has ever known. The author has also benefited from discussions with R. A. Ghez, T. S. Kuan, and D. V. Lang. G. R. Woolhouse suggested Fig. 3 and the calculation in Eqs. (5)–(15). This work was supported in part by the Air Force Office of Scientific Research under contract No. F49620-77-C-0005.

[1]J. W. Matthews and J. A. Van Vechten, J. Crystal Growth 35, 343 (1976).
[2]A. U. MacRae, Bull. Amer. Phys. Soc. 22, 306 (1977).
[3]H. Shiraki, in *Semiconductor Silicon 1977*, edited by H. R. Huff and E. Sirtl (Electrochemical Society, Princeton, 1977), pp. 546.
[4]R. B. Fair and J. C. C. Tsai, J. Electrochem. Soc. 124, 1107 (1977).
[5]E. Nes and J. Washburn, J. Appl. Phys. 44, 3682 (1973).
[6]E. Nes, Phys. Status Solidi A 33, K5 (1976).
[7]S. M. Hu, J. Vac. Sci. Technol. 14, 17 (1977).
[8]P. M. Petroff and A. J. R. deKock, J. Cryst. Growth 30, 117 (1975).
[9]A. Seeger, H. Föll, and W. Frank, in *Radiation Effects in Semiconductors 1976*, edited by N. B. Urli and J. W. Corbett, Conf. Ser. No. 31 (Institute of Physics, Bristol, 1977), p. 12.
[10]H. Föll and B. O. Kolbesen, Appl. Phys. 8, 319 (1975); H. Föll, U. Gösele, and B. O. Kolbesen, J. Crystal Growth 40
[11]A. J. R. deKock, in Ref. 3, p. 508.
[12]J. R. Patel, Ref. 3, p. 521.
[13]J. A. Van Vechten, J. Electrochem. Soc. 122, 1556 (1975).
[14]G. R. Woolhouse, IEEE J. Quantum Electron. 11, 556 (1975).
[15]L. C. Kimerling, P. M. Petroff, and H. J. Leamy, Appl. Phys. Lett. 28, 297 (1976).
[16]S. O'Hara, P. W. Hutchinson, and P. S. Dobson, Appl. Phys. Lett. 30, 368 (1977).
[17]G. D Watkins, in *Lattice Defects in Semiconductors 1974*, edited by A. Seeger, Conf. Ser. No. 23 (Institute of Physics, London, 1975), p. 1.
[18]C. Weigel, D. Peak, J. W. Corbett, G. D. Watkins, and R. P. Messmer, Phys. Rev. B 8, 2906 (1973).
[19]J. A. Van Vechten, in Ref. 9, p. 441.
[20]H. C. Casey, M. B. Panish, and L. L. Chang, Phys. Rev. 162, 660 (1967).
[21]J. A. Van Vechten, in Ref. 17, p. 212.
[22]J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B 14, 3551 (1976).
[23]J. C. Phillips and J. A. Van Vechten, Phys. Rev. Lett. 30, 220 (1973).
[24]J. A. Van Vechten, Phys. Rev. B 10, 1482 (1974).
[25]J. A. Van Vechten, J. Electrochem. Soc. 122, 419 (1975).
[26]S. Mader and A. E. Michel, Phys. Status Solidi A 33, 793 (1976) and (private communication).
[27]P. S. Dobson (private communication).
[28]J. Friedel, *Dislocations* (Pergamon, Oxford, 1964), p. 91.
[29]J. M. Cowley, Acta Crystallogr. A 29, 529 and 537 (1973).
[30]W. C. Dash, J. Appl. Phys. 31, 2275 (1960).
[31]V. A. Phillips and W. C. Dash, J. Appl. Phys. 33, 568 (1962).
[32]L. S. Milevskii, Sov. Phys.-Solid State 4, 1792 (1963).
[33]T. Iizuka, J. Appl. Phys. Jpn. 4, 1018 (1966).

[34]R. W. Balluffi, Phys. Status Solidi 31, 443 (1969), especially Fig. 1 of Part 2. Note the correction at the end of the caption to this figure.
[35]J. M. Silcock and W. J. Tunstall, Philos. Mag. 10, 361 (1964).
[36]R. Räty and H. M. Miekk-eja, Philos. Mag. 18, 1105 (1968).
[37]F. Seitz, Phys. Rev. 79, 723 (1950).
[38]L. M. Brown and G. R. Woolhouse, Philos. Mag. 21, 329 (1970).
[39]T. Y. Tan and W. K. Tice, Philos. Mag. 34, 615 (1976).
[40]N. H. Fletcher, J. Crystal Growth 28, 375 (1975).
[41]J. A. Van Vechten, J. Vac. Sci. Technol. 14, 992 (1977), especially pp. 994–995.
[42]A. A. Chernov and S. I. Budurov, Kristallografiya 9, 388 (1964) [Sov. Phys.-Crystallogr. 9, 309 (1965)].
[43]A. A. Chernov and S. I. Budurov, Kristallografiya 9, 466 (1964) [Sov. Phys.-Crystallogr. 9, 388 (1965)].
[44]K. A. Jackson, Philos. Mag. 7, 1117 (1962).
[45]T. L. Davis and J. P. Hirth, J. Appl. Phys. 37, 2112 (1966).
[46]J. J. Lander and J. Morrison, J. Appl. Phys. 34, 1403 (1963).
[47]J. J. Lander, in *Progress in Solid State Chemistry*, edited by H. Reiss (Pergamon, Oxford, 1965), Vol. 2, p. 26.
[48]F. Jona, IBM J. Res. Dev. 9, 375 (1965).
[49]P. Mark and W. R. Bottoms, in Ref. 47, Vol. 6, 1971, p. 17.
[50]F. Bäuerle, W. Mönch, and M. Henzler, J. Appl. Phys. 43, 3917 (1972).
[51]W. Mönch, in *Advances in Solid State Physics* (Pergamon-Vieweg, Braunschweig, 1973), Vol. 13, p. 241.
[52]J. C. Phillips, Surf. Sci. 40, 459 (1973).
[53]K. C. Pandey (unpublished).
[54]H. Brooks, *Impurities and Imperfections* (American Society for Metals, Cleveland, 1955), p. 84.
[55]N. D. Lang and W. Kohn, Phys. Rev. B 1, 4555 (1970).
[56]J. Schmit and A. A. Lucas, Solid State Commun. 11, 415 (1972).
[57]J. A. Van Vechten, Phys. Rev. B 11, 3910 (1975).
[58]J. A. Appelbaum and D. R. Hamann, Phys. Rev. B 6, 2166 (1972).
[59]J. E. Rowe, S. B. Christman, and H. Ibach, Phys. Rev. Lett. 34, 874 and 1298(E) (1975).
[60]J. A. Van Vechten, Appl. Phys. Lett. 26, 593 (1975).
[61]J. A. Van Vechten, J. Crystal Growth 38, 139 (1977).
[62]H. C. Abbink, R. M. Broudy, and G. P. McCarthy, J. Appl. Phys. 39, 4673 (1968).
[63]L. J. Cheng, J. C. Corelli, J. W. Corbett, and G. D. Watkins, Phys. Rev. 152, 761 (1966).
[64]C. A. Ammerlaan and G. D. Watkins, Phys. Rev. B 5, 3988 (1972).
[65]G. Saada and J. Washburn, J. Phys. Soc. Jpn. 18 Suppl. 1, 43 (1963).
[66]J. A. Van Vechten, Phys. Rev. B 7, 1479 (1973) especially p. 1491.
[67]J. Friedel, in Ref. 28, p. 74.
[68]H. Brooks, in Ref. 54, p. 1.
[69]P. N. Keating, Phys. Rev. 145, 637 and 149, 674 (1966).
[70]R. M. Martin, Phys. Rev. B 1, 4005 (1970).
[71]J. A. Van Vechten, Phys. Rev. B 10, 4222 (1974).
[72]F. A. Kröger, Ann. Rev. Mat. Sci. 7, 449 (1977).
[73]J. Friedel, in Ref. 28, p. 22.
[74]J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1968), p. 404.
[75]See Ref. 74, p. 762 (Note the error in the values of H and of μ quoted for Si in this table).
[76]H. J. McSkimin, W. L. Bond, E. Buelher, and G. K. Teal, Phys. Rev. 83, 1080 (1951).
[77]C. D. Thurmond, J. Electrochem.Soc. 122, 1133 (1975).
[78]J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B 14, 3539 (1976).
[79]N. A. Goryunova, *The Chemistry of Diamond-like Semiconductors* (MIT, Cambridge, Mass., 1965), p. 99.
[80]G. R. Woolhouse (unpublished).
[81]G. H. Olsen (private communication); H. Kressel, C. J. Nuese, and G. H. Olsen, J. Appl. Phys. (to be published).
[82]J. Friedel, in Ref. 28, Chap. 13.
[83]J. P. Hirth and J. Lothe, in Ref. 75, p. 462.
[84]A. Van Wieringen and N. Warmoltz, Physica (Utr.) 22, 849 (1956).
[85]A. J. R. deKock, Philips Res. Rep. Suppl. 1, 1 (1973).
[86]J. Silcox, Scanning Electron Microsc. 1 393 (1977).
[87]H. Koike, T. Namae, T. Watabe, and A. Miksjiri, Jpn. Electron Opt. Lab. News 10, 2 (1973).

VOLUME 41, NUMBER 4 PHYSICAL REVIEW LETTERS 24 JULY 1978

Threshold for Optically Induced Dislocation Glide in GaAs-AlGaAs Double Heterostructures: Degradation via a New Cooperative Phenomenon?

B. Monemar,[a] R. M. Potemski, M. B. Small, J. A. Van Vechten, and G. R. Woolhouse[b]
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 5 October 1977)

We have observed a sharp threshold for the process of optically induced glide at which the velocity changes by more than a factor of 10^3 when the excitation intensity changes only 20%. This threshold is insensitive to doping and to the presence of a p-n junction. The effect is shown not to be related to recombination-enhanced motion or to local heating. An explanation in terms of the reduction of frictional forces by interaction with unrecombined carriers is offered.

It has been reported previously[1,2] that optical excitations of intensity in excess of 10^5 W cm^{-2} can lead to rapid (50 μm s^{-1}) dislocation glide in AlGaAs double-heterostructure lasers. The observations were made using photoluminescence topography. A laser wafer, without the normal capping layer of GaAs, is illuminated by the focused 647.1- and 616.4-nm radiation of a Kr^+ laser, to which the upper waveguide layer is transparent but the active layer is not. The active layer may be observed by the photoluminescence produced in it. Defects are apparent as dark regions because of the higher probability of nonradiative recombination events in their vicinity. Thus the motion of a gliding dislocation may be observed directly as a moving dark spot in the photoluminescence field.

The glide process does not occur at dislocations which were present during growth, but only at fresh dislocations. These are introduced by gently scratching the surface of the upper waveguide layer. On applying an optical excitation of sufficient duration, a dense dislocation network develops within a few microns of the scratch. Similar networks have been observed in studies of moderate-intensity 10.6-μm-laser damage in GaAs.[3]

A secondary stage of damage may now be produced if the optical excitation is sufficiently intense. Then we find that dislocation loops will glide away from the damaged area. This results in a portion of the dislocation threading from the bottom waveguide layer through the active layer and the upper waveguide layer to the surface.

© 1978 The American Physical Society

This new threading dislocation may be made to glide in a ⟨110⟩ direction. As it does so, it lays down an increasing length of misfit dislocation. If one assumes the scratch punches out interstitial dislocations, then the extra half-planes may relieve the misfit between the active layer, which has a smaller lattice constant, and the underlying waveguide. Therefore, misfit dislocations are expected to occupy this interface. Transmission electron microscopy has been used to confirm this location.[4]

Dislocation glide in double-heterostructure lasers has also been reported under conditions of electrical excitation and simultaneous stress.[5,6] The same phenomenon has been observed in GaP light-emitting diodes.[7] In order to explain this remarkably enhanced dislocation velocity both local heating effects and recombination-enhanced motion have been suggested tentatively.[1,2,6] In this Letter we report additional observations of these effects. We find a sharp threshold of excitation intensity, above which the dislocation moves extremely rapidly. The nature of the change about the threshold is so great as to imply a critical phenomenon and to discount the previously suggested mechanisms. We provide both a description of the observations made on dislocation movement and a new tentative theory which may explain the sharp threshold we have observed. We conclude it is likely that some cooperative phenomenon is active in producing the observed critical phenomenon.

The materials used for the present study consisted of three layers grown on GaAs(100) substrates by liquid-phase epitaxy. These layers were the lower waveguide, the active layer, and the upper waveguide. The normal GaAs capping layer was omitted. Details of both the Al contents of the three layers and the doping of the layers are given in Table I.

The exciting radiation was focused to a spot whose diameter was 30 μm. The intensities used in this investigation are within the range of those at which semiconductor lasers are operated. (The absolute limit of optical flux which may be obtained from such a laser is that which causes catastrophic facet damage. This occurs when the emergent flux is greater than[8,9] 3×10^6 W/cm^2. The internal flux is greater than this.) It is also worth noting that the samples were not lasing. Although the local flux was above threshold, only a small fraction the total cavity was being pumped

As dislocations were seen to move in the irradiated region, the sample was tracked to keep the moving point central in the field.

The nature of the movement was complicated. At very high excitation levels, $\sim3\times10^5$ W cm^{-2}, the dislocation glides with a fairly uniform rate across the excited area. As the intensity is reduced a threshold region is reached where the dislocation velocity diminishes by at least three orders of magnitude for a 20% change in the exciting intensity. The motion of the dislocations also becomes irregular in this region. When the intensity is below threshold, the speed is 60 nm/s or less and is in a ⟨100⟩ rather than a ⟨110⟩ direction. This behavior of velocity as a function of intensity, as intensity is reduced, is shown in Fig. 1.

The growth in the ⟨100⟩ direction at low intensity has been investigated by transmission electron microscopy.[10] A complex structure is produced as a result of dislocation climb. Such structures are evidenced as dark line defects. Once the dislocation has grown a few microns by the climb mechanism, it will not glide again.

It has been proposed that the glide mechanism is thermally activated.[6] If this were the case, the dislocation velocity, v, would be given by

$$v = \nu b B \tau^m e^{-U/kT}, \quad (1)$$

where ν is an attempt frequency, b the Burgers vector of a dislocation, U the activation energy, τ the local stress, m an exponent in the range[11] from 1 to 2, and B is a constant. It has been estimated[12,13] that the temperature rise under

TABLE I. Sample parameters (layers numbered from substrate) and measured threshold intensities, I_t.

Wafer	I_t (10^5 W/cm^2)	Percentage Al (layers 1/2/3)	Carrier type, number/cm^3 (dopant)	Width (μm)
A	0.8	32/1/32	[*n*, 10^{15}(Sn)]/[*n*, 10^{18}(Te)]/[*n*, 10^{15}(Sn)]	3.1/0.7/0.9
B	1.4	41/1/41	[*n*, 10^{17}(Te)]/[*n*, 10^{17}(Sn)]/[*n*, 10^{16}(Sn)]	5.0/0.7/1.5
C	1.6	40/1/40	[*n*, 10^{17}(Te)]/[*n*, 10^{15}(—)]/[*p*, 10^{17}(Ge)]	4.5/0.5/0.8
D	2.0	40/1/40	[*n*, 10^{17}(Te)]/[*p*, 10^{18}(Ge)]/[*p*, 10^{17}(Ge)]	3.6/0.6/0.6

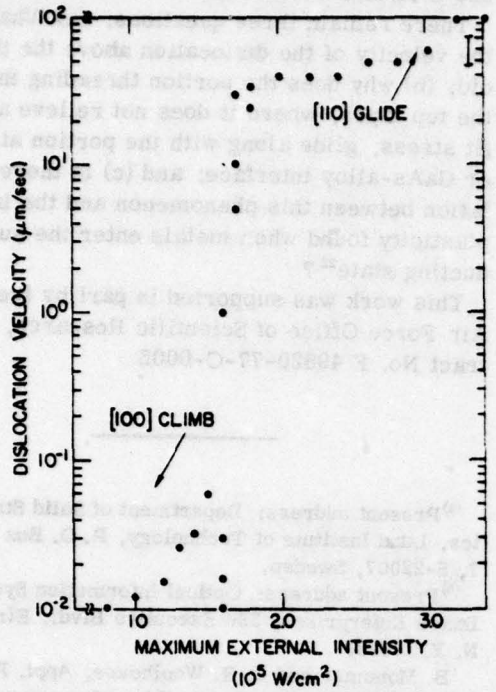


FIG. 1. Semilog plot of velocity of photoluminesence spot due to threading dislocation versus external intensity of 647.1- and 676.4-nm laser light. External intensity is controlled to about 10%; internal intensity may vary more due to variations in sample surface. At the highest intensities, velocity uncertainty becomes as large as indicated. Just below threshold the scatter in data is large, but it is difficult to separate any glide component of the predominant climb motion. Only the highest velocity values observed for several dislocations are shown below threshold whereas several measurements are shown for each intensity in the threshold region.

these experimental conditions may be as great as 200°C. Three scratched samples were heated to temperatures of 250, 400, and 500°C, respectively, for periods of 1 min. No dislocation glide was observed to have taken place. Upon optical excitation, the samples behaved exactly like unannealed samples. In Eq. (1) τ would not be expected to be more than linear in the intensity I, so that one would not expect the preexponential to be more than quadratic in I. Note that the effect of the nonlocal heating due to the absorption of the light only in the GaAs is to *reduce* the misfit stress that drives the glide. This is because the GaAs has the smaller lattice constant and expands as it is heated. A quantitative analysis of the nonlocal heating shows that it reduces the misfit stress less than 10% in our experiment.

A further characteristic of thermally activated glide is that it is strongly dependent upon the doping of the crystal.[14] Therefore, the different compositions of the structures listed in Table I ought to be expected to show large differences in threshold levels or velocities. Neither is observed.

The other previously proposed mechanism for glide is that energy from recombining electron-hole pairs[15] enhances the motion of the dislocation.[1,2] In this case, one may write

$$v = \nu b(e^{-U/kT} + Ae^{-(U-R)/kT}), \tag{2}$$

where A is a constant and R is the energy donated by a recombination event. According to this hypothesis v would simply be linear in I.

We will now present a tentative explanation for optically induced glide which we find consistent with our observations. It is proposed that the threading dislocation glides locally when the instantaneous concentration, N, of unrecombined electron-hole pairs in its vicinity fluctuates from its mean value, M, to such a high value, N_c, that the frictional forces of the lattice are weakened below the level required to prevent the dislocation from gliding under its resolved stress. We recall[16-21] that the excitation of an electron from the bonding states near the top of the valence band to the antibonding states of the conduction band severely reduces the frequency of transverse acoustic lattice modes and the shear strength. When $N_c \gg M$, the probability that an instantaneous fluctuation in N will drive it above N_c is a very rapidly increasing function of M, and thus of I, for virtually all plausible models of the statistical distribution.

We illustrate our proposal with an admittedly oversimplified model in which the number, n, of electron-hole pairs in a critical volume, V, about the dislocation obeys the Poisson distribution,

$$P(n) = m^n e^{-m}/n!, \tag{3}$$

where m is the mean number of pairs, and assume

$$v = \nu b \textstyle\sum_n P(n) e^{-U(n,T,\tau)/kT}, \tag{4}$$

where the activation energy, U, is now taken to be a function of n, of temperature, and of strain. Note that even at the highest velocities the dislocation jumps only very infrequently (its velocity is much less than the speed of sound). Taking $\nu = 5\times10^{12}$/sec and $b = 0.4$ nm, one would have $\nu b = 2.0\times10^{3}$ m/sec, a factor of 10^{7} faster than the

highest velocity observed above threshold and at least 10^{10} faster than below. With a 1-ns lifetime, M is of order 10^{19} cm^{-3} in the bulk of the active layer at threshold. Because of recombination at the dislocation, let us suppose it is 5×10^{18} in the region V about the dislocation. We may also suppose V has a radius of 1.0 nm and a length of 1.0 μm so that its volume is 3×10^{-18} cm^3 and therefore $m=15$. If we further suppose that the critical number $n=45$, the intrinsic value at the maximal melting point, we find $P(45)=2.1\times10^{-10}$. Now if the intensity and thus m is increased by 20% to 18, we find that $P(45)=7.9\times10^{-7}$, an increase of more than three orders of magnitude. While this choice of suppositions serves to produce from Eq. (4) a predicted onset of the glide consistent with the lower bound for the sharpness of the observed threshold, it is obvious that a number of other choices of approximations and parameters would also serve the purpose. We suspect that the threshold is actually much sharper than this lower bound. (Observation of lower rates of glide is obscured by the dislocation climb.) The threshold would be sharper if a cooperative phenomenon among the unrecombined carriers operates to cause their concentration to increase more than linearly with I near the critical intensity. Possible cooperative phenomena might include a saturation of the recombination channels about the dislocation or a polarization of the dislocation which might attract carriers from the bulk or strain or field gradients produced by the carriers that might separate electrons and holes spatially to prevent their recombination.

The difficulty in causing the dislocation to glide once it has climbed may be explained either by the reduction in carrier lifetime, and thus in M and m, in the vicinity of a climb network due to recombination at it, or by a pinning effect of the climb network upon the threading dislocation. The inability to glide grown-in dislocations may similarly be explained by the recombination due to defects which congregate about dislocations at growth temperatures, or by a difference in the geometries of the two dislocations. We note that the maximum n- or p-type doping in our samples was about 1×10^{18} cm^{-3}, which is a factor of 15 less than our assumed value for N_c, so that the observed insensitivity of threshold to doping seems also to be consistent. Moreover, the intrinsic carrier concentration reaches N_c only at the congruent melting point of the material; so the lack of thermally activated migration at 500°C is also consistent.

There remain three questions: (a) What limits the velocity of the dislocation above the threshold; (b) why does the portion threading through the top layer, where it does not relieve any misfit stress, glide along with the portion at the lower GaAs-alloy interface; and (c) is there any relation between this phenomenon and the increased plasticity found when metals enter the superconducting state[22]?

This work was supported in part by the U. S. Air Force Office of Scientific Research, Contract No. F 49620-77-C-0005.

[a]Present address: Department of Solid State Physics, Lund Institute of Technology, P. O. Box 725, Lund 7, S-22007, Sweden.

[b]Present address: Optical Information Systems, Exxon Enterprises, 350 Executive Blvd., Elmsford, N. Y. 10523.

[1]B. Monemar and G. R. Woolhouse, Appl. Phys. Lett. 29, 605 (1976). Note that in this reference and in the next intensities were stated in terms of the average power across the spot rather than as the peak intensity at the center cited here.

[2]B. Monemar and G. R. Woolhouse, Inst. Phys. Conf. Ser. 33a, 400 (1976).

[3]J. J. Comer, J. Appl. Phys. 47, 1780 (1976).

[4]G. R. Woolhouse, B. Monemar, and C. M. Serrano, to be published.

[5]T. Kamejima, K. Ishida, and J. Matsui, Jpn. J. Appl. Phys. 16, 233 (1977).

[6]H. Nakashima, S. Kishino, N. Chinone, and R. Ito, J. Appl. Phys. 48, 2771 (1977).

[7]M. Iwamoto and A. Kasami, Appl. Phys. Lett. 28, 591 (1976).

[8]B. W. Hakki and F. R. Nash, J. Appl. Phys. 45, 3907 (1974).

[9]Y. Shima, N. Chinone, and R. Ito, Appl. Phys. Lett. 31, 625 (1977).

[10]P. M. Petroff, W. D. Johnston, Jr., and R. L. Hartman, Appl. Phys. Lett. 25, 226 (1974).

[11]S. K. Choi, M. Mihara, and T. Ninomiya, Jpn. J. Appl. Phys. 16, 737 (1977).

[12]Y. Nannichi, J. Matsui, and K. Ishida, Jpn. J. Appl. Phys. 14, 1561 (1975).

[13]T. Kobayashi, T. Kowakami, and Y. Furukawa, Jpn. J. Appl. Phys. 14, 508 (1975).

[14]S. A. Erofeeva and Yu. A. Osip'yan, Fiz. Tverd. Tela (Leningrad) 15, 772 (1973) [Sov. Phys. Solid State 15, 538 (1973)].

[15]D. V. Lang and L. C. Kimerling, Phys. Rev. Lett. 33, 489 (1974).

[16]H. Brooks, Adv. Electron. 7, 121 (1955).

[17]E. Antoncik, Czech. J. Phys. 5, 449 (1955).

[18]J. A. Van Vechten, Inst. Phys. Conf. Ser. 23, 212

(1975).
[19]V. Heine and J. A. Van Vechten, Phys. Rev. B 13, 1622 (1976).
[20]J. A. Van Vechten, Phys. Rev. B 13, 946 (1976).
[21]J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B 14, 3539 (1976).
[22]H. Kojima and T. Suzuki, Phys. Rev. Lett. 21, 896 (1968); A. V. Granato, Phys. Rev. Lett. 27, 660 (1971).

Appendix C

GaAs is different from Si*

by
J. A. Van Vechten
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 100598

Defect thermochemistry in zinc-blende III-V's is different from, and more complicated than, that in Si or Ge because the two sublattices of the III-V are inequivalent. Thus a given impurity may occur on either the anion or cation sublattice and may transfer from one to the other with atomic diffusion. Moreover, simple anion vacancies, for example, convert to cation vacancies plus cation-on-anion-site antisite defects each time the vacancy migrates to a nearest neighbor site. In this paper two effects resulting from this complication will be treated. These are the variation in the rate of anion and cation atomic diffusion with the stoichiometry of the host material and the site preference and compensation ratio of Group IV impurities (C, Si, Ge, and Sn) in III-V's. The first problem has been highlighted by the recent observations of Small and Ghez that Al diffusion in GaAlAs is about 10^5 times faster under LPE conditions than the rate measured for bulk MBE material by Chang, et al. and by Dingle, et al. This discrepancy is in accord with the author's previous estimates of the equilibrium concentrations of anion and cation vacancies for the appropriate growth conditions (LPE is done on the metal rich side and MBE on the As rich side of the existence curve) and the conclusion that self-diffusion occurs by nearest neighbor vacancy migration in GaAs and most other III-V's. Thus the rate of Al diffusion is proportional to the As vacancy concentration. The donor or acceptor character of Group IV impurities is governed by two factors. The relative concentrations of anion and cation vacancies during growth implies more p-type behavior in LPE material than in MBE or VPE material. The second factor is the site preference of a given impurity independent of stoichiometry. Two factors contribute to this site preference -the relative size of the impurity and host atoms and the electronic structure of the impurity's core. It is argued that large size mismatches of either sign are more easily accommodated on an anion site than on a cation site, so that both C and Sn are acceptors in GaAs. However, Si, which is nearly the same size as Ga or As, prefers the cation site due to its lack of a filled d-core and is usually a donor. Ge has little preference because its size and core are nearly the same as those of Ga and As. It is a donor or acceptor according to stoichiometry.

* This work was supported in part by the U. S. Air Force Office of Scientific Research under contract No. F 49620-77-C-0005.

Appendix D

DN 11 **Mechanism for the Production of a Metastable Plasma in Pulsed Laser Annealing** Ellen J. Yoffa and J. A. Van Vechten, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598 * --High density electron and hole plasmas can be created by interband absorption of pulsed laser radiation. For sufficiently intense laser pulses, the energy of the plasma mode will be driven through the optic phonon branch to a value which exceeds phonon energies. The plasma excitations are then effectively decoupled from the phonons and rapid heating of the lattice cannot occur. At a comparable density ($\sim 10^{19}$ cm^{-3}) Auger recombination within the degenerate system is suppressed due to coulombic screening effects and phase-space restrictions. Consequently, the laser induced plasma becomes metastable and persists for times of order 10^{-7} sec, until decay occurs due to expansion and eventual recombination.

*Supported in part by AFOSR Contract No. F49620-77-C-0005

DN 12 **Defect Migration During Pulsed Laser Induced Plasma Annealing** J. A. Van Vechten and Ellen J. Yoffa, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598* -- Vacancies in Si are known to be mobile at 80 K with an activation energy varying with charge state from 0.2 to 0.33 eV. The reason that ion implanted or amorphous films do not anneal at room temperature and that normal thermal annealing leaves a large fraction of the dopant atoms in electrically non-active complexes, e.g., P^+V^- or B^-V^+, is that vacancies are trapped by coulombic binding to various charged centers. We argue that during pulsed plasma annealing, PPA, the coulomb interaction is screened and vacancies are allowed to migrate out of the lattice even if it has not been heated above room temperature. Depending on the experiment, they either migrate to a free surface or form small voids. Optically induced dislocation glide[1] is also expected and should aid the perfection of the material. In this way, the thermal shock associated with thermal melting and recrystallization is avoided and very high quality crystal with supersaturated doping is obtained.

*Supported in part by AFOSR Contract No. F 49620-77-C-0005

1. B. Monemar, et al., Phys. Rev. Lett. 41 260 (1978).

Appendix E

Self-Consistent Method for Point Defects in Semiconductors: Application to the Vacancy in Silicon

J. Bernholc, Nunzio O. Lipari, and Sokrates T. Pantelides
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 30 June 1978)

We report the development of a method to calculate self-consistently the electronic structure of neutral point defects in semiconductors. The method is an adaptation of the original Koster-Slater idea. Calculations become feasible, practical, and accurate at the level of current band-structure and surface calculations when an LCAO basis set is used instead of Wannier functions. A detailed study of the isolated vacancy in Si is used to illustrate the method.

Point defects (vacancies, interstitials, etc.) and impurities in semiconductors are known to introduce localized states with energy levels in the fundamental gap. While shallow levels and some moderately deep levels are adequately described by effective-mass theory,[1] the theoretical description of most deep levels, which play a dominant role in determining many properties of electronic devices, has been one of the major outstanding problems of semiconductor physics. A large number of methods have been introduced and used for a variety of defects and impurities.[2]

© 1978 The American Physical Society

The results have provided substantial qualit understanding of the nature of binding at deep levels, but no approach has established itself as capable of providing accurate and unambiguous solutions, comparable in accuracy and reliability to the calculations that are currently possible for perfect bulk solids (band theory), surfaces, and interfaces. In this Letter, we report the development of a method which is capable of producing results of precisely such accuracy and reliability and illustrate its power with a detailed description of an isolated neutral unreconstructed vacancy in silicon.

The problem at hand has two distinct aspects: the choice of *Hamiltonian* and that of a *method* to seek the corresponding eigensolutions. Even though a variety of Hamiltonian choices (semiempirical, superposition of atomic potentials, etc.) can provide useful information, we have elected to use our method in the context of a self-consistent local-density theory of electronic structure. This choice frees our results from any dependence on the assumed similarity of interactions in the perturbed system to those in the unperturbed bulk crystal. We show below that some of the results we obtained previously using a semiempirical tight-binding Hamiltonian in fact survive the iteration to self-consistency, but others do not.

Given the Hamiltonian, we now turn to the choice of method. The most common method, namely the cluster method, has recently[3] been used with self-consistent Hamiltonians. The conclusion from these studies was that even 54-atom clusters, the largest that could be handled, are not adequate to contain the bound-state wave functions and produce a large uncertainty.[3] More recently, two of us[4] demonstrated that, when a semiempirical Hamiltonian is used, the most powerful and accurate method available to solve the problem of neutral point defects is an adaptation of the Koster-Slater[5] method, first introduced in 1954 and subsequently used in several studies of point defects[6,7] and surfaces.[8] Our method has three clear-cut advantages over the cluster approach. First, our approach permits us to focus directly on *changes* in the electronic structure caused by the defect; related to this point is the fact that properties of the bulk, such as the band gap, are accurately built in to our calculations and are not affected by introduction of the defect. This is not true of even the largest clusters (54 atoms[3]) which have been studied self-consistently. Second, our formalism fully exploits the translational symmetry of the host material. The third important advantage of this approach is that the magnitude of the required numerical problem is governed by the spatial range of the *perturbation potential* rather than the very much greater range over which the electronic *wave function* is altered. In the particular application discussed below, for example, a full 30% of the charge associated with the localized state in the band gap lies outside of the region in which the perturbation potential is localized.

A description of the Koster-Slater method is available in the original paper and in a series of papers by Callaway and Hughes[6] and Callaway.[9] The most convenient formulation of the method is in terms of Green's functions. If $G^0(E)$ is the Green's-function operator for the perfect crystal, bound states in the gap introduced by a given perturbation U are given by the zeros of the determinant

$$D(E) = \det \| 1 - G^0(E)U \|. \tag{1}$$

The change in the density of states in the band continua is also given by an expression involving only $D(E)$. More importantly, the change in the charge density, which is needed for self-consistency, is given by

$$\Delta\rho = \frac{2}{\pi} \int_{\infty c} dE \{1 - [1 - G^0(E)U]^{-1}\} G^0(E). \tag{2}$$

For analytical as well as numerical work, the operators appearing in the above formulas must be represented in a convenient basis set. Traditionally, operators have been expressed in terms of Wannier functions as the natural basis set of localized functions. The first applications of the Koster-Slater approach,[6] however, proved extremely cumbersome, not for reasons related to the method *per se*, but for reasons related to the construction of the Wannier functions. The final results had to be empirically adjusted even to obtain a bound state in the gap. We avoid these serious difficulties by starting with the linear-combination-of-atomic-orbitals (LCAO) method for a self-consistent pseudopotential band-structure calculation.[10] We then use the same set of LCAO orbitals as basis states for all the operators of the Koster-Slater–Green's-function formalism.[11] While simple LCAO orbitals are not conveniently orthonormal, they are enormously easier to generate and the relatively small number of functions actually required to describe the defect-induced *change* in the electron density can be easily orthonormalized. Physically, the

implicit assumption is made that as long as the chosen set of LCAO orbitals on each atom of the perfect crystal is adequate to give an accurate band structure, then the same set of orbitals at all sites is also adequate to describe the infinite crystal containing a single vacancy.[11] Having chosen a set of LCAO orbitals on each atom,[10] we exploit symmetry and form symmetrized orbitals on each shell of atoms surrounding the defect site (shell orbitals). Finally, because shell orbitals are not orthogonal, we construct orthogonalized shell orbitals (OSO's) by orthogonalizing each shell orbital to the orbitals on all the shells closer to the defect site.

The Green's-function matrix elements between pairs of OSO's are then calculated using standard techniques of Brillouin-zone integration for the imaginary part and a Hilbert transform for the real part. Since the number of LCAO orbitals on each atom is finite, the Hilbert transform is uniquely defined in this case.[11] These quantities are calculated once for a given host crystal and stored. For a given U, one has simply to calculate its matrix elements between pairs of OSO's and construct the quantity $D(E)$ of Eq. (1), which now becomes

$$D(E) = \det \| \delta_{\alpha\beta} - \textstyle\sum_\gamma G_{\alpha\gamma}{}^0(E) U_{\gamma\beta} \|, \qquad (3)$$

where α, β, and γ label the OSO's. The self-consistent-field iteration is initiated with an arbitrary estimate of U. The change in the charge density $\Delta\rho(r)$ is then obtained by expressing (2) in the OSO representation. From this quantity, a new perturbation potential is constructed and so on. Details of the calculations will be given elsewhere.

We turn now to our results for the isolated vacancy in Si. The objectives of the calculations are to investigate the following: (a) bound-state energies and wave functions, (b) changes in the density of states within the band continua corresponding to resonances and antiresonances, and (c) the charge distribution in the vicinity of the vacancy.

The calculations were performed using LCAO's centered on the first three shells of neighbors surrounding the vacancy (total of 28 atoms) and on the vacancy site. From simple tight-binding arguments[4] for an unreconstructed vacancy, one expects bound states and the most important changes in the density of states to occur in the A_1 and T_2 symmetries. In Si, we find only one bound state within the fundamental gap; it has T_2 symmetry (threefold degenerate) and lies at 0.8 eV above valence bands, which is our reference energy. Since the crystal is neutral, this state contains only two electrons. The changes in the density of states for A_1 and T_2 symmetries are shown in Fig. 1. From the analytic properties of $D(E)$ [Eq. (1)], it can easily be shown that the total change in the density of states in the valence bands, for each symmetry, must integrate to an integer. In the case of T_2 states, the change is mostly negative and the integral is -6, thereby compensating the T_2 bound state in the gap. In the case of A_1 states, a sharp resonance is present at -0.6 eV. This resonance lies in an energy region where the local density of states of A_1 symmetry is very small so that the resonance is effectively a bound state. A second resonance is present at -8.1 eV. Antiresonances compensate so that the total integral is zero.

The contour plot in Fig. 2(a) shows that the total change in the charge density (and therefore also the corresponding potential) is localized almost entirely within the cavity defined by the nearest neighbors. Correspondingly, in our basis (which includes LCAO on the first three shells) the total change in the charge density integrates to its full value -4.0 (as we have removed an atom with

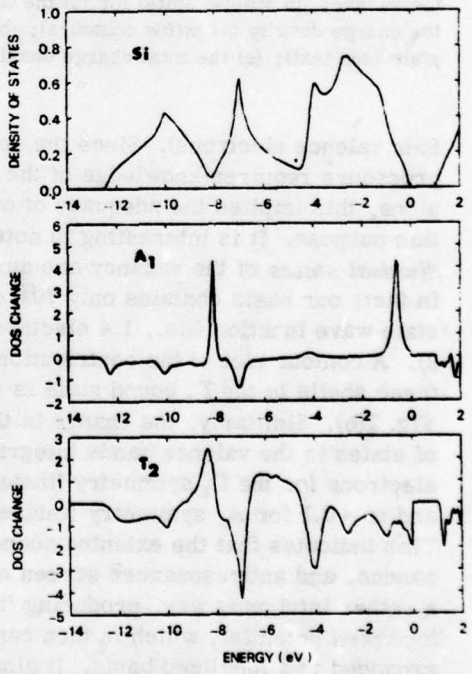


FIG. 1. The density of states and the change in the density of states of A_1 and T_2 symmetries. The curves are broadened by 0.2 eV and the reference energy is the top of the valence bands.

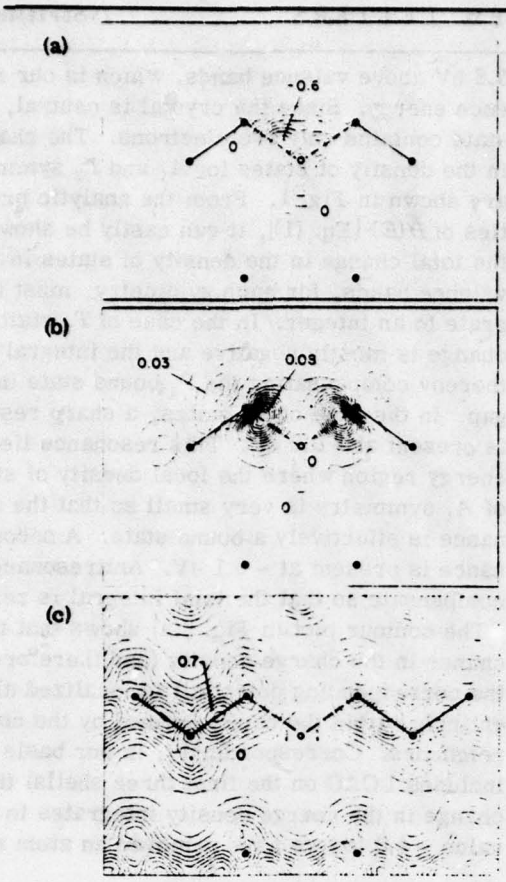


FIG. 2. Contours of constant electron density near the vacancy (in atomic units) for (a) the total change in the charge density (*negative* contours); (b) the T_2 bound state (see text); (c) the total charge density.

four valence electrons). Since the self-consistent procedure requires knowledge of the potential alone, this implies the adequacy of our basis for this purpose. It is interesting to note that the *individual states* of the vacancy are quite *extended*. In fact, our basis contains only 70% of the bound-state wave function (i.e., 1.4 electrons instead of 2). A contour plot of the contributions of the first three shells to the T_2 bound state is shown in Fig. 2(b). Similarly, the change in the density of states in the valence bands integrates to -5.2 electrons for the T_2 symmetry (instead of -6), and to -0.1 for A_1 symmetry (instead of 0). This indicates that the extented bound state, resonances, and antiresonances screen each other in a rather intriguing way, producing finally a very *localized* potential, which in turn can easily be expanded in a localized basis. It also clearly demonstrates the advantage of the present formalism. A cluster, which would adequately represent this infinite system, would need to contain fully these extended states individually.

The total charge density around the vacancy is shown in Fig. 2(c). We observe that all the bonds in the crystal remain virtually unchanged, and the dangling bonds remain pointing toward the missing atom. The contour plots reveal that the dangling bonds have a rather constant charge density near the atoms and extend appreciably into the vacant site. The bound-state orbital [Fig. 2(b)], although considerably delocalized, has a dangling-bond character.

Comparison of the present self-consistent results with those obtained earlier[4] using the semi-empirical tight-binding method (SETBM) reveals the following: (1) The state-density changes are strikingly similar. (2) The position of the bound state in the gap (0.8 eV versus 0.3 eV found earlier[4]) is appreciably different and probably reflects not only self-consistency, but mainly the effect of the more realistically broad conduction bands of the present work. (3) The electron-density maps *provided* by the present work confirm the basic *assumption* of the SETBM study, namely that the Si-Si bonds near the vacancy are not significantly affected. The present work also indicates that the simple "removal" model used in Ref. 4 to describe the vacancy is fundamentally correct.

Comparison with experiment is not yet possible because the vacancy in Si is known[12] to induce a significant lattice reconstruction, an effect which has not been included in the present calculations. The situation is comparable to free surfaces, where the determination of reconstruction is still an unsolved problem. However, the bound state we find at 0.8 eV is consistent with experimental observation[12] according to which the level for the neutral vacancy is near the bottom of the gap after Jahn-Teller stabilization by about 1 eV.[12]

In conclusion, we feel that the present work goes a long way towards bringing the isolated-defect problem to the level of sophistication which characterizes contemporary studies of crystal surfaces; we hope that theory will soon play an important role in analyzing the wealth of experimental data which already exists in this context.

It is a pleasure to thank S. G. Louie and A. R. Williams for valuable discussions. This work is supported in part by the U. S. Air Force Office of Scientific Research under Contract No. F49620-77-C-0005.

[1]W. Kohn, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York,

1957), Vol. 5, p. 257; S. T. Pantelides, Festkörperprobleme 15, 149 (1975).

[2]For a recent critical review of the field, see S. T. Pantelides, to be published.

[3]S. G. Louie, M. Schlüter, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. B 13, 1634 (1976). In this work, periodic boundary conditions were used (superlattice of defects). The bound state had a dispersion of 1 eV in the 1-eV gap of Si. Self-consistent calculations with even smaller clusters have been reported by B. Cartling, J. Phys. C 8, 3183 (1975) and by L. A. Hemstreet, Phys. Rev. B 15, 834 (1977).

[4]J. Bernholc and S. T. Pantelides, Phys. Rev. B 18, 1780 (1978).

[5]G. F. Koster and J. C. Slater, Phys. Rev. 95, 1167 (1954).

[6]J. Callaway and A. J. Hughes, Phys. Rev. 156, 860 (1967); and 164, 1043 (1967).

[7]M. Jaros and S. Brand, Phys. Rev. B 14, 4494 (1976). The method used in this work is a modification of the Koster-Slater approach introduced first by F. Bassani, G. Iadonisi, and B. Preziosi, Phys. Rev. 186, 735 (1969).

[8]J. Koutecky, Adv. Chem. Phys. 9, 85 (1965); J. Pollmann and S. T. Pantelides, Phys. Rev. B (to be published).

[9]J. Callaway, J. Math. Phys. (N. Y.) 5, 783 (1964).

[10]D. J. Chadi, Phys. Rev. B 16, 790 (1977).

[11]Ten orbitals per atom are used as in Ref. 10. The use of the same set of orbitals for the defect problem means that a total of twenty bands are used for the expansion of the defect wave functions. (It was established in Ref. 8 that ten bands are entirely adequate for convergence.) Our choice of orbitals for the defect problem also eliminates any ambiguities associated with Hilbert transforms which are necessary for the evaluation of Green's functions (see also Ref. 5). In the case of an impurity with an electronic structure significantly different from that of the host atoms, *additional* LCAO orbitals would have to be included on the impurity atom.

[12]G. D. Watkins, in *Gallium Arsenide and Related Compounds—1974*, Institute of Physics Conference Series No. 23, edited by J. Bok (American Institute of Physics, New York, 1975), p. 1.

VOLUME 40, NUMBER 3 PHYSICAL REVIEW LETTERS 16 JANUARY 1978

Electron Capture into the Two-Electron O^- State in GaP

T. N. Morgan

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

(Received 23 September 1977)

A reinterpretation of the ir spectra of Dean and Henry in oxygen-doped GaP strongly suggests that they involve radiative capture and excitation of the *second* electron bound to oxygen. Further, this electron is localized near the nearest-neighbor Ga ions, it is bound by ~0.9 eV as found also by Samuelson and Monemar, and its electron-phonon coupling is *small*—$\lambda \approx 0.6$, $\hbar\omega \approx 47.5$ meV—in disagreement with Henry *et al.* The analysis emphasizes the usefulness of the "phonon signature" in identifying deep levels from their optical spectra.

Dean and Henry (DH) have published a careful study of radiative electron capture onto deep oxygen donors in GaP.[1] Subsequently Henry and coworkers[2,3] discovered by photocapacitance techniques that each oxygen donor in GaP could bind a second electron to become O^-, with each electron bound by nearly 1 eV. Since this latter discovery both states have been studied extensively, and two conflicting models for the two-electron state have developed—that of Henry and co-workers[2-4] and that of Grimmeiss *et al.*[5] and of Morgan.[6] The former authors conclude that a very strong electron-lattice interaction is acting and, indeed, that GaP:O is a prime example of such interactions in semiconductors.[4] The latter authors conclude that a kind of selection rule must be operating, so that optical transitions between the X_1 conduction-band edge and the two-electron ground state of O^- are very weak. In this Letter I reexamine the DH data and propose that they involve the *second* electron in O^- and allow important properties of this anomalous state to be determined.

An indication that the original one-electron interpretation of the DH data is incorrect is given by a comparison of the shape of the phonon replica structure in the luminescence spectrum of DH with that of the donor-acceptor (D-A) pair spectra of Dean, Henry, and Frosch,[7] both of which are shown in Fig. 1. Monemar and Samuelson have recently found[8] that the same phonon energies [$\hbar\omega_1 = 19$ meV and $\hbar\omega_2 = 48$ meV (see also Ref. 7)] and relatively large coupling strengths [$\lambda_1 = 1.65$ and $\lambda_2 = 1.1$] which fit the pair (emission) spectra also explain the shape and temperature dependence of the photoneutralization (absorption) spectra for the O^0 ground state. [These are referred to as the configuration-coordinate (C-C) modes.] These phonons do *not* appear in other D-A pair spectra. Thus, this satellite structure may be considered the "signature" of the deep O^0 ground state with respect to its one-electron transitions.

The spectra in DH, however, are very different—in part because the zero-phonon transition O_0 (see Fig. 1) is forbidden and cooperation of a nonsymmetric (active) phonon is needed to relax the selection rule and produce a strong replica, and in part because the states are different. If the deep electron states involved in both of these spectra were the same, the C-C phonons and coupling strengths would be nearly the same (being only weakly dependent on the shallow state involved), and the same replicas of *any* strong features which appear in one spectrum would appear in both. Thus there would appear in Fig. 1(a) a strong broad 19-meV replica of O_{1cc}' and also of O_{TA} and of the optical phonon peaks, none of

© 1978 The American Physical Society

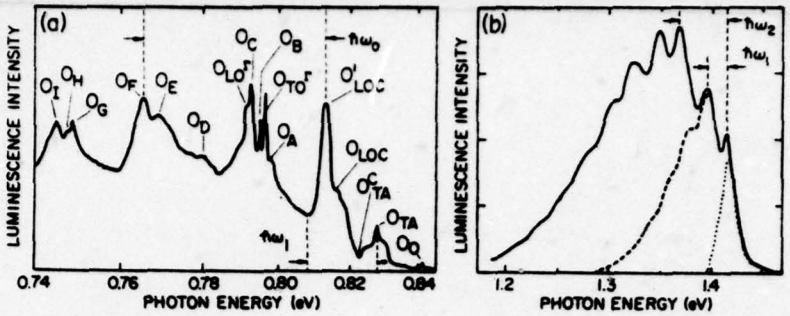


FIG. 1. Photoluminescence from O-doped GaP at low temperature: (a) capture spectrum from Ref. 1. showing C-C phonon ω_0 and the missing phonon ω_1; (b) pair spectrum from Ref. 8 showing the phonons and the stepwise deconvolution. Note the change in energy scale.

which is present. Therefore, the deep electron states responsible for the two spectra are not the same! Since the state involved in the pair spectra is known from many independent measurements to be the neutral oxygen donor, we conclude that the oxygen centers studied by DH are in the two-electron O^- state.

The conclusion that these spectra involve O^- is reinforced by another part of the DH study. In the n-type sample used for the excitation spectrum shown in Fig. 6(a) of Ref. 1, one expects the two-electron oxygen species to be dominant, $n_- \gg n_0 \gg n_+$, and the sharp-line spectrum shown in Fig. 6(a) to arise from the transition $O^- - (O^-)^*$ of the two-electron oxygen state. This conclusion is supported by the DH finding that this spectrum is observable *only* in n-type crystals containing oxygen, although this argument is not conclusive.[9] Since these zero-phonon transitions are optically allowed, strong phonon replicas are generated only by A_1-type C-C modes to which the state is coupled. There are two of these modes, B' and C', identified by DH in their Fig. 6(a) and listed in their Table II. The primes have been added to distinguish them from the non-A_1 active modes B and C found in the electron-capture spectrum, Table I of DH.) The energies and approximate coupling strengths are $\hbar\omega_{B'} = 46.42 \pm 0.07$ meV, $\lambda_{B'} \approx 0.4$, and $\hbar\omega_{C'} = 49.1 \pm 0.1$ meV, $\lambda_{C'} \approx 0.2$. There is, again, *no* indication of the prominant 19-meV mode associated with O^0, thus supporting the identification of this as an O^- spectrum! Further, the persistent phonon in the luminescence spectrum, Fig. 1 and Table I of DH, which has an energy of $\hbar\omega_0 = 47.53 \pm 0.09$ meV and a strength of $\lambda_0 \approx 0.6$, is simply the unresolved sum of B' and C'. The agreement in energy and strength of the phonons in these luminescence and excitation spectra confirms that the deep states involved are the same in both.

The above interpretation of the data provides a simple resolution of several anomalies reported in the DH paper. One of these was the quenching with increasing acceptor concentration of the electron-capture luminescence relative to the oxygen-acceptor pair luminescence. It was expected that the capture luminescence would be proportional to the pair luminescence, since electrons captured radiatively on the O donors should then recombine with holes on acceptors. Instead, DH found that with increased zinc doping the pair luminescence became dominant while the capture luminescence disappeared. Somehow, the electrons were being captured on the O donors without emitting the characteristic spectrum.

The resolution of this problem is now clear. Electrons are captured with no identified radiation [possibly through a deeper $(O^0)^*$ level] into the O^0 ground state from which they recombine rapidly with bound holes. Thus the neutral oxygen concentration is too low to give an observable rate of capture into the O^- state, which is the *radiative* one.

Further, the absence of saturation of the capture luminescence under strong excitation, for which the D-A pair luminescence saturates, is also easily understood. The flow of electrons through the radiative O^- channel—$O^0 + e^- \rightarrow O^- + h\nu$, followed by subsequent annihilation of one of the bound electrons by a hole—is independent of the pair process and is not limited by the small saturable transition probabilities of pair recombination.

By substituting O^{18} for the normal O^{16} isotope in their samples, DH were able to shift the 0.841-eV emission line by 0.67 ± 0.05 meV to lower energies— an amount equal, within experimental error, to the corresponding shift (0.71 ± 0.02

meV) to higher energy of the pair spectra reported by Dean, Henry, and Frosch in Ref. 7. This agreement suggested that the oxygen states observed in the two experiments were the same. This argument is not conclusive, however, because a similar reduction in the local elastic constants for the active oxygen modes (and, hence, a similar isotope shift) is expected from the binding of each (antibonding) electron in a deep state.

It is useful to construct models for the C-C modes, since these can give information about the localization of the electronic states. Briefly, one expects these to be breathing-type modes involving the nearer neighbors, with the lower-energy 19-meV mode consisting mainly of vibration of the four (heavier) nearest-neighbor Ga ions and with the 48-meV modes involving mainly the twelve second-nearest-neighbor P ions. Thus, the 19-meV mode is excited predominantly by a change in electron density in the immediate vicinity of the O impurity and couples to the deep O^0 ground state, as experiment confirms. The absence of this mode in the O^- spectra indicates that capture of the second electron does not alter appreciably the charge density in the "central cell" region and that the second electron is localized near the four nearest-neighbor Ga ions. The 48-meV modes appear in both the O^0 and O^- spectra, in agreement with this model and with the interpretation of these as "phosphorus" modes. The weaker coupling to these (by about half) for the second electron is consistent with a more extended wave function. A simple electrostatic model confirms that the coupling strengths are all reasonable.

The lowest excited two-electron states $(O^-)^*$ are expected to contain one electron in a localized state—similar to O^0— and one in an extended (effective mass) multivalley state. Hence, the ionization threshold of O^- should lie slightly above the zero-phonon lines of Figs. 1 and 6(a) of DH at ~ 0.9 eV. These conclusions agree well with the recent low-temperature photoionization experiments by Samuelson and Monemar[10] in which they established that the threshold for *double* ionization of the O^- state is about 1.8 eV, or ~ 0.9 eV above the O^0 ionization threshold, and that the corresponding phonon coupling is not large.

The evidence cited above confirms that the ionization energy of the second electron in O^- is about 0.9 eV, a value also consistent with the thermal activation energy of 0.89 eV (Ref. 4) and with optical values of 0.8–0.9 eV found by Ledebo and Ovren.[11] These values are in contradiction, however, with the interpretation of Grimmeiss *et al.*[5] (which assigns a photoionization threshold at 0.65 eV to oxygen) as well as with that of Henry and co-workers[2-4] (which admits of no significant thresholds below 2.0 eV).

Henry and Lang[4] have pointed out in their Appendix A that small concentrations of deep centers other than oxygen may account for some of the thresholds observed in GaP by Grimmeiss *et al.*[5] The evidence cited above appears to confirm this explanation, at least for the lowest 0.65-eV threshold. In a reinterpretation of these data, which will be published separately, I note that the weak 0.86- and 1.38-eV thresholds probably correspond to transitions from O^- into the X_1 and Γ_1 minima, while the 1.77- and 2.15-eV thresholds involve two-electron transitions.[6] Of the latter, the last (and strongest) agrees with the only transition ($E_{n_2} = 2.15$ eV at 300 K)[4] recognized in Ref. 4 as belonging to O^-, while the first matches the two-electron structure of Ref. 10.

It is apparent from the above discussion, as well as from Fig. 11 of Ref. 4, that the oscillator strengths are abnormally small for all O^- photoionization transitions within about 1 eV of the threshold, and that the rate does not become appreciable at room temperature until the "two-electron" transitions begin near 1.77 eV. This absence of the low-energy oscillator strength cannot be attributed to strong electron-phonon coupling, as has been proposed,[2-4] because of the small value of the coupling strength ($\lambda \approx 0.6$) established above. For the same reason, the usual explanation of the strong temperature dependence of the $\sigma_{n_2}{}^0$ photocapacitance data (and the somewhat weaker dependence found for $\sigma_{p_2}{}^0$) in terms of the electron-lattice coupling[4] is unacceptable.

These facts lead to the important conclusion that the temperature dependence of such spectra is not, by itself, a reliable indicator of the C-C electron-phonon coupling strength and that it must, therefore, have another explanation. I propose that the evidence strongly supports the suggestion of Refs. 5 and 6 that it is the smallness of the electronic matrix element which suppresses the one-electron transitions (those below about 1.8 eV) and that the relaxation of this "selection rule" by phonon cooperation as the temperature is raised gradually turns on the one-electron transitions and accounts for the strong temperature dependence. Note, further, that these properties are not adequately accounted for

by the recent theories of Jaros and co-workers.[12] I shall consider possible origins of these effects and of the large two-electron binding energy in a later publication.

The most compelling of the many facts supporting this new interpretation are the anomalous phonon structure in the DH spectra and the agreement with the two-electron excitation spectra of Samuelson and Monemar.[10] On the other side one could cite the argument of Dean[9] and the similarities found by Carter *et al.*[13] between the excited-state energies observed by DH and those expected and observed for shallow donors in GaP. It is hoped that this Letter will stimulate the crucial experiments which will clarify the nature of this most peculiar deep center.

It is a pleasure to acknowledge many valuable discussions with L. Samuelson and with L. -Å. Ledebo, B. Monemar, and C. Ovren and to thank Professor H. G. Grimmeiss and the Department of Solid State Physics of the Lund Institute of Technology for their hospitality and support during my visit there, where a part of this work was performed. I am indebted to P. J. Dean for helpful critical comments. This work has been supported in part by the U. S. Air Force Office of Scientific Research under Contract No. F49620-77-C-0005.

---

[1]P. J. Dean and C. H. Henry, Phys. Rev. 176, 928 (1968).

[2]H. Kukimoto, C. H. Henry, and F. R. Merritt, Phys. Rev. B 7, 2486 (1973).

[3]C. Henry, H. Kukimoto, G. L. Miller, and F. R. Merritt, Phys. Rev. B 7, 2499 (1973).

[4]C. Henry and D. V. Lang, Phys. Rev. B 15, 989 (1977).

[5]H. G. Grimmeiss, L.-Å. Ledebo, C. Ovren, and T. N. Morgan, in *Physics of Semiconductors, Proceedings of the Twelfth International Conference*, edited by M. H. Pilkuhn (Teubner, Stuttgart, 1974), p. 386.

[6]T. N. Morgan, J. Electron. Mat. 4, 1029 (1975).

[7]P. J. Dean, C. Henry, and C. J. Frosch, Phys. Rev. 168, 812 (1968).

[8]B. Monemar and L. Samuelson, J. Lumin. 12/13, 507 (1976), and to be published.

[9]P. J. Dean reports (private communication) that this spectrum could not be seen in intentionally *n*-doped crystals (though, for moderate doping, the capture luminescence could) and that this is consistent with the original interpretation. Unfortunately, this observation alone does not distinguish between the two models, since the shallow second electron in the long-lived excited $(O^-)^*$ state could easily tunnel to unoccupied donors in the doped crystals and quench the luminescence. However, under intense (band gap) photoexcitation, which fills most of the shallow donors, the weak radiative transition would dominate.

[10]L. Samuelson and B. Monemar, to be published.

[11]L.-Å. Ledebo and C. Ovren, private communication.

[12]M. Jaros and S. F. Ross, in *Physics of Semiconductors, Proceedings of the Twelfth International Conference*, edited by M. H. Pilkuhn (Teubner, Stuttgart, 1974), p. 401; M. Jaros, J. Phys. C 8, 2455 (1975); M. Jaros and S. Brand, in *Physics of Semiconductors, Proceedings of the Thirteenth International Conference*, edited by F. G. Fumi (Marves, Rome, 1976), p. 1090.

[13]A. C. Carter, P. J. Dean, M. S. Skolnick, and R. A. Stradling, to be published.

MULTIBAND AND MULTIVALLEY EFFECTIVE-MASS THEORY FOR IMPURITIES IN SEMICONDUCTORS

Sokrates T. Pantelides

Institute of Theoretical Physics, University of Lund, Lund, Sweden*
and IBM T.J. Watson Research Center
Yorktown Heights, New York 10598 USA

Typed by Linda P. Rubin (SP.2253)

ABSTRACT It is shown that the multivalley probiem for donors in Si and GaP is intrinsically a two-band problem. One-band multivalley equations for Ge and two-band multivalley equations for Si and GaP are derived by evaluating all terms to the same order. Intervalley kinetic-energy matrix elements, for which conflicting accounts appeared recently, are evaluated explicitly and shown to provide an excellent measure of the validity of the theory. The retention of Umklapp terms in potential-energy matrix elements, which has recently been demonstrated to be important for quantitative calculations, gives rise to additional terms, not considered previously, which are of the same order as other terms already included. It also leads to the requirement that the impurity pseudopotential must be internally consistent with the host-crystal pseudopotential. As a consequence, when an empirical-pseudopotential band-structure calculation is used, the screened point-charge potential no longer corresponds specifically to the isocoric impurity. Overall, the construction of appropriate impurity pseudopotentials is a more sensitive task than in earlier theories that neglected Umklapp terms.

---

*Temporary sabbatical address until Dec. 31, 1978.

Recently, a number of authors[1-4] independently showed that the conventional one-band multivalley effective-mass (EM) equations for impurities in semiconductors, which have been widely used[5-10] to study donors in Si and Ge, have a major shortcoming: The Umklapp terms in the intervalley potential-energy (PE) matrix elements, which were left out in previous theories, are in fact dominant. Shindo and Nara[1] further suggested that intervalley kinetic-energy (KE) matrix elements should be completely absent and used the resulting equation for some preliminary calculations with model impurity potentials describing shallow donors in Si. Altarelli et al.[3] adopted the same point of view about intervalley KE terms and carried out calculations using the screened point-charge potential, which was taken to be a good approximation for the impurity potential of isocoric impurities (Si:P and Ge:As), as shown earlier in the no-Umklapp theory of Ref. 9. Herbert and Inkson,[4] on the other hand, pointed out that the intervalley KE terms are not rigorously zero and suggested a new expression for them. They made use of model impurity potentials and presented results which demonstrated that the intervalley KE terms can play a crucial role in variational calculations.

In this paper we report the results of a systematic study of multivalley EM equations. In particular, we focus on the following results:

1. We derive one-band multivalley EM equations by evaluating all matrix elements to the same order. We show that, when Umklapp terms are retained, additional terms appear in the PE matrix elements which have not been considered before, but are of the same order as other terms already included. Intervalley KE matrix elements are not rigorously zero, but are negligibly small for shallow donors in Ge. Further, it is shown that the approximations of the theory break down when the intervalley KE terms become appreciable.
2. We point out that the multivalley problem for Si and GaP is intrinsically a two-band problem. We derive the appropriate two-band multivalley EM equations and discuss their solutions. For Si, these equations can be cast in a form which is analogous to that of one-band multivalley equations with an explicit intervalley KE term. They can be further reduced to those used in Refs. 1-4 if additional approximations are made. Finally, the equations reduce to the usual one-band one-valley EM equation[11] when all

interband, intervalley and Umklapp terms are left out. In contrast, for GaP, the two-band equations cannot be reduced to simpler forms.

3. We point out that, when Umklapp terms are retained, the impurity pseudopotential must satisfy a requirement which we refer to as internal consistency. As a consequence, when an empirical-pseudopotential band structure is used, the screened point-charge potential does not necessarily correspond to the impurity pseudopotential of the isocoric impurity, but could yield any binding energy within the range of shallow-donor binding energies. We conclude that the construction of impurity pseudopotentials and the solution of the equations are far more demanding tasks than they were in the case of previous no-Umklapp theories.[6-11]

We have developed one-band multivalley EM equations, which are appropriate for Ge, and two-band multivalley EM equations, which are appropriate for Si and GaP. In both cases, we start with a one-electron local-density[12] Hamiltonian H+U, where H is the perfect-crystal Hamiltonian and U is the impurity potential, and make only two additional approximations. It then becomes apparent what additional approximations and assumptions are necessary to reduce our results to any of the previous theories.

We begin by assuming that the perfect-crystal band-structure problem

$$H\,\psi_{n\underset{\sim}{k}}(\underset{\sim}{r}) = E_n(\underset{\sim}{k})\,\psi_{n\underset{\sim}{k}}(\underset{\sim}{r}) \tag{1}$$

has been solved. We then expand the bound-state wavefunction $\phi(\underset{\sim}{r})$ in the form

$$\phi(\underset{\sim}{r}) = \sum_{n\underset{\sim}{k}} F_n(\underset{\sim}{k})\;\psi_{n\underset{\sim}{k}}(\underset{\sim}{r}). \tag{2}$$

We treat the $F_n(\underset{\sim}{k})$ as variational functions and without loss of accuracy or generality we choose them to be of the form

$$F_n(\underset{\sim}{k}) = \sum_{\alpha=1}^{N} \lambda_n^\alpha F_n^\alpha(\underset{\sim}{k}), \tag{3}$$

where $F_n^\alpha(\underset{\sim}{k})$ is centered about $\underset{\sim}{k}_n^\alpha$, which are a few selected points in the Brillouin zone [e.g. where $E_n(\underset{\sim}{k})$ has its extrema]. The coefficients λ_n^α can be determined by symmetry only when the points $\underset{\sim}{k}_n^\alpha$ are equivalent.

The first approximation we wish to make in the case of donors is to retain only the lowest band n=1 and only the $F_1^\alpha(\underset{\sim}{k})$ that correspond to the N equivalent absolute minima. The criterion for the validity of this assumption is easily estab-

lished by perturbation theory to be $E_B/\Delta E << 1$, where E_B is the binding energy. In the case of subsidiary minima, ΔE is the energy separation between the minima, whereas in the case of other bands, ΔE is the minimum energy separation (direct or indirect) from the absolute minima.

In the case of Ge, the above assumptions are valid for shallow donors and we proceed to develop one-band multivalley EM equations. We immediately get for the expectation value of the energy

$$E[F] = \sum_{\alpha}\sum_{\beta} \lambda_1^{\alpha}\lambda_1^{\beta}\,[<H>^{\alpha\beta} + <U>^{\alpha\beta}] \;/\; \sum_{\alpha}\sum_{\beta} \lambda_1^{\alpha}\lambda_1^{\beta} S^{\alpha\beta} \quad (4)$$

where

$$<H>^{\alpha\beta} = \sum_{\underset{\sim}{k}} F_1^{\alpha}(\underset{\sim}{k})^{*} <\psi_{1\underset{\sim}{k}} | H | \psi_{1\underset{\sim}{k}}> F_1^{\beta}(\underset{\sim}{k}), \quad (5)$$

$$<U>^{\alpha\beta} = \sum_{\underset{\sim}{k}}\sum_{\underset{\sim}{k}'} F_1^{\alpha}(\underset{\sim}{k})^{*} <\psi_{1\underset{\sim}{k}} | U | \psi_{1\underset{\sim}{k}'}> F_1^{\beta}(\underset{\sim}{k}'), \quad (6)$$

and

$$S^{\alpha\beta} = \sum_{\underset{\sim}{k}} F_1^{\alpha}(\underset{\sim}{k})^{*} F_1^{\beta}(\underset{\sim}{k}). \quad (7)$$

We distinguish between intravalley ($\alpha=\beta$) and intervalley ($\alpha\neq\beta$) terms. Note that in general[13] $S^{\alpha\beta} \neq 0$ for all α, β.

We now proceed to evaluate eqs. (5) and (6) to the same order.[14] For this purpose, $\psi_{1\underset{\sim}{k}}(\underset{\sim}{r})$ is expanded about appropriate points $\underset{\sim}{K}_{\nu}$ using $\underset{\sim}{k}\cdot\underset{\sim}{p}$ theory.[15] The leading terms are

$$\psi_{1\underset{\sim}{k}}(\underset{\sim}{r}) \simeq \psi'_{1\underset{\sim}{k}}(\underset{\sim}{r}) = \Big[\psi_{1\underset{\sim}{K}_{\nu}}(\underset{\sim}{r}) + (\underset{\sim}{k}-\underset{\sim}{K}_{\nu})\cdot\sum_{n\neq 1}\frac{\underset{\sim}{p}_{1n}}{E'_{1n}}\psi_{n\underset{\sim}{K}_{\nu}}(\underset{\sim}{r})\Big] e^{i(\underset{\sim}{k}-\underset{\sim}{K}_{\nu})\cdot\underset{\sim}{r}}, \quad (8)$$

where we use atomic units,

$$\underset{\sim}{p}_{1n} = <u_{n\underset{\sim}{K}_{\nu}} | \underset{\sim}{p} | u_{1\underset{\sim}{K}_{\nu}}> = -i<u_{n\underset{\sim}{K}_{\nu}} | \underset{\sim}{\nabla} | u_{1\underset{\sim}{K}_{\nu}}> \quad (9)$$

[$u_{n\underset{\sim}{k}}(\underset{\sim}{r})$ is the cell-periodic part of $\psi_{n\underset{\sim}{k}}(\underset{\sim}{r})$], and $E'_{1n} = E_1(\underset{\sim}{K}_{\nu}) - E_n(\underset{\sim}{K}_{\nu})$. Our second approximation is that $F_1^{\alpha}(\underset{\sim}{k})$ is sufficiently localized about $\underset{\sim}{k}_1^{\alpha}$ to allow the use of $\underset{\sim}{k}\cdot\underset{\sim}{p}$ expansions about these points. Such expansions allow us to evaluate all but the intervalley KE matrix elements to order[11] $(\underset{\sim}{k}-\underset{\sim}{k}_{\alpha})^2$ [from now on we use $\underset{\sim}{k}_{\alpha}$ in place of $\underset{\sim}{k}_1^{\alpha}$ for clarity]. We define

$$f_1^{\alpha}(\underset{\sim}{r}) = \sum_{\underset{\sim}{k}} F_1^{\alpha}(\underset{\sim}{k})\, e^{i(\underset{\sim}{k}-\underset{\sim}{k}_{\alpha})\cdot\underset{\sim}{r}} \quad (10)$$

and get

$$<H>^{\alpha\alpha} = <f_1^{\alpha} | T^{\alpha}(-i\nabla) | f_1^{\alpha}>, \quad (11)$$

and

$$<U>^{\alpha\beta} = <f_1^{\alpha} | \tilde{U}^{\alpha\beta} | f_1^{\beta}>, \quad (12)$$

where

$$T^{\alpha}(-i\underset{\sim}{\nabla}) = -\frac{1}{2}\nabla^2 - \sum_{n\neq 1} |\underset{\sim}{p}_{1n}\cdot\underset{\sim}{\nabla}|^2 / E^{\alpha}_{1n}, \qquad (13)$$

[usually written in the form[16]

$$T^{\alpha}(-i\underset{\sim}{\nabla}) = -\sum_i (1/2m^*_i)(\partial^2/\partial x_i^2) \qquad (14)$$

with m^*_i being the effective masses at the α^{th} extremum], and

$$\tilde{U}^{\alpha\beta} = \psi^*_{1\underset{\sim}{k}_\alpha}(\underset{\sim}{r})U(\underset{\sim}{r})\,\psi_{1\underset{\sim}{k}_\beta}(\underset{\sim}{r}) +$$

$$-i\,[\psi^*_{1\underset{\sim}{k}_\alpha}(\underset{\sim}{r})U(\underset{\sim}{r})\sum_{n\neq 1}(\underset{\sim}{p}_{1n}/E^{\beta}_{1n})\psi_{n\underset{\sim}{k}_\beta}(\underset{\sim}{r})]\cdot\underset{\sim}{\nabla}$$

$$+i\,\underset{\sim}{\nabla}\cdot[\sum_{n\neq 1}(\underset{\sim}{p}^*_{n1}/E^{\alpha}_{1n})\,\psi^*_{n\underset{\sim}{k}_\alpha}(\underset{\sim}{r})U(\underset{\sim}{r})\psi_{1\underset{\sim}{k}_\beta}(\underset{\sim}{r})]. \qquad (15)$$

The second and third terms in (15), which have not been considered before, must be retained because they are of the same order[17] as the second term in (13), which is responsible for converting the free-electron mass to m^*. Their actual contribution will of course depend on the particular choice of $U(\underset{\sim}{r})$. In the limit of assuming a weak potential U and dropping all Umklapp[19] and intervalley terms,[11] (15) reduces to simply $U(\underset{\sim}{r})\delta_{\alpha\beta}$, and the one-band one-valley EM equation[11] follows.

Finally, for the intervalley KE matrix elements, which are clearly not rigorously zero, expansion of $\psi_{1\underset{\sim}{k}}(\underset{\sim}{r})$ about either $\underset{\sim}{k}_\alpha$ or $\underset{\sim}{k}_\beta$ would not be correct unless $(\underset{\sim}{k}-\underset{\sim}{k}_\alpha)^4$ and higher-order terms were kept. Instead, we note that the sum over $\underset{\sim}{k}$ is dominated by values of $\underset{\sim}{k}$ where $F^{\alpha}_1(\underset{\sim}{k})\,E_1(\underset{\sim}{k})\,F^{\beta}_1(\underset{\sim}{k})$ is large. A convenient point of expansion for $\psi_{1\underset{\sim}{k}}$ is therefore the midpoint between $\underset{\sim}{k}_\alpha$ and $\underset{\sim}{k}_\beta$, which we denote by $\underset{\sim}{k}_{\alpha\beta}$. Using eq. (8) with $\underset{\sim}{K}_\nu = \underset{\sim}{k}_{\alpha\beta}$, and keeping the lowest order, we get

$$\langle H\rangle^{\alpha\beta} = E_1(\underset{\sim}{k}_{\alpha\beta})\,S^{\alpha\beta},\ \alpha\neq\beta. \qquad (16)$$

In Ge, we can use one-valley envelope functions to estimate $S^{\alpha\beta}$ and find that $\langle H\rangle^{\alpha\beta}$ is of order 10^{-5} meV and thus totally negligible. In other materials with larger effective masses, the value of $\langle H\rangle^{\alpha\beta}$ given by eq. (16) can be appreciable. Higher-order terms would then have to be included in eq. (16), which, in general, reduce the value of $\langle H\rangle^{\alpha\beta}$. However, an estimate of the $(\underset{\sim}{k}-\underset{\sim}{k}_\alpha)^4$ and higher-order corrections to the intravalley KE matrix elements shows[18] that they are also of order $\eta E_1(\underset{\sim}{k}_{\alpha\beta})S^{\alpha\beta}$, where η is of order unity. Therefore, in each application, evaluation of $\langle H\rangle^{\alpha\beta}$, $\alpha\neq\beta$, using eq. (16) provides an excellent measure of the uncertainty of the calculation. Clearly, when it becomes appreciable, the $(\underset{\sim}{k}-\underset{\sim}{k}_\alpha)^4$ terms become important and a higher-order theory would have to be

developed or the procedure abandoned in favor of other methods.[20,21]

We now turn to Si. In Fig. 1(a) we show the lowest conduction band of Si along the (001) direction in the standard Brillouin zone centered at Γ. The dashed lines are a second band, which is not explicitly included in the derivation of the one-valley EM equation[11] since it is well-separated from the minimum at $\underset{\sim}{k}_0$. In a multivalley theory, however, it is the indirect interband separation that is relevant, so that Fig. 1(a) suggests that a two-band formulation may be necessary. The importance of the second band becomes even more compelling if we view the Si bands from a Brillouin zone centered at X, as shown in Fig. 1(b). It immediately becomes clear that the multivalley problem in Si is intrinsically a two-band problem.

The two-band theory is developed in two stages. First, states $\phi^\alpha(\underset{\sim}{r})$ are constructed from the two bands around the α^{th} X point, X_α. For this purpose, degenerate $\underset{\sim}{k}\cdot\underset{\sim}{p}$ theory[15] about X_α is employed, keeping terms up to order $(k-X_\alpha)^2$. The resulting dispersion describes the two-valley (divale) structure more accurately than individual expansions to second order about the two minima.[22] Depending on the choice of the two degenerate Bloch functions at X_α, different but equivalent EM equations are obtained. Making use of the full symmetry at the X point, the resulting single-divale EM equations are

$$[T_q(-i\nabla)I+T_c(-i\nabla)\sigma_z + T_\ell(-i\nabla)\sigma_y + \tilde{U}^{\alpha\alpha}]\, f^\alpha(\underset{\sim}{r}) = E\, f^\alpha(\underset{\sim}{r}), \tag{17}$$

where the zero of energy is now at the X point, I is the 2×2 unit matrix, σ_y and σ_z are Pauli spinors, and $f^\alpha(\underset{\sim}{r})$ is a two-component column vector.[23] The remaining quantities are defined by

$$T_q(\underset{\sim}{k}-\underset{\sim}{X}) = \frac{1}{2m_\perp}(k_x^2+k_y^2) + \frac{1}{2m_\parallel}(k_z-X)^2, \tag{18}$$

$$T_c(\underset{\sim}{k}-\underset{\sim}{X}) = \frac{1}{m_c}k_xk_y, \tag{19}$$

and

$$T_\ell(\underset{\sim}{k}-\underset{\sim}{X}) = \frac{1}{m_\parallel}(k_o-X)(k_z-X), \tag{20}$$

where we have chosen $X_\alpha = (001)(2\pi/a)$. Here $m_\perp$ and $m_\parallel$ are the usual transverse and longitudinal effective masses, respectively, at each minimum, and m_c is an interband effective mass which must either be calculated from the appropriate $\underset{\sim}{k}\cdot\underset{\sim}{p}$

expression or fit to calculated energy bands.[18] The operator $\tilde{U}^{\alpha\alpha}$ is also a 2×2 matrix, where $\tilde{U}^{\alpha\alpha}_{ij}$ is an appropriate generalization of eq. (15). If the interband KE term given by eq. (19) and all Umklapp terms are neglected, then $\tilde{U}^{\alpha\alpha}_{ij} = \delta_{ij}\, U(r)$, the two bands are decoupled exactly, and (17) reduces to two equations, one at each valley, which have the standard one-valley form[11].

The solutions of (17) may be classified as symmetric and antisymmetric with respect to z. Symmetric solutions can be obtained by taking $f^{\alpha}_1(\underset{\sim}{r})$ to be an even function and $f^{\alpha}_2(\underset{\sim}{r})$ to be an odd function. For antisymmetric solutions the opposite choice must be made. The difference in the energies of the two types of solutions arises mainly from the fact that, to leading order, $\tilde{U}^{\alpha\alpha}_{11}(r) \sim \cos^2 Xz$ and $\tilde{U}^{\alpha\alpha}_{22}(r) \sim \sin^2 Xz$. Symmetric states, therefore, sample the full strength of the potential near the origin and would be deeper for an attractive core.

In order to complete the theory for Si, eigensolutions $\phi(\underset{\sim}{r})$ are constructed by writing

$$\phi(\underset{\sim}{r}) = \sum_{\alpha=1}^{3} \lambda^{\alpha}\ \phi^{\alpha}(\underset{\sim}{r}) \qquad (21)$$

Symmetric ϕ^{α}'s which are non-degenerate in the space of a single X_{α} give rise to a singlet A_1 and a doublet E, whereas antisymmetric, similarly non-degenerate ϕ^{α}'s give rise to a triplet T_2.[24] The three λ^{α} for the latter are of the form 0,0,1, so that a single-divale theory is adequate. For A_1 and E states, multidivale equations are constructed following the same procedure used for the one-band multivalley theory with obvious generalizations.[18]

An alternative form of the two-band equations for Si can be obtained by making use of the symmetry present along the (001) axis instead of the full symmetry at X. The resulting equations can then be cast in a form that is analogous to the one-band multivalley equations derived above for Ge, except that now the term $T_c(\underset{\sim}{k}-\underset{\sim}{X})$ of eq. (19) appears explicitly as the intervalley KE for the two valleys on the same (001) axis.[18] The particular form of the one-band multivalley equations used in Refs. 1-4 can then be recovered by omitting T_c as well as the second and third terms in eq. (15) in all the PE matrix elements, and making some additional minor approximations.[25] The importance of the omitted terms and of the other approximations depends on the host band structure and on the choice of impurity pseudopotential. Before we address these subjects, however, we first complete our derivations of multivalley EM equations.

The two-band multidivale EM equations derived above for Si can immediately be extended to describe the camel's-back extrema of GaP. The lowest conduction bands of GaP near the X point look like those of Si (Fig. 1b) except that the states at X are split by the antisymmetric part of the crystal potential giving rise to a camel's-back shape (Fig. 1c).[26] The appropriate single-divale equation is

$$[T_q(-i\nabla)I+T_c(-i\nabla)\sigma_z + T_\ell(-i\nabla)\,\sigma_y + \frac{\Delta}{2}\sigma_x + \tilde{U}^{\alpha\alpha}]\, f^\alpha(\underset{\sim}{r}) = Ef^\alpha(\underset{\sim}{r}), \qquad (22)$$

where T_q, T_c and T_ℓ are the same as in eq. (17), the σ's are Pauli spinors, and Δ is the band splitting at X. We note that even if Δ is large compared with E_B, and even if Umklapp terms are dropped so that $\tilde{U}^{\alpha\alpha}_{ij}(\underset{\sim}{r}) \sim \delta_{ij}\, U(\underset{\sim}{r})$, eq. (22) cannot be decoupled into two one-band equations.[27] The generalization to a multidivale theory is straightforward.[18]

We turn now to the choice of impurity pseudopotentials which must be used with the above equations. For substitutional impurities, one may write

$$U(\underset{\sim}{r}) = v_i(\underset{\sim}{r}) - v_o(\underset{\sim}{r}) + \text{screening}, \qquad (23)$$

where $v_i(\underset{\sim}{r})$ and $v_o(\underset{\sim}{r})$ are pseudopotentials for the impurity and host ion, respectively. In a theory which neglects all Umklapp terms,[28] one must simply make sure that the choice of pseudopotentials is such that $U(\underset{\sim}{r})$ is weak for the EM approximations to be valid.[29] When Umklapp terms are included in the EM equations, however, one also needs Bloch functions for the host crystal. Since Bloch functions depend on the choice of crystal pseudopotential, internal consistency[29] requires that the crystal pseudopotential be given by

$$V(\underset{\sim}{r}) = \sum_j v_o(\underset{\sim}{r}-\underset{\sim}{R}_j) + V_v(\underset{\sim}{r}), \qquad (24)$$

where $v_o(\underset{\sim}{r}-\underset{\sim}{R}_j)$ is <u>the same</u> as in eq. (23). Here $V_v(\underset{\sim}{r})$ is the potential of the valence-electron cloud. The importance of violating the requirement of internal consistency in numerical calculations has not been investigated.[29] The requirement, however, limits the usefulness of the screened point-charge potential $U_{pc}(r)$. It was shown in Ref. 9 that $U_{pc}(r)$ is a good approximation for an isocoric impurity if one works within a true-potential representation or the Cohen-Heine[30] optimized-pseudopotential representation where both $v_o(\underset{\sim}{r})$ and $v_i(\underset{\sim}{r})$ retain their Coulombic spikes at the origin. If, however, the

host band-structure calculation is carried out using a model potential, as in the empirical-pseudopotential method,[31] $v_o(\underset{\sim}{r})$ is fixed[32] and $U_{pc}(r)$ cannot be unambiguously identified with the impurity pseudopotential of any specific shallow donor. Instead, its binding energy will depend on the choice of pseudopotential used for the band-structure calculation that yields the necessary Bloch functions. For commonly used choices, the calculations of Ref. 3 suggest that the resulting binding energies would lie in the range of experimental binding energies of shallow donors.

The task that lies ahead, therefore, is the construction of impurity pseudopotentials that are internally consistent with a pseudopotential band-structure calculation. Evidence that is available thus far indicates that the task is very sensitive. The calculations of Ref. 3 (see, e.g., Table 1 of Ref. 3) show that the net contribution of Umklapp terms in Si is the result of subtle cancellations. Almost the entire intervalley matrix element for valleys on the same axis comes from the $\underset{\sim}{G}=(331)$ contribution. The host potential v_o used for the band-structure calculation, however, and most commonly used pseudopotentials[31] have no Fourier components at such high $\underset{\sim}{G}$ values. In fact, in keeping with the empirical-pseudopotential spirit, Shindo and Nara[1] imposed a cutoff on the sum over $\underset{\sim}{G}$. If a similar cutoff were used in the calculations of Ref. 3,[33] the net value of the intervalley matrix element would be reduced by almost an order of magnitude. Calculations of the additional terms arising from the retention of Umklapp contributions [cf. eq. (15)] is even more sensitive, since it involves summing over bands as well.[34] Their net contribution will again depend on the choice of host and impurity pseudopotentials.

In conclusion, we have derived rigorous multivalley EM equations for impurity states in semiconductors. They can form the framework for accurate calculations if sufficiently accurate and internally consistent host and impurity pseudopotentials are constructed. If the latter task is accomplished, the calculations would still be highly demanding. In the case of shallow donors in Ge, the intervalley matrix elements which cause the ground-state splitting are of order 0.1-1.0 meV.[35] Since these quantities would have to be calculated by summing over reciprocal lattice vectors and bands, it appears that the theory will be more useful for double donors in Ge and moderately deep donors in Si and GaP (E_B ~100-300 meV). In such cases, one may have to include contributions from subsidiary minima in a straightforward generalization of the theory,[18]

while uncertainties of order 10-30 meV would not be important.

It is a pleasure to acknowledge valuable discussions on questions of symmetry with J.F. Janak, and helpful discussions with J. Bernholc, T.N. Morgan, and J. Pollmann. This work has been supported in part by the Air Force Office of Scientific Research under contract No. 49620-77-C-0005.

REFERENCES

* Temporary sabbatical address (July 1 - December 31, 1978).

1. K. Shindo and H. Nara, J. Phys. Soc. Jpn. 40, 1640 (1976).
2. L. Resta, J. Phys. C 10, L179 (1977).
3. M. Altarelli, W.Y. Hsu, and R.A. Sabatini, J. Phys. C 10, L605 (1977).
4. D.C. Herbert and J. Inkson, J. Phys. C 10, L 695 (1977).
5. H. Fritzsche, Phys. Rev. 125, 1560 (1962).
6. A. Baldereschi, Phys. Rev. B 1, 4673 (1970).
7. M. Altarelli and G. Iadonisi, Nuovo Cim. B5, 21 (1971); *ibid* p. 36.
8. T.H. Ning and C.T. Sah, Phys. Rev. B 4, 3468 (1971); *ibid*, p. 3482.
9. S.T. Pantelides and C.T. Sah, Solid State Comm. 11, 1713 (1972); Phys. Rev. B 10, 621 (1974); ibid, p. 638.
10. D. Schechter, Phys. Rev. B8, 652 (1973); *ibid 11*, 5043 (1975).
11. C. Kittel and A.H. Mitchell, Phys. Rev. 96, 1488 (1954); J.M. Luttinger and W. Kohn, Phys. Rev. 97, 869 (1955); W. Kohn, Solid State Phys. 5, 257 (1957).
12. W. Kohn and L.J. Sham, Phys. Rev. 140, A 1133 (1965).
13. $F_1^{\alpha}(\underset{\sim}{k})$ and $F_1^{\beta}(\underset{\sim}{k})$ must be allowed adequate variational freedom to overlap if the impurity potential requires it.
14. In the one-band case, eq. (5) can also be evaluated exactly using eq. (1). A systematic evaluation to the same order as PE matrix elements is preferred, especially in multiband cases (e.g. acceptors, Ref. 11, and the Si and GaP donors treated in this paper).
15. E.O. Kane, in *Semiconductors and Semimetals*, ed. by R.K. Willardson and A.C. Beer, (Academic, New York, 1966), p. 75.
16. $T^{\alpha}(\underset{\sim}{k}-\underset{\sim}{k}_{\alpha})$ is the usual expansion of $E_1(\underset{\sim}{k})$ about $\underset{\sim}{k}_{\alpha}$ to order $(\underset{\sim}{k}-\underset{\sim}{k}_{\alpha})^2$.
17. The second and third terms in eq. (15) are of order $(\underset{\sim}{k}-\underset{\sim}{k}_{\alpha})$ and must be retained. (Note that the definition of m^* that appears in T^{α} also involves the terms $\sum_{n\neq 1} p_{1n}/E_{1n}$). An *additional* term, which is of order $(k-k_{\alpha})^2$, is not shown in eq. (15) because it is actually of the same order as explicit interband coupling i.e. of order $E_B/\Delta E$. In actual applications, this additional term can be evaluated and used as another quantitative measure of the validity of the theory. See Ref. 18 for further discussion.
18. S.T. Pantelides, Phys. Rev. B, to be published.
19. The first term in eq. (15) is also the Fourier transform of $\sum_{\underset{\sim}{G}} C_{\alpha\beta}(\underset{\sim}{G})\, U(\underset{\sim}{k}-\underset{\sim}{k}'+\underset{\sim}{G})$, where $\underset{\sim}{G}$ are the reciprocal lattice vectors; hence the reference to Umklapp terms ($\underset{\sim}{G}\neq 0$).
20. For a recent review of other methods, see S.T. Pantelides, Rev. Mod. Phys. 50, 797 (1978).
21. G.A. Baraff and M. Schlüter, Phys. Rev. Lett. 41, 892 (1978); J. Bernholc, N.O. Lipari, and S.T. Pantelides, Phys. Rev. Lett. 41, 895 (1978).
22. The description of the bands of Si in the vicinity of the X point have been discussed in detail by J.C. Hensel, H. Hasegawa, and M. Nakayama [Phys. Rev. 138, A225 (1965)]. In the present notation, the dispersion for the bands is given by $E_{1,2}(\underset{\sim}{k}) = T_q\,(\underset{\sim}{k}-\underset{\sim}{X}) \pm [T_c(\underset{\sim}{k}-\underset{\sim}{X})^2 + T_{\ell}(\underset{\sim}{k}-\underset{\sim}{X})^2]^{1/2}$. Along the (100) axis we recover the simpler form $E_{1,2}\,(\underset{\sim}{k}) = T_q\,(\underset{\sim}{k}-\underset{\sim}{X}) \pm T_{\ell}(\underset{\sim}{k}-\underset{\sim}{X})$, given previously by P. Lawaetz [Solid State Comm. 16, 65 (1975)].
23. $f_1^{\alpha}(\underset{\sim}{r})$ and $f_2^{\alpha}(\underset{\sim}{r})$, the two components of $f^{\alpha}(\underset{\sim}{r})$, are the two envelope functions so that, to leading order, $\phi^{\alpha}(\underset{\sim}{r}) = \Sigma_{i=1}^{2} f_i^{\alpha}(\underset{\sim}{r})\psi_{iX_{\alpha}}(\underset{\sim}{r})$ where $\psi_{1X_{\alpha}}(\underset{\sim}{r})$ and $\psi_{2X_{\alpha}}(\underset{\sim}{r})$ are the two degenerate Bloch functions at X which transform like cos(Xz) and sin(Xz), respectively.
24. These states are in one-to-one correspondence with the states obtained from a symmetry analysis of the problem in the one-band approximation. (Ref. 11).
25. E.g. dropping the Umklapp terms in the intravalley PE matrix elements and the approximations involved in defining an "Umklapp renormalization factor" (Ref. 3), which depends on U.

26. F.H. Pollak, C.W. Higginbotham, and M. Cardona, J. Phys. Soc. Jpn. 21, Suppl., 20 (1966); P. Lawaetz, Ref. 22.

27. A 2×2 matrix KE operator has been used by M. Altarelli, R.A. Sabatini, and N.O. Lipari [Solid State Comm. 25, 1101 (1978)] in a study of excitons. In that work, T_c was left out entirely and a 2×2 matrix was chosen as a convenient analytical way to describe the dispersion of one band having a camel's-back shape. Unlike the form used in that work, eq. (21) describes the full two-band problem. See Ref. 18 for a detailed discussion.

28. In the theory of Ref. 8-10, all Umklapp terms were left out, but *additional* approximations were also made. Refs. 1-4 demonstrated that each approximation is not quantitatively valid *by itself*. The success of the theory was due to cancellations between overestimated intervalley KE and PE terms, an effect which was discussed in Ref. 9 and elsewhere [S. T. Pantelides, Festkörperprobleme 15, 149 (1975); also Ref. 29]. The totality of the approximations views the Si bands as consisting of six free-electron parabolas at the six $\underset{\sim}{k}_s$ having the effective mass m^* and plane-wave Bloch functions. The model thus assumes that the details of the actual bands at higher energies and the oscillatory part of the Bloch functions do not affect the bound-state energies. The quantitative accuracy of such a model is not assured from the outset, as assumed in Refs. 8-10, but its success for a point-charge potential (Ref. 9) provides a basis with which the systematic study of the rest of the donors in Si using rigorously-constructed impurity pseudopotentials remains valid. For further discussion see Ref. 18.

29. S.T. Pantelides, Ref. 20.

30. M.H. Cohen and V. Heine, Phys. Rev. 122, 1821 (1961).

31. M.L. Cohen and V. Heine, Solid State Phys. 24, 37 (1970).

32. From an empirical-pseudopotential band-structure calculation, one can actually extract an "atomic", not "ionic" pseudopotential, which is not defined uniquely at small $\underset{\sim}{r}$ values, because the crystal pseudopotential is known only at a few reciprocal lattice vectors.

33. Such a cutoff would correspond to ignoring the Coulombic spike of $U_{pc}(r)$ at small r values in the spirit of empirical model potentials (Ref. 31).

34. Comparison of eqs. (13)-(15) reveals that the evaluation of (15) is even more sensitive than the calculation of m^* from $\underset{\sim}{k}\cdot\underset{\sim}{p}$ theory. In fact, it would be more appropriate to use eq. (13) and evaluate the sum over bands, instead of using eq. (14) with the experimental value of m^*. Repeating the calculation with (14) instead of (13) would provide an excellent test of the adequacy of the band structure quantities used in the evaluation of (15).

35. J.H. Reuszer and P. Fisher, Phys. Rev. 135, A1125 (1964).

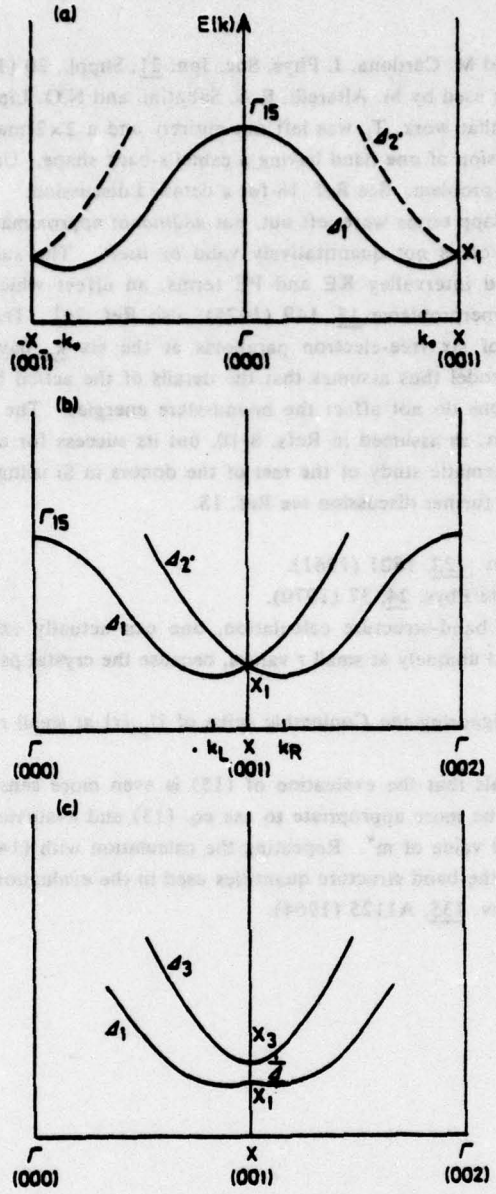


FIGURE CAPTION

Fig. 1. **(a) The standard view of the Si conduction bands in the Brillouin zone centered at Γ. The dashed curve is a second band, which is neglected in the one-valley approximation. (b) View of the same bands from a Brillouin zone centered at X, showing that the problem is intrinsically a two-band problem. (c) The camel's-back form of the bands of GaP.**

Solid State Communications, Vol. 25, pp. 665–668, 1978. Pergamon Press. Printed in Great Britain

INTERPRETATION OF ACCEPTOR SPECTRA IN SEMICONDUCTORS*

N.O. Lipari

IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

and

A. Baldereschi

Laboratoire de Physique Appliquee, EPF-Lausanne, Switzerland

(*Received* 8 *September* 1977 *by M.F. Collins*)

The acceptor Hamiltonian proposed in 1955 by Kohn and Luttinger is extended to introduce the q-dependence of electronic screening and the coupling with the split-off valence band. Accurate solutions of the resulting equations by the irreducible-tensor technique allow a good interpretation of the spectra of the shallow acceptors in Si and Ge.

DURING the last few years, the developments in the purification methods of various semiconductors have made possible higher resolution spectroscopy of impurities. The shallow acceptors in Si and Ge have received particular attention [1, 2] and the spectral position of many excited states has been determined with accuracy of 0.01 meV. On the theoretical side, even though the shallow-impurity problem has been formulated over twenty years ago by Kittel and Mitchell [3] and Luttinger and Kohn [4], good solutions to the proposed effective mass equations do not exist yet. The presently available theoretical results [5] have qualitative value only and are not adequate for the interpretation of the recent high resolution data [1, 2] and, sometimes they even fail in the description of the less accurate, older data [6–8]. For example, the excited acceptor states in Ge [6], the large ionization energy of acceptors in Si [7] and the existence of an excited state 23.4 meV above the ground state, observed in the Raman spectroscopy of Si : B [8], are not yet understood. These difficulties have often been considered as failures of the effective mass approximation.

In this letter we report accurate solutions of the acceptor effective-mass Hamiltonian, and show that how, after proper inclusion of q-dependent screening and coupling with the split-off valence band, many of the difficulties mentioned above are lifted. The present results, which have been obtained by the "irreducible-tensor technique" [5], can be considered for all practical purposes as the "exact" solution of the effective-mass acceptor.

The effective-mass Hamiltonian for acceptors in semiconductors can be written as [5, 9]:

$$H = \frac{1}{\hbar^2}p^2 - \frac{\mu}{3\hbar^2}(P^{(2)} \times I^{(2)}) + \frac{\delta}{3\hbar^2} \times \left\{ [P^{(2)} \times I^{(2)}]^4_4 + \frac{\sqrt{70}}{5}[P^{(2)} \times I^{(2)}]^4_0 + [P^{(2)} \times I^{(2)}]^4_{-4} \right\} + \frac{2}{3}\left(\frac{1}{2} - \mathbf{I}\cdot\mathbf{S}\right)\bar{\Delta} - \frac{e_\infty}{\pi^2}\int \frac{e^{iqr}}{\epsilon(q)q^2}d^3q. \quad (1)$$

This Hamiltonian includes coupling with the split-off valence band and q-dependent diagonal dieletric screening. In (1), p denotes the hole linear momentum operator, I is the angular momentum operator corresponding to spin 1, and S is the spin $\frac{1}{2}$ of the hole. The definitions for the tensor operators $P^{(2)}$, $I^{(2)}$ and their products are the same as in [5]. The parameters μ and δ describe the energy dispersion of the holes near the center of the Brillouin zone and are simply related [5] to the Kohn and Luttinger parameters γ_1, γ_2 and γ_3. The quantity $\bar{\Delta}$ is the valence-band spin-orbit splitting measured in units of the effective Rydberg where ϵ_∞ is the high frequency dielectric constant of the host crystal. Finally $\epsilon(q)$ is the q-dependent dielectric function which, for convenience, in actual calculations, has been expressed as [10].

$$\frac{1}{\epsilon(q)} = \frac{1}{\epsilon_\infty}\left[1 + \frac{A_l q^2}{B_l + q^2}\right] \quad (2)$$

where the parameters A_l and B_l in (2) are chosen so that expression (2) fits the computed values of $\epsilon(q)$.

Hamiltonian (1) has been solved with the variational technique using symmetrized combinations of the basis functions:

$$f_{L,J,F,F_z}(r)\,|L,(IS)J,F,F_z\rangle \quad (3)$$

which transform like a specific row of any given representation of the symmetry group O_h of the Hamiltonian. The angular and spin dependent factor in (3) is defined as in Edmonds [11]. L is the orbital

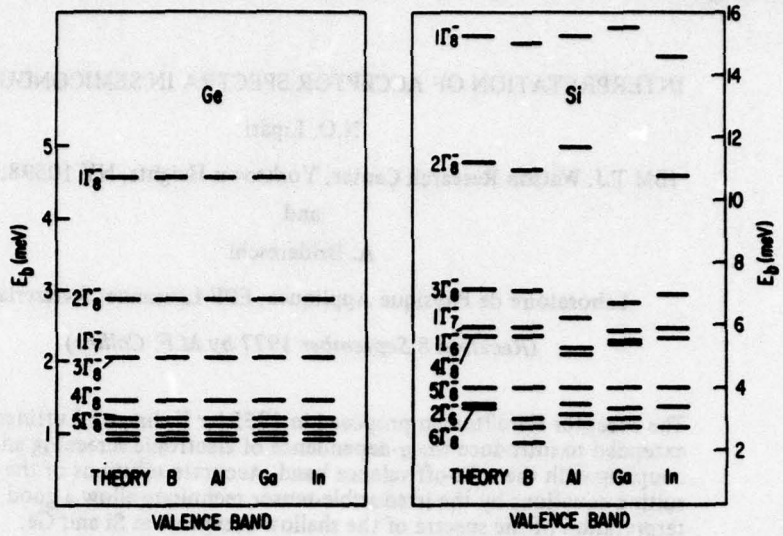


Fig. 1. Comparison between theoretical and experimental acceptor states in Si and Ge. The experimental data are those referenced in text.

Table 1. Ionization energy for single acceptors in silicon using a q *dependent and a* q *independent screening. The dielectric function used is that labelled* c *in* [10]. *The results using various truncations to the solution are shown*

L_{max}		Binding energy (meV)	
Main	Split-off	q independent screening	q dependent screening
2	0	33.3	37.6
4	0	33.8	38.5
6	0	33.9	38.7
8	0	33.9	38.7
2	2	37.8	46.7
4	4	44.3	63.3
6	6	44.3	73.2
8	8	44.4	74.3

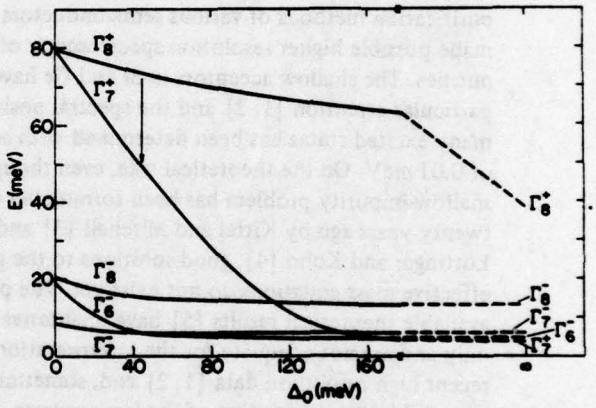


Fig. 2. Lowest acceptor states in silicon with symmetry Γ_8^+, Γ_7^+, Γ_8^-, Γ_7^- and Γ_6^- as function of the spin–orbit splitting Δ. The vertical arrow indicates the electronic transition observed in the Raman spectrum of Si : B.

angular momentum and I + S can assume the values 3/2 or 1/2 which correspond to the main and split-off valence bands, respectively. This shows how our procedure incorporates the coupling between different valence bands. The infinite set of basis functions is truncated by including all possible functions (3) with $L \leq L_{max}$. The radial functions $f_l(r)$ are expressed as linear combinations of a sufficiently large set of exponential functions with fixed exponents. Finally, the matrix elements of Hamiltonian (1) between the basis functions (3) are obtained using the "reduced-matrix-element technique" [11] and their actual evaluation is done by computer.

The spherical model previously proposed by the authors [5] considers the angular momentum F as a good quantum number, and corresponds to including in the basis functions only those orbital angular momenta L which are compatible with a given F. The resulting basis set contains at most two functions with $J = 3/2$ and one with $J = 1/2$. The small number of basis functions makes the spherical model very simple and attractive, but generally limits its validity to qualitative predictions only. The quantitative interpretation of experimental data requires the complete procedure reported in this letter. Convergence tests, which will be reported elsewhere, show that the choice of $L_{max} = 7$ is sufficient to provide results with 1% accuracy. This choice implies a basis set of 20–30 basis functions as defined in (3). We have applied the above procedure to Si and Ge. These materials, for which high resolution acceptor spectra are available [1, 2] and for which the parameters γ_1, γ_2, γ_3,

and ϵ_∞ are known very accurately, [10] are the ideal ground for comparison between theory and experiment. Figure 1 gives our theoretical predictions for the excited acceptor states in Si and Ge together with available experimental data.

The results represent the first accurate identification of the nature and symmetry of the final states of the experimentally observed infrared transitions. The relevance of the two new features introduced in the present treatment, i.e. q-dependent dieletric screening and coupling with the split-off band, is illustrated in Table 1 for the acceptor ionization energy in Si. The results shown in the table explain why, up to now, all effective mass calculations, which have neglected one or both of the above effects, did not predict binding energies of acceptors in Si larger than 54 meV [10], while the exact solution of the effective-mass Hamiltonian gives a much larger value. The value 74 meV reported in Table 1 was obtained using, for the dielectric constant, the set of parameters called c in [10]. Using the set of parameters labelled a and b in [10], we obtain 94 and 130 meV respectively. The calculated value for the binding energy is quite sensitive to the choice of the dieletric screening and an accurate determination of $\epsilon(q)$ is necessary. The experimental ionization energies for acceptor is Si are 44.4, 68.9, 72.7 and 156.2 for B, Al, Ga and In, respectively. The calculated binding energy should be compared, as shown for donors by Pantelides [12], to the isocoric acceptor in Si, i.e. Al, and we see that the theoretical value is considerably higher than the observed value. For acceptors in Ge the effective mass predicts for the ionization energy 11.2 meV and the experimental binding energy for Ge : Ga is 11.3 meV. We emphasize that screened coulomb potentials give a good description of isocoric-acceptor ground states only when accurate effective-mass solutions are used. Preliminary calculations of the acceptor binding energy in GaAs, indicate that the effective-mass prediction falls in between the ionization energies of the isocoric acceptors Ge and Zn substituting the anion and cation, respectively. A more precise prediction on this point, requires a better determination of the band parameters for this semiconductor and an accurate treatment of polaron effects.

A rather old puzzle in the interpretation of acceptor spectra in semiconductors is a mysterious line observed in the Raman spectrum of Si : B [8] and of various acceptors in GaP [13]. The difficulty in the interpretation of this line is that the final electronic state involved in the optical process is at lower energy than any of the states seen in infrared spectra, whereas existing theoretical models do not predict any state in this energy range. Our calculation for acceptors in Si predicts the existence of two-fold degenerate Γ_7^+ state 23.8 meV above the ground state. The experimental Raman line is at 23.4 meV and degeneracy and symmetry of the final state are consistent with our identification. The predicted Γ_7^+ state contains a strong mixture of d-like orbital state of the $J = 3/2$ valence band and s-like states of the split-off band. The low energy position of this state is a consequence of the small spin-orbit splitting in Si. In fact, as shown in Fig. 2, in the limit of vanishing spin-orbit splitting, the Γ_7^+ state would be degenerate with the acceptor ground state, and for large values of the spin-orbit splitting, higher than the p-like states. The analysis of the wave function of the Γ_7^+ state shows 57% s-like character from the split-off band, 31% d-like character from the $J = 3/2$ valence band, and the remaining from higher angular moments in the two bands. The large s-like component from the split-off band, explains the strong intensity of the observed Raman line. It is important to note that, even though the binding energy for the ground state in silicon strongly depends on the choice of the dielectric function used, the distance of the Raman level from the ground state is remarkably insensitive to its choice. In fact the calculated Raman level shifts by approximately the same amount (within 0.1 meV) as that of the ground state, thus maintaining the agreement reported above. In addition, Fig. 2 shows that the effect of the split-off band on the acceptor ground state is large even for large spin-orbit splittings. Furthermore even the excited states are affected by the spin-orbit splitting in Si, and the acceptor spectrum looks very different in the limit of small and large spin-orbit splitting.

In conclusion, we have shown that it is possible to obtain a good interpretation of the acceptor states in semiconductors within the effective mass approximation provided an accurate solution of these equations is obtained. This means that for s-like states one has to include in the basis set expansion terms up to $L = 6$, and up to $L = 7$ for p-like states. The inclusion of the split-off band is very important and, when considered, it allows an understanding of the acceptor Raman spectrum in silicon. Finally, the problem of an accurate determination of the ground state binding energy is still open. We have seen that small variations in the dielectric function for silicon produce large effects in this quantity. For the first time, however, the theoretical binding energy is larger than the experimental isocoric value. This is very encouraging since the use of pseudopotentials (weaker than real potentials) should now be able to provide a better description of the isocoric acceptor and of the dependence of the acceptor binding energy with the impurity atom.

*This work was supported in part by the AFOSR under contract F49620-77-C-005.

REFERENCES

1. HALLER E.E. & HANSEN W.L. *Solid State Commun.* **15**, 687 (1974).
2. SKOLNICK M.S., EAVES L., STRADLING R.A., PORTAL J.C. & ASKENAZY S. *Solid State Commun.* **15**, 1403 (1974).
3. KITTEL C. & MITCHELL A.H. *Phys. Rev.* **95**, 1488 (1954).
4. KOHN W. & LUTTINGER J.M. *Phys. Rev.* **97**, 883 (1955).
5. BALDERESCHI A. & LIPARI N.O. *Phys. Rev.* **B8**, 2697 (1973); *Phys. Rev.* **B9**, 1525 (1974); and references quoted in these two papers.
6. JONES R.L. & FISHER P. *J. Phys. Chem. Solids* **26**, 1125 (1965).
7. ONTON A., FISHER P. & RAMDAS A.K. *Phys. Rev.* **163**, 686 (1967).
8. WRIGHT G.B. & MOORADIAN A. *Phys. Rev. Lett.* **18**, 608 (1967).
9. A term linear in p should be added to (1) in the case of semiconducting compounds (see BALDERESCHI A. & LIPARI N.O. *J. Luminesc.* **12**, (1976)). The effects of this term have already been studied, but due to lack of space, the results will be presented elsewhere.
10. BERNHOLC J. & PANTELIDES S. *Phys. Rev.* **B15**, 4935 (1977).
11. EDMONDS A.R. *Angular Momentum in Quantum Mechanics.* (Princeton University Press (1960)).
12. PANTELIDES S.T. & SAH C.T. *Phys. Rev.* **B10**, 621 (1974).
13. MANCHON, Jr. D.D. & DEAN P.J. *Proc. 10th Int. Conf. Phys. Semicond.* p. 760 (1970).

PHYSICAL REVIEW B VOLUME 18, NUMBER 3 1 AUGUST 1978

Energy and polarizability of atoms in a magnetic field: Donors in silicon

N. O. Lipari
IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598

D. L. Dexter
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 27 February 1978)

The binding energy of donors as described by a simple model Hamiltonian in a magnetic field of arbitrary strength is calculated using an expansion of the wave function in terms of spherical harmonics. All the matrix elements are calculated analytically and the Hamiltonian is solved accurately. We then calculate the polarizability by applying a very small electric field. Results are given for P, As, and Sb donors in silicon and are discussed with reference to recent experimental data by Castner and Lee.

I. INTRODUCTION

Very recently,[1] Castner and Lee have performed static (actually low-frequency) magneto-capacitance measurements at low temperatures in Si with low concentration (N_D) of donor impurities, and have interpreted their results as indicating a reduction in the polarizability, $\alpha(H)$, of weakly interacting donors by a static magnetic field.

Theoretically, even though a great deal of work has been done regarding the energy of atoms and donor impurities in the effective-mass approximation (EMA)[2-4] in the presence of a magnetic field, not much exists regarding optical properties, such as dipole-matrix elements, oscillator strengths, etc.[5] In a previous paper,[6] one of us (D.L.D.) has calculated the magnetic-field dependence of the polarizability α for a hydrogenic atom in a weak magnetic field. A simple analytical expression for the ground-state wave function in the presence of a magnetic field was obtained by a variational method. The effect of an electric field, to determine the polarizability, was then calculated using a variational method introduced by Hassé.[7] Analytical expressions for $\alpha_\parallel$ $(\vec{E}\parallel\vec{H})$ and $\alpha_\perp$ $(\vec{E}\perp\vec{H})$ were given. These analytical expressions are however quite complicated, the variational minimization procedure in the presence of the external field is nonlinear, and terms of higher order than H^2 were impracticable to consider. In addition, the calculation was performed for hydrogenic atoms only, so that no impurity dependence in α could appear.

The treatment of the energy of donor impurities in multivalley semiconductors is quite complicated,[8-10] and very recently new complications have been pointed out.[11] The inclusion of external fields would make a precise treatment of the problem even more difficult, and is beyond the scope of the present investigation. In this paper we introduce an approximation to describe the spatial dependence of the ground state of the different donors; in view of the crudeness of this approximation the results must be regarded with reservation.

First, we describe a method which provides a good description of the polarizability of the hydrogenic atom for magnetic field of arbitrary strength. Section II describes this method and the solution. In Sec. III we introduce a simple model Hamiltonian for donors. An impurity-dependent potential is introduced and fixed so as to give the measured donor binding energy in the absence of an external field. Finally, in Sec. IV the results for the polarizability are presented and discussed using available experimental data.

II. HYDROGENIC ATOM: METHOD AND SOLUTION

The EMA leads[2] to the following Schrödinger equation for a hydrogenic impurity in a uniform magnetic field H_z, in the z direction:

$$\mathcal{H}_0 = -\nabla^2 - 2/r + \tfrac{1}{4}\gamma^2(x^2+y^2)\,. \tag{1}$$

Here we use the quantities

$$a^* = \epsilon\hbar^2/m^*e^2,\quad R^* = m^*e^4/2\epsilon^2\hbar^2,$$
$$\gamma = \epsilon^2\hbar^3 H_z/m^{*2}e^3c = e\hbar L_z/2m^*cR^*\,, \tag{2}$$

as the units of length, energy, and magnetic field; $\gamma=1$ is the magnetic field at which the diamagnetic energy is equal in magnitude to the Coulomb energy. [For the free hydrogen atom $\gamma=1$ would correspond to $H_z = 2.35\times10^9$ G, but for a donor impurity it would be reduced by a factor $(m^*/m_e\epsilon)^2 \sim 10^{-3}$ or less]. ϵ is the static dielectric constant, and m^* the (scalar) effective mass. For purposes of comparison with experiment we choose $m^* = 0.3m_0$ and $\epsilon = 11.4$. (The value $m^* = 0.3m_0$ has been chosen so as to give the same binding energy as that obtained in the single valley case with anisotropic masses.) In writing (1) we have omitted relativistic terms, which yield a spin-dependent polarizability. These terms have been shown to be negligible for most purposes.[6] To calculate the polarizability we introduce an additional elec-

tric-field term

$$\mathcal{H}_1 = -e a^* E_t t/R^* = -\xi t\,, \tag{3}$$

where t stands for z or x, parallel or perpendicular to the direction of the magnetic field H. The Hamiltonian in the presence of an external magnetic and electric field is therefore given by

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1. \tag{4}$$

The polarizability α is defined by

$$E(H_z, \xi) = E(H_z, 0) - \tfrac{1}{2}\alpha\xi^2, \tag{5}$$

i.e.,

$$\alpha = 2\lim_{\xi\to 0}\frac{E(H_z,0) - E(H_z,\xi)}{\xi^2}. \tag{6}$$

Depending on the direction of $\vec{\xi}$ we obtain the components $\alpha_\perp$ if $\vec{\xi}\perp\vec{H}$, and $\alpha_\parallel$ if $\vec{\xi}\parallel\vec{H}$.

We now seek the eigenfunctions of the Hamiltonian in Eq. (4). We write the solution $\psi(\vec{r})$ as

$$\psi(\vec{r}) = \sum_{i=1}^{n} f_i(r) Y_{l_i m_i}(\theta,\varphi), \tag{7}$$

i.e., we expand it in terms of spherical harmonics. In the absence of any external field, l and m are good quantum numbers and therefore there would be only one angular term in the expansion (7). Once the magnetic field is introduced, l is no longer a good quantum number but m and parity still are. Therefore for the ground state, the sum in (7) will contain only terms with fixed m and even l. When the electric field is turned on, parity is no longer a good quantum number and m will still be if the electric field lies in the same direction as the magnetic field, otherwise even m is no longer a good quantum number and the sum in (7) will contain terms with any l and m. In order to evaluate the expectation value of the Hamiltonian (4), we must calculate its matrix elements in the basis (7). To take full advantage of the present method, it is convenient to rewrite (4) in terms of irreducible components of tensor operators whose effect on the basis functions (7) can be calculated using standard angular-momentum relationships.[12] Equation (1) can be rewritten as

$$\mathcal{H}_0 = -\nabla^2 - 2/r + \tfrac{1}{12}\gamma^2(2r^2 - \sqrt{\tfrac{4}{5}}\,T_0^2), \tag{8}$$

where T_m^l represents the mth irreducible component of the tensor of rank l, as defined by Edmonds.[12] Obviously here

$$T_0^2 = \sqrt{\tfrac{4}{5}}\,(2z^2 - x^2 - y^2) = \sqrt{\tfrac{16\pi}{5}}\,r^2 Y_{2,0}. \tag{9}$$

The electric field term can be written

$$\mathcal{H}_1 = -\xi z = -\xi T_0^{(1)} = -\xi\sqrt{\tfrac{4}{3}}\,\pi r Y_{1,0} \tag{10}$$

or

$$\mathcal{H}_1 = -\xi x = -\xi\frac{T_{-1}^1 - T_1^1}{\sqrt{2}} = -\xi\sqrt{\tfrac{2}{3}}\,\pi r\,\frac{Y_{-1}^1 - Y_1^1}{\sqrt{2}}. \tag{11}$$

The following relationships can therefore be used

$$\left(\nabla^2 + \frac{2}{r}\right) f(r) Y_{l,m} = Y_{l,m}\left(\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2} + \frac{2}{r}\right) f(r), \tag{12}$$

$$T_0^2 f(r) Y_{l,m} = \sqrt{6} f(r) r^2 \sum_{l'} [(2l'+1)(2l+1)]^{1/2} \times \begin{pmatrix} 2 & L & L' \\ 0 & m & m' \end{pmatrix}\begin{pmatrix} 2 & L & L' \\ 0 & 0 & 0 \end{pmatrix} (-1)^{m'} Y_{l',m'}\,. \tag{13}$$

$$T_m^1 f(r) Y_{l,m} = f(r) r \sum_{l',m'} [(2l'+1)(2l+1)]^{1/2} \times \begin{pmatrix} 1 & L & L' \\ m'' & m & m' \end{pmatrix}\begin{pmatrix} 1 & L & L' \\ 0 & 0 & 0 \end{pmatrix} (-1)^{m'} Y_{l',m'}\,. \tag{14}$$

Using these relationships we can treat exactly and analytically the angular part of the matrix elements of (4) in the basis (7). One is therefore left with systems of radial differential equations for the $f_i(r)$. These systems can be solved very accurately by making the following ansatz,

$$f_i(r) = \sum_{j=1}^{M} C_{ij} e^{-\eta_j r}\,. \tag{15}$$

The eigenvector coefficients C_{ij}, and the corresponding eigenvalues are obtained by solving the secular determinant of

$$\sum C_{ij}(H_{ij} - E S_{ij}) = 0\,. \tag{16}$$

The size of the secular determinant in (16) is $N \times M$, where N is the number of terms in (7) and M the number of terms in (15). The parameters η_j are not treated as variational parameters. With a sufficiently large number of terms (M) describing the physically interesting spatial range of the wave function, the specific choice of the η_i's does not matter, as discussed more extensively in Sec. IV. We also note that it is not necessary to include any powers of r with each exponential in (15).

III. MODEL HAMILTONIAN FOR DONOR IMPURITIES

The first successful theoretical investigation of donor spectra in semiconductors was given by Kohn and Luttinger[13] using the EMA. In indirect-gap

TABLE I. Values of the parameter K (in atomic units), which are obtained by requiring that the calculated donor binding energies equal the experimental ones. We have used $\epsilon = 11.4$ and $m^* = 0.3m_0$. The energy unit is meV.

Material	K	E_b (calc.)	E_b (expt.)
Hydrogenic	∞	31.41	
P	0.574	45.45	45.47
As	0.515	53.70	53.69
Sb	0.608	42.68	42.69

materials, the multivalley nature of the conduction band gives rise to a multiplet structure of the donor-impurity ground state, because the intervalley components of the impurity potential remove the degeneracy between the zero-order states obtained independently for each valley.[8] This "valley-orbit" interaction is quite substantial, as it increases the binding energy of the ground state by about 35% for P in Si,[14] and should therefore be included in the theoretical description of these states. An accurate description of this effect is very difficult, however, as has long been appreciated and recently emphasized anew.[11] Furthermore, the donor-binding energy depends on the given impurity, e.g., P, Sb, and As donors in Si have different binding energies.[14] A proper analysis of these "chemical shifts" is not yet available, and we do not attempt one here. Instead, we use the simplest formulation possible which incorporates in some average way the effects described previously. We do so by introducing a phenomenological impurity dependent potential of the form

TABLE II. Energy of the hydrogen ground state as a function of the magnetic field. Comparison with Cabib *et al.* (Ref. 2) is also shown. The energy unit is the effective Rydberg.

γ	E_{1s} (present)	E_{1s} (Cabib *et al.*)
0.0	−1.000 00	−1.000 00
0.1	−0.995 05	−0.995 08
0.2	−0.980 76	−0.980 76
0.3	−0.958 37	−0.958 41
0.4	−0.929 21	−0.929 23
0.5	−0.894 42	−0.894 47
0.6	−0.854 92	−0.854 94
0.7	−0.811 45	−0.811 42
0.8	−0.764 57	−0.764 57
0.9	−0.714 74	−0.714 73
1.0	−0.662 34	−0.662 41
1.5	−0.370 71	−0.370 76
2.0	−0.044 43	−0.044 50
2.5	0.304 93	0.304 90
3.0	0.670 94	0.670 87
4.0	1.438 4	1.438 4
5.0	2.239 2	2.239 2

TABLE III. Ground-state binding energy for P, As, and Sb donor impurities in Si, using the model potential described in Sec. III. The energy unit is the effective Rydberg.

γ	P	As	Sb
0.0	1.447 29	1.709 83	1.359 08
0.2	1.437 10	1.702 09	1.347 75
0.4	1.408 21	1.679 79	1.315 86
0.6	1.364 01	1.644 95	1.267 53
0.8	1.307 69	1.599 70	1.206 41
1.0	1.241 70	1.545 83	1.135 23
2.0	0.817 36	1.188 07	0.682 62
3.0	0.301 28	0.740 47	0.138 00
4.5	−0.265 73	0.241 07	−0.457 52
5.0	−0.865 67	−0.292 55	−1.085 62
10.0	−4.123 49	−3.237 00	−4.479 14

$$V(r) = -(e^2/\epsilon r)[1 + (\epsilon - 1)e^{-Kr}] \ . \tag{17}$$

The coefficient K determines the strength of the screening of the potential V. In fact when $K \to \infty$ Eq. (17) reproduces the static screened potential, while for $K = 0$ it gives the unscreened Coulomb potential. The value of K for each impurity is determined by requiring the calculated binding energy to agree with the observed one for that impurity. The use of this potential does not complicate the solution of the problem because it does not involve any new type of integrals. In Table I we give the values of K for As, P, and Sb in silicon.

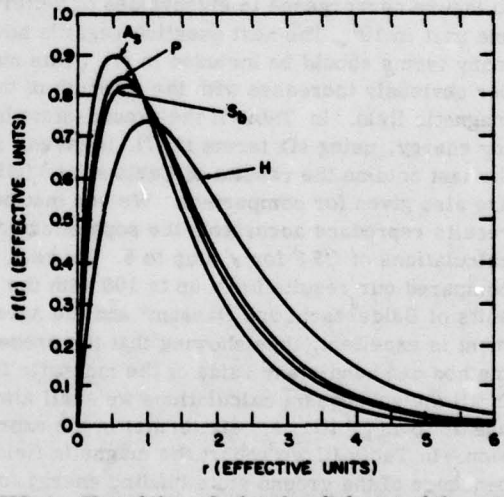


FIG. 1. Plot of the calculated radial part of the ground-state wave function [actually $r \times f(r)$] as a function of r in the absence of applied fields for various impurities in Si.

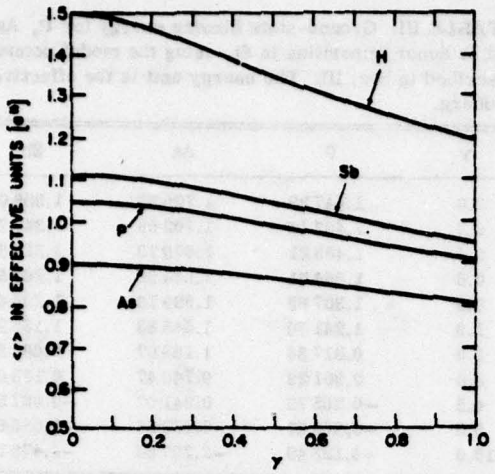


FIG. 2. Plot of the expectation value of r, $\langle r \rangle$, as a function of reduced magnetic field for H, P, As, and Sb impurities in Si.

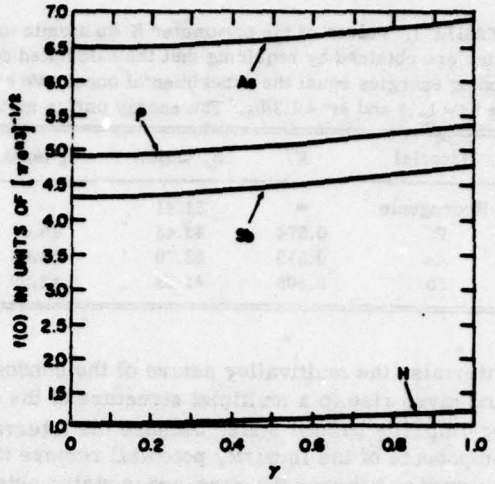


FIG. 3. Calculated magnitude of the radial wave function at the origin for various impurities in Si as a function of reduced magnetic field.

IV. NUMERICAL RESULTS AND DISCUSSION

We first address the question of convergence of the binding energy as a function of the basis functions in (15). In principle, the exponents η_i should be allowed to change in a variational way and to depend on any given state. In practice, however, as we have previously seen,[15] the η_i can be fixed and not considered as variational parameter, if one chooses a sufficient number of terms in (15) and if the η_i's are selected in a sensible way. We have seen that it suffices (i) to include 15 terms in (15), and (ii) to choose the largest and smallest η_i to be 50 and 0.05, respectively, with the intermediate ones chosen in geometrical progression, to insure convergence in eigenvalues to better than one part in 10^6. The next question regards how many terms should be included in (7). This number obviously increases with the strength of the magnetic field. In Table II the ground-state binding energy, using six terms in (7), is given. In the last column the results of Cabib *et al.*[2] (CFF) are also given for comparison. We see that our results reproduce accurately the sophisticated calculations of CFF for γ's up to 5. We have also compared our results for γ up to 100 with the results of Baldereschi and Bassani[5] and the agreement is excellent, thus showing that the present method can handle any value of the magnetic field. In all the subsequent calculations we shall always use up to $L_{max} = 10$ (i.e., six terms) in the expression. In Table III we report the magnetic field dependence of the ground state binding energy for P, As, and Sb in Si, respectively, using the model potential described in the previous section. In Fig. 1 we show the ground-state wave function in the absence of any external field for the four cases considered in this paper. The expectation value of r for the ground state, as a function of the reduced magnetic field, is shown in Fig. 2, for the various impurities while the value at the origin, $|\psi(0)|$ is given in Fig. 3.

To determine the polarizability α we calculate the binding energy using a very small value for the electric field and then use expression (6). To ensure that the electric field is indeed "infinitesimal," that is, only virtually perturbs the wave functions, we have performed calculations with $\xi = 0.05, 0.025, 0.01$ with essentially the same results (to four significant figures). The introduction of

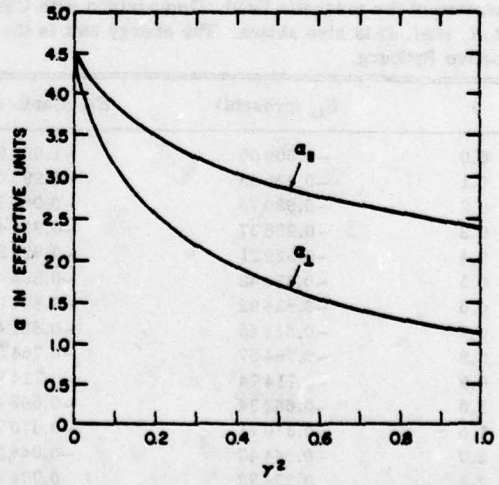


FIG. 4. Calculated polarizability components in effective units as a function of γ^2 for the hydrogenic impurity in Si.

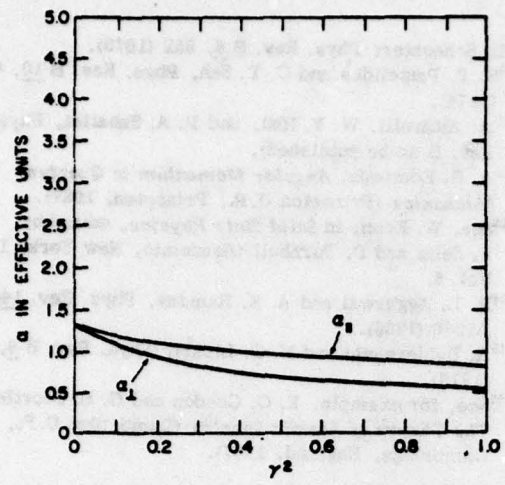


FIG. 5. Same as Fig. 4 for the P impurity.

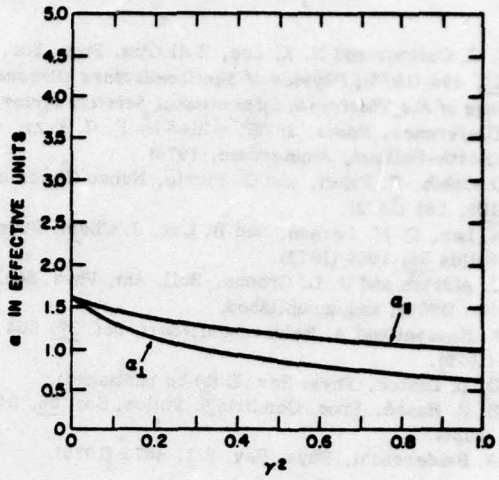


FIG. 7. Same as Fig. 4 for Sb impurity.

the electric field requires the odd L terms in expression (7) and in the calculation we have used all odd terms up to $L_{max}+1$. For $H_z=0$, the calculated polarizability α for hydrogen can be compared with the exact value.[16] We obtain for α a value which compares very well (i.e., to four significant figures) with the exact value $\frac{9}{2}$. In Fig. 4 we show the magnetic field dependence of $\alpha_\perp$ and $\alpha_\parallel$. These results compare reasonably well with those obtained in a previous paper in the low γ region by one of us (D.L.D.), and in addition, they are valid also for large γ.

In Figs. 5–7 the polarizabilities for P, As and Sb are given and we see that the magnetic-field dependence gets weaker with increasing binding energy. This is reasonable since the wave function becomes more localized with increasing binding energy and therefore the external fields are less effective in perturbing the wave functions.

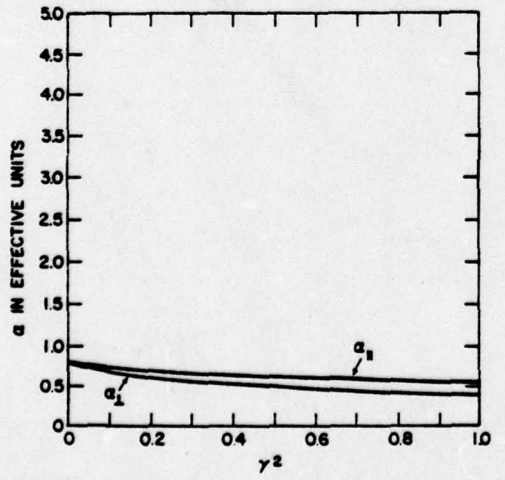


FIG. 6. Same as Fig. 4 for the As impurity.

The comparison with experiment is to be performed with caution since we have used here a crude approximation for the description of the donor-impurity dependence. The calculated decrease of α for $H_z=0$ in going from Sb to As is in agreement with observation. In fact, we predict $11\times10^5a_0^3$ for Sb, $8.2\times10^5a_0^3$ for P, $4.9\times10^5a_0^3$ for As, while the experiment gives $21\times10^5a_0^3$, $16\times10^5a_0^3$, and $6.7\times10^5a_0^3$, respectively. Also in agreement with experiment is the weaker dependence of α with magnetic field in going from Sb to As. The calculated absolute value of these dependences however is quite different from the observed ones. Specifically, the calculated dependences are much weaker than for the hydrogenic atom which in turn are much weaker than the experimental ones. Even though these results are obtained using a crude approximation, the fact that the experimentally observed dependence of α on H_z is much larger than even the hydrogenic case is puzzling. Ideally one should perform calculations using a more realistic model for the donor impurities which include a proper description of the electron anisotropy, intervalley coupling, and central-cell effects but in fact one would expect that, since the observed binding energies for the various donors are always greater than for the hydrogenic case, any calculation, even a more sophisticated one, would give the same qualitative results as those obtained here. More experimental information is also required to interpret the disagreement between theory and experiment.

ACKNOWLEDGMENTS

We are indebted to T. G. Castner for useful discussions. This work was supported in part by the AFOSR under Contract No. F49620-77-C-005.

[1]T. G. Castner and N. K. Lee, Bull. Am. Phys. Soc. 21, 405 (1976); *Physics of Semiconductors (Proceedings of the Thirteenth International Semiconductor Conference, Rome, 1976)*, edited by F. G. Fumi (North-Holland, Amsterdam, 1976).
[2]D. Cabib, E. Fabri, and G Fiorio, Nuovo Cimento 10B, 185 (1972).
[3]N. Lee, D. M. Larsen, and B. Lax, J. Chem. Phys. Solids 34, 1059 (1973).
[4]C. Aldrich and R. L. Greene, Bull. Am. Phys. Soc. 21, 354 (1976); and unpublished.
[5]F. Bassani and A. Baldereschi, Surf. Sci. 37, 304 (1973).
[6]D. L. Dexter, Phys. Rev. B (to be published).
[7]H. R. Hassé, Proc. Cambridge Philos. Soc. 26, 542 (1930).
[8]A. Baldereschi, Phys. Rev. B 1, 4673 (1970).
[9]D. Schechter, Phys. Rev. B 8, 652 (1973).
[10]S. T. Pantelides and C. T. Sah, Phys. Rev. B 10, 621 (1974).
[11]M. Altarelli, W. Y. Hsu, and P. A. Sabatini, Phys. Rev. B (to be published).
[12]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957).
[13]See, W. Kohn, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic, New York, 1959), Vol. 5.
[14]R. L. Aggarwal and A. K. Ramdas, Phys. Rev. 140, A1246 (1965).
[15]A. Baldereschi and N. O. Lipari, Phys. Rev. B 8, 2697 (1973).
[16]See, for example, E. O. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U.P., Cambridge, England, 1967).

Solid State Communications, Vol. 28, pp. 203–206. 0038–1098/78/1008–0203 $02.00/0
© Pergamon Press Ltd. 1978. Printed in Great Britain.

QUENCHING OF EXCITON DIAMAGNETIC SHIFTS IN POLAR, LAYERED MATERIALS*

J. Pollmann† and N.O. Lipari

IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, U.S.A.

and

H. Büttner

Physikalisches Institut der Universität Bayreuth, 8580 Bayreuth, West Germany

(*Received* 23 *June* 1978 *by J. Tauc*)

A theory for Wannier excitons in a magnetic field of arbitrary strength, including both anisotropy and exciton–phonon coupling, is presented. Both the anisotropy and the exciton–phonon coupling are shown to considerably quench the shifts of the energy levels in the magnetic field. Results are presented for PbI_2 and compared with recent experimental data.

THE INTEREST in layer structure materials has increased appreciably during recent years [1]. Optical properties, in particular, have been studied in order to better understand the electronic structure. One of the most studied cases has been PbI_2 in which band-edge excitons have been observed by several investigators [2]. Theoretical interpretations of the experiment differ considerably with regard to both the nature of the states responsible for the observed lines and the ground state binding energy for which estimates range from 16 meV [3] to 180 meV [4]. Recently, in an effort to obtain more information about the exciton spectrum, its variation with an applied external magnetic field has been studied [4]. Unexpectedly, the excitonic lines did not show any shifts.

In this letter we show theoretically that the absence of shifts is a direct consequence of the exciton–phonon interaction and the anisotropy. On the basis of our results we conclude that previously used isotropic descriptions of the exciton, which neglect either of the above two effects, are inadequate in polar, layered materials. We present the first theory which incorporates both the exciton–phonon interaction and the anisotropy and which is valid for arbitrary magnetic field strength. Both the exciton ground and excited states are calculated accurately. The available experimental data for PbI_2 are interpreted.

In [5] it was shown, that the interaction of LO-phonons with an isotropic exciton can be described by spherically symmetric, exciton-radius-dependent screening of the electron–hole Coulomb interaction. The generalization of this result to anisotropic materials is extensively discussed in [6], where it is shown that for not too large anisotropies ($0 \leqslant \alpha \leqslant 0.6$) one can obtain reliable energies using the spherically symmetric screening function in which averaged dielectric constants and masses are used. The results of [5] and [6] can be transferred to the present case because the magnetic field terms in the operator are not changed by the derivation of the effective potential, as can be seen in [7].

As an effective Hamiltonian, therefore, for the case of Faraday geometry (with a symmetrical gauge) we use

$$H = -\Delta_r + \alpha\partial_z^2 - \frac{2}{r} + \Delta V_{eff}(r) + E_g + \frac{\gamma^2}{4}(x^2+y^2) - \frac{1}{\mu_\perp}\gamma\mu_B l_z \qquad (1)$$

with the anisotropy constant $\alpha = 1 - \epsilon_{1\infty}\mu_\perp/\epsilon_{0\parallel}\mu_\parallel$. The z-component of the angular momentum operator in units of $\hbar$ is denoted as l_z. We have used as energy and length units:

$$R\hat{y} = Ry\frac{\mu_\perp}{\epsilon_{0\perp}\epsilon_{0\parallel}}; \quad a_0 = a_B\frac{\sqrt{\epsilon_{0\perp}\epsilon_{0\parallel}}}{\mu_\perp},$$

where Ry and a_B are the atomic Rydberg and the Bohr radius. Perpendicular "$\perp$" and parallel "$\parallel$" refer to quantities perpendicular and parallel to the preferred axis, which we define to be the z-axis. The field strength is given in terms of the field energy over the effective Rydberg energy which is $\gamma = \hbar\omega_c/R\hat{y} = \mu_B\mathcal{H}/R\hat{y}\mu_\perp$, and μ_B is the Bohr magneton. The reduced masses are abbreviated as $\mu_\perp$ and $\mu_\parallel$ and the coefficient in the

* Work supported in part by the AFOSR under contract F49620-77-C-005.

† Permanent address: Institut für Physik der Universität Dortmund, 4600 Dortmund 50, Pf. 500500.

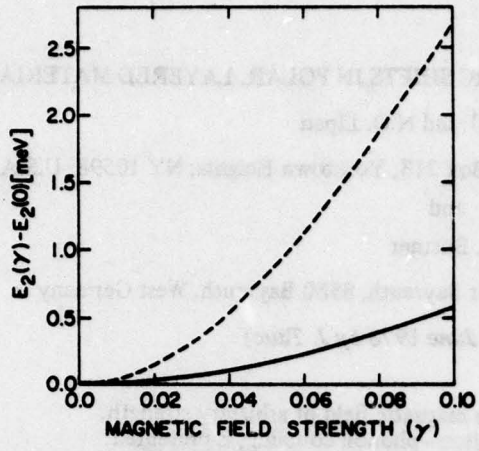


Fig. 1. Diamagnetic shift of the first excited state as a function of the magnetic field strength γ for PbI_2. The full line is the present result, including anisotropy and exciton–phonon interaction; the dashed line represents the hydrogenic approximation.

Zeeman term is $1/\mu_\perp' = 1/m_{e\perp} - 1/m_{h\perp}$. The additional electron–hole potential $\Delta V_{eff}(r)$ due to phonons is given [5] as (in our units):

$$\Delta V_{eff}(r) = -\frac{\mu_\perp}{\bar{\mu}}\left(1-\frac{\bar{\epsilon}_0}{\bar{\epsilon}_\infty}\right)\left\{\frac{2}{r}\sum_{i=1}^{4} a_i e^{-b_i r} + a_5 e^{-b_5 r}\right\}, \tag{2}$$

where $1/\bar{\mu} = 1/\bar{m}_e - 1/\bar{m}_h$. The coefficients a_i, the exponents b_i and the self energy E_s are given in [5] as functions of m_e, m_h, ϵ_0, ϵ_∞ and $\hbar\omega_{LO}$. In the present paper we can use the same analytical form of the potential, replacing those material parameters by the following averages [6].

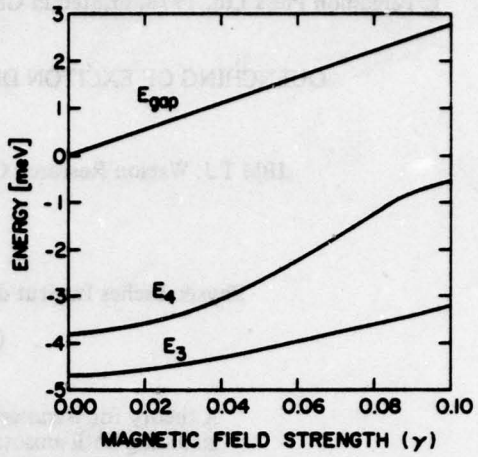
$$\frac{1}{\bar{m}_e} = \frac{1}{3}\left\{\frac{2}{m_{e\perp}} + \frac{1}{m_{e\parallel}}\right\}; \quad \bar{\epsilon}_0 = \sqrt{\epsilon_{0\perp}\epsilon_{0\parallel}}\,.$$

The quantities $\bar{m}_h$ and $\bar{\epsilon}_\infty$ are defined correspondingly. It is important to point out, that we use averaged material data only in the additional potential $\Delta V_{eff}(r)$, whereas in the other terms the anisotropy is treated exactly.

The solution of (1) can be expanded in terms of spherical harmonics

$$\psi_{N,m}(\mathbf{r}) = \sum_l F^m_{N,l}(r)|l, m\rangle \tag{3}$$

where m and the parity are good quantum numbers as can be simply seen from (1). We use N as a labelling index of the states and all l values in the sum are either odd or even. The states with $m = 0$ and $N = 1, 2, 3, 4$ are therefore the hydrogenic $1s$, $2s$, $3s$, $3d$ states if we approach the isotropic limit in the zero field case. In the resulting system of radial differential equations we use



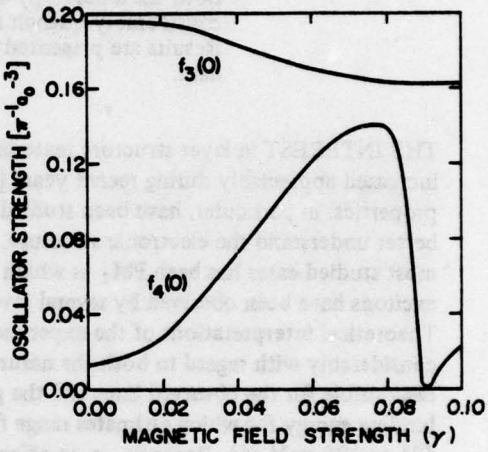


Fig. 2. (a) Diamagnetic shift of the second and third excited states for PbI_2. For comparison, we also give the linear shift of the band edge (lowest Landau level). (b) Square of the envelop function at the origin, $f_N = |\psi_{N,0}(0)|^2$ in units of $1/\pi a_0^3$ for the second ($N = 3$) and third ($N = 4$) excited state as functions of the magnetic field strength γ.

for the coefficient functions the following expansion:

$$F^m_{N,l}(r) = \sum_i^{N_{max}} C^l_{N,l,m}\, e^{-d_i r}. \tag{4}$$

The minimization with respect to the linear coeffecients C reduces the problem to the diagonalization of the secular matrix $H_{ij} - ES_{ij}$. With 7 terms in the spherical harmonics expansion and 20 exponentials, the convergence of the resulting eigenvalues is better than one part in 10^6. An extensive analysis of the convergence and the choice of the parameters d_i will be given elsewhere [8].

Before presenting our numerical results for PbI_2, we briefly discuss our choice of the material parameters given in Table 1. From the interpretation of the "$2p\pm_1$", "$3p\pm_1$", and the "$2p_0$" states, Fröhlich *et al.* [9]

obtained the anisotropy constant $\alpha = 0.37$, the band gap $E_g = 2.527$ eV and an effective Rydberg $R_y^* = 28$ meV. Using the value [12] $\epsilon_{0\perp} = 26.4$ and [3] $\mu_\parallel/\mu_\perp = 4.4$, we find $\epsilon_{0\parallel} = 9.5$, which compares reasonably well with the estimates given in [10], and $\mu_\perp = 0.512 m_0$. The LO-phonon frequency $\hbar\omega = 13.4$ meV and $\epsilon_{\infty\perp} = \epsilon_{\infty\parallel} = 6.1$ are also taken from [10]. The determination of the average masses $\bar{m}_e$ and $\bar{m}_h$ require knowledge of four individual masses. From experiment we know only $\mu_\perp = 0.516 m_0$ and $\mu_\parallel = 2.276 m_0$. Band structure calculations [11, 12] suggest a large, almost isotropic hole mass, and they show that the parallel electron mass is smaller than the hole mass. Therefore, we use in $\Delta V_{\text{eff}}(r)$ an isotropic hole mass $m_{h\perp} = m_{h\parallel} = \bar{m}_h = 6 m_0$. This gives $m_{e\parallel} = 3.67 m_0$ and $m_{e\perp} = 0.56 m_0$ for the electron masses. A smaller value for $\bar{m}_h$, e.g. $2.5 m_0$ would have resulted in a value $m_{e\parallel} = 25.7 m_0$ which is unreasonable. It is important to emphasize that the above choice for $\bar{m}_h$ is not critical with respect to the energy shifts in the magnetic field. In fact, as will be pointed out, other choices for $\bar{m}_h$ make the quenching due to the exciton–phonon interaction even more pronounced. The averaged masses used in $\Delta V_{\text{eff}}(r)$ and E_e are $\bar{m}_e = 0.7764 m_0$ and $\bar{m}_h = 6.0 m_0$.

To study the level shifts in a magnetic field, we have chosen a range of $0 \leqslant \gamma \leqslant 0.10$, which corresponds to a magnetic-field range of $0 < \mathcal{H} < 250$ kG. (The highest fields ever applied to PbI_2 is $\mathcal{H} = 180$ kG.) The ground state is practically unaffected by the field. The same result would also be true for the ground state in a hydrogenic description, because of the small value of γ.

The relatively greater sensitivity of excited states to the magnetic field (even for small γ values) exhibits the crucial role played by the exciton–phonon interaction and anisotropy. Figure 1 shows substantial quenching of the diamagnetic shift $E_2(\gamma) - E_2(0)$ of the first excited state. It is to be noted that, even for the highest applied field of 180 kG, i.e. $\gamma = 0.072$, the energy shift is only 0.3 meV, which is below the resolution of the experiment [4]. The hydrogenic description would have predicted a shift of 1.5 meV. Figure shows that the hydrogenic model is completely inadequate for all values of the magnetic field.

The results for the second and third excited states are shown in Figs. 2(a) and 2(b), where we have plotted $E_3(\gamma)$ and $E_4(\gamma)$. The conduction band edge in the zero field case is used as the zero of energy. Its linear shift (lowest Landau level) in the field is also plotted. The lower state $E_3(\gamma)$ is predominantly *s*-like at zero field strength. For this state, also, the diamagnetic shift is below the experimental resolution. The hydrogenic model predicts a shift of 4 meV (for $H = 180$ kG) which again reveals its inadequacy. The higher state $E_4(\gamma)$ shows a larger diamagnetic shift, which in principle could be observed. This state, however, is the "3*d*" state, which has only 12% *s*-character (due to the anisotropy) in the absence of a magnetic field. With increasing magnetic field the intensity of this state increases as shown in Fig. 2(b). Since, however, the energy position of this state is very near the quasicontinuum of the spectrum, it is probably very difficult to resolve. The peculiar behaviour of the oscillator strength for this state near $\gamma = 0.08$ is direct evidence of the complex interaction of this level with higher levels. In fact, for γ near 0.08 the "3*d*" and "4*s*" states would become degenerate if their crossing were not forbidden by symmetry.

In summary, we have presented here a theory, which properly takes into account the influences of the anisotropy and the exciton–phonon interaction on exciton levels in a magnetic field. Ground and excited states have been calculated and for all levels, a very large quenching of the diamagnetic shift is found. The magneto-optical data [4] for PbI_2 are in agreement with the theoretical predictions. The above effects play an important role in a large class of materials and the present theory should prove useful in their description.

REFERENCES

1. See, e.g. *Proc. Int. Conf. on Layered Semiconductors, and Metals, a Satellite Conference of the I.C.P.S.* 1976, *Nuovo Cimento* **38B**, 153–167 (1977).
2. A recent detailed discussion of the experimental results and a compilation of the corresponding literature can be found in HARBEKE G. & TOSATTI E., *J. Phys. Chem. Solids* **37**, 126 (1976).
3. BORDAS J. & RICCO B., *Phys. Status Solidi* **67**, 577 (1975).
4. SKOLNIK M.S., THANH Le CHI, LEVY F. & HARBEKE G., *Physica* **89B**, 143 (1977).
5. POLLMANN J. & BÜTTNER H., *Phys. Rev.* **B16**, 4480, Nov. (1977).
6. FOCK H., KRAMER B. & BÜTTNER H., *Phys. Status Solidi (b)* **72**, 155 (1976).
7. BEHNKE G., BÜTTNER H. & POLLMANN J., *Solid State Commun.* **20**, 873 (1976).

8. LIPARI N.O. & POLLMANN J. (to be published).

9. FRÖHLICH D. & KENKLIES R., *Nuovo Cimento* 38B, 433 (1977).

10. LUCOVSKY G., WHITE R.M., LIANG W.Y., ZALLEN R. & SCHMID Ph., *Solid State Commun.* 18, 811 (1976).

11. SCHLÜTER I.C. & SCHLÜTER M., *Phys. Rev.* B9, 1652 (1974).

12. DONI E., GROSSO G. & SPAVIERI G., *Solid State Commun.* 11, 493 (1972).

13. POLLMANN J. & BÜTTNER H., *Solid State Commun.* 17, 1171 (1975).