Approximation to density functional theory for the calculation of band gaps of semiconductors

Luiz G. Ferreira*
Instituto de Física, Universidade de São Paulo, Caixa Postal 66318, 05315-970 São Paulo, São Paulo, Brazil

Marcelo Marques† and Lara K. Teles‡
Instituto Tecnológico de Aeronáutica, 12228-900 São José dos Campos, São Paulo, Brazil
(Received 20 May 2008; revised manuscript received 29 August 2008; published 30 September 2008)

The local-density approximation (LDA) together with the half occupation (transition state) is notoriously successful in the calculation of atomic ionization potentials. When it comes to extended systems, such as a semiconductor infinite system, it has been very difficult to find a way to half ionize because the hole tends to be infinitely extended (a Bloch wave). The answer to this problem lies in the LDA formalism itself. One proves that the half occupation is equivalent to introducing the hole self-energy (electrostatic and exchange correlation) into the Schrödinger equation. The argument then becomes simple: The eigenvalue minus the self-energy has to be minimized because the atom has a minimal energy. Then one simply proves that the hole is localized, not infinitely extended, because it must have maximal self-energy. Then one also arrives at an equation similar to the self-interaction correction equation, but corrected for the removal of just 1/2 electron. Applied to the calculation of band gaps and effective masses, we use the self-energy calculated in atoms and attain a precision similar to that of GW, but with the great advantage that it requires no more computational effort than standard LDA.

DOI: 10.1103/PhysRevB.78.125116 PACS number(s): 71.15.−m, 31.15.−p, 71.20.Mq

I. INTRODUCTION

The well-known *density functional theory* (DFT) (Ref. 1) is an approach to the theory of electronic structure in which the electron-density distribution, rather than the many-electron wave function, plays a central role. The practical applications of DFT are based on approximations for the so-called exchange-correlation potential, which describes the effects of the Pauli principle on the many-electron system. If we had the exact exchange-correlation potential, we could solve the many-body problem exactly for the ground state. Although the potential is unknown, approximations are made. The most common is the so-called *local-density approximation* (LDA), which locally uses the exchange-correlation energy density of a homogeneous system. The LDA to the Kohn and Sham DFT (Ref. 1) is still one of the most reliable methods for condensed-matter calculations, having successfully predicted and explained a wide range of ground-state properties in solid-state physics and chemistry.[2] Lately, but very slowly, it is being progressively abandoned in favor of the many generalized gradient approximations (GGAs).[3] However, while LDA and GGA have predicted many ground-state properties with good accuracy, the electronic properties such as band gaps are significantly smaller than those from experiment. These discrepancies are caused by the lack of the discontinuity of the exchange-correlation potential[2] in going from the valence to the conduction band. Several methods for overcoming these limitations have been proposed. One of them is the GW approximation, in which one considers the energies of quasiparticles and calculate the electron self-energy in terms of perturbation theory.[4,5] This procedure has been quite successful, achieving good accuracy, but it goes beyond the DFT. Other procedures were also proposed, among them we can mainly cite the self-interaction correction (SIC) (Ref. 6); the atomic SIC (ASIC) applied to solids,[7–9] which is perhaps the procedure closest to ours; hybrid functionals;[10] screened exchange LDA (SX-LDA) (Ref. 11); the so-called exact-exchange approach;[12] the well-known LDA+U (Ref. 13); and the work of Liberman[14] and others. Most of these approaches are computationally very demanding, which prohibits their application to large systems of atoms.

The Slater half-occupation scheme[15–17] was very successful for valence states. One example is that one could obtain energies which were comparable to the experimental ionization energies,[17] though, at that time, good spin-polarized exchange-correlation approximations, as those based on the approach of Ceperley and Alder, did not exist.[6] In order to illustrate to the reader the quality of the results that can be obtained, we present in Table I the first and second ionization potentials of 12 atoms, measured and calculated with LDA with 1/2 occupation.

Though the precision of the calculated results shown in Table I is much better than the precision one reaches in the calculation of band gaps, either by LDA or GGA, it has been difficult to find a way to make the ionization of 1/2 electron in extended systems as crystals. Of course the problem is that a crystal is described by means of Bloch waves and removing the population of just one Bloch state is of no consequence. In this paper we present a solution to this problem. We are especially concerned with the calculation of band gaps in semiconductors, for which we obtain calculated results that compare very favorably with experiment and are not computationally demanding. We report the results for 14 semiconductors, including groups II-VI, III-V, and IV. Our method is inspired by the LDA and by the half ionization but, at some point, it has to be postulated. The quality of the results and the ease with which they are obtained show that our assumptions are very good. Now we develop our method, which could be properly named as LDA-1/2.

TABLE I. First and second ionization potentials (IPs) for some atoms (eV). These results were obtained with spin polarization but assuming spherical charge densities for ions and atoms. We used a code originally written by Froyen, modified by Troullier and Martins, and modified and maintained by Garcia.

	First IP		Second IP	
Atom	Calc.	Expt.	Calc.	Expt.
C	11.60	11.26	24.58	24.38
N	14.81	14.53	30.01	29.60
O	13.89	13.62	35.38	35.12
Al	5.94	5.99	18.97	18.83
Si	8.19	8.15	16.30	16.35
P	10.44	10.49	19.80	19.73
S	10.57	10.36	23.25	23.33
Zn	9.70	9.39	18.65	17.96
Ga	6.00	6.00	20.83	20.51
Ge	7.99	7.90	15.88	15.93
As	9.90	9.81	18.63	18.63
In	5.73	5.78	18.56	18.97

II. LDA AND HALF IONIZATION IN SOLIDS

Accepting the LDA as a valid approximation to the DF, the theorem of Janak[19] follows,

$$\frac{\partial E}{\partial f_\alpha} = e_\alpha(f_\alpha), \tag{1}$$

where E is the total energy of the system and f_α is a function of the occupation of the one-particle Kohn and Sham state α. It is a well-known fact that the eigenvalue $e_\alpha(f_\alpha)$ is almost precisely linear with the occupation f_α.[17] Then integrating

$$\int_{-1}^{0} df_\alpha$$

between the ground state (f_α=0) and the ion (f_α=−1), one obtains

$$E(0) - E(-1) = e_\alpha(-1/2) = -\text{ ionization potential}. \tag{2}$$

At this point we must explain that f_α=0 means the occupied one-electron state of the neutral ground state, f_α=−1 means the state depleted of one electron, and f_α=−1/2 refers to the state of the half ion depleted of 1/2 electron. Taking another derivative,

$$\frac{\partial e_\alpha}{\partial f_\alpha} = 2S_\alpha, \tag{3}$$

where

$$S_\alpha = \int\int d^3r d^3r' \frac{n_\alpha(\vec{r}) n_\alpha(\vec{r}')}{|\vec{r}-\vec{r}'|} + \frac{1}{2}\int\int d^3r d^3r' n_\alpha(\vec{r}) \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')} n_\alpha(\vec{r}') + \int\int d^3r d^3r' \frac{n_\alpha(\vec{r})}{|\vec{r}-\vec{r}'|} \sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha} + \frac{1}{2}\int\int d^3r d^3r' n_\alpha(\vec{r}) \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')} \sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha} \tag{4}$$

is named "self-energy" because of the first term in the right. In LDA the functional derivatives become common derivatives times delta functions. We maintain the functional derivative notation because the final formulas can have extended use. Because of the linearity of $e_\alpha(f_\alpha)$,[17] we may write

$$e_\alpha(-1/2) = e_\alpha(0) - S_\alpha \tag{5}$$

and

$$E(0) = E(-1) + e_\alpha(0) - S_\alpha. \tag{6}$$

Equation (6) is quite surprising in its simplicity. The equation tells us that to restore the ground state, with total energy $E(0)$, from an ion with a hole at state α we add an electron whose energy is the eigenvalue $e_\alpha(0)$ minus the hole self-energy. The self-energy is large when the function is much localized as an atomic wave function, and it is small and zero when it is much spread as a Bloch function. Since the energy of the restored ground state must be a minimum, the hole self-energy must be a maximum. Thus the hole should be representable by a very localized wave function. *This is a demonstration of the hole localization*, though this proof is based on an approximation (LDA) to the DF theory and on the linearity assumption. Of course we cannot say that the localized hole state is truly stationary, especially if its energy is inside the band continuum of the Bloch states, into which the localized hole would be scattered.

The self-energy may be thought of as the quantum-mechanical average of a "self-energy potential" $V_S(\vec{r})$ such that

$$S_\alpha = \int d^3r n_\alpha(\vec{r}) V_S(\vec{r}), \tag{7}$$

where $n_\alpha = \psi_\alpha^* \psi_\alpha$ and

$$V_S(\vec{r}) = \int d^3r' \frac{n_\alpha(\vec{r}')}{|\vec{r}-\vec{r}'|} + \frac{1}{2}\int d^3r' \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')} n_\alpha(\vec{r}') + \int d^3r' \frac{\sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha}}{|\vec{r}-\vec{r}'|} + \frac{1}{2}\int d^3r' \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')} \sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha}, \tag{8}$$

depends on the state α. From now on, in equations such as Eq. (8), we will not write the last two terms, those depending

on the derivative of the wave functions with respect to the occupation f_α.

To derive Eq. (6) we assumed linearity, aside from the Janak theorem. The linearity results when the Kohn and Sham eigenfunctions of the ground state are equal to those of the ion, which is correct to a large extent.[17] Coherently we may neglect the last two terms in equations such as Eq. (8), and make the difference $E(0)-E(-1)$ between two minima an extremum. Then we minimize (extremize) $e_\alpha(0)-S_\alpha$ as suggested in Eq. (6). To do so we must write a variational expression that, upon minimization (extremization), leads to a differential equation for the hole wave function $\psi_\alpha(-1/2,\vec{r})$. At this point we must set clearly what we imply by the term "hole." From Eq. (6) we see that we are adding an electron to a hole state of the ion. The hole state might be in a valence or conduction band, the only requirement being that the state is empty in the ion. Thus by a "hole" we mean "an electron filling an empty state" or "a particle excitation." This particle excitation may be in the valence or in the conduction band.

The expression $e_\alpha(0)-S_\alpha$ to be minimized is mixed in the sense that the first term is an average for the wave function $\psi_\alpha(0,\vec{r})$ of the neutral ground state, $\langle\psi_\alpha(0)|H_0|\psi_\alpha(0)\rangle$, while the second term S_α is the self-energy of the hole state $\psi_\alpha(-1/2,\vec{r})$ belonging to the half ion. Then, as our first attempt at a variational expression, we write

$$e_\alpha - S_\alpha \simeq \langle\psi_\alpha(-1/2)|-\nabla^2-2\sum_I\frac{Z_I}{|\vec{r}-\vec{r}_I|}+2\int\frac{n_0(\vec{r}')}{|\vec{r}-\vec{r}'|}d^3r' + \frac{\delta E_{\rm xc}}{\delta n_0(\vec{r})}-V_S|\psi_\alpha(-1/2)\rangle, \tag{9}$$

where $n_0(\vec{r})$ does *not* include the hole wave function $\psi_\alpha(-1/2,\vec{r})$ because it comes from the one-particle Hamiltonian H_0 of the neutral ground state.

Later we will describe a more convenient all-electron variational expression, instead of the one-electron expression in Eq. (9). For the time being, aiming at comparing our method with SIC, we perform the extremization and then insert Eq. (8) to obtain the equation below with the top entries in the brackets { }:

$$\left[-\nabla^2-2\sum_I\frac{Z_I}{|\vec{r}-\vec{r}_I|}+2\int\frac{n_0(\vec{r}')}{|\vec{r}-\vec{r}'|}d^3r'+\frac{\delta E_{\rm xc}}{\delta n_0(\vec{r})} - \begin{Bmatrix}1\\2\end{Bmatrix}\int\frac{\psi_\alpha(-1/2,\vec{r}')^*\psi_\alpha(-1/2,\vec{r}')}{|\vec{r}-\vec{r}'|}d^3r' - \begin{Bmatrix}1/2\\1\end{Bmatrix}\int\frac{\delta^2E_{\rm xc}}{\delta n_0(\vec{r})\delta n_0(\vec{r}')}\psi_\alpha(-1/2,\vec{r}')^* \times\psi_\alpha(-1/2,\vec{r}')d^3r'\right]\psi_\alpha(-1/2,\vec{r})=\lambda_\alpha\psi_\alpha(-1/2,\vec{r}). \tag{10}$$

If we insert Eq. (8) before extremization, we obtain Eq. (10) with the bottom entries in { }. In this latter case, the next to last term of the operator is exactly the term in the SIC equation.[6] Its effect is to exclude the electron being considered from the Hartree interaction. The last term in the operator, the exchange-correlation term, is very different from the corresponding SIC term, since it depends on the whole density of the system and not only on the density of the α state. The SIC equation, Eq. (10) with the bottom entries in { }, is not what we want because the eigenvalue $\lambda_\alpha=e_\alpha-2S_\alpha$ and not $\lambda_\alpha=e_\alpha-S_\alpha$ as the half ionization requires. It is worth mentioning that in the calculation of band gaps, SIC overcorrects and halving it seems to be a better procedure.[9]

TABLE II. Atomic self-energies (eV) [Eq. (7)] for some third-row atoms and different valence electrons, calculated with self-energy potentials [Eq. (18)] derived from half ionization of different states. The fact that the self-energy of a given state has no important dependence on the self-energy potential (that is, the entries in each column for a given atom do not differ much) shows that, for practical purposes, one can neglect the state dependence of the self-energy potential. In the table below, only the self-energies of *d* states seem to be much dependent on the way the self-energy potential is derived.

Half ionized		Self-energy (eV)		
Valence state		*s*	*p*	*d*
Zn	*s*	3.63		4.62
	d	4.62		7.43
Ga	*s*	3.15	3.67	4.67
	p	3.44	3.02	3.91
	d	4.75	4.02	8.12
Ge	*s*	4.13	3.71	5.03
	p	3.70	3.35	4.22
	d	5.11	4.30	9.00
As	*s*	4.11	3.73	5.26
	p	4.26	3.93	4.73
	d	5.33	4.57	9.79

Except for atoms, solving Eq. (10) is very difficult, either with the top or bottom entries in { }. One important problem is that the solutions of Eq. (10) are not orthogonal. The SIC solution for atoms is used in the ASIC method, which is excellently reviewed in Ref. 9. In our case we proceed differently. We introduce a parametrized self-energy potential and use a variational expression that is an extremum for variations in the parameter(s). The first question to be answered is whether it is possible to define a unique self-energy potential that is state independent. In Table II we show a study of how the atomic self-energy of many states vary with the assumed self-energy potential. One sees that for *s* and *p* orbitals, the self-energy does not vary much whether it is calculated with *s*, *p*, and even *d* self-energy potentials. Of course the self-energy potential that we will use is the one corresponding to the atomic orbital dominating the crystal energy bands around the gap. In extreme cases we can define a self-energy potential that is angular momentum dependent,

$$V_S=\sum_l V_{S,l}(r)\sum_{m=-l}^{l}|l,m\rangle\langle l,m|. \tag{11}$$

This possibility was explored in the case of diamond, as shown in the discussion of our results.

Thus, assuming a self-energy potential that is state independent, we use the following variational expression:[20,21]

$$E[n,v,\rho] = K[n] - \int V[p]\rho + \frac{1}{2}\int V[\rho]\rho - \int V_S\rho + \int v(n-\rho) + E_{\rm xc}[\rho], \tag{12}$$

where p is the proton number density, $n=\Sigma_\beta f_\beta \psi_\beta \psi_\beta^*$ is the electron number density made out of the squares of the wave functions, f_β are the occupation numbers, v is the Kohn-Sham potential, ρ is the model density, V_S is the given parametrized self-energy potential, K and $E_{\rm xc}$ have their usual meaning of kinetic and exchange correlations of the Kohn-Sham DFT. The functional $V[\rho]$ is defined as

$$V[\rho(\vec{r})] = 2\int \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|}d^3r'. \tag{13}$$

The functional E is an extremum for variations in any of the three functions n, v, and ρ:

(1) $\delta E/\delta v=0$ leads to $\rho=n$.

(2) $\delta E/\delta\rho=0$ leads to

$$v = -V[p] + V[\rho] - V_S + \delta E_{xc}/\delta\rho. \tag{14}$$

(3) $\delta E/\delta n=0$ leads to Schrödinger equations with potential v and eigenvalues e_α, rewritten as

$$\left[-\nabla^2 - 2\sum_I \frac{Z_I}{|\vec{r}-\vec{r}_I|} + 2\int \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|}d^3r' + \frac{\delta E_{\rm xc}}{\delta\rho(\vec{r})} - V_S(\vec{r})\right]\psi_\alpha(\vec{r}) = e_\alpha\psi_\alpha(\vec{r}). \tag{15}$$

Using Eq. (14) to determine ρ for given v and V_S and solving the Schrödinger equations, we find

$$E = \sum_\beta f_\beta e_\beta - \frac{1}{2}\int V[\rho]\rho + E_{\rm xc}[\rho] - \int \rho\frac{\delta E_{\rm xc}}{\delta\rho}. \tag{16}$$

It must be understood that both n and ρ are number densities of N electrons, not $N-1/2$. But the eigenvalues correspond to a situation where 1/2 electron is removed if V_S is well chosen.

We want to find band gaps by taking the difference in total energies due to different occupations f_α. Maintaining the Kohn and Sham potential v and the model density ρ, solution of Eq. (14), the band gap becomes a difference between eigenvalues e_α. Now, because the total energy is a variational functional, that is, an extremum for variations in v, resulting from variations of the self-energy potential V_S, one should look for extreme eigenvalue differences,

$$\frac{\delta(e_\alpha - e_\beta)}{\delta V_S} = 0. \tag{17}$$

III. LDA-1/2 METHOD

Consider the case of an atom. We first prove that the self-energy potential is given by

$$V_S \simeq -V(-1/2,r) + V(0,r), \tag{18}$$

namely, the difference between the all-electron potentials of the atom and of the half-ion.

We begin by writing the potential difference as

$$\begin{aligned} V(-1/2,r) - V(0,r) &= \int_0^{-1/2} df_\alpha \frac{\partial}{\partial f_\alpha}\left\{-2\frac{Z}{r} + 2\int d^3r' \frac{n(\vec{r}')}{|\vec{r}-\vec{r}'|} + \frac{\delta E_{\rm xc}}{\delta n(\vec{r})}\right\} \\ &= \int_0^{-1/2} df_\alpha \left\{2\int d^3r' \frac{n_\alpha(\vec{r}')}{|\vec{r}-\vec{r}'|} + \int d^3r' \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')} n_\alpha(\vec{r}')\right\} \\ &\quad + \int_0^{-1/2} df_\alpha \left\{2\int d^3r' \frac{\sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha}}{|\vec{r}-\vec{r}'|} + \int d^3r' \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')}\sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha}\right\}, \end{aligned} \tag{19}$$

or for a certain value of f_α in $[-1/2,0]$

$$\begin{aligned} -V(-1/2,r) + V(0,r) &= \int d^3r' \frac{n_\alpha(\vec{r}')}{|\vec{r}-\vec{r}'|} + \frac{1}{2}\int d^3r' \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')} n_\alpha(\vec{r}') \\ &\quad + \int d^3r' \frac{\sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha}}{|\vec{r}-\vec{r}'|} + \frac{1}{2}\int d^3r' \frac{\delta^2 E_{\rm xc}}{\delta n(\vec{r})\delta n(\vec{r}')}\sum_\beta f_\beta \frac{\partial n_\beta(\vec{r}')}{\partial f_\alpha}. \end{aligned} \tag{20}$$

Figure 1 depicts r times the self-energy potential for the nitrogen atom, a typical case, for degrees of ionization I ranging from 0.5 to −0.2. Observe that the ratio

$$\frac{V(f_\alpha,r) - V(0,r)}{f_\alpha} \tag{21}$$

has a very poor dependence on f_α, meaning that in Eq. (20) we can take the exchange-correlation (xc) functionals at the full occupation $f_\alpha=0$. Then, comparing Eqs. (8) and (20), our proof is completed.

We will leave to another paper a discussion on the lone hole solution we can get out of Eq. (10). Here we are interested in calculating band gaps of semiconductors. For that purpose we will repeat the atomic self-energy potential [Eq.

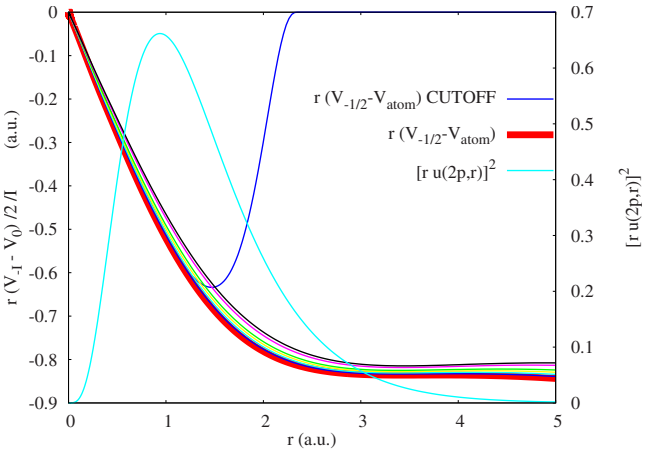

FIG. 1. (Color online) Self-energy potential (rV_S) calculated for the N atom at different ionizations I ranging from 0.5 to −0.2. The lines bunch around that of I=0.5, which is made thicker. The wave function $u(2p,r)$ for the ionized state ($2p$) is also shown. Also shown is the potential after the cutoff by $\Theta(r)$.

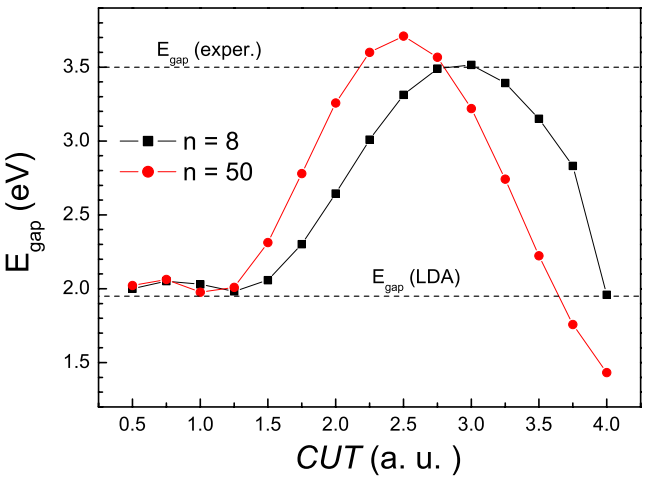

FIG. 2. (Color online) Band gap of GaN as function of the parameter CUT of the cutoff function $\Theta(r)$ applied to the self-energy potential of N $2p$. The band-gap extreme values depend to some extent on the exponent n being used (n=8 and 50 are shown).

(20)] in the whole lattice and calculate eigenvalues for "hole bands." But observe that the first term on the right of Eq. (20), when repeated in the whole crystal, diverges because it is Coulomb type. On the other hand, as Fig. 1 shows, the Coulomb tail of the atomic V_S has no importance because the wave function never goes far. Then, in using the self-energy potential V_S defined in the atoms, we first trim the potential with a function as

$$\Theta(r) = \begin{cases} \left[1 - \left(\frac{r}{CUT}\right)^n\right]^3, & r \le CUT \\ 0, & r > CUT. \end{cases} \quad (22)$$

The idea is that the atomic self-energy potential is meaningful only where the atomic wave function is not negligible. Of course the trimmed self-energy potential is repeated throughout the infinite crystal, so that we are actually calculating "filled hole bands." Due to the trimming, the Coulomb tail (of −1/2 electrons) of the atomic V_S does not penetrate into the neighboring atoms. With the trimming, the eigenvalues e_α in Eq. (15) become dependent on the trimming parameter CUT. However, Eq. (17) sets a recipe for choosing the value of CUT: one should make the energy gaps extreme.

The function in Eq. (22) has some important properties: (1) Its derivative is also zero at r=CUT, so that its electric field is zero at that point and the cutoff does not add to the total charge of the atom. (2) The trimmed self-energy potential $\Theta(r)V_S(r)$ is wholly contained inside a sphere of radius CUT, which facilitates its use in band-calculation methods such as SIESTA and augmented plane wave–like (APW-like). The power n should be large so that the cutoff is sharp. In actual practice we tried n=8 and n=50 with similarly good results. Thus we adopted n=8, which is less abrupt and does not introduce numerical problems into the programs. Figure 2 shows a typical behavior of a band gap as function of the parameters defining the cutoff function. The first increase in the band gap with CUT only means that we are getting more of the valence-band self-energy as the cutoff is made at larger radii. In principle the larger CUT is, the more we get of the valence self-energy. After reaching a peak, the gap decreases because: (i) the potential V_S is penetrating into neighboring atoms, tending to a uniform negative potential everywhere in space, shifting downward all bands, valence and conduction alike; and (ii) the self-energy potential perturbation, being broad, diffuses the excitation wave function, thus making it to loose locality and self-energy. In other words, the cutoff function should be broad enough so as to include most of the excitation wave function and thin enough so as not spread it. Thus, the procedure for determining CUT is based on Fig. 2, namely, we look for the extreme band gap according to Eq. (17).

For a given atom and bonding type, the value of CUT depends little on the chemical environment. Because we are using CUT values that make the gaps extreme, small deviations from the optimal values produce only second-order deviations in the gaps. In Fig. 3 we present the values of anion p-state CUT optimized for arsenides, phosphides, and nitrides. The anion CUT value has a small dependence on the chemical environment, which is approximately linear with the compound bond length. However, the relative variation produced in the energy gap is very small. This can be easily verified in Fig. 2. Around the energy-gap maximum, the range of CUT found in the optimization for all nitrides leads to a change of only 0.05 eV in the energy-gap value. This behavior was also verified for all other calculated compounds. Therefore, we conclude that is very reasonable to consider the same CUT value for the anion potentials. In Table III we show the optimal values of the parameter CUT of the trimming function Θ in Eq. (22). The values for CUT in the table reminds one of a table of ionic, covalent, or atomic radii, but are not equal to any.

IV. RESULTS AND DISCUSSION

We calculated, within the LDA-1/2 approach, the electronic structure for several semiconductors. By comparing

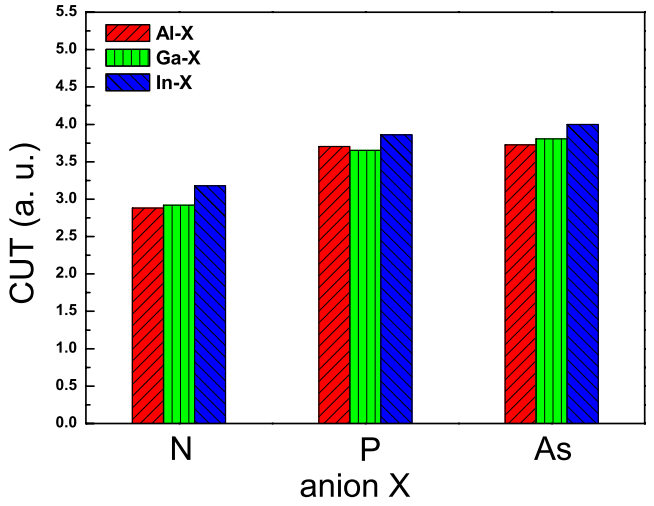

FIG. 3. (Color online) The optimized values of the parameter *CUT*, in which we correct only the anion *p* state for nitrides, phosphides, and arsenides compounds.

the LDA and LDA-1/2 calculation procedures, the LDA-1/2 calculations lead to not more computational effort than standard LDA, because the values of *CUT* depend only on the atoms, not on their environment, and are calculated just once. This is a great advantage of our method. Most of the calculations were made with the code Vienna *Ab initio* Simulation Package (VASP) using the ultrasoft pseudopotential[22,23] to which we added the trimmed self-energy potential. In some instances we repeated the calculations with the SIESTA code[24] and the results differed by no more than 0.1 eV. The two codes are so different, for they use different basis functions and pseudopotentials, that the agreement of their results runs in favor of the reliability of our LDA-1/2 procedure. The **k**-space integrals were approximated by sums over a $9\times9\times9$ special point of the Monkhorst-Pack type within the irreducible part of the Brillouin zone.[25] The number of plane waves for the expansion of wave functions was optimized for each system, and it was basically the same value obtained for optimization of the equivalent standard LDA calculation. The lattice parameters used were the experimental ones.

TABLE III. Values of *CUT* that make the band gaps extreme, that is, when the self-energy potential is defined by Eq. (18) and trimmed by Eq. (22). The optimal value of *CUT*, as is the case of an ionic or covalent radius, is typical of each atom and the orbital that was half ionized. In most cases only the anion matters.

Atom	Half ionized Orbital	*CUT* (a.u.)
Si	*p*	3.67
N	*p*	2.90
As	*p*	3.81
O	*p*	2.67
Ga	*d*	1.23
Ge	*p*	3.46
P	*p*	3.86
Zn	*d*	1.665
S	*p*	3.39
In	*d*	2.126

The present LDA-1/2 proposal assumes that in promoting an electron from the valence band to the conduction band, the hole thus created is similar to the hole created in the atomic photoionization. In other words, the hole has the extent of an atomic hole. If the hole in the extended system overlapped N equal atoms, its self-energy would be $1/N^2$ that of the atomic hole, and the self-energy potential V_S would be $1/N$ that of the atom, or it would have to be calculated with an ion with $1/2N$ electrons removed. The results for the semiconductors III-V and II-VI, to be presented shortly, definitely point to $N=1$, meaning that the hole in the solid resembles much the hole in the atom. In fact, the valence band of these semiconductors is known to be made of the anion wave functions. On the other hand, for the IV elements Ge and Si (and also for diamond), the results point to $N=2$, meaning that the hole covers the two atoms with covalent bond.

The band gaps calculated with LDA-1/2 are presented in Table IV. Here we must remember that LDA-1/2 is still a scheme for calculating excitations, not the total energy and the equilibrium lattice parameter. Whereas the LDA results exhibit the well-known underestimation of the energy gap, LDA-1/2 results present excellent agreement with experiment. In general, by comparing the theoretical LDA and LDA-1/2 band structures, we observe that, as in LDA-1/2 the self-energy is removed, the valence states are now more localized and are pulled down in energy in comparison with the LDA, which results in a larger energy gap.

The LDA-1/2 entries in Table IV require a further explanation. In the cases marked with superscript "e," we add the trimmed self-energy potential derived from the half-ionized anion *p* state and the trimmed self-energy potential derived from the half-ionized cation *d* state. The questions then are why are we adding the *p* correction to the anion and not the *s* correction, and the *d* correction and not the *s* one to the cation. The case of C, Si, and Ge, when we used $-1/4$ and not $-1/2$ ionization, has been discussed above. Thus there seems to be a certain degree of arbitrariness in a LDA-1/2 scheme. But that is not so because, from what is known from the chemical bonding of these compounds, we could not proceed differently. Further we are always keeping in mind the criterion of an extreme band gap [Eq. (17)]. The case of diamond (C) is even more puzzling because we are adding trimmed self-energy *s* and *p* potentials to a single atom. In this case we are defining the self-energy potential as in Eq. (11) and approaching the method of Filippetti and Spalding.[8] Again, Eq. (17) is our guide.

Figures 4–6 depict the corresponding band structures (BSs) along the main symmetry directions of the Brillouin zone (BZ) for Si, ZnO, and InN, comparing the LDA-1/2 with LDA. The zero of energy was placed at the top of the valence band. We chose to show the BSs for these semiconductors for two reasons: First, silicon is one of the most important semiconductors. Second, we would like to show the results for cases where the LDA fails completely, such as InN, for which LDA gives a semimetal instead of a semiconductor, and ZnO, which became a very interesting material

TABLE IV. Band energy gaps (eV) for several semiconductors obtained with the LDA-1/2 at experimental lattice constant, by using the VASP code and SIESTA (S), compared with pure LDA, GW, and experimental results in Ref. 32 except where noted. Direct energy gaps are denoted as (d) and the indirect ones as (i). The majority of the LDA-1/2 calculations were obtained using only the trimmed self-energy potential of *p* anion; exceptions are noted.

	LDA-1/2	LDA	Expt.	GW
C (i)	5.25 (S)[a]	4.13	5.47[b]	5.48–5.77[c]
C (d)	6.75 (S)[a]	5.54	7.3[b]	
Si (i)	1.137, 1.21 (S)	0.51	1.17[b]	1.32,[d] 0.95–1.10[c]
Si (d)	2.9, 2.94 (S)	2.54	3.05, 3.40[b]	
Ge (i)	0.70	0.08	0.66–0.74[b]	0.66–0.83[c]
AlN (d)	6.06	4.27	6.23	5.83–6.24[c]
GaN (d)	3.52[e]	1.95	3.507	3.15–3.47[c]
InN (d)	0.95[e]	−0.29	0.7–1.9	0.20–0.33[c]
AlP (i)	2.79	1.47	2.52	2.59[d]
GaP (i)	2.36($\Gamma-L$) [e]	1.49($\Gamma-X$)	2.35	2.55[d]
InP (d)	1.12[e]	0.50	1.42	1.44[d]
AlAs (i)	2.73	1.34	2.24	2.15[d]
GaAs (d)	1.41	0.41	1.519	1.22,[d] 1.40–1.70[c]
InAs (d)	0.75	−0.34	0.417	0.31[d]
ZnO (d)	3.29[e]	0.83	3.4[b]	2.51–3.07[c]
ZnS (d)	3.68[e]	2.02	3.91[b]	3.21–3.57[c]

[a]$-1/4p-1/4s$.
[b]Reference 33.
[c]Reference 5.
[d]Reference 18.
[e]$-1/2p$ anion–$1/2d$ cation.

with large band gap, and the LDA predicts an energy gap much smaller than the experimental value. Moreover, for ZnO it is difficult to obtain the correct BS, even if performing quasiparticle calculations using GW if the starting point is the standard LDA wave functions.[26] From our results, we observe that for Si, the LDA-1/2 dispersion relations are similar to the LDA but with the correct band-gap energy. For InN and ZnO, the same behavior as that for Si occurs, but with some differences concerning the cation *d* states. In both InN and ZnO, the semicore cation *d* states play an important role. In the nitride, the states derived from the atomic $4d$ (In) orbital lie close to the bottom of the valence band $2s$ (N) orbital and hybridize with it.[27] On the other hand, in ZnO, the cation *d* states lie approximately in the middle of the valence band. Moreover, recently it was shown that in both cases the *d* states interact and hybridize with the top of va-

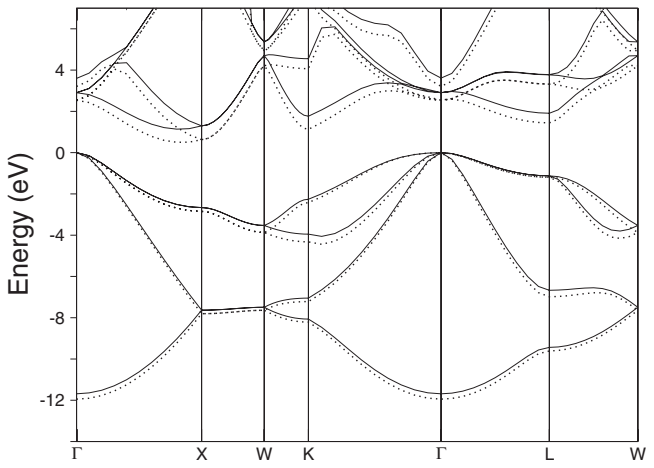

FIG. 4. Calculated band structures for Si (in eV). Dashed lines display LDA and solid lines represent LDA-1/2 results. The zero of energy was placed at the top of the valence band.

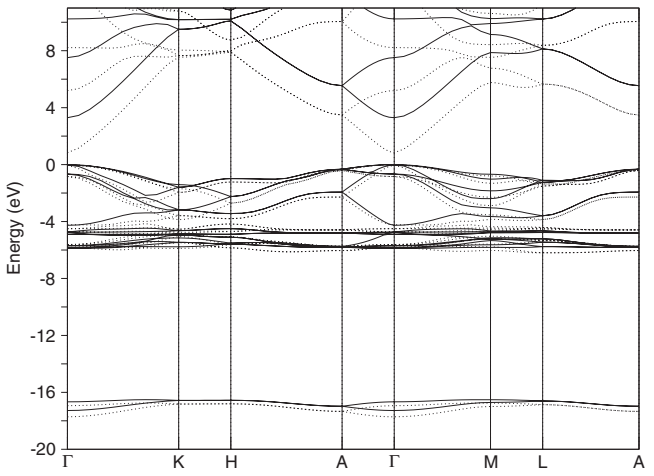

FIG. 5. Calculated band structures for ZnO (in eV). Dashed lines display LDA and solid lines represent LDA-1/2 results. The zero of energy was placed at the top of the valence band.

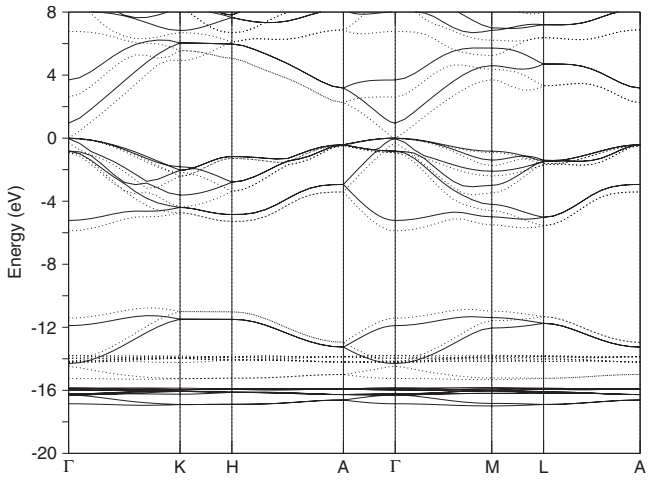

FIG. 6. Calculated band structures for InN (in eV). Dashed lines display LDA and solid lines represent LDA-1/2 results. The zero of energy was placed at the top of the valence band.

lence band, and in DFT-LDA there is an underestimation of the binding energies of these semicore d states and consequently an overestimation of their hybridization with the anion. The enhanced p-d coupling then pushes up the valence-band maximum and the energy gap becomes smaller.[28–31] By taking these facts into account, for InN and ZnO, LDA-1/2 corrects not only the top of valence but also the cation d states. Thus, by comparing the LDA with LDA-1/2 band structures, we observe that the latter has the cation d orbitals deeper in energy. This effect is more pronounced in InN. In both cases, ZnO and InN, LDA-1/2 is remarkable, leading to values very near the experimental ones. In order to study deeply the influence of the cation d state in both InN and ZnO, in Table V we present the results for the VBW and energy gap at different levels of the LDA-1/2 procedure. We observe that in the case of ZnO, the O p-state correction increases the VBW and the Zn d-state correction decreases the VBW. The combination of (O p)+(Zn d) corrections presents a smaller VBW than the pure LDA calculation. However, both corrections increase the value of the energy gap. The combination of O p and Zn d corrections results in energy gap in very good agreement with experiment, which again states the importance of taking into account the cation d state. It is worth to point out that we obtain this good result, in spite of the fact that the position of Zn d state ($\sim$5 eV below the top of valence band) is higher in energy than the experimental data ($\sim$7.8 eV).[34] In the case of InN, as the d state is deeper than in the case of ZnO, the In d-state correction is more important for the VBW value, while the N p correction is more important for the energy-gap value. Particularly, with (N p)+(In d) correction we obtain a value for the energy gap which is in good agreement with experiment. Moreover, our full N p+In d LDA-1/2 calculation is in precise agreement with the measured value obtained from x-ray photoemission spectroscopy experiments,[35] from which the d state of In atom is found to lie 16.0 eV below the valence-band maximum.

TABLE V. InN and ZnO valence-band width (VBW) and band-gap energy values at different levels of the LDA-1/2 calculation procedure. The levels presented are: (i) standard LDA calculation, (ii) LDA-1/2 anion p correction only, (iii) LDA-1/2 cation d correction only, and (iv) full LDA-1/2 anion p+cation d corrections.

	Correction	VBW (eV)	Band gap (eV)
ZnO	None	17.72	0.83
	O p	19.44	2.14
	Zn d	17.01	1.48
	(O p)+(Zn d)	17.28	3.29
InN	None	15.25	−0.29
	N p	14.49	1.16
	In d	18.23	−0.49
	(N p)+(In d)	16.85	0.95

TABLE VI. Effective masses (units of electron free mass m_e) for several semiconductors obtained with the LDA-1/2 at experimental lattice constant, compared with pure LDA and experimental results. The calculations were made using the VASP code. The experimental data were extracted from Ref. 32 except where noted. The same trimmed self-energy potential used in Table IV are also used here.

	Electron effective mass		
	LDA-1/2	LDA	Expt.
AlN	0.38	0.30	
GaN	0.30	0.17	0.18–0.29
InN	0.12		0.11–0.23
AlP	0.256	0.18	
GaP	0.17	0.10	0.09–0.17
InP	0.088	0.04	0.077–0.081
AlAs	0.064	0.022	0.06–0.15
GaAs	0.064	0.026	0.065–0.07
InAs	0.047	0.033	0.023–0.03
ZnO	0.39	0.14	0.3–0.36[a]
ZnS	0.26	0.16	

[a]Reference 36.

In order to analyze the band dispersion in more detail, we also performed calculations to obtain the conduction-band effective masses. Thus, now we focus our attention on the electronic structure around the conduction-band minima. We fit a parabola to the curves of energy versus **k** around the conduction-band minimum up to 1.0% along the main symmetry directions of the Brillouin zone. Considering the degeneracies and making weighted averages, we obtain the electron effective masses. Table VI summarizes the effective conduction-band masses for several semiconductors. Since a negative value for the LDA InN band gap was obtained, it was not possible to calculate an effective mass in that case and only the LDA-1/2 value is shown. We note from the table that the LDA-1/2 method systematically gives larger electron effective masses than LDA. This is due to the fact that with the LDA underestimation of band-gap energy, the $\vec{k}\cdot\vec{p}$ interaction between valence band (VB) and conduction band (CB) is stronger, leading to smaller effective masses. Therefore, in the cases where the correction of the energy

gap is more pronounced, the difference between the LDA and LDA-1/2 electron effective masses is larger. This is the same case as, e.g., the GaAs and ZnO, for which the LDA values agree rather poorly with experimental data, and the LDA-1/2 gives excellent agreement with experiment. Moreover, if we take a look at the whole table, we will observe that the LDA-1/2 effective masses are generally in very good agreement with experimental data. Therefore, the LDA-1/2 not only improves the band gaps as a "scissors operator" approach, but also provides reliable important band structure-derived properties, such as the effective masses.

V. SUMMARY

The very important problem concerning the calculation of excitations in solids is addressed and a method to overcome this problem is developed. The method is inspired by the simple half-ionization method. The localization of the hole created by promoting an electron from the valence band to the conduction band follows naturally from the method. The hole is shown to be representable by a square-integrable wave function, instead of the usual Bloch wave hole of band-structure calculations.

The major success of this method is its reliable description of excited states in solids, giving band-gap energies, effective masses, and band structures in very good agreement with experiment, even in the cases for which the LDA markedly fails, such, e.g., ZnO and InN. The method is not more computationally demanding than the LDA calculations. Moreover, the method is general and can be applied to a broad class of DFT self-consistent methods, all-electron and pseudopotential based.

ACKNOWLEDGMENTS

This work was supported by the Brazilian funding agencies FAPESP (Processos No. 2006/05858-0 and No. 2006/61448-5) and CNPq.

*guima00@gmail.com
†mmarques@ita.br
‡lkteles@ita.br

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[2] R. O. Jones and O. Gummarsson, Rev. Mod. Phys. **61**, 689 (1989).
[3] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[4] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).
[5] M. van Schilfgaarde, T. Kotani, and S. V. Faleev, Phys. Rev. Lett. **96**, 226402 (2006); Phys. Rev. B **74**, 245125 (2006).
[6] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[7] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **54**, 5495 (1996).
[8] A. Filippetti and N. A. Spaldin, Phys. Rev. B **67**, 125109 (2003).
[9] C. D. Pemmaraju, T. Archer, D. Sánchez-Portal, and S. Sanvito, Phys. Rev. B **75**, 045101 (2007).
[10] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[11] R. Asahi, W. Mannstadt, and A. J. Freeman, Phys. Rev. B **59**, 7486 (1999).
[12] M. Städele, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. Lett. **79**, 2089 (1997).
[13] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys.: Condens. Matter **9**, 767 (1997).
[14] D. A. Liberman, Phys. Rev. B **62**, 6851 (2000).
[15] J. C. Slater and K. H. Johnson, Phys. Rev. B **5**, 844 (1972).
[16] J. C. Slater, Adv. Quantum Chem. **6**, 1 (1972).
[17] J. R. Leite and L. G. Ferreira, Phys. Rev. A **3**, 1224 (1971).
[18] X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14142 (1991).
[19] J. F. Janak, Phys. Rev. B **18**, 7165 (1978).
[20] L. G. Ferreira and J. R. Leite, Phys. Rev. A **20**, 689 (1979); L. G. Ferreira, A. Fazzio, H. Closs, and L. M. Brescansin, Int. J. Quantum Chem. **16**, 1021 (1979).
[21] J. Harris, Phys. Rev. B **31**, 1770 (1985).
[22] D. Vanderbilt, Phys. Rev. B **32**, 8412 (1985).
[23] G. Kresse and J. Hafner, J. Phys.: Condens. Matter **6**, 8245 (1994).
[24] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[25] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[26] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).
[27] L. E. Ramos, L. K. Teles, L. M. R. Scolfaro, J. L. P. Castineira, A. L. Rosa, and J. R. Leite, Phys. Rev. B **63**, 165210 (2001).
[28] S. H. Wei and A. Zunger, Phys. Rev. B **37**, 8958 (1988).
[29] Su-Huai Wei, X. Nie, I. G. Batyrev, and S. B. Zhang, Phys. Rev. B **67**, 165209 (2003).
[30] C. G. Van de Walle and J. Neugebauer, Appl. Phys. Lett. **70**, 2577 (1997).
[31] A. Janotti, D. Segev, and C. G. Van de Walle, Phys. Rev. B **74**, 045202 (2006).
[32] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**, 5815 (2001); I. Vurgaftman and J. R. Meyer, *ibid.* **94**, 3675 (2003).
[33] S. Adachi, R. Blachnik, R. P. Devaty, F. Fuchs, A. Hangleiter, W. Kulisch, Y. Kumashiro, B. K. Meyer, and R. Sauer, in *Semiconductors: Intrinsic Properties of Group IV Elements and III-V-II-VI, and I-VII Compounds*, Landolt-Börnstein, New Series, Group III: Semiconductors, subvol. A1–Part β, and R. Blachnik, J. Chu, R. R. Galazka, J. Geurts, J. Gutowski, B. Hönerlage, D. Hofmann, J. Kossut, R. Lévy, P. Michler, U. Neukirch, T. Story, D. Strauch, and A. Waag, subvol. B, edited by U. Rössler (Springer, Berlin, 1999).
[34] Ü. Özgür, Ya. I. Alivov, C. Liu, A. Teke, M. A. Reshchikov, S. Doğan, V. Avrutin, S.-J. Cho, and H. Morkoç, J. Appl. Phys. **98**, 041301 (2005).
[35] L. F. J. Piper, T. D. Veal, P. H. Jefferson, C. F. McConville, F. Fuchs, J. Furthmüller, F. Bechstedt, Hai Lu, and W. J. Schaff, Phys. Rev. B **72**, 245319 (2005).
[36] D. L. Young, T. J. Coutts, V. I. Kaydanov, A. S. Gilmore, and W. P. Mulligan, J. Vac. Sci. Technol. A **18**, 2978 (2000).