Title
DEEP LEVEL DEFECTS IN SEMICONDUCTORS: STRESS AND PRESSURE STUDIES Thesis

Author
Nolte, D.D.

Publication Date
1988-05-01

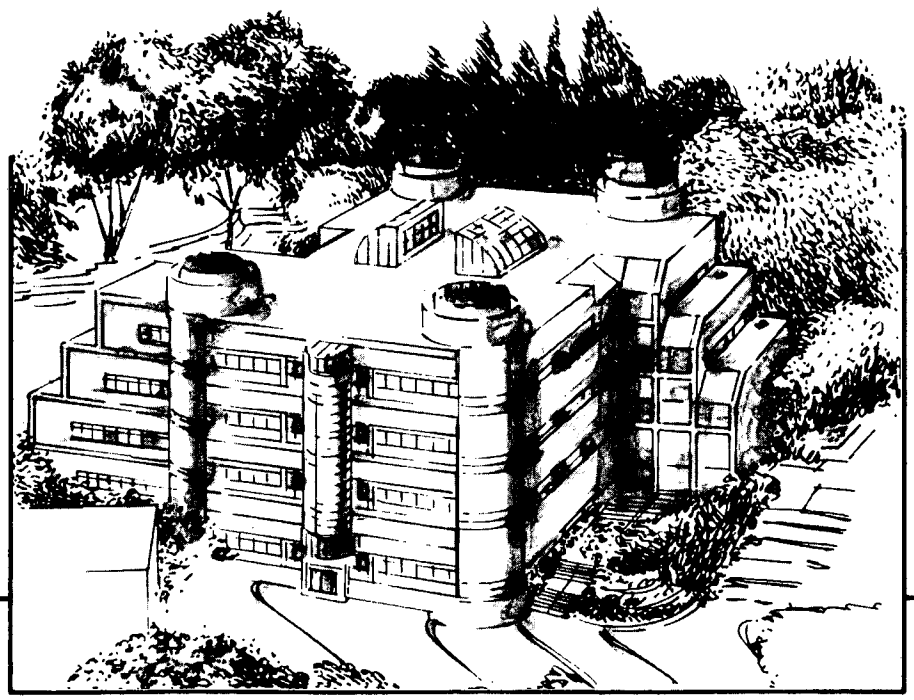
Center for Advanced Materials

CAM

Deep Level Defects in Semiconductors: Stress and Pressure Studies

D.D. Nolte
(Ph.D. Thesis)

May 1988



Materials and Chemical Sciences Division

Lawrence Berkeley Laboratory • University of California

ONE CYCLOTRON ROAD, BERKELEY, CA 94720 • (415) 486-4755

Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

LBL-25242

Deep Level Defects in Semiconductors: Stress and Pressure Studies

David D. Nolte

Ph.D. Thesis

Lawrence Berkeley Laboratory
University of California
Berkeley, Ca. 94720

May 1988

This work was supported in part by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Science division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Deep Level Defects in Semiconductors: Stress and Pressure Studies

David D. Nolte

Submitted in partial fullfilment of requirements for the degree of Doctor of Philosophy, to the Department of Physics, University of California, Berkeley, CA

Abstract

Deep-level defects in semiconductors have strongly localized wavefunctions, which sample near-neighbor potentials around the defect center. These defects therefore can coupling strongly to the local lattice, leading to many interesting phenomena, including large lattice relaxation. One method to probe the defect-lattice coupling is by applying stress or pressure. The application of stress or pressure to a semiconductor sample containing deep-level defects alters the lattice spacings around the defects and consequently alters their energy position in the bandgap. The magnitude of the energy shift under stress, as well as the dependence of the energy shift on uniaxial stress direction, can provide information concerning the structure of the deep level and its symmetry.

Uniaxial stress applied in conjunction with deep level transient spectroscopy is a relatively new technique for probing the symmetry of deep-level defects. Uniaxial stress lowers the symmetry of the crystal and can break the degeneracy of defects, leading to energy splittings. The effect of these splittings on the thermal emission of carriers from defects is discussed in detail. The limit of energy resolution for this thermal ionization technique are also investigated.

Pressure and uniaxial stress can also affect the band-edges of the bandgap. Using iron-acceptor pairs in silicon as stress-insensitive reference levels, the effects of uniaxial stress on the valence-band top is studied. It is found that the effective masses of the valence band must be redefined as a function of stress to be consistent with the splitting of the top of the valence

band. The new effective masses roughly conserve the center of gravity of the stress-split valence band.

Substitutional transition-metal defects in semiconductors have the interesting property that the conduction band and valence band contribute nearly equally to their electronic structure. These defects therefore are only sensitive to the gross features of the band structure and are not strongly influenced by the details of the band extrema near the edges of the bandgap. For this reason, these defects can be used as reference levels from which to measure changes in band structure. Specifically, they can be used to measure the band-edge hydrostatic deformation potentials. These deformation potentials have important consequences for carrier mobility (through acoustic deformation-potential scattering) and breathing-mode lattice relaxation around defects.

Acknowledgements

The pursuit of a doctorate of philosophy, though a personal endeavor, cannot be carried to fruition without the help and support of many people. I would foremost and above all like to express my heartfelt thanks to Eugene Haller and Leo Falicov for the guidance and support that they provided me during the progress of my thesis work. Much of my professional growth and maturation I attribute to them.

The long path of my education began ultimately with my parents Nancy and Stuart Nolte. They instilled in me, whether by inheritance or by example, the initial skills and curiosity that have driven my quest for understanding. They have been constant companions, providing advice and encouragement and keen interest in my progress (as well as the incalculable value of a college education).

Of teachers and professors, there are several who stand out in my mind as having influenced my development towards and through physics, and who I would especially like to thank. These are: Shirley Durant for her early confidence in me; Kurt Gottfried for opening up the higher levels of physics for me; Paul Richards for his support during my time in his laboratory; Eicke Weber for interesting discussions and strong support with EPR experiments and letters of recommendation; and Peter Yu and Angelica Stacey for taking the time to read my thesis. I would in addition like to thank Wladyslaw Walukiewicz for the many hours of exciting discussions of semiconductor physics.

A graduate student learns most of his "practical" skills from other graduate students, and this has been no exception in my case. Among these, I

would first like to thank Andrew Lange for introducing me to experimental physics. I have enjoyed watching him rise in the ranks from graduate student to assistant professor of physics at Berkeley. Robert McMurray, Jr., played an introductory role in the choice of my thesis work and thesis advisors. I would like to thank him for his enthusiasm and advice during my work. Nancy Haegel and Joseph Kahn were also helpful in my early education in semiconductors; these are two most competent researchers, and I am grateful for their examples of excellence. I would also like to thank John Trombetta for his patience and willingness to pursue the interesting properties of ESR on iron-acceptor pairs.

Friends have been a great help during my time as a graduate student. I shall remember fondly the times spent with Jeff and Carline Beeman, especially the era of "Thursday Nights"! One friend in particular has been instrumental in his help and support through the years of my study of physics at Berkeley: R. K. Owen. We have commiserated the ups and downs of graduate school from the beginning days as teaching assistants through course work and thesis work. To him, I would like to express special thanks and good luck.

Finally, at this unique and prominent position in the acknowledgements, I would like to thank my wife Laura, my companion spirit with whom I shared the thrills and trials of Berkeley and graduate school. Her inspiration and love, and the great confidence she has in me, have been invaluable. At the end of this one great stage in our life, we face the future together.

Table of Contents

Chapter I: Introduction to Deep-Level Defects in Semiconductors

A cubic centimeter of a semiconductor crystal contains on the order of 10^{23} strongly interacting particles. The complexity of this extreme N-body problem is reduced considerably by the periodicity of the crystal lattice. Bloch's theorem, based on translational invariance of the crystal potential, places a strong restriction on the form of the electron eigenfunctions, reducing them to a superposition of plane waves with discrete wavevectors. These electrons are scattered by the periodic potential of the crystal ions. Gaps in the energy spectrum of the crystal eigenvalues open at wavevectors that satisfy the Bragg condition. In the case of semiconductors and insulators, the Fermi level falls within one of the gaps. This gap is consequently called the fundamental bandgap. The exponential dependence of the intrinsic carrier concentrations on the bandgap is the classic signature of the semiconductor's electrical properties.

When crystal atoms are removed, or displaced, or when impurities are incorporated into the crystal, the translational invariance of the crystal is destroyed and the simplification of Bloch's theorem is lost. The energies of the defect electron eigenstates are not excluded from the bandgap, and localized electron states can result. The task of understanding defect structure is a quantum mechanical many-body problem. Simplifications can still be made for some types of defects; if the defect potential is not locally very strong, and if it has a long-range coulombic field, then the defect electronic structure is

dominated by the band structure of the otherwise periodic crystal. This is the case of the shallow donor and acceptor levels. The electronic structure of these defects can be derived by considering the defect potential as a weak perturbation of the perfect crystal states. The defect wavefunctions are mostly composed of the Bloch states at the edges of the bandgap, and the energies are pinned to the respective band edges.

The structures of defects that have strong local potentials, the so-called "deep level" defects, have been much more difficult to understand. The strength of the local potential prevents the use of perturbation theory, and the energies of these defects are typically not pinned to any easily recognizable band edge. The localization of the defect wavefunctions can also cause strong local coupling of the defect energy to the lattice, leading to many interesting but complicated phenomena related to large lattice relaxations around the defects. Furthermore, because of the many possible strong local potentials that can be produced, the properties of deep-level defects vary widely. These aspects of deep-level defects make them interesting but difficult defects to study. Only in the past decade has strong experimental and theoretical progress been made in attempts to elucidate the structure of these crystal imperfections. The purpose of this thesis work is to use uniaxial and hydrostatic stress as a probe of deep-level defect structure.

I. Classification of Deep-Level Defects and Defect Coupling to the Lattice

A. Classification of Deep-Level Defects

The many varieties of deep-level defects can be roughly classified into three basic categories. There is some overlap between the categories, and

some defects resist classification altogether, but the categories cover most of the common deep levels and give an insight into similarities and differences among them. The three defect classes are: effective-mass-like defects, deep substitutional defects, and vacancy-related defects.

Effective-mass-like defects are defects that have long-range coulombic potentials with central-cell potentials of weak or moderate strength. These defects are typically the double and triple donors and acceptors. The excited p-states are effective-mass states with hydrogenic energy spacings. The ground state, on the other hand, may experience a strong central-cell correction. The levels that tend to be effective-mass-like are the first ionization stages of these defects. These levels are usually not far removed in energy from the band edges and have pressure derivatives that are similar to the pressure derivative of the band edges. The ground state of the second or third ionization level of these defects can be deep in the bandgap and lose its special relationship to a specific band edge.

Deep substitutional defects result when a host atom is replaced by an impurity leading to a strongly localized defect wavefunction. Examples of this class are transition-metal atoms or isoelectronic defects such as GaP:N. Considering only the transition-metal defects, the dominant characteristic of these defects is the partially filled d-orbital. The high occupancy of the d-shell frequently leads to high degeneracy (with accompanying Jahn-Teller distortion), and multiple charge states in the bandgap. The transition-metal defects can be classified into two subcategories depending on the site of the transition-metal atom: substitutional or interstitial. In the substitutional configuration, the transition-metal atom replaces a host atom and bonds strongly with the nearest neighbors. In the interstitial configuration, the transition-metal atom occupies the tetrahedral interstitial site in the anti-bonding directions from the nearest

neighbors. The difference in the bonding properties of these two configurations leads to different defect structure and properties.

Vacancy-related defects occur when one or more host atoms are displaced and their sites remain vacant. These defects are characterized by the presence of dangling bonds, either reconstructed or unreconstructed. The dangling bonds occur at sites that are next-nearest neighbors. Therefore the interactions between dangling bonds tend to be rather weak. This is especially true for the Coulomb energy. The small Coulomb energy allows several different charge states to occur in the bandgap. The vacancy is the canonical deep-level defect: It is the limiting case for the other two categories mentioned above. The presence of deep substitutional multiple donor or acceptor atoms, as well as substitutional transition metals within the vacancy do not strongly alter the dangling bonds. For this reason, the categories mentioned above are rather "soft" and involve considerable overlap in the classification of deep-level defects.

A final category includes the complexes that are formed by associates of individual defects from these three basic classes. These complexes can have a combination of properties derived from the individual constituents. Because of the unlimited possibilities of combination, this category is in fact the largest, and includes the least understood defects. Such scientifically and technologically important defects as the so-called EL2 defect in GaAs or the DX center in AlGaAs may fall under this category.

B. Defect-Lattice Interaction

The electronic energy of a defect is a function of the lattice coordinates of the neighboring atoms. As these lattice sites are perturbed (either through phonons or applied stress) the defect energy will change. In addition to

external perturbations of the local lattice sites, the defect itself can affect the positions of the neighboring atoms. The change in defect electronic energy per unit change in the lattice spacing has the units of an energy gradient and defines the force acting on the atoms. These atoms will respond to the force, relaxing until the elastic energy counterbalances the electronic energy. This is the phenomenon of lattice relaxation. Lattice relaxation, especially when the relaxation is particularly large, can lead to many interesting and complicated dynamical properties including persistent photoconductivity and metastability. Defects with these properties, such as EL2 in GaAs and the DX center in AlGaAs, can have great technological and scientific importance. Therefore electron-lattice coupling is a key aspect of the structure of deep-level defects. Stress is one particularly direct probe of electron-lattice coupling, and is the primary motivation behind this thesis. Various aspects of lattice coupling will be discussed next.

1. Defect Energy as a Function of Lattice Coordinate

The total defect Hamiltonian can be written as a combination of electronic and ionic contributions to kinetic and potential energy

$$H = T_e + T_N + V_{ee} + V_{NN} + V_{eN} ,$$

where T_e and T_N are the electronic and ionic kinetic energy terms, and V_{ee} and V_{NN} are the electron-electron and ion-ion interactions. The last term V_{eN} is the electron-lattice interaction. The adiabatic (or Born-Oppenheimer) approximation simplifies this Hamiltonian by neglecting the kinetic energy T_N of the ions compared to the electronic energy. For a review, see Stoneham[1975].

This approximation is justified by considering the ratio of electronic to nuclear kinetic energy; which varies as

$$T_e / T_N \approx (m / M)^{1/2},$$

where M is the mass of the nucleus and m is the electron mass. Usually M is many orders of magnitude larger than m and the neglect of the nuclear energy is justified. However in defects that contain hydrogen, the nuclear mass is sufficiently small so that the Born-Oppenheimer approximation breaks down [Haller, 1980; Muro, 1986; Kahn, 1986]. For defects that do obey the Born-Oppenheimer approximation, the coordinates of the nuclei **Q** are considered to be classical variables (since there is no dependence on the ionic momenta). These coordinates may be varied continuously during the intermediate step of solving the electronic quantum-mechanical problem for the electronic energy $E_e(\mathbf{Q})$ and the electronic wavefunction $\phi(\mathbf{r},\mathbf{Q})$ where the **Q** are parameters. The minimum of the electronic energy occurs at the equilibrium lattice coordinates $\mathbf{Q_0}$. The energy can be expanded about the minimum to give

$$E_e(\mathbf{Q}) = E_e(\mathbf{Q_0}) + 1/2 \sum_\alpha k_\alpha (Q_\alpha - Q_{\alpha 0})^2,$$

where the summation is over different lattice normal modes and the k_α are the local force constants for the different modes. The ionic motion is that of an harmonic oscillator and is found by solving $[T_N + E_e(\mathbf{Q})]\ \psi(\mathbf{Q}) = E_N(\mathbf{Q})$, where $E_e(\mathbf{Q})$ is the potential experienced by the ion, $\psi(\mathbf{Q})$ is the wavefunction for the ionic motion, and $E_N(\mathbf{Q})$ are the eigenstates of an harmonic oscillator.

When a defect is present, a term linear in the lattice displacement is introduced and the energy is expressed as

$$E = \Sigma_\alpha [\ 1/2\ k_\alpha\ (Q_\alpha - Q_{\alpha 0})^2 + A_\alpha(Q_\alpha - Q_{\alpha 0})\ ] + E_e(\mathbf{Q_0}),$$

where the first term is the elastic energy and the second term is the defect-lattice interaction. A_α is the defect-lattice coupling parameter for mode α. For non-zero lattice-coupling the defect energy can be lowered by relaxing the lattice coordinates from Q_0. Two types of lattice relaxation can occur: symmetry-preserving, and symmetry-breaking relaxation. When the lattice relaxation preserves the symmetry of the defect, the lattice relaxation is called a breathing-mode or isotropic relaxation. All defects can couple isotropically to the lattice. Defects that have electronic degeneracy can in addition couple to symmetry-breaking lattice distortions. The symmetry-breaking distortions remove the defect degeneracy and lower the energy of one of the degenerate manifold states. Spontaneous symmetry-breaking is called in this case the Jahn-Teller effect and has been studied in detail for molecules and crystal defects [Engelman,1972; Stoneham,1975].

A defect can couple to several different lattice modes, which can lead to complicated distortions of the defect. To understand the properties of a defect, these modes are often lumped into a single configuration coordinate with a single effective coupling parameter. This coupling parameter will in general be different, depending on whether the defect is occupied by a carrier or not. The total energy of a defect in the occupied and unoccupied states can be presented graphically in a configuration-coordinate diagram. A typical configuration-coordinate diagram for a deep level is shown in Fig.(1.1). The parabolas are the energy surfaces of the defect for the case of the defect ionized with the charge carrier in the conduction band, and the defect occupied

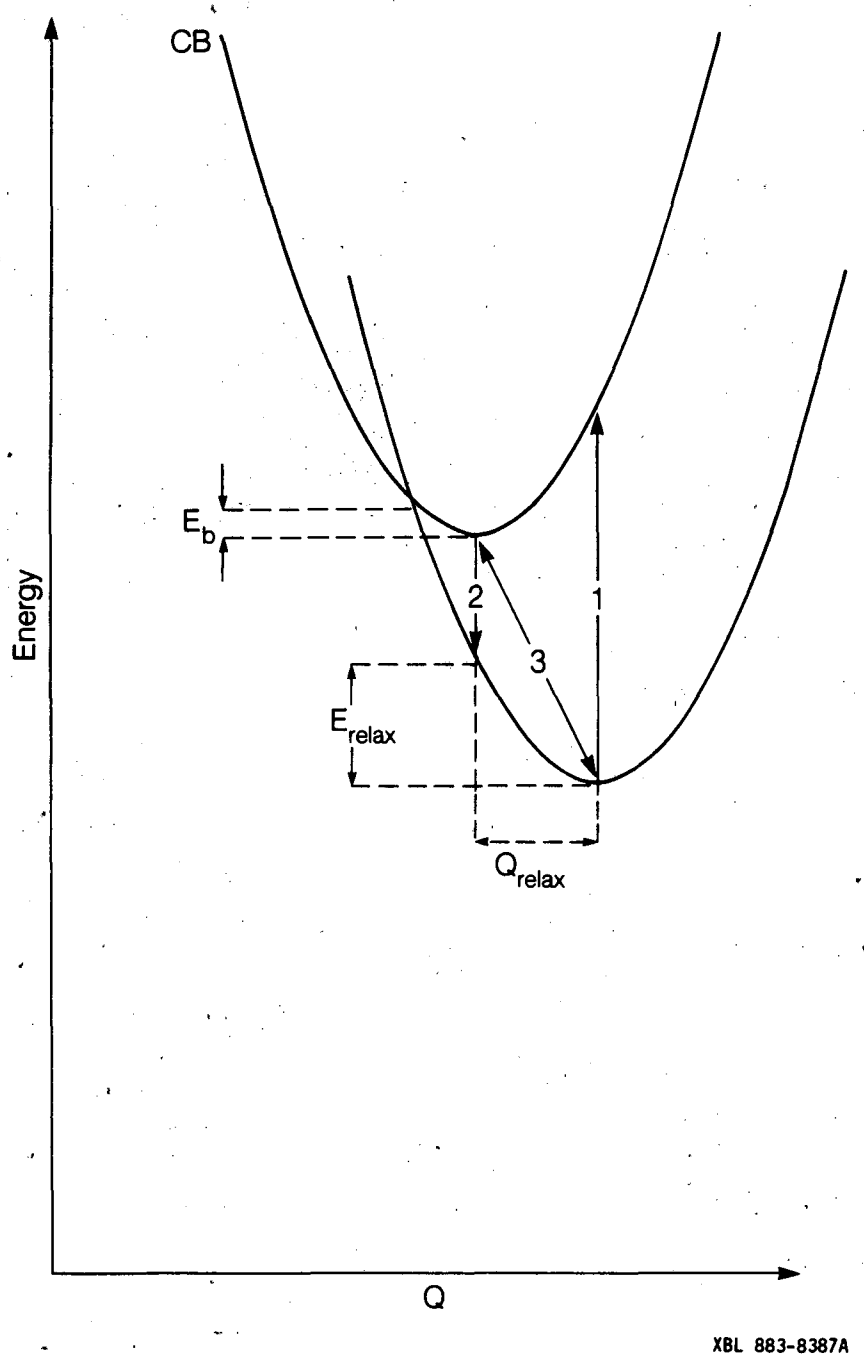


Fig. 1.1 Configuration-coordinate diagram for a defect showing optical and thermal transitions between two charge states of the defect.

by the carrier. The lattice coordinate relaxes upon carrier emission by the amount

$$Q_{relax} = (A_{unocc} - A_{occ}) / k$$

with the corresponding relaxation energy

$$E_{relax} = -1/2 (A_{unocc} - A_{occ})^2 / k ,$$

where A_{unocc} and A_{occ} are the unoccupied and occupied lattice coupling parameters and k is the local force constant.

Three transitions are depicted in Fig. 1.1 by the arrows. Transition 1 is an optical ionization transition in which a carrier is removed to the conduction band, while transition 2 is a photoluminescence transition in which the carrier is captured radiatively from the band into the defect. Both are "sudden" transitions in what concerns the lattice coordinate. These optical transitions are drawn as vertical lines on the configuration-coordinate diagram because they occur too quickly for the lattice coordinate to respond. This is the Franck-Condon principle. The difference between the absorbed energy in transition 1 and the emitted energy in transition 2 is called the Franck-Condon shift, which is

$$w_{FC} = (A_{unocc} - A_{occ})^2 / k = 2 E_{relax} .$$

The Franck-Condon shift is related to the normal-mode vibration frequency ω through the expression

$$w_{FC} = 2 S \hbar\omega ,$$

where S is the Huang-Rhys factor [Huang,1950] and is a measure of the strength of the lattice relaxation. The third transition in Fig. 1.1 is a thermal transition; it is an adiabatic transition with respect to the lattice degrees of freedom. The energy difference in a thermal transition is equal to the difference in equilibrium free energies between the occupied and unoccupied defect. Thermal-transition energies therefore involve total energies: elastic and electronic energies. Optical transitions involve only electronic energies. The comparison between optical and thermal ionization energies can therefore yield information concerning the lattice relaxation around a defect.

2. Large Lattice Relaxation

When lattice relaxation is very large, several interesting effects can occur. These effects include multiphonon capture and metastability. Multiphonon capture occurs during non-radiative recombination of carriers with defects [Huang,1950; Henry,1977; Stoneham,1981]. The capture occurs when the energy surfaces of the occupied and unoccupied states intersect at a certain value of the configuration coordinate. One of the signatures of multiphonon capture is the exponential dependence of the capture cross section on an energy barrier E_B,

$$\sigma = \sigma_\infty \exp[-E_B / k_B T] + \sigma_0 ,$$

where σ_∞ is the high-temperature multiphonon capture cross section and σ_0 is the capture cross section of a competitive capture process that short-circuits the multiphonon capture process. For low temperatures ($k_B T << \hbar\omega$), the cross

section tends to σ_0 , which varies typically between 10^{-17} cm^2 to 10^{-21} cm^2 for defects in GaAs [Henry,1977].

Thermal emission rates depend on the capture cross section and the trap energy E_T through the equation derived from detailed balance (see Chapter II)

$$e = 1/\tau = c\,\sigma \exp[\,-E_T / k_B T\,] ,$$

where c depends on bulk and defect properties. In the temperature range where σ depends exponentially on E_B this becomes

$$1/\tau = c\,\sigma_\infty \exp[\,-(E_T + E_B) / k_B T\,] .$$

The thermal activation energy is therefore the sum of the trap energy and the barrier to capture. This point must be considered when comparing thermal emission energies (as obtained from deep level transient spectroscopy, or DLTS) with thermal equilibrium energies (as obtained from Hall effect measurements). In stress measurements E_T and E_B are both affected by the stress. Care must therefore be taken when interpreting energy shifts [Li,1987] measured by stressed DLTS.

Perhaps the most interesting result of large lattice relaxation is defect metastability. Two technologically important deep levels are EL2 in GaAs and the DX center in AlGaAs. Both of these defects exhibit metastable properties that have major impacts on devices that contain them. The electrical activity of EL2 (which is the compensating deep level responsible for semi-insulating GaAs) can be photo-quenched at 77 K into a metastable state [Bois,1977; Vincent,1982]. The DX center in AlGaAs leads to persistent photoconductivity (PPC) in which a photo-ionized DX defect has a large barrier to recapture of an

electron [Lang,1979]. These metastable properties are suspected to be the result of large lattice relaxations in which the configuration-coordinate relaxes to such an extent in the metastable state that large energy barriers must be overcome to regain the stable state.

3. Defect Stress Derivatives and Local Force Constants

Lattice relaxation involves two quantities: the defect-lattice coupling parameter, and the local force constant. To fully understand lattice relaxation each of these quantities must be measured separately. The application of stress alters the lattice coordinates around a defect, thus altering the defect energy, the latter measured experimentally as a shift in a transition energy. If one can know how the lattice sites around the defect are perturbed by an applied stress, then the measured stress derivative of the defect energy can be related directly to the defect-lattice coupling. However, local force constants are not equal to the force constants in the bulk, and therefore stress derivatives (like the relaxation energies) involve both the defect-lattice coupling and the local force constant. Although the defect-lattice coupling and the local force constants cannot be determined separately from stress experiments, lattice relaxation and stress derivatives are directly related, and lattice relaxation around defects are measurable [Samara,1986,1987a,1987b]. A major complication arises in the case of isotropic relaxation because of the participation of the band edge in the measured pressure derivative of the defect. The hydrostatic band-edge deformation potential was recently determined for several semiconductors [Nolte,1987c,1987e]. This work and its application to the calculation of isotropic lattice relaxation is described in Chapters V and VI of this thesis.

It is still desirable to know the defect-lattice coupling and the local force constants separately, and attention has to be paid to how these two quantities affect the structure of the defect and how they can be either measured experimentally or predicted theoretically. The magnitudes of defect-lattice coupling parameters span a wide range for the case of deep levels. Two deep-level defects observed with DLTS in Si, the thermal donor [Benton,1984; Henry, 1984] and the oxygen-vacancy [Meese,1983], have axial stress derivatives of nearly 100 meV/GPa. The oxygen-vacancy defect observed by EPR has comparable stress derivatives [Watkins,1961], while other vacancy-related defects have smaller, but still significant stress derivatives [Watkins,1965,1975]. Transition-metal defects, on the other hand,have axial stress derivatives [West,1980; Nolte,1987a] on the order of 5 meV/GPa. There is likewise a wide range of hydrostatic stress derivatives among deep-level defects. Defects in silicon all have relatively small isotropic stress derivatives [Nolte,1987e], while some defects in GaAs (such as EL2 and EL6) couple very strongly to isotropic lattice modes [Nolte,1987d]. Yet a large number of defects in GaAs (including all the transition-metal defects and many radiation-induced defects) have small (or vanishing) isotropic lattice coupling.

There have been strong theoretical advances in the prediction of axial lattice couplings for some specific defects, most notably the vacancy in silicon [Baraff,1980] as well as vacancy-related complexes [DeLeo,1984]. But these approaches are heavily computational and cannot easily be generalized to other defect classes. The theoretical prediction of isotropic defect-lattice coupling has been almost non-existent in view of the difficulty in defining absolute defect pressure derivatives.

It is interesting to compare experimentally predicted Jahn-Teller energies (derived from experimental stress derivatives) with theoretical values. This has

been done for the vacancy in Si in a review article by Watkins [see Pantelides,1985]. He finds that the experimental values are larger than the theoretical values [Baraff,1980] by a factor of three, much of this difference being attributed to the local force constants around the vacancy. It is therefore clear that local force constants play an appreciable role in the comparison of experiment with theory. Because of this importance, considerable theoretical [Baraff,1980; DeLeo,1984] work has been performed on the force constants around specific defects with substantial disagreement in the results. The local force constant remains an open problem.

II. Transition-Metal Impurities in Semiconductors

Transition-metal defects are perhaps the best understood deep-level defects in semiconductors. This state of understanding has come about for several reasons: these defects are point defects that do not dramatically alter the local bonding structure of the semiconductor; the availability of different chemical species that differ only slightly from their immediate neighbors in the periodic table has allowed systematic trends to be understood; their wavefunctions are localized, which make them prime candidates for the use of Green's function calculations; and finally they can be introduced easily into semiconductor crystals either at growth or by diffusion. The ease of introduction of transition metals into semiconductors, though a boon for defect identification, is a serious technological problem because these impurities are the primary defects introduced during high-temperature processing of integrated circuits. The technological importance of these defects has therefore provided an additional impetus for their detailed study. Because the structure of transition-metal defects is so well understood, they are ideal test cases in which to study

the effects of stress and pressure. Most of the experimental work of this thesis is undertaken on transition-metal impurities and complexes, and important conclusions of this thesis are based on the electronic structure of transition-metal defects. Therefore the present chapter includes a survey of the experimental and theoretical state of knowledge of transition-metal defects.

A. Incorporation into the Crystal

Transition-metal atoms have high diffusivities and high solubilities in both the group IV and group III-V semiconductors. In silicon all the transition metals diffuse interstitially and tend to remain in interstitial sites in their electrically active form. Exceptions to this rule include the noble and near-noble transition-metal elements such as Cu, Au, Pt, and Pd. These defects occur substitutionally in some of their electrically active forms. Among the 3d transition metals the diffusivities and solubilities increase as the atomic number Z increases from Ti to Ni [Weber,1983]. As an example of typical diffusivities and solubilities, the results for Fe in silicon are

$$D(Fe_i) = 1.3\times10^{-3} \exp[\, -0.68 \text{ eV} / k_B T \,] \quad \text{cm}^2/\text{sec}$$

$$S(Fe_i) = 6.7\times10^{22} \exp[\, 7.9 \,] \exp[\, -3.0 \text{ eV} / k_B T \,]. \quad \text{cm}^{-3}$$

In III-V semiconductors transition-metal impurity atoms occupy the substitutional cation site. The diffusivities and solubilities in the III-V compounds have not been studied in as much detail as they have in silicon. However, the same trends have been observed, namely, increasing Z increases both the diffusivity and the solubility. The solubility limit in the III-V semiconductors, around 10^{17} cm^{-3} , is rather low compared to silicon.

Transition-metal impurities can easily pair with other impurities or defects in the crystal. This is seen dramatically in the case of the iron-acceptor pairs in silicon. The total concentration of iron in silicon increases with increasing acceptor concentration. Furthermore, after iron diffusion, most of the electrically active iron occurs in the iron-acceptor pairs. At room temperature, interstitial iron is mobile enough so that it either forms the iron-acceptor pair, or it precipitates with other interstitial iron and loses its electrical activity. Iron-acceptor pairs exhibit interesting metastable properties, in which two configurations can be transformed among themselves [Chantre,1985]. This property has been attributed to the small barriers to iron diffusion, although the chemical trends in the activation energies as the acceptor species is changed have yet to be explained. In Chapter IV of this thesis the results of stressed DLTS on the iron-acceptor pairs are described and several possibilities for the defect structure are discussed.

B. Experimental Observation of Defect States

1. Electron Paramagnetic Resonance

Transition-metal defects in Si were first studied extensively using electron paramagnetic resonance (EPR) by Ludwig and Woodbury [1962]. Transition-metal impurities are especially well suited for study by EPR because the d-shell is only partially occupied and therefore has unpaired spins. In addition, several charge states are possible, providing a rich spectrum of signals for even a single transition-metal impurity. The d-orbitals are split by the tetrahedral crystal field into a doublet of the so-called e symmetry and a triplet of the so-called t_2 symmetry. The occupancy of these states are determined from the d-shell occupancy of the free atom. In the case of interstitials, the neutral d-

shell occupancy is d^{n+2}, where n is the neutral free atom d-orbital occupancy, and the two outer s electrons are promoted to the d-shell. For substitutional species, the neutral d-shell occupancy is d^{n+2-v}, where v is the valence of the atom that the transition metal replaces. The substitutional transition metal uses v electrons to satisfy the local sp^3 bonds. The valence in Si is v=4. In GaAs, the valence v=3 for the Ga-site. In ionic notation, the neutral charge state in GaAs is denoted by TM^{3+}, which signifies the use of three electrons in the local bonds. This notation is merely used as a convenience and does not accurately reflect the chemical oxidation state of the transition-metal atom.

Ludwig and Woodbury were able to explain the observed g-tensors with a simple one-electron model of the d-orbitals. The orbitals are filled according to Hund's rule of high spin. In the case of substitutional transition metals, the e-state is below the t_2 state. The e state is filled with two electrons, and the t_2 state filled with three electrons before the e state is occupied by electrons with the reverse spin. This indicates that the pairing energy is larger than the crystal-field splitting.

EPR probes the local magnetic structure of a specific charge state. Through spin-orbit interaction the g-tensor reflects the defect site symmetry. Information on the site symmetry of a defect is important for the development of a microscopic model of the defect. The hyperfine interaction gives additional information on local lattice sites and this aspect has been exploited in the technique of electron nuclear double resonance (ENDOR) in which the displacements of successive shells of atoms around the defect can be mapped out. The effects of stress [Watkins,1961,1965] or optical absorption can be included with EPR to give even more information concerning the electron-lattice coupling parameter or the energy level position of the defect in the bandgap. Because of the high sensitivity of EPR to local symmetry, it is an excellent

technique for probing lattice relaxations, especially Jahn-Teller distortion [Stauss,1977; Krebs,1977]. Despite this strength, it is often difficult to match EPR signals with a corresponding energy level in the bandgap. One means of providing a correlation between a defect observed by EPR with a defect level observed in the bandgap is to look for similar behavior under perturbation, such as stress. This was particularly successful for the oxygen-vacancy, observed in stressed DLTS and stressed EPR [Watkins,1961; Meese,1983].

2. Optical Techniques

Optical absorption and photoluminescence studies of transition metals have been successful in the III-V semiconductors. This success can be attributed to the relatively large bandgaps (>1eV) compared to the relatively small crystal-field splitting (≈1eV). This situation makes it possible to observe intracenter absorption and luminescence at photon energies below the bandgap energy. The intracenter transitions occur between states split by the crystal field. These techniques therefore provide the most accurate experimental values for the magnitude of crystal-field splittings. The high energy resolution of optical techniques furthermore allows the detection of small splittings of initial or final states, which yield information on site symmetry and Jahn-Teller relaxations. Optical techniques are also easily used in conjunction with uniaxial stress [Hayes,1979; Hayes,1980; West,1980] and hydrostatic stress [Devaud,1984] to yield stress couplings.

Photo-ionization transitions give values for the energy level of a defect, but the ionization threshold for deep levels tends to be broad and the absorption weak. Electrical characterization is better suited for measuring defect energy levels. This is also true for measuring stress effects. In Chapter

V, the effects of stress on ionization energies of several transition metals in GaAs are discussed.

3. Energy Levels

Several reviews give experimental energy-level positions for many transition-metal impurities in semiconductors. (For a review of transition metals in III-V compounds see Clerjaud[1985] and for Si see Chen[1980] and Weber[1983]). Two notations in common use describe the charge state involved in a defect energy level: ionic notation and charge-state notation. In charge-state notation, the total electronic charge of the defect is considered. A donor level involves a neutral to positive charge-state transition, denoted as (0/+). An acceptor level is denoted as (-/0). The terms "donor" or "acceptor" for transition metals are anachronistic and refer only to the sign of the charge states involved.

It is interesting to compare the chemical trends of the 3d transition metals in GaAs with the simple model of Ludwig and Woodbury. This is done in Fig.(1.2) for the single acceptor levels of 3d transition metal impurities in GaAs. These energy levels involve the TM^{3+}/TM^{2+} charge states. The d-shell occupancy for the TM^{2+} charge state is presented at the bottom of the figure for comparison. Several interesting features are immediately apparent. First, the electronic states are deeper for higher Z. This leads to a decrease of the energy level in the bandgap as one goes from Ti to Ni. Second, the effect of the crystal-field splitting Δ_{cf} is seen as one finishes filling the e states at Ti and Co and begin filling the t_2 states at V and Ni. Third, the exchange interaction J affects the energy difference between Mn and Fe as the d-shell is half filled with one spin, and a reverse spin is added to the e state. The Coulomb energy, U, not shown in the figure, determines the energy differences between different charge

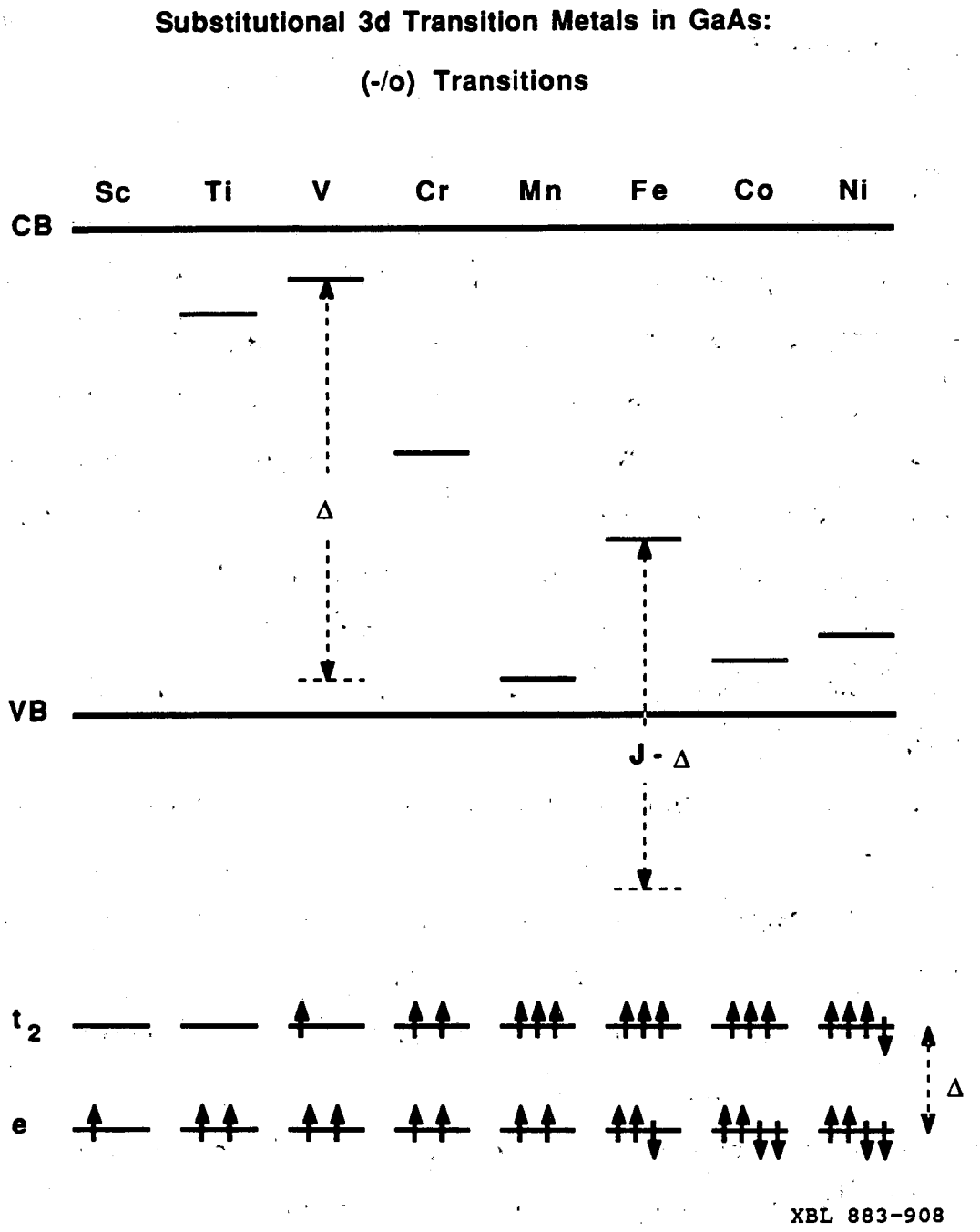


Fig. 1.2 Substitutional 3d transition metal single acceptor levels in GaAs. The figure at the bottom shows the d-shell occupancy for the TM^{2+} ionization states.

states of a given impurity. This Coulomb energy is U ≈ 0.2 eV, which is very close to the Coulomb energy of the vacancy [Picoli,1984].

The comparison between transition metals and the vacancy is not accidental. Many of the properties of the transition metals can be understood from the viewpoint of an impurity occupying a vacancy. This interpretation in fact plays a fundamental role in defining a transition-metal-based reference level, which is discussed in section III of this chapter, and which has been extended to apply to stress properties as well [Nolte,1987c]. These new aspects are discussed in length in Chapter V.

C. Theoretical Predictions

1. Cluster Calculations

The first realistic theoretical studies of the properties of transition metals in semiconductors were performed by cluster calculations. The accuracy of cluster calculations is limited by the difficulty in defining the bandgap. However, these calculations give reasonable chemical trends and they do include the interaction of the impurity states with the bulk states of the host atoms. This interaction is largely ignored in tight-binding calculations. Calculations of substitutional transition metals have been performed for 3d metals in Si [Hemstreet,1977] and GaAs [Hemstreet,1980] and 5d metals in Si [Alves,1986]. The interstitial species has been calculated in Si for 3d transition metals [DeLeo,1981; DeLeo,1982].

One of the important results of these calculations is the identification of the degree of mixing between the impurity atom states and the host states. Fully occupied states that are highly localized on the transition-metal atom and have strong atomic-like character are buried deep in the valence band. States of t_2

and e symmetry at the top of the valence band interact with these deep atomic-like states and are repelled into the bandgap. These bandgap states are partially occupied and produce the electrically active energy levels observed experimentally. The recognition that the gap states are host-like, while the atomic-like levels are buried deep in the valence band was an important conceptual breakthrough that highlighted the importance of the band structure in determining the structure of deep-level defects. Furthermore, for substitutional transition metals, the defect states in the gap may take on much of the character of the vacancy. This leads to the vacancy model of Watkins[1983] in which the structure of substitutional transition metals could be understood largely as a vacancy occupied by a transition-metal impurity that did not strongly alter the dangling bonds of the vacancy.

The weak interaction of the transition-metal states with the vacancy has been explained using inter-electron Coulomb energy in the context of tight-binding theory [Picoli,1984]. The Coulomb energy change caused by adding an electron to the transition-metal d-shell is roughly 8 eV, while the Coulomb energy change caused by adding an electron to a vacancy dangling bond is only 0.2 eV. It is therefore much more energetically favorable to add or remove electrons to and from the dangling bonds than from the d-orbitals. There is mixing between the transition-metal orbitals and the dangling bonds of the vacancy, but the defect eigenstates are mostly vacancy-like. The vacancy Coulomb energy of 0.2 eV is reflected in the energy difference between successive ionization stages of a single impurity and explains the occurrence of several different charge states in the bandgap.

2. Green's Function Method

While the results of cluster calculations provide important physical insight into the electronic structure of transition-metal defects, the results are not highly accurate. The cluster calculations also have difficulty in defining the role that the conduction band plays in the defect structure. These deficiencies are mostly overcome in Green's function calculations. One of the main strengths of the Green's function approach is the automatic inclusion of the correct band structure of the host crystal. This overcomes the difficulty that the cluster approximations have in defining the bandgap and gives an accurate representation of the conduction band. The Green's function theory of 3d transition metals in semiconductors has been carried out in detail by Zunger et al.[1982,1983,1986]. The results are qualitatively similar to many of the properties of the cluster calculation, although more accurate. In one aspect the Green's function approach is superior; this is the identification of the role of the conduction band in the transition-metal impurity structure.

The results of the detailed Green's function calculations can be summarized in terms of a three-level model that describes the gross features of the electronic structure of substitutional transition metals. The three levels in this model are: 1) the energy of the maximum density of states in the upper valence band; 2) the energy of the maximum density of states in the lowest conduction band; and 3) the transition-metal d-electron energy. In this model three linear combinations of these states result: a bonding combination, an antibonding, and a non-bonding combination. The bonding and antibonding combinations are called crystal-field resonances (CFR's) in the valence and conduction bands respectively. The nonbonding states are the gap states and are called dangling-bond hybrids (DBH's). The transition-metal gap states have strong dangling-bond character, similar to the vacancy. This is the same

conclusion arrived at by tight-binding and cluster calculations, with an important exception. For high or low Z, the transition-metal level does not pin to the vacancy level, as assumed by earlier models.

The three-level model plays a central role in the explanation of the universal pressure derivative of substitutional transition metals, discussed in Chapter V. The connection between the three-level model and the band deformation potentials is made in Chapter VI.

III. Hydrostatic Deformation Potentials and Heterojunction Band Offsets

The band structure of a solid is altered by perturbation. The perturbation can include changes in composition, as for the tertiary compound $Al_xGa_{1-x}As$, or changes in lattice constant, as under pressure. Because the bandgap of a semiconductor plays the central role in determining the electronic properties of the material, it is especially important to know how a perturbation will affect the edges of the bandgap.

A. Band-Edge Deformation Potentials

Stress, pressure, or thermal vibrations can alter the lattice spacings of the semiconductor host atoms. When the symmetry of the lattice is reduced, degenerate band states may split. These splittings are easy to measure experimentally. For this reason, the shear and normal deformation potentials of the states at the edge of the bandgap have been well characterized. For symmetry-preserving compression or expansion, there are no splittings, only shifts. It has been easy to measure the difference in the shifts of the edges of the bandgap induced by pressure, but the hydrostatic shifts of the individual band edges had eluded direct measurement. The hydrostatic deformation

potentials of the band edge contribute to several important phenomena including deformation-potential scattering of carriers.

1. Deformation-Potential Scattering of Carriers

The theory of deformation-potential scattering was developed by Bardeen and Shockley[1950]. In the limit of long-wavelength acoustic phonons, the phonon displacement field can be replaced by a slowly modulated strain tensor. In this limit, the effect of the phonons on the band structure and the effects of stress are equivalent. Therefore, the change in band energy measured by applied stress is proportional to the scattering potential experienced by an electron in the presence of a phonon field.

The perturbation potential is

$$\delta\varepsilon = \Xi_{eff} (\nabla \cdot \delta \mathbf{r}) ,$$

where Ξ_{eff} is the effective acoustic deformation potential and $(\nabla \cdot \delta \mathbf{r})$ is the strain induced by the phonons. The perturbation Hamiltonian for scattering an electron from the electron state with k-vector **k** to k-vector **k'** is

$$| H_{k,k'} | = \Xi_{eff} [k_B T / 2 V c_l]^{1/2} ,$$

where c_l is the longitudinal elastic constant and V is the volume of a unit cell of the crystal. Inserting the scattering Hamiltonian into the Golden Rule leads to a mean electron-phonon scattering time. The expression for acoustic-phonon-limited mobility becomes

$$\mu = \frac{2\sqrt{2\pi}\, e\, \hbar^4 c_l}{m^{*5/2} k_B^{3/2} \Xi_{eff}^2} T^{-3/2}$$

The effective deformation potential is a combination of normal, shear and hydrostatic deformation potentials. For the case of the valence band this is

$$\Xi_{eff}^2 = a^2 + \left(\frac{c_l}{c_t}\right)\left[b^2 + \frac{1}{2}d^2\right]$$

While the normal and shear deformation potentials b and d are well known, the hydrostatic component, a, must be fitted from mobility data. Such fitting has lead to a wide scatter of values that vary by more than a factor of 2 for the valence band of GaAs. For the technology of high-purity and high-mobility field-effect transistors it is important to know the hydrostatic deformation potential in order to place limits on acoustic-phonon-limited mobility. With the lack of experimental evidence for the hydrostatic deformation potential, the values have been estimated theoretically.

2. Theories of Hydrostatic Deformation Potentials

Early theories of hydrostatic deformation potentials were based on the dielectric theory of the chemical bond [Phillips,1970; Camphausen, 1971] and on the linear combination of atomic orbitals (LCAO) [Harrison, 1980]. These theories yield good values for differences, but are not reliable for absolute deformation potentials. It was sometimes assumed, without justification, that the conduction-band deformation potential contributed the majority to the bandgap deformation potential. Though this assumption was unfounded, it was in fact valid [Van de Walle,1987a]. More realistic calculations of absolute deformation

potentials involved band structure calculations using orthogonalized plane waves (OPW) [Collins, 1970] and linearized muffin-tin orbitals (LMTO) [Verges, 1982] methods, but the deformation potential values obtained were too large to be consistent with mobility experiments.

A breakthrough in the theory of hydrostatic deformation potentials came in 1987 from two independent sources. Van de Walle and Martin[1987b] succeeded in performing first-principles, self-consistent pseudopotential calculations of band-edge hydrostatic deformation potentials. Simultaneously, Cardona[1987] argued that the earlier LMTO deformation potentials must be screened by the dielectric constant of the material before they can be used to calculate mobility. These two theoretical approaches yielded similar results and were consistent with mobility experiments. In Chapter V, these values will be shown to be consistent with the universal pressure derivative of transition-metal impurities.

B. Heterojunction Band-Edge Line-Up

A problem similar to finding the band-edge displacements under stress is the problem of finding the band-edge offset at a heterojunction interface. When two dissimilar semiconductors are in contact along a defect-free interface, the conduction-band and valence-band edges will not in general line up across the interface. The difference between the band edges is called the band offset or band-edge discontinuity. This offset occurs over only a few lattice constants, compared to the thousand angstrom range of band-bending caused by the equilibrium of the Fermi level. The band-edge offset is therefore an intrinsic property of the semiconductor interface and is not related to doping or electrical properties.

The band offset is the fundamental component of bandgap engineering [Capasso, 1986,1987]. It defines the potential step experienced by charge carriers at the interface. This potential step can be used for confinement of carriers to two, one or even zero dimensions. It can also accelerate or decelerate hot carriers that travel perpendicular to the interface. The band offset need not be discontinuous, but varies continuously in compositionally graded semiconducting devices. For these reasons it is of fundamental importance to know the values of band offsets between the many different semiconductors. However the band-edge offset has been difficult to define both theoretically and experimentally. Experimental difficulties involve space-charge effects, lattice mismatch, interface defects and other characteristics of the interface that are difficult to control. The theoretical approaches have concentrated on the problem of defining a suitable reference point from which to measure the change in the band structure as the material is altered at the interface.

1. The Neutrality Level

Early attempts at defining band offsets analyzed the problem from the aspect of surface polarization at the interface. A dipole layer at the interface would adjust the line-up across the interface [Tejedor,1978;Flores,1979]. The theory involved the analysis of the two specific materials in contact. The calculation of a surface dipole was generalized by Tersoff so that a reference energy could be defined for each material which would line up across an interface. This reference energy was called the charge neutrality level. The neutrality level was originally defined by the energy for which the contribution of the conduction band to the crystal Green's function equaled the contribution of the valence band [Tersoff,1984a,1984b]. The computation of the Green's

function could be replaced by a simple rule of thumb that located the neutrality level at the midpoint of the indirect gap [Tersoff,1985].

2. Transition-Metal Line-Up across Heterojunction Interfaces

An interesting feature of the energy levels of transition-metal impurities is the fact that the relative energy spacings are the same in all isovalent semiconductors [Ledebo,1982; Caldas,1984]. The source of the invariance was not well understood, though there was the suggestion that the transition-metal levels tracked with the vacuum level. Langer[1985] recently proposed that transition-metal levels could be used to predict heterojunction band offsets. This conjecture was tested by comparison with experimental offset data from AlGaAs-GaAs heterojunctions, and the agreement was favorable. This specific heterojunction is well behaved because the materials are lattice matched. Data from other heterojunctions are not as consistent and the transition-metal-based reference level was only tested for this one heterojunction system. A theoretical attempt was made to explain the relation between the transition metals and the reference level [Tersoff,1987] and it was found that the transition metals are locked to the neutrality level within an additive constant if the Coulomb energy of the d-electrons is much larger than the Coulomb energy associated with the vacancy.

In Chapter V, the transition-metal-based reference level is extended to include the effects of stress. Transition-metal defects in GaAs possess a universal pressure derivative. This pressure derivative is compared against the best first-principles calculations of band-edge deformation potentials and against the most recent mobility measurements. Excellent agreement is found in both cases.

Chapter II. Stressed Deep Level Transient Spectroscopy

One of the methods to probe the structure of a defect is to perturb the environment of the defect and measure how the defect reacts. Among the common perturbations are: electric fields, magnetic fields, polarized electromagnetic fields, and strain fields. Uniaxial-strain fields, in particular, are useful to determine defect symmetry because the strain can reduce the local symmetry of the defect and thereby break possible defect degeneracies. The magnitudes of the strain splittings yield information about the axial defect-lattice coupling which is closely related to the mechanism of Jahn-Teller distortion. Uniaxial stress has been commonly used in conjunction with optical spectroscopy and with electron paramagnetic resonance. The combination of uniaxial stress with deep level transient spectroscopy (DLTS),on the other hand, has been a recent development [Meese,1983]. One reason for the late appearance of stressed DLTS is that DLTS suffers from limited energy resolution (roughly 10% of the defect binding energy). Therefore defect splittings must be large for the stress effects to be measurable. Despite this handicap, DLTS has the advantage of high sensitivity to small concentrations of deep-level defects, an advantage which is not shared by optical techniques or by EPR. In addition, some deep-level defects do have extraordinarily large stress splittings that are easily measured by DLTS. In this chapter, the details of stressed DLTS will be developed, including optimization of the energy resolution of DLTS and the treatment of thermal emission from stress-split states.

I. Space-Charge Spectroscopy

One of the fundamental goals of the experimentalist involved with deep-level defects is to observe the physical properties of a specified defect and from these properties deduce the structure of the defect. The defects can have macroscopic properties (such as total concentration), or microscopic properties (such as symmetry or electron-lattice coupling). Several techniques are commonly used to measure the properties of deep-level defects. Each technique has its own strengths and weaknesses and specificity for some of the properties that a deep level can exhibit. The space-charge techniques are sensitive to all electrically active deep levels and are able to measure very low defect concentrations. Space-charge techniques rely on the space-charge layer in a semiconductor beneath a rectifying contact. Deep-level defects modify the space charge and this modification can be measured as current, admittance, or capacitance of a semiconductor diode. The best known space-charge spectroscopy is deep level transient spectroscopy (DLTS). Space-charge spectroscopy is a thermal ionization spectroscopy; it measures the thermal emission of carriers trapped in deep levels. The thermal-emission probability depends on the temperature and on the activation energy for ionization. The free variable of this spectroscopy is temperature (related to the average phonon energy in the crystal) in the analogous manner that photon energy is the free variable of optical spectroscopy. Because space-charge spectroscopy is an ionization spectroscopy, it is only capable of measuring the differences in properties between two different charge states. The defect energy level in the bandgap is therefore related to a difference in energy between the two specific charge states.

A. Biased Junctions in Semiconductors

Rectifying contacts are produced when the Fermi level of a material in contact with a semiconductor is different from the Fermi level of the semiconductor. To reach equilibrium, charge is transferred between the materials until the Fermi levels match across the interface. The charge transfer depletes the free carriers from a region below the contact. This is called the depletion layer. The width of this layer depends on the voltage difference between the two materials and on the dopant concentration of the semiconductor. In the case of a Schottky contact, the depletion of the metal is negligible and the junction is one-sided on the semiconductor side. The width of the depletion layer is given by (MKS units)

$$w = \sqrt{\frac{2\,\varepsilon\,\varepsilon_0\,(V_{bi} + V_{ap})}{|N_A - N_D|\,e}} \quad , \qquad (2.1)$$

where ε is the dielectric constant, V_{bi} is the original voltage difference between the semiconductor and the contact material, V_{ap} is the externally applied voltage, and $|N_A - N_D|$ is the density of the space-charge in the depletion layer arising from the ionized dopants. Because the depletion layer supports no charge carriers, it behaves as a dielectric layer of width w between the contact and the bulk semiconductor -- both of which are conducting materials. The capacitance of this dielectric slab is

$$C = \frac{A\,\varepsilon\,\varepsilon_0}{w}$$

$$= A\sqrt{\frac{\varepsilon\,\varepsilon_0\,e\,|N_A - N_D|}{2\,(V_{bi} + V_{ap})}} \quad , \qquad (2.2)$$

where A is the area of the contact.

Deep-level defects present in the depletion layer modify the steady-state space-charge density. If carriers are injected into the depletion region, the deep levels will trap the carriers. After the carrier injection pulse stops, the trapped carriers on the deep levels will be thermally emitted with a mean time τ and will be swept out of the depletion region by the electric field. For the case of an electron trap, the capacitance as a function of time is

$$C = C_0 \sqrt{1 - \frac{N_T \exp(-t/\tau)}{|N_D - N_A|}} \quad , \qquad (2.3)$$

where C_0 is the steady-state capacitance and t is measured from the time the injection pulse stops. The presence of deep levels can therefore be detected by a relaxation of the junction capacitance after a forward bias pulse. The relaxation time τ depends on the defect activation energy and on the temperature.

B. Detailed Balance and Emission Rates

The relaxation rate is derived from the principle of detailed balance. In equilibrium, the rate of thermal emission from a deep level is equal to the rate of carrier capture

$$e_b = n v \sigma s \, , \qquad (2.4)$$

where e is the emission rate, b and s are the occupied and unoccupied defect fractions, n is the number of carriers in the band with thermal velocity v, and σ is the capture cross section. The occupation fractions are related through

$$b/s = \gamma \exp(-(E_T - E_f) / k_B T) , \qquad (2.5)$$

where γ is the degeneracy factor of the defect level. The emission rate is therefore

$$e = \gamma^{-1} n v s \exp (-E_f - E_T) / k_B T) . \qquad (2.6)$$

This is rewritten using the effective density of states

$$n = N \exp (E_{band} - E_f) / k_B T) \qquad (2.7)$$

yielding

$$e = \gamma^{-1} N v \exp(- \Delta G / k_B T) , \qquad (2.8)$$

where ΔG is the free energy of ionization to the band edge. The free energy consists of an enthalpy term and an entropy term. The temperature dependence of the emission rate yields the activation enthalpy of the defect.

C. Deep Level Transient Spectroscopy (DLTS)

The space-charge techniques were developed in broad form by Sah[1970]. A convenient method of analyzing the transient response of the sample diode as a continuous function of temperature was developed by Lang[1974]. This analysis technique is called the rate-window technique and in

this form the space-charge technique became a true spectroscopy, called deep level transient spectroscopy. This spectroscopy of deep levels is described by Miller and Lang [Miller,1977; Lang,1979]. Many varieties of transient spectroscopy have been developed [Sah,1981] which measure capacitance [Miller,1977], current [Borsuk,1980], or charge [Farmer,1982] transients and these systems have been optimized with respect to noise [Miller,1975], spatial resolution [G. Li,1985] and accuracy of the energy determination [Thurber,1982]. Despite the fact that DLTS is a thermal spectroscopy which relies on probabilistic thermal averages rather than the absolute energy differences observed in optical transitions, the energy determinations can be quite accurate when temperature and electric-field effects are taken into account. However, the broad response linewidth of standard DLTS as a function of temperature severely restricts the energy resolution of the technique when defect levels are closely spaced in energy. In this case, the DLTS signal is composed of multiple exponentials. The problem of DLTS resolution has been studied in detail [Nolte,1987b] and certain aspects will be discussed in a following section under signal analysis.

II. Experimental Apparatus

A. Temperature Control

Cooling is provided by a CTI-Cryogenics refrigerator, Model 21 Cryodyne® Cryocooler. This system is a closed-cycle sterling-engine refrigerator that uses helium as the working gas. The refrigerator can reach a minimum temperature of 10 K. Above 80 K it has between 5 - 10 Watts of cooling power. Below 80 K the cooling power drops continuously to zero at 10 K. The closed-cycle refrigerator has the advantage that it requires no liquid refrigerant. The base

temperature of 10 K is sufficient to observe thermal emission from defect levels with binding energies greater than 20 meV. This allows for the observation of all deep levels and even some effective-mass levels. Only shallow levels are inaccessible to observation in this temperature range.

Heating is controlled by introducing helium exchange gas into the vacuum chamber around the CTI cold head. When the cold head is below room temperature, the gas conducts heat from the outside. The exchange gas also plays the important role of equilibrating temperatures within the sample chamber. Additional heating is supplied by a 25 Watt power resistor. The maximum temperature can reach 430 K.

The temperature is measured by a Lakeshore Cryotronics DT-470 series temperature diode. This sensor is calibrated to a base temperature of 1.4 K and a maximum temperature of 475 K. The diode is calibrated through a standard voltage vs. temperature curve. All DT-470 series sensors track this curve with a tolerance between 0.25 K and 2 K. Repeatability is within ±10 millikelvin over multiple thermal cycles. The V vs. T curve is contained in a programable read only memory (PROM) in a DRC-80 digital thermometer. The output from the DRC-80 is either an analog voltage proportional to the temperature, or a binary coded digital (BCD) output with 16 binary data channels.

B. DLTS Stress Apparatus

The uniaxial stress experiments described in this thesis were performed by a DLTS uniaxial stress apparatus. The stress apparatus was designed with two constraints: the thermal and mechanical constraints of the CTI cold head, and the mechanical constraints of 1 GPa.

The first constraint arises from the attachment of the stress apparatus to the top of the CTI cold head within the system vacuum chamber. The means of

applying stress must be communicated to the sample stage through the vacuum, requiring a vacuum feedthrough. Also, the CTI cold head is supported by thin-walled stainless steel which cannot support any significant load. The stress rig therefore was designed to contain the stress internal to a stress box, with no load on the cold head. The cold head is thereby isolated from any load which might occur if there is a mechanical failure. Because of the low cooling power of the CTI, the stress-rig mass must be kept to a minimum using a material with a high thermal conductivity. The second constraint is the maximum load which must be delivered to the semiconductor sample. Of the common semiconductors, silicon has the highest yield strength of roughly 1 GPa. This pressure exceeds the yield strength of the 300 series of stainless steels.

A cross section of the DLTS stress apparatus is shown in Fig.(2.1). The stress box is situated on top of the CTI cold head. Stress is applied through a spring and lever system. The spring is capable of delivering a maximum force of 100 Newtons. This is multiplied by a factor of 10.7 by the lever arm. The sample is vertical in the sample holder and is compressed by a piston of hardened steel. The sample cross section is 1 mm^2. The stress apparatus can therefore supply up to 1 GPa of uniaxial stress. The spring is compressed by a screw pushing on a thrust bearing on top of the spring. The screw is turned by a screw driver that enters the vacuum chamber through a vacuum feedthrough. As force is applied to the lever, the lever flexes and the endpads between the piston and the sample compress. Therefore to measure the deflection of the spring, a test rod runs through the screw driver to the bottom of the spring. By measuring the relative displacements of the test rod and the top of the screw driver, the absolute compression of the spring can be measured without interference from internal flexing or compressing.

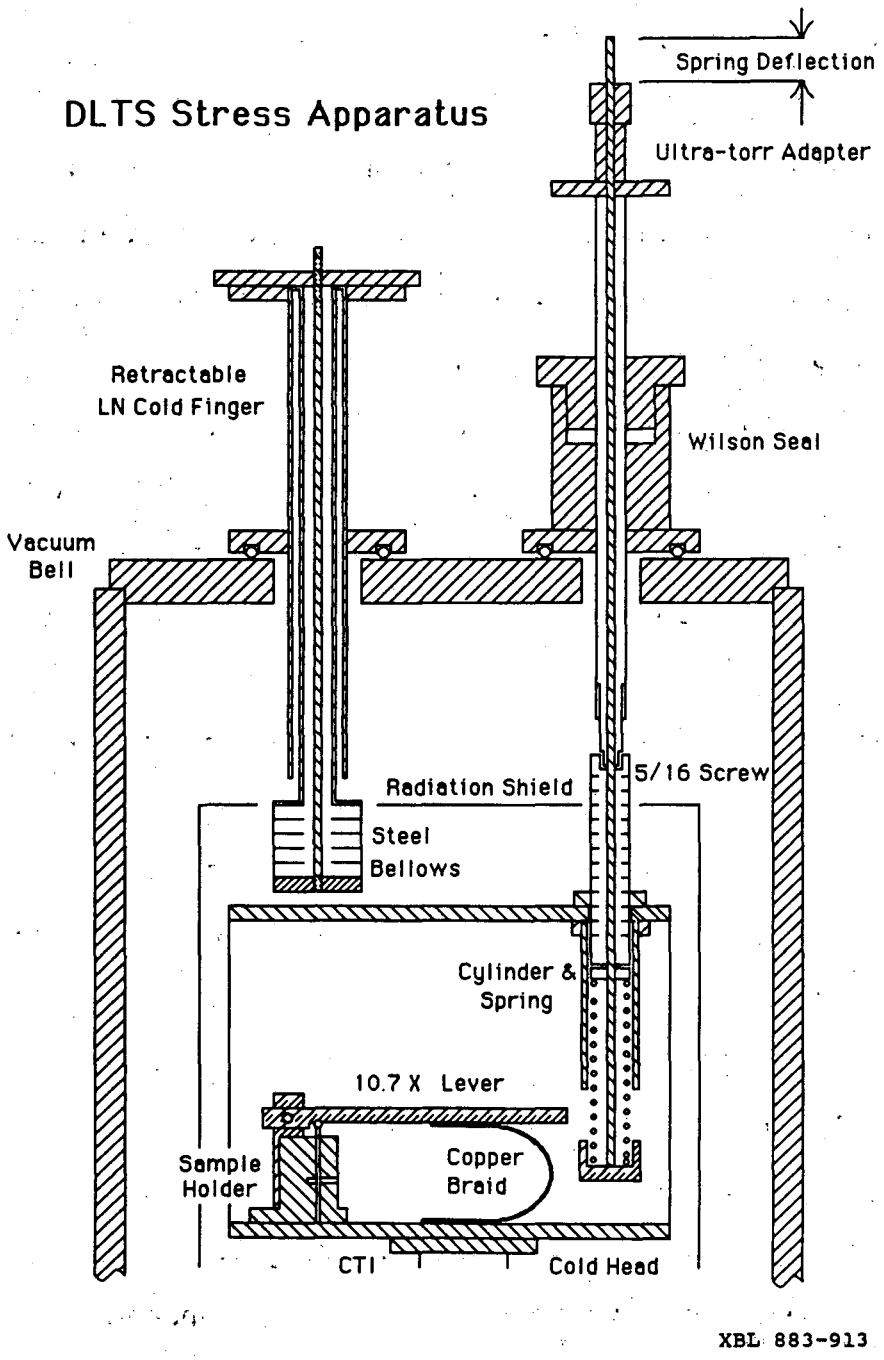


Fig. 2.1 Schematic of the stressed DLTS apparatus.

The requirement of low thermal mass, high thermal conductivity and high strength was satisfied by aluminum 6061 alloy. The stress box is constructed of this aluminum, except for the steel sample holder and the steel spring. Stainless steel extrudes at 1 GPa, so the piston pushing on the sample was made of 17-4 PH steel, thermally hardened at 900 ^{0}C for 1 hour to a Rockwell hardness of 50. The thermal properties of the stress rig were improved by the addition of a copper braid to aid the cooling of the lever arm, and by a retractable liquid nitrogen cold finger.

The stress calibration is an important aspect of the stress apparatus because one of the main thrusts of this thesis work was to obtain absolute stress derivatives. The temperature dependence of the spring constant was measured externally by applying a direct load on the spring and measuring the deflection at several temperatures. The spring-constant data are shown in Fig.(2.2). The data were taken at room (295 K), dry ice (195 K), liquid nitrogen (77 K) and liquid helium (4 K) temperatures. Within the apparatus, this calibration is checked by applying stress to a standard sample. The oxygen thermal donor in silicon is an excellent test sample because it shows dramatic splitting under moderate stresses. Stress data for the thermal donor are shown in Fig.(2.3). The model for the thermal donor predicts a separation of 92 meV/GPa between the peaks. This pressure derivative was verified experimentally by two independent groups [Benton,1984; Henry,1984]. From this standard sample it is possible to observe a slow increase of the spring constant with repeated use caused by stress hardening. The estimates of the errors in the parameters that determine the value of the applied stress are: deflection of spring = 2%, sample area = 1%, spring constant = 3% and lever arm = 5%. These errors add to a 6% error in external calibration. By using the thermal-donor internal calibration, the last two errors are systematically reduced. Though these errors are small, a

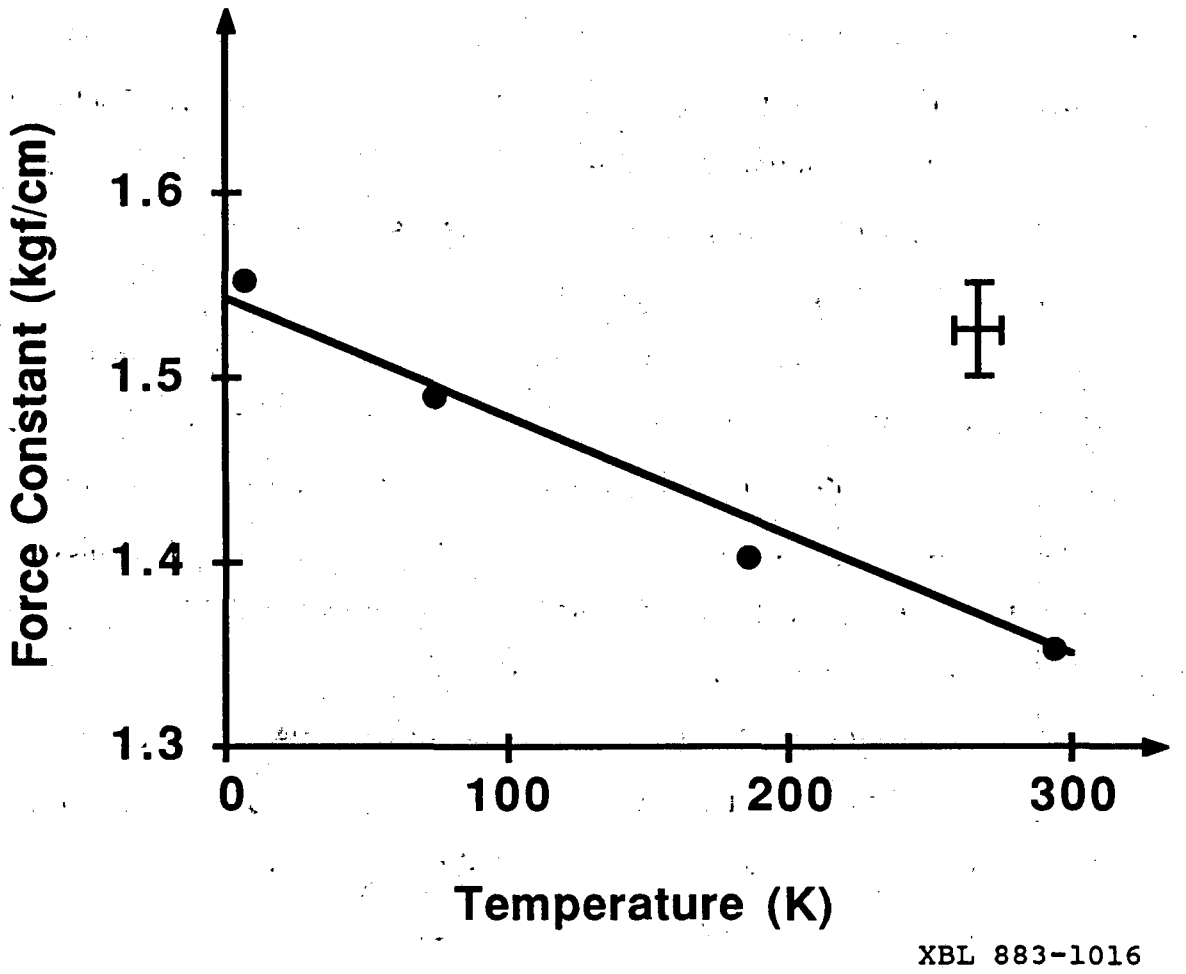


Fig. 2.2. Spring constant as a function of temperature.

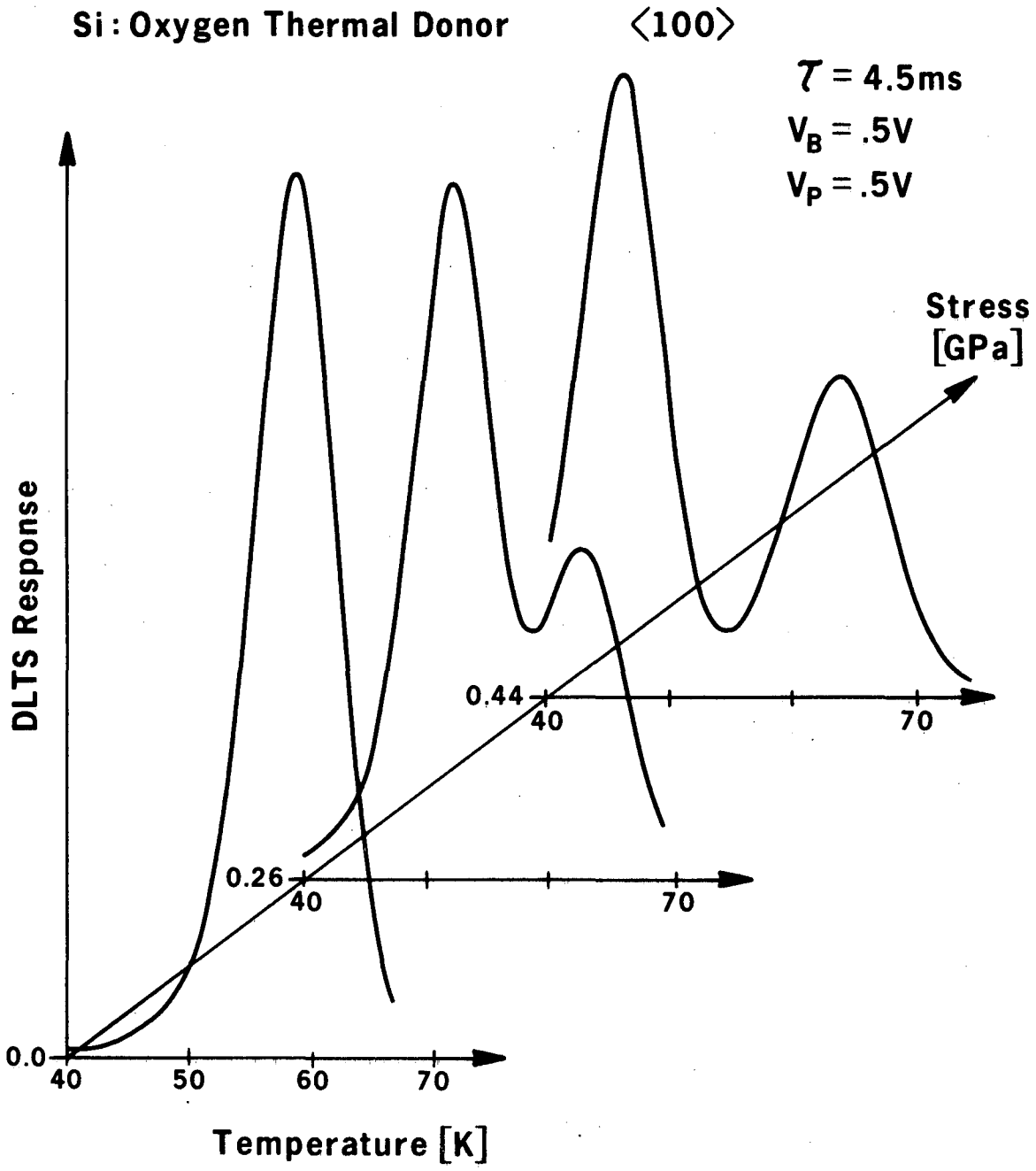


Fig. 2.3 Stressed DLTS data for the oxygen thermal donor in Si. The sample is stressed along the [100] axis. The splitting is 92 meV/GPa.

substantially larger error is associated with the mounting of the sample. This emphasizes the importance of experimental technique.

C. Experimental Technique

The stress apparatus applies up to 100 Newtons on a sample with an area of 1 mm^2 and a height of 6 mm. The alignment of this sample is therefore difficult, but of critical importance. A 10% inhomogeneity in the stress on the sample introduces a 10% additional shear strain that would not be present if the sample were perfectly loaded. This 10% effect can overwhelm the smaller measurement errors for systems that are sensitive to shear stress. To ensure proper vertical mounting of the stress sample, the sample is lightly pressed against the side of a vertical anvil, machined with a tolerance of 2 mil. Endpads of index card paper cushion the ends of the brittle sample from the steel pistons. These endpads compress and distribute the load evenly over the ends of the sample. The pistons and their slots are machined vertical, again within 2 mil tolerance, to remove any play in the vertical alignment. This close tolerance required that the sample holder be made from steel rather than 6061 aluminum because steel more closely matches the thermal contraction of the hardened steel pistons.

DLTS probes the near surface of the semiconductor sample. Therefore, to obtain an accurate measurement of the stress shift of a defect in the sample it is necessary to measure the defect signal on both sides of the sample. This is done by forming rectifying contacts on two opposing surfaces and switching the sign of the bias. For a given bias, one contact will be under reverse bias and the other will be under forward bias. Though the forward-biased contact contributes to the total capacitance of the sample, it does not contribute to the ΔC which arises from the deep levels in the depletion layer of the rectifying

contact. In this way, a stressed DLTS spectrum is taken of both surfaces of the diode and averaged to yield the stress derivative of the defect. This technique also gives a direct measurement of the stress inhomogeneity. The typical inhomogeneity is around 5%, although it can be as high as 10%.

Temperature tracking is the last important technical problem with stressed DLTS. Because the sample is subject to such strict constraints in terms of alignment and stress, the temperature diode cannot be placed near the sample. Furthermore, the steel sample holder has low thermal conductivity and some amount of steel lies between the sample and the temperature sensor. When the system is under vacuum, this steel is the only thermal link between the two. Therefore, during cooling under vacuum there is a lag in the temperature experienced by the sample and the temperature read by the thermometer. During warming, on the other hand, the helium exchange gas provides a good thermal link that equilibrates the temperatures within the stress box. For this reason, data are only taken while warming up under a helium exchange gas.

D. Signal Analysis

The capacitance of the sample diode is monitored by a 1 MHz Boonton model 72B capacitance bridge. A schematic of the bridge is shown in Fig.(2.4). The DC bias is applied to the HI terminal, while the pulse is applied to the LO terminal of the center-tap coil. The output signal can be analyzed by either analog circuitry or by computer data acquisition. The advantage of the analog circuitry is the high repetition rate and the corresponding high signal-to-noise ratio. The advantage of the data acquisition is the ability to capture and analyze the capacitance transient. Because of these relative strengths, the two analysis methods are used for different purposes. For broad temperature scans to survey the deep levels present in a sample, the analog system is used because

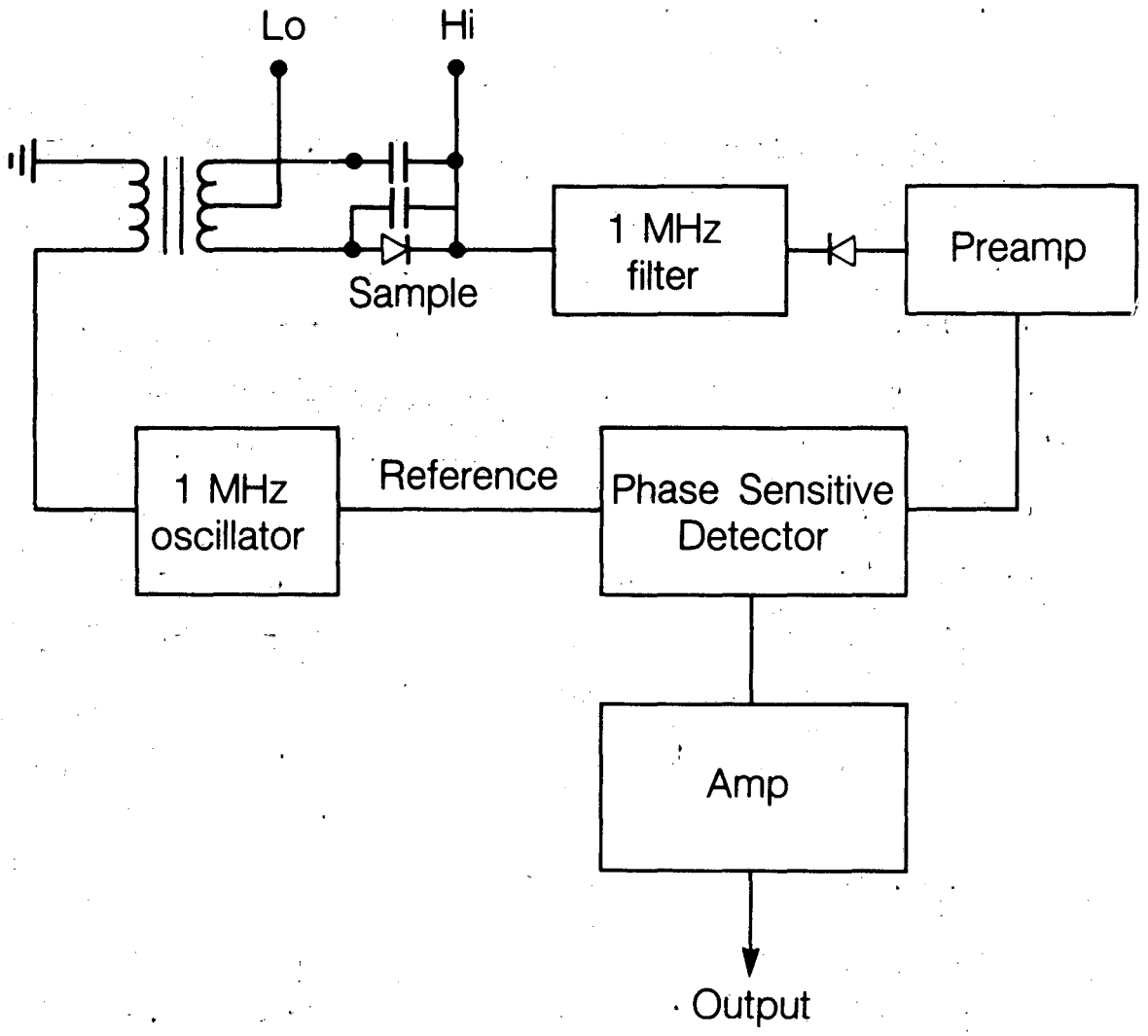


Fig. 2.4 Schematic diagram of the Boonton capacitance bridge.

it can run unattended and has a superior signal-to-noise ratio which can pick out low concentrations of defects. For precise stress measurements, the computer system is used because the transient can be observed directly and the decay-time constant measured as a function of stress. The computer system is also superior for finding activation energies because a single thermal scan will give the full Arrhenius plot. With the analog system, several thermal scans would be required to get several points of the plot.

1. Computer Data Acquisition

In the computer data acquisition mode the transient signal is captured and digitized by a Tektronix Model 2230 digital storage oscilloscope. This digitizer has 8-bit resolution. Successive acquisitions can be averaged in the data buffer in the scope which reduces the effect of the discrete signal. The scope has a minimum time-base of 5 ms/div for single sweep signals, and 0.05 ms/div for repetitive storage. The record length for a single sweep can be 1 Kbyte or 4 Kbyte. The 1 Kbyte record fills the scope screen, so this record length is used for the DLTS application. The averaged signal is transferred to an IBM AT personal computer on a general purpose interface bus (GPIB). The temperature from the DRC-80 thermometer is read from the BCD output of the thermometer by the IBM PC data acquisition and control adapter. A block diagram of the computer-controlled data acquisition is shown in Fig.(2,5). The transient signal time constant is fitted by the Marquardt-Levenberg non-linear least squares method to the function

$$S(t) = A \exp[-t/\tau] + B . \qquad (2.9)$$

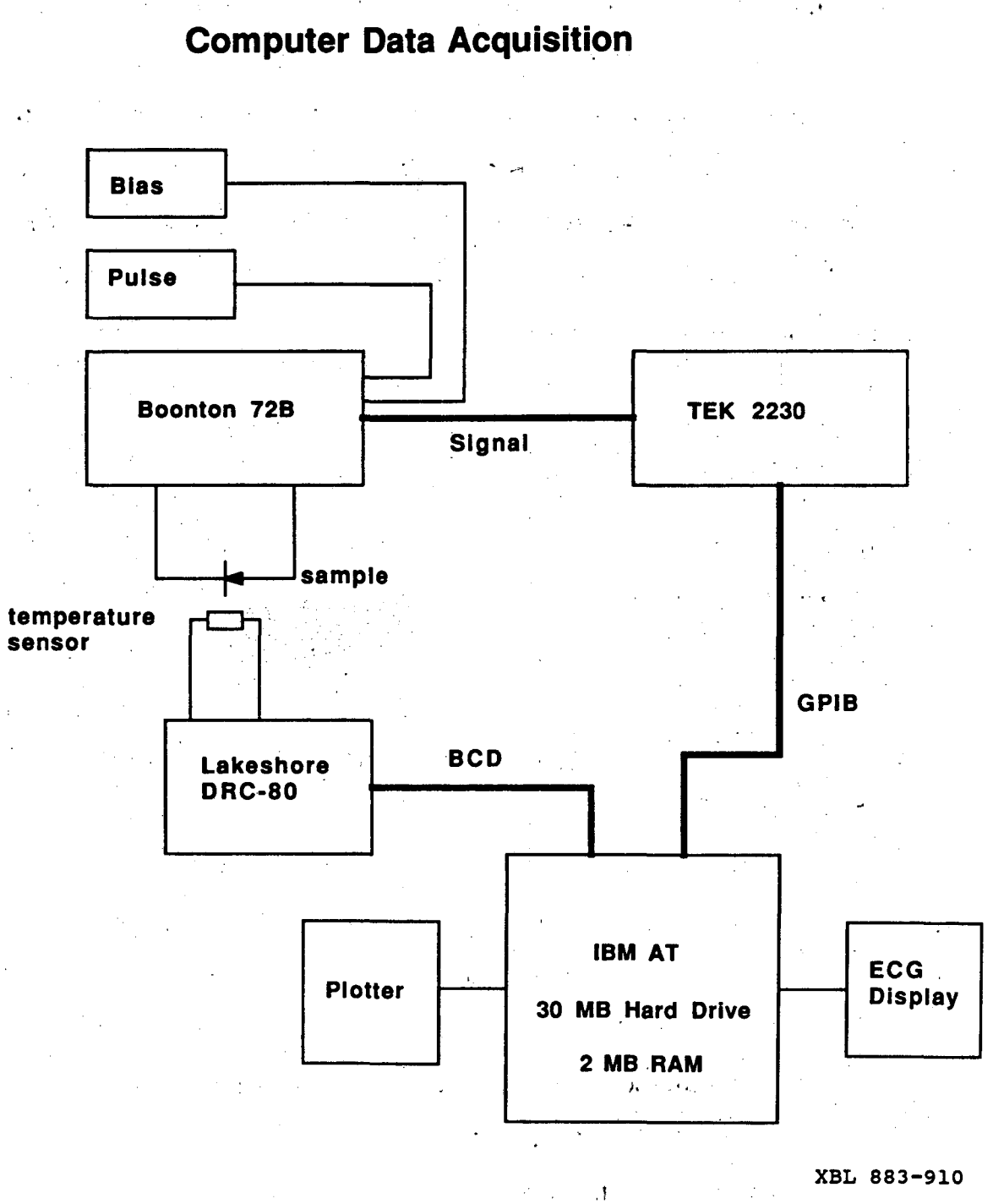


Fig. 2.5 Schematic diagram of the computer data acquisition.

The baseline shift B must be included because the scope time-base may not capture the full transient in the 1 Kbyte record.

2. Analog Rate-Window

A schematic of the analog rate-window circuit is shown in Fig.(2.6). The signal correlator triggers the pulse generator at a rate dependent upon the operator-set time constant and receives the raw signal from the Boonton capacitance bridge. The raw signal is amplified and the baseline is subtracted from the signal to yield the baseline-restored signal. The restored signal is multiplied by an exponential decay with a decay time equal to the operator-set time constant. The multiplied signal is integrated and output to an x-y recorder. The response signal is a maximum when the thermal-emission rate is equal to the inverse of the time constant and is zero when the emission rate is much larger or much smaller than this value. This is the origin of the expression "rate-window". Ramping the sample temperature leads to the deep-level defect line-shape characteristic of the rate-window analysis technique.

The details of the response lineshape have not been studied in detail before. Therefore the effects of the various operations on the lineshape are investigated and some general trends and conclusions are described which are valid for the many types of weighting functions that are used in the rate-window techniques. An absolute low bound on the resolution exists which is independent of the energy of the deep level. The low bound does depend on the density of states of the bulk material as well as on the capture cross section of the defect.

After a deep-level defect in the depletion region of a diode is filled by a forward-bias filling pulse, it will thermally emit the captured carriers into the valence or conduction band with the rate

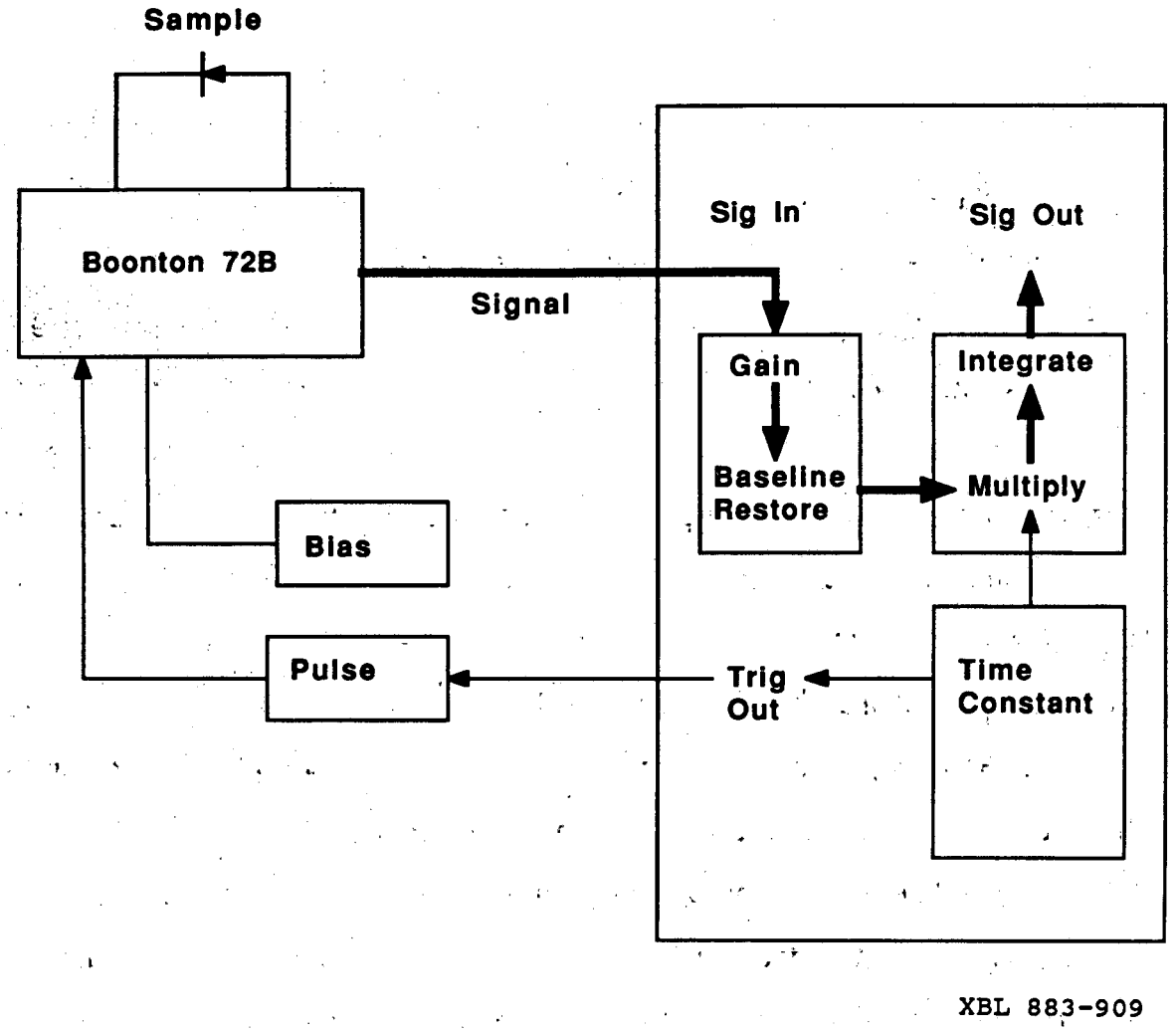


Fig. 2.6 Schematic diagram of the analog rate-window circuitry for DLTS.

$$e_n = \sigma_0 N_V v_{th} e^{-\Delta H/k_B T} , \qquad (2.10)$$

derived by means of detailed balance, where σ_0 is the capture cross section, N_V is the effective density of states at the band edge, v_{th} is the thermal velocity of carriers in the band, and ΔH is the emission activation enthalpy of the defect. The rate-window technique gives a response that peaks when the emission rate is equal to a specified value. The response is formed by applying three linear operations to the raw, measured signal. First, a baseline time t_0 is defined and the value of the signal at this time is subtracted from the raw signal. Second, the baseline-restored signal is multiplied by a weighting function with a characteristic time constant related to t_0. Third, the resultant function is integrated over time. Because the DLTS signal is composed of decaying exponentials, the second and third operations together are equivalent to performing a Laplace transform of the weighting function. The high-temperature side of the response lineshape can therefore be understood by looking at the trends of the Laplace transforms as the weighting functions are varied. The baseline restoration, however, modifies the Laplace transform for slow emission rates at the lower temperatures. The low-temperature side of the response lineshape is therefore dominated by the effect of baseline restoration.

A general, analytic expression can be derived in a straightforward manner for the full-width half-maximum (FWHM) linewidth of the response as a function of temperature. If e_1, T_1 and e_2, T_2 are the emission rates and temperatures at which the response is half of the maximum, and e_0, T_0 are the emission rate and temperature at which the response is a maximum, then the temperature resolution is given by

$$\frac{\Delta T}{T} = \ln(\sigma_0 N_0 v_0/e_0)\left\{\frac{1}{\ln(\sigma_1 N_1 v_1/e_1)} - \frac{1}{\ln(\sigma_2 N_2 v_2/e_2)}\right\} \quad , (2.10)$$

where the subscripts 0, 1, and 2 refer to the values of σ, N, v and e evaluated at the temperatures T_0, T_1, and T_2 respectively.

To facilitate the comparison between the various weighting functions and baseline restoration times, this equation can be rewritten as

$$\frac{\Delta T}{T} = \ln(\sigma_0 N_0 v_0 \tau^* t_0)\left\{\frac{1}{\ln(\sigma_1 N_1 v_1 \tau^* t_0) + \ln(x_1)} - \frac{1}{\ln(\sigma_2 N_2 v_2 \tau^* t_0) + \ln(x_2)}\right\} \quad (2.12)$$

where τ^* is the effective time constant measured in units of t_0, and x_1 and x_2 are the ratios of the half-maximum time constants divided by $\tau^* t_0$.

This equation is still quite general, yet several observations can be made regarding its qualitative behavior. First, the linewidth does not depend explicitly on the binding energy of the deep level. Second, the linewidth depends only weakly on the temperature through the temperature dependence of the density of states and the thermal velocities. Third, the exact value of the linewidth depends sensitively on the cancelation between the two terms in parentheses. Typical values for $\ln(\sigma N v \tau^* t_0)$ are of the order of 20, while $\ln(x_1)$ and $\ln(x_2)$ are typically of the order of 1. Therefore the linewidth is determined predominantly by the defect capture cross section and the baseline restoration time t_0 (both of which can vary by many orders of magnitude) and depends rather weakly on the the specific choice of weighting function (which defines τ^*).

As mentioned earlier, the details of the high-temperature side of the response lineshape is determined from the properties of the Laplace transform

of the weighting functions. Neglecting baseline restoration, the response for a weighting function w(s,t) is

$$R(s,\tau) = s\int \exp(-t/\tau)\ w(s,t)\ dt\ , \qquad (2.13)$$

where s is the weighting function rate, and τ is the emission rate of the deep level as a function of temperature. This response is just the Laplace transform of the weighting function. The transforms of three common weighting functions are given in Table (I). The high temperature lineshape is determined by the numerators in Table (I) because the product $s\tau$ vanishes as the temperature increases. Clearly, the response for the sine function vanishes faster at high temperatures than the responses for the cosine and decreasing exponential functions.

Table I. Transforms of exponential decay

$w(s,t) = \exp(-st)$	$R(s,\tau) = (s\tau) / (1 + s\tau)$
$w(s,t) = \sin(st)$	$R(s,\tau) = (s\tau)^2 / (1 + (s\tau)^2)$
$w(s,t) = \cos(st)$	$R(s,\tau) = (s\tau) / (1 + (s\tau)^2)$

The functional form of the high-temperature response of these and all other Laplace transforms is simply the Laplace transform of the leading term in the Taylor's expansion of the weighting function. Therefore if the leading term in the weighting function goes as t^n, then the high temperature dependence of the response goes as $(s\tau)^{n+1}$. To increase the steepness of the slope, which decreases the high temperature half-maximum half-width of the final response, the leading term in the Taylor's expansion of the weighting function should have

as high an order as possible. This is nothing more than *weighting to later times* which reflects the fact two decaying exponentials separate hyperbolically with increasing time. The price, of course, for the decrease in linewidth is the corresponding decrease in the signal-to-noise ratio.

Baseline subtraction is necessary both to remove any DC drifts in the signal and to make measurements feasible in a reasonable length of time. Baseline restoration does not affect the high-temperature response of the Laplace transform, so the remark about weighting to later times is still valid. However the baseline subtraction does cut off the low-temperature response of the Laplace transform, providing the characteristic DLTS lineshape.

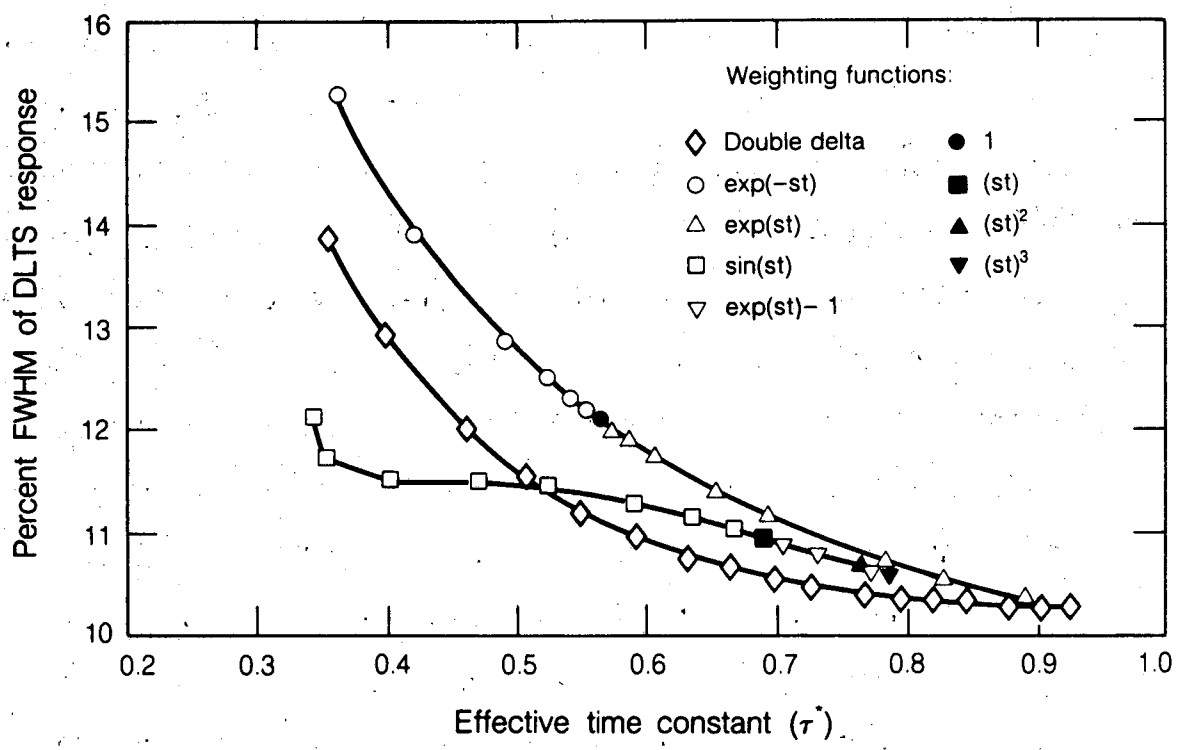
The time constant of the weighting function 1/s and the baseline restoration time t_0 both provide characteristic measurement times. They are not equal in general, though they usually are both of the same order of magnitude. As s or t_0 or the functional form of the weighting functions are changed, the emission rate e_0 for which the response function is a maximum changes. To compare the various response functions, a dimensionless time constant is defined which relates e_0 to t_0 explicitly, and to s implicitly. The dimensionless effective time constant is $\tau^* = 1/e_0 t_0$. As a rule of thumb $\tau^* \approx 2/s t_0$, i.e. $e_0 \approx s/2$, but this is not true in general, especially when $1/s > t_0$. For a given weighting function, as s decreases the function weights to later times with respect to t_0, decreasing the linewidth. This causes τ^* to increase towards unity, but it can never exceed it.

Changing either s or t_0, or both, changes the temperature at which the DLTS response peaks. This artificially affects the calculated linewidth through the temperature dependence of the density of states. Therefore to compare the results for various weighting functions with different values of s and t_0, it is necessary to "normalize" the responses such that the peak always occurs at the

same temperature. This is done, in principle, by choosing a set function and a set ratio $1/st_0$, and then changing s and t_0 simultaneously until the peak occurs at the same temperature T_0. Since the ratio $1/st_0$ determines the effective time constant τ^* for the function, the results to be compared can all be plotted versus τ^*.

The full-width half-maximum linewidths for nine functions covering a range of $1/st_0$ are plotted in Fig.(2.7) vs. τ^*. The magnitude of the linewidths depend on $\sigma N v \tau$, which is taken equal to 10^9 in this example. Several features can be noted. First, the linewidth for a set function decreases as τ^* increases; this is nothing more than the effect of decreasing the linewidth by weighting to later times. Also, the linewidth from several different functions which have the same τ^* decreases as the order of the leading term in the Taylor's expansion increases. A deviation from this last observation occurs for the sine function with low τ^*. This is due to the oscillatory nature of the sine function leading to an additional effective baseline subtraction when $st_0 > 2$. Furthermore, the weighting functions which are monomials have only one τ^*, regardless of the choice of s, and they occur at the limits of the relevant weighting functions as s goes to zero such that only the first term in the Taylor's series of the weighting function remains.

The most important result from Fig.(2.7) is that there is an absolute lower bound to the linewidth. This limit is approached as τ^* approaches unity. Of course the linewidth limit, which is always finite, is approached with vanishing signal to noise. The Taylor's expansion of $\partial(t-t_0)$ around $t = 0$ has the highest order leading term of any function defined on the interval $(0,t_0)$, producing the minimum linewidth. Therefore the delta function produces the minimum linewidth of all weighting functions.



XBL 865-10121

Fig. 2.7 Full width at half maximum of the DLTS peak for several weighting functions as a function of the effective time constant.

The resolution limit as a function of $\sigma N v t_0$ is plotted in Fig.(2.8). The linewidth decreases for increasing σ and t_0. The capture cross section σ can vary in magnitude from 10^{-12} cm^2 for defects with coulombic excited states, to 10^{-22} cm^2 for deep repulsive centers. Therefore semi-shallow, effective-mass-like states typically have quite narrow linewidths. Although σ is not a variable property of the deep-level defect, t_0 is set by the experimenter, and can be increased from microseconds to seconds. The values in Fig.(2.8) are only the low bound on the possible linewidth. The actual linewidth will be larger than this value, depending upon the choice of weighting function.

As stated previously the minimum linewidth is approached only at a cost of signal-to-noise ratio. For a single shot, the signal is

$$S = \int_0^{t_0} w(t)\, S(t)\, dt \quad , \qquad (2.14)$$

where $S(t) = s(t) - s(t_0)$, $s(t)$ being the raw signal. The noise is

$$N = \left\{ \int_0^{t_0} | w(t) |^2 \, dt \right\}^{1/2} \quad . \qquad (2.15)$$

For repetitive signal averaging the signal-to-noise ratio becomes

$$S/N|_{rep} = \sqrt{\text{rep rate}} * S/N|_{\text{one shot}} \quad , \qquad (2.16)$$

where $\sqrt{\text{rep rate}} = 1/\sqrt{t_0}$.

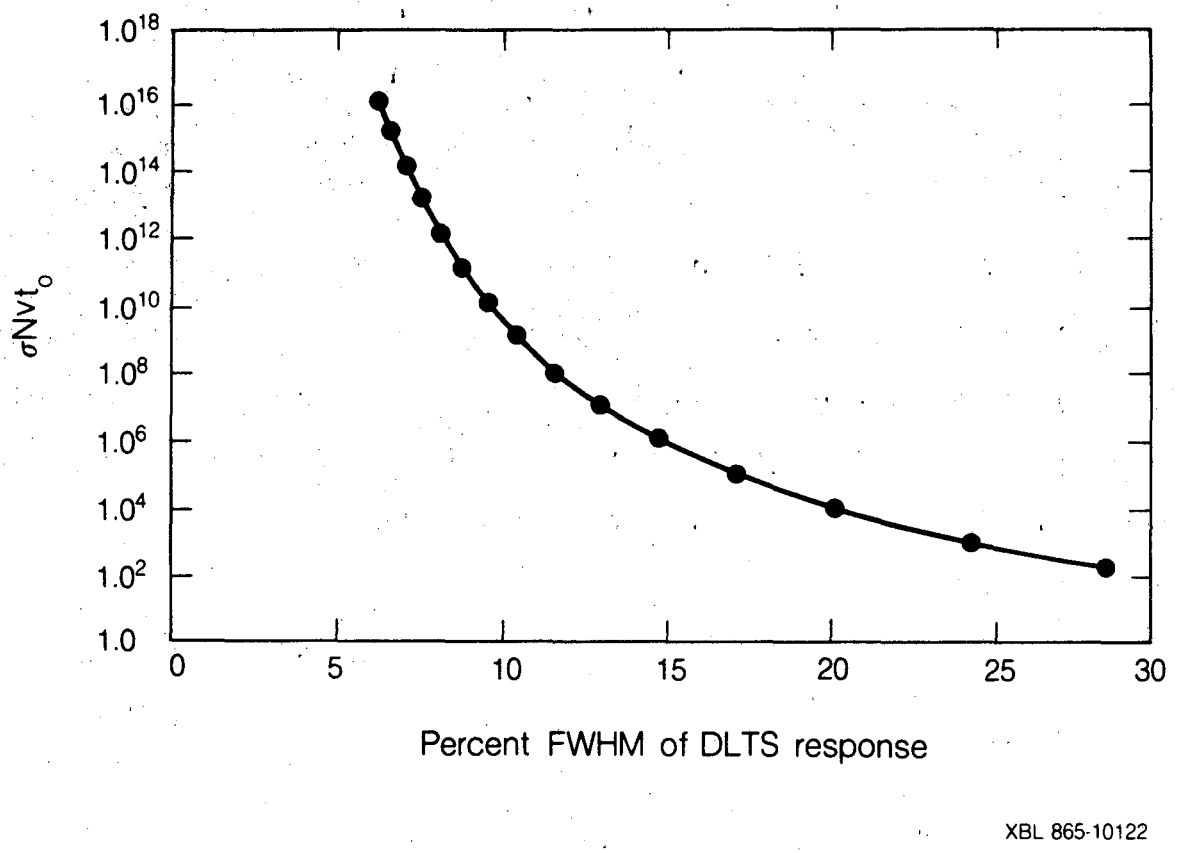
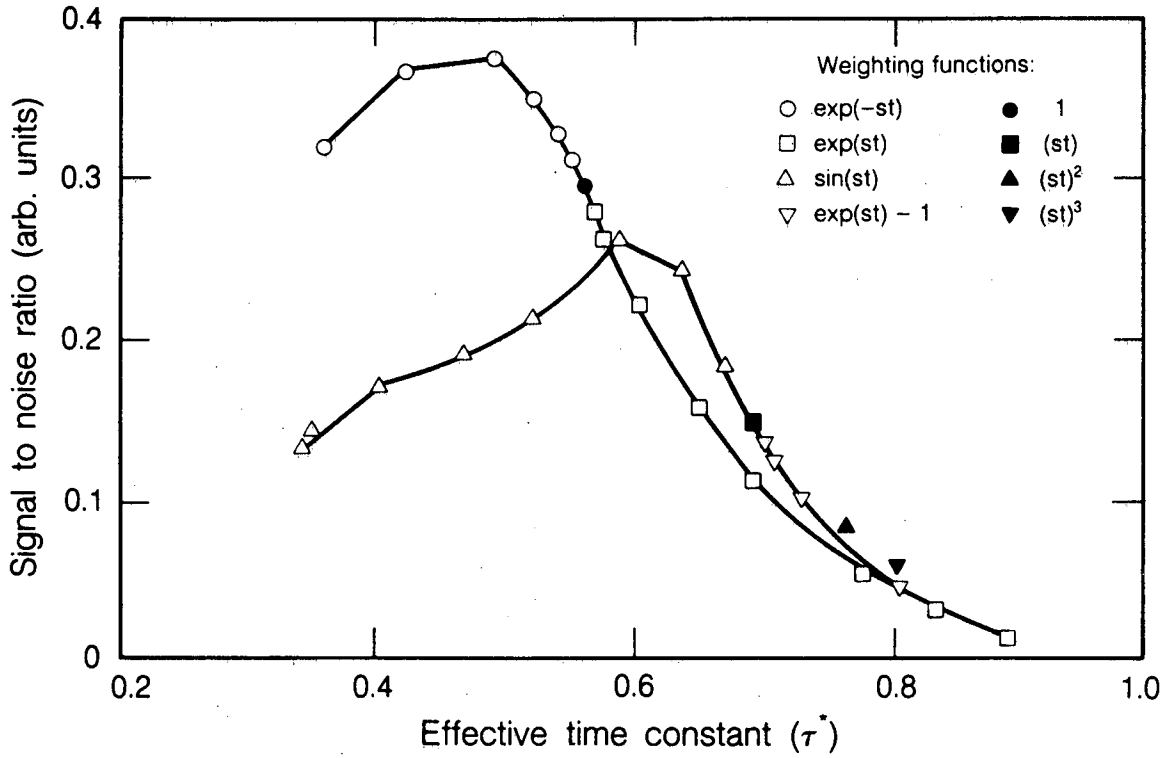


Fig. 2.8 The limit of DLTS temperature resolution as a function of σNvt_0.



XBL 865-10123

Fig. 2.9 The signal-to-noise ratio for several weighting functions.

The signal-to-noise ratios for eight of the weighting functions in Fig.(2.7) are plotted in Fig.(2.9). The signal-to-noise ratio decreases roughly exponentially with increasing τ^*. This reflects the fact that weighting to later times decreases the signal-to-noise ratio by weighting the tails of exponentially decreasing signals. At small effective time constants the signal-to-noise ratio also falls off because the time constant of the weighting function is considerably smaller than the measurement time t_0 creating considerable "dead" time before the next repetition. As pointed out by Miller et al. [Miller,1975], the largest signal-to-noise is obtained for the decaying exponential weighting function. The largest signal-to-noise ratio occurs when the time constant of the exponential is roughly half of t_0.

In this section the efficiency of the rate-window technique has been investigated. An absolute lower bound on the temperature linewidth is found which was independent of the defect activation energy. This lower bound does depend on the capture cross section and on the baseline restoration time, both of which can vary by many orders of magnitude. The smallest achievable energy linewidth is about 7%. The most usual situations give linewidths that vary from 9% to 15%. The smallest linewidths are given by using a weighting function that has a high-order first term in its Taylor's expansion. For this reason the best practical resolution is obtainable with a double-boxcar integrator using narrow windows and a long delay after the pulse.

III. Carrier Emission Rate Shifts under Uniaxial Stress

Thermal-emission rates from deep-level defects can be shifted by stress. The shifts arise from several sources: the removal of electronic or orientational degeneracy of the defect; the removal of electronic or orientational degeneracy of the band edge; and the change in the capture cross section of the defect.

Any combination of these effects may also occur. Only in the case of orientational degeneracy of the defect is there an emission-rate splitting leading to multi-exponential decays.

In the following section the case of emission from a split ground-state level to a single band is considered.

A. Defect-Ground-State Splitting

The ground state of a defect can have degeneracy that is lifted by a reduction in the crystal symmetry from uniaxial stress. The breaking of the ground-state symmetry will have associated with it a linear splitting as a function of stress. The generation and capture processes between a split ground state and a single band are shown in Fig.(2.10).

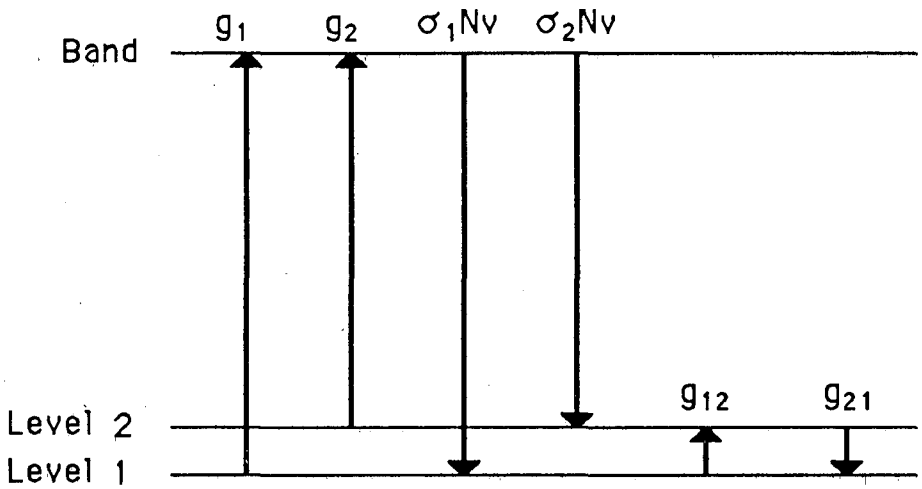


Fig. 2.10. Emission and capture processes between a split ground state and a single band.

The total number of occupied defects as a function of time, N(t), is measured experimentally as the change in the space-charge of a junction. At a given time, $N_1(t)$ defects will have a carrier occupying the lower ground state, while $N_2(t)$ defects will have a carrier occupying the higher ground state. These are related by the equation:

$$N(t) = N_1 + N_2. \qquad (2.17)$$

If the probability for the emission of a single carrier per unit time from the lower state is e_1, and the equivalent probability from the upper state is e_2, then the rate of change of N(t) is

$$dN(t)/dt = -e_1N_1(t) - e_2N_2(t)\ . \qquad (2.18)$$

The expressions for e_1 and e_2 can be obtained through the principle of detailed balance applied to each of the ground-state levels. Two equations result, corresponding to the rates into and out of each level. These are

$$\begin{aligned} e_1b_1 + g_{12} &= nv\sigma_1s_1 + g_{21} \\ e_2b_2 + g_{21} &= nv\sigma_2s_2 + g_{12}\,, \end{aligned} \qquad (2.19)$$

where e_1 and e_2 are the probabilities of emission per unit time to the band edge from levels 1 and 2 respectively, σ_1 and σ_2 are the capture cross sections, g_{21} is the rate from level 2 to level 1, and g_{12} is the complementary rate from level 1 to level 2. The fractions of populated levels are b_1 and b_2, while s_1 and s_2 are the

fractions of the unpopulated levels. The b's and s's are related by the expression

$$b_k/s_k = \gamma e^{-(E_k-E_f)/k_BT} , \qquad (2.20)$$

where γ is the degeneracy factor of the originally unsplit level, E_f is the Fermi level, and E_k is the free energy of the k^{th} level. Solving for the probabilities of emission to the band edge yields

$$\begin{aligned} e_1 &= \gamma^{-1} n v \sigma_1 e^{-(E_f-E_1)/k_BT} + (g_{21} - g_{12})/b_1 \\ e_2 &= \gamma^{-1} n v \sigma_2 e^{-(E_f-E_2)/k_BT} + (g_{12} - g_{21})/b_2 . \end{aligned} \qquad (2.21)$$

The argument of detailed balance is carried out under the conditions of equilibrium, therefore $g_{21}/g_{12} = (N+1) / N$, where N is the expectation value of the number of phonons of energy $(E_1 - E_2)$. In all cases $k_BT >> (E_1 - E_2)$, so $N >> 1$ and $g_{12} = g_{21}$. This causes the last terms in the above equations to vanish. Using the definition of the density of states of the bandedge leads to the equations for the emission probabilities

$$\begin{aligned} e_1 &= \gamma^{-1} N v \sigma_1 e^{-H_1/k_BT} \\ e_2 &= \gamma^{-1} N v \sigma_2 e^{-H_2/k_BT}, \end{aligned} \qquad (2.22)$$

where H_1 and H_2 are the enthalpies, and the entropy terms have been absorbed into the capture cross section. These probabilities are just what is obtained if the two levels were considered to be independent.

Eq. (2.18) can now be expanded to provide a differential equation for N(t), which is the physically measured quantity. $N_1(t)$ and $N_2(t)$ can be expressed as a ratio

$$N_2(t) / N_1(t) = f(t) , \tag{2.23}$$

where f(t) is an arbitrary function of time. The use of this relation and eq.(2.18) yields

$$\frac{dN(t)}{dt} = -e_1\left(\frac{1}{f(t)+1}\right)N(t) - e_2\left(\frac{f(t)}{f(t)+1}\right)N(t) \tag{2.24}$$

This differential equation cannot be solved for a general function f(t), but it is easy to solve for several specific functions f(t) which correspond to common physical situations.

Case # 1: If f(t) = 1 (the levels are actually degenerate) then

$$N(t) = N(0)e^{-e_0 t} , \tag{2.25}$$

where e_0 is the constant emission rate for no splitting.

Case # 2: If f(t) = const (the coupling rate g_{12} between N_1 and N_2 is faster than the emission rate to the band) then

$$N(t) = N(0)e^{-e^* t} , \tag{2.26}$$

where e^* is a constant effective emission rate.

Case # 3: If $f(t) = e^{-(e_1-e_2)t}$ (the two levels are entirely uncoupled, $g_{12}= 0$) then

$$N(t) = N_1(t)e^{-e_1 t} + N_2(t)e^{-e_2 t} , \qquad (2.27)$$

which is a true bi-exponential and would yield a double DLTS peak.

Case # 4: If f(t) = general function of time (there is some coupling between the states, but not enough to maintain equilibrium during the emission process) then N(t) is in general not an exponential function and the DLTS peak would be warped. These cases are discussed in more detail below.

1. Electronic Splitting

The situation of case # 2 occurs for an electronically split deep level. This is because the ratio between N_1 and N_2, for relatively small splitting, is constant (equal to the Boltzman factor). Assuming a Boltzman distribution for the ratio of the split levels $N_1/N_2 = g \exp(-\Delta E/kT)$, where g is the ratio of level degeneracies for the split levels, the emission rate from a split state is

$$e^* = \frac{e_0 \exp(-\Delta E_1 / k_B T)}{1 + g \exp(-\Delta E / k_B T)} (1 + g) \quad , \qquad (2.28)$$

where $\Delta E = \Delta E_1 + \Delta E_2$ is the sum of the separate shifts from the center of gravity. DLTS produces a peak response at the temperature at which e^* equals a preset rate. For a constant preset rate, the peak temperature shifts as stress is applied. At small stresses and energy splittings, the shift in temperature is non-linear in the applied stress. This non-linearity is caused by the exponential dependence of the Boltzman factor on stress. The details of the temperature

shift at low stress might give some indication of the type of electronic degeneracy. The interpretation of low stress data can be complicated, however, by the effects of stress on the band structure of the material.

When the Boltzman factor becomes negligible at sufficiently large splittings, the emission only occurs from the lowest energy level of the split state. The change in temperature of the DLTS peak is then linear in stress. Therefore high-stress data are required for easy interpretation of the energy shift of the lower energy component of a split level. It is interesting to note that in the high-stress limit, when emission only occurs from the lowest energy level, the effective emission rate becomes $e^* = e_1(1+g)$. The factor (1+g) correctly compensates for the loss of degeneracy g of the originally degenerate ground state.

2. Orientational Splitting

If defects are orientationally degenerate, distinct defects will have different activation energies; there will be no communication between the different populations N_1 and N_2. This corresponds to case #3 for which a biexponential decay results. Therefore, DLTS will exhibit multiple peaks for orientational splitting, but only a single peak, shifted in temperature for electronic splitting. The possible orientational degeneracies of point defects in tetrahedral semiconductors have been catalogued by Kaplyanskii[1964] along with the relative splittings and magnitudes. The final case, case # 4 above, would occur if there were weak thermal communication between the two levels. This case could occur for an orientationally degenerate defect that could reorient during a time comparable to the mean time for thermal emission.

B. Other Stress Effects

In addition to the splitting of defect ground states, the band states to which the trapped carriers are emitted can also be split by uniaxial stress. This occurs for the conduction band of semiconductors with indirect bandgaps, and for the valence band in all the tetrahedral semiconductors. The splitting of the band edge is typically as large as or larger than the splitting of the defect state, so this effect must be considered when analyzing thermal emission under stress. In Chapter III, this effect is investigated for general split band edges and then the results are applied to the specific case of the valence band in silicon. This analysis includes changes in the effective masses.

Finally, stress can affect properties of a defect other than its energy position in the bandgap. The prefactor of the thermal-emission equation contains the term

$$\sigma = \sigma_\infty \exp[-E_B / k_B T] \exp[\Delta S/k] . \qquad (2.29)$$

Stress can affect the capture cross section σ_∞, the barrier to capture E_B, and the change in entropy ΔS. Changes in the capture cross section have been observed for defects in GaAs [Barnes,1982]. Stress naturally alters the barrier to capture because this barrier is a direct result of defect-lattice coupling. This effect has been used to determine lattice relaxation properties of the DX center in AlGaAs [Li,1987]. Finally, changes in ΔS have been measured in pressure experiments [Samara,1987a,1987b]. In chapter VI, a relationship between $d\Delta S/dp$ and changes in local force constants that occur upon carrier emission from a defect is derived.

Chapter III. Treatment of the Valence Band under Uniaxial Stress

I. Introduction

Semiconductors are placed in states of uniaxial stress for many different applications. Far-infrared detectors based on shallow-level absorption are often placed under uniaxial stress to extend the sensitivity of the detectors to longer wavelengths [Haller,1979]. Carrier statistics, and in particular the density of states at the band edge, play an important role in these devices by defining defect occupancies and carrier detrapping lifetimes. Uniaxial stress is also commonly used as a probe of defect symmetry and structure. In particular, there has been recent work in the use of uniaxial stress in conjunction with thermal emission of carriers from deep-level defects [Meese,1983; Benton,1984] Defect energies are derived by measuring the rates of thermal emission of carriers from defects to the band edge. The emission probability is defined through detailed balance, which includes the density of states in the band and the thermal velocites of the carriers. For these applications and techniques, it is crucial to understand how the uniaxial stress affects the thermal emission of carriers from defects.

There have been decades of research covering the stress properties of the band edges (and shallow levels) in semiconductors. Much of the theoretical [Hasegawa,1963; Aspnes, 1978] and experimental work [Dresselhaus,1955], has been focused on defining the effective masses and the deformation potentials of the band extrema. The typical analysis techniques include

excitonic recombination luminescence [Merle,1978; Laude,1971] and cyclotron resonance [Hensel,1963], all studied successfully under uniaxial stress. The structure of the valence band has presented a challenge because of its non-parabolic energy bands. The valence band in silicon is also complicated by the small spin-orbit splitting. The deformation potentials of the valence-band edge have been measured accurately using luminescence, and are found to vary non-linearly with stress because of interaction with the split-off band. The curvature of the valence band has been studied through cyclotron resonance under uniaxial stress. Good agreement has been found between theory and experiment for large stresses [Hasegawa,1963; Balslev,1965].

These analysis techniques investigate the stress properties of the band extrema alone; they do not give a good measure of the effects on the energy band structure at wave vectors away from the extremum point in k-space. Yet the thermal properties of carriers in the bands at finite temperatures depend strongly on the band structure away from the extremum. In this chapter, the thermal-emission probability of holes from electrically active defects in the bandgap to the valence-band edge is emphasized. It therefore must be considered how stress affects valence-band states with energies several k_BT from the band edge. The results derived for the band-edge effective density of states and the thermal velocities are general and can be used for any problem involving carrier statistics under stress. This chapter begins by describing the zero-stress structure of the valence band and the difficulty of defining an effective mass. In section III the derivation of thermal emission from defects to multiple, independent bands is presented and generalized to include the continuous energy dispersion of a band with a complicated structure such as the valence band under uniaxial stress.

II. Structure of the Valence Band

The top of the valence band in silicon originates from the three valence p-orbitals of silicon. At the center of the Brillouin zone the band is three-fold degenerate (in the absence of spin-orbit effect) and transforms as the $\Gamma_5(O_h)$ representation of the point group of the crystal. For finite k-vector this degeneracy is broken, resulting in energy dispersion with a small and a large curvature. Under the influence of the spin-orbit interaction, the Γ_5 symmetry at the Γ point is broken into a Γ_8 quartet corresponding to j = 3/2 and a Γ_7 doublet corresponding to j = 1/2, which is split off from the Γ_8 energy by an amount Δ_{so}, called the spin-orbit splitting. Silicon is special among the common semiconductors in that the spin-orbit splitting is relatively small compared to the bandgap; Δ_{so} = 0.044 eV compared to E_{gap} = 1.15 eV. This small splitting has important consequences for the structure of shallow acceptor wavefunctions, and plays a significant role in high stress experiments, as will be discussed later in this thesis.

The energy dispersion, including the spin-orbit interaction, is shown for finite k-vector in Fig.(3.1). The structure of the valence band is composed of a heavy-hole band and a light-hole band (which are degenerate at the center of the Brillouin zone) and a split-off band. The energies of the heavy and light holes are given by

$$E(k) = E_0 + (\hbar^2k^2/2m_0)\,(\,A \pm (\,B^2 + sC^2)^{1/2}\,)$$
$$s = (\,k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2)/k^4\,, \qquad (3.1)$$

where the plus sign is for light holes, and the minus sign is for heavy holes. The parameters A, B and C determine the curvature of the bands, and hence the effective masses. The curvatures of the bands at the origin are not unique, but

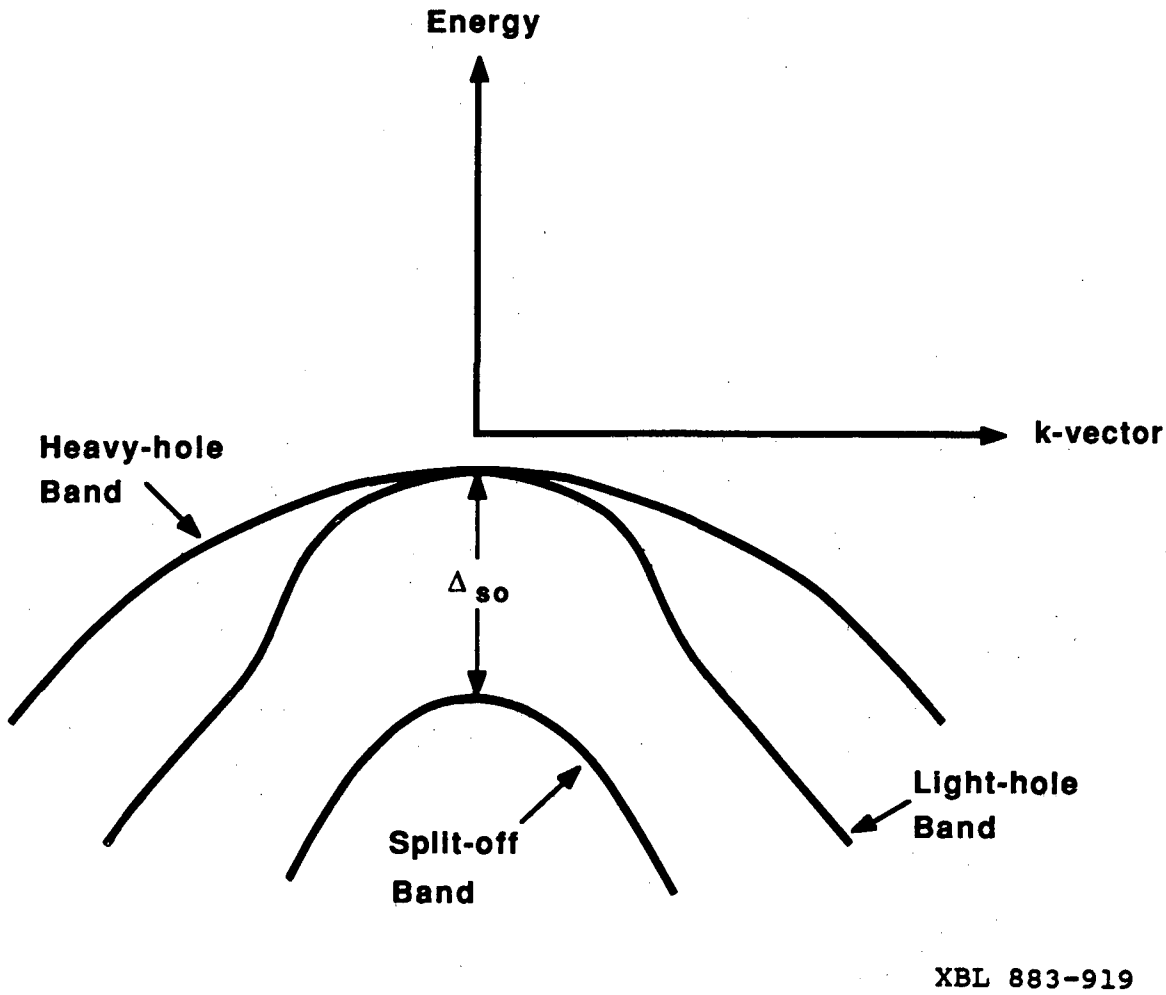


Fig. 3.1 Structure of the valence band at zero-stress. The spin-orbit interaction splits the original Γ_5 symmetry at k = 0 into Γ_8 and Γ_7 symmetries, separated by Δ_{so}.

depend on the direction of the k-vector. Effective masses can still be defined by expanding the fluted energy surfaces in spherical harmonics [Lipari,1970]. In this spherical approximation, the average effective masses are given by

$$1/m^*_{av} = (1/m_0)\,(\,A \pm (\,B^2 + C^2/6)^{1/2}\,)\,. \qquad (3.2)$$

The parameters A, B and C have been found from cyclotron resonance experiments [Hensel,1963; Balslev,1965] to be A = -4.27 , B = -0.63, and |C| = 4.93 at 1.26 K. The corresponding zero-stress heavy- and light-hole masses are $m^*_h = 0.47$ and $m^*_l = 0.16$.

III. Thermal Emission of Carriers to Multiple Bands

The total probability for the thermal emission of a carrier from a single defect to multiple bands is the sum of various independent emission probabilities. These bands may be degenerate, or not. For heuristic purposes the case for two distinct bands is considered first, denoted as band 1 and band 2. Since no distinction can be made whether the carrier is emitted to band 1 or to band 2, the emission processes to these bands are independent and the total emission probability per unit time becomes

$$e_0 = g_1 + g_2\,, \qquad (3.3)$$

where the g_i are the generation rates of carriers from the defect level to band i. Since all the defects with the same ground state are equivalent, they will all have this same emission probability, which will lead to a single exponential decay of the DLTS signal

$$N(t) = N(0)e^{-e_0 t} \quad . \tag{3.4}$$

The expression for e_0, can be obtained through detailed balance. The detailed balance equation is

$$\sigma_1 v_1 n_1 s + \sigma_2 v_2 n_2 s = g_1 b + g_2 b \, , \tag{3.5}$$

where the subscripts 1 and 2 refer to the two different bands, v_j is the thermal velocity of the carriers and n_j is the number of carriers in the respective band, s is the fraction of unoccupied defects, b is the fraction of occupied defects, and σ is the capture cross section which may be different for capture from different bands. The transition processes are shown in Fig.(3.2).

The ratio of occupied traps to unoccupied traps is

$$b/s = \gamma \exp[-(E_T - E_f) / k_B T] \, . \tag{3.6}$$

where E_T is the defect binding energy, E_f is the Fermi level and γ is the defect degeneracy factor. For a single band the number of carriers occupying that band can be approximated as

$$n = n_i \exp((E_f - E_i) / kT) = N \exp((E_{band} - E_f) / k_B T) \, . \tag{3.7}$$

where n_i is the intrinsic carrier concentration in the band, and E_i is the intrinsic Fermi level. The band effective density of states is given by

$$N = 2 \, (m_{ds}^* \, k_B \, T / 2 \, \pi \, \hbar^2)^{3/2} \tag{3.8}$$

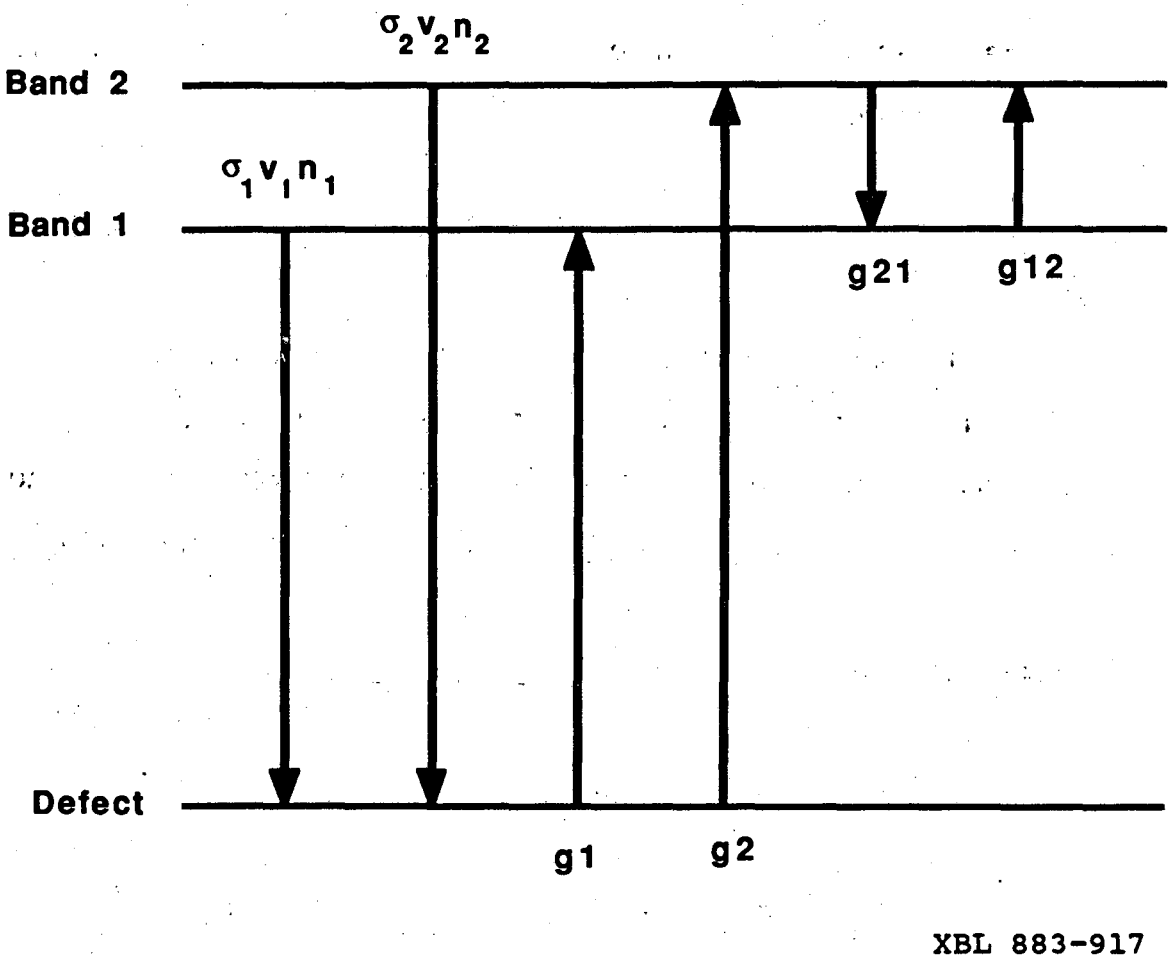


Fig. 3.2 Transition processes between a defect and multiple bands. The capture and thermal emission probabilities from and to two different bands are presented.

where m_{ds}^* is the density-of-states effective mass. Equation (3.7) can be retained for each band separately in the multiple band situation

$$n_{1,2} = n_{i1,2} \exp[(E_f - E_i) / k_B T] = N_{1,2} \exp[(E_{1,2} - E_f) / k_B T] , \qquad (3.9)$$

where $N_{1,2}$ is the density of states defined in the same manner as for a single or degenerate band, and $E_{1,2}$ is the respective band-edge energy.

Solving directly for $g_1 + g_2$ gives the total emission probability

$$e(\tau) = c\, T^2 \{ \sigma_1(\tau)\, m^*_1(\tau) \exp[(E_T(\tau) - E_1(\tau)) / k_B T] + \sigma_2(\tau)\, m^*_2(\tau) \exp[(E_T(\tau) - E_2(\tau)) / k_B T] \} . \qquad (3.10)$$

This equation is the independent-band model for emission of carriers to two bands. The stress dependence τ of the effective masses, energies and capture cross sections have been explicitly included.

The quantities $m^*_1(\tau)$ and $m^*_2(\tau)$ are averaged effective masses arising from the mass dependences of the thermal velocities and density of states. This average effective mass is

$$m^* = (m^*_{ds})^{3/2} (m^*_{th})^{-1/2} , \qquad (3.11)$$

where m^*_{ds} is the density-of-states mass defined by

$$m^*_{ds} = (m_1\, m_2\, m_3)^{1/3} , \qquad (3.12)$$

where the m_i are the principal values of the effective-mass tensor. The thermal effective mass m^*_{th} is defined by the average thermal velocities $\langle v \rangle_{rms}$ as

$$m^*_{th} = 3\,k_B T / \langle v \rangle_{rms}^2$$
$$= 1/3\,[\,1/m_1 + 1/m_2 + 1/m_3\,]\,. \qquad (3.13)$$

The emission probability in the independent-band model of eq.(3.10), though a bi-exponential in temperature, yields a single exponential relaxation in time, and hence a single DLTS peak. It is interesting to note, however, that the system discussed above will have a non-linear Arrhenius plot. Although all the transients may be single exponentials, the emission rate is determined by several emission probabilities, and each probability will have a different activation energy. This presents some difficulty when attempting to determine an activation energy by plotting the emission rate as a function of 1/T. In order to find the defect activation energy E_T, all parameters describing the various bands must be known accurately and used to fit the data to the above equation.

The conduction band minima in the indirect-gap semiconductors represent truly independent bands with well-defined density of states and carrier effective masses. In the zero-stress limit these minima are degenerate and the carrier emission occurs to each equivalent minimum with equal probability. Under the application of uniaxial stress, the degeneracy is lifted and the minima become inequivalent. The emission process in this case is determined simply by extending eq.(3.10) to include the correct number of minima and their respective energy shifts. The effective masses are not altered (to first-order) by the application of stress.

The thermal emission of holes to the valence band, on the other hand, cannot be defined rigorously in terms of emission to multiple, independent bands. This is especially true for the zero- and low-stress case. The difficulty originates from the spin-orbit interaction and the complicated structure it

imposes on the top of the valence band (described in section II). This structure cannot be described exactly by two independent bands at zero or low stress because the two bands mix. Furthermore, it is impossible to uniquely define an effective-mass tensor for the valence band. These features make it difficult to apply eq.(3.10) to the thermal emission of carriers to the valence band. In this situation, the detailed balance eq.(3.5) can be made more precise by replacing it with:

$$\sigma_0 \, s \int v(E,\tau) \, n(E,\tau) \, dE = b \int g(E) \, dE \, , \qquad (3.14)$$

where the number of carriers at energy E is given by

$$n(E,\tau) = N(E,\tau) \exp(-(E - E_f) / k_B T) \qquad (3.15)$$

and N(E) is the density of states at energy E. For complete generality, the stress dependence of the density of states and thermal velocity is included explicitly. The average energy- and stress-dependent thermal velocity $v(E,\tau)$ is

$$v(E,\tau) = (1 / \hbar) \, (\mathbf{n} \cdot \nabla_k E)_{E,\tau} \, ,$$

where

$$(\mathbf{n} \cdot \nabla_k E)_{E,\tau} = \frac{\int dS_E}{\int (\mathbf{n} \cdot \nabla_k E)^{-1} dS_E}$$

$$= \frac{\int dS_E}{4 \pi^3 \hbar \, N(E,\tau)} \qquad (3.16)$$

and **n** is the unit vector pointing along the direction of the k-vector. The term dS_E is a mono-energetic surface element in k-space. The thermal emission rate is therefore

$$e_0 = \sigma_0 (\hbar \gamma)^{-1} \int (\mathbf{n}\cdot\nabla_k E)_{E,\tau} \; N(E,\tau)\, e^{-(E_T-E)/k_B T}\, dE$$

$$= \sigma_0 (4 \hbar \gamma \pi^3)^{-1} \int \left[\int dS_E\right] e^{-(E_T-E)/k_B T}\, dE\,. \qquad (3.17)$$

Though this expression has the advantage of being rigorous, it must be evaluated numerically.

The problem can be simplified considerably by developing an independent-band model for the valence band. This model approximates the valence band as two independent bands that displace rigidly with changing stress. When this approximation is extended to the case of non-zero stress, eq.(3.17) reduces to the previous eq.(3.10) after substitution of appropriate effective masses and energy splittings of the top of the valence band under uniaxial stress. In the next section the energy splittings of the valence band are described in detail and the application of the independent-band approximation is discussed.

IV. The Valence Band under Stress

A. Energy splittings under stress

There are three deformation potentials describing the strain couplings of a Γ_8 state. These are: a, the hydrostatic; b the normal; and d the shear deformation potentials. The splittings in the limit of small stresses are given by

$$\Delta E(\tau)^2 = b^2/2\,[\,(\varepsilon_{xx} - \varepsilon_{yy})^2 + (\varepsilon_{xx} - \varepsilon_{zz})^2 + (\varepsilon_{yy} - \varepsilon_{zz})^2] + d^2[\varepsilon_{xy}^2 + \varepsilon_{xz}^2 + \varepsilon_{yz}^2]\,. \qquad (3.18)$$

The deformation potentials are [Merle,1978] b = -2.1 eV d = -5.1 eV at 77 K. The splittings described by eq.(3.18) are only valid in the limit of small stresses such that the valence-band splitting is negligible compared to the spin-orbit splitting Δ_{SO}. For stresses greater than 0.1 GPa, however, the coupling of the the top of the valence band with the split-off band is strong enough to produce significant corrections to the shifts described in eq.(3.18). The coupling of the top of the valence band to the split-off band under uniaxial deformation can be calculated directly for all stress directions and magnitudes by extending the stress Hamiltonian of Bir and Pikus [Pikus,1960] to include the split-off band. This Hamiltonian represents the valence bands according to the j=3/2 and j=1/2 angular monemta. The eigenvalues of the full 6x6 Hamiltonian matrix reduce to three values with double (Kramer's) degeneracy, because stress does not break time-reversal symmetry. For stress along either the <100> or <111> directions the z-component of the angular momentum quantum numbers m_J remain good quantum numbers, and if the stress direction is taken as the quantization axis, then there are no matrix elements of the stress Hamiltonian between states with m_J and $-m_J$. The Hamiltonian for [001] stress reduces to a 1x1 matrix for the $m_J = \pm 3/2$ and a 2x2 matrix for the $m_J = \pm 1/2$. The 2x2 matrix has the following eigenvalues:

$$E = \frac{E_{100}}{4} - \frac{\Delta_{so}}{2} \pm \frac{\Delta_{so}}{2}\sqrt{1 + \frac{9}{4}\left(\frac{E_{100}}{\Delta_{so}}\right)^2 + \left(\frac{E_{100}}{\Delta_{so}}\right)}\,, \qquad (3.19)$$

where $E_{100} = 2b(s_{11} - s_{12})T$. For the case of stress in the [111] direction, the quantization axis can be rotated to coincide with the stress axis to obtain similar eigenvalues. For the case of arbitrary stress direction the m_J are no longer good quantum numbers and the Hamiltonian reduces to a 3x3 matrix which yields three eigenvalues. The energy shifts for stress parallel to [100] are shown in Fig.(3.4) compared to the shifts described by eq.(3.18) in the absence of coupling to the split-off band. Under [100] compression the Γ_8 degenerate valence band splits into X_6 and X_7 Kramer's doublet representations with the X_7 band moving into the bandgap. This X_7 band couples with the X_7 split-off band, producing nonlinear shifts for the top of the valence band. The interaction under 1 GPa produces nearly a 50% correction to the value predicted by eq.(3.18). The angular dependence of the energy shifts for 1 GPa is shown in Fig.(3.3) in comparison with the splittings of the Γ_5 state in the absence of spin-orbit splitting. One of the remarkable features of the valence band under stress is the fact that the energy shifts are isotropic within 10%.

The energy splittings defined by eq.(3.18) are valid only for the valence band extrema at the Γ point for zero k-vector. Points on the dispersion curve for finite k-vector will shift by different amounts, depending on the magnitude as well as direction of the k-vector. In other words, the bands are not rigidly displaced in energy, but are warped by the application of stress. Yet, the energy splittings of eq.(3.18) are convenient to use in the independent-band approximation as the energies by which two independent bands are rigidly displaced. The effect of the additional warping under stress of the originally fluted energy surfaces must therefore be included as changes in the effective masses describing the density of states and the thermal velocities. The appropriate effective masses which enter the independent-band approximation are discussed in the next section.

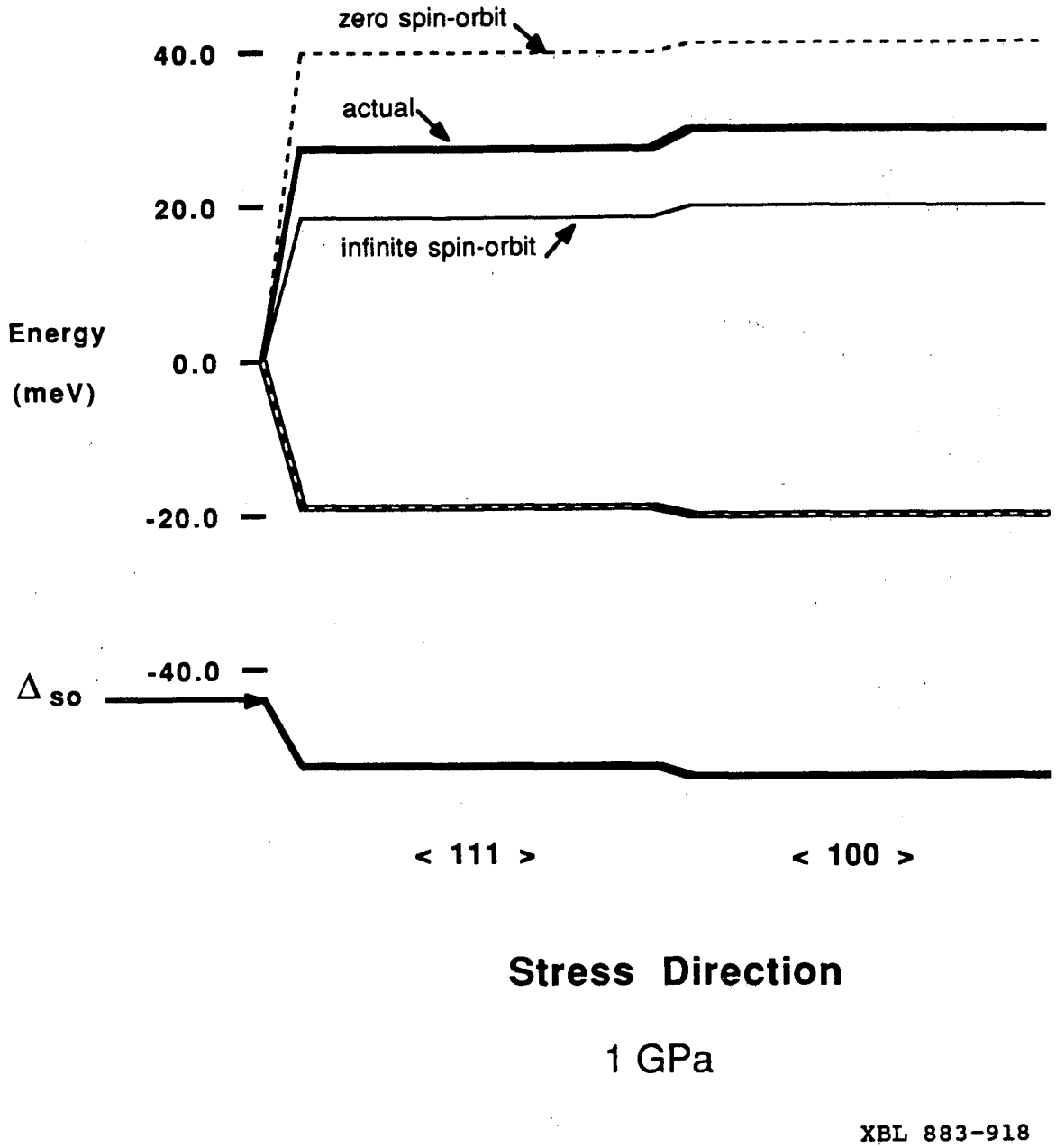


Fig. 3.3 Splitting of the valence band under 1 GPa compressive uniaxial stress. The dotted line is for the case of no spin-orbit interaction; the thin line is for the case of infinite spin orbit coupling; and the bold line is the actual valence band shift under 1 GPa.

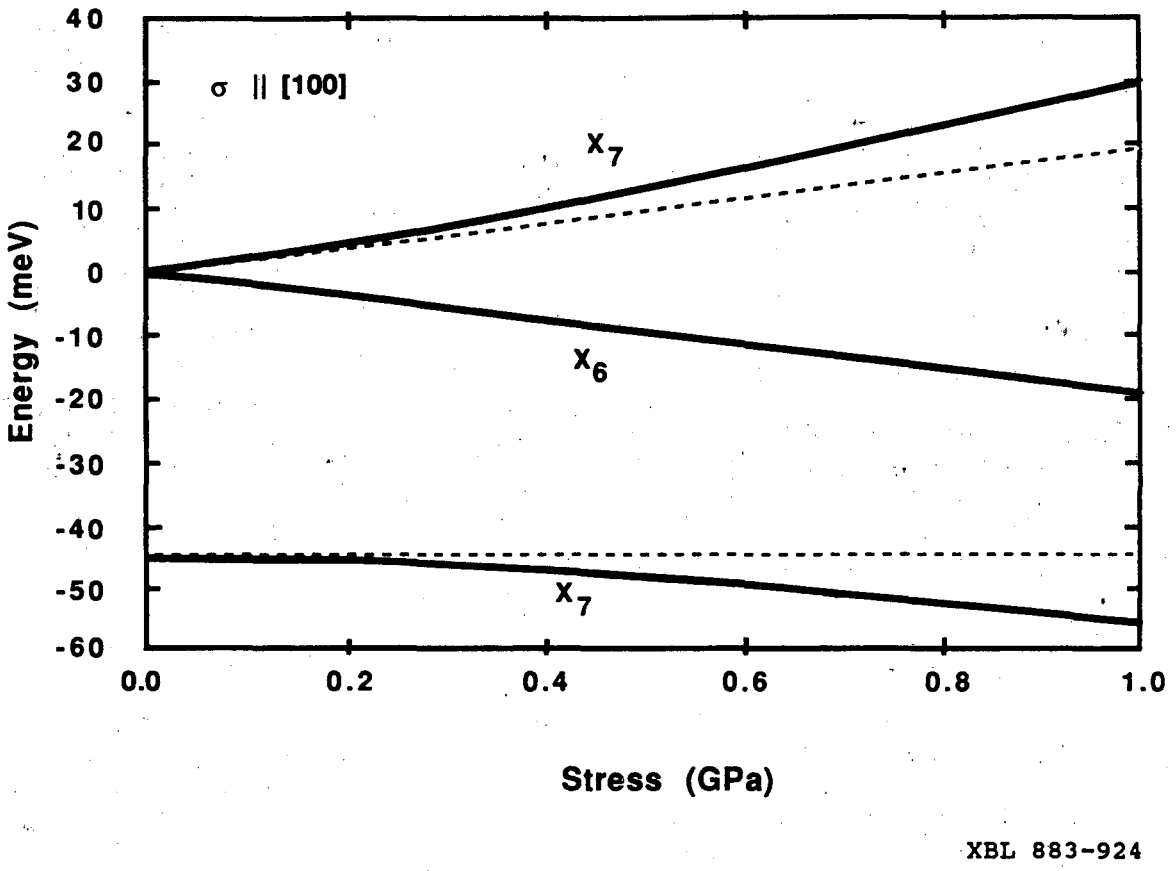


Fig. 3.4 Energy splitting and shifts of the top of the valence band and the split-off band for compressive stress along [100]. The parameters used in eq.(3.19) are b = -2.1 eV and Δ_{so} = 44 meV.

B. Effective Masses under Stress

The effective mass of a band extremum is proportional to the inverse curvature of the electron energy as a function of k-vector. Bands described by spherical energy surfaces have a unique effective mass, while bands described by elliptical or spheroidal energy surfaces can be described uniquely by effective-mass tensors. The density-of-states effective mass m^*_{ds} is defined in eq.(3.12) as the cube root of the product of the principal values of the mass tensor. The energy surfaces of the valence band at zero-stress, however, are not elliptical or spheroidal, but are described by fluted or warped spheres. This makes it impossible to uniquely define an effective-mass tensor, as stated before.

The structure of the valence band becomes much simpler for high stresses applied along the <100> or <111> directions. For large stresses in these directions (strain energy >> k_BT), the energy surfaces are described by prolate and oblate spheroids which do have well-defined effective-mass tensors. The principal values of the effective masses for these stress directions are given in Table I where $N^2 = 3\,(B^2 + C^2)$ [Hasagawa,1963]:

Table I

<100>	<111>
$m_{1/2,\|\|} = 1/(A + B\,Z(x))$	$m_{1/2,\|\|} = 1/(A + 1/3\,N\,Z(x))$
$m_{1/2,\perp} = 1/(A - 1/2\,B\,Z(x))$	$m_{1/2,\perp} = 1/(A - 1/6\,N\,Z(x))$
$m_{3/2,\|\|} = 1/(A - B) = 0.28$	$m_{3/2,\|\|} = 1/(A - 1/3\,N) = 0.75$
$m_{3/2,\perp} = 1/(A + 1/2\,B) = 0.22$	$m_{3/2,\perp} = 1/(A + 1/6\,N) = 0.18$

The function

$$Z(x) = (1/2)\,(\,1 + (\,1 - 9x)\,(\,1 - 2x + 9x^2)^{-1/2}\,) \qquad (3.20)$$

describes the effect of coupling to the split-off band on the principal values of the effective-mass tensor, where the quantity $x = \Delta E(\tau) / \Delta_{SO}$.

The function $Z(x) = 1$ in the absence of coupling to the split-off band. For 1 GPa, the value for silicon is $Z(x=0.4) = -0.6$, which dramatically alters the principal values of the effective-mass tensor, and has important consequences in ESR [Hensel,1963]. The density-of-states mass, however, remains surprisingly unaffected (<5% at 1 GPa) and therefore this extra mixing of the effective mass can be neglected in stressed DLTS. It is interesting to point out that these high-stress effective masses have the property that the average thermal inverse effective mass is exactly

$$1/m^*_{th} = (1/3)\,\Sigma_i\,(\,1/m_i\,) = A\ , \qquad (3.21)$$

independent of the strength of the coupling to the split-off band [Hasegawa,1963]. The valence band under large stress is accurately described by two rigidly displacing, independent bands with well-defined effective masses; no approximations are required.

The situation for small stress (strain energy $<< k_BT$) is substantially more complicated than the case of large stress because of the strong mixing of the originally degenerate bands. The curvature of the energy dispersion becomes a function of the magnitude of the k-vector as well as its direction. Again, as in

the zero-stress case, there is no unique method to define an effective density-of-states mass. Two methods are described next to define average stress-dependent density-of-states effective masses. The first method defines the effective masses through the curvature of the energy dispersion. The second method (which is presented in this chapter) defines the valence band as two rigidly displacing parabolic bands and finds the appropriate density-of-states effective masses for these bands. Both of these methods begin with the Bir-Pikus strain Hamiltonian describing the energy shifts under stress of the electron-energy dispersion curves.

In the absence of coupling to the split-off band, the electron Hamiltonian including stress and k-vector is [Pikus,1960]

$$E(k) = Ak^2 + a\Delta \pm (E_k^2 + E_{\varepsilon k} + E_\varepsilon^2)^{1/2} ,$$

$$E_k^2 = B^2k^4 + C^2(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2) ,$$

$$E_\varepsilon^2 = 1/2\, b^2 [(\varepsilon_{xx} - \varepsilon_{yy})^2 + (\varepsilon_{xx} - \varepsilon_{zz})^2 + (\varepsilon_{yy} - \varepsilon_{zz})^2] + d^2 [\varepsilon_{xy}^2 + \varepsilon_{xz}^2 + \varepsilon_{yz}^2] ,$$

$$E_{\varepsilon k} = Bb[3 (k_x^2 \varepsilon_{xx} + k_y^2 \varepsilon_{yy} + k_z^2 \varepsilon_{zz}) - k^2\Delta] + 2Dd [k_xk_y \varepsilon_{xy} + k_xk_z \varepsilon_{xz} + k_yk_z \varepsilon_{yz}] , \qquad (3.22)$$

where $\Delta = (s_{11} + 2s_{11})T$ is the dilatation (T is negative for compressive strain). The spherical approximation of the valence band [Lipari,1970] can be extended to include stress by deriving the appropriate irreducible-spherical-tensor operator that represents the strain field caused by applied uniaxial stress [Broeckx,1987]. Stress-dependent effective masses can also be obtained

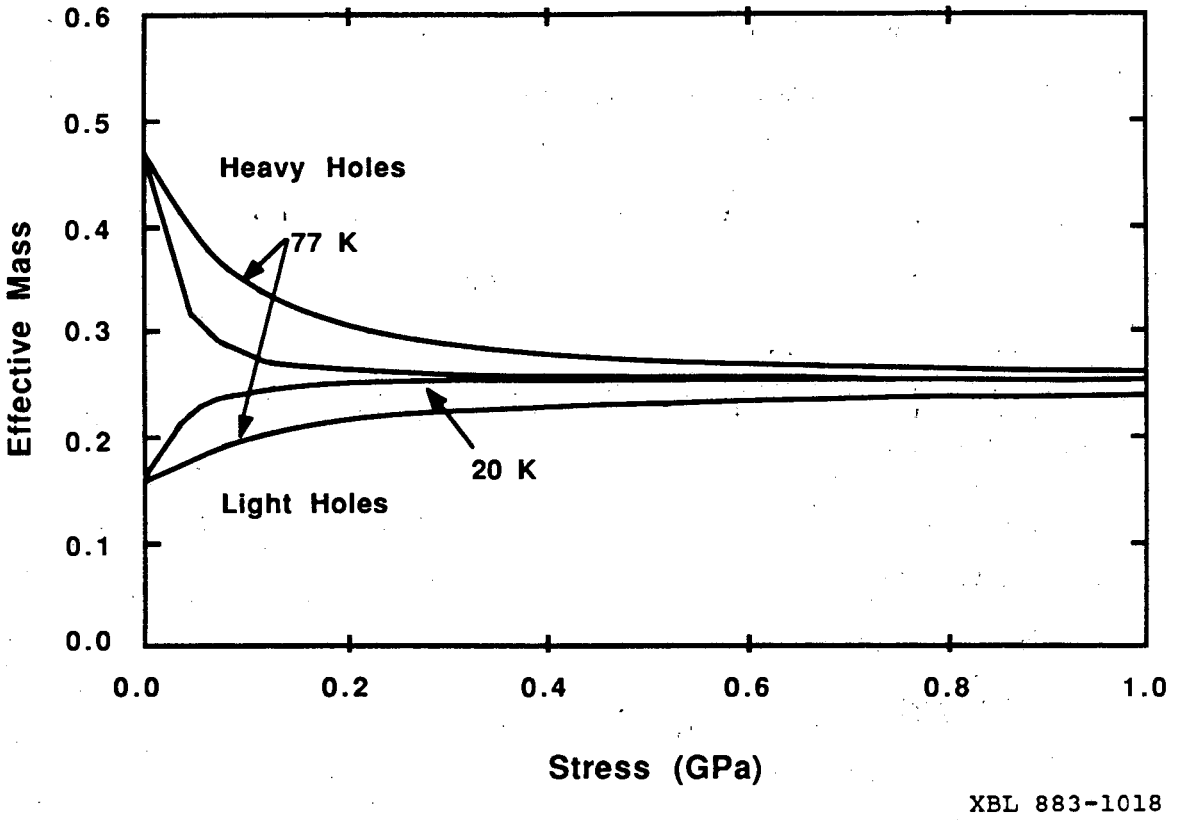


Fig. 3.5 Mixing of the effective masses calculated from the spherical approximation under compressive stress along [100].

through numerical calculations by finding the curvature in the two branches [Kelso,1982] of eq.(3.22). Qualitatively similar results to Kelso are obtained by replacing B, C and D respectively in the above equation by

$$B' = (B^2 + (1/6)C^2)^{1/2} ,$$
$$C' = 0 \quad ,$$
$$D' = \sqrt{3}\, B' \quad . \qquad (3.23)$$

The new Hamiltonian gives (by definition) the correct effective masses at zero-stress. For finite strain the effective masses are found by calculating the curvature of the two branches of eq.(3.22) for directions parallel and perpendicular to the applied stress direction and applying eq.(3.12). Thermal averages are taken by weighting the curvature with the Boltzman factor. The effective masses are therefore functions of both stress and temperature. The results for the density-of-states effective masses at 77 K and 20 K are given in Fig.(3.5). The light and heavy holes mix quickly with increasing stress, and saturate to constant values defined by eq.(3.21) for high stress. The mixing rate is faster for lower temperatures because the effective mass is dominated by the curvature at smaller k-vector and the high-stress limit is approached more quickly.

The second approach for defining the stress-dependent effective masses (the independent-band model) obtains the stress and energy-dependent density of states of the valence band. This density of states is used with the energies given by eq.(3.18) to fit the effective masses for the rigid, independent bands. The density of states of a band is defined by the electron-energy dispersion equation $E(\mathbf{k},\tau)$ through the integral

$$N(E,\tau) = \int \delta(E - E(\mathbf{k},\tau))\, d\mathbf{k} . \qquad (3.24)$$

This integration was carried out numerically for the case of compressive stress applied along the [100] and [111] axes. By symmetry, for [100] stress, only 1/16 of the Brillouin Zone is considered. For [111] stress only 1/12 of the Brillouin Zone is sampled. Furthermore, it is only necessary to consider the density of states within several k_BT of the band extremum, which limits the magnitude of the k-vector to only 1/20 of the value at the Brillouin Zone boundary. Therefore, the volume of k-space sampled in the integration was approximately 1/400 of the volume of the Brillouin Zone. Within this volume, 50,000 points were sampled. The results of the integration for [100] stress are shown as the data points in Fig.(3.6a-c) for τ = 0.0, 0.3, and 0.6 GPa respectively. Results for [111] oriented stress are quantitatively similar. Coupling to the split-off band was neglected in this analysis. The solid curves are the result of the independent-band model. This model predicts density of states that vary as the square root of energy, with a multiplicative prefactor that depends on the effective mass of the respective band raised to the 3/2 power (from eq.(3.8)). The model fits the numerical data well for zero-stress, but there are deviations from the independent-band model for finite stress. The deviations occur within the region of energy in which the two bands are strongly mixed. This interaction modifies the curvature of the energy dispersion relation through avoided crossings, which directly affect the density of states.

The dependence of the thermal-emission probability on the density of states is weighted exponentially towards the band edge. Therefore, to compare the results of Fig.(3.5) with the results of the numerical calculation of $N(E,\tau)$, it is convenient to define an effective density of states by

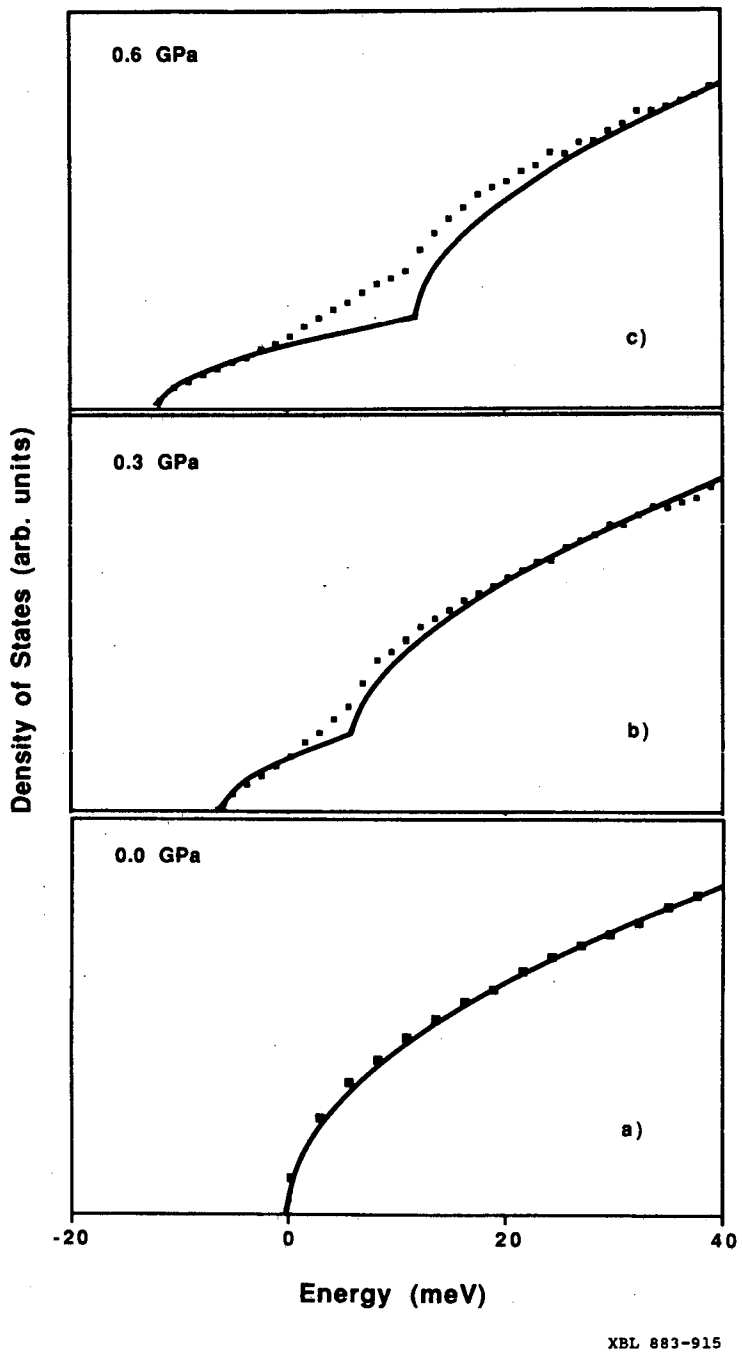


Fig. 3.6 The density of states of the valence band for a) 0 GPa, b) 0.3 GPa, and c) 0.6 GPa of compressive uniaxial stress applied along [100]. The best fit of two independent parabolic bands is included as the solid line.

$$N = \int N(E,\tau)\, \exp(-E / k_B T)\, dE \,/\, N(E,0) \,. \qquad (3.25)$$

The inverse of this quantity as a function of stress is plotted in Fig.(3.7) for the numerical calculations. In the independent-band model, eq.(3.25) is equivalent to

$$N = \frac{(\, m_l(\tau)\,)^{3/2} \exp(E_\varepsilon / k_B T) + (\, m_h(\tau)\,)^{3/2} \exp(-E_\varepsilon / k_B T)}{(\, m_l(0)\,)^{3/2} + (\, m_h(0)\,)^{3/2}} \,, \qquad (3.26)$$

where $m_l(\tau)$ and $m_h(\tau)$ are the stress-dependent effective light-hole and heavy-hole masses, and E_ε is the energy shift of the respective band edge defined in eq.(3.18). The result of eq.(3.26), using the effective masses from the spherical approximation, is compared in Fig.(3.7) to the exact numerical result from eq.(3.25). This function is proportional to the fractional change in the emission time constant of a defect that is insensitive to stress. The result of the spherical approximation clearly disagrees with the exact numerical calculation for low stresses. The spherical approximation would predict an increase of the emission time for small stresses. For low stresses the emission would still be predominantly to the heavy-hole band, which is moving away from the defect energy (although the thermal emission is weighted by a Boltzman factor), because the heavy-hole band has the larger effective density of states. This produces the initial increase in the emission time constant. With increasing stress, however, the Boltzman factor increasingly favors emission to the light-hole band, which is approaching the defect energy, and therefore the time

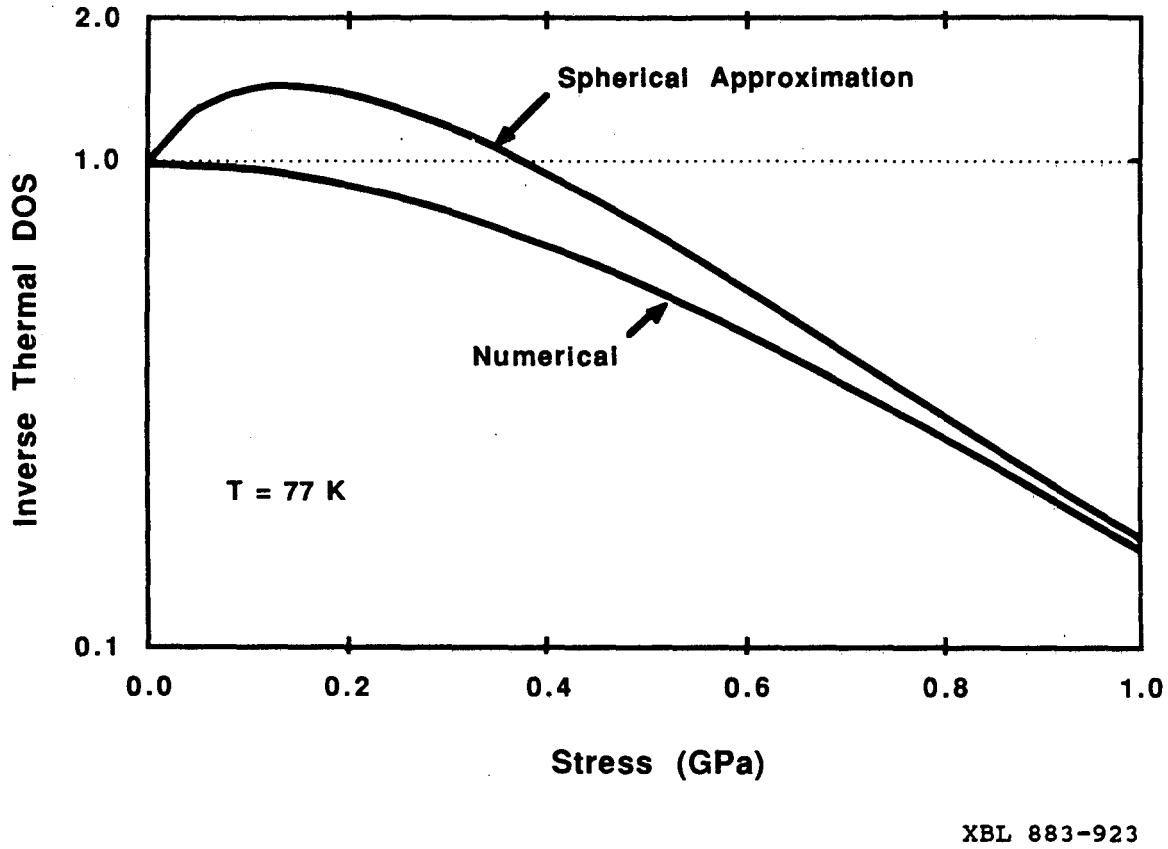


Fig. 3.7 The inverse thermal density of states defined by eq.(3.25) as a function of compressive stress at T = 77 K. The curve labeled Spherical Approximation uses the effective masses from Fig. (3.5). The curve labeled Numerical is the exact result using the accurate density of states from eq.(3.24).

constant eventually decreases. Under the higher stresses, the values of the spherical approximation approach the exact numerical values. This demonstrates that the spherical approximation does in fact give the correct density of states in the limit of zero-stress, as well as the correct effective mass for the light hole in the limit of large stress. However, the predicted partitioning of the density of states between the two bands by the spherical approximation is incorrect. The error has arisen from the (incorrect) use of the effective masses from the spherical approximation with the band-edge energy splittings (eq.(3.18)). The energies and effective masses cannot be defined separately. Once one of these quantities is defined, the other quantity is completely determined. Because the energy splittings of the band edges are well established, it is convenient to retain these energies and then find the appropriate effective masses for the density of states. The correct effective DOS masses as a function of stress can be derived by applying the exact results of Fig.(3.7) to eq.(3.26). In eq.(3.26) the "rigid" bands are forced to shift in energy by E_ε, and then the effective masses are adjusted to yield the value from eq.(3.25). The resulting DOS effective mass values are shown in Fig.(3.8) for 77 K. At zero stress, the two masses are nearly equal, in strong contradiction to the spherical approximation, yet the total density of states is the same for both models. At high stresses, the light-hole effective mass approaches the asymptote 1/A correctly described by eq.(3.21).

The near equality of the DOS light- and heavy-hole effective masses at low stress has the important consequence that the center of gravity of the splitting valence bands is approximately conserved as stress is increased. This fact is reflected by the zero-slope asymptote of the thermal density of states in Fig.(3.7) for zero stress. This would not be the case if the masses from the spherical approximation were used; in that case the center of gravity would shift

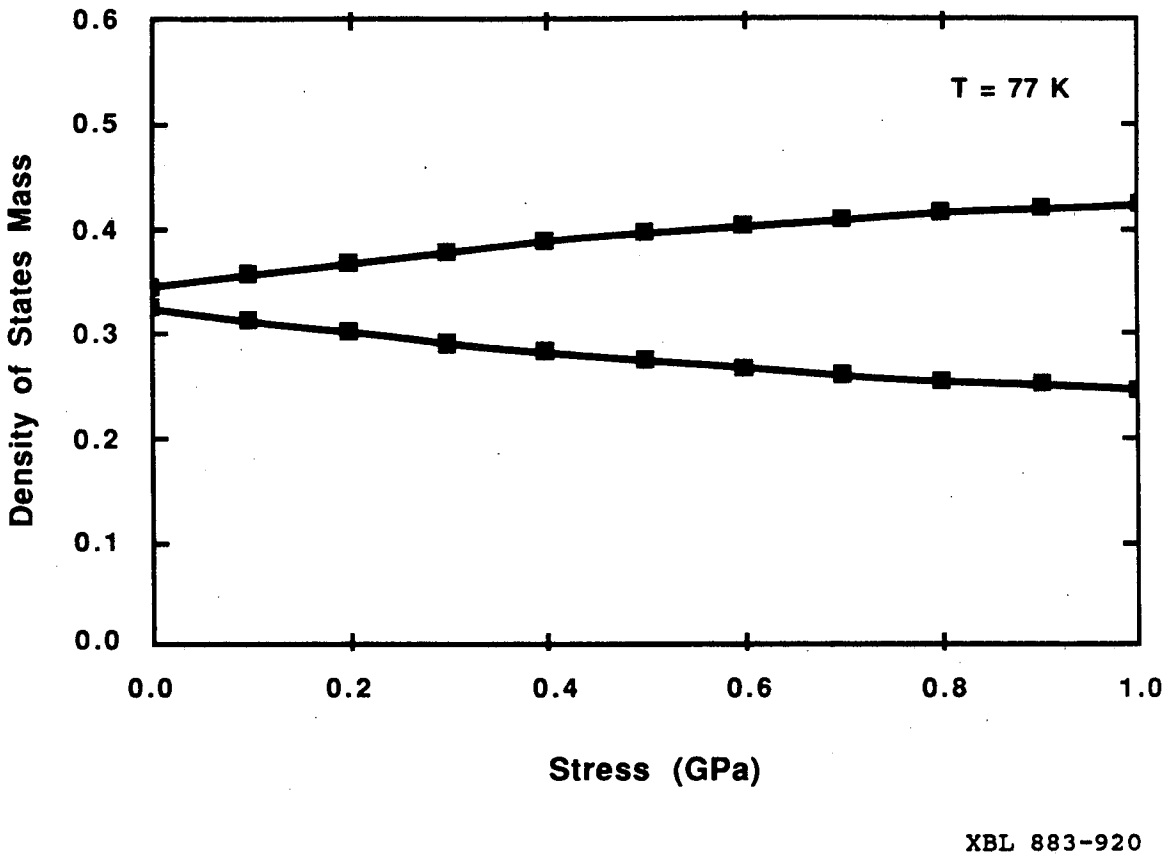


Fig. 3.8 Density-of-states effective masses as functions of stress derived from the exact thermal density of states in Fig.(3.7) using eq.(3.26).

away from the band edge. Shifts in the center of gravity of electronic states is caused by the trace (or hydrostatic component) of the strain tensor, which has already been included by the term $a\Delta$ in eq.(3.22).

The density of states plays a major role in the analysis of the change in thermal emission rates. However, the effective mass that is used to define the effective density of states has a different origin than the effective mass that describes the thermal velocity of carriers in the band. Though the thermal and density-of-states masses are certainly related through the curvature of the energy dispersion, it remains to establish the connection between the two masses and find the combined effective mass of eq.(3.11). This is done numerically by integrating eq.(3.16) over k-space in an analogous fashion to the evaluation leading to the density of states N(E). The calculated average velocity as a function of energy is plotted in Figs.(3.9a-c) for 0.0 GPa, 0.3 GPa, and 0.6 GPa of compressive uniaxial stress oriented along [100]. The best fit from the independent-band model is included as the solid line. The independent-band model clearly overestimates the velocity in the region near the strong mixing of the two bands, just as the model underestimates the density of states in Fig.(3.6) in the same energy region. This is a direct consequence of the interaction of the two bands which results in avoided crossing (from symmetry considerations). The avoided crossing changes the curvatures and slopes of the bands near the crossing point and therefore alters the density of states and thermal velocities. The independent-band model therefore has difficulty modelling the the density of states and the energy-dependent velocities separately near the crossover between the light- and heavy-hole bands for moderate stresses. Thermal-emission probabilities, however, do not depend on these properties separately, but depend on the product of the density of states with the carrier velocity (see eq.(3.5)). This product is

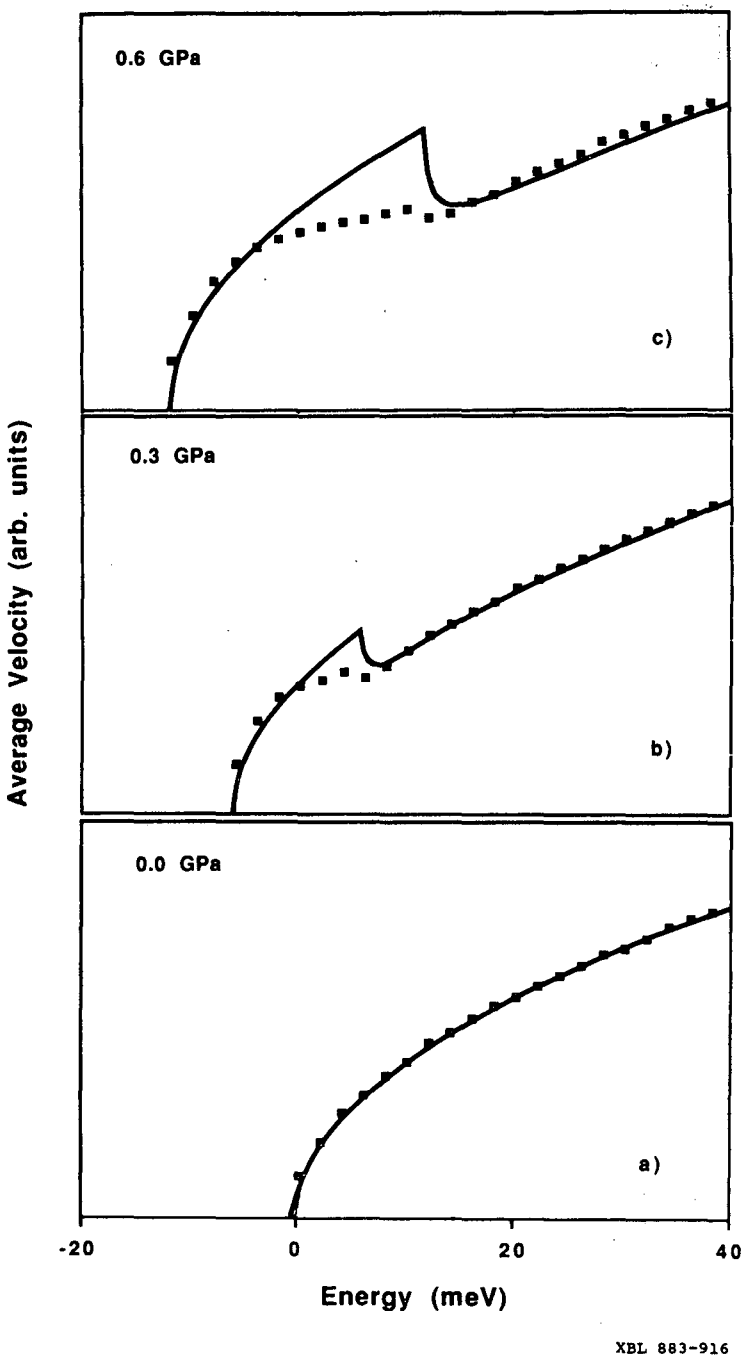


Fig. 3.9 The average energy-dependent velocity for a) 0-stress, b) 0.3 GPa and c) 0.6 GPa. The solid curves are the best fit from the independent-band model.

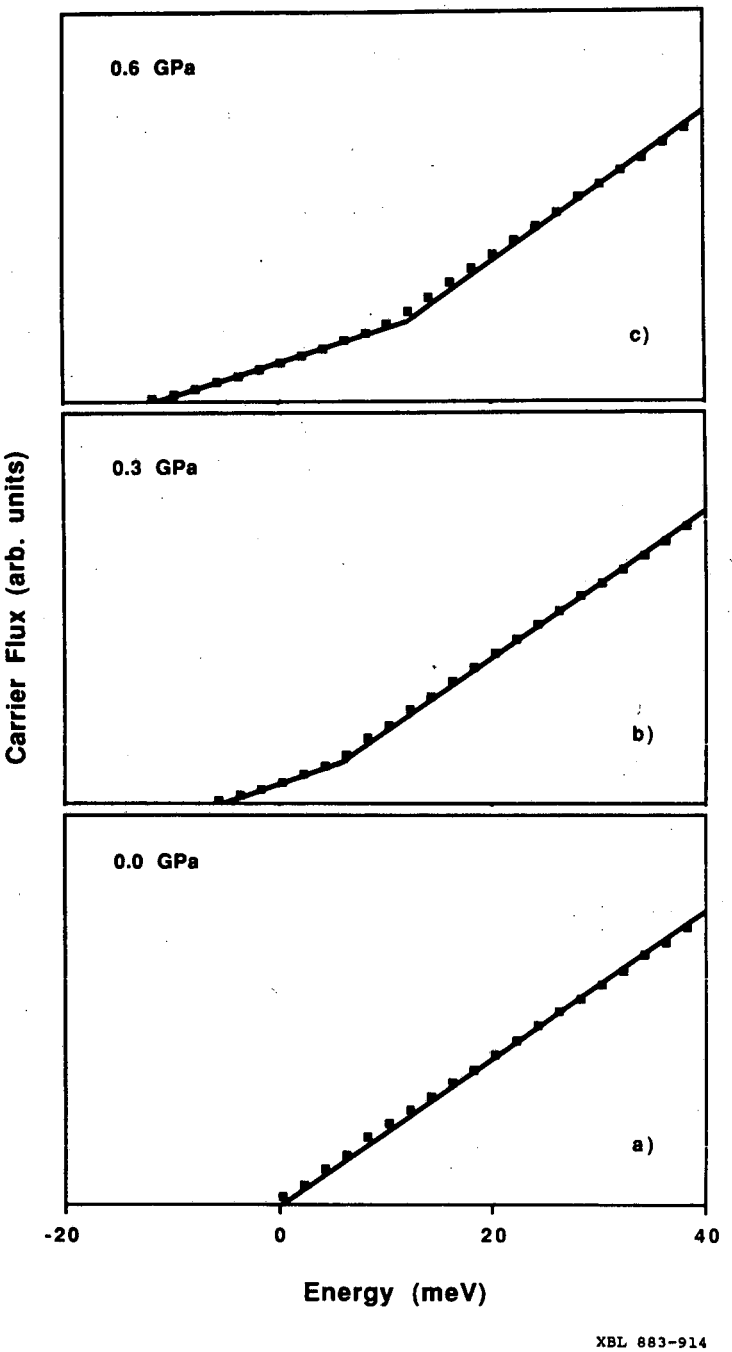


Fig. 3.10 The product of Figs.(3.6) and (3.9) for a) 0-stress, b) 0.3 GPa and c) 0.6 GPa. This product is directly proportional to the probability for the emission of a carrier to the given energy. The independent-band model gives a good fit for all stresses.

presented in Fig.(3.10) for the three stresses of Figs.(3.6) and (3.9). Both the density of states and the velocity depend on the square root of the energy, so the independent-band model predicts a linear dependence on energy for the product of the two values. The best fit from the independent-band model is included as the solid curves. The independent-band model gives an excellent fit to the numerical calculations for all values of stress, even in the energy region in which there is strong mixing of the two bands. This is because the thermal emission is insensitive to the dynamics of the holes (and hence the avoided crossing) and only samples the phase space available through the term $\left[\int dS_E\right]$ in eq.(3.17). The independent-band model is therefore able to model thermal-emission probabilities very accurately.

To find the appropriate combined effective mass of eq.(3.11) the full thermal-emission-rate equation eq.(3.17) is evaluated numerically and compared to the normalized thermal emission rate based on the independent-band model

$$\frac{e_0(\tau)}{e_0(0)} = \frac{m_l(\tau)\ \exp(E_\varepsilon / k_B T) + m_h(\tau)\ \exp(-E_\varepsilon / k_B T)}{m_l(0) + m_h(0)} . \qquad (3.27)$$

The resulting values for the combined effective masses are shown in Fig.(3.11). From the DOS and combined effective masses, the thermal effective mass is derived from eq.(3.11). All three effective masses for the light holes are given in Fig.(3.12) for compressive uniaxial stress oriented along [100] at T = 77 K . The heavy-hole masses are approximately given by $m_h^*(\tau) = 2m_l^*(0) - m_l^*(\tau)$.

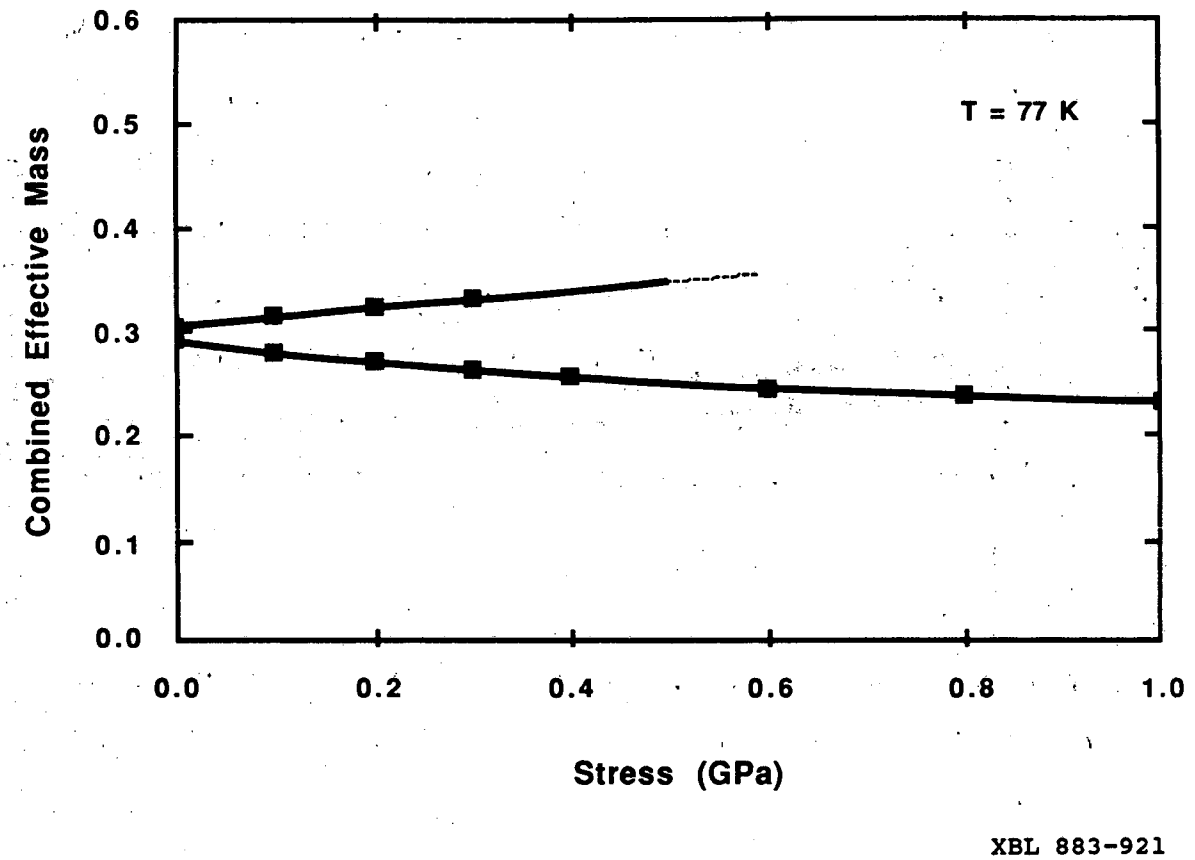


Fig. 3.11 The combined effective mass fitted to the numerical results Fig.(3.10) based on the independent-band model of eq.(3.27).

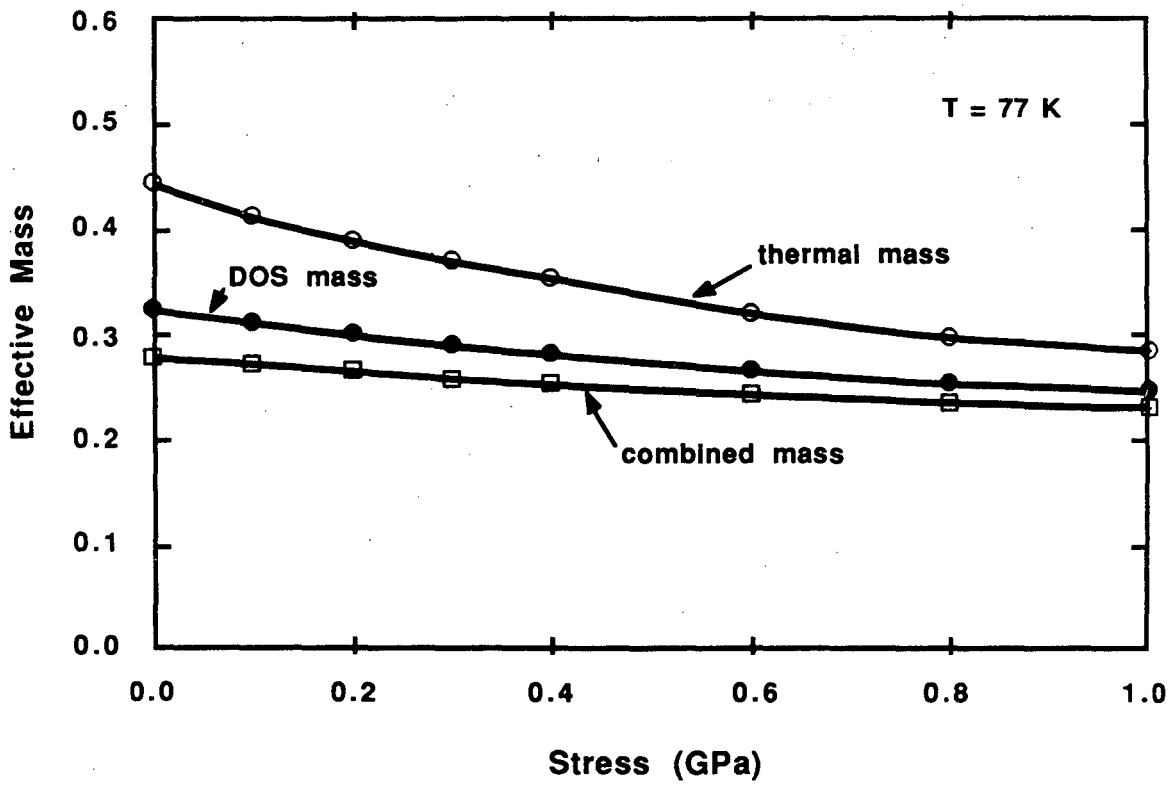


Fig. 3.12 The density-of-states, thermal, and combined effective masses for the light-hole band as function of stress. The corresponding heavy hole masses are given roughly by $m_h^*(\tau) = 2m_l^*(0) - m_l^*(\tau)$.

In this chapter, the energy shifts and the effective masses have been described which are appropriate for the independent-band model describing the heavy- and light-hole valence bands under uniaxial stress. The understanding of how these parameters vary with stress is essential to derive the energy shifts of a defect under stress from the independent-band model of eq.(3.10). To test the completeness of this understanding of the valence band, it is useful to observe the DLTS spectrum of a defect which is insensitive to stress. This defect level can act as a reference level from which the properties of the valence band under stress can be measured. Iron-aluminum pairs in p-Si are chosen as insensitive reference levels. In the next chapter the structure of the iron-acceptor pairs are described and a justification is given of the insensitivity of the pairs to stress. Using the FeAl level as a reference, the validity of the independent-band model is tested using the stress-dependent effective masses shown in Fig.(3.11).

Chapter IV. Iron-Acceptor Pairs in Silicon

I. Previous Research and Pairing Reactions

Iron-boron pairs are the dominant defects present in Si:B following iron diffusion and quenching. They constitute part of a larger class of iron-acceptor pairs including FeAl, FeGa, and FeIn. The pairing reaction occurs in p-type material in which the Fermi level lies in the lower half of the gap. For this position of the Fermi level, both the iron and the acceptor are ionized, but with opposite charge. The resulting Coulomb attraction between the ionized donor and acceptor provides the mechanism for the capture of the mobile interstitial iron by the substitutional Group III element.

Iron-acceptor pairs in silicon were first detected by Ludwig and Woodbury[1962] using electron paramagnetic resonance (EPR). The defects were found to be highly anisotropic and to exhibit trigonal symmetry, except for FeIn which exhibited orthorhombic C_{2V} symmetry (ordinarily called rhombic I). Recently a rhombic I configuration has also been identified [van Kooten,1984] for FeAl using EPR. Detailed studies of the EPR of FeB pairs in silicon have been carried out independently by Gehlhof and Segsa[1983] and van Kooten et. al.[1984] They both arrive at a strong negative trigonal field. On physical grounds the trigonal field would be expected to be strongly repulsive because of the negative charge of the boron. The model for the iron-acceptor pairs (derived exclusively from the EPR data) places the iron in the tetrahedral interstitial site adjacent to the substitutional boron in the trigonal centers (2.35 Å nearest neighbor separation). In the rhombic I center of FeAl the iron is

assumed to be situated in the next-nearest-neighbor interstitial site with respect to the boron (2.72 Å next-nearest-neighbor separation). Of course, the iron atom is likely to relax towards the substitutional site because of the attractive Coulomb field.

Iron-acceptor pairs in silicon have been studied extensively using deep level transient spectroscopy [Wünstel,1982; Kimerling,1983]. After iron diffusion in silicon a level is observed at E_V + 0.44 eV. This level is assigned to the isolated interstitial iron. A level is also observed at E_V + 0.10 eV in Si:B after iron diffusion which has been ascribed to the FeB pair. Two defect levels have been observed in Si:Al after iron diffusion; one at E_V + 0.13 eV and another at E_V + 0.20 eV. These have tentatively been identified as the rhombic I and trigonal FeAl configurations observed in EPR [Chantre,1985]. Two levels associated with FeGa pairs occur at E_V + 0.24 eV and E_V + 0.17 eV in Si:Ga, while a level associated with FeIn occurs at E_V + 0.2 eV in Si:In.

As proposed by Ludwig and Woodbury, the 4s electrons of interstitial transition-metal elements in silicon are not required for bonding and are transferred to the 3d shell. The iron interstitial energy level (E_V + 0.44 eV) is assigned to the first ionization stage of iron, Fe^+/Fe^0, or a $3d^7$-to-$3d^8$ configuration transition. The iron-acceptor pair energy levels, which lie between the interstitial level and the acceptor levels, are considered to involve the second ionization stage of iron, Fe^{++}/Fe^+, with the acceptor remaining ionized and thus negatively charged. This is a $3d^6$-to-$3d^7$ electron configuration transition. The second ionization stage of iron is normally buried deep within the valence band [Zunger,1986], but the presence of the negatively charged acceptor pushes the energy level up, out of the valence band and into the bandgap by Coulomb repulsion.

Iron-acceptor complexes have been investigated with photoluminescence. The luminescence from the isoelectronic trap observed in Si:FeIn has been shown to be consistent with a <100> orientation of the constituent atoms [Sauer,1983] which is also the symmetry observed by EPR for FeIn pairs. An isoelectronic trap observed in Si:FeTl shows similar structure as the FeIn trap [Thewalt,1982], but stress and magnetic field perturbations indicate that the FeTl related trap may undergo temperature induced transformations between trigonal and rhombic configurations [Watkins,1985]. On the other hand the luminescence from the isoelectronic trap observed in Si:FeB exhibits no axial symmetry under the application of a magnetic field [Mohring,1984]. This lack of orientational splitting of the no-phonon line, along with other less direct evidence [Schlesinger,1983], leads to the conclusion that the FeB related defect observed in photoluminescence is not the FeB pair observed in EPR.

Perhaps the most interesting feature of the iron-acceptor pairs is their metastability. Chantre et al.[1985] have demonstrated that the FeAl(2) defect is a metastable state of the FeAl(1) defect. A metastable configuration has also been found for FeGa [Benton,1987]. The transformation between the two configurations is controlled by charge-state dependences of the total energy. The activation energy for the transformation is about 0.5 eV. The phenomenological model proposed to explain this metastability is based on the electrostatic energy of the Fe^+ in the electric field of the ionized acceptor. The iron atom in the stable state is assumed to lie in the nearest-neighbor interstitial site, while in the metastable state it is assumed to lie in the next-nearest-neighbor interstitial site (this model is identical to the model proposed to explain the EPR results of FeAl). The model gives reasonable predictions for the energy differences between the stable and metastable states, and the activation energy for transformation is of the same order as the activation energy for

diffusion of interstitial iron. On the other hand, the model cannot explain the chemical dependence of the metastable state. In the purely electrostatic model the local strain induced by the mismatch in the ionic size of the acceptor atoms compared to the silicon atoms should push the iron atom further from the substitutional acceptor site, thereby reducing the activation energy as the acceptor atom ionic radius increases. This is directly opposite to the observed trend. The influence of the central-cell potential of the acceptor on the iron wavefunction should be negligible when the iron is situated at the next-nearest-neighbor site, leading to the roughly same activation energy for all the metastable states. This is not the case. The metastable state activation energy increases with the binding energy of the isolated acceptor, suggesting substantial overlap of the iron wavefunction with the central-cell potential of the acceptor.

It is clear from the foregoing review of the experimental observations of iron-acceptor complexes in silicon that the defects can exhibit a large variety of sometimes contradictory symmetries, and show interesting metastable behavior which is not yet fully understood. To attempt to determine the symmetries of the iron-acceptor pairs, uniaxial stress was applied in conjunction with DLTS.

II. Stress Data

A. Experimental Details

Iron diffusion into samples of 5×10^{15} cm^{-3} boron or aluminum doped silicon samples was performed by first coating the sample with $FeCl_3$ and then heating for two hours at 1000 °C in an argon atmosphere. The chlorine had no measureable effect on the iron diffusion or on the electrical characteristics of the sample after diffusion. The use of the $FeCl_3$ was not essential; some samples

which were cleaned carefully and annealed at 1000 °C were found to have comparable iron concentrations. This iron may have been present already in the sample and was simply activated by the high temperature anneal. The samples were quenched in air, then annealed at 80 °C for two hours. The quench is necessary to capture the iron as interstitials without precipitation, and the subsequent 80 °C annealing stage facilitates the movement of the iron interstitials to speed up their capture by the substitutional acceptors. The samples were etched and allowed to oxidize before evaporation of 300 Å of gold onto one surface to form the ohmic contact. The rectifying contact was formed by evaporating 1000 Å of Al onto the opposite side. The evaporation was performed with minimal current in order to avoid heating of the contact. The aluminum was covered by 300 Å of gold to protect the aluminum from the indium electrical contacts. Typical concentrations of the FeB and FeAl levels after preparation ranged from $5x10^{12}$ cm^{-3} to $5x10^{13}$ cm^{-3}.

The samples to be used for stress experiments were cut from the contacted sample into 1 x 1 x 6 mm^3 parallelepipeds. The ends of the cut samples were etched to remove the sharp edges. Copper wire (32 guage) was attached to the samples using pressed indium contacts. Samples were mounted into a DLTS stress rig capable of achieving 100 Newtons per mm^2, or 1 GPa of uniaxial stress.

The stress data were taken with the time constant of the DLTS correlator set to a fixed value, usually around 3ms. Zero-stress data were taken, then the stress was increased by intervals of about 0.1 GPa to nearly 1 GPa. The zero-stress measurement was repeated afterwards to check for hysteresis. Data were always taken for increasing temperature under 1/100 atm of helium exchange gas in order to equilibrate the temperatures within the stress rig and remove any effects from thermal hysteresis. A representative example of stress

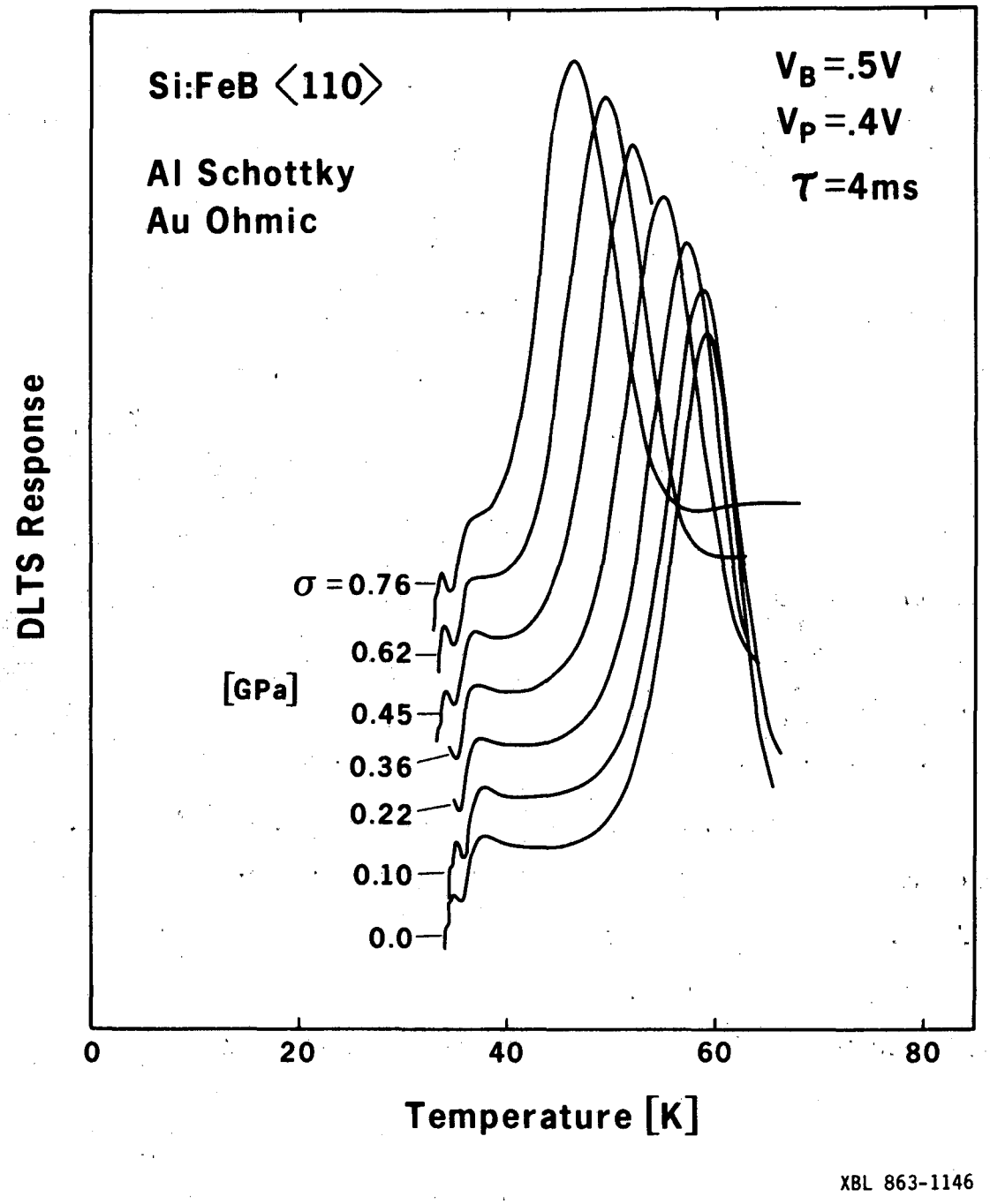


Fig. 4.1 DLTS spectrum of FeB under [110] uniaxial compression. The data were taken for a constant emission time constant of 3 ms, with the stress varying from zero to 0.76 GPa. The period of each scan was approximately 3 minutes.

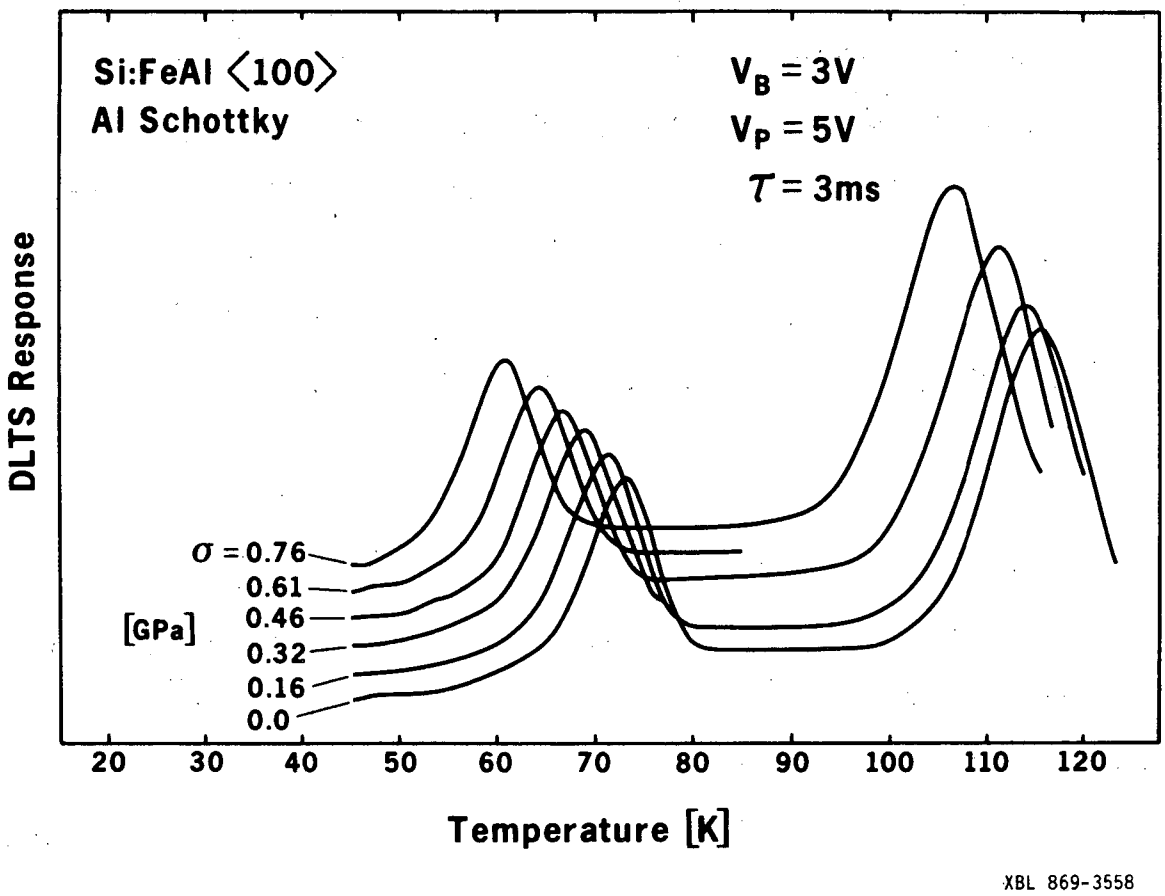


Fig. 4.2 DLTS spectrum of FeAl in Si under [100] uniaxial compression. Data were taken for a constant emission time constant of 3 ms.

data for FeB is shown in Fig.(4.1) for increasing stress along a <110> direction at constant emission rate, and for both FeAl defects in Fig.(4.2) for a <100> stress. Two features are evident: first, there are no splittings greater than 2 K/GPa; second, the peak shifts in temperature by roughly 15 K/GPa. The effect of thermal emission to a stress-split valence band has been considered in chapter III and explains much of the large observed shift in temperature of the DLTS peak. Using this analysis, the shifts in energies of the defects can be extracted. These shifts in defect energy may provide information on the possible defect symmetries.

B. Shear Dependence

The shifts in the temperature of the defect peak positions for increasing stress are shown in Fig.(4.3a-d) for the two defects under [100] and [111] stresses. The solid lines are the best fit of the data to eq.(3.10) using the combined effective masses of Fig.(3.11) of chapter III. The only free parameter is the shift of the defect energy under stress, which in all cases is small compared to the shift of the valence band edge. The data fit the theory well in all cases. It is important to note that there is no evidence of an increase in the peak temperature under small stresses as is incorrectly predicted by the spherical approximation.

The results of this analysis are shown in Fig.(4.4) for the FeAl defects. Energy shifts under 1 GPa are plotted against stress direction. The origin of energy is taken at the center of gravity of the top two valence bands. The zero stress energies of the iron-acceptor pair are also indicated. There are two contributions to the shift in temperature of the DLTS peak. The predominant contribution stems from the splitting of the valence band. The remaining contribution is from the total energy shift of the FeB pair, which can be broken

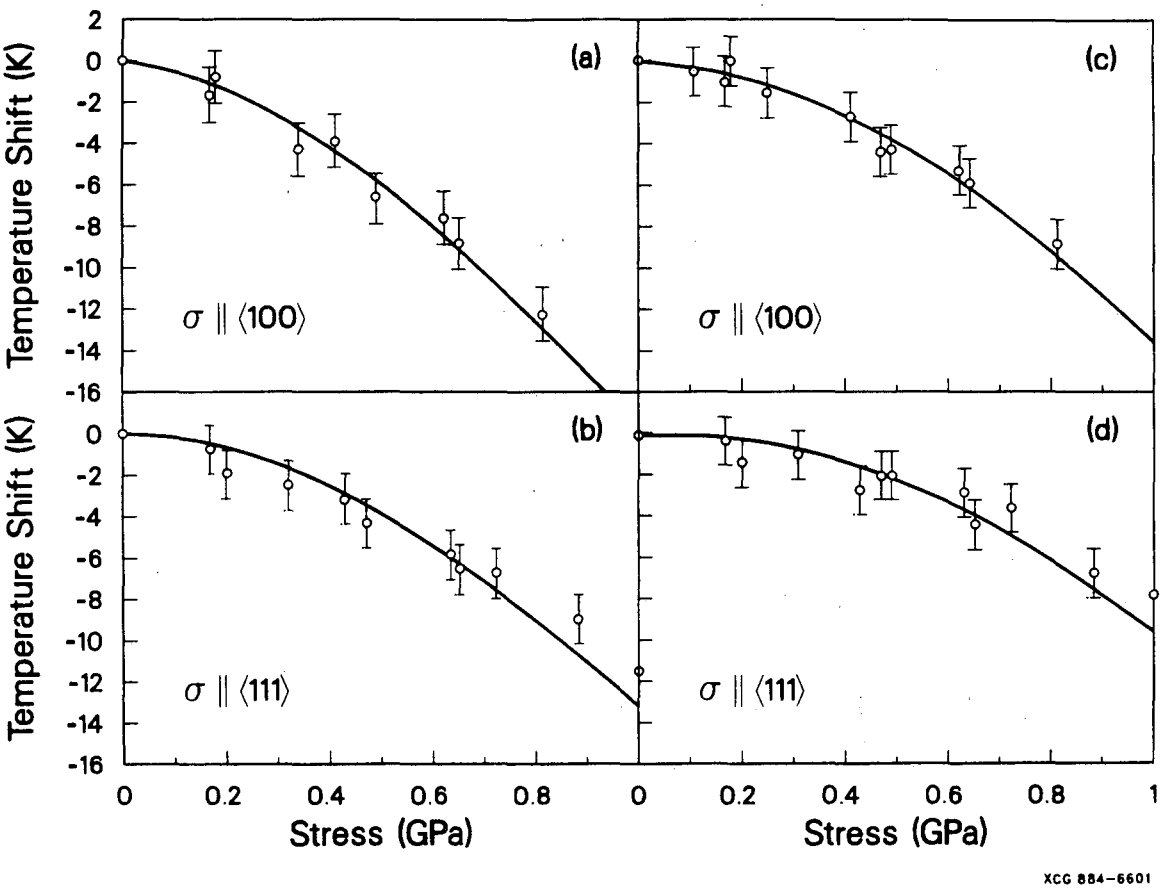


Fig. 4.3 Change in temperature of the DLTS peak with increasing stress for a constant emission rate for a) FeAl(2) under [100] stress, b) FeAl(2) under [111] stress, c) FeAl(1) under [100] stress, and d) FeAl(1) under [111] stress.

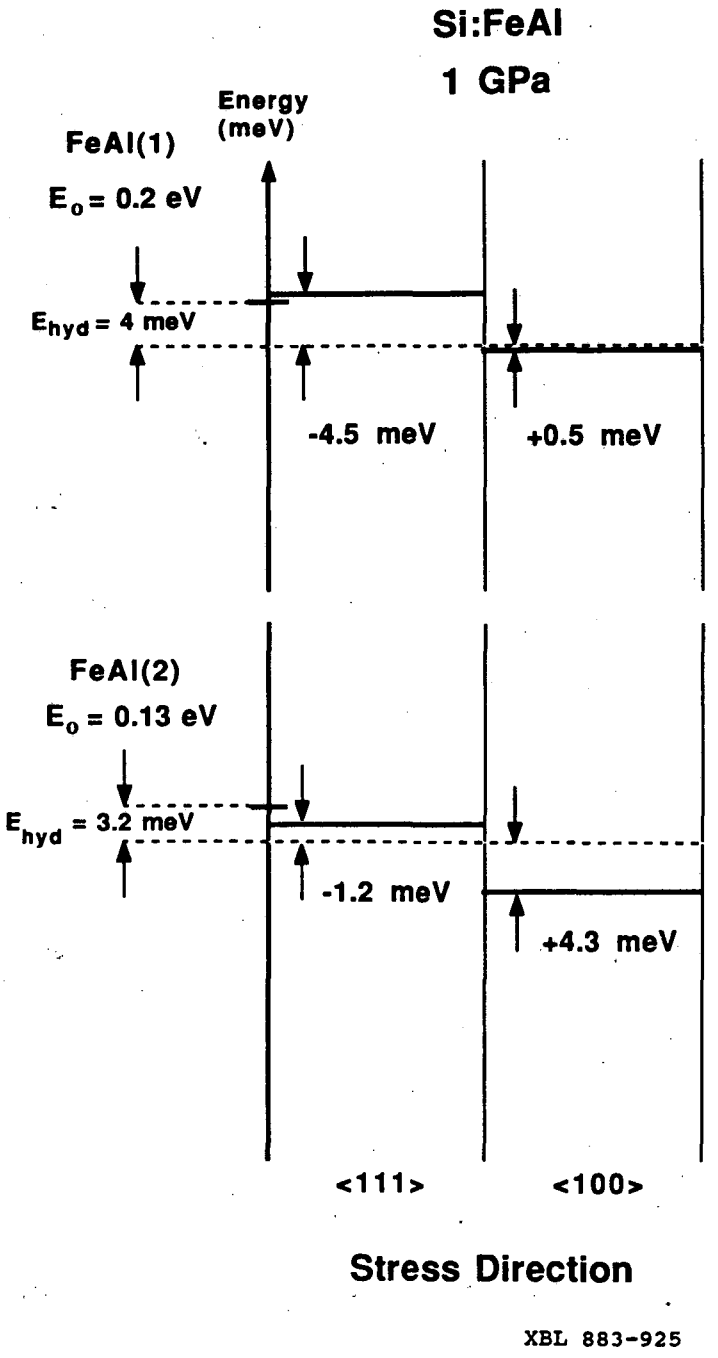


Fig. 4.4 Shear energy dependence of the FeAl(1) and FeAl(2) defects under 1 GPa as fitted from Fig.(4.3). The hydrostatic shifts are from Wünstel[1982b] and Weber[1987]. The deviation from the hydrostatic shift is the anisotropic dependence. The error on these values is ± 2 meV/GPa.

into a hydrostatic component and a shear component. The activation energy for thermal emission is therefore the difference between the shifted defect energy, and the energy of the topmost edge of the split valence band. The hydrostatic components are -4.0 meV.GPa for FeAl(1) and -3.2 meV.GPa for FeAl(2). When these values are subtracted from the total energy shifts of the defect, the shear energy shifts remain.

III. Defect Symmetries

The energy shifts of the iron-acceptor complexes show qualitatively similar anisotropy. There is a consistent anisotropy of 5 meV/GPa for both defects, but the baseline is shifted for the FeAl(1) defect relative to the FeAl(2) defect. The uncertainty in the measurements and the complicated analysis produce error bars that make this anisotropy a one-standard-deviation effect. Such uncertainty makes any conclusions concerning the symmetry of the defect impossible. However, it is interesting to understand why the stress sensitivity of the iron-acceptor pairs is so small. To help elucidate the problem, the expected stress properties of several possible structures for the iron-acceptor pairs are presented next. These are: trigonal symmetry; rhombic symmetry; and a dipole in T_d symmetry.

A. Trigonal Defect

The energy shifts of a trigonal defect are shown in Fig.(4.5). Trigonal defects experience largest shifts when stressed parallel to <111> directions, and zero shift for stresses along <100> directions. The defect levels associated with the $3d^6 \rightarrow 3d^7$ transition on the iron atom can be written as

$$3d^6 + e_V \rightarrow 3d^7$$

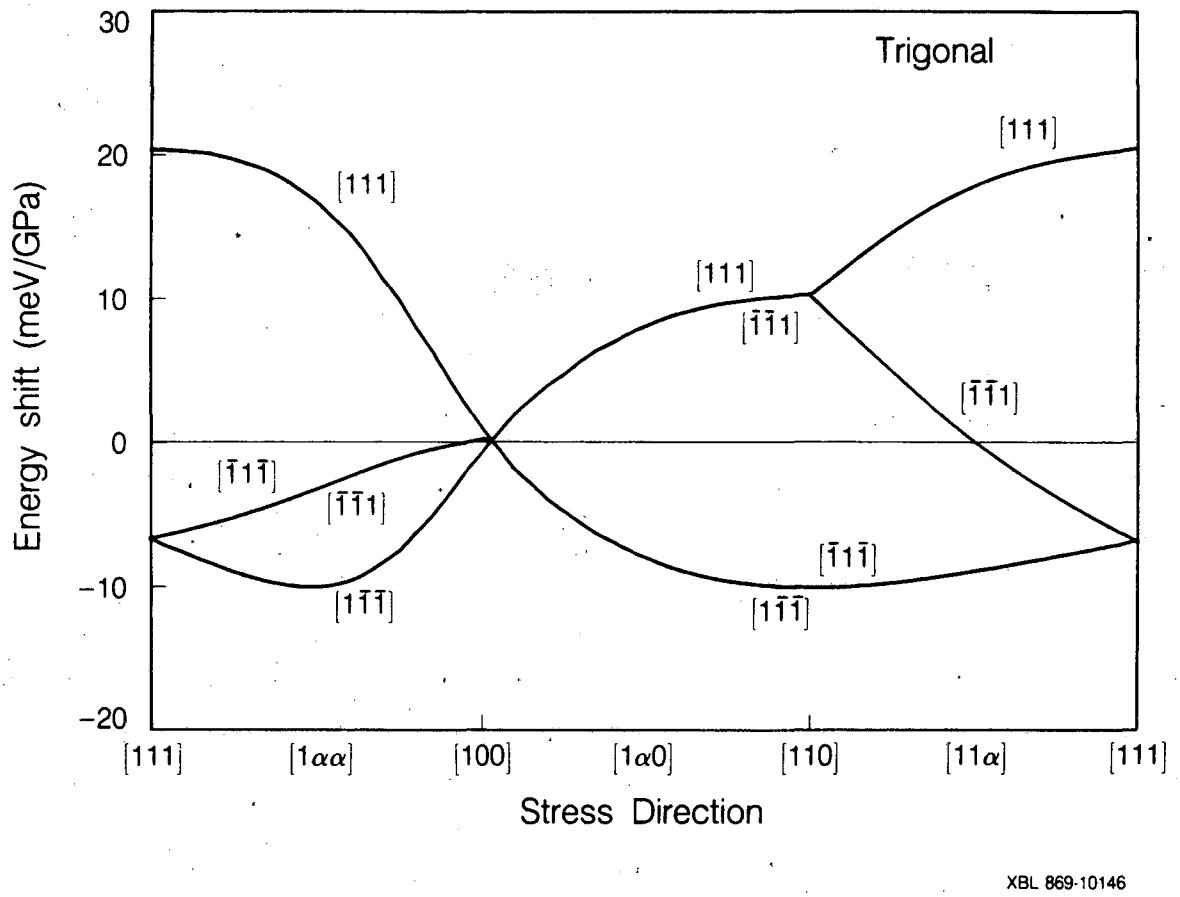


Fig. 4.5 Splittings of a trigonal defect for several stress directions.

where an energy ΔE is required to promote an electron from the top of the valence band to produce the $3d^7$ configuration from the $3d^6$ configuration. The energy balance for the transition is

$$E(3d^6) - 2\, e^2/\varepsilon_{av}d + E(vb) + \Delta E = E(3d^7) - e^2/\varepsilon_{av}d\,, \qquad (4.1)$$

where $E(3d^6)$ and $E(3d^7)$ are the total energies of the $3d^6$ and $3d^7$ configurations of the isolated, interstitial iron atom respectively when occupied in their ground states. The terms $e^2/\varepsilon_{av}d$ are the contributions of the electrostatic energy to the total energies of the initial and final states, where d is the separation between the iron atom and the acceptor atom, and ε_{av} is the dielectric permitivity integrated from $\varepsilon(\infty)$ to $\varepsilon(d)$. These terms arise from

$$e^2/\varepsilon_{av}d = -\int_{\infty}^{d} \left(e^2/\varepsilon(r)\, r^2 \right) dr\,. \qquad (4.2)$$

This energy is on the order of 1 eV (assuming that $\varepsilon \approx 10$), and is responsible for pushing the energy of the $3d^6$ to $3d^7$ transition of isolated interstitial iron up out of the valence band. E(vb) is the energy of the valence band, and ΔE is the activation energy. Usually E(vb) is taken as the origin and ΔE is simply the difference between the total energies of the two configurations. Under applied uniaxial stress the change in the activation energy as a function of stress τ becomes

$$\Delta E(\tau) = \Delta E(3d^7;\tau) - \Delta E(3d^6;\tau) - \Delta E(vb;\tau) + e^2/\varepsilon_{av}d(\tau)\,, \qquad (4.3)$$

where the iron-acceptor separation d is explicitly considered as a function of the stress. The effect of stress on the valence band has already been considered in chapter III. From the above considerations, it is clear that the splitting of the initial state as well as the final state under uniaxial stress must be considered. This is demonstrated for the trigonal iron-acceptor pairs in in Fig.(4.6). The total energy of the pair oriented along the [111] stress axis decreases more for the $3d^6$ initial state than for the $3d^7$ state because of the extra charge in the initial state. If the defect is free to reorient at the elevated temperatures of DLTS, then the orientation parallel to the stress direction will be preferentially populated and the total defect energy difference between the $3d^6$ and the $3d^7$ states will increase by the amount proportional to $e^2/\varepsilon_{loc}d$. The defect is insensitive to [100] oriented stress and there would be no shear shift for this stress direction. Therefore, the shear shifts shown in Fig.(4.4) are consistent with a reorienting trigonal defect.

The shear stress derivative for the defect can be calculated for this electrostatic model. The expression is simply

$$dE_{<111>}/d\tau = 3/4\ (e^2/\varepsilon_{loc}\, d)\ s_{44}\,, \qquad (4.4)$$

where $\varepsilon_{loc} = \varepsilon(d)$ is the local dielectric permitivity at the site of the iron atom, and s_{44} is the shear compliance constant. The experimental value is 5 meV/GPa. If d=2.35Å (nearest-neighbor interstitial distance), then the local dielectric constant would be $\varepsilon_{loc} \approx 10$. This is near the value of the full bulk dielectric constant for Si. Therefore the small shear stress shift of the FeAl(1) pair is consistent with the electrostatic model of Chantre. However, the justification of such a large dielectric constant for such short distances remains an open problem. Furthermore, the defect must be required to reorient to explain the

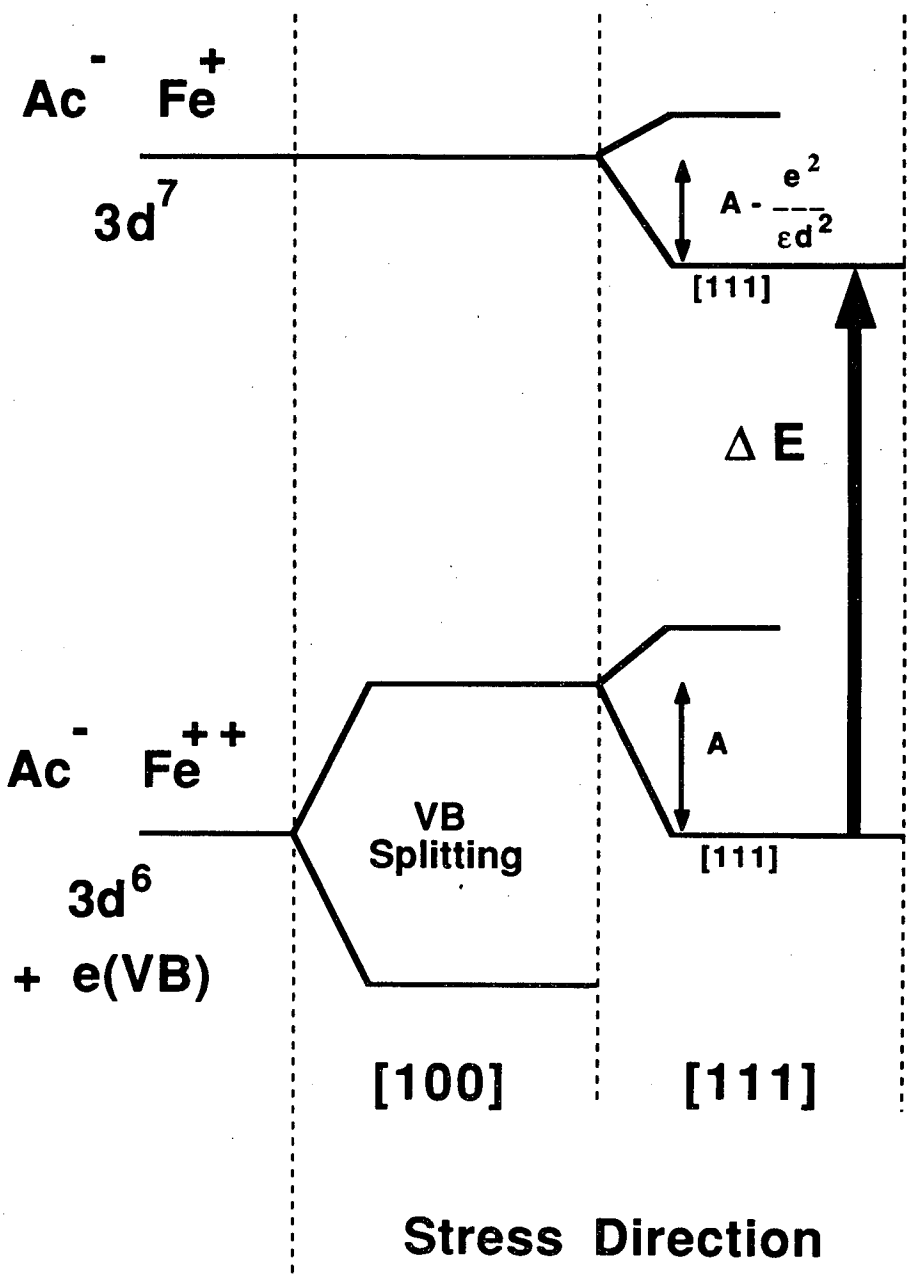


Fig. 4.6 The dependence of the thermal ionization energy on applied stress after the effects of stress on the valence band are removed. The splitting from the electrostatic interaction of the Fe with the acceptor is larger for the $3d^6$ configuration than for the $3d^7$ configuration.

sign of the shift for [111] stress if that shift originates with the defect. The possibility that this shift is an artifact of the valence band cannot be ruled out.

B. Rhombic Defect

The model for the defect observed in EPR with rhombic I symmetry places the iron in the next-nearest-neighbor interstitial site. The position of the iron is shown in Fig.(4.7) nested between a pair of bonds at a distance of 2.72Å from the substitutional site. In the actual defect, the iron is likely to relax towards the acceptor site. There are six equivalent orientations for the iron atom in this case corresponding to the six combinations of bond pairs. The symmetry of this defect is C_{2v} and it has associated with it two deformation potentials. These can be calculated in a similar fashion to the deformation potential calculated for the trigonal defect. The results are

$$dE_{<111>}/d\tau = 1/2\ \Xi\ s_{44}\ ,$$
$$dE_{<100>}/d\tau = -2/3\ (e^2/\varepsilon_{loc}d)\ (s_{11} - s_{12})\ , \qquad (4.5)$$

where Ξ is the transverse deformation potential describing the interaction of the iron with the pair of bonds. The splittings of C_{2v} symmetry are shown in Fig.(4.8). In a similar manner to the case of the trigonal defect, the defects oriented along the [100] stress direction have lower energy than those oriented perpendicular to the stress direction. The decrease in the activation energy for [100] oriented stress can only be explained if the defects are not allowed to reorient. In this case, the DLTS peak will be broadened slightly, with more weight in the defects perpendicular to the stress. The energy separation between these orientations decreases with increasing stress and would explain

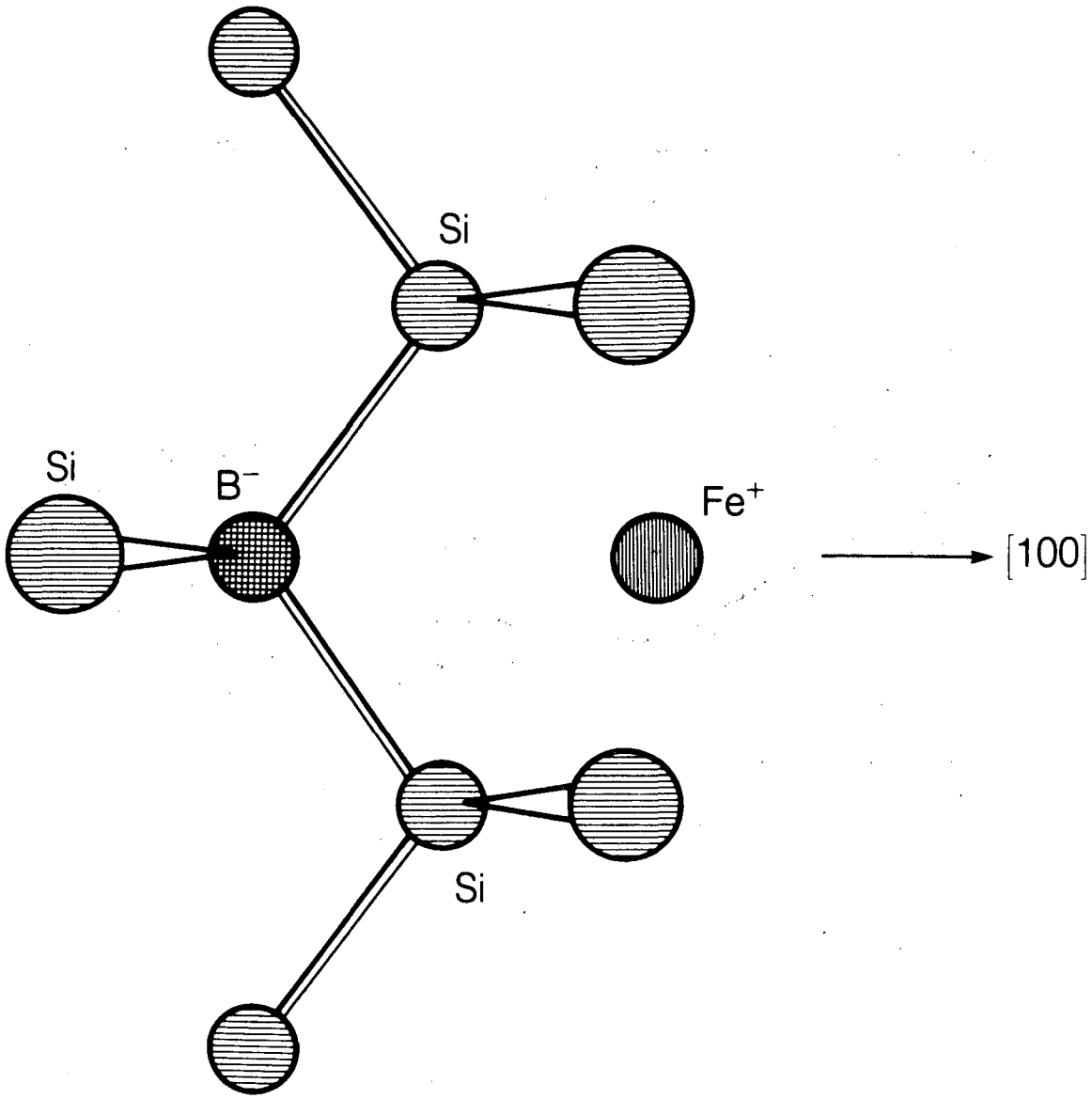
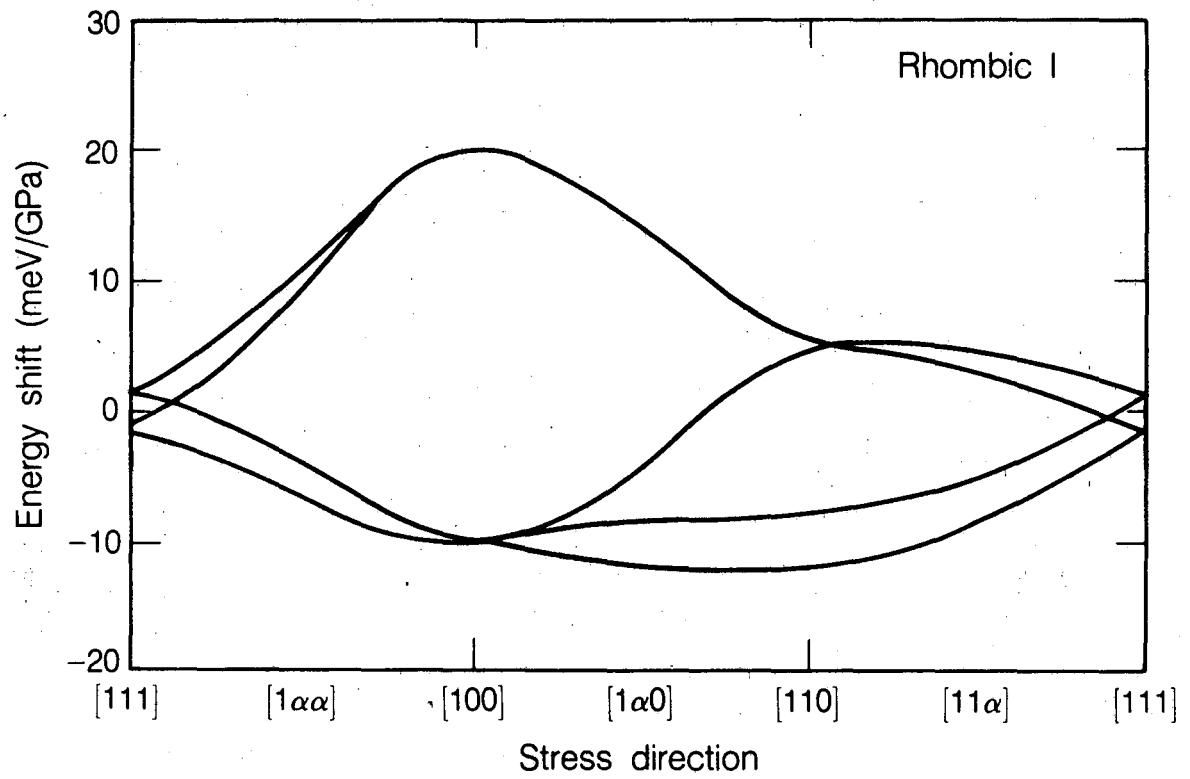


Fig. 4.7 Rhombic I model for the FeB pair in silicon seen in the (110) plane. The iron has six configurations among all six equivalent bond pairs.



XBL 865-10143

Fig. 4.8 Stress splittings for a rhombic I defect.

the decrease observed for the FeAl(2) defect for [100] stresses. The sensitivity of these rhombic defects to [111] oriented stress would be small.

The magnitudes of the shear shifts for the FeAl(1) and FeAl(2) defects are therefore seen to be consistent with the electrostatic model of Chantre et al. The change in sign of the shear contribution is furthermore only consistent with the respective trigonal and rhombic symmetries for these defects if the trigonal FeAl(1) defect is allowed to reorient and the rhombic FeAl(2) defect is not able to reorient at the temperatures of DLTS. Of course, the similarity in the anisotropy between the FeAl(1) defect and the FeAl(2) defect suggests that this anisotropy may arise from the valence band and not from the defect at all. A 0.5 eV error in the relative values of the deformation potential parameters b and d for the valence band could account for the observed shifts.

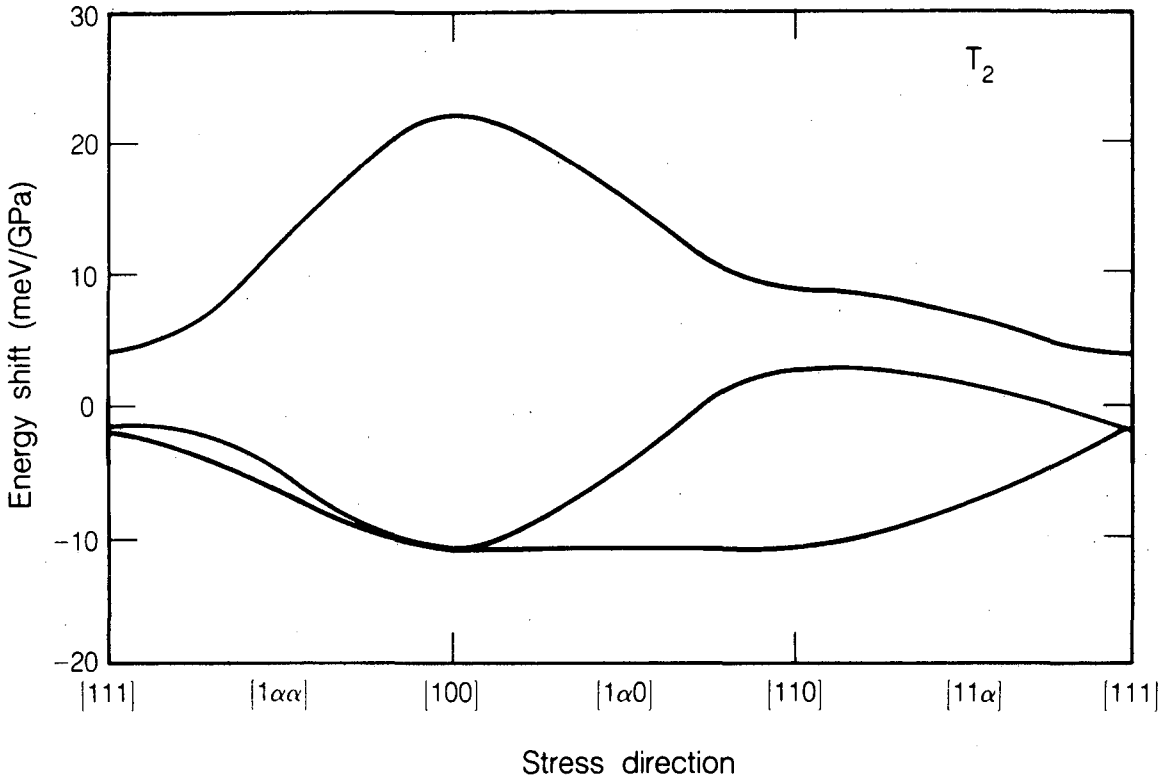
C. Classical Dipole in a Tetrahedral Environment

As a final possibility for the defect structure, the stress properties of an electric dipole in a tetrahedral field are considered. This model is motivated by the fact that, to lowest order, the iron-acceptor pairs can be represented as simple electric dipoles: the iron is positively charged, the acceptor is negatively charged, and there is some small separation between them. The dipole operator in the tetrahedral group T_d transforms as the T_2 representation of that group, which is a threefold-degenerate representation. Neglecting coupling to higher order T_2 multipole moments of the tetrahedral crystal field, a dipole can orient in any direction with the same total energy, with no potential barriers; the dipole is a free rotator. When uniaxial stress is applied, a dipole located at the tetrahedral site must exhibit anisotropic energy shifts which depend on the direction of the stress. Specifically, the total energy of the dipole has an angular variation compatible with that of the T_2 symmetry. The energy variation of the

three components of the dipole with stress orientation is shown in Fig.(4.9) for specific values of normal and shear deformation potentials B=10 meV/GPa and C=3 meV/GPa respectively. It must be emphasized that this is a classical result. The three curves shown in Fig.(4.9) are the principal values of the energy ellipsoid describing the energy surface of the dipole in the strain field. As the stress direction is changed, the principal axes of the energy surface are altered and the dipole reorients accordingly to the direction having the lowest energy.

In fact, the dipole described by an iron-acceptor pair has a spatial extent which would couple the dipole to higher T_2 multipole moments of the tetrahedral crystal field. This results in a hindered rotator. In general terms, there should be a threshold energy below which the rotator becomes a librator, and the tetrahedral symmetry is then lowered: the isotropy of the system is destroyed. The potential energy must have minima in certain directions, possibly in the <111> or <100> directions. If these minima are shallow, the defect retains much of the stress behavior of a tetrahedral T_2 state. Yet the localization of the iron atom in these minima would explain the anisotropic symmetry observed in EPR and photoluminescense. Such a dipole model could in fact explain the large variety of symmetries observed by different techniques at different temperatures.

To carry this model further, the effect of trigonal minima on the stress behavior of a T_2 dipole can be calculated to compare with the experimental results to see if the model might be consistent with observation. A dipole of T_2 symmetry will exhibit four equivalent minima in this case. This effect can be modeled by considering the T_2 state to be split by a trigonal field into an A_1 singlet and E doublet of C_{3v} symmetry. The size of the splitting, Δ, is a measure of the potential hindering free orientation of the dipole in the crystal. If this



XBL 865-10142

Fig. 4.9 Stress splittings of a defect with T_2 symmetry. In the illustration, the piezospectroscopic constants are B=10 meV/GPa and C=3 meV/GPa.

dipole model is correct, then the iron-acceptor pairs should reorient easily. Reorientation was also invoked to explain the sign of the anisotropy for the FeAl(1) defect. To test for reorientation at low temperature, uniaxial stress was applied in conjunction with EPR. This technique has the advantage of high sensitivity to small changes in defect structure under stress, and all orientations of the defect can be observed individually.

IV. Stressed EPR

The paramagnetic center observed in EPR is attributed to the $3d^7$ configuration of the iron. The hyperfine splitting from the interaction with the spin 3/2 on B^{10} is clearly observable and gives a positive identification of the defect. The FeB and FeAl EPR signals show strongly anisotropic g-tensors with C_{3v} symmetry. This symmetry is consistent with the iron occupying the next-nearest interstitial site adjacent to the ionized boron. Stress was applied in conjunction with EPR to attempt to observe electronic redistribution or reorientation of the C_{3v} axis of FeB pairs.

In order to observe the FeB defect in the paramagnetic $3d^7$ configuration, the Fermi level in the material must be pinned above the FeB level. This can be achieved by diffusing Fe into the material in excess of the B concentration, and quenching rapidly to capture the iron as electrically active interstitials. This pins the Fermi level at the $E_v + 0.40$ iron interstitial level. Iron from $FeCl_3$ was diffused into the EPR samples at 1150 °C for 30 minutes in a vertical oven and quenched rapidly by dropping the samples into diffusion-pump oil. The quench rate on the order of 200 °C/sec was sufficient to prevent the iron from precipitating and the material remained semi-insulating (Fermi level pinned around mid-gap) even at room temperature. The side surfaces of the [111] oriented 1x1x6 mm^3 samples were lapped after the diffusion and the

ends faced off perpendicular to the vertical orientation, followed by a polish etch. The samples were attached to the bottom of a 14 GHz EPR (TE_{011}) cavity mounted in a Janis varitemp cryostat which could be stabilized at temperatures from room temperature to 4.2 K. A maximum of 0.1 GPa could be applied to the sample with a polystyrene plunger.

The B-field was set parallel to the [1 -1 0] direction for which the EPR signal consisted of two closely spaced peaks from the [1 1 1] and [-1 -1 1] oriented defects, and two peaks from the other two directions. The maximum stress of 0.1 GPa was applied to the sample at the measurement temperature of 11 K and no change in the relative amplitudes of the parallel and off-parallel defects was observed. This rules out reorientation at low temperature. To test for reorientation at higher temperatures, the maximum stress was applied at around 230 K, then the sample was cooled for thirty minutes to 180 K and then the temperature dropped in five minutes to the 11 K measurement temperature at which the stress was removed. The stress was removed before the measurement to ensure that there was no stress broadening of the EPR signals. No observable changes in amplitudes were seen. Finally, the maximum stress of 0.1 GPa was applied to the sample at 270 K for sixty minutes, after which the sample was cooled slowly to 230 K at a rate of 1 K/min maintaining the stress, and finally quenched to the measurement temperature of 11 K. The EPR spectra were taken after the stress had been released at the base temperature of 11 K. No significant changes in the amplitudes of the EPR signals originating from the defects parallel and off-parallel were observed. From the magnitude of the stress and the temperature at which the stress was applied, an upper limit of 0.3 eV per unit strain can be set for the strain-coupling of the defect. From this upper limit, we can conclude that the FeB pairs will show energy splittings under uniaxial stress less than 3 meV/GPa. Such small splittings are below the

resolution of DLTS and are not likely to account for the common anisotropy in the FeAl(1) and FeAl(2) defects. Therefore, this anisotropy is most likely an artifact from the splitting of the top of the valence band.

V. Conclusions

In this chapter, the results of Chapter III have been used to find the shear dependence of iron-acceptor pair defect energy on stress orientation. A common anisotropy of 5 meV/GPa was observed for both the FeAl(1) defect and the FeAl(2) defect. The magnitude of this anisotropy is consistent with the electrostatic model of Chantre which requires the dielectric constant to be $\varepsilon(r) \approx 10$ for a separation of 2.35 Å. The sign of the anisotropy is only consistent if the FeAl(1) defect has trigonal symmetry and reorients, while the FeAl(2) defect has rhombic symmetry and is not able to reorient. Although this symmetry assignment is consistent with the symmetry assignments for these defects from EPR, there is no evidence from stressed EPR that the FeB defect reorients under uniaxial stress. Therefore no conclusions can be made concerning the symmetries of the stable and metastable iron-acceptor pairs.

Because of the insensitivity of the iron-acceptor pairs to shear stress, they become a convenient reference point from which to measure thermal emission of holes to the stress-split valence band. The shift in temperature of the FeAl defect DLTS peaks confirms the calculations of Chapter III that the density of states is divided roughly equally between the two displacing bands. This partition conserves the center of gravity of the split bands and accounts for the zero-slope asymptote of the temperature shift of the DLTS peak at low stresses. Furthermore, the 5 meV/GPa anisotropy observed for both the FeAl(1) and FeAl(2) defects may reflect a relative error between the accepted values of b and d.

Chapter V. Hydrostatic Band-Edge Deformation Potentials

I. Band Offsets

The energies of extended electronic (Bloch) states of a crystal change adiabatically with changes in crystal volume. This fact is expressed conventionally in terms of the tight-binding model of solids: as atoms are brought together from infinite separation, the individual atomic states overlap, forming bonding and anti-bonding combinations which have energies that spread into energy bands. The equilibrium lattice separation is determined by the trade-off between electronic bonding energy and repulsive nuclear-nuclear energy. The total energy is minimized at the equilibrium lattice spacing R_0. In the case of a semiconductor, a gap in the energy spectrum between occupied and unoccupied states occurs at the equilibrium lattice spacing. The edges of this gap, the valence and conduction band edges, vary in energy as the lattice spacing is changed.

The dependences of the band-edge energies on the crystal volume can be defined in several equivalent ways. The change of the band-edge energy with lattice spacing is given by the slope of the line tangent to the energy curve at R_0, and would be expressed in the units of eV/Å. A more convenient measure of the energy dependence on crystal volume is the hydrostatic deformation potential, denoted by a, which describes the energy change per unit volume strain, or equivalently

$$a = dE / d(\ln V) . \qquad (5.1)$$

Electron states which have energies that decrease under compression have a positive deformation potential. An important distinction must be made between relative and absolute deformation potentials. The relative change of the conduction and valence band edges under pressure can be measured easily using photo-excitation across the bandgap as the bandgap is altered by applied pressure. However, this technique measures only the *difference* between the individual band-edge shifts; the absolute energy shifts of the separate band edges cannot be measured this way.

Absolute hydrostatic deformation potentials describe the isotropic coupling of electronic states to the lattice. This chapter is concerned with the isotropic coupling of the states at the band edges to the lattice. This isotropic electron-lattice coupling plays an important role in several seemingly disparate aspects of the physics of semiconductors: acoustic deformation-potential scattering of charge carriers, band offsets in strained-layer superlattices, and volume relaxation around deep defect centers upon carrier emission.

Despite the importance of isotropic electron-lattice coupling, the absolute hydrostatic deformation potentials of the band edges have been elusive parameters that have evaded direct measurement. The difficulty in measuring the absolute band-edge deformation potentials arises from the fact that the band edges are buried roughly 5 eV in energy below the vacuum energy level, which is a possible reference level from which to measure absolute energies. This large separation in energy between the reference level and the band edges introduces a large uncertainty in the measurement of small perturbations on the energy levels. This uncertainty has precluded the direct measurement of absolute deformation potentials.

An alternative approach to measuring the band-edge deformation potentials would be to measure the changes in the band-edge energies with respect to an 'internal' reference level. Small changes in the energy separation between the band edges and the reference level could therefore be measured accurately. The need to define an internal reference level in semiconductors has arisen lately in the context of heterojunction band offsets. At the interface between two dissimilar semiconductors there are discontinuities (or offsets) in the energies of the valence and conduction band edges. These band discontinuities are the basic building block of the novel semiconducting devices designed through bandgap engineering. It is therefore of primary importance that these discontinuities be known accurately. Accurate experimental measurement of these discontinuities has only been possible, however, for the limited class of interfaces in which the lattice constants of the two materials match closely. For lattice-mismatched interfaces, local strain fields complicate the growth of the interfaces and the measurement of the discontinuities.

In the absence of good experimental values for the band offsets, there has been much theoretical activity focused on predicting the band lineups. Several approaches have been taken, from first-principles calculations [Van de Walle,1987] to empirical rules-of-thumb [Tersoff,1985]. One encouraging approach to predicting band lineups has been to define an intrinsic energy in each semiconductor material which lines up across an interface between two different materials [Flores,1979; Tersoff,1984]. This intrinsic energy has been called the charge neutrality level, because this is the energy which must match up across an interface in order to minimize surface polarization [Tersoff,1984]. Tersoff has defined the neutrality level as an inflection point in the real part of the Green's function describing the perfect crystal. The inflection point is interpreted as the energy of a localized electronic state in the bandgap for

which the contribution of the conduction bands to the structure of the localized state equals the contributions from the valence bands. It has been suggested that this construction of the neutrality level is equivalent to defining the energy position of the neutral vacancy. This connection between intrinsic deep-level defects and the neutrality level suggests that deep-level defects, or at least vacancies, could be used to predict band lineups. Namely, the band lineups obtained by lining up the energy levels of the vacancy between two different (but isovalent) materials should be the correct band discontinuities.

The energy levels of vacancies are difficult to characterize because of the tendency for vacancies to form complexes and because of strong reconstruction of the dangling bonds leading to large lattice relaxations. This complicates the use of vacancies to predict band lineups. However, it has been established that many deep substitutional impurity defects have strong vacancy-like structure. In particular, the energy structure of substitutional transition-metal defects have been described in terms of the Watkins' vacancy model [Watkins,1983]. Before the connection was made between deep-level defects and band discontinuities, however, it was found experimentally that the relative spacing of the transition-metal levels in one III-V semiconductor was the same in several other isovalent semiconductors [Ledebo,1982]. This observation was substantiated through Green's function calculations of transition metals in III-V and II-VI compounds [Caldas,1984]. Later, Langer[1985] amassed experimental data for the energy levels of several transition metals in III-V and II-VI semiconductors, and again found striking evidence for the invariance of the relative energy spacings between the transition-metal energy levels within isovalent semiconductors. By lining up these energy levels among the semiconductors, the band offsets between various semiconductors were predicted. This lineup hypothesis was tested for the single case of the lattice-matched GaAs-AlGaAs system, and was

found to agree well with experimental measurement of the band discontinuities. A physical explanation for the lineup of transition-metal defect levels across a hetero-interface was provided by Tersoff and Harrison[1987] based on the charge transfer between the d-shell of the transition metal and the dangling bonds of the vacancy. They found that the transition-metal defects were locked to the neutrality level within an additive constant.

The establishment of an internal reference level from which changes in band structure can be measured has direct application to the problem of measuring the absolute band-edge deformation potentials. The principle of the lineup of transition-metal defect levels across a hetero-interface is general and should apply as well to the interface between two materials that differ only by a perturbation. This is illustrated by the stress homojunction in Fig.(5.1). The interface is between two materials that differ only by their state of strain. The neutrality level, and hence the transition-metal level, remains constant across the interface. The band discontinuities, in this case, are the band-edge deformation potentials, which are precisely the electron-lattice perturbation potential that enters into the theory of acoustic deformation potential scattering. To measure the band-edge deformation potential, a stress homojunction is not required. By measuring the pressure derivative of a transition-metal defect level by an ionization technique such as deep level transition spectroscopy (DLTS), one is in fact measuring the isolated band-edge pressure derivative.

In this chapter, evidence is presented that the transition-metal-based reference level is valid for measuring band-edge deformation potentials. GaAs is selected as the ideal test material and DLTS stress data are provided from several different transition metals in GaAs among several different charge states. It is found that they possess a universal pressure derivative. This pressure derivative is compared to the deformation potentials obtained from

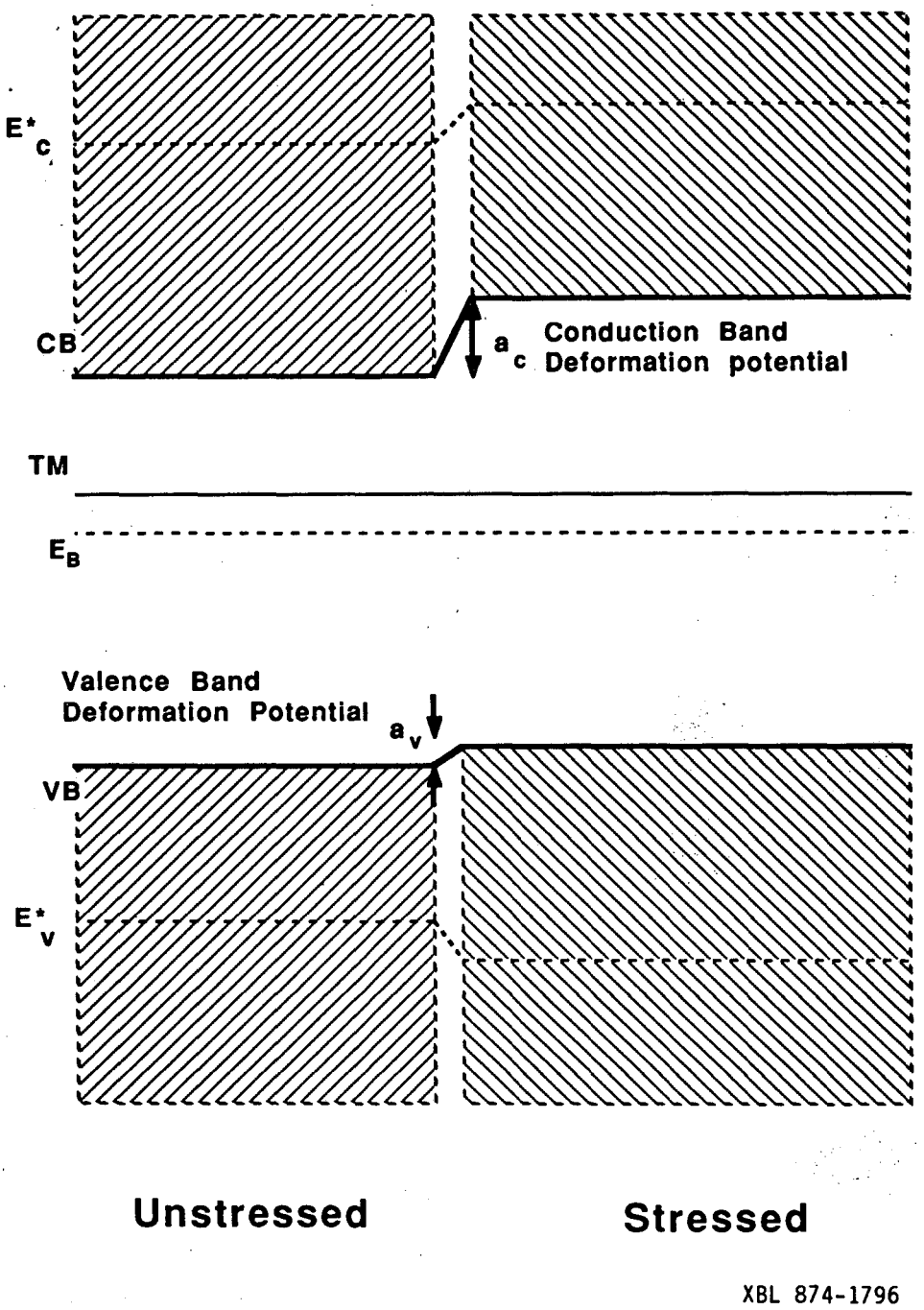


Fig. 5.1 Strain-induced homojunction. The semiconductors on the left and right differ only by their state of strain. The transition-metal level is taken as the reference. The band discontinuity caused by the strain is directly proportional to the band-edge deformation potential.

acoustic-phonon-limited mobility, and from first-principles theoretical calculations. All three values agree within experimental error. This cross-correlation among the values from three very different techniques provides a strong internal consistency. In section III the principle is extended to include other semiconductors including InP and Si. Again the correlation of pressure data with mobility data and theoretical calculations is emphasized to give a strong, internally consistent argument for each material.

II. Band-Edge Deformation Potentials in GaAs and InP

If transition-metal impurity levels are to provide a valid reference level from which pressure-induced changes in band structure can be measured, then the transition-metal levels in a given material must possess a universal pressure derivative. This transition-metal pressure derivative must be independent of chemical species and charge state. This universality hypothesis was tested by measuring the pressure responses of several substitutional transition metals in GaAs. The choice of a III-V semiconductor as the test material over group IV materials is based on two points. First, the III-V semiconductors incorporate transition-metal impurities only substitutionally, which is necessary for the transition-metal impurity levels to behave as reference levels (vacancy-like defects). Transition metals in the group IV semiconductors can occur either substitutionally or interstitially. This introduces complications that remove group IV semiconductors as unambiguous test materials. This complication is not insurmountable, and the case of Si is discussed in section III of this chapter. Second, the relatively wide bandgaps in the III-V's, compared to Si or Ge, allow several different charge states of the same transition-metal impurity to produce defect levels in the bandgap. This point is important for the investigation of possible charge state dependences of the defect pressure derivatives. The

specific choice of GaAs over the other III-V semiconductors is based on the fact that the identification of isolated transition-metal impurity levels in GaAs has been more successful than in other III-V compounds because GaAs is available in higher purity, and because the technological importance of transition metals in GaAs has driven a wealth of research on this topic. This point is important for the investigation of possible chemical trends in the pressure derivatives.

A. Transition-Metal Pressure Derivatives

Transition-metal impurities in III-V compounds have been observed to occur only substitutionally and to occupy only the cation site, Ga in the case of GaAs. Because of the high occupancy of the transition-metal d-shell, these impurities can introduce many different charge-state levels into the bandgap. The charge states are conventionally labeled in ionic notation, with the 3+ charge state denoting the neutral defect after three electrons are used to satisfy the valence of the missing group III atom. In this notation, single acceptor levels are associated with the 3+/2+ charge-state transition, and single donor levels are associated with the 4+/3+ charge-state transitions. The d-electrons experience an environment with tetrahedral symmetry. This tetrahedral crystal field splits the d-orbitals into a triplet of t_2 symmetry above a doublet of e symmetry. These manifolds are occupied according to Hund's rules, i.e. high spin is favored, except in special circumstances [Caldas,1986]. The electronic structure of transition-metal impurities have been studied in detail through Green's function calculations [Zunger,1986].

1. Transition-Metal Impurity Levels

The properties of Ti and V transition-metal impurity levels in GaAs have been tested using deep level transient spectroscopy (DLTS). The energy levels

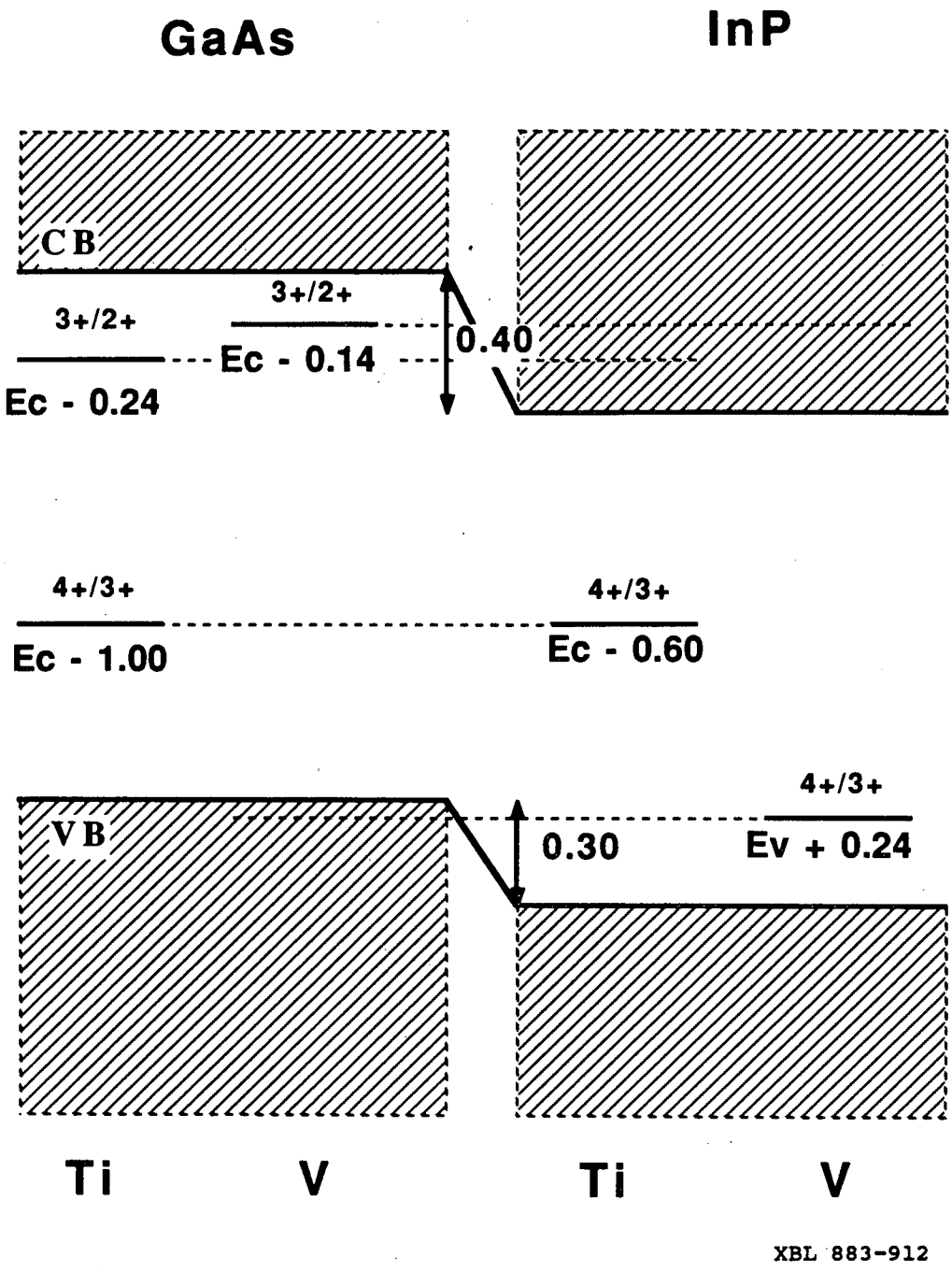


Fig. 5.2 Energy levels of Ti and V in GaAs and InP. The band-offset obtained by lining up the transition-metal levels should provide the band-offsets that would occur at a heterojunction interface between these two materials.

in the bandgap are shown in Fig.(5.2). All of the defect levels shown in GaAs act as electron traps and therefore their energies are referred to the conduction band enedgy. The Ti impurity introduces two defect levels[Hennel,1986], an acceptor at E_C - 0.24 eV and a donor at E_C - 1.00 eV. The V impurity has one acceptor level [Brandt,1986] at E_C - 0.14. These two chemical species include two different charge states with both t_2 and e state occupancies.

2. Uniaxial Stress

To study the stress properties of the transition-metal energy levels, compressive uniaxial stress was applied in conjunction with DLTS. Uniaxial stress contains both shear and hydrostatic components. To separate out the contribution from the hydrostatic components to the energy shifts it is necessary to analyze the symmetries of the initial and final states in the thermal ionization transition observed by DLTS. The electronic structure of the acceptor level of Ti is shown in Fig.(5.3). The initial charge state after the defect has trapped an electron during the filling pulse is Ti^{2+}. This corresponds to a $3d^2$ configuration with a two-electron atomic F-term. The F symmetry is split by the tetrahedral crystal field into manifolds which transform as the A_1, T_1 and T_2 representations of the tetrahedral group, with the A_1 state as the ground state. The final state in the thermal ionization involves the Ti^{3+} charge state with a $3d^1$ configuration and an electron in the conduction band. The one-electron D-term is split by the crystal field into T_2 and E manifolds. This crystal-field splitting of the $3d^1$ configuration is in fact the crystal-field splitting of one-electron theories. The ground state of the final state is the E state.

Under the application of uniaxial stress, the totally symmetric A_1 initial state experiences no contribution from the shear stress. Therefore all shear contributions arise in the final state from the shear dependences of the electron

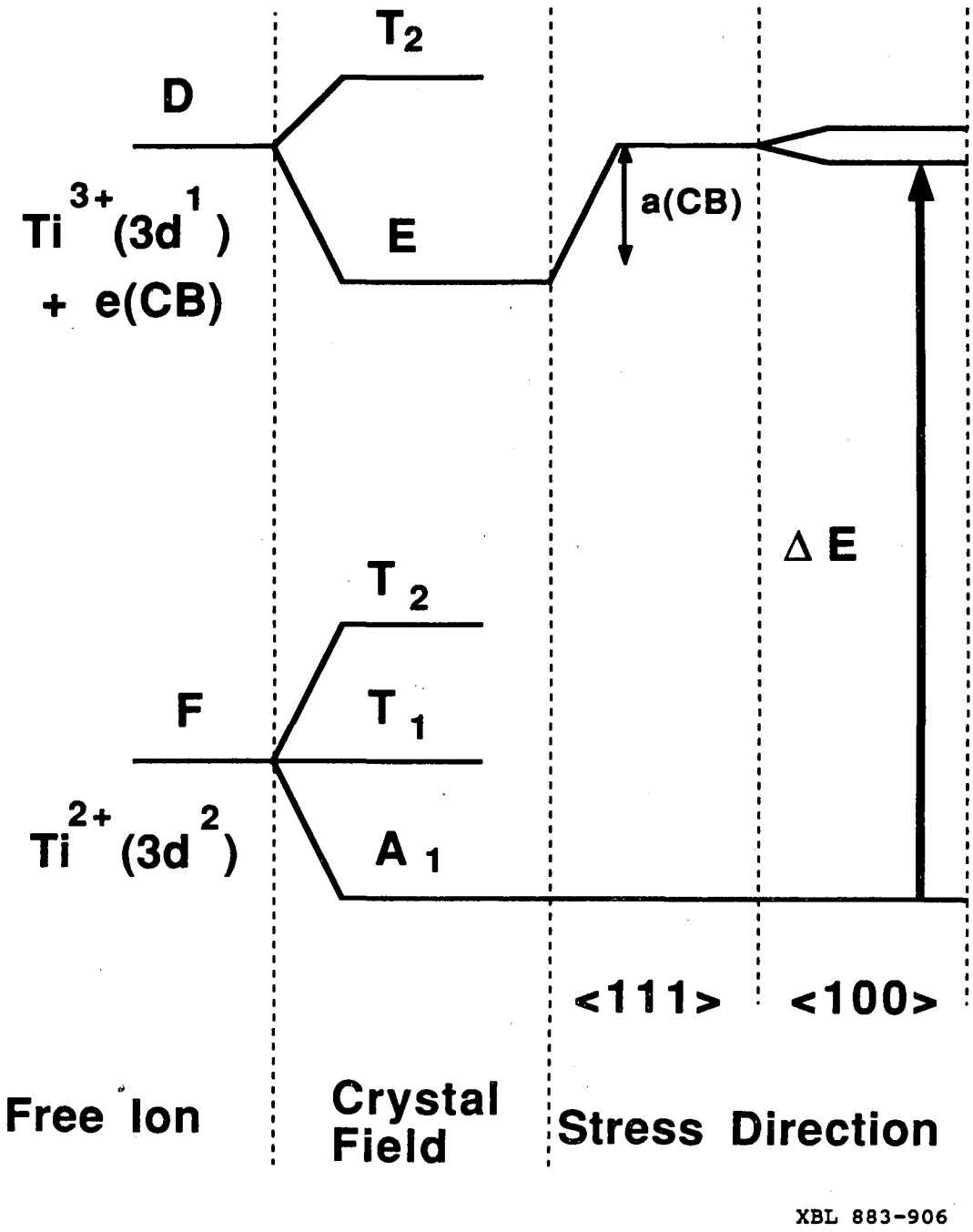


Fig. 5.3 Level diagram for Ti in GaAs. The initial and final state symmetries are shown along with the stress shifts and splittings appropriate for [111] and [100] oriented stresses.

in the conduction band and the E-state. For the case of GaAs, the conduction band minimum is non-degenerate and the energy shift of the conduction band is therefore also insensitive to shear stress. Therefore all shear contributions arise from the E symmetry of the Ti^{3+} charge state. Finally, E-symmetry is sensitive only to <100> oriented stresses. For stress oriented in the <111> directions there is no contribution from the shear components of the uniaxial stress to the energy shift of the defect level. For this case, the observed energy shift of the defect arises entirely from the hydrostatic component.

The Ti acceptor level is one of a small number of special cases in which <111> oriented stress removes the shear contribution to the defect energy shift. For general transition-metal defects, there will be a contribution from shear stress, even for <111> oriented stresses. Therefore, to perform hydrostatic experiments with uniaxial stress, the magnitude of the shear contributions must be known. This magnitude is measured easily for the case of the Ti acceptor by comparing the stress response of the defect for [111] oriented stress with the response for [100] oriented stress. The difference in the energy shifts between these two directions is from the shear deformation potential of the one-electron E-state. This shear contribution is roughly 5 meV/GPa [Hayes,1983] for transition-metal defects. This must be compared to the hydrostatic pressure derivative of the bandgap, which is 110 meV/GPa. Therefore the contributions from the shear component of the uniaxial stress is less than 5% of the hydrostatic shift in the worst case. To eliminate or reduce any shear component, samples are always stressed in the <111> directions. For the Ti(3+/2+) level the energy shift originates only from the hydrostatic component of the stress.

The stress data from the three transition-metal levels of Ti and V in GaAs in Fig.(5.2) are given in Fig.(5.4). These levels possess a universal pressure

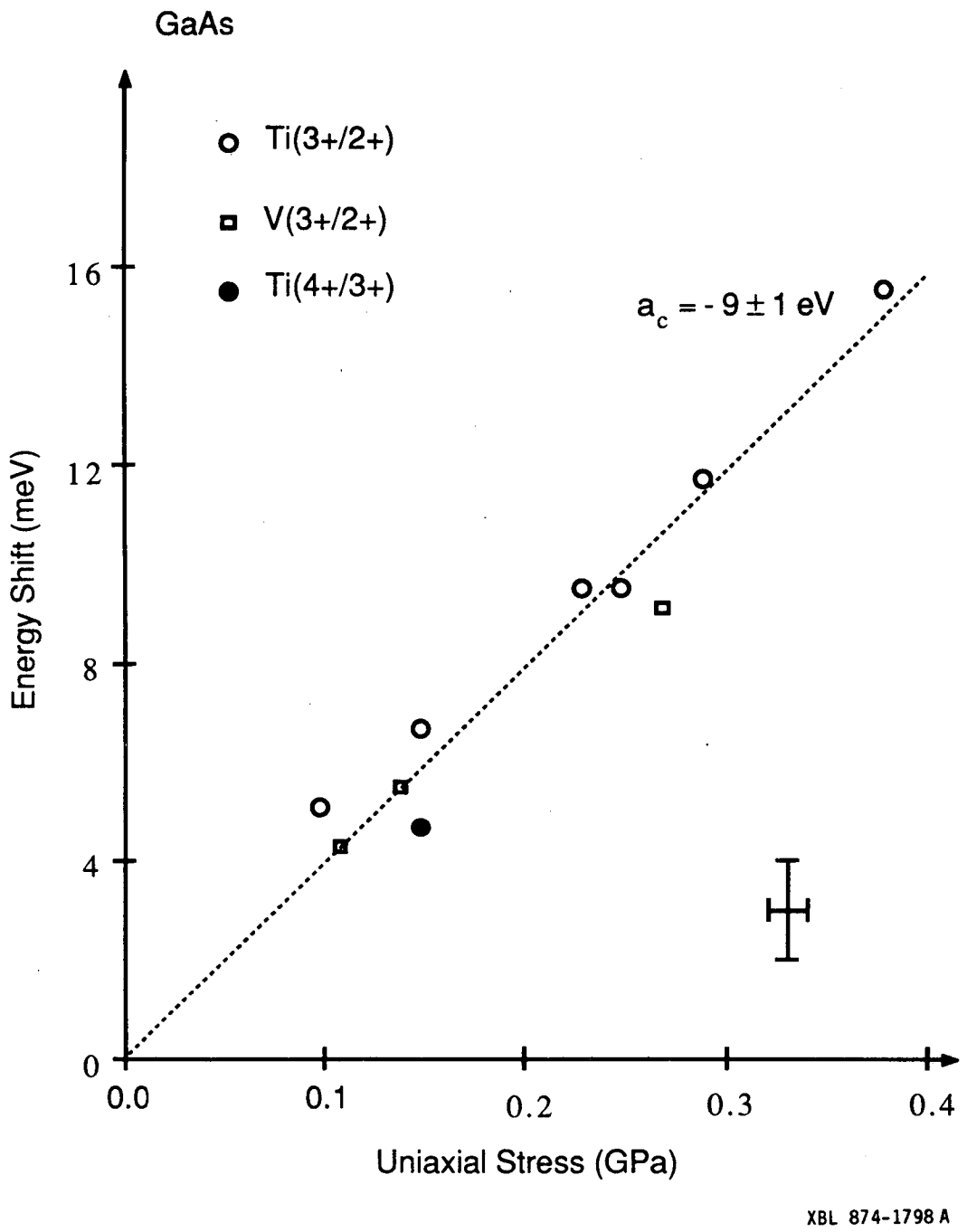


Fig. 5.4 Energy shifts of the DLTS peaks as a function of [111] oriented uniaxial stress. The corresponding deformation potential for the bottom of the conduction band in GaAs is -9 eV ± 1 eV.

derivative within a scatter of values. The best fit line gives a value for the conduction band deformation potential in GaAs of a_c = -9 eV ± 1 eV. By considering the bandgap deformation potential [Wolford,1985] of 8.6 eV, this gives the top of the valence band a deformation potential of a_v = -0.7 eV ± 1 eV. The levels of Ti and V in InP are given in Fig.(5.2). The pressure derivative of the Ti(4+/3+) level gives a deformation potential for the bottom of the conduction band of a_c = -7 eV ± 1 eV. The valence-band deformation potential is obtained to be a_v = -0.6 eV through subtraction of the bandgap deformation potential [Müller,1980] of 6.4 eV from the conduction-band value.

B. Acoustic-Phonon-Limited Mobility

The existing experimental values of the deformation potentials determined from the analyses of carrier transport or free carrier absorption are highly unreliable as they range from -7 eV to -17.5 eV in GaAs [Wolfe,1970; Lee,1978] and from -3.4 eV to -18 eV in InP [Takeda,1984; Nag,1978]. Such large variations of a_c lead to more than one order of magnitude spread in the relative contribution of deformation-potential scattering to the total scattering rate. The high values of the deformation potentials were in most cases obtained from fitting of the electron mobility [Lee,1978] or free carrier absorption [Pfeffer,1984] to theoretical models in which the effect of compensation was not accounted for. It has been shown recently [Look,1983] that the best fit to electron mobility in high-purity, low-compensation GaAs is obtained with a_c in the range -9 eV to -10 eV. The direct measurement of a_c is in excellent agreement with this result.

The introduction of AlGaAs/GaAs modulation doped heterostructures allowed for a substantial reduction of ionized impurity scattering, enhancing the relative contribution of acoustic phonons in limiting the low-temperature

electron mobility. However, even in this case the values of the deformation potential used to explain the temperature dependence of the electron mobility [Walukiewicz,1985; Mendez,1984] ranged from -7.0 eV to -13.5 eV. Lately, there is growing evidence based on new mobility calculations [Lei,1985] that the value of a_c = -8.5 eV satisfactorily accounts for the acoustic phonon deformation potential contribution to the electron scattering in n-GaAs/AlGaAs heterostructures. In addition, an analysis of the hole mobility in p-type GaAs/AlGaAs modulation doped heterostructures provided a value [Walukiewicz,1987] of a_v = -1.0 eV again in excellent agreement with the present determination.

Much fewer data are available for the deformation potentials of InP. The present determination of -7 eV is consistent with the values [Taguchi,1987; Walukiewicz,1980] a_c = -6.5 eV and -6.8 eV to explain electron mobility. Evidently, the measurements cannot be reconciled with such large values [Nag,1978] as -18 eV or such small values [Takeda,1984] as -3.4 eV for the conduction-band deformation potential reported in the literature. The acoustic-phonon scattering of the valence-band holes can be described in terms of the effective deformation potential

$$\Xi^2_{eff} = a_v^2 + (c_l/c_t)[b_v^2 + (1/2)d_v^2] \qquad (5.2)$$

where a_v, b_v, d_v are valence band deformation potentials and c_l and c_t are longitudinal and transverse elastic constants. The constants b_v and d_v are well known from uniaxial stress experiments [Wiley,1975]. Using the values of a_v, the values Ξ_{eff}(GaAs) = 6.1 eV and Ξ_{eff}(InP) = 6.6 eV are obtained which compare very favorably with the values 6.6 eV and 7.0 eV reported by Wiley.

It is evident from the above discussion that the values of the hydrostatic deformation potentials obtained from the transition-metal reference levels can explain consistently experimental data on electron and hole mobility in GaAs and InP. This provides additional support for the proposed method of directly measuring the band-edge deformation potentials using the transition-metal defect levels. It should be noted finally that the present experimental data confirm recent theoretical calculations which report values [Cardona,1987] of a_c(GaAs) = -8.8 eV and a_c(InP) = -5.9 eV. These results also agree with calculations [Van de Walle,1987a] that predict that a_v is typically an order of magnitude smaller than a_c. In conclusion, it has been shown, for the case of GaAs, that different transition-metal species among various charge states possess a universal pressure derivative within a scatter of values. This pressure derivative is in fact the pressure derivative of the band edge of the bulk material. The corresponding values for the band-edge deformation potentials are much more accurate than those determined by indirect methods, and give a consistent explanation of a variety of experimental data on electron and hole mobility in GaAs and InP. This technique is simple, and universal, and should be applicable to both III-V and II-VI semiconductors as well as group IV materials.

III. Extension to Silicon

The transition-metal reference level translates naturally from the III-V compounds to silicon. In practice, however, the application to silicon is complicated by several difficulties experienced in silicon, but not in most of the III-V compounds. First, most transition-metal impurities in silicon occupy interstitial sites and do not qualify as suitable reference levels, as will be shown later. The transition metals that do assume substitutional sites are heavy

elements with 4d or 4f electrons in the core, and it is not clear *a priori* how these inner-core electrons influence the electronic structure of these defects. The second difficulty arises from the fact that the pressure derivative of the indirect bandgap in silicon is an order of magnitude smaller than the pressure derivatives of the direct gaps in most of the III-V semiconductors. While this should present no fundamental difficulties to the theory, the ability to compare theoretical calculations of band-edge deformation potentials to experimental values is hampered by the current accuracy of the calculations. Despite these difficulties, sufficient experimental and theoretical evidence exists to define an absolute reference level in silicon from which defect as well as band properties can be measured.

Deep-level defects which are candidates for defining a reference level should share a common pressure derivative that is independent of charge state, chemical species and binding energy. This fact was established experimentally for the substitutional transition-metal impurities in GaAs and it was shown that the corresponding band-edge deformation potentials were consistent with mobility experiments and with theoretical calculations. The analogous system in Si are the substitutional transition elements Pd and Pt. The pressure derivatives of these isolated elements, as well as several complexes, have been measured previously [Stöffler,1986]. The acceptor level of both impurities (Pd(-/0) and Pt(-/0)) have equal pressure derivatives, which are similar to the pressure derivative of the Au acceptor level [Li,1985; Samara,1987]. On the other hand, the pressure derivatives of the levels assigned as Pt(0/+) and Au(0/+) have substantially different values. At first sight, this would cast doubt on the validity of using these heavy transition-metal elements as reference levels. However, additional defect pressure derivatives are available which can explain the discrepancy. Taking the value of the pressure derivative of Pd(-/0)

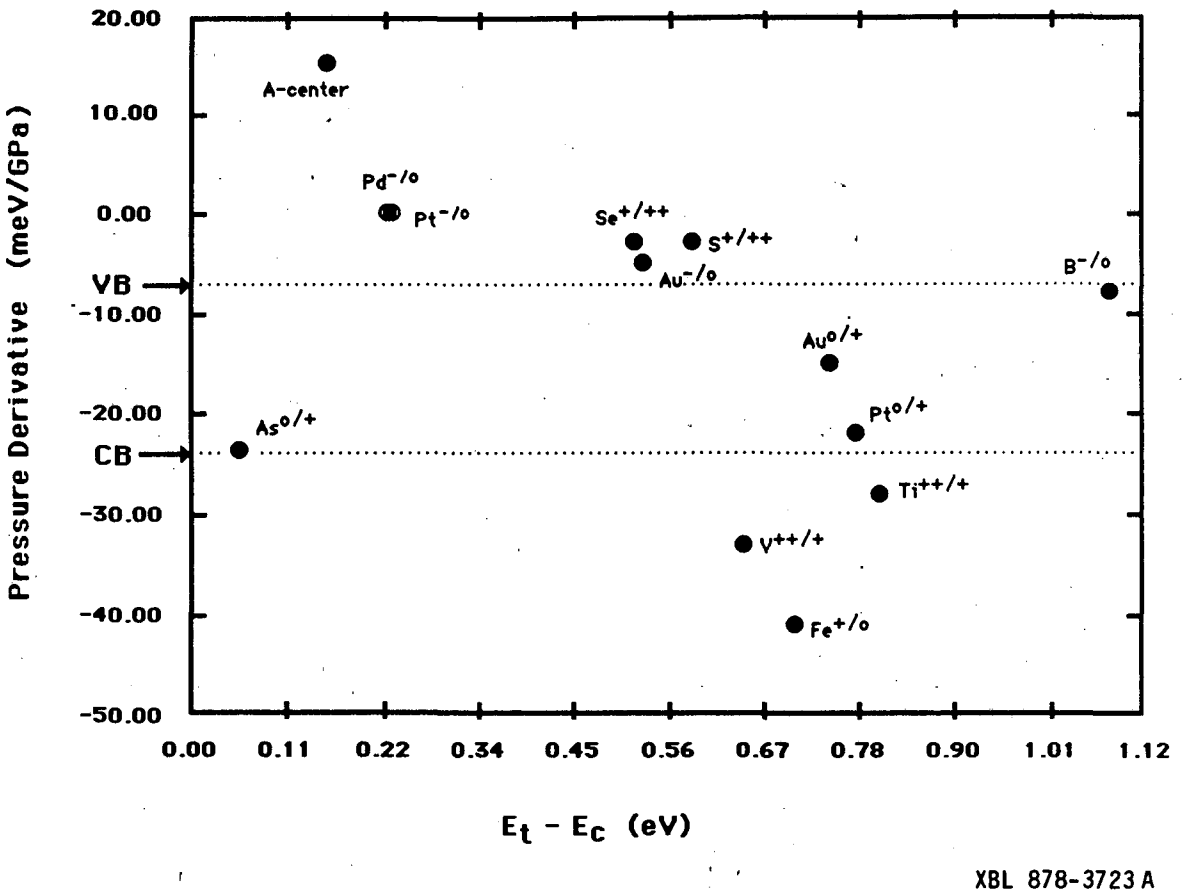


Fig. 5.5 Pressure derivatives of several defects in Si. The origin is taken as the pressure derivative of the Pt and Pd single acceptor levels. The pressure derivatives of the band edges are also included. Data are from Jantsch[1983], Stöffler[1986], Samara[1987a,1987b], and M. Li[1985].

and Pt(-/0) as the origin, the pressure derivatives of many deep-level defects in Si are plotted in Fig.(5.5) along with the pressure derivatives of the band edges. The defect values fall into three general classes: 1) shallow levels with pressure derivatives pinned to the nearest band edge; 2) interstitial transition metals with large negative pressure derivatives depending on the chemical species; and 3) deep substitutional impurities including transition metals and deep double donors which all have similar pressure derivatives. This third class of defects satisfies the criterion for providing an absolute reference level. The Pt level at Ev + 320 meV, which had previously been assigned as the substitutional Pt(0/+) donor, has a pressure derivative consistent with the interstitial transition metals. Considering the propensity of the transition metals in Si to occur interstitially, this Pt level is likely to arise from the interstitial configuration and therefore explains the discrepancy between the pressure derivatives of the two Pt levels. Pressure derivatives of transition-metal defects in silicon should therefore provide a criterion for distinguishing substitutional from interstitial impurities (or complexes).

The inclusion of the chalcogens Se and S in the class of substitutional impurities which define a reference level has not previously been suggested. Their structure would seem to be substantially different from the structure of substitutional transition metals which locks the transition-metal defect levels to the neutrality level. In fact, the levels arising from Te do not have pressure derivatives similar to those of Se and S. However, the chemical trend in the pressure derivatives of the chalcogens shows a remarkable feature: the pressure derivatives for decreasing atomic number Z approach the value defined by the substitutional transition metals. Such behavior may be reconcilable with theories of substitutional deep traps in covalent semiconductors [Hjalmarson,1980] which do predict asymptotic behavior for

strongly electronegative substitutional elements. These defects take on vacancy-like wavefunctions, as do the substitutional transition metals which are locked to the neutrality level. Therefore, the near-equality of the low Z chalcogen pressure derivative with the pressure derivative of substitutional Pd and Pt provides additional experimental evidence for the existence of a reference level in Si.

The deformation potentials of the band edges, measured with respect to the Pd and Pt levels, are $a_C = \Xi_d + (1/3)\Xi_u = +2.4$ eV and $a_V = +0.9$ eV. From the reported value [Laude,1971] of $\Xi_u = 8.6$ eV, the value $\Xi_d = -0.5$ eV is determined. These values can now be compared to theoretical calculations and mobility experiments to check for consistency. The most accurate calculation of the band-edge deformation potentials in Si comes from ab initio, self-consistent pseudopotential calculations [Van de Walle,1987b] of the valence band which yield a value $a_V = +0.8$ eV. The conduction band deformation potential is obtained by adding the deformation potential of the bandgap to this value. This gives $a_C = +2.3$ eV. This excellent agreement between the values from the transition metals and the calculated values is a further indication that the pressure derivatives of the substitutional transition metals in Si provide an accurate reference level, although such remarkable close agreement between experiment and theory is somewhat fortuitous since the error on the calculation is estimated to be ±0.5 eV.

A more important test of the values for the band-edge deformation potentials comes from mobility measurements. Acoustic-phonon-limited mobility is given by [Herring,1956]

$$\mu = (1/3)\, e\, [\, \tau_{\|} / m_{\|}^{*} + 2\, \tau_{\perp} / m_{\perp}^{*}\,], \qquad (5.3)$$

where

$$\tau_{\perp} = \tau_0 \, T^{-3/2} \, (E / k_B T)^{-1/2} \qquad (5.4)$$

and

$$\tau_0 = \frac{2\sqrt{2}\,\pi\,\hbar^4 c_l}{3\sqrt{m_{\perp}^{*2} m_{\parallel}^{*}}\, k_B^{3/2}\, \Xi_u^2 \left[1.33(\Xi_d/\Xi_u)^2 + 1.15(\Xi_d/\Xi_u) + 1.18\right]} \qquad . (5.5)$$

In this expression c_l is the longitudinal elastic constant, and E is the average electron energy. It has been found [Norton,1973] that a best fit to experimental data on electron mobility in silicon is obtained for $\tau_0 = 3.56 \times 10^{-9}$ sec $K^{3/2}$. Using eq.(5.5) with the empirical condition $\Xi_d + (1/3)\Xi_u = 2.4$ eV the values $\Xi_u = 9$ eV and $\Xi_d = -0.6$ eV are found. This value of Ξ_u compares very favorably with the reported value of 8.6 eV from the effects of uniaxial stress on indirect excitons [Laude,1971]. Therefore the determination of Ξ_d using the transition-metal pressure derivatives is consistent with the electron mobility data, and is certainly more accurate than previous values obtained from mobility data alone.

The effective deformation potential describing scattering of holes by acoustic phonons is

$$\Xi_{eff} = a_v^2 + (c_l/c_t)[b_v^2 + d_v^2/2] \, , \qquad (5.6)$$

where c_t is the transverse elastic constant. The deformation potentials $b_v = -2.2$ eV and $d_v = -5.1$ eV are well established [Merle,1978]. Using the value of $a_v = -0.9$ eV one obtains $\Xi_{eff} = 7.3$ eV which again is very close to the value

of 7.5 eV found from hole mobility analysis in p-type silicon [Wiley,1975]. Therefore, the values of the hydrostatic band-edge deformation potentials in silicon determined from substitutional transition-metal pressure derivatives have been found to provide a basis for a consistent explanation of electron and hole mobilities. In conclusion, it has been established that strong evidence exists for a pressure reference level in silicon based on heavy substitutional transition metals as well as on low-Z chalcogens. From this reference level the first accurate value for Ξ_d is provided which is consistent with the best mobility data.

Chapter VI. Isotropic Defect-Lattice Coupling in Semiconductors

I. Introduction

The total energy of a defect in a semiconductor crystal is a function of the lattice positions of the atoms surrounding the defect. The equilibrium lattice positions are determined by the minimum energy configuration. The lattice positions in this configuration are usually distorted from the lattice positions of the perfect crystal in the absence of the defect. Different charge states of the same defect may have different equilibrium lattice configurations. Therefore after the emission or capture of a carrier, the lattice coordinates around the defect will relax to the new equilibrium configuration. Jahn-Teller systems are examples of lattice relaxation that is driven by the spontaneous breaking of degeneracy arising from defect symmetry. Considerable research, both theoretical and experimental, has been conducted on Jahn-Teller systems [Engleman,1972]. Of primary interest in the Jahn-Teller distortion is the Jahn-Teller coupling parameter describing the change in defect electronic energy with changes in the lattice coordinates around the defect. In the case of symmetry-lowering distortions, this axial coupling parameter can be measured by observing the splitting of the ground state. The measurement involves usually small differences between energy levels and can be measured accurately.

In addition to axial, symmetry-lowering coupling to the lattice, defects can also couple isotropically to the lattice. This coupling involves no energy

splitting. Rather, the defect ground-state energy is lowered by a symmetric inward or outward breathing-mode relaxation. To measure a shift in energy, a reference must be defined from which to relate the change in the ground-state energy. Before the definition of such a reference energy, isotropic defect-lattice coupling parameters had not been accessible to experiment, and estimates of the coupling were restricted to the realm of theory [Scheffler,1982; Lindefelt,1983; Lindefelt,1984]. Despite the difficulty in measuring isotropic defect-lattice coupling, breathing-mode lattice relaxation can contribute to scientifically and technologically interesting phenomena such as multiphonon capture [Henry,1977] and metastability (as for the EL2 and DX defects in GaAs and AlGaAs respectively).

Defect energy levels in the bandgap of a semiconductor are defined by the energies required to change the charge states of the defects and promote carriers to the conduction or valence band edge. These activation energies change under applied pressure because the pressure affects the band edge energies as well as the defect energies. If the band-edge deformation potential can be determined, then the effect of pressure on the defect alone can be isolated and related to the problem of lattice relaxation. The topic of the change in the band-edge energy under pressure has been discussed in chapter V. In the present chapter the physical basis for the use of transition-metal defects (or more generally the use of the average sp^3 hybrid energy of a semiconductor) as an absolute reference level is developed. The magnitudes of chemical trends in the pressure derivatives of transition-metal defect levels, which would give the limits of validity for using these levels as a reference, are estimated. The dependence of the deformation potentials of defects on the localization of the defect wavefunction is discussed in section III. Two extremes of behavior can be recognized between defects with highly localized wavefunctions that couple

strongly to the their local environment, and defects with weakly localized wavefunctions for which the local band extrema contribute significantly to the defect deformation potentials and the local coupling of the defect energy to the lattice is negligible. In section IV breathing-mode relaxations around defects are discussed and a relationship is established between changes in local force constants and changes in entropy which occur upon carrier emission. The results of this section are applied to calculate volume relaxations around defects in GaAs, most notably the EL2 defect in GaAs.

II. Transition-Metal Pressure Derivatives

It has been demonstrated in the last chapter [Nolte,1987c,1988] that transition-metal defect levels possess a universal pressure derivative. The band-edge deformation potentials derived from this pressure derivative agree well with mobility measurements [English,1987; Mendez,1985] and with recent theoretical calculations [Cardona,1987; Van de Walle,1987b]. This pressure derivative can therefore be used as a pressure derivative reference level from which the absolute pressure derivatives of other defects can be measured. Although the evidence for this pressure derivative reference level is compelling, the physical basis for the universal pressure derivative was not discussed. The answers to these questions are found in the relationship between the structure of deep-level defects and the band structure of the host crystal.

A. Transition-Metal Defect Structure

Substitutional transition-metal defects in semiconductors can be visualized as transition-metal atoms occupying simple vacancies [Watkins,1983]. The basic validity of this viewpoint has been borne out by many different theoretical techniques ranging from tight-binding models [Picoli,1984] to first principle

Green's function calculations [Zunger,1986]. First-principles Green's function calculations have been the most successful theoretical technique for describing transition-metal defect structure. Though these calculations are computationally intensive, a simple descriptive model has been developed [Zunger,1986] based on the accurate Green's function results. This is the three-level model of the transition metals. In this model, the conduction bands are condensed into a single level whose energy coincides with the energy of the maximum density of states in the conduction band, E^*_c, about 3 eV above the top of the valence band in GaAs. Likewise, the valence band is represented as a single level that occurs at the energy of the maximum density of states in the valence band, E^*_v, roughly 2 eV below the top of the valence band. The transition-metal atom introduces an atomic level as the third level in the three-level model whose energy shifts as a function of chemical species. The transition-metal levels interact with the conduction and valence bands equally through an interaction matrix element V_d. The mixing of levels produces a bonding combination that is resonant with the valence bands, a non-bonding combination that produces localized levels in the bandgap, and an antibonding combination that is resonant with the conduction bands. These levels are shown in Fig.(6.1). A value of 2 eV for V_d reproduces many of the trends deduced from the accurate Green's function calculations.

The condensation of the conduction and valence bands into single levels at E^*_c and E^*_v respectively to represent the effect of the band structure on the electronic properties of these deep must be justified before the three-level model can be used for any predictive purposes. The energy separation between E^*_v and E^*_c can be identified with the Jones' zone gap. The Jones' zone gap is the fundamental bandgap (in the extended zone scheme) that opens up at the Fermi level. The magnitude of the bandgap is roughly constant

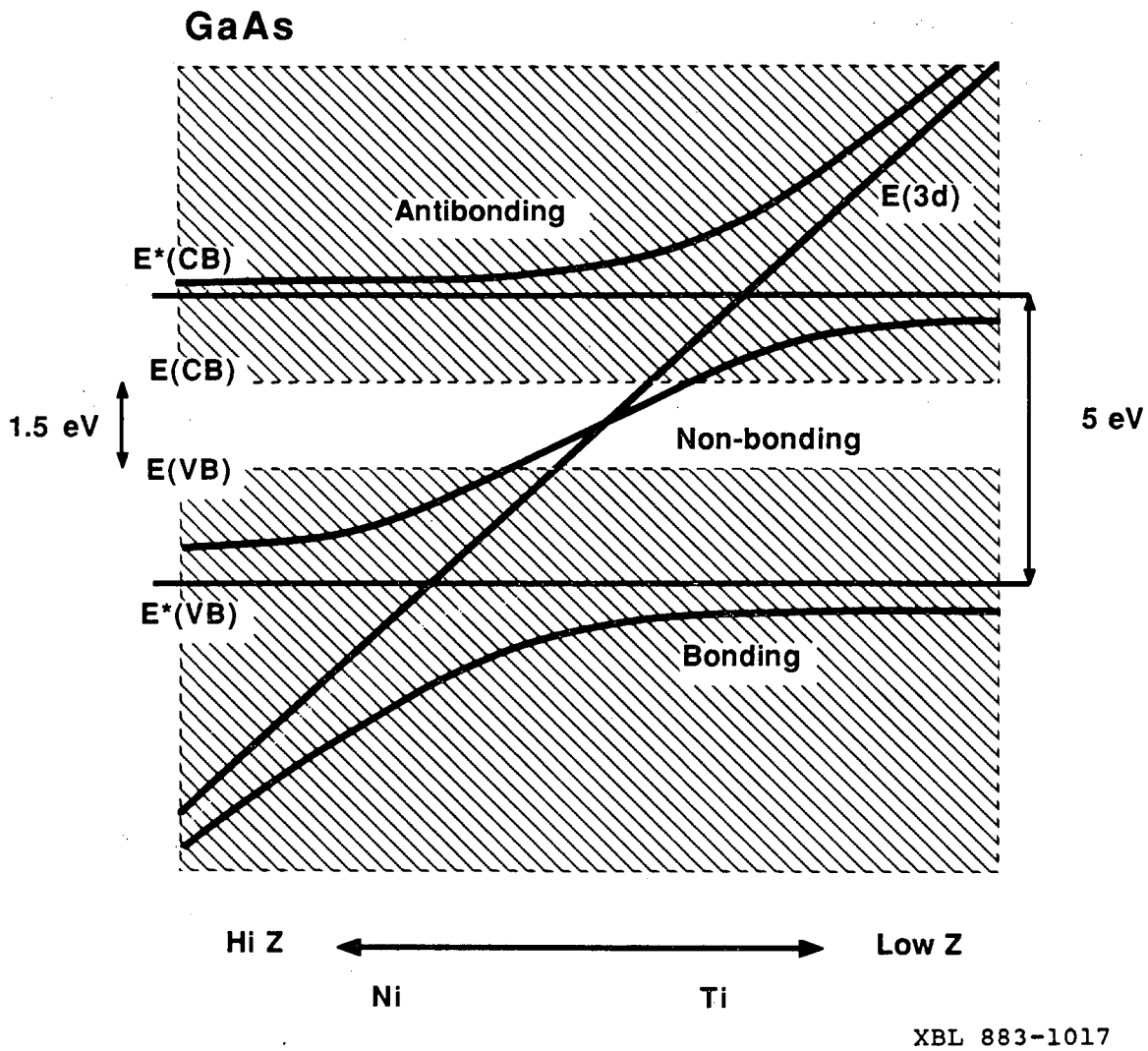


Fig. 6.1 Three-level model of transition metals.

among isovalent semiconductors, and in the III-V compounds is about 5 eV. Because of the two-atom basis in the unit cell of the zinc-blende structure, this fundamental gap is folded back into the first Brillouin zone in the reduced zone scheme. The weight of the density of states in a given band is carried by the edge of the Jones' zone gap. This gap is often referred to as the dielectric, or optical, gap because transitions between the edges of this gap have large oscillator strengths and therefore dominate the frequency-dependent dielectric response of the materials. This optical gap is not to be confused with the fundamental bandgap which is the gap over which thermal excitation of carriers dominates the electrical properties of the semiconductor.

The Jones' zone gap can also be identified as a result of the tight-binding picture of the band structure. In the tight-binding model, the sp^3 hybrids of individual atoms overlap with nearest neighbors, producing bonding and anti-bonding combinations. The periodicity of the lattice further spreads out these levels into the valence and conduction bands respectively. The initial energy separation between the bonding and anti-bonding combinations of the sp^3 hybrids is roughly 5 eV, i. e., the Jones' zone gap. With the removal of nearest-neighbor interactions, as in a vacancy, the states reduce to localized dangling sp^3 hybrids. The Jones' zone is therefore seen to be directly associated with the sp^3 hybrid energies of the semiconductor atoms. This dominance of the gross features of the band-structure on the electronic structure of deep substitutional impurities is general and applies to impurities other than transition metals. It has been shown previously [Nolte,1987e] that even the deep chalcogen levels in Si and the isoelectronic N level in GaAs [Nolte,1988] (which is actually resonant with the direct conduction band) are insensitive to the proximity of local band extrema, and have pressure derivatives that behave in much the same manner as the transition metals.

B. Limits of Universality

The question concerning the pressure derivatives of the transition metals remains to be addressed. Experimental data have been given previously (chapter V) describing the similarity between the pressure derivatives of Ti, V and Ni. To give a qualitative model of the chemical trends in the pressure derivatives of the transition metals, the three-level model is used. This model is already known to produce qualitatively correct chemical trends in the transition-metal binding energies. In this model, the pressure responses of various band extrema are ignored, and the effect of the pressure dependences of the band structure on the defect energy levels is contained entirely in the pressure response of the Jones' zone gap. This gap has been measured in GaAs under pressure [Zallen,1967] to increase by 60 meV/GPa. For convenience, the center of the Jones' zone gap is taken as the origin. The importance and validity of this choice of origin will be discussed later. With this choice, the energy E^*_c increases by 30 meV/GPa and the energy E^*_v decreases symmetrically by 30 meV/GPa. The lattice dependence of the interaction matrix element V_d is assumed to vary as an inverse power of the lattice spacing d;

$$V_d \propto d^{-n} . \qquad (6.1)$$

For illustration, the exponents n = 0 and 2 are considered. In the case when n=0, there is no dependence of V_d on the lattice. The choice of an exponent of 2 is motivated by suggestions from tight-binding arguments [Harrison,1980]. The three-level model is solved for zero stress and 1 GPa. The resulting pressure derivatives of the non-bonding states in the Jones' zone gap are shown in Fig.(6.2) where the pressure derivative of the center of the Jones'

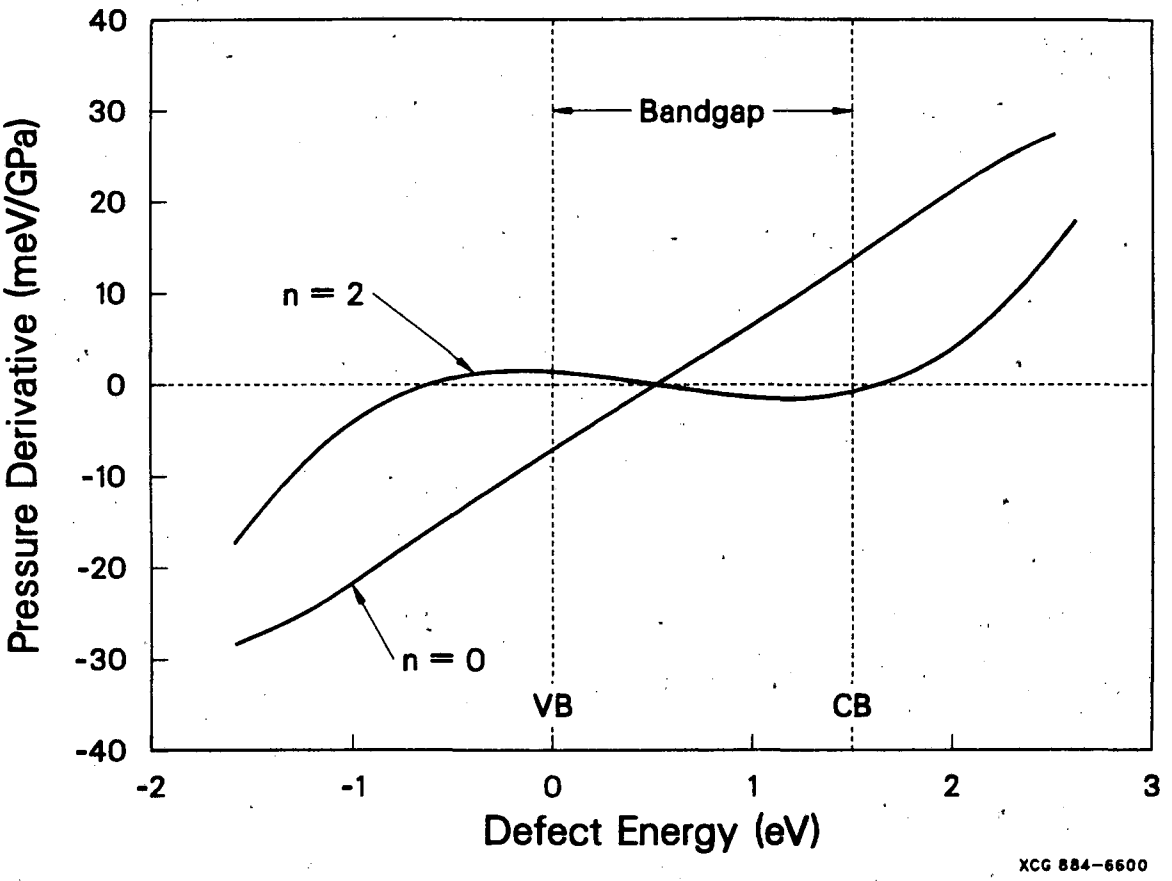


Fig. 6.2 Pressure derivatives from the three-level model.

zone is taken as the zero. For the case of n = 0, the pressure dependences of the non-bonding states are constrained by the pressure derivatives of the edges of the Jones' zone gap with limiting values of ± 30 meV/GPa. For states near the center of the Jones' zone gap, these effects roughly cancel. The fundamental bandgap covers a relatively small range of energies near the center of the Jones' zone gap, so the transition-metal defects all have relatively small pressure derivatives within this range. When the values of n = 2 is used, the lattice dependence of V_d can cancel or even overpower the effects of the band structure. For all cases, the pressure effects of either the band structure or the interaction between the transition metal and the lattice roughly cancel for defects near the center of the Jones' zone gap. This crude model may therefore help to explain the universal pressure derivative of the transition-metal defects. The dependence on the specific model is not crucial. Simple tight-binding calculations that include the s and p orbitals of the transition metal also give values for the defect hydrostatic deformation potentials that are under 0.77 eV. These simple models (which neglect electron-electron interaction) all conclude that the pressure derivatives of the transition-metal defects are similar to the pressure derivative of the center of the Jones' zone gap.

III. Defect Hydrostatic Deformation Potential Equation

The total energy of a defect depends on the lattice positions of the neighboring atoms. If stress alters the local lattice positions, the defect energy will respond. Therefore, stress perturbation can give a good measure of the defect-lattice coupling through the defect deformation potential. However, the situation is not this easy in practice. An example of this are shallow donors in GaAs. These defects have large hydrostatic deformation potentials, yet do not couple to the local lattice. For the more general case of deep-level defects, it is

crucial to separate out the effects of the band structure on the defect deformation potential from the effects of local lattice coupling. This can be done qualitatively by deriving the expression for the defect deformation potential.

In the one-electron approximation, the Hamiltonian governing the defect electron consists of a kinetic energy term T, and a potential energy term that involves the potentials of all the atoms in some neighborhood of the origin of the defect:

$$H = \Sigma_m \; V(\mathbf{a}_m) + T \,, \qquad (6.2)$$

where $V(\mathbf{a}_m)$ is the potential of the crystal with the defect and $\mathbf{a}_m$ is the position vector of the m^{th} neighbor to the defect origin. Under the application of external stress, the position vectors are altered such that the new Hamiltonian is

$$H = \Sigma_m \; V(\mathbf{a}_m + \Delta\mathbf{a}_m) + T \,, \qquad (6.3)$$

where $\Delta\mathbf{a}_m$ is the change of the position vector due to the strain. The deformed Hamiltonian can be expanded to lowest order to give

$$H = \Sigma_m \; V(\mathbf{a}_m) + \Sigma_m \; (\Delta\mathbf{a}_m \cdot \nabla) V(\mathbf{a}_m) + T \,, \qquad (6.4)$$

where the $V(\mathbf{a}_m)$ are functions of position and the differential operator is a position operator. Clearly, if the strain is small, the term $\Sigma_m \; (\Delta\mathbf{a}_m \cdot \nabla) V(\mathbf{a}_m)$ can be taken as a perturbation describing the response of the electronic energy to applied stress. The defect deformation-potential operator can now be defined in terms of the crystal potential:

$$\Xi_{ij} = \Sigma_m a_{jm} \nabla_i V(\mathbf{a}_m) . \qquad (6.5)$$

The index m refers to the m^{th} neighbor, and the indices i and j are the coordinates.

To investigate the effect of the perturbation Hamiltonian on the defect wavefunction, first-order perturbation theory can be applied for either degenerate or non-degenerate defect states. The first-order matrix elements of the perturbation Hamiltonian acting on the unperturbed defect ground state $\Psi(r)$ are:

$$dE = \Sigma_m \int \Psi^*(r) [(\Delta\mathbf{a}_m \cdot \nabla) V(\mathbf{a}_m)] \Psi(r) d^3r . \qquad (6.6)$$

This expression is general, but is far from transparent for explaining how the magnitudes of defect deformation potentials vary with the details of the binding mechanism and with the radius of the defect wavefunction. A more obvious expression can be derived through partial integration, in which the differential position operator is transposed to operate on the defect wavefunction rather than on the potential term. The new expression for the energy matrix elements is

$$dE = \Sigma_m \Delta\mathbf{a}_m \cdot [-\int \nabla\Psi^*(r) V(\mathbf{a}_m) \Psi(r) d^3r] + C. C. \qquad (6.7)$$

where the integrated term evaluated over the infinite surface has vanished because the wavefunction decays exponentially far from the defect center.

The effect of defect wavefunction localization can now be investigated by considering the product of the defect envelope wavefunction with its gradient $\Psi(r)\nabla\Psi(r)$. When this product is large at the sites $\mathbf{a}_m$ of neighboring potentials,

the contribution to the defect deformation potential from this charge state will be large. This product will be maximum at a site a_m when the envelope wavefunction radius is approximately equal to the neighbor distance: $r \approx a_m$. This can be demonstrated explicitly by choosing a 1s envelope function and placing a delta function potential a distance d from the defect origin. The delta function potential is taken as a perturbation and represents the contribution of the local lattice sites to the defect energy. The change in the defect energy due to the effects of pressure on the site of the delta function is proportional to the contribution of the coupling of the local lattice to the defect deformation potential. The relative deformation potentials as a function of the radius a_0 of the defect wavefunction are shown in Fig.(6.3) for d = 1, 1.5, and 2 Å. The striking results of this analysis are: first, that the deformation potential decreases roughly exponentially for a given d when the defect radius a_0 is much larger or much smaller than d; and second, that the maximum defect deformation potential decreases roughly exponentially as the perturbation potential distance d increases.

The first point concerning the dependence of the defect deformation potential on the localization radius of the defect wavefunction has important consequences for comparing the deformation potentials of shallow levels, which are weakly localized, with the deformation potentials of deep levels, which are strongly localized. If the lattice sites that contribute strongly to the defect binding energy are near neighbor to the center of the defect, then those sites will contribute strongly to the defect deformation potential only if the defect wavefunction radius is approximately equal to the near neighbor separation. Because the deformation potential depends exponentially on the defect wavefunction radius, the deformation potential rapidly vanishes as the wavefunction radius increases. In the limit of the shallow level defects, the local

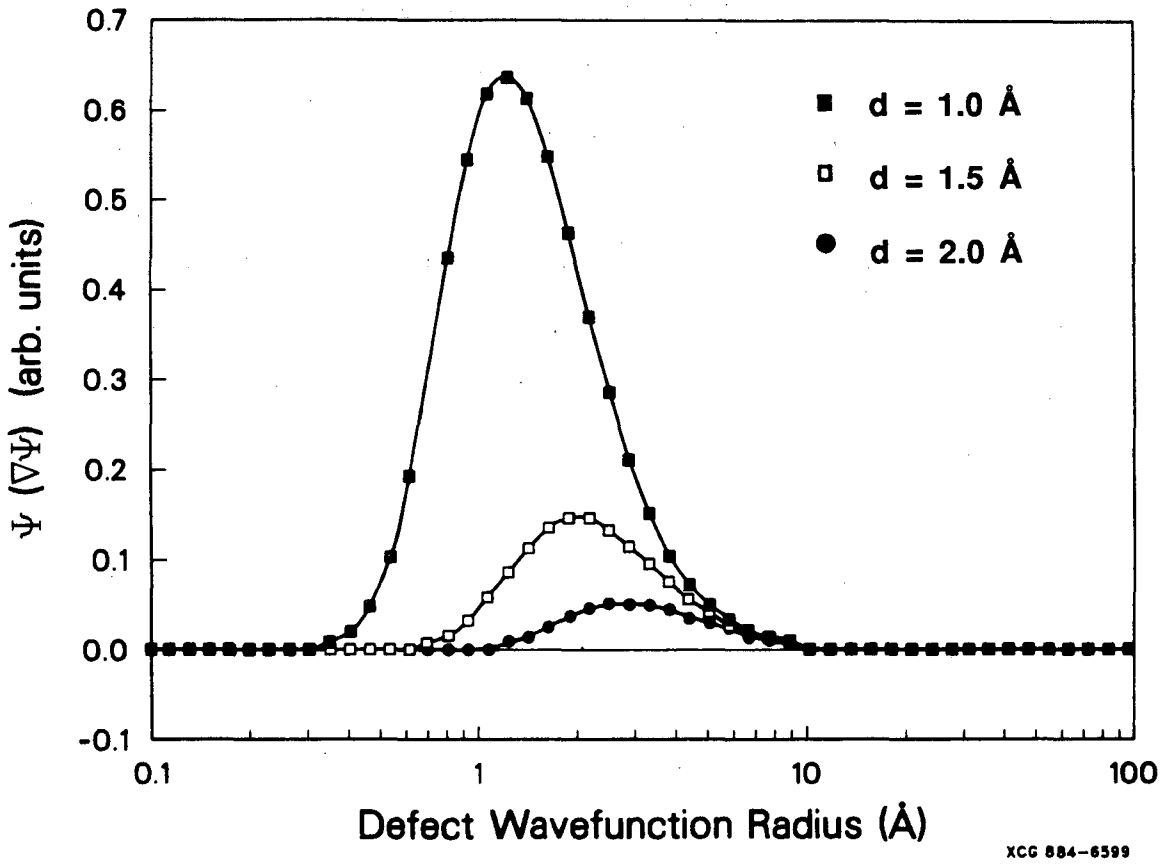


Fig. 6.3 The magnitude of $\Psi\nabla\Psi$ as a function of wavefunction radii evaluated at distances of 1Å, 1.5Å and 2Å from the defect center.

contribution to the defect deformation potential vanishes. The second point concerning the dependence of the maximum deformation potential on the distance of a contributing potential from the defect center is important for comparing the deformation potentials of different deep levels. The smaller the deep level wavefunction radius is, the potentially larger the deformation potential can be. It must be reemphasized that only a single charge state of a defect is being considered in the above analysis. The experimentally measured deformation potential is the difference between the deformation potentials of two different charge states. In fact the transition metal defects provide a good example of strongly localized defects that have a zero net deformation potential. The coupling of a specific transition metal charge state to the lattice can be strong, yet the coupling remains constant as the charge state is changed.

The localization of the defect envelope wavefunction is not the only contribution to the defect-lattice coupling. The band structure of the crystal also influences the deformation potential. To include the effects of band structure on the defect deformation potential the defect wavefunction can be described in a perturbative approach as a linear combination of Bloch states of the perfect crystal

$$\Psi(r) = \Sigma_{n,k}\, b_{n,k}(r)\, \phi_{n,k}(r)\,, \qquad (6.8)$$

where $\phi_{n,k}(r)$ are Bloch states, and the indices n and k refer to the band and k-vector. The coefficients $b_{n,k}(r)$ describe the localization of the defect in real space. Using this wavefunction in eq.(6.7) and considering only diagonal terms in the Hamiltonian yields

$$dE \approx \Sigma_m \Delta a_m \{ \int V(a_m) \Sigma_{n,k} |\phi_{n,k}(r)|^2 \quad [b_{n,k}(r)] \nabla [b^*_{n,k}(r)] \, d^3r$$
$$+ \int V(a_m) \Sigma_{n,k} | b_{n,k}(r)|^2 \quad [\phi_{n,k}(r)] \nabla [\phi_{n,k}(r)] \, d^3r \}$$
$$+ \text{ C. C.} \qquad (6.9)$$

The first term on the right hand side involves the gradient of the envelope functions, and therefore is the contribution to the defect energy from the local interaction of the defect with its immediate environment. The second term involves the gradient of the Bloch functions and is the contribution to the defect energy from the band structure. This expression is approximate, and the neglect of off-diagonal components has not been justified (though the expression *is* correct for effective-mass defects for which the summation over n and k is dropped). However, if it is taken as a crude approximation of the effects of stress on a general defect, then the origin of various contributions to the defect energy shift can be roughly identified by rewriting it qualitatively as

$$\Xi_{defect} \approx \Xi_{envelope} + \Sigma_{n,k} a_{n,k} \Xi_{n,k} \quad , \qquad (6.10)$$

where the first term on the right-hand side is the contribution from the local interaction between the defect electronic energy and the neighbor ions around the defect, and the second term is the contribution from the band structure. The role of these two terms can be understood in two opposite limits: the limit of shallow effective-mass levels; and deep substitutional defects (such as transition-metal impurities).

In the case of an effective-mass level, the sum over n and k in the second term of eq.(6.9) is restricted to a single band extremum. The second term therefore reduces to a value related to the band-edge deformation potential.

The contribution from the first term is negligible in the case of an effective-mass defect because the envelope wavefunction is too delocalized to have an appreciable amplitude on the nearest-neighbor lattice sites, and therefore the local coupling of the defect to the lattice vanishes. Shallow-level defects therefore have deformation potentials dominated by the deformation potentials of the band edge, and do not couple strongly to the local lattice.

The other extreme is the case of deep substitutional defects. For these defects, the conduction band and the valence band contribute approximately equally to the electronic structure. If the two-level model of the crystal band structure is adopted (i.e. the Jones' zone gap) then the contributions roughly cancel. Therefore, for deep-level defects, the deformation potential is dominated by the interaction of the defect energy with the near-neighbor lattice sites. An important example to consider are the transition-metal defects as well as other vacancy-related defects. These defects have wavefunctions that extend over several near-neighbor lattice positions, and by eq.(6.9) should therefore have strong coupling to the lattice. Yet precisely these defects define the absolute zero of the defect deformation potentials. This apparent contradiction can be resolved by considering two important points concerning lattice coupling and its experimental measurement. First, the defect -lattice coupling can have different signs for different shells of neighbor atoms: one shell might move inwards while the next shell moves outwards. These lattice displacements decrease rapidly as distance from the defect center increases, heavily weighting the effects of the first shell. But the net effect of lattice distortion on the defect may be relatively small despite a large coupling to a particular shell. Second, defect deformation potentials are the result of the *difference* in lattice coupling between two different charge states. In the case of the transition metals, the universal pressure derivative is only the statement that

all charge states of all chemical species have the same lattice coupling. This coupling to the lattice can in fact be non-zero. On the other hand, some deep-level defects exhibit strong differences in the coupling to the lattice between two charge states. For these defects the local coupling of the defect to the lattice contributes predominantly to the defect deformation potential.

Some defects fall between the two extremes of effective-mass defects and strongly localized defects. For these more general defects it is important to consider the effects of the band structure on the lattice coupling. This contribution to the measured defect deformation potential does not influence the lattice relaxation around the defect. This caution extends mainly to deep effective-mass-like defects for which the extended Coulomb potential may contribute as much or more to the electron binding energy than the strong central-cell potential. For many of the vacancy-related or intrinsic defects, on the other hand, band-structure effects are likely to cancel and the measured defect deformation potentials can be related to breathing-mode relaxations around the defects.

IV. Breathing-Mode Relaxation around Defects

The proportionality constant between the change in defect energy and the change in the lattice positions around the defect in the electron-lattice coupling parameter of a defect. This electron-lattice coupling has the units of an energy gradient or force, and is in fact the force acting on the atoms neighboring the defects. The neighboring atoms respond to this force by relaxing to new equilibrium positions determined by the balance of electronic energy and elastic energy. The elastic energy is supplied by the work performed during the relaxation on the local bonds. The relationship between the electron-lattice

coupling and lattice relaxation for a defect in a given charge state can be described in general terms by the simple force equation for a spring

$$F = k\,x\,. \tag{6.11}$$

In ionization spectroscopy, single charge states cannot be measured separately (as they can in EPR), rather, differences between the initial and final charge states are measured. The total differential of eq.(6.11) is therefore the relevant equation for ionization spectroscopy

$$dF = k\,dx + x\,dk\,. \tag{6.12}$$

For the case of a defect coupling isotropically to a surrounding elastic medium, the expression becomes

$$\Xi = B\,\Delta V + V\,\Delta B\,, \tag{6.13}$$

where Ξ is the isotropic deformation potential describing the change in electron-lattice coupling between the initial and final state, B is the local bulk modulus around the defect, and V is the volume occupied by the defect. The first term on the right-hand side of the equation corresponds to the elastic energy change associated with volume relaxation upon carrier emission. The second term corresponds to the change in the local force constants upon carrier emission. Solving eq.(6.13) for the change in volume yields

$$\Delta V = \frac{\Xi}{B} - V\frac{\Delta B}{B} \quad , \qquad (6.14)$$

which is the volume relaxation of a defect associated with the emission of a carrier.

This expression was derived from simple microscopic arguments relating the defect electronic energy gradient to the elastic forces around the defect. An equivalent expression for volume relaxation upon carrier emission has been derived based entirely on thermodynamic arguments [Samara,1986]. This is done by taking into account that the pressure derivative of a Gibb's free energy at constant temperature is equal to a volume change

$$(d\Delta G/dp)_T = \Delta V \; . \qquad (6.15)$$

The Gibb's free energy for the thermal emission of carriers from deep-level defects is given by

$$\Delta G = \Delta H - T\,\Delta S \; , \qquad (6.16)$$

where ΔH is the activation enthalpy for carrier emission, T is the temperature, and ΔS is the change in entropy upon carrier emission. By combining eq.(6.16) with eq.(6.15) and solving for the change in volume yields

$$\Delta V = \frac{d\Delta H}{dp} - T\frac{d\Delta S}{dp} \quad . \qquad (6.17)$$

By comparing eq.(6.14) (which is based on microscopic arguments) with this equation (based on purely thermodynamic arguments) it is found that the first term on the right side of each equation is identical, by definition. This allows for the identification of the second term in eq.(6.17) involving the change in entropy with the term in eq.(6.14) involving the relative change in the local force constants, namely

$$V\,\Delta B/B = T\,d\Delta S/dp. \qquad (6.18)$$

Therefore, by measuring the pressure dependence of the activation enthalpy and entropy of a deep-level defect, the volume relaxation around the defect and the relative change in local force constant upon carrier emission can be calculated.

Two complications arise in the calculation of the volume relaxation: the first involves the the contribution of the band-edge deformation potential to the measured activation enthalpy; the second concerns the problem that local force constants may differ from the force constants of the bulk material. The activation enthalpy for a defect is defined as the difference in total energy between the initial and final state

$$\Delta H = E(2,Q_2) + E(e_{band}) - E(1,Q_1)\,, \qquad (6.19)$$

where $E(2,Q_2)$ and $E(1,Q_1)$ are the total energies of the defect in the final and initial states at the equilibrium configurations Q_2 and Q_1 respectively. The final state includes a carrier at the band edge with an energy $E(e_{band})$. Usually the band energy is taken as the origin, which causes no difficulty in the interpretation of activation energies. As perturbations are applied, however, the

perturbations affect the band-edge energies discussed in chapter V as well as the energies of the different defect charge states. Under pressure, the activation varies as

$$d\Delta H/dp = \{ dE(2,Q_2)/dp - dE(1,Q_1)/dp \} + dE(e_{band})/dp \, . \qquad (6.20)$$

In this case the effect of applied pressure on the band edge enters on equal footing with the effect of pressure on the defect states alone. The quantity in brackets is the absolute pressure derivative of the defect, which is proportional to the defect deformation potential. The last term in eq.(6.20) is the band-edge pressure derivative. To measure the absolute pressure derivative of a defect, the band-edge pressure derivative must be known. It was shown in chapter V that band-edge pressure derivatives can be measured with respect to transition-metal defect levels in semiconductors, and therefore this last term no longer presents a difficulty for measuring the electron-lattice coupling of deep-level defects.

The second complication that arises in the measurement and calculation of lattice couplings is the fact that the local force constants around a defect can be different than the force constants in the bulk material. The measured defect deformation potential is given by

$$\Xi_{defect} = \{ dE(2,Q_2)/dp - dE(1,Q_1)/dp \}/B \, , \qquad (6.21)$$

where B is the bulk modulus of the crystal. However, if the local force constants around the defect are different from those in the bulk material, then the measured deformation potential Ξ will differ from the "true" deformation potential Ξ' by

$$\Xi' = (B'/B)\,\Xi\,, \qquad (6.22)$$

where B' is the local bulk modulus at the defect. Eq.(6.14) for the calculation of the volume relaxation around a defect on carrier emission was derived from a microscopic model relating defect electronic and elastic energies. In this equation, the terms on the right hand side should involve the "true" deformation potential and the local bulk modulus

$$\Delta V = \frac{\Xi'}{B'} - V\frac{\Delta B'}{B'}\,, \qquad (6.23)$$

rather than the values that are measured externally. This possible difficulty in calculating the volume relaxation around the defect is eliminated by the fact that the effect of the local force constants affect B and Ξ equally, and therefore

$$\Xi'/B' = \Xi/B\,. \qquad (6.24)$$

These effects also cancel for the second term in eq.(6.14), and eq.(6.14) reduces to eq.(6.17). The important conclusion here is that the calculation of the volume relaxation around the defect on carrier emission is insensitive to the local force constant. This is not an accident, because eq.(6.17) has also been derived from thermodynamic principles which are general and must be insensitive to the microscopic details of the problem.

V. Deep-Level Defects

Having developed the relationship between absolute pressure derivatives and volume relaxation around defects upon carrier emission, the pressure derivative data for several defects in GaAs can now be presented and the corresponding lattice relaxations calculated. The absolute pressure derivatives of several defects in GaAs are presented in Fig.(6.4). The origin on the vertical axis is defined by the pressure derivatives of the Ti(3+/2+) transition-metal levels. Included in the figure are the pressure derivatives of the important band extrema, as well as the shallow acceptor and donor levels which are pinned to the valence- and conduction-band edges. In the case of the shallow donor levels, the large absolute pressure derivative does not arise from strong coupling to the local lattice, but from the coupling of the extended conduction-band states to the lattice via the band-edge deformation potential. This distinction between the contributions to the defect deformation potential from local coupling and the coupling of the band states was made in eq.(6.10) of section III. The envelope wavefunctions for shallow levels are too delocalized to contribute to the defect deformation potential, the only contribution coming from the nearest band edge. This distinction is an important one to make because the large pressure derivative of the shallow donors, if used blindly in eq.(6.15) would suggest a large volume relaxation upon electron emission. This is clearly not the case for the shallow donors, and some caution must be taken when calculating the volume relaxation of effective-mass-like deep levels. These defects may be dominated in their pressure response by a nearby band extremum.

For localized defects, the proximity of band extrema to a defect level can be unimportant to the pressure response of the defects. This is especially evident in the case of deep defects that produce energy levels resonant with the direct

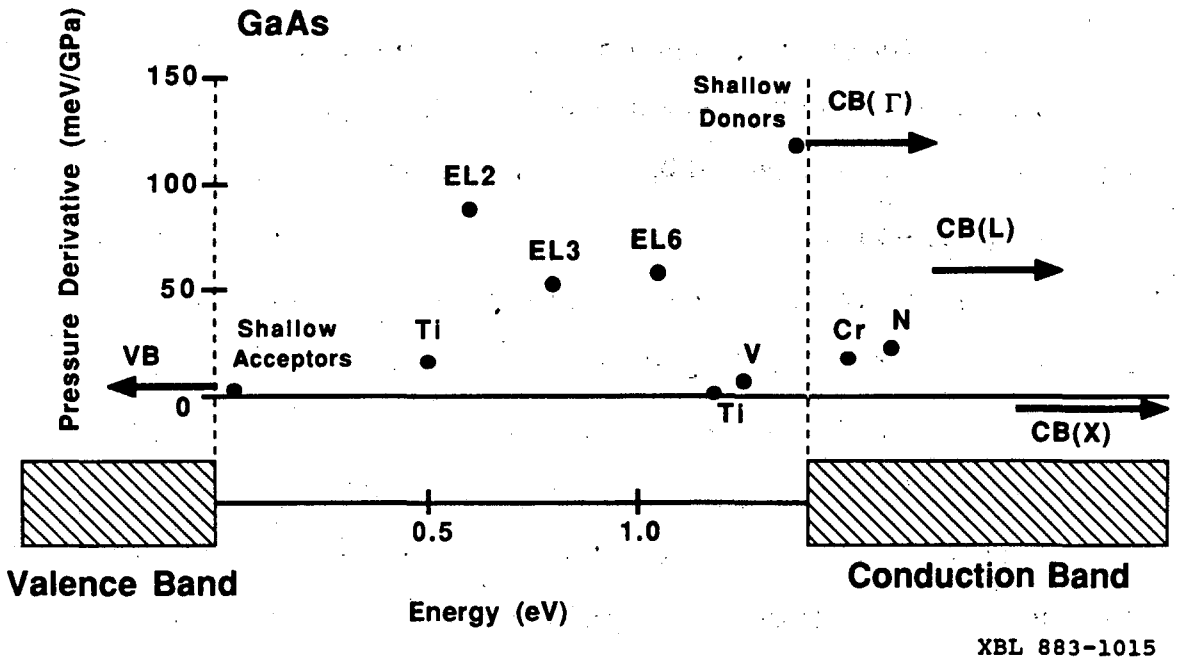


Fig. 6.4 Absolute pressure derivatives of defects in GaAs. Data are from Bastide[1980], Wolford[1985], Hennel[1982] and Nolte[1987c,1987d].

conduction band in GaAs. Despite their proximity to the conduction band, the pressure derivatives of two resonant levels, the Cr(2+/1+) transition-metal level [Hennel,1982] and the nitrogen isoelectronic level [Wolford,1985], fall in the range of small pressure derivatives shared by the deep transition-metal levels. These isolated substitutional impurities do not couple strongly to the lattice, nor do local band extrema affect their pressure response. For these defects, the electronic structure is controlled by the gross features of the conduction and valence bands equally.

In contrast to the isolated substitutional defects, the deep grown-in defects EL2, EL3 and EL6 do couple strongly to the lattice. The especially large coupling of the EL2 defect is intriguing in light of the interesting metastable properties that this defect exhibits. The volume relaxation around a defect upon carrier emission, as calculated from eq.(6.14), is the result of changes in the electron-lattice coupling as well as changes in the local force constants. It has been demonstrated [Samara,1987a,1987b] in Si that the contribution from the change in the local force constants to the pressure derivatives of defect activation energies is only several meV per GPa. This contribution is negligible compared to the large pressure derivatives for these grown-in defects, so the volume relaxation can be calculated from the pressure derivatives alone.

Because of the scientific and technical interest in the EL2 defect, the lattice coupling of this specific defect is discussed next. With the definition of a convenient reference level for measuring absolute pressure derivatives, published data on the pressure derivatives of several different transitions associated with EL2 can be reviewed and used to study the isotropic electron-

lattice coupling parameters of the defect. The pressure derivative of EL2 has two main features: it is uncharacteristically large, and it is non-linear. The magnitude of the pressure derivative leads to an extraordinarily large inward lattice relaxation of 23% upon electron emission. Such a large lattice relaxation seems to be incompatible with the relatively small relaxation (~6% volume) calculated for the isolated As-antisite defect [Scheffler,1984]. The non-linearity has been observed directly [Dobaczewski,1987] in the pressure dependence of the thermal ionization transition of DLTS, but was also inferred from the pressure-dependent Frank-Condon shift [Zylbersztejn,1978].

Three transitions for the EL2 defect have been measured under pressure. The pressure derivatives of these transitions are [White,1977; Nolte,1987d]

$$\begin{aligned}
&\text{I:} && \partial E_+/\partial p\,\big|_{R_0} + \partial E_{cb}/\partial p\,\big|_{R_0} - \partial E_0/\partial p\,\big|_{R_0} = +12 \pm 10 \text{ meV/GPa}, \\
&\text{II:} && \partial E_+/\partial p\,\big|_{R_+} + \partial E_{cb}/\partial p\,\big|_{R_+} - \partial E_0/\partial p\,\big|_{R_0} = +30 \pm 10 \text{ meV/GPa}, \\
&\text{III:} && \partial E_0/\partial p\,\big|_{R_+} - \partial E_{vb}/\partial p\,\big|_{R_+} - \partial E_+/\partial p\,\big|_{R_+} = +17 \pm 10 \text{ meV/GPa}, \qquad (6.25)
\end{aligned}$$

where the pressure derivatives of the separate charge states are evaluated at the relevant equilibrium lattice positions. If the lattice-coupling were linear, then the transition energies from transition I and III would add to the pressure derivative of the bandgap of 110 meV/GPa. Actually, the sum falls short by a factor of four. This has caused some confusion in the literature, in which it has been stated that the EL2 defect tracks with both the valence band and the conduction band.

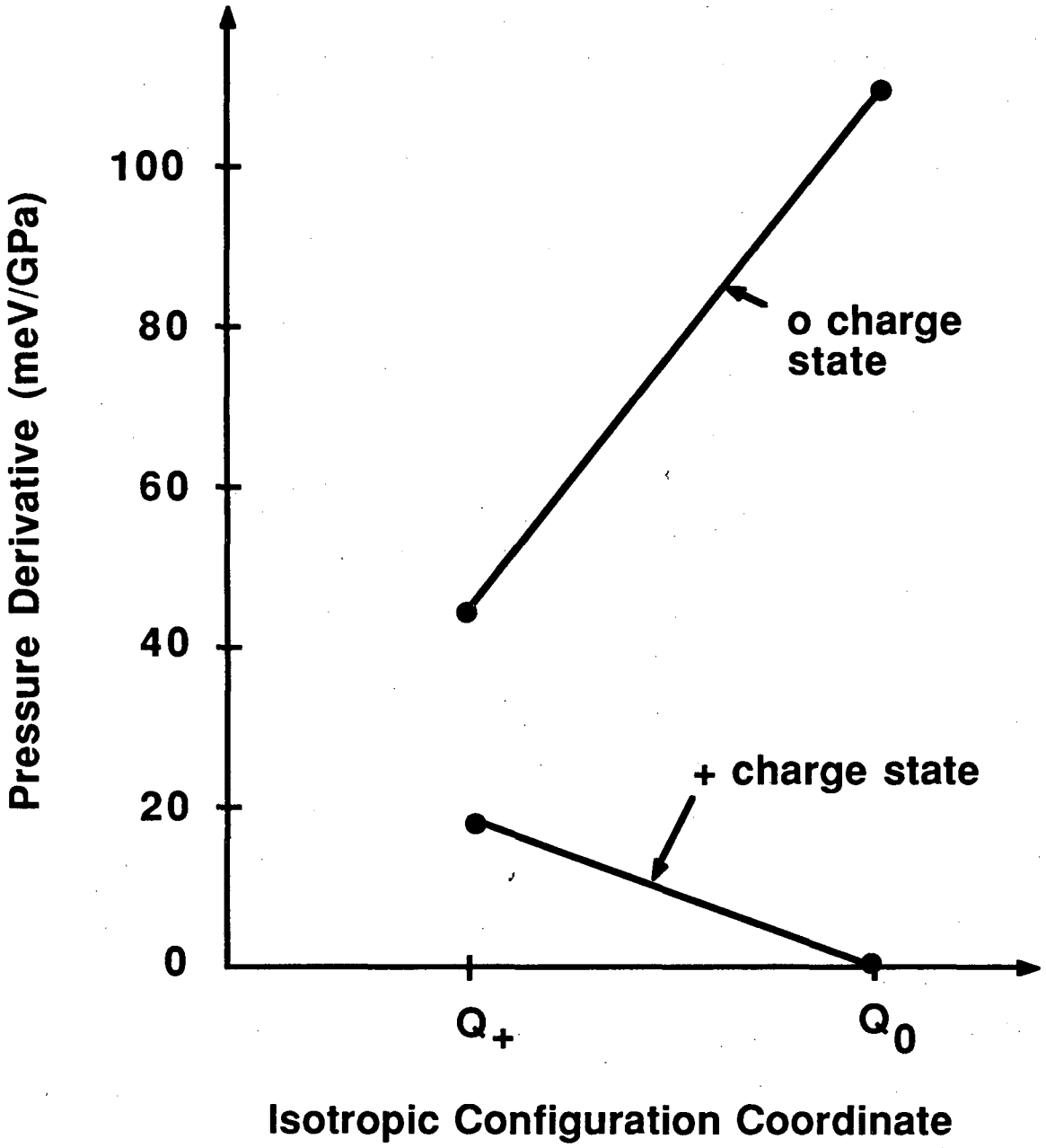


Fig. 6.5 Relative pressure derivatives for the occupied and unoccupied charge states of EL2. The non-linear coupling is predominantly in the neutral charge state.

To lowest order, the non-linearity of the lattice coupling of the neutral and charged charge states can be included in the pressure derivatives as

$$\partial E_0/\partial p = A_0 + B_0 (R - R_0) ,$$
$$\partial E_+/\partial p = A_+ + B_+ (R - R_0) , \qquad (6.26)$$

for which the pressure derivatives are a function of lattice coordinate. Eqs.(6.25) can now be solved for the lattice coupling parameters A_0, A_+, B_0 and B_+:

$$B_+ = - 90 \text{ meV/GPaÅ} \pm 75 \text{ meV/GPaÅ} ,$$
$$B_0 = + 315 \text{ meV/GPaÅ} \pm 75 \text{ meV/GPaÅ} ,$$
$$A_+ - A_0 = -108 \text{ meV/GPa} \pm 15 \text{ meV/GPa} . \qquad (6.27)$$

The relative pressure derivatives of the two charge states are plotted in Fig.(6.5). The non-linear lattice coupling occurs primarily in the neutral charge state. This strong non-linearity in the coupling of the neutral state may have implications for understanding the metastable properties of the defect, because it is the neutral state that is transformed into the metastable state through optical quenching.

V. Conclusions

Isotropic lattice relaxation has been one of the defect properties most difficult to measure experimentally. The lack of a reference from which to measure absolute shifts had restricted this study primarily to theory. The pioneering work of Samara and Barnes on the thermodynamic relationship

between pressure derivatives and volume relaxations provided the first estimates of volume relaxations around deep-level defects upon carrier emission. However, the lack of a well-defined origin left broad limits for the magnitudes and even signs of the deep-level volume relaxations. With the work on transition-metal-based reference levels discussed in this thesis, an accurate estimate for the absolute origin of pressure derivatives has been provided. With this origin, accurate calculations of the volume relaxations of deep-level defects in GaAs are now possible.

Before calculating the volume relaxations of defects based on the pressure derivatives of substitutional transition-metal defects, it was first neccessary to explain the universal pressure derivative of the transition metals from physical arguments. From simple models it was found that that the well-established vacancy-like structure of the transition metals locks them to the center of the Jones' zone gap. This gap carries the weight of the density of states of the bonding and anti-bonding combinations of the sp3 hybrids of the semiconductor atoms. The center of the Jones' zone gap therefore plays the role of the average vacancy energy in the crystal. The effects from the lattice dependence of the interaction of the d-shell electrons with the vacancy dangling bonds, as well as from the lattice dependence of the bonding and antibonding separation, roughly cancel for the defect states near the center of the Jones' zone gap. Transition-metal defects therefore follow the average sp^3 energy of the crystal.

Defect structure is influenced by local coupling of the defect electronic energy to the lattice, as well as by the band structure of the crystal. Both of these effects determine the defect deformation potential. Volume relaxation, on the other hand, is only a function of the local coupling to the lattice. To separate out the effects of the band structure from the local lattice-coupling, a general expression was derived for the defect hydrostatic deformation potential. One

result of this derivation was the relationship between the magnitude of the local coupling to the lattice with the localization of the defect envelope wavefunction. For envelope wavefunctions that are either too extended or too localized, the magnitude of $\Psi\nabla\Psi$ is too small on the near neighbors to allow for strong coupling to these lattice sites. Another other important result urges caution to be used when considering the volume relaxations of defects that may have moderate or strong effective-mass-like structure. For these defects, the local band structure may contribute as much or more to the defect deformation potential than the local coupling to the lattice.

The thermodynamic connection between pressure derivatives and volume relaxation is a macroscopic relationship that does not involve the microscopic origin of the defect deformation potential. From simple considerations of the force constants around the defect and energy gradients, the enthalpy and entropy pressure derivatives were identified with the local deformation potential and local force constants respectively. It was also shown that the calculation of volume relaxation is independent of the local forces constants. With the establishment of a sound experimental and theoretical basis for the calculation of defect volume relaxations, several defects in GaAs were considered. The important defect EL2 in GaAs was found to have an extraordinarily large isotropic coupling to the lattice, with an associated inward volume relaxation of roughly 23% upon electron emission. Furthermore, the large non-linear electron-lattice coupling of the defect was found to be predominantly in the neutral state. These aspects may have important implications for understanding the metastable properties of the defect.

Several problems remain unsolved concerning the isotropic coupling of defects to the lattice in semiconductors. One difficulty in the calculation of defect volume relaxation is the problem of precisely defining a relaxed volume.

Different shells of near neighbors can experience large distortions with different signs. It remains to define clearly how the measured volume relaxation is related to these multiple distortions. Perhaps the most important problem is the experimental measurement of isotropic lattice relaxation in a single charge state. All defect characterization techniques that rely on ionization transitions can only measure changes in lattice-coupling parameters and local force constants between two different charge states. These techniques cannot measure absolute shifts in the separate charge states. Charge-state specific techniques, such as EPR or ENDOR, can provide some information about near-neighbor lattice positions. But these techniques cannot separate the effects of local force constants and electron-lattice coupling. In view of the simplification of many outstanding problems that has followed the introduction of transition-metal reference levels, it may yet be possible to find an analogous reference from which absolute shifts in single charge states can be measured.

Bibliography:

Alves, J. L. A. and J. R. Leite, Phys. Rev. B**34**, 7174(1986)

Aspnes, J. D. E. and M. Cardona, Phys. Rev. B**17**, 726(1978)

Balslev, I. and P Lawaetz, Phys. Lett. **19**, 6(1965)

Baraff, G. A., E. O. Kane and M. Schlüter, Phys. Rev. B**21**, 3563(1980); Phys. Rev. B**21**, 5662(1980)

Bardeen, J. and W. Schockly, Phys. Rev. **80**, 72 (1950)

Barnes, C. E., G. A. Samara, L. R. Dawson, and T. E. Zipperian, Inst. Phys. Conf. Ser. **65**, 493 (1982)

Bastide, G., G. Sagnes, and C. Merlet, Rev. Phys. Appl. **15**, 1517 (1980)

Benton, J. L., K. M. Lee, P. E. Freeland, and L. C. Kimerling, AIME, Thirteenth International Conference on Defects in Semiconductors, vol. 14a:647 (1984)

Bois, D. and G. Vincent, J. Phys. Lett.**38**, L351(1977)

Borsuk, J. A. and R. M. Swanson, IEEE Trans. Electron. Devices **ED-27**, 2217(1980)

Brandt, C. D., A. M. Hennel, L. M. Pawlowicz, F. P. Dabkowski, J. Lagowski and H. C. Gatos, Appl. Phys. Lett.**47**, 607(1985)

Broeckx, J. and J. Vinnik, Phys. Rev. B**35**, 6165(1987)

Caldas, M. J., A. Fazzio, and A. Zunger, Appl. Phys. Lett. **45**, 671 (1984)

Camphausen, D. L., G. A. N. Connel and W. Paul, Phys. Rev. Lett.**26**, 184(1971)

Capasso, F., Ann. Rev. Mat. Sci. **16**, 263(1986)

Capasso, F., Science **235**, 172(1987)

Cardona, M. and N. E. Christensen, Phys. Rev. B**35**, 6182(1987)

Chantre, A. and D. Bois, Phys. Rev. B **31**, 7979 (1985)

Chen, J. W. and A. G. Milnes, Annu. Rev. Mater. Sci. **10**, 157(1980)

Clerjaud, B., J. Phys. C**18**, 3615(1985)

Collins, T. C., D. J. Stukel and R. N. Euwema, Phys. Rev. B**1**, 724(1970)

DeLeo, G. G., G. D. Watkins and W. B. Fowler, Phys. Rev. B**23**, 1851(1981)

DeLeo, G. G., G. D. Watkins and W. B. Fowler, Phys. Rev. B**25**, 4962(1982a)

DeLeo, G. G., G. D. Watkins and W. B. Fowler, Phys. Rev. B**25**, 4972(1982b)

DeLeo, G. G., W. B. Fowler and G. D. Watkins, Phys. Rev. B**29**, 3193(1984)

Devaud, B., G. Picoli, B. Lambert and G. Martinez, Phys. Rev. B**29**, 5749(1984)

Dobaczewski, L. and A. Sienkiewicz, Act. Phys. Polon.**A71**,341(1987)

Dresselhaus, G., A. F. Kip and C. Kittel, Phys. Rev. **98**, 368(1955)

Engelman, R., The Jahn-Teller Effect in Molecules and Crystals (London: Wiley-Interscience, 1972)

English, J. H., A. C. Gossard, H. L. Störmer and K. W. Baldwin, Appl. Phys. Lett.**50**, 1826(1987)

Farmer, J. W., C. D. Lamp and J. M. Meese, Appl. Phys. Lett.**41**, 1063(1982)

Farmer, J. W., J. M. Meese, P. M. Henry and C. D. Lamp, AIME, Thirteenth International Conference on Defects in Semiconductors, vol. 14a:639,(1985)

Flores, F. and C. Tejedor, J. Phys. C**12**, 731(1979)

Gehlhof, W. and K.H. Segsa, phys. stat. sol.(b) **115**, 443(1983)

Haller, E. E., M. R. Heuschen and P. L. Richards, Appl. Phys. Lett.**34**, 495(1979)

Haller, E. E., B. Joos and L. M. Falicov, Phys. Rev. B**21**, 4729(1980)

Harrison, W., Electronic Structure [San Francisco: W. H. Freeman and Co, 1980]

Hasegawa, H., Phys. Rev. **129**, 1029(1963)

Hayes, W., J. F. Ryan, C. L. West and P. J. Dean, J. Phys. C**12**, L815(1979)

Hayes, W., J. F. Ryan, C. L. West and P. J. Dean, J. Phys. C**13**, L149(1980)

Hemstreet, L. A., Phys. Rev. B**15**, 834(1977)

Hemstreet, L. A., Phys. Rev. B**22**, 4590(1980)

Hennel, A. M., C. D. Brandt, Y. T. Wu, T. Bryskiewicz, K. Y. Ko, J. Lagowski, and H. C. Gatos, Phys. Rev. **B33**, 7353(1986)

Hennel, A. M. and G. Martinez, Phys. Rev. B**25**, 1039(1982)

Henry, C. H. and D. V. Lang, Phys. Rev. B**15**, 989(1977)

Henry, P. M., J. W. Farmer, and J. M. Meese, Appl. Phys. Lett. **45**, 454(1984)

Hensel, J. C. and G. Feher, Phys. Rev. **129**, 1041(1963)

Herring, C. and E. Vogt, Phys. Rev.**101**, 944(1956)

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

Huang, K. and A. Rhys, Proc. R. Soc. A**204**, 406(1950)

Jantsch, W., K. Wünstel, O. Kumagai and P. Vogl, Physica **117**B, 188(1983)

Kahn, J. M., L. M. Falicov and E. E. Haller, Phys. Rev. Lett.**57**,2077(1986)

Kaplyanskii, A.A., Opt. Spectro. (USSR) **16**, 329(1964)

Kelso, S. M., Phys. Rev. B**25**, 1116(1982)

Kimerling, L.C. and J.L. Benton, in Defects in Semiconductors, 1982 [Physica(Utrecht) 116B, 297(1983)]

Krebs, J. J. and G. H. Stauss, Phys. Rev. B**16**, 971(1977)

Lang, D. V., J. Appl. Phys.**45**, 3023(1974)

Lang, D. V., Thermally Stimulated Relaxation in Solids, ed. , Springer-Verlag (1979) pp.109-134

Lang, D. V., R. A. Logan and M. Jaros, Phys. Rev. B**19**, 1015(1979)

Langer, J. M. and H. Heinrich, Phys. Rev. Lett. **55**, 1414(1985)

Lannoo, M., J. Bourgoin, Point Defects in Semiconductors I, Springer-Verlag, 1981.

Laude, L. D., F. H. Pollak and M. Cardona, Phys. Rev. B**3**, 2623(1971)

Ledebo, L. and B. K. Ridley, J. Phys. C. **15**, L961(1982)

Lee, H. J., J. Basinski, L. Y. Juravel, J. C. Wooley, Can. J. Phys. **57**,419(1978)

Lei, X. L., J. Phys. C**18**,L593(1985)

Levinson, M. and J. A. Kafalas, Phys. Rev. B**35**,9383(1987)

Li, G. P. and K. L. Wang, J. Appl. Phys.**57**, 1016(1985)

Li, M.-F., Y.-S. Yao and G. Bai, J. Appl. Phys. **58**, 2599(1985)

Li, M. -F. and P. Y. Yu, Sol. St. Commun. **61**, 13(1987)

Lindefelt, U., Phys. Rev. B**28**, 4510(1983)

Lindefelt, U. and A. Zunger, Phys. Rev. B**30**, 1102(1984)

Lipari, N. O. and A. Baldereschi, Phys. Rev. Lett. **25**, 1660(1970)

Look, D. C., P. C. Colter, Phys. Rev. B**28**,1151(1983)

Ludwig, G.W. and H.H. Woodbury, Solid State Physics 13, 223 (1962)

Meese, J. M., J. W. Farmer, and C. D. Lamp, Phys. Rev. Lett. **51**, 1286 (1983)

Mendez, E. E., P. J. Price, and M. Heilblum, Appl. Phys. Lett. **45**, 294 (1984)

Mendez, E. E. and W. I. Wang, Appl. Phys. Lett.**46**, 1159(1985)

Merle, J.-C., M. Capizzi, P. Fiorini and A. Frova, Phys. Rev. B**17**, 4821(1978)

Miller, G. L., J. V. Ramirez, D. A. H. Robinson, J. Appl. Phys. **46**, 2638 (1975)

Miller, G. L., D. V. Lang, L. C. Kimerling, Ann. Rev. of Mat. Sci., Annual Reviews, Inc., 1977, pp. 377-448

Mohring, H. D., J. Weber and R. Sauer, Phys. Rev.B**30**, 894(1984)

Müller, H. , R. Trommer, M. Cardona and P. Vogl, Phys. Rev. B**21**, 4879(1980)

Muro, K. and A. J. Sievers, Phys. Rev. Lett.**57**, 897(1986)

Nag, B. R. and G. M Dutta, J. Phys. **C11**, 119(1978)

Nolte, D. D., W. Walukiewicz, and E. E. Haller, Bul. Am. Phys. Soc. **32**, 515(1987a)

Nolte, D. D. and E. E. Haller, J. Appl. Phys.62, 900(1987b)

Nolte, D. D., W. Walukiewicz, and E. E. Haller, Phys. Rev. Lett. **59**, 501 (1987c)

Nolte, D. D., W. Walukiewicz, and E. E. Haller, Phys. Rev.B**36**,9374(1987d)

Nolte, D. D., W. Walukiewicz, and E. E. Haller, Phys. Rev.B**36**,9392(1987e)

Nolte, D. D., W. Walukiewicz and E. E. Haller, Proc. Fall Meeting of MRS(Boston,1987) Symposium E

Norton, P., P. Braggins and H. Levinstein, Phys. Rev.B**8**, 5632(1973)

Pantelides, S. T., ed. Deep Centers in Semiconductors, (Gordon and Breach:New York, 1985)

Pfeffer, P., I. Gorczyca, and W. Zawadzki, Sol. St. Comm. **51**, 179 (1984)

Phillips, J. C., Phys. Rev. B**1**, 1540(1970)

Phillips, J. C., Rev. Mod. Phys.**42**, 317(1970)

Picoli, G., A. Chomette and M. Lannoo, Phys. Rev. B**30**, 7138(1984)

Pikus, G. E. and G. L Bir, Sov. Phys. - Solid St. **1**, 1502(1960)

Ren, S. Y., J. D. Dow, and D. J. Wolford, Phys. Rev. **B25**, 7661 (1982)

Sah, C. T., L. Forbes, L. L. Rosier and A. F. Tasch, Jr., Solid-State Electron.**13**, 759(1970)

Sah, C. T., Semiconductor Silicon 1977 (Electrochemical Society, Pannington, NJ, 1981), p.861

Samara, G. A. and C. E. Barnes, Phys. Rev. Lett. **57**,2069(1986)

Samara, G. A. and C. E. Barnes, Phys. Rev. B**35**, 7575(1987a)

Samara, G. A., Phys. Rev. B**36**, 4841(1987b)

Sauer, R and J. Weber, Physica **116B**, 195(1983)

Scheffler, M., J. P. Vigneron and G. B. Bachelet, Phys. Rev. Lett.**49**, 1765(1982)

Scheffler, M., F. Beeler, O. Jepsen, O. Gunnarsson, O.K. Andersen and G. B. Bachelet, 13th ICDS(Coronado) pg.45 (1985)

Schlesinger, T. E. and T. C. McGill, Phys. Rev. B**28**, 3643(1983)

Stauss, G. H., J. J. Krebs and R. L. Henry, Phys. Rev. B**16**, 974(1977)

Stöffler, W. and J. Weber, Phys. Rev. B**33**, 8892(1986); Proceedings of the 14th ICDS, ed. H. J. von Bardeleben(Tran Tech., Clausthal-Zellerfeld, 1986), pp. 705-710

Stoneham, A. M., Theory of Defects in Solids (Oxford Univ. Press, London, 1975)

Stoneham, A. M., Rep. Prog. Phys. **44**, 79(1981)

Taguchi, A., S. Yamada, J. Appl. Phys.**61**,2412(1987)

Takeda, Y. and A. Sasaki, Sol. St. Electr. **27**,1127(1984)

Tejedor, C. and F. Flores, J. Phys. C**11**, L19(1978)

Tersoff, J., Phys. Rev. **B30**, 4874(1984a)

Tersoff, J., Phys. Rev. Lett.**52**, 465(1984b)

Tersoff, J., Phys. Rev. **B32**, 6968(1985)

Tersoff, J., Phys. Rev. Lett. **56**, 675(1986)

Tersoff, J. and W. A. Harrison, Phys. Rev. Lett.**58**, 2367(1987)

Thewalt, M. L. W., U. O. Ziemlis, S. P. Watkins and R. R. Parsons, Can. J. Phys.**60**, 1691(1982)

Thurber, W. R., R. A. Forman and W. E. Phillips, J. Appl. Phys.**53**, 7397(1982)

Van de Walle, C. G. and R. M. Martin, Phys. Rev. B**35**, 8154(1987a)

Van de Walle, C. G. and R. M. Martin, unpublished

van Kooten, J.J., G.A. Weller, and C.A.J. Ammerlan, Phys. Rev. B **30**, 4564(1984)

Verges, J. A., D. Glötzel, M. Cardona and O. K. Andersen, phys. stat. sol.(b)**113**, 519(1982)

Vincent, G., D. Bois and A. Chantre, J. Appl. Phys. **53**, 3643(1982)

Wallis, R. H., A. Zylberstejn, and J. M. Besson, Appl. Phys. Lett. **38**, 698(1981)

Walukiewicz, W., J. Lagowski, L Jastrzebski, P. Rava, M. Lichtensteiger, C. H. Gatos, J. Appl. Phys.**51**,2659(1980)

Walukiewicz, W., H. E. Ruda, J. Lagowski, and H. C. Gatos, Phys. Rev. **B32**, 2645(1985)

Walukiewicz, W., J. Appl. Phys. **59**,3577(1987)

Watkins, G. D. and J. W. Corbett, Phys. Rev.**121**, 1001(1961)

Watkins, G. D. and J. W. Corbett, Phys. Rev.**138**,A543(1965)

Watkins, G. D., Phys. Rev. B**12**, 4383(1975)

Watkins, G. D., Physica 117&118B, 9(1983)

Watkins, S. P. and M. L. W. Thewalt, Can. J. Phys.**63**, 1074(1985)

Weber, E. R., Appl. Phys. A**30**, 1(1983)

Weber, J., (1987) unpublished

West, C. L., W. Hayes, J. F. Ryan and P. J. Dean, J. Phys. C**13**, 5631(1980)

White, A. M., P. Porteous, W. F. Sherman and A. A. Stadtmuller, J. Phys. C**10**,L473(1977)

Wiley, J. D. in: Semiconductors and Semimetals, ed. R. K. Willardsen and A. C. Beer (Academic Press, New York, 1975) vol. 10, chapt. 2

Wolfe, C. M., G. E. Stillman, J. D. Dimmock, J. Appl. Phys. **41**,504(1970)

Wolford, D. J., J. A. Bradley, K. Fry and J. Thompson, Proc. 17th Int. Conf. Phys. Semicond., Springer-Verlag, New York(1985) p. 627

Wünstel, K. and P. Wagner, Appl. Phys. A **27**, 207(1982a)

Wünstel, K., O. Kumagai, P. Wagner, and W. Jantsch, Appl. Phys. A **27**, 251(1982b)

Wünstel, K., K.H. Froehner and P. Wagner, Physica **116**B, 301(1983)

Zallen, R. and W. Paul, Phys. Rev.**155**, 703(1967)

Zunger, A. and U. Lindefelt, Phys. Rev. B**26**, 5989(1982)

Zunger, A. and U. Lindefelt, Phys. Rev. B**27**, 1191(1983)

Zunger, A., Solid State Physics, vol. 39, Academic Press (1986)

Zylberstejn, A., R. H. Wallis and J. M. Besson, Appl. Phys. Lett. **32**,764(1978)

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720