Mesoscopic pointlike defects in semiconductors: Deep-level energies

D. D. Nolte
Department of Physics, Purdue University, West Lafayette, Indiana 47907-1396
(Received 20 February 1998)

Mesoscopic pointlike defects are a class of extended defects with surfaces of minimal curvature that span in size from the point-defect limit (multivacancy, antisite, impurity complex, etc.) to macroscopic inclusions or voids within a semiconductor host. The structural, electronic, and optical properties of these defects evolve continuously from the quantum-mechanical limit to the classical limit. Mesoscopic defects share some features in common with quantum dots, such as Coulomb-charging energies, but unlike quantum dots their electronic properties are dominated by the covalent bond energies of the defect-semiconductor interface. The deep-level energies of spheroidal mesoscopic defects are calculated self-consistently in the unrestricted Hartree-Fock approximation using an extension of the many-electron model of Haldane and Anderson. The calculations in GaAs reveal a high-multiplicity Coulomb ladder of discrete charge states distributed across the semiconductor band gap and centered on the charge neutrality level. [S0163-1829(98)07836-9]

I. INTRODUCTION

The history of semiconductor research over the past century has been characterized by steadily increasing material heterogeneity. The first half century concentrated on bulk crystals of elemental and compound semiconductors, while the latter half saw forays into epitaxy. Epitaxy evolved from simple homoepitaxy into more sophisticated forms, culminating in the versatile heteroepitaxy of molecular beam epitaxy and related growth techniques. Today, semiconductor heteroepitaxy is continuing to evolve, expanding beyond all-semiconductor heterostructures to include metal-semiconductor,[1] insulator-semiconductor, and superconductor-semiconductor heterostructures and composites.[2] With the proliferation of diverse new forms of heterogeneous composite materials, it is important to identify unifying principles and concepts that encompass the physics of a wide range of seemingly unrelated materials within a single class.

For instance, the physical properties of inclusions within semiconductors may be viewed as a single class of mesoscopic defect that includes the simple limit of point defects as well as the electronic properties of extended defects. The accumulation of vacancies into increasingly larger multivacancy complexes is perhaps the best example of the continuous extension from the point-defect limit (the isolated vacancy) to the mesoscopic limit (voids).[3] Impurity atoms or nonstoichiometric concentrations can also accumulate into precipitates within the semiconductor.[1] These voids and precipitates are mesoscopic pointlike defects that exchange charge with the semiconductor host and have well-defined ''deep-level'' energies.

The term ''mesoscopic'' describes those defects that have a characteristic energy scale that brings them continuously from the point-defect (quantum) limit to the classical limit. This characteristic energy is the Coulomb-charging energy, reduced by hybridization with the semiconductor host. Mesoscopic defects have Coulomb-charging energies comparable to or smaller than the semiconductor band gap. The term ''pointlike'' describes defects that are primarily spheroidal, with possible faceting. They are unlike other extended defects, such as line defects (dislocations, etc.) or plane defects (platelets, stacking faults, etc.), because pointlike defects have bulk volumes enclosed by surfaces of minimal curvature. The small curvature minimizes Coulomb energies and maximizes the number of charge states that can exist within the semiconductor band gap.

Mesoscopic pointlike defects retain many of the features of point defects, such as the presence of discrete charge-state energy levels within the semiconductor band gap, and their ability to participate in compensation mechanisms within the semiconductor. But they also have important differences from point defects, such as extremely high charge-state multiplicity. It is the high multiplicity, in particular, that makes these mesoscopic pointlike defects important constituents in compensated semiconductors.

The importance of mesoscopic defects for charge compensation was proposed by Warren *et al.*[4] to explain the semi-insulating properties of nonstoichiometric GaAs. Nonstoichiometric GaAs forms during molecular beam epitaxy on substrates held at low temperatures (near 200–300 °C) rather than at the usual growth temperature of 600 °C. Low-temperature growth incorporates excess arsenic into the GaAs.[5,6] A key discovery by Purdue researchers was that the excess arsenic undergoes precipitation into small arsenic inclusions upon postgrowth anneal.[7] The accumulation process and size of the precipitate are controlled by growth or postgrowth material processing.[1] Therefore, the precipitates could be ''engineered,'' and their size and spacing could be controlled and chosen simply by playing off growth temperature against anneal temperature and time.[8] Nonstoichiometric GaAs and related compounds are semi-insulating under most growth and postgrowth processing conditions and are playing an increasingly important technological role. They are currently being developed as ultrafast photodetectors[9,10] with subpicosecond lifetimes,[11] as far-infrared photodetectors,[12] in the fabrication of GaAs-based integrated circuits,[13] and in photorefractive quantum wells,[14–16] among other applications.

Since the discovery of the arsenic precipitates, the semi-

insulating nature of nonstoichiometric GaAs has been a source of controversy, until recently. The semi-insulating properties were initially attributed to the high concentration of arsenic-related point defects.[17] Warren *et al.*[4] proposed that the arsenic precipitates were acting as internal Schottky barriers with associated depletion regions surrounding the precipitates. Subsequent studies have supported either the defect or the Schottky models, depending on the growth and anneal conditions. The production of mesoscopic voids in heavily radiation-damaged GaAs (Ref. 3) raised a complication in the defect vs Schottky argument because these voids cannot be viewed as Schottky barriers with a high density of available internal states. However, the materials with the voids are semi-insulating with many of the same properties as nonstoichiometric GaAs containing arsenic precipitates. It was shown recently that the voids play virtually an identical role as the precipitates for their ability to compensate charge,[18] and that voids and precipitates can be viewed within a broader class of mesoscopic pointlike defects that share many common properties, independent of their internal constituents.

Because the energy levels of mesoscopic pointlike defects in semiconductors attain high charge multiplicity, they must be calculated within a many-electron theoretical framework. One simple many-electron model that has been a valuable heuristic tool is the model of Haldane and Anderson.[19] This model was originally developed to explain the occurrence of multiple charge states of transition-metal impurities in semiconductors. It was later used to calculate effective U properties of the dangling bond in silicon,[20] and to explain the band lineup of heterostructures with respect to transition-metal energy levels.[21] More recently, it has been used to calculate the charge-state splitting of deep energy levels in narrow-gap semiconductors,[22] to track the formation of Schottky barriers by atoms adsorbed on a semiconductor surface,[23] and to explain the compensation of doped semiconductors by mesoscopic voids.[18]

We extend the Haldane-Anderson model to calculate the deep energy levels of mesoscopic pointlike defects in GaAs. Two extensions to the Haldane-Anderson model are necessary. First, the intrasite Coulomb interaction energy U is identified. This is done in a manner that is consistent with both the classical limit of the charging energy of a classical metallic sphere, as well as with the point-defect limit. The second extension to the model is the identification of the role played by the charge neutrality level.[24] The neutral mesoscopic defect level coincides with the charge neutrality level of the semiconductor band structure. This behavior is included explicitly in the many-electron model. These two extensions are described in Sec. II. In Sec. III, the calculations are presented for the deep energy levels for both spherical, as well as spheroidal defects. The classical limit is discussed in Sec. IV as the limiting behavior for large defect radius.

II. HALDANE-ANDERSON MODEL OF MESOSCOPIC DEFECTS

The basic mechanics of the Haldane-Anderson model for deep-level point defects is introduced first. The model is then extended to treat mesoscopic defects.

A. Haldane-Anderson model

The Haldane-Anderson model for transition-metal impurities in semiconductors begins with a model Hamiltonian solved in the unrestricted Hartree-Fock approximation in which only the dominant Coulomb term U in the two-particle energy is retained, and the exchange term J is omitted.[19] The model Hamiltonian is

$$H_d^{\rm eff}=E_{m\sigma}n_{m\sigma}+\sum_k \varepsilon_k n_k+\sum_k V_{ksp}c_k^+ c_{sp}+\text{c.c.}, \quad (1)$$

where $E_{m\sigma}$ is the effective defect energy, $n_{m\sigma}$ is the fractional occupancy of state $|m\sigma\rangle$, ε_k is the dispersive energy of an electron in state k with a state occupancy n_k, and V_{ksp} is the interaction matrix element between extended states and the sp bonds of the localized defect. The effective defect energy is determined self-consistently from

$$E_{m\sigma}=E_d+\frac{N_s-1}{N_s}U\sum_{m\sigma\neq m\sigma'}{}' \langle n_{m\sigma'}\rangle, \quad (2)$$

where E_d is the ''bare'' defect energy, N_s is the total number of surface states, and U is the Coulomb interaction energy. The sum is over both occupied and unoccupied surface states, and the average occupancies $\langle n_{m\sigma}\rangle$ are calculated from the Green's function

$$G_{m\sigma}(\omega)=\frac{1}{[\omega-E_{m\sigma}-\Sigma(\omega)]}, \quad (3)$$

where the self-energy is given by

$$\Sigma(\omega)=\sum_k \frac{|V_{ksp}|^2}{\omega-\varepsilon_k}, \quad (4)$$

which contains both real and imaginary parts given by

$$\Sigma(\omega)=\Sigma'(\omega)+i\Delta(\omega), \quad (5)$$

where

$$\Delta(\omega)=\pi|V_{ksp}|^2\rho_k \quad (6)$$

and ρ_k is the density of states of the band structure. The original Haldane-Anderson model assumed a constant density of states to illustrate the basic principles. In our analysis, we go a step further and use the total density of states derived for GaAs by Cohen and Chelikowsky.[25] Following Haldane and Anderson, we assume an energy-independent interaction energy V_{ksp}. While these gross approximations affect the details of the calculations, they are useful to illustrate the general trends and show the physical content of the theory.

The average state occupancies are expressed as

$$\langle n_{m\sigma}\rangle=\frac{1}{\pi}\,\text{Im}\int_{-\infty}^{E_F}G_{m\sigma}(\omega)d\omega=Z_1(E_{m\sigma})+Z(E_{m\sigma}), \quad (7)$$

where Z_1 and Z are given by

$$Z_1(E_{m\sigma})=\frac{1}{\pi}\,\text{Im}\int_{-\infty}^{E_v}G_{m\sigma}(\omega)d\omega, \quad (8)$$

$$Z(E_{m\sigma}) = \frac{1}{\pi} \operatorname{Im} \int_{E_v}^{E_c} G_{m\sigma}(\omega)\, d\omega, \tag{9}$$

where Z_1 is the contribution to $\langle n_{m\sigma} \rangle$ from hybridization between the defect states and the states of the occupied valence band, and Z is the contribution from the pole of the Green's function $G_{m\sigma}$ in the gap. The charge contributions can be solved numerically[19] using

$$Z_1(E_{m\sigma}) = \frac{1}{\pi} \int_{-\infty}^{E_v} \frac{\Delta}{[\omega - E_{m\sigma} - \Sigma'(\omega)]^2 + \Delta^2}\, d\omega \tag{10}$$

and

$$Z(E_{m\sigma}) = \frac{1}{\left| 1 - \frac{d}{d\omega} \Sigma'(\omega) \right|_{\omega_p}}. \tag{11}$$

The defect energies and average occupancies must be solved self-consistently because the effective defect energy $E_{m\sigma}$ depends on the average occupancies $\langle n_{m\sigma} \rangle$ in Eq. (2), which in turn depends on the effective defect energy $E_{m\sigma}^{\text{eff}}$ by Eq. (7). The first step in this procedure is to calculate the self-energy term $\Sigma(\omega)$. This calculation is performed once for a given interaction energy V_{ksp} and density of states ρ_k. The second step is to find the locations of the poles in the Green's function of Eq. (3), which satisfy the condition

$$\omega_p - \Sigma(\omega_p) - E_{m\sigma} = 0. \tag{12}$$

Both occupied and unoccupied poles are found, and the average state occupancies $\langle n_{m\sigma} \rangle$ are calculated by Eq. (7).

In the Hartree-Fock approximation, it is necessary to calculate *total* energies for a given charge state because the two-particle energies are counted twice. Half of these terms must be subtracted from the total sum. The total energy is finally given by

$$E^{\text{tot}} = \sum_{m\sigma} \int_{E_q}^{E_{m\sigma}} Z_1(E)\, dE + \sum_{m\sigma\ \text{occ}} \omega_p - \tfrac{1}{2} U \sum_{m\sigma \neq m\sigma'}{}' \langle n_{m\sigma} \rangle \langle n_{m\sigma'} \rangle, \tag{13}$$

in which the sums are over both occupied and unoccupied states, except for the sum over the occupied poles.

B. Extension to mesoscopic defects

The Haldane-Anderson model requires two extensions before it can be applied to mesoscopic pointlike defects. The first extension handles the large number of states that arise at the internal surface of the void or precipitate, and ensures that the neutral defect energy falls near the charge-neutrality level of the semiconductor. The second extension defines an intradefect Coulomb energy for the electrons occupying the surface states of the defect.

1. Charge neutrality

The number of electrons required to occupy the total N_s surface states to produce a neutral defect state is defined as n_v^0, and Δn is defined as the electron occupancy that deviates from charge neutrality. Rewriting Eq. (2) with these definitions gives

$$x = E_d + U\{(n_v^0 + \Delta n - 1)[Z_1(x) + Z(x)] + (N - n_v^0 - \Delta n) Z_1(y)\}, \tag{14}$$

$$y = E_d + U\{(n_v^0 + \Delta n)[Z_1(x) + Z(x)] + (N - 1 - n_v^0 - \Delta n) Z_1(y)\}, \tag{15}$$

where x is the occupied effective defect energy $E_{m\sigma}^{\text{occ}}$, y is the unoccupied effective defect energy $E_{m\sigma}^{\text{unocc}}$, and E_d is the ''bare'' defect energy associated with N total surface states.

There is on average one surface bond per surface atom for a semiconductor surface of random orientation. Therefore the number of surface states of even a moderately sized void with a radius of 30 Å or larger can approach a thousand. The ''bare'' defect energy E_d in Eqs. (14) and (15) in this case will be hundreds of eV below the valence-band maximum. Such a large negative defect energy must be balanced precisely by a Coulomb energy U to cause low-charge-state energy levels to occur near the middle of the band gap, which is known to be the physical situation for semiconductor surface states. The Haldane-Anderson model cannot guarantee that the neutral defect energy will lie near the middle of the band gap because E_d cannot be easily quantified. Therefore our first extension of the Haldane-Anderson model to apply to mesoscopic defects is to *a priori* choose the energy of the charge-neutral defect to be equal to E_q, the charge-neutrality level of the bulk semiconductor.

The charge-neutrality level E_q of a semiconductor has been identified by several researchers beginning with Heine[26] to describe the nature of the metal-derived gap states, which were later called metal-induced gap states or MIGS.[27] It can be approximately identified as the branch point in the complex band structure where virtual gap states have equal contributions from the conduction and valence bands,[24] and has been calculated by Tersoff for several of the common semiconductors.[28–30]

By specifically requiring that the neutral defect energy coincide with the charge-neutrality level, Eqs. (14) and (15) simplify to $x = E_q$ and $y = y_q$ for the case when $\Delta n = 0$. This condition allows the difficult quantity E_d to be cancelled from the equations to give the extended equations

$$x = E_q + U\{(n_v^0 - 1)[Z_1(x) - Z_1(E_q) + Z(x) - Z(E_q)] + (N - n_v^0)[Z_1(y) - Z_1(y_q)] + \Delta n[Z_1(x) + Z(x) - Z_1(y)]\} \tag{16}$$

and

$$y = x + U\{Z_1(x) + Z(x) - Z_1(y)\}. \tag{17}$$

The new equations are expressed in terms of the charge-neutrality level E_q and the deviation from charge neutrality Δn. An important feature of these equations is that they are also expressed as differences of the functions Z_1 and Z, which gives small numbers that counterbalance the large numbers defined by N and n_v^0. This process produces a lad-

der of deep donor and acceptor energy levels centered in the band gap around the energy E_q.

Equations (16) and (17) are solved self-consistently for $E_{m\sigma}$ for both occupied and unoccupied poles in the band gap. Once the poles and defect energies are obtained, the energy levels of the defect in the band gap are found by calculating differences in total energy. The total energy of a state with Δn electrons is now given by

$$E^{\mathrm{Tot}}(\Delta n)=\Delta n Z_1(E^x_{m\sigma})[E^x_{m\sigma}-E_q]-\Delta n Z_1(E^y_{m\sigma})[E^y_{m\sigma}-y_q] + \Delta n \omega^x_p - \tfrac{1}{2}U\Delta n^2[Z_1(E^x_{m\sigma})+Z(E^x_{m\sigma})-Z_1(E^y_{m\sigma})]^2, \tag{18}$$

where the first line is the change in band energy with change in occupancy, the second line is the change in defect energy, and the third line is the correction to the Hartree-Fock total energy. Equation (18) replaces the Haldane-Anderson form in Eq. (13) and represents the primary extension of the current model to treat mesoscopic pointlike defects with large numbers of surface states. The deep energy level of the defect for the charge-state transition from Δn to $\Delta n-1$ is given by

$$E^{\mathrm{HA}}(\Delta n,\Delta n-1)=E^{\mathrm{Tot}}(\Delta n)-E^{\mathrm{Tot}}(\Delta n-1). \tag{19}$$

These total energy differences are the energies that can be compared to values measured experimentally by deep-level spectroscopy within the semiconductor host.

2. Nature of the surface states

The types of mesoscopic defect of concern here are either metallic or semimetallic inclusions or voids within the semiconductor host. The natures of the semiconductor surface states are similar in both cases in spite of the different boundary conditions. In the case of mesoscopic voids, the surface states are the unreconstructed dangling bonds. It was shown in Ref. 18 that if even 90% of the surface bonds are reconstructed, the void will still produce a ladder of energy levels within the band gap and will provide sufficient states to compensate doped material. Conversely, in the case of metallic or semimetallic inclusions, the surface states can be either dangling surface bonds of the semiconductor, or can be metal-induced gap states.[27] A considerable literature, with noted controversies, has developed concerning the nature of the states at the metal-semiconductor interface.[24] Despite the opposing views, the theories of the metal-semiconductor interface converge on similar behavior. The surface states arise from hybridization or interaction between the surface states and the bulk states of the semiconductor. Most of the theories of the metal-semiconductor interface admit this interaction, although they may differ on the nature of the interaction.

Given the presence of surface states of the metallic inclusion or void, it is necessary to identify the Coulomb energy U of Eq. (2). In the case of transition-metal point defects in semiconductors this is taken to be an effective Coulomb interaction[31] that is reduced from the free transition-metal atom by hybridization with the host and screening by long-range interactions.[21] In the case of spherical metallic inclusions or voids, we adopt a semiclassical expression for the effective intrasite Coulomb interaction for a spherical defect of radius a, given by

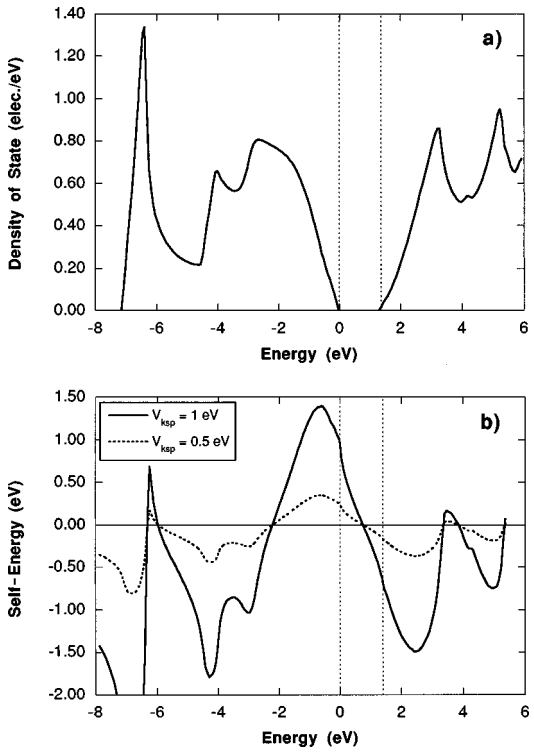

FIG. 1. (a) The density of states for GaAs calculated in Ref. 25. (b) The self-energy calculated from the density of states for two values of the interaction matrix element. The zero crossing of the self-energy defines the charge neutrality level E_q.

$$U(a)=\frac{e^2}{C}, \tag{20}$$

where the capacitance of the sphere imbedded in the semiconductor is $C=4\pi\varepsilon\varepsilon_0 a$.

III. DEEP-LEVEL ENERGIES

The theoretical density of states of GaAs from Ref. 25 are shown in Fig. 1 with the calculated self-energy for matrix elements $V_{ksp}=1$ and 0.5 eV. To solve Eq. (2) self-consistently, the poles of the Green's function are found first. The poles occur when $\omega_p-E_{m\sigma}=\Sigma(\omega_p)$, which is shown graphically in Fig. 2. The self-energy has a discontinuous slope at the edges of the band gap, but poles can occur outside the gap, corresponding to resonance states. Both the occupied and unoccupied poles are found by using $E_{m\sigma}=x$ or $E_{m\sigma}=y$, respectively, in Eq. (12).

The charge neutrality position within the band gap occurs when $\omega_p=E_{m\sigma}$ and the self-energy $\Sigma(E_q)=0$, giving the solution $E_{m\sigma}=E_q$. Because the matrix element V_{ksp} is assumed to be a constant, the charge neutrality position is independent of the magnitude of V_{ksp}. In our calculations, E_q is taken at the zero crossing of the self-energy. This was calculated numerically from the density of states[25] to be $E_q \approx 0.75$ eV, which is in close agreement with more fundamental calculations of E_q.[30]

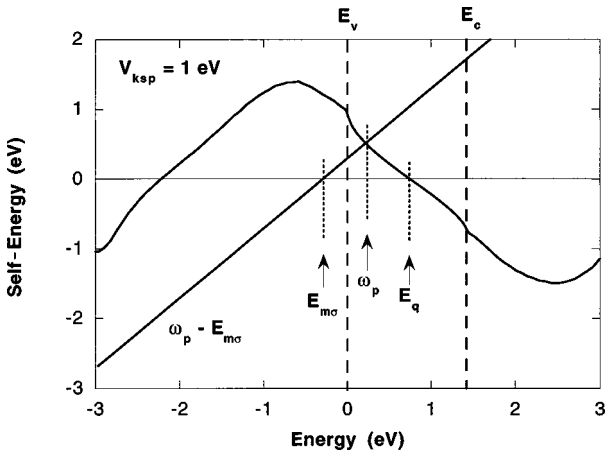

FIG. 2. Graphical solution to find the occupied pole in the band gap for a defect energy $E_{m\sigma}$.

A. Spherical defects

The full self-consistent calculations were performed for spherical mesoscopic defects with radii between 5 and 30 Å. The calculations were not performed for defect radii smaller than 5 Å (corresponding approximately to a pentavacancy or five-atom metal cluster) for several reasons. The Haldane-Anderson (HA) model neglects lattice relaxation energies, which can contribute substantially to defect energy levels. In addition, there are quantum size effects and chemical effects that become appreciable in this limit, but are ignored by the HA model. For instance, in the case of spherical voids, the chemical species of the dangling bond (Ga or As bonds) will affect the energy position in the band gap. In addition, the dangling bonds will themselves hybridize, leading to crystal-field splitting of the defect levels, such as the A_1 and T_2 states of the monovacancy.

Several of the energies that enter into the self-consistent solutions to the total energies for the defect levels in the semiconductor are shown in Fig. 3 for $Q=-1e$ and $V_{ksp}=1$ eV. The Hartree-Fock defect energy $E_{m\sigma}$ overcounts the two-particle interaction energy. It is used in Eq. (12) to find the occupied pole. These two energies, with the Coulomb interaction U, combine to give the total defect energy for this

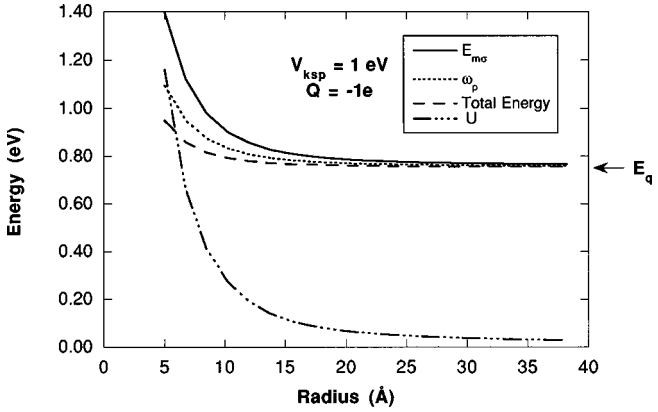

FIG. 3. Energy contributions to the $Q=-1$ charge state for $V_{ksp}=1$ eV showing the Coulomb interaction energy U, the occupied pole ω_p, the defect energy $E_{m\sigma}$, and the Haldane-Anderson deep-level energies for the $(-1,0)$ charge-state transition.

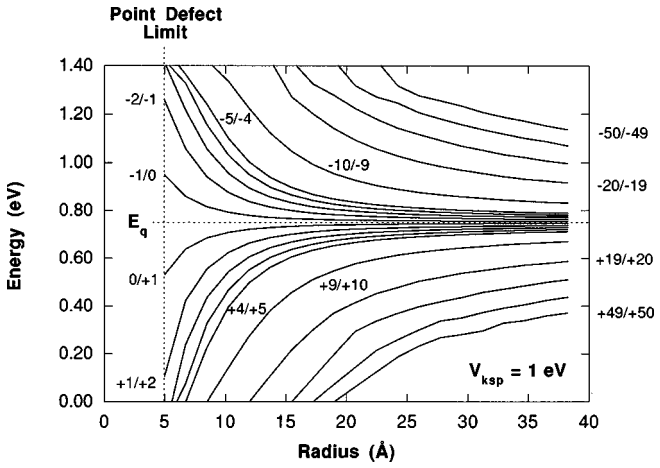

FIG. 4. Haldane-Anderson deep-level energies for spherical voids calculated for charge states up to $Q=\pm 5e$, and higher charge states in intervals of $\pm 10e$.

charge state. The occupied pole energy lies midway between the total energy and the defect energy $E_{m\sigma}$. This trend continues for higher charge states for both charge signs. For high charge-state transitions $(Q,Q-1)$, the deep-level energy is the average of the occupied poles for the two charge states. This suggests that simplified calculations may use the occupied pole alone, which is relatively simple to find, rather than needing to calculate the total energies from Eq. (18).

The deep-level energies of spherical defects of successively higher charge states are shown in Fig. 4. The transitions for $Q=\pm 5e$ are shown, as well as higher charge-state transitions in intervals of $\pm 10e$. With increasing defect radius, the defect levels fan into the band gap as the Coulomb energy U rapidly decreases. The distribution of acceptor and donor levels is relatively symmetric about the charge neutrality level E_q. Over 100 charge states can exist within the band gap for a defect with a radius of only 20 Å.

B. Ellipsoidal defects

A semiclassical capacitance can also be defined for prolate and oblate spheroidal defects. Oblate spheroids ($a=b>c$ ''pancakes'') have

$$C=4\pi\varepsilon\varepsilon_0\frac{\sqrt{a^2-c^2}}{\cos^{-1}(c/a)} \tag{21}$$

and prolate spheroids ($a>b=c$ ''needles'') have

$$C=4\pi\varepsilon\varepsilon_0\frac{\sqrt{a^2-b^2}}{\cosh^{-1}(a/b)}. \tag{22}$$

Prolate or oblate spheroids of increasing eccentricity have increasing surfaces of curvature, which increases the Coulomb-interaction energy. The limit $c=0$ for oblate spheroids reduces to a planar defect analogous to a stacking fault.

The HA energies for two prolate spheroids with large eccentricities are compared in Fig. 5 with the energy of a spherical defect. The radius used for plotting is the semimajor axis of the spheroid. Increasing eccentricity turns the sphere into a needle and the capacitance increases significantly. Strongly eccentric spheroids are not ''pointlike,'' and

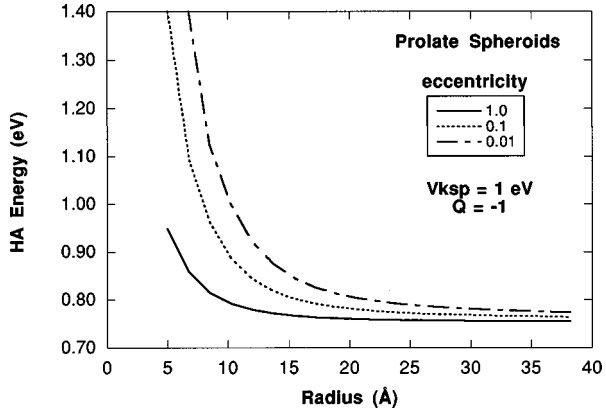

FIG. 5. HA energy levels for the $(-1,0)$ transition for prolate spheroids of increasing eccentricity.

do not fit well within the mesoscopic formalism described here. Fundamental quantum-mechanical calculations need to treat the energy levels of dislocations and platelets. However, weakly eccentric spheroids will be described well by the present model, and may be represented by real systems. For instance, spheroidal rather than spherical precipitates of silver were seen in TEM micrographs in silver-implanted GaAs.[32]

IV. THE CLASSICAL LIMIT

In the classical limit of a spherical metal inclusion the difference in total energies between defects that carry Δn and $\Delta n-1$ charges is

$$E^{\text{class}}(\Delta n,\Delta n-1)=(\Delta n-\tfrac{1}{2})\,\frac{e^2}{C}, \tag{23}$$

where the defect capacitance C is the quantity used for the Haldane-Anderson calculations. The Haldane-Anderson calculations for the mesoscopic defects should approach this dependence for sufficiently large defects when quantum size effects become negligible. To make the connection between

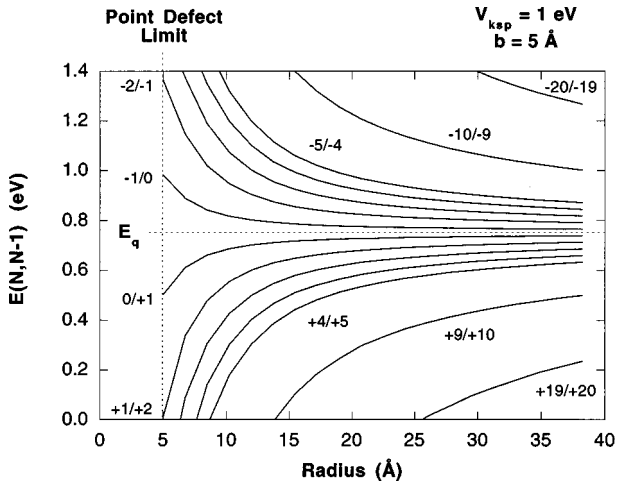

FIG. 6. Deep-level energies of spherical voids corrected for long-range potentials with $b=5$ Å showing the development of the Coulomb ladder.

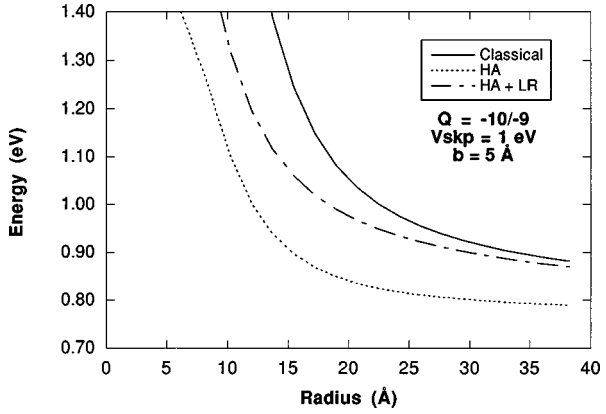

FIG. 7. HA energy level for the $(-10,-9)$ transition compared with the classical and HA-corrected levels for $b=5$ Å showing the transition from quantum to classical behavior with increasing radius.

the Haldane-Anderson model and the classical limit, it is necessary to correctly interpret the Haldane-Anderson results.

The Haldane-Anderson model describes the charge-dependent covalent hybridization of the defect surface states with the bulk states of the semiconductor. The bare Coulomb-interaction energy is reduced by charge density ''leaking'' into the bulk to form a long-range charge density tail. For point defects, the total energy is dominated by the intrasite Coulomb energy, and the long-range charge density has a negligible effect. However, in the case of mesoscopic defects, the long-range density contributes a significant portion of the total energy.

The energy levels of the Haldane-Anderson calculation can be corrected for the long-range density by including the density in the deep-level energy. The charge density at the defect surface is approximately given by $Z(E_{m\sigma})$. If the charge density decays exponentially beyond the defect surface as

$$\rho(r)=\rho_0\,\exp(-a/b) \tag{24}$$

with a decay length b, then the potential that the long-range charge density contributes to the deep-level transition is given by

$$U_{\text{LR}}=[1-Z(E_{m\sigma})^2]\,\frac{b(b+a)\rho_0}{\varepsilon\varepsilon_0}. \tag{25}$$

This long-range potential is added to the Haldane-Anderson deep-level energies $E(\Delta n,\Delta-1)$ to account for the long-range charge density. The defect energy level in the band gap is therefore given by

$$E^{\text{LR}}(\Delta n,\Delta n-1)=E^{\text{HA}}(\Delta n,\Delta n-1)+(\Delta n-\tfrac{1}{2})U_{\text{LR}}. \tag{26}$$

The long-range potential is due to both the wave function that extends beyond the surface of the mesoscopic defect, as well as to the long-range Coulomb tail of the defect potential. Highly localized defect wave functions would favor a small value for b of only several angstroms, while the long-range Coulomb tail would favor a value for b of tens of angstroms.

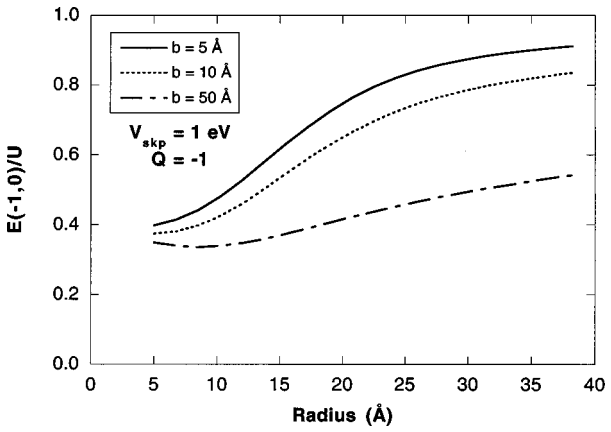

FIG. 8. Ratio of the corrected HA energy to the classical energy for the $(-1,0)$ transition for $b=5$, 10, and 50 Å. The asymptotic approach to the classical limit marks the mesoscopic regime.

The deep-level energies for a spherical pointlike defect within the Haldane-Anderson model, and corrected for the long-range potential, are shown in Fig. 6 for all charge states up to $\pm 5e$, and in steps of $\pm 10e$ for higher charge states. The decay length for the long-range charge was assumed to be $b=5$ Å in the calculations, which assumes that the deep defect wave function is restricted to only one or two nearest-neighbor distances, neglecting the long-range Coulomb tail. Figure 6 can be contrasted with Fig. 4, which were the uncorrected HA energies. The long-range potential contributes significantly to the deep-level energies, reducing the number of possible charge states that can exist inside the band gap. However, high charge multiplicity is still possible, with over 50 charge states on a defect with a radius of 40 Å.

The deep energy level before and after correction for the long-range potential is shown in Fig. 7 and compared with the classical energy for the transition $E(-10,-9)$, assuming $b=5$ Å. The corrected energy approaches the classical energy for radii larger than 40 Å, but significant deviations from the classical result occur for smaller defect radii. The ratios of the corrected HA deep-level energies to the classical energies are shown in Fig. 8 for $b=5$, 10, and 50 Å. The *a posteriori* long-range potential depends on the value chosen for the decay length b. The local nature of the deep-level wave functions would favor $b=5$ Å. However, if the long-range charge is accommodated in an effective-mass Coulomb tail, then $b=50$ Å may best describe the long-range corrections to the model of Haldane and Anderson. Even in the short-range case with $b=5$ Å, there is a clear crossover in behavior for defect radii between 10 and 30 Å. This crossover regime is, by definition, the mesoscopic regime that continuously connects the classical and the quantum limits.

V. CONCLUSIONS

In conclusion, the many-electron model of Haldane and Anderson that was originally developed to describe the charge states of point defects in semiconductors has been extended to describe the energy levels of mesoscopic pointlike defects. The low curvature of the defects and their size allows many charge states to exist inside the band gap of GaAs, producing a Coulomb ladder of energy levels centered on the charge neutrality level.

This extended Haldane-Anderson model is primarily heuristic, and helps describe the nature of the Coulomb ladder, but probably cannot be relied upon to give quantitative energies. It neglects many details, such as crystal-field splitting, chemical shifts, and lattice relaxation, which contribute to defect total energies. It is also necessary to correct the HA energies with contributions from long-range potentials because these corrections cannot be neglected for larger radii defects. The length scale of the long-range potential remains an unanswered question for these defects, and is bounded below and above by 5 and 50 Å, depending on whether it is composed primarily of the localized deep defect wave function, or by the long-range Coulomb potential.

In spite of these quantitative deficiencies in this extended Haldane-Anderson model, the qualitative trends are expected to give a good physical description of the origin and approximate magnitudes of the multiple energy levels that can exist for a single mesoscopic pointlike defect in a semiconductor. Because it also gives an estimate for the size and extent of the transition regime between the quantum limit and the classical limit for these defects, it should provide a convenient frame of reference for more sophisticated calculations.

ACKNOWLEDGMENTS

D. D. Nolte gratefully acknowledges support by the NSF MRSEC for Technology Enabling Heterostructures Grant No. DMR-9400415. Helpful discussions with P. Muzikar are also appreciated.

[1] M. R. Melloch, J. M. Woodall, E. S. Harmon, N. Otsuka, F. Pollak, D. D. Nolte, R. M. Feenstra, and M. A. Lutz, Annu. Rev. Mater. Sci. **25**, 547 (1995).
[2] M. L. Cohen, Phys. Scr. **T68**, 84 (1996).
[3] W. Walukiewicz, Z. Lilienthal-Weber, J. Jasinski, M. Almonte, A. Prasad, E. E. Haller, E. R. Weber, P. Grenier, and J. F. Whitaker, Appl. Phys. Lett. **69**, 2569 (1996).
[4] A. C. Warren, J. M. Woodall, J. L. Freeouf, D. Grischowsky, D. T. McInturff, M. R. Melloch, and N. Otsuka, Appl. Phys. Lett. **57**, 1331 (1990).
[5] K. Mahalingam, N. Otsuka, M. R. Melloch, J. M. Woodall, and A. C. Warren, J. Vac. Sci. Technol. B **9**, 2328 (1991).
[6] Z. Lilienthal-Weber, A. Claverie, J. Washburn, F. W. Smith, and A. R. Calawa, Appl. Phys. A: Solids Surf. **53**, 141 (1991).
[7] M. R. Melloch, N. Otsuka, J. M. Woodall, A. C. Warren, and J. L. Freeouf, Appl. Phys. Lett. **57**, 1531 (1990).
[8] M. R. Melloch, J. M. Woodall, N. Otsuka, K. Mahalingham, C. L. Chang, and D. D. Nolte, Mater. Sci. Eng., B **22**, 31 (1993).
[9] F. W. Smith, H. Q. Le, V. Diadiuk, M. A. Hollis, A. R. Calawa, S. Gupta, M. Frankel, D. R. Dykaar, G. A. Mourou, and T. Y. Hsiang, Appl. Phys. Lett. **54**, 890 (1989).
[10] A. C. Warren, N. Katzenellenbogen, D. Grischkowsky, J. M. Woodall, M. R. Melloch, and N. Otsuka, Appl. Phys. Lett. **58**, 1512 (1991).

[11] Gupta, Appl. Phys. Lett. **59**, 3276 (1991).
[12] A. Srinivasan, K. Sadra, J. C. Campbell, and B. G. Streetman, J. Electron. Mater. **22**, 1457 (1993).
[13] F. W. Smith, A. R. Calawa, C. Chen, M. J. Mantra, and L. J. Mahoney, IEEE Electron Device Lett. **EDL-9**, 77 (1988).
[14] I. Lahiri, K. M. Kwolek, D. D. Nolte, and M. R. Melloch, Appl. Phys. Lett. **67**, 1408 (1995).
[15] I. Lahiri, M. Aguilar, D. D. Nolte, and M. R. Melloch, Appl. Phys. Lett. **68**, 517 (1996).
[16] I. Lahiri, D. D. Nolte, M. R. Melloch, and M. B. Klein, Opt. Lett. **23**, 49 (1998).
[17] M. Kaminska, E. R. Weber, Z. Liliental-Weber, and R. Leon, J. Vac. Sci. Technol. B **7**, 710 (1989).
[18] D. D. Nolte, Appl. Phys. Lett. **70**, 3401 (1997).
[19] F. D. M. Haldane and P. W. Anderson, Phys. Rev. B **13**, 2553 (1976).
[20] W. B. Fowler and R. J. Elliott, Phys. Rev. B **34**, 5525 (1986).
[21] M. Hamera, W. Walukiewicz, D. D. Nolte, and E. E. Haller, Phys. Rev. B **39**, 10 114 (1989).
[22] W. Li and J. D. Patterson, Phys. Rev. B **53**, 15 622 (1996).
[23] S. Y. Davydov, Semiconductors **31**, 1062 (1997).
[24] W. Mönch, *Semiconductor Surfaces and Interfaces* (Springer-Verlag, Berlin, 1995).
[25] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors* (Springer-Verlag, Berlin, 1988).
[26] V. Heine, Phys. Rev. **138**, A1689 (1965).
[27] S. G. Louie and M. L. Cohen, Phys. Rev. B **13**, 2461 (1976).
[28] J. Tersoff, Phys. Rev. Lett. **52**, 465 (1984).
[29] J. Tersoff, Surf. Sci. **168**, 275 (1986).
[30] J. Tersoff, Phys. Rev. Lett. **56**, 2755 (1986).
[31] S. Froyen, Phys. Rev. B **22**, 2898 (1980).
[32] D. Crouse, D. D. Nolte, J. C. P. Chang, and M. R. Melloch, J. Appl. Phys. **81**, 7981 (1997).