Ab initio calculation of the binding energy of impurities in semiconductors: Application to Si nanowires

Y. M. Niquet, L. Genovese, C. Delerue, and T. Deutsch

1CEA-UJF, INAC, SP2M/L-Sim, 38054 Grenoble Cedex 9, France
2European Synchrotron Radiation Facility, 6 rue Horowitz, BP 220, 38043 Grenoble, France
3Institut d’Électronique, de Micro-électronique et de Nanotechnologie (UMR CNRS 8520), Département ISEN, 41 boulevard Vauban, F-59046 Lille Cedex, France

(Dated: April 21, 2010)

Abstract

We discuss the binding energy $E_b$ of impurities in semiconductors within density functional theory (DFT) and the GW approximation, focusing on donors in nanowires as an example. We show that DFT succeeds in the calculation of $E_b$ from the Kohn-Sham (KS) Hamiltonian of the ionized impurity, but fails in the calculation of $E_b$ from the KS Hamiltonian of the neutral impurity, as it misses most of the interaction of the bound electron with the surface polarization charges of the donor. We trace this deficiency back to the lack of screened exchange in the present functionals.
The binding energy $E_b$ of donors and acceptors is a key quantity in semiconductor physics because it determines the doping efficiency. In semiconductor nanostructures for example, confinement and electrostatics tend to shift the impurity levels deeper in the gap, which decreases dopant activity [1, 2]. Therefore, *ab initio* calculations of impurity binding energies are highly desirable to assess the performances of ultimate nanodevices. Besides, donors and acceptors are the prototypes of charged defects in semiconductors, and a fundamental understanding of the strengths and weaknesses of present *ab initio* approaches such as density functional theory (DFT) and the $GW$ approximation [3, 4] would open the way to a more accurate modeling of complex defects.

So far, the calculation of $E_b$ in bulk semiconductors has been possible only with semi-empirical methods [5, 6]. However, calculations based on DFT have become practicable in ultimate nanostructures with a smaller number of atoms. Recently, the case of donors in Si nanowires (Si NWs) has been addressed with both semi-empirical methods and DFT, with contradictory results. Tight-binding [7, 8] and effective mass calculations [9], supported by experiments [1, 2], indeed suggest that $E_b$ increases as $1/R$ with decreasing wire radius $R$, due to the interaction of the bound electron with the surface polarization (or “image”) charges of the impurity, resulting in a significant decrease of the doping efficiency in the $R < 10$ nm range. In contrast, DFT calculations [10, 11] predict that $E_b$ decreases much faster than $1/R$, and is about $3 − 4$ times lower than found in Refs. [7, 9]. In this Letter, we show that present DFT approaches, based on the Kohn-Sham (KS) hamiltonian of the neutral donor, can not predict $E_b$ correctly in bulk and nanostructures, because they miss most of the interactions of the carriers with the polarization charges of the impurity. We propose an alternative strategy based on the KS hamiltonian of the ionized donor which circumvents this deficiency.

For a donor, $E_b$ is the energy needed to ionize the neutral impurity and bring the electron to the conduction band edge far away. It can be defined as the difference $E_b = I^d(N + 1) − A^p(N)$ between the ionization energy $I^d(N+1)$ of the neutral impurity (with $N+1$ electrons) and the affinity $A^p(N)$ of the pristine system (with $N$ electrons and no dopant). Since $A(N) = I(N + 1)$, the binding energy can also be computed as an isoelectronic difference of ionization energies, $E_b = I^d(N+1)−I^p(N+1)$, or affinities, $E_b = A^d(N)−A^p(N)$. In practice, the ionization energies and affinities can be computed either as total energy differences $I(N) = E(N − 1) − E(N)$ and $A(N) = E(N) − E(N + 1)$ [10], or as “quasiparticle” energies...
i.e., as the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbital energies. However, the quasiparticle problem should in principle be addressed with many-body perturbation theories (MBPTs) such as the \( GW \) approximation, since DFT is known to miss the HOMO-LUMO gap \[3, 4, 12, 13\]. We actually show hereafter that the above definitions of \( E_b \) are consistent in the \( GW \) approximation, but not in DFT. Using the insight gained from many-body theory, we conjecture that DFT should succeed in the calculation of \( E_b \) from the KS LUMO of the ionized impurity, but fails in the calculation of \( E_b \) from the KS HOMO of the neutral impurity, due to the lack of explicit \textit{screened} exchange in the present functionals. We support these conclusions with DFT calculations on Si NWs.

The binding energy in many-body theory – In MBPT, the quasiparticle energies \( E_n \) and wave functions \( \varphi_n \) of the \( N \)-electron system are the solutions of the quasiparticle equation:

\[
-\frac{1}{2}\Delta r \varphi_n(r) + v_{\text{ion}}(r) \varphi_n(r) + v_h(r) \varphi_n(r) + \int d^3r' \Sigma_{\text{xc}}(r, r', E_n) \varphi_n(r') = E_n \varphi_n(r)
\]

(1)

where \( v_{\text{ion}}(r) \) is the ionic potential, \( v_h(r) = \int d^3r' \rho(r')/|r - r'| \) is the Hartree potential created by the ground-state electronic density \( \rho(r) \), and \( \Sigma_{\text{xc}}(r, r', E_n) \) is the “self-energy” that describes exchange and correlation effects. The ionization energy is \( I(N) = -E_N \), while the affinity is \( A(N) = -E_{N+1} \).

The \( GW \) approximation has become the reference for the calculation of the band structure of semiconductors \[3, 4\]. For illustrative purposes, we shall use hereafter the simpler static COHSEX form (COulomb Hole and Screened EXchange) of the \( GW \) self-energy \[3\]. \( \Sigma_{\text{xc}} \) can then be split in two parts \( \Sigma_{\text{COH}} + \Sigma_{\text{SEX}} \):

\[
\Sigma_{\text{COH}}(r, r') = \frac{1}{2}\delta W_N(r, r)\delta(r - r')
\]

(2a)

\[
\Sigma_{\text{SEX}}(r, r') = -W_N(r, r')\gamma(r, r')
\]

(2b)

\[
\gamma(r, r') = \sum_{n' \in \{\sigma\}} \varphi^*_n(r) \varphi_{n'}(r')
\]

(2c)

where the sum runs over the occupied states with a given spin \( \sigma \). \( W_N(r, r') \) is the \textit{screened} Coulomb interaction, i.e. the total potential created at point \( r' \) by a test unit charge at point \( r \) \[14\]. It can also be split in the bare potential \( v(r, r') = 1/|r - r'| \) created by this test charge, plus the response \( \delta W_N(r, r') = W_N(r, r') - v(r, r') \) of the valence electrons. \( \Sigma_{\text{SEX}} \) has the same functional form as the Hartree-Fock exchange, but with a screened instead of a
bare Coulomb interaction. $\Sigma_{\text{COH}}(r,r')$ describes the interaction of a carrier at point $r$ with the valence electrons which dynamically respond to its motion.

Before addressing the impurity problem, we shall discuss the form of $W_N(r,r')$ in bulk materials and nanowires. In a solid, a test charge $q_t = +1$ at point $r$ attracts valence electrons in a small “cloud” around (over $\sim$ a bond length). This cloud contains a total charge $q_c = -(1 - 1/\kappa)$, where $\kappa$ is the static dielectric constant of the material. The electrons are actually dragged from the surface of the system, where they leave an opposite polarization (or “image”) charge $q_s = -q_c$. In bulk, these image charges are infinitely far away, so that the long-range potential created by the test charge is simply $W_N(r,r') \sim (q_t + q_c)/|r - r'| = 1/(\kappa|r - r'|)$. In a nanowire, however, the electrons are dragged within a few $R$’s only from $q_t$, so that the transfer of charges from the surface to the cloud becomes shorter and shorter-ranged with decreasing $R$. The screening is therefore reduced by $q_s$ and the potential ultimately tends to $W_N(r,r') \sim 1/|r - r'|$ when $R \to 0$ (i.e., the test charge mostly sees vacuum around for small $R$’s). This simple picture is consistent with classical electrostatics (where the surface polarization charges are given by the discontinuity of the electric field), and fully supported by quantum calculations [15, 16].

As discussed previously, the many-body binding energy of an impurity can be computed as $E_b = A^d(N) - A^p(N)$, the difference between the affinities of the ionized impurity and pristine systems. They fulfill the equation $H^{p,d}(N)\phi^{p,d}_{N+1} = -A^{p,d}(N)\phi^{p,d}_{N+1}$, where $H^p(N)$ and $H^d(N)$ are the respective quasiparticle hamiltonians:

$$H^{p,d}(N) = -\frac{1}{2}\Delta + v_{\text{ion}}^{p,d} + v_{h}^{p,d} + \Sigma_{\text{SEX}}^{p,d} + \Sigma_{\text{COH}}^{p,d}.$$  

The physics of the impurity is most easily brought out from the difference between $H^p(N)$ and $H^d(N)$. On one hand, the extra proton of the ionized impurity is screened by the valence electrons through the Hartree potential $v_h^d$. Neglecting short range chemical corrections in a first approximation [17], we can therefore write:

$$[v_{\text{ion}}^d + v_h^d] - [v_{\text{ion}}^p + v_h^p] \simeq -W_N^d(r_i, r),$$  

where $r_i$ is the impurity position. On the other hand, we do not expect significant differences between the screened Coulomb interactions $W_N^d$ and $W_N^p$, nor between the one-particle density matrices $\gamma^d$ and $\gamma^p$, except possibly right around the donor and surface (image charges), on length scales much shorter than the Bohr radius of the impurity. Hence, $\Sigma_{\text{SEX}}^d \simeq \Sigma_{\text{SEX}}^p$. 


\[ \Sigma^d_{\text{COH}} \simeq \Sigma^p_{\text{COH}}, \text{ and:} \]
\[ H^d(N) \simeq H^p(N) - W_N(r_i, r). \]  \hfill (5)

In a first approximation, the quasiparticle hamiltonian of the ionized impurity is the quasiparticle hamiltonian of the pristine system plus the screened Coulomb potential of a unit charge at the impurity position. This is the usual "hydrogenic model" [7] used in Refs. [7–9] to calculate \( E_b \) in Si NWs.

The electron is therefore bound to the impurity by the screened Coulomb interaction \( W_N(r_i, r) \). In bulk silicon, \( W_N(r_i, r) \sim 1/|\kappa| |r_i - r| \) and \( E_b \simeq 50 \text{ meV} \). In a nanowire, however, the electron also interacts with the image charges of the donor. Since the total surface polarization charge is \( q_s = (1-1/\kappa) \gg 1/\kappa \), this leads to a large \( \propto 1/R \) enhancement of \( E_b \) with decreasing \( R \) [7].

Let us now compute the binding energy \( E_b = I^d(N+1) - I^p(N+1) \) from the ionization energy of the neutral impurity. \( I^d(N+1) \) and \( I^p(N+1) \) fulfill the equation \( H^p,d(N+1) \phi^d_{N+1} = -I^p,d(N+1) \phi^p_{N+1} \), where, as before, \( H^p(N+1) \) and \( H^d(N+1) \) are the quasiparticle hamiltonians of the \((N+1)\)-electron pristine and impurity systems. In the latter, the HOMO \( \phi^d_{N+1} \) is the occupied bound state of the impurity. The neutral impurity as a whole now introduces a localized perturbation of the pristine system which is screened by the valence electrons. We can therefore write:

\[ [v^d_{\text{ion}} + v^d_h] - [v^p_{\text{ion}} + v^p_h] \simeq -W^d_{N+1}(r_i, r) + v_b(r), \]  \hfill (6)

where:

\[ v_b(r) = \int d^3r' W^d_{N+1}(r, r')|\phi^d_{N+1}(r')|^2 \]  \hfill (7)

accounts for the screening of the bound state potential. Assuming again that \( W^p_{N+1} \simeq W^d_{N+1} \), and that the valence band wave functions \( \varphi_1, \ldots, \varphi_N \) are little affected by the neutral impurity, we further get:

\[ \Sigma^d_{\text{SEX}}(r, r') - \Sigma^p_{\text{SEX}}(r, r') \simeq -W_{N+1}(r, r') \times [\varphi^{d*}_{N+1}(r) \varphi^d_{N+1}(r') - \varphi^{p*}_{N+1}(r) \varphi^p_{N+1}(r')]. \]  \hfill (8)

The second term can be neglected in bulk and nanowires where \( \varphi^p_{N+1} \) is an extended state. The first term cancels \( v_b(r) \) when applied to the HOMO \( \phi^d_{N+1} \). The effective hamiltonian for the bound electron therefore reads:

\[ H^d(N + 1) \simeq H^p(N + 1) - W_{N+1}(r_i, r). \]  \hfill (9)
In principle, \( I(N + 1) = A(N) \) and we should have recovered the same equation as before [Eq. (5)]. Here \( H^p(N) \) is however replaced with \( H^p(N + 1) \) and \( W_N \) with \( W_{N+1} \). Since \( \varphi_{N+1}^p \) is an extended state, \( H^p(N + 1) \) and \( H^p(N) \) also primarily differ by the substitution \( W_N \to W_{N+1} \). The appearance of \( W_{N+1} \) introduces a residual “self-correlation” error in the \( GW \) ionization energies [18], which is however expected to be limited in solids. We can therefore conclude that \( GW \) provides a consistent description of the binding energies, whether computed from \( A^d(N) \) or \( I^d(N + 1) \).

This paragraph clearly demonstrates the importance of screened exchange in the calculation of \( I^d(N + 1) \). Screened exchange indeed cancels the unphysical screened interaction of the bound electron with itself which arises from \( v_b(r) \) in Eq. (6). \( H^d(N + 1) \) is therefore the hamiltonian of a charged system as expected (the bound electron interacts with \( N + 1 \) ionic charges but \( N \) electrons). Such spurious self-interactions are a serious issue in self-consistent descriptions of occupied localized states. In this respect, we would like to point out that the Hartree-Fock (HF) bare exchange \( \Sigma_x(r, r') = -v(r, r') \gamma(r, r') \) does not properly correct the screened self-interactions appearing in solids. Following the same lines as before, the HF hamiltonian of the HOMO of the neutral impurity can indeed be written \( H_{HF}^d(N + 1) \simeq H_{HF}^p(N + 1) - W_{N+1}(r_i, r) + v_{st}^v(r) \), where:

\[
v_{st}^v(r) = \int d^3r' \left[ W_{N+1}(r, r') - v(r, r') \right] |\varphi_{N+1}^d(r')|^2.
\]

(10)

\( v_{st}^v(r) \) is the spurious potential created by the valence electrons in response to the bound state density \( |\varphi_{N+1}^d(r)|^2 \), i.e. the potential created by a diffuse charge \( \rho_{eff}(r) \simeq (1 - 1/\kappa) \times |\varphi_{N+1}^d(r)|^2 \) plus the opposite surface polarization charge \( q_s = -(1 - 1/\kappa) \). These surface polarization charges balance those embedded in the impurity potential \( W_{N+1}(r_i, r) \) (equivalent, as discussed before, to the potential of a net charge \( 1/\kappa \) at \( r_i \) and \( q_s = (1 - 1/\kappa) \) at the surface). \( H_{HF}^d(N + 1) \) is therefore approximately equal to the hamiltonian of the pristine system plus the bare Coulomb potential of a unit charge spread around the impurity (the charge \( 1/\kappa \) at the impurity position plus the diffuse charge \( \rho_{eff} \) around). As a consequence, \( \rho_{eff} \) plays the role in the HF approximation of an effective polarization charge, mislocalized within the scale of the Bohr radius instead of the surface. The relative error on \( E_b \) should be limited in thin nanowires (where the Bohr radius is comparable to \( R \)), and maximum in bulk. The implications for hybrid functionals in DFT will be discussed in the next paragraph.

The binding energy in DFT – We now discuss the binding energy of the donor within
DFT. For the sake of simplicity, we first focus on the local density (LDA) and generalized gradients (GGA) approximations, then address the case of hybrid functionals. In DFT, the ground-state density $\rho(r)$ of the $N$-electron system is computed from the eigenstates of the Kohn-Sham hamiltonian [19]:

$$-\frac{1}{2}\Delta r \varphi_n(r) + [v_{\text{ion}} + v_h + v_{xc}](r)\varphi_n(r) = E_n\varphi_n(r) \quad (11)$$

where $v_{xc}(r)$ is the exchange-correlation potential. In LDA and GGA, $v_{xc}(r) \equiv v_{xc}(\rho(r))$ is a function of the local density $\rho(r)$ and of its derivatives. DFT is known to underestimate the band gap energy of semiconductors [13]. Still, we show below that DFT should succeed in the calculation of the binding energy from the LUMO of the ionized impurity, but that present functionals fail on the neutral impurity.

Let us first compute $E_b = A^d(N) - A^p(N)$ from the LUMOs of the KS hamiltonians $H_{KS}^p(N)$ and $H_{KS}^d(N)$. The previous arguments are also valid in DFT: The extra proton of the donor is screened by the valence electrons, so that Eq. (11) still holds. We do not, moreover, expect much differences between $v_{xc}^p(r)$ and $v_{xc}^d(r)$, except possibly right around the impurity. Therefore, in a first approximation:

$$H_{KS}^d(N) \simeq H_{KS}^p(N) - W_N(r_i, r) \quad (12)$$

We hence recover the hydrogenic model as before [Eq. (5)]. The KS hamiltonian of the ionized impurity thus embeds the same extra physics (with respect to the hamiltonian of the pristine system) as the GW approximation: *Although the LUMO energies are typically underestimated by DFT, the binding energies computed as the difference between the KS LUMOs of the ionized impurity and pristine systems should be reasonably accurate.* This only holds, of course, as long as the binding energy is not too large with respect to the DFT band gap.

Let us now compute $E_b = I^d(N+1) - I^p(N+1)$ from the HOMOs of the KS hamiltonians $H_{KS}^p(N+1)$ and $H_{KS}^d(N+1)$. The KS wave function $\varphi_{N+1}^d$ is the occupied bound state of the impurity. The neutral impurity as a whole is again screened by the valence electrons [Eq. (13)]. The exchange-correlation potential $v_{xc}(r)$ is also affected by the extra bound state density around the impurity. We hence get :

$$H_{KS}^d(N+1) \simeq H_{KS}^p(N+1) - W_{N+1}(r_i, r) + v_b(r) + \Delta v_{xc}(r) \quad (13)$$
where $\Delta v_{xc}(r) = v_{xc}^d(r) - v_{xc}^p(r)$. At variance with the $GW$ approximation, $\Delta v_{xc}(r)$, a local density correction within the Bohr radius, can not be expected to cancel $v_b(r)$ [Eq. (7)], a long-range Coulomb term. This results in $i)$ a self-interaction error, and $ii)$, an almost complete cancellation of the interaction of the electron with the image charges of the impurity. Indeed, $W_{N+1}(r, r)$ and $v_b(r)$ are the potentials created by two opposite charges (the ionized impurity and bound electron), leaving no net charge in the hamiltonian. Both errors give rise to an increase of the impurity level and to a decrease of the binding energy. Although this is especially sensitive in thin nanowires, where the enhancement of $E_b$ is mostly due to the interaction with the image charges, the LDA and GGA would fail up to the bulk [where the impurity potential decreases exponentially instead of $1/(|r - r_i|)$]. We stress that the calculation of the ionization energy or affinity of the impurity as a difference of total energies [10], which involves the neutral impurity as the initial or final state, suffers from the same deficiencies in the LDA or GGA.

Application to Si nanowires – The binding energies of dopant impurities in Si NWs have been previously computed from the KS HOMO of the neutral impurity using GGA and a hybrid functional (HGH) [11], i.e. a mixture of Hartree-Fock bare exchange with GGA. As discussed previously, bare exchange does not localize the polarization charges properly, the error being however likely limited in thin nanowires (the total charge being correct). The GGA results of Ref. [11] are therefore expected to completely miss image charge effects, while the HGH results, which include 12% bare exchange, are expected to account for $\simeq$ 12% of the interactions with the image charges (even though mislocalized). As a consequence, the difference between the GGA and HGH results of Ref. [11] should be approximately 12% of the image charge correction given by Eq. (3) of Ref. [7], that is, 0.12 eV for $R = 1$ nm, 0.17 eV for $R = 0.75$ nm, and 0.25 eV for $R = 0.5$ nm. This is actually in good agreement with the data of Table I of Ref. [11].

To further support the above conclusions, we have computed the binding energy of a P impurity at the center of a hydrogen passivated, [110]-oriented Si nanowire with diameter $d = 1.73$ nm, either as $E_b^I = I^d(N + 1) - A^p(N)$, or as $E_b^A = A^d(N) - A^p(N)$, using KS HOMOs and LUMOs as ionization energies and affinities. The LDA was used in a wavelet basis set as implemented in the BigDFT code [20]. The neutral impurity was first relaxed in a 660 atoms supercell. Since the treatment of a charged system is still problematic within such a supercell approach, $A^d(N)$ (as well as $I^d(N + 1)$ and $A^p(N)$ for consistency) were
actually computed in finite rods with lengths $l$ up to 9.2 nm (1584 atoms). These rods were built from the original 660 atoms supercell by connecting segments of pristine nanowires and hydrogen passivated ends. The binding energies computed from the charged and neutral impurities are respectively $E_A^b = 0.93$ eV and $E_I^b = 0.06$ eV for $l = 9.2$ nm. As expected, $E_A^b$ is much larger than $E_I^b$, and in good agreement with the semi-empirical model of Ref. [7] ($E_b = 0.92$ eV when $l \to \infty$). This confirms that present functionals are able to predict the binding energies of impurities or defects from the KS hamiltonian of the charged defect.

To conclude, we have shown, by a formal comparison with the GW approximation, that the donor binding energies computed from the Kohn-Sham hamiltonians of neutral impurities can be strongly underestimated in semiconductor nanostructures (even with hybrid functionals). This is due to the lack of screened exchange in the present functionals, and explains the discrepancies between Refs. [10, 11] and previous works [7, 9]. The binding energy of a donor should preferably be computed as the difference between the Kohn-Sham LUMOs of the ionized impurity and pristine systems. This provides a reasonable substitute for much more expensive GW calculations of defect bound states in solids.

We thank L. Wirtz, X. Blase and H. Mera for fruitful discussions. This work was supported by the french ANR project “QuantaMonde” (contract ANR-07-NANO-023-02). The calculations were run at the CCRT and CINES.

* Electronic address: yniquet@cea.fr

[1] M. T. Björk et al., Nature Nanotechnology 4, 103 (2009).
[2] J. Yoon et al., Appl. Phys. Lett. 94, 142102 (2009).
[3] L. Hedin and S. Lundqvist, Solid State Physics 23, ed. by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, London 1969).
[4] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. 74, 601 (2002).
[5] W. Kohn and J. M. Luttinger, Phys. Rev. 98, 915 (1955).
[6] A. S. Martins et al., Phys. Rev. B 65, 245205 (2002).
[7] M. Diarra et al., Phys. Rev. B 75, 045301 (2007).
[8] M. Diarra et al., J. Appl. Phys. 103, 073703 (2008).
[9] B. Li et al., Phys. Rev. B 77, 115335 (2008).
More precisely, the screened coulomb interaction in the random phase approximation (RPA).

We also neglect non-linear and exchange-correlation effects beyond the RPA in Eq. (1). They would not change the conclusions drawn in this letter.

P. Romaniello, S. Guyot and L. Reining, J. Chem. Phys. 131, 154111 (2009).

L. Genovese et al., J. Chem. Phys. 129, 014109 (2008)