Bound exciton model for an acceptor in a semiconductor

Yong Zhang[*] and Jianwei Wang

Electrical and Computer Engineering Department, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA

We attempt to clarify an ambiguity in the understanding of the electronic structure of an acceptor in a semiconductor. Instead of using only a single quantity acceptor binding energy E_A as in the literature, there is in fact an impurity level E_I that plays perhaps a more important role when dealing with the electronic transitions involving the acceptor. Together they determine an excitonic transition energy $E_{A,\mathrm{ex}} = E_I - E_A$. We discuss the underlying physics and approximations of using three different approaches, all in the framework of a density-functional theory, for calculating the acceptor electronic structure, and point out the different meanings of the results yielded from these approaches and how they should be compared with experimental data to extract E_I and E_A. Furthermore, the discussions provide a unified view of "deep" and "shallow" impurities: The characteristics of deep and shallow reflect two different aspects of a complete description of a typical impurity, either known as isoelectronic or acceptor (donor), rather than the spatial extension of the impurity potential.

I. THE CONCEPT OF AN ACCEPTOR IN A SEMICONDUCTOR

When an impurity with one or more valence electron(s) less than that of the replaced host atom is introduced into an otherwise perfect semiconductor, it typically introduces a partially occupied state near the top of the valence band. Such an impurity is often referred to as an acceptor, because it can accept one or more electron(s) from the valence band by thermal excitation, assuming these states are relatively close to the valence band maximum (VBM). Here we do not consider the trivial case where the impurity level turns out to be below the VBM, thus the acceptor is self-ionized, effectively resulting in a metallic material. At least conceptually, one would think that the electronic structure of an acceptor impurity in a semiconductor is well understood. In a typical textbook description, an acceptor would introduce an empty level at energy E_A above the VBM. E_A is known as *acceptor binding energy*, and understood as the energy needed to promote an electron from the VBM and thus leave a free hole in the valence band. Consequently, the transition energy for an electron in the conduction band to the acceptor level, known as a free-to-bound transition, would be $E_{\mathrm{F-B}} = E_g - E_A$, where E_g is the band gap. With this understanding, the energy diagram of an acceptor center and the related transition energies are illustrated in Fig. 1, as appeared virtually in all textbooks. In an idealistic situation, E_A would be given by the solution of a Schrödinger equation resembling that of a hydrogen atom, with an effective mass m_h of the valence band and a screened Coulomb interaction between the acceptor and hole [1]:

$$\left(\frac{\hbar^2}{2m_h^*} \nabla^2 + \frac{e^2}{\epsilon r} \right) F(r) = E_A^{\mathrm{eff}} F(r), \tag{1}$$

where E_A^{eff} is the acceptor binding energy in an effective mass approximation.

The above standard textbook description is conceptually incorrect [2]. Note that the acceptor impurity is in fact charge

neutral if the sample temperature is sufficiently low. The attractive Coulomb potential in Eq. (1) only arises after one electron has "jumped" into the acceptor site. This process involves an electronic transition for an electron to be excited into a higher energy level that is provided by the acceptor impurity. Quite surprisingly, there is rarely any mention in the literature about this level when one discusses about the acceptor electronic structure. We assume that this impurity state is at E_I above the VBM, and E_I is referred to as *impurity energy level* thereafter. The value of E_I is largely determined by the difference in the atomic orbitals between the host and impurity atom, and thus should depend sensitively on the impurity species, whereas E_A is originated from the Coulomb attraction between the ionized impurity core and the hole left in the valence band, and ideally independent of the impurity species. An extra energy E_A is required to set this hole free. A tight-binding model can make this picture easier to understand. Let us consider an acceptor impurity, with one less valence electron, having a higher p orbital than the host. If one electron is moved from the host to the impurity site by applying an excitation energy of the difference between the p orbitals, a hole is then generated in the valence band, but remaining attracted to the ionized impurity through the Coulomb interaction. Additional energy is needed to allow this bound hole to break away from the acceptor becoming a free hole.

It has long been known that the experimentally determined acceptor binding energy E_A^{expt} is usually substantially larger than that given by the hydrogen model, and also varies significantly with the dopant species. For instance, in Si, E_A^{expt} varies from 45.8 meV for B to 247.7 meV for Tl among the group III elements [3], whereas $E_A^{\mathrm{eff}} = 31.6$ meV from an improved version of Eq. (1) taking into account the degeneracy and nonparabolicity of the valence band [4]. The primary contribution to the discrepancy between E_A^{expt} and E_A^{eff} has generally been believed to be the so-called "chemical shift" or "central cell correction" that could be mitigated by introducing a semiempirical short-range potential in Eq. (1) [5,6]. However, this practice is conceptually rather ambiguous,
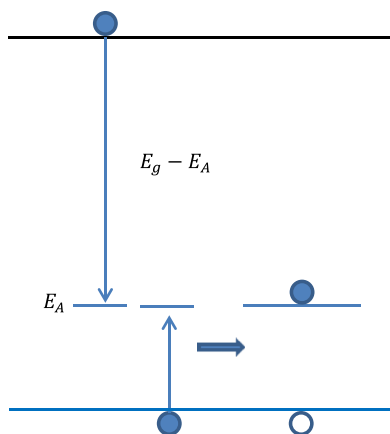
---

[*]yong.zhang@uncc.edu

FIG. 1. (Color online) Textbook band diagram and related transitions for an acceptor in a semiconductor: E_A is the electron transition energy from the valence band to the acceptor level, $E_g - E_A$ is the transition energy for an electron in the conduction band to the acceptor level (i.e., free-to-bound).

One-electron picture **Many-electron picture**
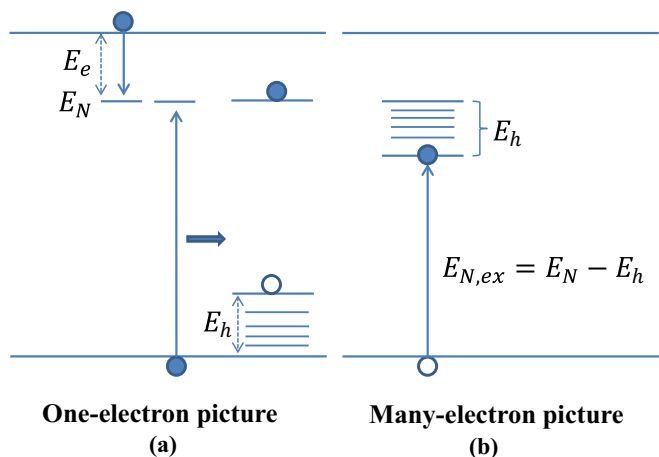**(a)** **(b)**

FIG. 2. (Color online) Band diagrams for an exciton bound to an isoelectronic impurity, known as an acceptorlike bound exciton, such as in GaP:N, where E_N is the energy level of the bare electron bound state measured from the VBM, E_h the hole binding energy to the N^- center, and $E_{N,ex} = E_N - E_h$ the lowest bound exciton transition energy. (a) In one-electron picture, and (b) in many-electron picture.

and may lead to misinterpretation of experimental data. The intent of this work is to clarify the ambiguity in the literature about the electronic structure of the acceptor, and discuss how to correctly interpret the theoretical, in particular from an *ab initio* calculation, and experimental results, and compare them. In short, E_I has nothing to do with E_A, and E_A^{expt} actually corresponds to E_I.

We first describe the electronic structure of an exciton bound to an isoelectronic impurity of an electron trap, such as GaP:N, where the bound exciton is known as an "acceptorlike bound exciton" based on the model proposed by Hopfield, Thomas, and Lynch (HTL model) [7,8]. This bound exciton problem is viewed as a classic example of the "deep" impurity that is thought to be profoundly different from the "shallow" impurity, either an acceptor or a donor, in terms of the extension of the impurity potential [9]. We point out that the electronic structures of an isoelectronic impurity and an acceptor actually bare much more similarity than one might think. As illustrated in Fig. 2, for an isoelectronic impurity N in GaP, the nitrogen atom generates an electron bound state within the band gap. The energy level of the electron bound state is denoted as E_N measured from the VBM, and the separation from the conduction band minimum (CBM) $E_e = E_g - E_N$ is known as electron binding energy. Being an empty state far away from the VBM, E_N can be considered as a deep acceptor level. In the one-electron picture, the formation of a bound exciton on the N impurity could be viewed as a two-step process (HTL model): (1) One electron is excited into the electron bound state at E_N (or captured from the conduction band if the electron was already in the conduction band), forming a so-called bare electron bound state or a negatively charged center N^-; and (2) through the Coulomb interaction, a hole is attracted to the N^- center, forming a bound exciton with a hole binding energy E_h (with respect to a hole at the VBM), as shown in Fig. 2(a), where the transition energy $E_{N,ex} = E_N - E_h = E_g - (E_e + E_h)$ corresponds to the zero-phonon absorption or emission energy of the bound exciton. The bound exciton formation is ultimately a many-electron problem that should be treated

as the transition between two states of the whole system. Figure 2(b) shows the energy diagram in the many-electron picture where E_N can be viewed as the upper limit of the bound exciton states corresponding to the hole in different hydrogenlike bound states $1s$, $2s$, etc. The many-electron picture makes it easier to understand why the formation of a bare electron state is not a necessary precursor to the formation of a bound exciton. Rather a bound exciton can be formed with a resonant excitation of energy $E_{N,ex}$. The relationship between E_N and $E_{N,ex}$ is analogous to that between E_g and the free exciton band gap $E_{g,ex}$ in a semiconductor, and E_h corresponds to the free exciton binding energy $E_{ex,b}$ (e.g., $E_{ex,b} = 20.6$ meV for Si). Although the electron bound state is typically found to be more close to the CBM than VBM for an isoelectronic impurity with stronger electron negativity than the host atom, it could in principle be very close to the VBM and therefore may be excited thermally. If we do have a system like this, the acceptorlike bound exciton will be essentially the same as a conventional acceptor state, except that the acceptor level for the former is usually s-like and empty, and for the latter p-like and partially occupied.

With above preparation, we now discuss the conventional acceptor impurity problem. As illustrated in Fig. 3, an acceptor generates an impurity level E_I in the band gap that can accept an electron excited from the valence band, forming a negatively charged center A^-, and if it does happen the acceptor is considered ionized (step 1); then through Coulomb interaction, this ionized impurity center can now attract a hole (step 2), resulting in a bound exciton. These two steps are illustrated in Fig. 3(a) in the one-electron picture, where the energy needed to promote the electron from the VBM in step 1 is E_I, and the hole binding energy to the A^- center is E_A, similar to the case of the acceptorlike bound exciton of the isoelectronic impurity with E_I and E_A corresponding to, respectively, E_N and E_h in the former case in Fig. 2. Alternatively, if we describe the impurity system in the many-electron picture, i.e., consider the total energy change of the whole system, the corresponding
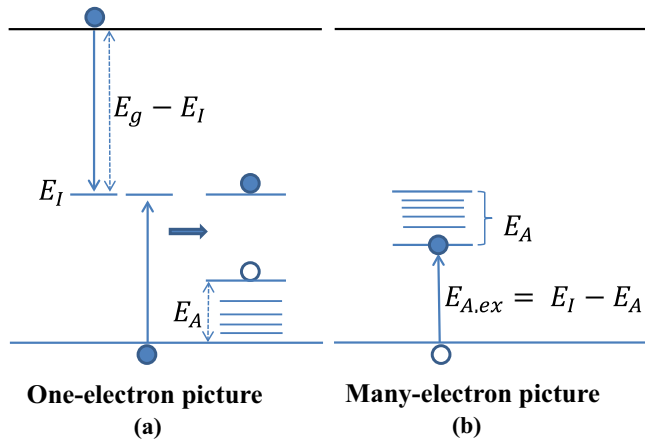
FIG. 3. (Color online) Band diagrams for an acceptor, where E_I is the impurity energy level or the energy needed to create a free hole in the valence band, E_A the hole binding energy to the ionized acceptor, $E_g - E_I$ is the free-to-bound transition energy, and $E_{A,\mathrm{ex}} = E_I - E_A$ the minimum energy required to create a bound exciton. (a) In one-electron picture, and (b) in many-electron picture.

transition energies can then be illustrated in Fig. 3(b) for various excitonic transitions. If the E_I state is highly localized such that its charge density can be approximated by a point charge, the solution of Eq. (1) or a more generalized effective mass equation could be used as an approximation for E_A, although there are other complications that will be discussed later.

How does the above description compare to the familiar textbook description about the acceptor state? The qualitative differences are evident by comparing Fig. 1 with Fig. 3, and can be summarized in three aspects.

First, the minimum energy required for generating a free hole in the valence band by thermalization or optical excitation is not E_A but E_I. As a matter of fact, the experimentally derived "acceptor binding energy" E_A^{expt}, by either IR absorption or temperature-dependent electrical conductivity measurement, should be interpreted as E_I rather than E_A. As pointed out earlier, E_I is primarily determined by the chemical difference, for the specific system being considered here Si:III, largely the p-orbital energy difference of the valence electrons, between the dopant and the host atom, thus has nothing to do with the solution of Eq. (1). This understanding explains the large deviation between E_A^{expt} and E_A^{eff} and the strong species dependence.

Second, the energy $E_{A,\mathrm{ex}} = E_I - E_A$ will be the minimum energy needed to generate a hole bound to the ionized acceptor or an exciton bound to the acceptor, resembling $E_{\mathrm{N,ex}}$ in the isoelectronic impurity problem GaP:N. Therefore, roughly speaking, taking E_A as the ionization energy instead of E_I is equivalent to taking the exciton binding energy $E_{\mathrm{ex},b}$ or hole binding energy E_h as the interband transition energy instead of E_g or E_N in each case to generate a free hole in the valence band. These comparisons are more meaningful if one further envisions an array of impurities with a high enough density that just barely forms an energy band, then E_I and E_N could be viewed as the new conduction band of the material. Thus, the bound exciton problem would resemble the free exciton

problem, with E_A or E_h just like the free exciton binding energy $E_{\mathrm{ex},b}$. In fact, the absorption spectrum of an acceptor is qualitatively similar to that of an isoelectronic impurity bound exciton or even a free exciton: A series of discrete absorption peaks converged to E_I for the acceptor [10], to E_N for the isoelectronic bound exciton [11], and to E_g for the free exciton, although they follow somewhat different selection rules.

Third, the free-to-bound transition energy should be given approximately by $E_{\mathrm{F\text{-}B}} = E_g - E_I$ instead of the commonly used $E_g - E_A$.

II. UNDERSTANDING THE *AB INITIO* COMPUTATIONAL RESULTS

Next we discuss the different meanings of various calculated transition energies associated with an acceptor using different density-functional theory (DFT) based approaches. In virtually all DFT calculations for impurities or point defects, the calculated transition energies were either explicitly or implicitly treated as E_A or compared to experimental results that have been interpreted as E_A. We first offer some qualitative discussions based on a Hartree-Fock (H-F) approximation that seems to be conceptually more transparent than a DFT for illustrating the underlying physics. Within the H-F approximation, the total energy difference between the two states of the system, the excited state (one electron has been moved to the impurity state from the VBM) and the ground state (the valence band is fully occupied), is given as [12]

$$\delta E_{\mathrm{tot}} = E_I - E_{\mathrm{VBM}} - \left[\left\langle \varphi_I \varphi_{\mathrm{VBM}} \left| \frac{e^2}{\epsilon r} \right| \varphi_I \varphi_{\mathrm{VBM}} \right\rangle \right.$$
$$\left. - \left\langle \varphi_I \varphi_{\mathrm{VBM}} \left| \frac{e^2}{r} \right| \varphi_{\mathrm{VBM}} \varphi_I \right\rangle \right], \quad (2)$$

where E_I and E_{VBM} are the absolute values of H-F one electron eigenenergies for the impurity and VBM state, and ϕ_I and ϕ_{VBM} are the corresponding wave functions, respectively. We should assume $E_{\mathrm{VBM}} = 0$ so that the meaning of E_I in Eq. (2) is consistent with the same quantity introduced above. The first term in the square brackets is the Coulomb interaction between the impurity state and VBM, and the second term is the exchange interaction. The dielectric function ε is added empirically to the Coulomb interaction term to include the screening effect, but not to the exchange term because of its short-range nature. From now on, we will refer both Coulomb and exchange interaction together as Coulomb contribution for simplicity. Conceptually, this Coulomb contribution is really what the E_A in Eq. (1) is about, because it occurs only after the transition of one electron from VBM to E_I. If we take this term as an approximation for E_A in Eq. (1), the total energy difference will then be $\delta E_{\mathrm{tot}} \approx E_{A,\mathrm{ex}} = E_I - E_A$. A DFT version of Eq. (2) is given by Eq. (15) of Ref. [13]. Evidently, δE_{tot} given by Eq. (2) or its DFT equivalent merely evaluates the static Coulomb interaction between the electron and hole, and neglects the kinetic energy of the hole, thus yielding only an approximate E_A. If the kinetic energy of the hole or the $k \neq 0$ component of the Coulomb potential is taken into account, as in Eq. (1), the Coulomb contribution will not be as simple as that only between ϕ_I and ϕ_{VBM}, which will be

discussed later. The most important message of Eq. (2) is that δE_{tot} and E_I are two different physical quantities.

Similarly we can write the free-to-bound transition energy as the total energy difference between two states of the system:

$$
\begin{aligned}
E_{\text{F-B}} = E_{\text{CBM}} - E_I - \Bigg[& \left\langle \varphi_I \varphi_{\text{CBM}} \left| \frac{e^2}{\epsilon r} \right| \varphi_I \varphi_{\text{CBM}} \right\rangle \\
& - \left\langle \varphi_I \varphi_{\text{CBM}} \left| \frac{e^2}{r} \right| \varphi_{\text{CBM}} \varphi_I \right\rangle \Bigg]
\end{aligned} \tag{3}
$$

Therefore, $E_{\text{F-B}} = E_g - E_I - E_A{}'$, where $E_A{}'$ is given by the terms in the square brackets for the Coulomb contribution involving the CBM instead of the VBM in E_A. $E_A{}'$ is expected to be in the order of free exciton binding energy, and will not be so significant if $E_g - E_I$ is relatively large. Therefore, $E_g - E_I$ could be taken as an approximation for $E_{\text{F-B}}$ in the situations where $E_g \gg E_I \gg E_A$ is valid.

We will discuss below the three representative approaches for computing the transition energies associated with the acceptor within the framework of DFT. Rather than trying to judge which method is more accurate, our intent here is to highlight the different meanings of the results obtained from these different approaches. The comparison is made for the results all obtained within the local density approximation (LDA), which is sufficient to serve the purpose—revealing the differences in the underlying physics. Si:In will be used as an example for quantitative comparison.

A. Total energy difference between the excited and ground state

In the literature, the total energy difference δE_{tot} is commonly used or implied as the quantity to be compared with the experimentally derived "acceptor binding energy" E_A^{expt} or as a more accurate version of the acceptor binding energy E_A in Eq. (1). However, as pointed out above, δE_{tot} is actually an approximate value for the transition or formation energy of the acceptor bound exciton $E_{A,\text{ex}}$, and thus should not be compared with E_A^{expt} that actually measures the single particle energy E_I, neither should it be viewed as E_A that describes the Coulomb interaction. Here we calculate δE_{tot} in order to compare it with the results to be obtained using the other two approaches.

The conceptually most straightforward way to evaluate the transition energy between the excited and ground state would be, with the total number of the valence electrons (N) fixed, calculating the total energy difference between them with one electron being forced to occupy the E_I level (the so-called constrained DFT). However, more often in the literature, the excited state of the system is simulated by a system with one extra valence electron added to the original one or ($N + 1$) valence electrons, and simultaneously a uniform positive background introduced to compensate the charge of the extra electron [14,15]. There are some subtle differences between the two methods, but the discussion is beyond the scope of this work. Nevertheless, either way, this total energy difference approach yields an approximation for $E_I - E_A$. Because the kinetic energy of the hole is neglected, E_A is overestimated. To correctly describe the Coulomb contribution and explain those abundant discrete transitions in absorption [10], one would need to convert Eq. (2) into an excitonic equation (also known as a Bethe-Salpeter equation) by taking into account the kinetic

energy of the hole [12,16]. If this last step is carried out, we should have the most rigorous treatment for the acceptor problem. A simplified treatment of the excitonic problem will be given later along with the second DFT based approach.

Taking Si:In as an example, the previous DFT-LDA calculation yielded $\delta E_{\text{tot}} = 39$ meV by evaluating the total energy difference between the $N + 1$ and N electron system [13]. Our constrained DFT-LDA calculation has yielded $\delta E_{\text{tot}} = 36$ meV, by extrapolating a fitting curve with results of supercell sizes varying from 64 to 1000 atoms to the infinite supercell size using a polynomial $a + b/n + c/n^2$, where n is the supercell size. Apparently δE_{tot} is much smaller than $E_A^{\text{expt}} = 153$ meV. The reason for this large discrepancy is not simply due to the deficiency of the computational method, for instance, the LDA. As mentioned above, they simply represent two physical quantities that should not be directly compared with each other. The distinction will be clearer after the discussions to be given for the other two DFT based approaches.

B. Total energy calculation of the ground state

From the ground state calculation, one can obtain the partially occupied impurity state E_I and its wave function ϕ_I. One can go one step further to solve the whole bound exciton problem. If the Coulomb interaction is relatively weak, the Coulomb contribution can be described by an effective mass equation with the point charge in Eq. (1) replaced by a charge density $d(r)$ and an exchange term, as given below for an isotropic and parabolic single valence band [17]:

$$
\left(\frac{\hbar^2}{2m_h^*} \nabla^2 + \frac{d(r)e^2}{\epsilon r} - J\rho(r) \right) F(r) = E_A^{\text{eff}} F(r), \tag{4}
$$

with $d(r)/r = \Sigma_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{r}) f(k) s(\mathbf{k})$ being the Coulomb potential with its Fourier component $f(k)$ weighted by $s(\mathbf{k}) = \Sigma_{\mathbf{k}'} a^*(\mathbf{k}') a(\mathbf{k}' - \mathbf{k})$, where $a(\mathbf{k})$ is the $\mathbf{k}$ component of the impurity wave function φ_I expanded in the basis of the bulk states; J is approximately the exchange term in Eq. (2), and $\rho(r) = \Sigma_{bfk} \exp(i\mathbf{k}\cdot\mathbf{r}) s(\mathbf{k}) \approx |\varphi_I|^2$. Apparently, if $|\varphi_I|^2$ is a δ function, we have $d(r) = 1$, and Eq. (4) is essentially the same as Eq. (1). Because of the finite extent of the impurity state, the binding energy will be smaller than E_A^{eff} from the idealistic effective mass equation, which is exactly what has been observed experimentally for the hole bound states of the excitons bound to various nitrogen impurity centers: E_h depends on the electron binding energy but always $E_h < E_A^{\text{eff}}$, for all NN_i and N centers [11,17]. The reduction was initially interpreted as due to the central cell correction [11], but is more correctly explained as due to the finite extension of the electron bound state [17]. As an approximation, one could neglect the finite extension of the E_I state or skip Eq. (4) by simply taking the multiband effective mass solution as the upper bound of E_A [4]. Therefore, the burden of solving the acceptor problem lies mostly on the ability of getting the accurate one electron impurity state E_I. For Si:In, our DFT-LDA calculation has yielded $E_{I,g} = 49$ meV, where "g" stands for "ground state." This result was obtained with a 4096-atom supercell by applying a charge patching method [18] based on a direct DFT-LDA calculation with a 512-atom supercell (using the same DFT code, PEtot, and parameters as in Ref. [13]). The charge patching method takes advantage

of the fact that the impurity induced perturbation in the total electron charge is much more localized than the potential, thus will converge sufficiently well in a moderate supercell size (e.g., 64–512 atoms), as applied for a number of impurity systems with great success, for instance GaAs:N and GaAs:Bi [19,20]. If taking $E_A \approx E_A^{\mathrm{eff}} = 27$ meV (calculated with the LDA band structure [13]), we have an estimate for the excitonic transition energy for In in Si as $E_{A,\mathrm{ex}} \approx E_{I,g} - E_A \approx 23$ meV. $E_{A,\mathrm{ex}}$ and δE_{tot} (=36 meV) can be viewed as two different approximations for the excitonic transition energy, and are physically different from E_I.

This two-step approach is expected to be a reasonably good approximation for solving the acceptor bound exciton problem for many real systems. One potential shortcoming of this approach lies in that it does not account for the difference in the lattice configurations between the excited and the ground state. This effect will be examined below.

C. Total energy calculation of the excited state

One may also perform the total energy calculation for an excited state, in particular with the single-electron impurity level being occupied at $E_{I,e}$, where "e" stands for "excited state." This can be accomplished with either the $N + 1$ or N electron system. Recently, with the $N + 1$ system, such calculation was done for all the group III acceptors in Si without and with GW correction to the LDA error and with supercell sizes up to 64 000 atoms [13,21]. Very good agreement between $E_{I,e}$ and E_A^{expt} has been achieved for all the elements after making the GW and other corrections to the impurity potential [21]. For Si:In, the DFT results were $E_{I,e} = 88$ meV in LDA (but with other corrections) [13], and 139 meV after further correcting the LDA error [21]. The latter value agrees quite well with $E_A^{\mathrm{expt}} = 157$ meV [3]. In order to make fair comparison with δE_{tot} of the first approach, we have re-done the calculation without applying the corrections of Ref. [13], and obtained $E_{I,e} = 58$ meV in LDA. Using this value, the excitonic transition energy is expected to be $E_{A,\mathrm{ex}} = E_{I,e} - E_A^{\mathrm{eff}} = 58 - 27 = 31$ meV, which is close to the total energy difference $\delta E_{\mathrm{tot}} = 39$ meV.

There is clearly a qualitative correlation between the E_I energy calculated by DFT and the p-orbital energy of the valence electron with respect to the Si $3p$ orbital for Si:III [13,21,22], as expected based on our understanding about the nature of the impurity state. As a matter of fact, the spatial extension of the impurity state wave function, plotted by spherically averaged radial distribution of $|\varphi_I|^2$, is found to be highly localized, and does not resemble at all a hydrogenic state, for all the group III elements, including the shallowest acceptor B [13]. One might be tempted to interpret this wave function localization in terms of the "central cell correction" to Eq. (1). However, we should realize that E_I fundamentally is an eigenvalue of the single particle Kohn-Sham equation that will never produce the abundant discrete absorption lines in the IR absorption spectrum of an acceptor, because it does not address the excitonic nature of the acceptor problem.

Using the constrained DFT with the N electron system and applying the same charge patching method as for the ground state, we obtain $E_{I,e} = 48$ meV for the excited state with the 4096-atom supercell, compared to $E_{I,g} = 49$ meV. The

TABLE I. DFT-LDA results for Si:In (in meV). e is excited state; g is ground state. The first lines are the results of our calculation, the second lines of Ref. [13].

Approach	Total energy calc.		
	(1)	(2)	(3)
Constrained DFT	$E_{A,\mathrm{ex}} \approx \delta E_{\mathrm{tot}} = 36$	$E_{I,g} = 49$	$E_{I,e} = 48$
$E_{\mathrm{tot},e}(N)$, $E_{\mathrm{tot},g}(N)$		$E_A^{\mathrm{eff}} = 27$	$E_A^{\mathrm{eff}} = 27$
		$E_{A,\mathrm{ex}} = 22$	$E_{A,\mathrm{ex}} = 21$
$E_{\mathrm{tot}}(N+1)$, $E_{\mathrm{tot}}(N)$	$E_{A,\mathrm{ex}} \approx \delta E_{\mathrm{tot}} = 39$		$E_{I,e} = 58$
			$E_A^{\mathrm{eff}} = 27$
			$E_{A,\mathrm{ex}} = 31$

difference between $E_{I,e}$ and $E_{I,g}$ is mostly due to the difference in lattice relaxation, which is apparently rather small for Si, but could be larger for other systems. Note that using either the $N + 1$ or N electron system, after obtaining the impurity state $E_{I,e}$, in order to account for those discrete absorption features observed experimentally [10], one has to go one step further to treat the excitonic problem as in Eq. (4) or in a more rigorous manner beyond the effective mass approximation.

Table I summarizes the results of the three different approaches. The numerical results are qualitatively and more or less quantitatively consistent, considering the variations in computational details and approximations involved.

III. EXTENSION TO A DONOR IN A SEMICONDUCTOR

Analogous to the acceptor problem, we may also develop a revised understanding to the donor problem. Figure 4 shows the band diagram for the donor in the same fashion as Fig. 3 for the acceptor. The corresponding total energy difference between the excited state (with one electron in the conduction band and the donor state empty) and the ground state (the donor state being occupied) will be

$$\delta E_{\mathrm{tot}} = E_{\mathrm{CBM}} - E_I - \left[\left\langle \varphi_I \varphi_{\mathrm{CBM}} \left| \frac{e^2}{\epsilon r} \right| \varphi_I \varphi_{\mathrm{CBM}} \right\rangle \right.$$
$$\left. - \left\langle \varphi_I \varphi_{\mathrm{CBM}} \left| \frac{e^2}{r} \right| \varphi_{\mathrm{CBM}} \varphi_I \right\rangle \right], \quad (5)$$
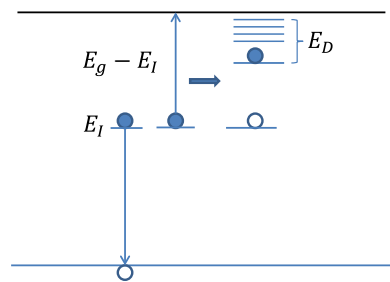

FIG. 4. (Color online) Band diagrams for a donor, where E_I is the energy level of the impurity state, the energy needed to create a free electron in the conduction band is $E_g - E_I$, E_D is the electron binding energy to the ionized donor, and E_I is the bound-to-free transition energy. In donor, the one-electron picture and the many-electron picture are essentially the same, because only the electron levels are considered instead of both electron and hole for acceptor.

where the terms in the square brackets can be viewed as an approximation for the Coulomb contribution or *donor binding energy* E_D, and the transition energy can be written as $\delta E_{\text{tot}} \approx E_{\text{CBM}} - E_I - E_D$. Again, the thermalization energy relevant to the electron conductivity should be $E_{\text{CBM}} - E_I$ not E_D, which are, respectively, the impurity binding energy and the Coulomb energy. The donor problem is very much the same as the "donorlike bound exciton" for another isoelectronic impurity problem GaP:Bi where Bi can be viewed as a deep donor [7].

The corresponding bound-to-free transition energy for the transition from the occupied donor state to VBM is given as

$$
\begin{aligned}
E_{\text{B}-\text{F}} = E_I - &\left[\left\langle \varphi_I \varphi_{\text{VBM}} \left| \frac{e^2}{\epsilon r} \right| \varphi_I \varphi_{\text{VBM}} \right\rangle \right. \\
&\left. - \left\langle \varphi_I \varphi_{\text{VBM}} \left| \frac{e^2}{r} \right| \varphi_{\text{VBM}} \varphi_I \right\rangle \right] \\
= E_I - &E_D', \tag{6}
\end{aligned}
$$

where E_D' is the Coulomb contribution given by the terms in the square brackets, which is expected to be in the order of E_A.

IV. SUMMARY

We have attempted to clarify the ambiguity in the conventional understanding on the electronic structure of an acceptor or a donor in a semiconductor, and underlying physics for the different results calculated with different DFT based approaches. We point out that the commonly called acceptor (donor) binding energy E_A (E_D) due to the Coulomb interaction between the ionized acceptor (donor) and the hole left behind (electron released) is only (usually small) part of the excitation energy for generating a free hole (electron) in the valence (conduction) band. If the impurity level is at E_I measured from the top of valence band, the activation energy to generate a free hole (electron) should be E_I ($E_g - E_I$) instead of E_A (E_D). However, $E_{A,\text{ex}} = E_I - E_A$ ($E_{D,\text{ex}} = E_g - E_I - E_D$) is the minimum energy required to create a hole (electron) bound to the charged impurity center or a bound exciton to the neutral impurity, for instance, corresponding to the absorption threshold in light excitation. An additional energy of E_A (E_D), for instance, provided by thermal activation, can release the bound hole (electron) into the valence (conduction) band. The corresponding free-to-bound transition energy between the conduction band and the acceptor level will be approximately $E_g - E_I$ instead of the commonly used $E_g - E_A$, and the

bound-to-free transition between the donor level and the valence band will be approximately E_I instead of $E_g - E_D$.

We have offered a unified understanding to the impurity problem for isovalent and nonisovalent impurities. The distinction between a "deep" impurity (which is said to have a short-range impurity potential [9]) and "shallow" impurity (to have a long-range Coulomb potential) does not lie in the difference of the impurity type (e.g., isovalent vs acceptor) but in which states of the impurity: In the ground state the impurity potential tends to be highly localized, whereas in the excited state, there is always a long-range Coulomb interaction between the ionized impurity core and the corresponding carrier. For the isovalent impurities, the deep aspect is often emphasized, because it tends to play the dominant role in the most important properties of the impurity, such as, perturbation to the host electronic structure [19], and electron-phonon coupling [23]. However, the shallow aspect is also important in describing the overall properties of the bound exciton, for instance, thermal quenching of the radiative recombination of the bound exciton [8]. For the acceptor and donor impurities, the deep aspect has largely been neglected or overlooked, and the focus has been nearly all on the shallow aspect, even though a vast amount of studies, both experimental and theoretical, were actually dealt with the consequences of the deep aspect. The primary aim of this work is to offer a comprehensive description of the so-called shallow impurities, and stimulate reexaminations of many closely related properties that have been interpreted in the framework of the conventional understanding.

We have discussed the differences in underlying physics and approximations for three different DFT based approaches used for evaluating the transition energies involving an acceptor or a donor, in particular, pointed out the need to develop an excitonic theory to fully account for the optical transitions associated with the acceptor and donor impurity.

Furthermore, we note that the discussions for the impurities are equally applicable to most point defects that typically behave like either an acceptor or a donor in a semiconductor.

ACKNOWLEDGMENTS

[1] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors* (Springer, Heidelberg, 1996).

[2] Y. Zhang and J. Wang, Bull. Am. Phys. Soc. **59**, (http://meetings.aps.org/link/BAPS.2014.MAR.B44.3) (2014).

[3] O. Madelung, *Semiconductors-Basic Data* (Springer, Berlin, 1996).

[4] A. Baldereschi and N. O. Lipari, Phys. Rev. B **9**, 1525 (1974).

[5] S. T. Pantelides, Rev. Mod. Phys. **50**, 797 (1978).

[6] N. O. Lipari, A. Baldereschi, and M. L. W. Thewalt, Solid State Commun. **33**, 277 (1980).

[7] J. J. Hopfield, D. G. Thomas, and R. T. Lynch, Phys. Rev. Lett. **17**, 312 (1966).

[8] M. D. Sturge, E. Cohen, and K. F. Rodgers, Phys. Rev. B **15**, 3169 (1977).

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

[10] D. W. Fischer and J. J. Rome, Phys. Rev. B **27**, 4826 (1983).

[11] E. Cohen and M. D. Sturge, Phys. Rev. B **15**, 1039 (1977).

[12] F. Bassani and G. P. Parravicini, *Electronic States and Optical Transitions in Solids* (Pergamon, Oxford, 1975).

[13] L.-W. Wang, J. Appl. Phys. **105**, 123712 (2009).

[14] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[15] S.-H. Wei and Y.-F. Yan, in *Advanced Calculations for Defects in Materials*, edited by A. Alkauskas, P. Deak, J. Neugebauer, A. Pasquarello, and C. G. Van de Walle (Wiley-VCH, New York, 2011).

[16] M. Rohlfing and S. G. Louie, Phys. Rev. Lett. **80**, 3320 (1998).

[17] Y. Zhang, Phys. Rev. B **45**, 9025 (1992).

[18] L. W. Wang, Appl. Phy. Lett. **78**, 1565 (2001).

[19] Y. Zhang, B. Fluegel, M. C. Hanna, A. Mascarenhas, L.-W. Wang, Y. J. Wang, and X. Wei, Phys. Rev. B **68**, 075210 (2003).

[20] Y. Zhang, A. Mascarenhas, and L. W. Wang, Phys. Rev. B **71**, 155201 (2005).

[21] G. Zhang, A. Canning, N. Grønbech-Jensen S. Derenzo, and L.-W. Wang, Phys. Rev. Lett. **110**, 166404 (2013).

[22] S. H. Wei and A. Zunger, Phys. Rev. B **60**, 5404 (1999).

[23] Y. Zhang, W. K. Ge, M. D. Sturge, J. Zheng, and B. Wu, Phys. Rev. B **47**, 6330 (1993).