Native defects and self-compensation in ZnSe

D. B. Laks*
Division of Metallurgy and Materials Science, Columbia University, New York, New York 10027 and IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

C. G. Van de Walle†
Philips Laboratories, Briarcliff Manor, New York 10510

G. F. Neumark
Division of Metallurgy and Materials Science, Columbia University, New York, New York 10027

P. E. Blöchl‡ and S. T. Pantelides
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 26 August 1991; revised manuscript received 27 December 1991)

Wide-band-gap semiconductors typically can be doped either *n* type or *p* type, but not both. Compensation by native point defects has often been invoked as the source of this difficulty. We examine the wide-band-gap semiconductor ZnSe with first-principles total-energy calculations, using a mixed-basis program for an accurate description of the material. Formation energies are calculated for all native point defects in all relevant charge states; the effects of relaxation energies and vibrational entropies are investigated. The results conclusively show that native-point-defect concentrations are too low to cause compensation in stoichiometric ZnSe. We further find that, for nonstoichiometric ZnSe, native point defects compensate *both* *n*-type and *p*-type material; thus deviations from stoichiometry cannot explain why ZnSe can be doped only one way.

I. INTRODUCTION

Wide band-gap semiconductors, such as ZnSe, ZnTe, ZnS, and diamond, have potential technological applications, especially for optical devices involving green or blue light.[1–3] Despite decades of research, many problems remain, mostly related to doping difficulties; some wide-band-gap materials can easily be made *n* type but not *p* type; others can be made *p* type, but not *n* type.[4] The cause of this difficulty remains a puzzle. At least five different explanations have been suggested,[5–9] but there is no firm evidence for any of them. One of the oldest and most popular explanations is that the doping of wide-band-gap semiconductors is compensated by native point defects.[2–12] According to this mechanism, the wide band gap could promote the formation of compensating native point defects because the formation energy of the defect is offset by the energy gained when electrons are transferred between the defect's electronic state in the gap and the Fermi level. For example, *p*-type doping may be compensated by defects that introduce electrons in levels near the conduction band. When the electrons drop from the level in the gap to the Fermi level (which is near the valence-band edge), the net formation energy for the compensating defect would be reduced by nearly the width of the band gap. This mechanism would be universal: it is independent of the dopant and the growth method used. The native-point-defect properties would directly determine the behavior of the material. A wide-band-gap semiconductor would tend to be *n* type if the dominant native point defects introduce full states near the conduction-band edge. It would be *p* type if the dominant defects introduce empty electronic states near the valence bands.

Our goal is to examine the native-point-defect mechanism using first-principles theoretical techniques based on density-functional theory and *ab initio* pseudopotentials. We will study native point defects in ZnSe, which is the wide-band-gap semiconductor that has received the most attention in the past decade. ZnSe can be grown *n* type, but only limited progress has been made growing *p*-type material.[13,14] Theoretical tools have been very useful in elucidating the properties of common semiconductors such as Si and GaAs.[15–18] Much less has been done for ZnSe, or any of the other II-VI semiconductors. For these materials, the plane-wave pseudopotential method, the standard for semiconductor defect calculations, does not work well. This is because the *d* electrons of the group-II metals are too tightly bound to be represented as valence electrons with a plane-wave basis set. In all previous pseudopotential calculations for ZnSe,[6,12,19] the zinc 3*d* electrons were treated as core states. Using this method, Jansen and Sankey[12] suggested that native-point-defect compensation is the cause of doping difficulties in ZnSe and ZnTe and on the same basis explained why ZnTe (which prefers to be *p* type) is different from ZnSe (which prefers to be *n* type). Unfortunately, the "*d*-in-the-core" pseudopotential approach is inaccurate: it cannot predict the experimental bulk properties of ZnSe,[20] and is therefore highly suspect for defect calculations.

We solve the *d*-electron problem by using a mixed-basis scheme, which adds to the plane-wave basis a set of tight-binding functions that can represent the *d* electrons as valence states. The mixed-basis scheme is implemented in a program that is efficient enough for large-scale defect calculations. Our defect calculations are the first for a II-VI semiconductor that include a proper treatment of the *d* electrons, and reach the level of accuracy previously attained for Si and GaAs. We calculate the formation energies of all native point defects in ZnSe. Using these formation energies we derive upper bounds for the defect concentrations. The results show clearly that native defect compensation in *stoichiometric* ZnSe is insignificant. Additional support for this conclusion is provided by calculations of native-point-defect concentrations in another wide-band-gap semiconductor, namely diamond. Here we derive the concentrations from published native-point-defect energies.[21] In *nonstoichiometric* ZnSe, native-point-defect compensations *will* occur, but will compensate *n*-type as well as *p*-type material. Deviations from stoichiometry, therefore, do not explain why it is easy to make *n*-type but not *p*-type ZnSe. Our results clearly indicate that native defects are not responsible for self-compensation in ZnSe and thus impose no *intrinsic* limitation on the ability to obtain both *n*-type and *p*-type conduction.

This paper is organized as follows: In Sec. II we describe the details of our mixed basis scheme. By relying on fast-Fourier-transform (FFT) routines and the convolution theorem,[22] total-energy calculations for a defect (which require a cell with a large volume) can be performed efficiently. A description of our test calculations follows; these establish the credibility of our theoretical methods. In Sec. III we describe our own total-energy calculations for the native point defects, and a discussion of the structure of each defect. Because ZnSe is a compound semiconductor, the formation energy of a single defect is not well defined. In Sec. IV we show how chemical potentials can be related to stoichiometry, yielding an unambiguous definition of formation energies in terms of the calculated total energies. Defect concentrations can then be obtained as a function of temperature, stoichiometry, and the Fermi level of the crystal. We then present our calculated native-point-defect concentrations (Sec. V), which show clearly that the native point defects do not affect the doping of ZnSe. We also show that the same is true of diamond. Having shown quantitatively that native point defects are not responsible for compensation, we present a qualitative analysis of whether native-point-defect concentrations increase with the width of the band gap.

II. THE MIXED-BASIS METHOD

In this section we describe our implementation of the mixed-basis method. Our formalism is based on density-functional theory in the local-density approximation (LDA),[23] using *ab initio* pseudopotentials.[24] In traditional pseudopotential calculations with a plane-wave basis set, the bulk of the computation consists of solving the eigenvalue problem for the Hamiltonian matrix. The mixed-basis scheme produces a much smaller Hamiltonian matrix by replacing many high-frequency plane waves with a few tight-binding functions. The price paid for the smaller Hamiltonian is that introducing tight-binding functions complicates the matrix elements and the integration of the charge density. We handle the added complications of the tight-binding functions by expanding them over the reciprocal lattice. The tight-binding functions ϕ are written as

$$\phi_{\mathbf{k}i}(\mathbf{r})=\sum_{\mathbf{G}}\phi_i(\mathbf{k}+\mathbf{G})e^{i(\mathbf{k}+\mathbf{G})\cdot r}\ , \tag{1}$$

where $\mathbf{G}$ is a reciprocal-lattice vector. Functions of this form automatically have the correct translational symmetry (Bloch's theorem). There are two principal advantages to the reciprocal-space expansion. One is that the choice of tight-binding functions is not restricted to any particular analytic form, such as Gaussians or Slater orbitals. This will allow us to use so-called pseudoatomic wave functions as basis functions, as discussed below. The second is that the reciprocal-space expansion is particularly well suited for treating the tight-binding functions and the plane waves in a unified fashion. (Note that the exponentials in the expansion for ϕ are simple plane waves.) This simplifies the calculation of the matrix elements between tight-binding functions and plane waves, as well as the calculation of the charge density. The scheme is similar to that used by Louie, Ho, and Cohen.[25] The programs are all new, and both the programs and the algorithms were carefully optimized to make large-scale defect calculations possible. We now discuss the details of our methods, using the work of Louie, Ho, and Cohen[25] as a starting point. A detailed description of the evaluation of the various matrix elements is presented in the Appendix. Further details on the method are given elsewhere.[26]

A. Basis set

Although a mixed basis allows great freedom in the choice of the basis set, our approach has been to keep our ZnSe calculations as similar as possible to the plane-wave calculations for Si and GaAs. Consequently, we use tight-binding functions to represent only the rapidly changing part of the zinc *d* orbitals. All other contributions to the wave functions are represented by plane waves. In this way we recover most of the advantages of a pure plane-wave calculation, and reduce the effort needed to choose and optimize the basis set. We have performed ZnSe calculations using two different forms for the zinc 3*d* orbitals: Gaussians and pseudoatomic wave functions. The pseudoatomic functions are the 3*d* pseudo-wave functions for the isolated zinc atom, as calculated by the program that generates the atomic pseudopotentials. We multiply the pseudoatomic basis functions by a smooth radial cutoff function that goes to zero for large *r* to remove the long-range tail (Fig. 1). (Basis functions with long-range tails become numerically unstable as the overlap between basis functions on different sites causes the basis set to become linearly dependent.) Using Gaussians requires at least two basis functions for each

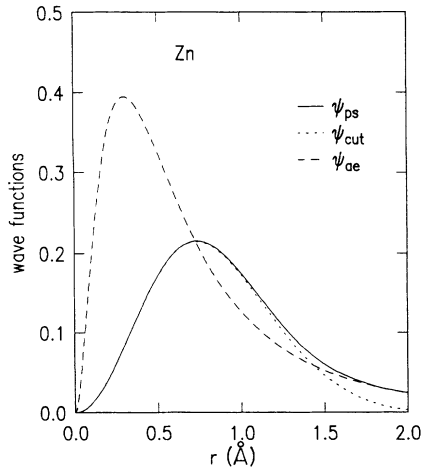

FIG. 1. Zn 3*d* wave function and pseudoatomic basis functions. The all-electron wave function is given by the dashed line. The pseudoatomic wave function is given by the solid line. The basis function (pseudoatomic wave function multiplied by a cutoff function) is given by the dotted line.

zinc *d* orbital (ten functions per zinc atom to represent five *d* orbitals). With the pseudoatomic basis functions, one orbital per state suffices (five functions per zinc atom). The total energy of a perfect ZnSe unit cell calculated with one set of pseudoatomic orbitals is lower than that calculated with two sets of Gaussians, even when the decay constants of the Gaussians are optimized. We conclude that the pseudoatomic basis functions provide a good description of the *d* states in the solid.

The reciprocal-space grid represents a parallelepiped placed about the origin in **G** space. Because of the shape of the parallelepiped, rotations that are symmetry operations for the crystal will map some **G** vectors that are inside the FFT grid into **G** vectors that are outside the grid, and vice versa. As a result, the functions represented on the FFT grid will no longer have the correct rotational symmetry. To correct this problem we set the Fourier coefficients to zero for all **G** vectors that lie outside the largest sphere that fits inside the FFT grid. This substantially reduces the number of nonzero **G** vectors. For example, a simple cubic lattice has a FFT grid that is a cube and the ratio of the volume of the inscribed sphere to the volume of the grid is $\pi/6=0.5236$. For other lattices, the ratio is even smaller. To avoid unnecessary storage, the program maps the full FFT **G**-vector grid onto a smaller **G**-vector grid containing only the points in the sphere. This grid is mapped back onto the FFT grid whenever an FFT is needed. No such reduction is possible for the real-space grid; the full FFT grid must be used in this case. Because of the asymmetry between real and reciprocal space, it is advantageous to store the functions and perform operations in reciprocal space whenever possible. Thanks to the convolution theorem and high-speed FFT routines, this can be done in most places.

The tight-binding functions used in our basis set are not orthogonal; as a result, a generalized eigenvalue problem must be solved to find the eigenstates of the Hamiltonian matrix. This is commonly done using the Cholesky decomposition.[27] Using this technique requires simultaneous storage of three matrices the size of the Hamiltonian matrix. In addition, the mixed-basis set with a large number of plane waves can suffer from overdetermination problems: the basis set may become nearly linearly dependent, which makes the generalized eigenvalue problem ill-conditioned. To avoid this problem and to save storage space, we have made the tight-binding functions orthogonal to the plane waves.[12] Using the reciprocal-space expansion of the tight-binding functions, this is done simply by setting the Fourier components of the tight-binding functions to zero for every reciprocal-lattice vector that is included in the plane-wave part of the basis set. Because the tight-binding functions are now orthogonal to all of the plane waves, and because the plane waves themselves are mutually orthogonal, the overlap matrix of the Cholesky decomposition is reduced to $n_{\mathrm{TB}}\times n_{\mathrm{TB}}$, where n_{TB} is the number of tight-binding functions in the basis set.

Although the mixed-basis scheme reduces the size of the basis set by several orders of magnitude, for a supercell calculation the Hamiltonian matrix is still on the order of 2000×2000. To reduce the computation time for the eigenvalue problem, we use group theoretical methods to block diagonalize the Hamiltonian matrix.[28] We also use an iterative diagonalization scheme to solve the eigenvalue problem.[29]

B. Test calculations

We performed a great number of calculations to test the reliability and accuracy of the programs, the basis set, and the pseudopotentials. Test calculations were performed for the two-atom unit cell of ZnSe and for a series of supercells. Because these are the first accurate pseudopotential calculations for a II-VI semiconductor, special care was devoted to these tests.

The two-atom cell tests were performed for three basic material parameters: the lattice constant a_{lat}, the bulk modulus B, and the transverse optical (TO) phonon frequency, ν_{TO}. The lattice constant and the bulk modulus are then derived from a fit of E_{Tot} (a_{lat}) for five or six different lattice constants to the Murnaghan equation of state.[30] We calculated more than 50 sets of Murnaghan equation fits, determining the lattice constant and the bulk modulus as we changed different calculational parameters. In these tests we varied such things as the form of the tight-binding functions (either pseudoatomic functions or Gaussians with different radii), the plane-wave cutoff, the size of the FFT grid, the local component, and the cutoff radii of the pseudopotentials. In all of our tests, the lattice constant was predicted to within a few percent of experiment, and the bulk modulus to within 30%. The ability of these tests to reproduce small energy differences (about 1 meV) guarantees the accuracy of our defect calculations. Based on these tests, we have chosen for our defect calculations a basis set of all plane waves

with energies up to 9 Ryd, and one set of five pseudoatomic basis functions per zinc atom. The calculated lattice constant and bulk modulus are 5.65 Å and 0.62 Mbar, compared with the experimental values of 5.67 Å and 0.63 Mbar.

The first-principles norm-conserving pseudopotentials used in this work were generated according to the Hamann-Schlüter-Chiang scheme.[24] The *s*, *p*, and *d* cutoff radii were 1.6, 1.56, and 1.01 a.u. for the zinc potential and 1.40, 1.40, and 1.51 a.u. for the selenium potential. The Zn *d* cutoff radius lies beyond the maximum of the zinc 3*d* wave function (which is at 0.56 a.u.). We tested the effects of this large cutoff radius on the pseudopotential by comparing it to a pseudopotential with a zinc *d* cutoff radius, 0.5 a.u., within the wave-function maximum. In a comparison of the bulk properties of ZnSe, the only one that was affected by the change in cutoff radius was the zone-center TO phonon frequency. The calculated TO phonon frequency using the smaller radius was 26.2 meV (=6.33 THz) compared with experimental values of 25–26 meV;[31] using the larger cutoff radius induces a 10% error in the calculated frequency. In our defect calculations, we used the larger core radius, which produces a smoother pseudopotential and pseudoatomic function. (This allows us to use a smaller FFT grid.) We have confirmed in supercell defect tests that increasing the core radius changes defect formation energies by less than 0.1 eV. We have also calculated the band structure of ZnSe, and find agreement to within 0.25 eV with the band structure calculated using all-electron methods.[32] Figure 2 shows our calculated band structure. We note that the band gap is smaller than its experimental value, due to the well-documented LDA error. The implications of this deficiency for our defect calculations will be discussed where appropriate.

We also performed a series of defect supercell tests to check the effects of the basis set, pseudopotentials, and FFT grid on defect formation energies. The error bar due to the combined effects of plane-wave cutoff, the FFT grid size, and the pseudopotential is 0.1–0.2 eV. We have also checked our results with respect to supercell convergence. Comparative tests were performed for 8-, 16-, and 32-atom supercells. We calculate the cell-size correction between an N_2-atom supercell and an N_1-atom supercell as

$$\Delta(N_2/N_1)=E^{N_2}_{\text{defect}}-E^{N_2}_{\text{perfect}} -E^{N_1}_{\text{defect}}+E^{N_1}_{\text{perfect}}\ , \tag{2}$$

where E^{N}_{defect} and E^{N}_{perfect} are the calculated total energies of an N-atom supercell with a defect and a perfect N-atom supercell, respectively. Cell-size corrections were calculated for the zinc vacancy and the zinc interstitial in different charge states. (As discussed later, these two defects are the most abundant native defects in stoichiometric *p*-type ZnSe.) Two trends emerge from these calculations. One is that the defects in the neutral charge state are well converged in a 32-atom cell, but that the charged defects (either positive or negative) may have cell-size errors of up to 0.4 eV. The second is that the correction terms are positive when going from a smaller to a larger supercell, indicating that the true defect-formation energies are larger than those calculated in the 32-atom supercell. (Cell-size corrections are not included in our results, however.) Since our main conclusion will be that the defect-formation energies are too large to allow for substantial compensation, the supercell tests strengthen our results by showing that the true formation energies are likely to be even larger than our calculated values.

III. RESULTS FOR INDIVIDUAL NATIVE POINT DEFECTS IN ZnSe

Using the methods described above, we have calculated the energies of all of the basic native point defects in ZnSe: Zn_i, Se_i (interstitials), V_{Zn}, V_{Se} (vacancies), Zn_{Se}, and Se_{Zn} (antisites). Interstitial energies were calculated

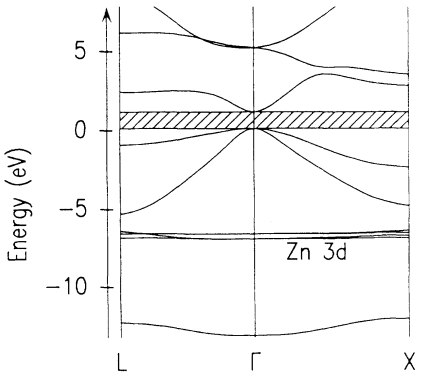

FIG. 2. ZnSe band structure. Calculated ZnSe band structure along high-symmetry directions in the Brillouin zone. The band gap is hatched. [Note that the theoretical (LDA) value of the direct band gap is 1 eV, compared to the experimental value of 2.7 eV. LDA problems are discussed in the text.] Note the set of narrow bands associated with the zinc 3*d* electrons.

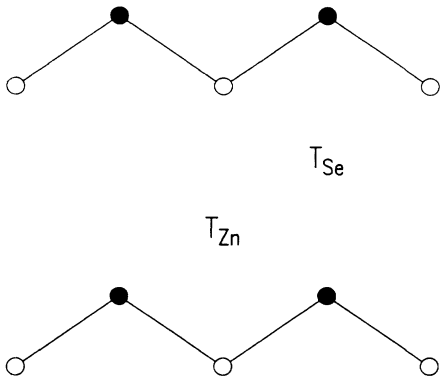

FIG. 3. Location of the two tetrahedral interstitial sites in ZnSe. Zinc atoms are represented by solid circles, selenium atoms by open circles. The *x* axis is in the [110] direction and the *y* axis in the [001] direction. Each tetrahedral site is surrounded by four atoms of the same type (two of which are out of the plane of the figure and are not shown).

at the two tetrahedral interstitial sites in ZnSe. The one site (T_{Zn}) is tetrahedrally surrounded by four Zn atoms while the other site (T_{Se}) is surrounded by four Se atoms (Fig. 3). These two sites are the most favorable interstitial sites for large atoms because they are surrounded by a large empty space. In fact, the nearest-neighbor configuration of the tetrahedral interstitial sites is the same as that of the atomic sites of the perfect crystal. Separate calculations were performed for the different charge states of each defect. Our calculated energies for each charge state of each defect (E_i) are presented in Table I.

A. Relaxation

The formation energy of a point defect can be reduced by relaxation of the atoms surrounding the defect. The lattice-relaxation energy is the energy difference between the unrelaxed defect (all atoms around the defect in their ideal lattice sites) and the relaxed defect. To find these relaxations, we must map out the total energy as a function of the positions of the surrounding atoms, and find the energy minimum. This is an arduous task because the relaxation of each atom is a function of the relaxation of the others. For practical applications, we limit the number of degrees of freedom in our relaxation calculations. The minimum of the total energy is found by calculating the total energy with different relaxations, and fitting the total-energy surface to a parabolic form about the minimum. We limit our calculations to symmetric relaxations in which each shell of atoms relaxes by the same amount ("breathing-mode" relaxations). A possible cause of nonsymmetric relaxation is the Jahn-Teller effect, which occurs when a degenerate electronic state in the band gap is partially filled with electrons.[33] Our pri-

TABLE I. Native-point-defect energies, in eV. E_i is the calculated energy for a supercell containing the defect in a 32-atom cell geometry (excluding relaxation energy). The energy of a perfect ("bulk") 32-atom supercell is $-27\,363.522$ eV. $\epsilon_i = E_i - (N_{Zn} + N_{Se})E_{ZnSe}$; ϵ_i includes the appropriate shift for charge states (referred to a state with the Fermi level at the top of the valence band). E_i and ϵ_i individually should *not* be interpreted as carrying physical meaning; in particular, they depend on the pseudopotential and on the choice of reference for the Fermi level. F_i is the formation energy for the defect in stoichiometric p-type ZnSe (doped with 10^{18} cm^{-3} Li) at $T=600$ K; it is based upon specific reference energies for individual Zn and Se atoms, calculated by solving the complete set of reaction equations, as described in the text. F_i includes the relaxation energy (which is set to 1 eV where not calculated). R_i is the calculated relaxation energy. Although F_i is a physically meaningful energy, it should not be construed as *the* formation energy of a single defect; as explained in the text, such a concept is not defined in a compound semiconductor.

Defect	Charge	n_i	E_i	ϵ_i	F_i	R_i
V_{Zn}	2−	1−	−25 906.228	598.343	2.20	0.00
V_{Zn}	1−	1−	−25 908.114	598.378	2.09	
V_{Zn}	0	1−	−25 909.881	598.532	1.81	
Zn_i (T_{Se})	0	1+	−28 809.490	−590.856	3.87	
Zn_i (T_{Se})	1+	1+	−28 813.860	−592.538	2.97	0.43
Zn_i (T_{Se})	2+	1+	−28 818.018	−594.007	1.80	0.34
Zn_i (T_{Zn})	0	1+	−28 810.117	−591.483	3.24	
Zn_i (T_{Zn})	1+	1+	−28 814.075	−592.753	2.82	0.22
Zn_1 (T_{Zn})	2+	1+	−28 817.795	−593.784	2.16	0.20
V_{Se}	0	1+	−27 099.998	−591.585	3.14	
V_{Se}	1+	1+	−27 102.872	−592.381	2.55	
V_{Se}	2+	1+	−27 105.503	−592.935	2.21	
Se_i (T_{Zn})	2−	1−	−27 609.480	604.089	6.94	
Se_i (T_{Zn})	1−	1−	−27 613.571	602.530	5.59	
Se_i (T_{Zn})	0	1−	−27 617.084	601.550	4.83	
Se_i (T_{Zn})	1+	1−	−27 620.356	600.810	4.30	
Se_i (T_{Zn})	2+	1−	−27 623.412	600.287	3.98	
Se_i (T_{Zn})	3+	1−	−27 626.280	599.950	3.86	
Se_i (T_{Zn})	4+	1−	−27 628.975	599.787	3.91	
Zn_{Se}	2−	2+	−28 542.164	−1183.563	6.46	
Zn_{Se}	1−	2+	−28 546.076	−1185.014	5.22	
Zn_{Se}	0	2+	−28 549.775	−1186.252	4.20	
Zn_{Se}	1+	2+	−28 553.036	−1187.051	3.61	
Zn_{Se}	2+	2+	−28 556.086	−1187.640	3.61	0.62
Se_{Zn}	2−	2−	−26 159.399	1199.827	6.96	
Se_{Zn}	1−	2−	−26 163.706	1197.699	5.01	
Se_{Zn}	0	2−	−26 167.784	1195.739	3.29	
Se_{Zn}	1+	2−	−26 171.377	1194.295	2.06	
Se_{Zn}	2+	2−	−26 174.689	1193.132	1.95	0.16

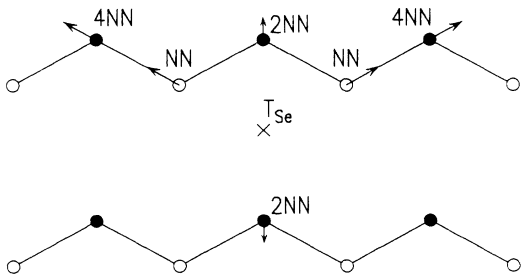

FIG. 4. Relaxations around the T_{Se} site. Outward relaxations are shown for the first-, second-, and fourth-nearest neighbors (NN, 2NN, and 4NN, respectively) around a zinc interstitial on the T_{Se} site. Four of the fourth-nearest neighbors relax in the same direction as the first-nearest neighbors. The magnitude of the relaxations is exaggerated for clarity.

mary objective is to study the behavior of defects in doped ZnSe, where defect states in the gap will either be completely full (in *n*-type material) or completely empty (in *p*-type material). Consequently, we will calculate relaxations only for defects that do not have partially filled states in the gap. For these cases, no Jahn-Teller relaxation will occur.

For substitutional site defects, we have relaxed the shell of four nearest-neighbor atoms. (The nearest-neighbor distance in ZnSe is 2.45 Å.) We have found the relaxations to be small in all cases (smaller than 0.2 Å), and the second-nearest-neighbor relaxation should be even smaller. (The second-nearest-neighbor distance is 4.01 Å.) For the tetrahedral interstitial sites, we relaxed both the first- (consisting of four atoms) and the second-neighbor (six atoms) shells simultaneously. For the relaxation of the nearest-neighbors around a tetrahedral interstitial site (T_{Zn} and T_{Se}), it turned out to be important to also include relaxations of fourth-nearest-neighbor atoms that are located on a line through the tetrahedral interstitial site and the first neighbors (Fig. 4). The reason is that a breathing relaxation of the nearest neighbors will change the length of the bond to these fourth-nearest neighbors. Since bond-stretching forces are larger than bond-bending forces, it is energetically more favorable to move the fourth-nearest neighbors outwards. This effect was not included in previous calculations.[12]

As a rule, the calculated relaxations were small: the largest relaxation energy that we found was about 0.6 eV and the typical relaxation distance was 0.1 Å, which is only 4% of the ZnSe bond length of 2.54 Å. Our calculated relaxations are listed in Table II. We will now describe our results for the individual native defects.

TABLE II. Calculated relaxations for native point defects in ZnSe. Calculated energies E_{relax} and relaxations of nearest (NN) and next-nearest (NNN) neighbors. A positive relaxation indicates relaxation outward from the defect. All relaxations in the table are symmetric. For relaxations about interstitial defects, the fourth-nearest neighbors relaxed by the same amount as the NN's.

Defect	E_{relax} (eV)	Relaxation (Å) NN	NNN
Zn_i^{+} (T_{Se})	0.43	0.11	0.05
Zn_i^{2+} (T_{Se})	0.34	0.06	0.09
Zn_i^{+} (T_{Zn})	0.22	0.09	0.03
V_{Zn}^{2-}	0.0	0.0	
Zn_{Se}^{2+}	0.62	0.2	
Se_{Zn}^{2+}	0.16	0.1	

B. Zinc self-interstitial

We start with the zinc self-interstitial (Zn_i). The neutral zinc interstitial has two electrons occupying a single level in the band gap. The possible charge states are therefore 0, 1+, and 2+, making the defect a double donor in *p*-type material. The zinc interstitial in ZnSe is a particularly interesting defect because it was the first isolated native interstitial directly observed in a semiconductor.[34] Using optically detected magnetic resonance, Rong and Watkins identified the isolated zinc self-interstitial in the 1+ charge state. The defects were produced by electron irradiation of ZnSe at a temperature of 4.2 K. They found that the interstitial occupied the T_{Se} site, and that there were no asymmetric relaxations of either the nearest-neighbor Se atoms or the second-nearest-neighbor Zn atoms. They also found the transition level from the 1+ to the 2+ charge state to occur when the Fermi level is at 1.9 eV above the valence-band edge. (This energy is the thermodynamic level in the gap.) The interstitials were observed to be mobile at temperatures above 260 K.[35] Although experiment can determine the site of the defect and its symmetry, the magnitudes of the relaxations and their energies must be determined from theory. In our calculations for the zinc self-interstitial, we have performed relaxations for the T_{Se} site in the 2+ and the 1+ charge states and for the T_{Zn} site in the 1+ charge state. The calculated relaxations are listed in Table II. The calculated valence-charge-density contours for the T_{Se} site interstitial (in the 1+ charge state) are shown in Fig. 5. Including relaxations, the energies of the 1+ charge state at the two interstitial sites are the same to within the accuracy of our calculations. Rong and Watkins actually also found a signal which they tentatively identified as the Zn self-interstitial at the T_{Zn} site. Although this defect is not stable, its energy may be only slightly higher than that of the self-interstitial at the T_{Se} site. We calculate (including relaxation energies) a value of 1.4 eV for the level in the gap between 2+ and 1+ interstitial on the T_{Se} site. The agreement with experiment is reasonable, in light of the large errors in the band gap inherent in LDA. For the 1+ T_{Se} site, Van de Walle and Laks[36] have calculated the values of the hyperfine parameters for the central Zn atom, and the first- and third-nearest-neighbor Se atoms. The hyperfine calculations included the relaxations of the neighboring atoms. The agreement between the theoretical and experimental hyperfine parameters[34] is very good. This confirms both the experimental identification of the defect and the accuracy of the calculated relaxations.

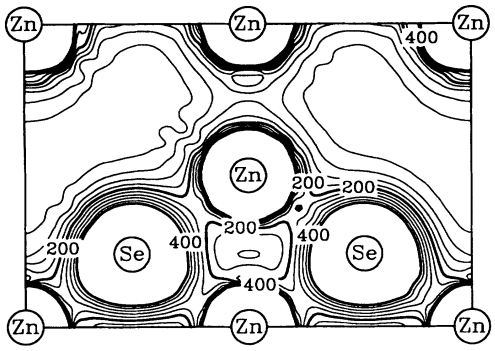

FIG. 5. Valence-charge density of the Zn self-interstitial. Contour plot of the valence-charge density around a zinc interstitial at the T_{Se} site, in the 1+ charge state. Relaxations of neighboring atoms are included. The x axis is along the [110] direction and the y axis along the [001] direction. The interstitial atom is at the center of the plot. The charge density is given in units of electrons per 32-atom cell volume ($=728.2$ Å^3) and the contour spacing is 40.

C. Zinc vacancy

The other native point defect in ZnSe that has been positively identified is the zinc vacancy. This defect was also observed in electron-irradiated ZnSe at low temperatures by Watkins.[37] The neutral zinc vacancy has a threefold-degenerate level in the band gap (with a capacity of six electrons), of which four are occupied. The possible charge states are 1− and 2−, making the vacancy an acceptor in n-type ZnSe. Watkins observed the 1− charge state using electron paramagnetic resonance and found that it undergoes a Jahn-Teller distortion. The 1− vacancy fills five electrons out of the six electron states in the gap. The remaining hole is localized by the Jahn-Teller effect on one of the four nearest-neighbor Se atoms. The Se atom with the hole moves in toward the vacancy and the symmetry of the defect is lowered from tetrahedral (point group T_d) to trigonal (C_{3v}). The energy lowering from the Jahn-Teller relaxation is estimated by Watkins[35] to be 0.35 eV. The level in the gap between the 2− and the 1− charge states is found to be at 0.66 eV above the valence-band edge. We have calculated the relaxation for the 2− charge state, which is expected to be symmetric. No relaxation (to within 0.02 Å) was found for the nearest neighbors. We did not explicitly calculate the low-symmetry relaxations for the 1− charge state. To compare with experiment, we can look at the level in the gap for the 2− to (unrelaxed) 1− transition, which we find at −0.03 eV. Adding in Watkins's estimate of the Jahn-Teller energy of the 1− vacancy (0.35 eV) brings the level to 0.32 eV. Taking the LDA deficiency into account, this value is once again in reasonable agreement with experiment (0.66 eV).

D. Other defects

There is no direct experimental evidence about the Se interstitial or the Se vacancy. The neutral Se interstitial has four electrons in a threefold-degenerate level. The possible charge states range from 4+ to 2−. Of the two tetrahedral interstitial sites, the T_{Zn} site is preferred. The neutral Se vacancy has a single level in the gap that is fully occupied by two electrons. The possible charge states are 1+ and 2+. We find that the formation energy of either of these defects is so high that they do not play any important role in ZnSe.

Nothing is known experimentally about the two antisite defects, either. Both the neutral Zn-on-Se antisite and the Se-on-Zn antisite have two electrons in a threefold-degenerate level in the gap. Possible charge states range from 2+ to 4−. For the neutral Se on Zn, we find a large lattice relaxation, in which the antisite lowers its energy by about 0.7 eV by moving about 1 Å along the ⟨111⟩ direction toward a tetrahedral interstitial site. This relaxation is favorable because it lowers the energy of the electrons in the states in the gap; it does not occur for the 2+ charge state, where the states in the gap are empty. This relaxation is similar to that found theoretically for the As-on-Ga antisite defect in GaAs.[38,39] The large lattice relaxations of the antisite in GaAs have explained the puzzling properties of the defect known as $EL2$. The occurrence of a similar relaxation in ZnSe may also be observable experimentally.

IV. DETERMINATION OF DEFECT CONCENTRATIONS

In this section we will describe how to determine the concentration of the native point defects from their calculated formation energies. Determining defect concentrations for a compound semiconductor is more difficult than for an elemental system, where the total energy of a single bulk atom is well defined. In the latter case the formation energy of a native point defect can be unambiguously determined from an N-atom defect supercell calculation: the defect formation energy is the difference between the calculated supercell energy and N times the energy of a single bulk atom. In the case of a Si self-interstitial, for instance, an extra Si atom is placed inside the crystal. This Si atom can be thought of as taken "from the surface," a process which does not change the nature of the surface; the crystal simply becomes one atom larger, and the reference energy is simply the energy of a bulk Si atom, which can be determined from a bulk calculation. This analysis cannot be applied to a compound semiconductor like ZnSe. Here, the energy of a *pair* of zinc and selenium atoms is well defined, but the energy of a *single* zinc or selenium atom depends on interactions between the crystal and its external environment. Hence energies and concentrations of the native defects will also depend on the environment.

Well-defined energies for defects in a compound semiconductor, such as ZnSe, can be determined in one of two ways. The first way is to define the energies of reactions that conserve the relative number of zinc and selenium atoms. For example, the reaction energy for forming a *pair* of zinc and selenium interstitials can be defined in the same way as the formation energy of a single silicon self-interstitial. This is true because removing a pair of zinc and selenium atoms from the surface does not

change the nature of the surface. The energy of a pair of zinc and selenium atoms can be determined from a bulk calculation. The second way to define defect formation energies is to introduce an external reservoir of zinc atoms. Zinc atoms may be added to the crystal from the reservoir, or removed from the crystal and added to the reservoir. The energy of zinc atoms in the reservoir is constant, and in thermal equilibrium with the crystal. The reservoir allows us to assign an energy to the zinc atoms. Since the sum of the zinc and selenium energies is determined by the total energy of the perfect ZnSe cell, the zinc energy determines the selenium energy. This, in turn, allows us to determine the formation energy of any defect.

The one problem with the latter prescription is that we must choose a zinc reservoir and calculate its energy. The choice of the reservoir depends on the conditions under which the crystal is grown. Instead of limiting our choice to a single reservoir energy, we picture a reservoir in which we can change the energy of the zinc atom to any value that we choose. Or, equivalently, we can set the difference between the zinc and selenium energies, δE, to be any value that we want. The formation energy for the ith defect F_i can then be expressed as

$$\begin{aligned} F_i &= E_i - N_{\mathrm{Zn}} E_{\mathrm{Zn}} - N_{\mathrm{Se}} E_{\mathrm{Se}} \\ &= E_i - (N_{\mathrm{Zn}} + N_{\mathrm{Se}}) E_{\mathrm{ZnSe}} - (N_{\mathrm{Zn}} - N_{\mathrm{Se}}) \delta E \\ &= \epsilon_i - n_i \delta E \ . \end{aligned} \tag{3}$$

Here E_i is the total energy of the supercell for the ith defect, containing N_{Zn} zinc atoms and N_{Se} selenium atoms,

$$\delta E = (E_{\mathrm{Zn}} - E_{\mathrm{Se}})/2 \ ,$$

$$E_{\mathrm{ZnSe}} = (E_{\mathrm{Zn}} + E_{\mathrm{Se}})/2 \ ,$$

$$\epsilon_i = E_i - (N_{\mathrm{Zn}} + N_{\mathrm{Se}}) E_{\mathrm{ZnSe}} \ ,$$

and $n_i = N_{\mathrm{Zn}} - N_{\mathrm{Se}}$. E_{ZnSe} is determined from a calculation of the energy of a perfect ZnSe supercell. (At $T=0$ K, E_{Zn} and E_{Se} can be identified with the chemical potentials for Zn and Se.) n_i is the number of extra Zn atoms that must be added to form the defect ($1+$ for V_{Se}, $2-$ for $\mathrm{Se_{Zn}}$, etc.), independent of the size of the supercell.

From the formation energy of the defect and its entropy S we can determine its fraction C_i by

$$\begin{aligned} C_i \equiv M_i^{\mathrm{defect}} / M_i^{\mathrm{site}} &= e^{S/k_B} e^{-F_i/k_B T} \\ &= e^{S/k_B} e^{-(\epsilon_i - n_i \delta E)/k_B T} \\ &= e^{S/k_B - \epsilon_i/k_B T} y^{n_i} = a_i y^{n_i} \ . \end{aligned} \tag{4}$$

where M_i^{defect} and M_i^{site} are the total number of defects and the total number of defect sites in the crystal, $y = \exp(\delta E / k_B T)$, and $a_i = \exp(S/k_B - \epsilon_i / k_B T)$. The concentration of defects per unit volume is found from the fraction by multiplying by the site concentration, which, for ZnSe, is 2.2×10^{22} cm^{-3}. (The site concentration is the number of atoms *of each type* per unit volume, not the *total* number of atoms per unit volume.)

The stoichiometry parameter is defined as

$$X \equiv \frac{M_{\mathrm{Se}} - M_{\mathrm{Zn}}}{M_{\mathrm{Se}} + M_{\mathrm{Zn}}} = -\tfrac{1}{2} \sum_i n_i C_i = -\tfrac{1}{2} \sum_i n_i a_i y^{n_i} \ . \tag{5}$$

M_{Zn} and M_{Se} are the total numbers of zinc and selenium atoms in the crystal. $X=0$ for perfect stoichiometry, and $X>0$ for Se-rich. The factor of $\frac{1}{2}$ enters this equation because the stoichiometry parameter is defined by dividing by the total number of atoms in the crystal, while the fractions are divided by the number of sites of each type. The stoichiometry parameter defined in this way only takes into account the deviations from perfect stoichiometry due to native point defects. In real crystals, deviations from stoichiometry may also be present because of higher dimensional defects, surfaces, and precipitates. Substitutional impurities are counted as the host species that they replace, since this replacement does not directly introduce native defects.

This formulation allows us to determine native-point-defect formation energies and concentrations for any value of δE. In practice, it is more convenient to fix the stoichiometry parameter X and determine from X the value of δE. We can do this simply by solving for y given X, using Eq. (5). The problem is essentially finding a root of a polynomial, which can be done quickly and easily using standard algorithms.[22] For our purposes, it will be clearer to talk about the stoichiometry X rather than the value of δE. However, we stress that our approach is quite generally valid for describing a system in equilibrium with other solids or gases which impose certain conditions on the chemical potential and thus determine δE.

The defect-formation energies for charged defects depend on the Fermi level. (The Fermi level is used here as the chemical potential of the electrons.) Consider a reaction in which a neutral defect D^0 with formation energy E^0 is ionized to its positive charge state D^+ with energy E^+:

$$D^0 \rightarrow D^+ + e^- \ . \tag{6}$$

The energy of this reaction is $E^0 - (E^+ + E_F)$ where E_F is the Fermi level. We can treat the combination of $D^+ + e^-$ as a single entity with energy $E^+ + E_F$. The quantity E^+ is the energy of the charge-state defect when the Fermi level is at zero. To follow the convention of choosing the zero of the Fermi level in a semiconductor at the valence-band maximum, we must change E^+ to $E^+ + E_V$ and E_F to $E_F - E_V$, E_V being the valence-band maximum. This can be generalized to any charge state: the defect energy for charge state m is $E^{(m)} + mE_F$, where we change $E^{(m)}$ to $E^{(m)} + mE_V$ to place the zero of the Fermi level at the top of the valence band.

Dealing with charged defects thus requires that we know the energy of the top of the valence band in the defect cell. The quantity that we want is (the energy of) the top of the bulk valence band in the defect cell. One cannot simply use the $\mathbf{k}=0$ band structure of the defect supercell because it includes the distortion of the band structure in the immediate vicinity of the defect. We should use the valence-band energy far away from the defect, which would correspond to a pure bulk calculation.

However, there is no absolute reference for potentials or eigenvalues in supercell calculations;[40] we therefore need additional information in order to "line up" the bulk band structure with the defect supercell. Here we use the "model solid" theory of Van de Walle and Martin,[41,42] which allows us to calculate the average electrostatic potential of a system of atoms on an absolute energy scale. Since the defect supercell and the bulk supercell in general contain a different set of atoms, the average electrostatic potentials will be shifted; the magnitude of the shift is predicted by the model solid theory. We have verified, by inspection of the locally averaged self-consistent electrostatic potential in the defect cell, that the model-solid lineup indeed provides an adequate description of the potential shift.

V. RESULTS AND DISCUSSION

In this section we present our calculated native-point-defect concentrations in ZnSe and discuss the general question of native-point-defect compensation in wide-band-gap semiconductors. As described in the preceding section, we need to know the formation energy and entropy of the defect to determine its concentration. Entropy calculations are unnecessary because our results are insensitive to values in the range $0 \le S \le 10k_B$ (see Fig. 6). By comparison, a recent accurate calculation[43,44] of the formation entropy of the Si self-interstitial found a formation entropy of $(5-6)k_B$ for the ground state. The Si self-interstitial represents an extreme case in that the ground-state configuration has low symmetry, which accounts for half of the formation entropy. It is therefore highly unlikely that the entropies for native point defects in ZnSe could be larger than $10k_B$.

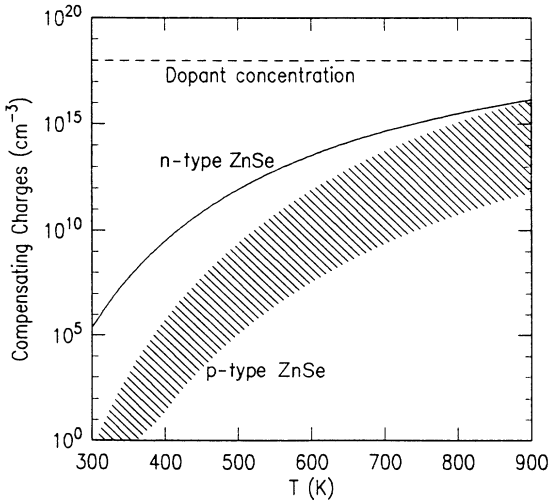

FIG. 6. Native-point-defect compensation in stoichiometric ZnSe. For p-type ZnSe the net number of electrons produced by all native point defects is shown. For n-type material, the net number of holes is shown. The range of values shown for p-type ZnSe is bounded by assuming relaxations of 1 eV and entropy of 10 k_B per defect for an upper bound and k_B for a lower bound. For n-type ZnSe, the uncertainty of the results is increased by the LDA band-gap error, and no error estimate is included.

We have explicitly calculated relaxation energies for the defects which are dominant in p-type ZnSe. As shown in Table I, the relaxation energies are all smaller than 0.7 eV, which is of the same order as calculated relaxations in other semiconductors including Si,[15,16] diamond,[21] and GaAs.[38,39,45] The defect-formation energies of other defects are high enough that, even with a generous estimate of the atomic relaxation energies (we assume 1 eV), the concentrations remain very low. Even a relaxation of 2 eV does not change our results (that is, the native-point-defect concentrations are still too low to compensate in stoichiometric material).

It is important to assess to what extent the LDA band-gap problem affects the formation energies. The band-gap problem has no direct effect on the concentrations of defects in p-type material, where electron levels in the band gap are empty. For n-type material, the position of occupied electron levels in the gap is uncertain due to the LDA band-gap error. We will treat this uncertainty by using the worst-case value of the defect energy.

Figure 6 shows the concentrations of minority carriers produced by native point defects for p-type stoichiometric ZnSe. The individual native-point-defect concentrations are given in Table III. The error bar is determined by allowing the formation entropy to range from 0 to 10 k_B. The results shown are for material with 10^{18} cm^{-3} dopants. The dopants are used to determine the position of the Fermi level. As the temperature increases, the Fermi level moves closer to the middle of the band gap. (This is because the intrinsic carrier concentration increases with temperature.) This effect slows the increase of the defect concentrations with increasing temperature. (Jansen and Sankey,[12] in their determination of the defect concentrations at $T=1658$ K, set the Fermi level at the valence-band edge. Taking into account the shift of the Fermi level with temperature would substantially lower their concentrations.) The dominant native point defects are Zn_i (a double donor) and V_{Zn} (an acceptor). At molecular-beam epitaxy growth temperatures ($T=600$ K) the concentration of minority carriers produced is less than 10^{12} cm^{-3}. For ZnSe grown at higher temperatures and not rapidly quenched, excess native point defects will be annihilated during cooling, as long

TABLE III. Native-point-defect concentrations in stoichiometric p-type ZnSe, at $T=600$ K. Only defects with concentrations greater than 10^5 cm^3 are shown. A formation entropy of 5 k_b is assumed for each defect.

Defect	Charge	Concentration (cm^{-3})
Zn_i (T_{Se})	2+	2.48×10^{9}
V_{Zn}	0	2.14×10^{9}
Se_{Zn}	2+	1.46×10^{8}
Se_{Zn}	1+	1.71×10^{7}
V_{Zn}	1−	8.70×10^{6}
V_{Zn}	2−	1.17×10^{6}
Zn_i (T_{Zn})	2+	2.21×10^{6}
V_{Se}	2+	8.58×10^{5}

as the defects are free to move. The temperature that determines the native-point-defect concentration is that at which the defects become immobile. The dominant native point defects in *p*-type material, V_{Zn} and Zn_i, are known experimentally to be mobile at temperatures above 400 and 260 *K*, respectively.[35] At 400 K the native-point-defect concentrations in *p*-type ZnSe are at most 10^5 cm^{-3}.

We have also determined the concentrations of native defects in *n*-type ZnSe (Fig. 6 and Table IV). The dominant native point defects are V_{Zn}^{2-} and Zn_{Se}^{-}. It is an experimental fact that *n*-type ZnSe can be produced much more easily than *p* type. If native-point-defect compensation were the cause, we would expect that defect concentrations would be much larger in *p*-type material than in *n* type. Instead, we find that defect concentrations are actually somewhat larger in *n*-type ZnSe. This is an additional proof that native point defects do not compensate *p*-type ZnSe. (For *n*-type ZnSe the levels in the band gap were shifted up by the LDA band-gap error, which increases the defect-formation energies. Actual defect concentrations may be higher than shown; this would further support the notion that native-point-defect compensation is no greater in *p*-type ZnSe than in *n* type.) We conclude that, in stoichiometric ZnSe, native-point-defect compensation will be insignificant.

To further support our conclusions, we have derived native-point-defect concentrations for diamond (Fig. 7) based on the first-principles defect energies of Bernholc *et al.*[21] The doping level is again 10^{18} cm^{-3}. The calculations show that only the vacancy is found in significant concentrations. Experimentally, diamond is easy to make *p* type but difficult to make *n* type. In our results for *n*-type material, we once again made a worst-case assumption about the LDA error (the LDA band-gap error is about 1 eV here): the Fermi level was shifted rigidly with the conduction-band edge, but the levels in the gap for the vacancy were not shifted at all. This assumption significantly increases the defect concentrations. The true concentrations are probably much smaller still. At a chemical-vapor-deposition-growth temperature of 1100 K, the number of holes produced in *n*-type diamond by native point defects is at most 2×10^{13} cm^{-3}. Clearly, the concentrations of native point defects in both stoichiometric ZnSe and diamond are far too low to produce significant compensation.

TABLE IV. Native-point-defect concentrations in *n*-type ZnSe. Same conditions as Table III.

Defect	Charge	Concentration (cm^{-3})
V_{Zn}	2−	1.37×10^{13}
Zn_{Se}	1−	5.23×10^{12}
Zn_{Se}	0	1.26×10^{12}
Zn_{Se}	2−	3.54×10^{11}
V_{Se}	0	3.65×10^{9}
Zn_i (T_{Zn})	0	5.07×10^{8}
Zn_{Se}	1+	6.19×10^{7}
V_{Se}	1+	1.70×10^{5}

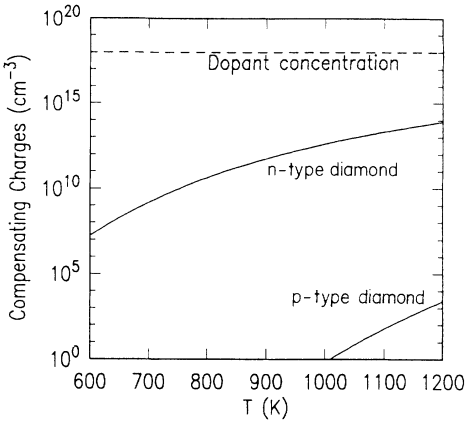

FIG. 7. Native-point-defect compensation in diamond. The native-point-defect concentrations are shown for both *n*-type and *p*-type diamond containing 10^{18} dopants. For *n*-type diamond, a worst-case treatment of the LDA band-gap error is used; the Fermi level is shifted up by the band-gap error, while the defect levels in the gap are not shifted. This gives an upper bound on the defect concentrations; actual defect concentrations in *n*-type diamond are probably much lower.

Jansen and Sankey[12] have calculated the formation energies of native defects in ZnSe and ZnTe, using pseudopotentials that treat the Zn *d* electrons as frozen-core states. From their defect-formation energies they derived defect concentrations as a function of temperature and stoichiometry. To derive defect concentrations from their calculated energies, Jansen and Sankey impose the stoichiometry parameter as an external constraint. This is equivalent to using our own method with an unknown chemical potential that produces the same stoichiometry. Their results exhibit the same trends as our own, although the actual defect concentrations are different, due in part to their approximate treatment of the *d* electrons. They also find the zinc interstitial and the selenium-on-zinc antisite defect to be the dominant defects in *p*-type ZnSe, which are both donors. In *n*-type ZnSe, they find that the zinc vacancy and the selenium antisite are dominant (both acceptors).

Jansen and Sankey suggest that their results explain why ZnSe prefers to be *p* type. Their calculated defect concentrations for *n*-type ZnTe are higher than those for *n*-type ZnSe, while their defect concentrations for *p*-type ZnSe are higher than those for *p*-type ZnTe. Based on these results, Jansen and Sankey propose that native point defects hamper the doping in *p*-type ZnSe and *n*-type ZnTe. Careful examination reveals that this conclusion is doubtful. For ZnSe, their numbers indicate that native-point-defect concentrations are 3,000 times *lower* in *p*-type material than in *n* type. Thus, if anything, native-point-defect compensation should prevent the growth of *n*-type ZnSe. Furthermore, their results were reported for a very high temperature (T=1658 K), and do not apply to the question of compensation for material that is grown at 600 K and never thermally annealed at higher temperatures. At the lower temperature, the native-point-defect concentrations derived from their

calculated energies are only $\sim 10^8$ cm^{-3}, even lower than our own predictions, and far too small to compensate doping.

Our conclusion that the concentrations of native point defects in stoichiometric ZnSe are very low does not mean that native-point-defect compensation in ZnSe never occurs. If the sample is grown with even a slight deviation from perfect stoichiometry, the concentration of native point defects will necessarily be very large, even at $T=0$ K (assuming that deviations from stoichiometry are accommodated by native point defects alone). Because the density of atomic sites in ZnSe is 4×10^{22} cm^{-3}, a deviation from stoichiometry as small as 10^{-4} implies a defect concentration of about 10^{18} cm^{-3}. Our major conclusion for nonstoichiometric material is that the native point defects that accommodate deviations from stoichiometry are *always* those that compensate the majority carriers. For *p*-type ZnSe, the dominant defect is Zn_i in Zn-rich material, and Se_{Zn} in Se-rich material; we find that both are double donors. For *n*-type ZnSe the dominant (acceptor) defects are Zn_{Se} and V_{Zn} for Zn- and Se-rich materials, respectively. Similar results were found by Jansen and Sankey.[12]

This defect structure is much richer than that used in many previous analyses of native point defects in II-VI semiconductors. Ray and Kröger,[10] for example, studied the properties of ZnSe as a function of Zn partial pressure, and analyzed their results in terms of only two native defects: V_{Zn} (an acceptor) in Se-rich material and V_{Se} (a donor) in Zn-rich material. Their model predicts that Zn-rich material should be *n* type and Se-rich material *p* type. Our results show that this model is oversimplified; changing the stoichiometry from Zn-rich to Se-rich will not convert *n*-type ZnSe to *p*-type ZnSe. Instead, the greater the deviation from stoichiometry in either direction, the greater the level of compensation.

Having addressed the native-point-defect compensation issue quantitatively, we now reexamine the notion that native-point-defect compensation increases with the width of the band gap. Let us state precisely the standard argument for this trend: For *p*-type material, imagine a prototypal compensating native-donor defect that, when neutral, introduces one electron into a state in the gap; the formation energy for this defect, E^0, is assumed not to depend on the width of the band gap. The energy gained by transferring the electron from the level in the gap E_L to the Fermi level E_F should, in contrast, increase with the width of the gap; thus, the net energy needed to form compensating defects, $E^0-(E_L-E_F)$, should decrease as the band gap increases. The flaw in this argument is that it assumes that E_L and E^0 are independent of one another. Actually, the level in the gap is *defined by* $E_L=E^0-E^+$, where E^++E_F is the (Fermi-level dependent) energy of formation of the positive charge-state defect (Fig. 8). Substituting this definition into the formula for the net energy of compensation, we find

$$E^0-(E_L-E_F)=E^0-(E^0-E^+-E_F)=E^++E_F\ ,$$

independent of the energy of formation of the neutral defect. We see that native-point-defect compensation will increase with the width of the band gap if and only if E^++E_F *decreases* with increasing band gap. For this to be true, additional assumptions would have to be made about how the formation energy of the charged defect changes as the band gap widens. In particular, there is no *a priori* reason to assume that E^+ would be lower in wide-band-gap materials. The first-principles results reported in this paper definitely show that, whatever the qualitative trends may be, the native-point-defect concentrations in stoichiometric ZnSe and in diamond are far too small to be a source of compensation.

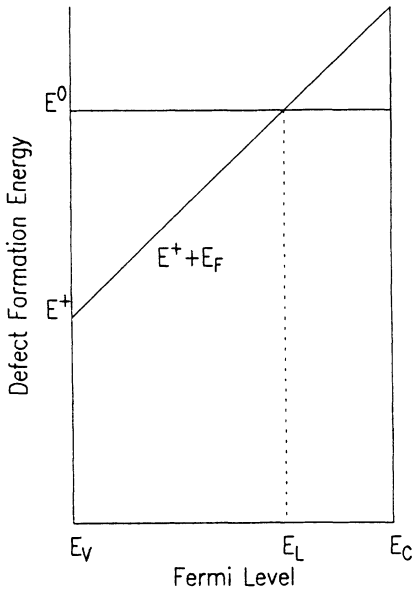

FIG. 8. Level in the gap for a donor defect. Total energy as a function of the Fermi level for the positive and neutral charge state of a prototypal donor defect. The level in the gap (E_L) is the value of the Fermi level at which the two charge states have the same energy.

VI. CONCLUSIONS

We have described a mixed-basis pseudopotential total-energy scheme which is fast and efficient enough for supercell calculations (Sec. II). These programs are capable of accurately describing the structural properties of ZnSe, including the important effects of the zinc 3*d*-electron states. We use these techniques to examine native-point-defect compensation in ZnSe; we calculate the total energies of the native point defects in ZnSe (Sec. III) and show how to extract defect concentrations from these energies (Sec. IV). We have shown that native-point-defect concentrations are very low in stoichiometric ZnSe; in nonstoichiometric material, both *n*- and *p*-type doping would be compensated (Sec. V). These results indicate that native-point-defect compensation is *not* responsible for the doping problems in ZnSe (and other wide-band-gap semiconductors). Efforts at understanding these problems should be aimed at investigating the behavior of individual dopant impurities.

ACKNOWLEDGMENTS

We are indebted to D. Vanderbilt for his iterative diagonalization program. We acknowledge helpful conversations with R. Bhargava, J.M. DePuydt, T. Marshall, J. Tersoff, and G.D. Watkins. D.B. Laks acknowledges partial support from IBM. This work was supported in part by NSF grant No. ECS-89-21159 and ONR Contract No. N00014-84-0396.

APPENDIX

The mixed-basis set results in a much smaller Hamiltonian matrix than a plane-wave basis, but requires more effort for matrix element evaluation. This appendix describes the techniques used to calculate various matrix elements. Setting up the Hamiltonian matrix for the Kohn-Sham equations requires evaluation of three types of matrix elements:[46] (1) kinetic energy and overlap, (2) local potential, V^L (pseudopotential plus screening potentials), and (3) nonlocal pseudopotential.

Each type of matrix element must be evaluated between (1) two plane waves (PW-PW), (2) two tight-binding functions (TB-TB), and (3) a plane wave and a tight-binding function (TB-PW). The PW-PW matrix elements are evaluated in the same way as they are in standard calculations with a pure plane-wave basis set.[46] We will limit our discussion to (TB-TB) and (TB-PW) matrix elements.

A tight-binding function centered on atomic site $\mathbf{T}$ of the crystal's unit cell can be written as

$$\phi_{\mathbf{k}}(\mathbf{r})=\frac{1}{\sqrt{\Omega}}\sum_{\mathbf{R}} f(\mathbf{r}-\mathbf{R}-\mathbf{T})e^{i\mathbf{k}\cdot(\mathbf{R}+\mathbf{T})} , \tag{A1}$$

where Ω is the unit cell volume, $\mathbf{R}$ is a direct lattice vector, and f is a localized real function (the pseudoatomic Zn $3d$ wave function in this work), which is of the form

$$f(\mathbf{r})=f(r)Z_{lm}(\hat{\mathbf{r}}) \tag{A2}$$

where Z_{lm} is a Bethe Kubic-harmonic function. The tight-binding function may be Fourier transformed to give

$$\phi_{\mathbf{k}}(\mathbf{G})=e^{-i\mathbf{G}\cdot\mathbf{T}}f(\mathbf{k}+\mathbf{G}) , \tag{A3}$$

$$f(\mathbf{K})=\frac{4\pi(-i)^l}{\Omega}Z_{lm}(\hat{K})\int_0^\infty j_l(Kr)f(r)dr , \tag{A4}$$

where j_l is a spherical Bessel function.

The number of TB-TB matrix elements is proportional to n_{TB}^2. (For our supercell calculations, n_{TB} is typically 80.) The TB-TB matrix elements require an integration over the crystal unit cell, making their evaluation a numerically intensive process. Instead of simply summing over a real-space grid, it is more efficient to perform these operations in reciprocal space.

1. Local matrix elements

The overlap and the kinetic-energy matrix elements are calculated in the manner of Louie, Ho, and Cohen.[25] The TB-TB matrix elements of the local potential are complicated by the presence of the $V^L(\mathbf{r})$ term. We can write the matrix element in reciprocal space as:

$$H_{ij}^L=\sum_{\mathbf{G}}\sum_{\mathbf{G}'}\phi_i^*(\mathbf{G})V^L(\mathbf{G}-\mathbf{G}')\phi_j(\mathbf{G}') . \tag{A5}$$

This formula is of order n_G^2, where n_G is the number of reciprocal-lattice vectors, and its evaluation is prohibitively expensive. To convert this expression into a more convenient form, we define

$$F_i(\mathbf{G}')=\sum_{G}\phi_i(\mathbf{G})V^L(\mathbf{G}'-\mathbf{G}) , \tag{A6}$$

so that

$$H_{ij}^L=\sum_{\mathbf{G}'}F_i^*(\mathbf{G}')\phi_j(\mathbf{G}') . \tag{A7}$$

We can now use the convolution theorem to evaluate $F_i(\mathbf{G}')$:

$$F_i(\mathbf{r})=\phi_i(\mathbf{r})V^L(\mathbf{r}) . \tag{A8}$$

This procedure is very efficient because the real-space operations are limited to n_{TB} sets of multiplications and $2n_{\mathrm{TB}}$ FFT's [for the convolution in Eq. (A8)]. The only operations that are performed $n_{\mathrm{TB}}(n_{\mathrm{TB}}+1)/2$ times are a multiplication and a summation over the reciprocal lattice [Eq. (A7)]. Only a small amount of extra computer memory is needed to store one copy of the function F_i.

2. Nonlocal matrix elements

The nonlocal pseudopotential TB-TB matrix elements can, in principle, be evaluated by applying projection operators to the reciprocal-space expansion of the tight-binding functions. Instead, we take advantage of the localized nature of the tight-binding functions by using the so-called on-site approximation. In the on-site approximation, the nonlocal pseudopotential acts only on tight-binding functions on the same site as the potential. This approximation is well justified because both the nonlocal potential and the tight-binding functions used in our calculations are very short ranged. Thus the product of the nonlocal potential on one site and the tight-binding function on a different site will be extremely small. The on-site approximation reduces the nonlocal matrix element to a single radial integral.[25]

The nonlocal matrix elements involve one additional complication. The on-site approximation is based on the assumption that the tight-binding functions are short-ranged functions. This is true only for the full tight-binding functions; the tight-binding functions used in our calculations are orthogonalized to the plane waves by setting their low-frequency Fourier components to zero. This orthogonalization leads to tight-binding functions with long-range oscillations, for which the on-site approximation is no longer valid. To correct for this we write the tight-binding functions in the following form:

$$|\phi_i\rangle=|\phi_i^F\rangle-|\phi_i^C\rangle \tag{A9}$$

where $|\phi_i\rangle$ is the actual tight-binding function of the basis set, $|\phi_i^F\rangle$ is the "full" tight-binding function before

orthogonalization, and $|\phi_i^C\rangle$ consists of the orthogonalization terms. The $|\phi_i^C\rangle$ are just the components of the full tight-binding function for all reciprocal-lattice vectors in the plane-wave basis.

$$|\phi_i^C\rangle = \sum_{|\mathbf{k}+\mathbf{G}|^2 < E_{\text{cut}}} \phi_i^F(\mathbf{k}+\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} . \tag{A10}$$

In this form, the nonlocal matrix element becomes

$$H_{ij}^{NL} = \langle \phi_i | V | \phi_j \rangle = \langle \phi_i^F | V | \phi_j^F \rangle - \langle \phi_i^C | V | \phi_j^F \rangle - \langle \phi_i^F | V | \phi_j^C \rangle + \langle \phi_i^C | V | \phi_j^C \rangle , \tag{A11}$$

where the nonlocal potential is now represented by V. The on-site approximation can now be applied to the first term. The on-site approximation is also applied to the second term:

$$\langle \phi_i^C | V | \phi_j^F \rangle = \frac{1}{\Omega} \sum_{|\mathbf{k}+\mathbf{G}|^2 < E_{\text{cut}}} \phi_i^{F*}(\mathbf{k}+\mathbf{G}) \int e^{-i(\mathbf{k}+\mathbf{G})\cdot(\mathbf{r}+\mathbf{T}_\beta)} V_{l,\beta}(r) f_i(r) Z_{lm}(\hat{\mathbf{r}}) d\mathbf{r} . \tag{A12}$$

The $\mathbf{T}_\beta$ term appears because the tight-binding function is centered about a specific atomic site, while the plane wave is defined with respect to the origin. By applying both the angular-momentum expansion of the plane waves and the orthogonality of the Kubic-Harmonic functions, we can reduce this expression to

$$\langle \phi_i^C | V | \phi_j^F \rangle = \frac{4\pi(-i)^l}{\Omega} \sum_{K^2 < E_{\text{cut}}} e^{-i\mathbf{K}\cdot\mathbf{T}_\beta} \phi_i^{F*}(\mathbf{K}) Z_{lm}(\hat{\mathbf{K}}) \int r^2 V_{l,\beta}(r) j_l(Kr) f_i(r) dr , \tag{A13}$$

where $\mathbf{K}=\mathbf{k}+\mathbf{G}$. We are again left with a set of radial integrals in real space. The third term is of the same form as (the complex conjugate of) the second term. The fourth term is treated as an expansion in terms of plane waves:

$$\langle \phi_i^C | V | \phi_j^C \rangle = \sum_{\mathbf{G}} \sum_{\mathbf{G}'} \phi_i^{F*}(\mathbf{k}+\mathbf{G}) \langle \mathbf{k}+\mathbf{G} | V^{NL} | \mathbf{k}+\mathbf{G}' \rangle \times \phi_j^F(\mathbf{k}+\mathbf{G}') . \tag{A14}$$

Because both summations are limited to reciprocal-lattice vectors in the plane-wave basis set, evaluation of the fourth term takes about the same amount of time as the nonlocal plane-wave matrix elements.

3. TB-PW matrix elements

We now describe the calculation of the TB-PW mixed-basis terms. These terms are relatively simple. The overlap and kinetic-energy matrix elements are all zero because of the orthogonalization of the tight-binding functions to the plane waves. The local-potential matrix elements between the ith plane wave and the jth tight-binding function are

$$H_{ij}^{L} = \frac{1}{\Omega} \int_{uc} e^{-i(\mathbf{k}+\mathbf{G}_i)\cdot\mathbf{r}} V^L(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r} = \frac{1}{\Omega} \int_{uc} e^{-i(\mathbf{k}+\mathbf{G}_i)\cdot r} F_j(\mathbf{r}) d\mathbf{r} = F_j(\mathbf{G}_i) . \tag{A15}$$

The function F_j is the same function that was introduced in the calculation of the local TB-TB matrix elements [Eq. (A6)]. Thus we get all of the local TB-PW matrix elements without any extra calculations. We see here the convenience of expanding the tight-binding functions in reciprocal space. For the nonlocal TB-PW matrix elements we will use the on-site approximation and an appropriate correction term once more:

$$H_{ij}^{NL} = \langle \mathbf{k}+\mathbf{G}_i | V^{NL} | \phi_j \rangle = \langle \mathbf{k}+\mathbf{G}_i | V^{NL} | \phi_j^F \rangle - \langle \mathbf{k}+\mathbf{G}_i | V^{NL} | \phi_j^C \rangle . \tag{A16}$$

The $\langle \mathbf{k}+\mathbf{G}_i |$ represents the ith plane wave. The two terms here are just the same as the third [Eq. (A13)] and fourth [Eq. A14)] terms in the TB-TB nonlocal matrix elements described above, where

$$\phi_i^F(\mathbf{k}+\mathbf{G}) = \delta_{\mathbf{G},\mathbf{G}_i} . \tag{A17}$$

Thus, these terms do not require any extra computation either.

*Present address: National Renewable Energy Laboratory, Golden, CO 80401.
†Present address: Xerox Palo Alto Research Center, Palo Alto, CA 94304.
‡Present address: IBM Zurich Laboratories, Rüschlikon, Switzerland.

[1]R. A. Reynolds, J. Vac. Sci. Technol. A 7, 269 (1989).
[2]Y. S. Park and B. K. Shin, *Topics in Applied Physics* (Springer, New York, 1977), Vol. 17, p. 133.
[3]H. Hartmann, R. Mach, and B. Selle, Curr. Top. Mater. Sci. **9**, 1 (1982).
[4]R. Bhargava, J. Cryst. Growth **59**, 15 (1982).
[5]S. Y. Ren, J. D. Dow, and S. Klemm, J. Appl. Phys. **66**, 2065 (1989).
[6]D. J. Chadi and K. J. Chang, Appl. Phys. Lett. **55**, 575 (1989).
[7]G. F. Neumark, Phys. Rev. Lett. **62**, 1800 (1989).
[8]G. F. Neumark, J. Appl. Phys. **51**, 3383 (1980).
[9]G. Mandel, Phys. Rev. **134**, A1073 (1964).
[10]A. K. Ray and F. A. Kröger, J. Electrochem. Soc. **125**, 1348 (1978).

[11]Y. Marfaing, Prog. Cryst. Growth Charact. **4**, 317 (1981).
[12]R. W. Jansen and O. F. Sankey, Phys. Rev. B **39**, 3192 (1989).
[13]H. Cheng, J. M. DePuydt, J. E. Potts, and M. A. Haase, J. Cryst. Growth **95**, 512 (1989).
[14]R. M. Park, M. B. Troffer, C. M. Rouleau, J. M. DePuydt, and M. A. Haase, Appl. Phys. Lett. **57**, 2127 (1990).
[15]R. Car, P. J. Kelly, A. Oshiyama, and S. T. Pantelides, Phys. Rev. Lett. **52**, 1814 (1984).
[16]Y. Bar-Yam and J. D. Joannopoulos, J. Electron. Mater. **14A**, 261 (1985).
[17]G. A. Baraff and M. Schlüter, Phys. Rev. Lett. **55**, 1327 (1985).
[18]C. G. Van de Walle, P. J. H. Denteneer, Y. Bar-Yam, and S. T. Pantelides, Phys. Rev. B **39**, 10 791 (1989), and references therein.
[19]T. Oguchi, T. Sasaki, and H. Katayama-Yoshida, in *Impurities, Defects, and Diffusion in Semiconductors: Bulk and Layered Structures,* edited by D. J. Wolford, J. Bernholc, and E. E. Haller, MRS Symposia Proceedings No. 163 (Materials Research Society, Pittsburgh, 1990), p. 81.
[20]S.-H. Wei and A. Zunger, Phys. Rev. B **37**, 8958 (1988).
[21]J. Bernholc, A. Antonelli, T. M. Del Sol, Y. Bar-Yam, and S. T. Pantelides, Phys. Rev. Lett. **61**, 2689 (1988).
[22]W. H. Press, B. P. Flannery, S. A. Teukolsy, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986).
[23]P. Hohenberg and W. Kohn. Phys. Rev. **136**, B864 (1964).
[24]D. R. Hamann, M. Schluter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[25]S. G. Louie, K.-M. Ho, and M. L. Cohen, Phys. Rev. B **19**, 1774 (1979).
[26]D. B. Laks, Ph.D. thesis, Columbia University, 1990.
[27]R. S. Martin and J. H. Wilkinson, Numer. Math. **11**, 99 (1968).
[28]J. C. Slater, *Quantum Theory of Molecules and Solids* (McGraw Hill, New York, 1965), Vol. 2.
[29]R. Natarajan and D. Vanderbilt, J. Comput. Phys. **82**, 218 (1989).
[30]F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **30**, 244 (1944).
[31]H. E. Gumlich, D. Theis, and D. Tschierse, in *Numerical Data and Functional Relationships in Science and Technology,* edited by O. Madelung, Landolt-Börnstein, New Series, Group III, Vol. 17, Pt. b (Springer-Verlag, Berlin, 1982), p. 126.
[32]A. Continenza, S. Massidda, and A. J. Freeman, Phys. Rev. B **38**, 12 996 (1988).
[33]J. Calaway, *Quantum Theory of the Solid State* (Academic, New York, 1974).
[34]F. Rong and G. D. Watkins, Phys. Rev. Lett. **58**, 1486 (1987).
[35]G. D. Watkins, in *Defect Control in Semiconductors, Proceedings of the International Conference on the Science and Technology of Defect Control in Semiconductors, Yokohama, 1989,* edited by K. Sumino (North-Holland, Amsterdam, 1990), p. 933.
[36]C. G. Van de Walle and D. B. Laks, in *Proceedings of 20th International Conference on the Physics of Semiconductors, Thessaloniki, 1990,* edited by E. M. Anastassakis and J. D. Joannopoulis (World Scientific, Singapore, 1990), p. 722.
[37]G. D. Watkins, in *Radiation Defects in Semiconductors 1976,* Proceedings of the International Conference on Radiation Effects in Semiconductors, edited by N. B. Urli and J. W. Corbett, IOP Conf. Proc. No. 31 (Institute of Physics and Physical Society, London, 1977), p. 95.
[38]J. Dabrowski and M. Scheffler, Phys. Rev. Lett. **60**, 2183 (1988).
[39]D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **60**, 2187 (1988).
[40]L. Kleinman, Phys. Rev. B **24**, 7412 (1981).
[41]C. G. Van de Walle and R. M. Martin, Phys. Rev. B **35**, 8154 (1987).
[42]C. G. Van de Walle, Ph.D. thesis, Stanford University, 1986.
[43]P. E. Blöchl and S. T. Pantelides (unpublished).
[44]P. E. Blöchl, D. B. Laks, S. T. Pantelides, E. Smargiassi, R. Car, W. Andreoni, and M. Parrinello, in Proceedings of the 20th International Conference on the Physics of Semiconductors (Ref. 36).
[45]D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).
[46]J. Ihm, A. Zunger, and M. L. Cohen. J. Phys. C **12**, 4409 (1979).