Applications of Density Functional Theory to Semiconductors

José Luís Martins
Instituto de Engenharia de Sistemas e Computadores, Rua Alves Redol 9, 1000 Lisboa, Portugal
and
Departamento de Física, Instituto Superior Técnico, Av.Rovisco Pais 1, 1096 Lisboa, Portugal

Abstract

Density functional theory has been extensively applied to the study of the electronic and structural properties of semiconductors. Whereas the structural properties are fairly well calculated with the local density approximation, this approximation fails for the value of the band gap, the most important of the electronic properties. The generalized gradient approximation does not improve the picture, except for the binding energy of the crystal. Only more complicated approaches like the GW approximation give accurate values for the band gap in semiconductors. We will review briefly the situation, and discuss the physical origin of the band-gap problem.

1. Introduction

Semiconductors are the key material for the electronic industry. Therefore the electronic and structural properties of semiconductors have been the subject of many applications of density functional theory. Bulk properties, such as band gaps or lattice constants, of the most common semiconductors are accurately known from experiment, and provide excellent reference data to test the physical accuracy of different approximations used within the framework of density functional theory. In the case of less common semiconductors, or for defect and surface and interface properties, density functional theory has been used as a predictive tool, and many phenomena have been elucidated from an interaction between simulation and experiment.
In the next section we will discuss how the local density approximation of density functional theory describes structural properties of semiconductors, for bulk materials and for defects and surfaces. In section 3 we will discuss how and why the local density approximation fails for the band gap in semiconductors, and review how methods beyond local-density can improve the calculation of the band structure.

2. Structural properties

2.1. Bulk semiconductors

Density functional theory allows the calculation of the ground state total energy, E, of a system bound by electrons. In the case of crystals this energy can be calculated as a function of volume, E(V). The minimum of that function gives the equilibrium volume and hence the lattice constant, subtracting the energy of separate atoms gives the binding energy and the second derivative at the minimum gives the bulk modulus $B = V(d^2E/dV^2)$. The first derivative gives the pressure $p(V) = -dE(V)/dV$ and therefore the pressure--volume equation of state.
The calculation of the structural properties of bulk semiconductors is a good test of the accuracy of the Local Density Approximation (LDA) of the density functional theory [1-4]. The general trend is for the lattice constant to be slightly under-estimated, the binding energy to be over-estimated by five to ten percent and the bulk modulus to be overestimated by a few percent. In contrast to the results for first row molecules, there is no systematic improvement in the use of the Generalized Gradient Approximation (GGA), except in

an improvement of the binding energy [5-7]. That improvement is related to an increase in the binding of the isolated atom, as the GGA prefers situations where regions of space have low electronic density, and therefore it does not seem that GGA is an improvement over LDA for the bulk semiconductor properties. Table 1 shows an example of a few calculations for Si. The earlier LDA calculations seem to be the most accurate, but that is due to the use of a small basis set. Computer resources at that time did not allow the use of larger basis sets. (It is interesting to note that those calculations were performed on a computer that was a 1000 times slower and had 500 times less memory than a good personal computer in 1999.) A smaller basis set decreases the binding energy, increases the bond-length and decreases the bulk modulus, improving systematically the agreement with experiment.

Table 1 – Structural properties of Si, lattice constant a, binding energy E_B and bulk modulus B, calculated with the LDA and GGA (PW91) as compared to experiment.

	LDA[a]	LDA[a]	GGA[b]	Experiment[c]
a ()	5.45	5.41	5.50	5.429
E_B (eV)	4.67	5.28	4.65[d]	4.63
B (GPa)	98	96	83	99

[a] Ref.[1]
[b] Ref. [5]
[c] Cited in ref.[5]
[d] Ref.[7]

For non-zincblende semiconductors, the determination of the zero pressure structure is complicated, as the total energy may depend on the lattice bond lengths and angles, and internal atomic coordinates, and therefore one must find the minimum of the total energy with respect to several parameters instead of just the volume. For more than a couple of free parameters that search should be automatic, and variable cell shape (VCS) molecular dynamics (MD) algorithms have been developed for the optimization of crystal structures [8, 9]. Table2 shows an example of the calculation of lattice constants and internal parameters ofternary calcium nitrides [10] compared with the experimental results [11, 12]. The structure of these crystals correspond to an orthorhombic distortion of a cubic perovskite structure. One can see that the subtle distortions of the lattice are well described by the LDA, and that VCS-MD is quite effective in the determination of equilibrium geometries.

Table 2 – The experimental values for for the lattice and distortion parameters of the compounds $AsNCa_3$ and $PNCa_3$ as compared with experiment. These compounds have an orthorhombic distorted perovskite structure, with lattice constants a, b, c, and distortion parameters, Δ_1, Δ_2, δ, γ_1, γ_2, λ_1, and λ_2. For the ideal perovskite structure $a = b = c/\sqrt{2}$, and all distortion parameters are zero. Experimental values for the distortion parameters of $PNCa_3$ are not available.

	Experiment[a]	Simulation[b]	Experiment[c]	Simulation[b]
Compound	$AsNCa_3$	$AsNCa_3$	$PNCa_3$	$PNCa_3$
T (K)	15	0	305	0
a (Å)	6.716	6.720	6.709	6.707
b (Å)	6.711	6.715	6.658	6.659
c (Å)	9.520	9.526	9.452	9.451
Δ_1	0.0329	0.0400	-	0.0464
Δ_2	0.0321	0.0400	-	0.0459
δ	0.0209	0.0265	-	0.0396
γ_1	0.0399	0.0510	-	0.0747
γ_2	0.0048	0.0100	-	0.0220
λ_1	0.0000	0.0032	-	0.0084
λ_2	0.017	0.0263	-	0.0449

[a] Ref. [12]
[b] Ref. [10]
[c] Ref. [11]

2.2. Defects and interfaces in semiconductors

Defects and interfaces in semiconductors have been extensively studied with LDA. From that vast literature two representative examples were chosen among many to illustrate the interplay between theory and experiment required to reach an improved knowledge of semiconductors.
In GaAs there is an important deep defect labeled EL2 that is related to the semi-insulating properties of that crystal. EL2 is responsible for a sub-band-gap optical absorption around 1.2 eV. Below 100 K this absorption is quenched in semi-insulating samples (that is, its intensity decreases with exposure to light). Heating above that temperature restores the absorption. This behavior indicates the presence of a meta-stable defect. The EL2 defect was known to be associated with an As anti-site defect, that is, a defect where one Ga atom is replaced in the lattice by an As atom. However it was believed

that a simple anti-site defect could not have a complex behavior with meta-stable configurations and several complex defects like an anti-site As associated with an interstitial As atom were proposed. Uniaxial pressure studies [13] indicated that the defect had tetrahedral symmetry and should therefore be a simple point defect, but the overall experimental evidence was not clear. Calculations showed that the As anti-site defect was stable against small displacements but that there was a meta-stable state characterized by a ruptured As — As bond. Furthermore that meta-stable state did not have electronic states in the band gap, providing an explanation for the quenching of the optical absorption [14]. Calculations for the competing As anti-site As interstitial model failed to find configurations stable with respect to a split interstitial formation, ruling against that model [15]. While LDA calculations cannot be used as proof, the identification of the source of the EL2 defect, is an example of a problem where calculations have been extremely useful.
Recently the isolated As anti-site buried below the surface has been identified in scanning tunneling microscopy (STM) images of the GaAs(110) surface [16]. Some of the images exhibited intriguing "satellite peaks". Total energy calculations (using super-cells with 80 and 200 atoms) elucidated how these satellite peaks were related to the electronic structure of the defects [17].

3. Band Structure

One of the biggest failures of LDA is that it predicts the incorrect band gap for semiconductors, giving half the experimental value for Si, and even predicting that Ge is a metal. Of course, when one goes back to the foundations of density functional theory, one finds that the eigenvalues of the Kohn-Sham equation are just Lagrange multipliers without physical meaning, so one should not be too surprised if they do not agree with experiment. Real band-structure should be calculated by a many-body quasi-particle approach such as the GW method [18]. In the GW method one must calculate the linear response of the system, and that is usually based in the LDA static dielectric matrix (a ground state property). GW calculations require more computing resources than the simpler LDA calculations.
As can be seen in Table 3, GW calculations of the band gap of semiconductors [18] are quite close to experiment, whereas the GGA eigenvalues are very close to LDA and far from the experimental values. Once again for bulk properties GGA is not an improvement over LDA.

Table 3 – Band gaps of selected semiconductors (in eV) calculated with LDA, GGA (PW91) and GW as compared with experiment. Both LDA and GGA fail, while the quasi-particle GW approach is quite accurate.

	LDA[a]	GGA[b]	GW[a]	Experiment[c]
C	3.9		5.6	4.48
Si	0.5	0.47	1.16	1.17
Ge	-0.26	-0.16	0.73	0.74
GaAs	0.12	0.29	1.42	1.52
AlAS	1.28		2.01	2.24

[a] Ref. [18]
[b] Ref. [5]
[c] Ref. [18]

Table 4 – The experimental ionization potentials of C and Si, I, as compared to the differences in calculated LDA total energies between the neutral and ionized atom, ΔE , the LDA eigenvalues of the neutral atom, $\varepsilon(1)$, and the LDA eigenvalues of the "half-ionized" atom, $-\varepsilon(0.5)$ (Slater's transition state). The differences in total energies and Slater's transition state are quite accurate, while the eigenvalues give incorrect ionization energies. The transition state also predicts correctly the ionization of the deeper s state.

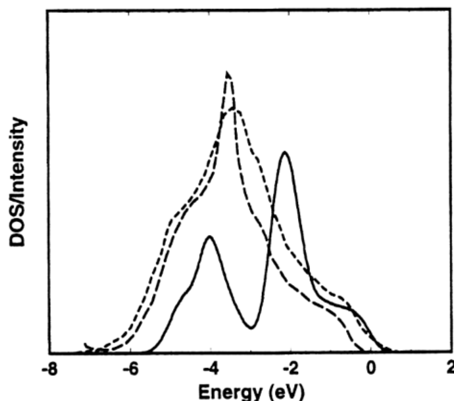
		I (eV)	ΔE (eV)	$-\varepsilon(1)$ (eV)	$-\varepsilon(0.5)$ (eV)
C	p	11.26	11.72	6.15	11.60
	s		16.53	11.94	16.67
Si	p	8.15	8.25	4.60	8.19
	s		13.54	9.91	13.59

The bands obtained from the Kohn-Sham eigenvalues are close enough to the one-particle excitation energies that it is worth investigating their meaning, and why and when the LDA eigenvalues fail to give the right one-particle excitation spectra. The exact density functional gives the correct ground state electron density. For finite systems, one can relate the asymptotic behavior of the charge density to both the real ionization potential and to the highest occupied Kohn-Sham eigenvalue, leading to the conclusion that they must be equal for the exact Kohn-Sham functional. If we compute atomic ionization potentials from the difference in total energy values between the neutral and the ionized atom, we obtain values quite close to experiment with the local den-

sity approximation as can be seen in Table 4. However the highest occupied Kohn-Sham eigenvalue is very far from the experimental value. This is mainly due to the incomplete cancellation of the orbital self energy in the Hartree term by local-density exchange [19]. We can calculate accurate atomic ionization energies with the LDA, but we cannot use the orbital eigenvalues for that.

For a semiconductor the band gap is the difference in energy between removing and adding an electron to the neutral solid; for a finite system this corresponds to the difference between the ionization potential and electronic affinity, that is the difference in ionization potential of the neutral system and the system with one extra electron, which should be given by the difference between the energies of the highest occupied Kohn-Sham orbitals of the neutral and the anionic systems. One could expect that for a large system adding one electron does not change the band structure and therefore a good approximate functional for the energy such as LDA should give gaps from orbital eigenvalues that are close to experiment. However it has been shown that for an extended system (where there is no privileged energy reference) this is not the case, which means that the exact exchange and correlation potential must have a jump when one adds an electron to a neutral insulator [20, 21].

Figure 1 – The calculated density of states for the valence band of AgCl (solid line) is compared to the intensity (in an arbitrary scale) of two different photoemission experiments (dashed lines). The shape of the calculated density of states does not agree with the experimental data because the relative energies of the Ag 4d and Cl 3p orbitals are incorrectly predicted by the LDA.

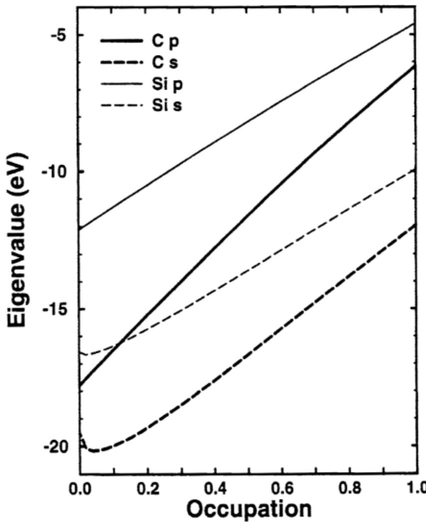


Because of the emphasis on the LDA band gap problem, one common misconception is that the band structure calculated from the LDA Kohn-Sham eigenvalues is correct except for the value of the gap itself, and that adding a single number to all unoccupied bands (the scissors operator) one should get the correct picture. That is not always the case: In graphite, which is a layered metal, the top of the occupied s band is predicted by LDA to be 3.0 eV below the Fermi level [22], while in experiment it is found 4.3 eV below the Fermi level, a large 1.3 eV discrepancy. In the II-VI semiconductor ZnS LDA calculations predict the d-band of Zn to be around 6.5 eV below the valence band maximum [23] whereas it is found experimentally 10 eV below. Figure 1 shows a comparison of the LDA density of states with photo-emission intensity for AgCl showing again that the d-electrons of silver are not predicted to appear in the right energy range [24]. The problem with LDA Kohn-Sham eigenvalues is not just for the band gap.
While the meaning of fractional occupation numbers in Kohn-Sham density functional formalism is not clear, in a calculation one can occupy the states with arbitrary real numbers. With this fractional occupation the calculated total energy is a function of the occupation numbers, $E(f_1,\dots f_n)$, and the eigenvalues acquire a new meaning,

$$E(f_1,\dots f_n) = \frac{\partial E}{\partial f_i} \tag{1}$$

that is, they are the partial derivatives of the energy with respect to occupation. One could expect that the dependence of the eigenvalues on their own occupation would be weak. In fact it is very strong in LDA due to the incomplete cancellation of the orbital self energy in the Hartree term by local-density exchange [19]. Figure 2 shows how the eigenvalues of atomic carbon and silicon depend on occupation when one removes the last s or p electron. Occupation 1 corresponds to the neutral system, and occupation 0 to the singly ionized atom. Integrating equation 1 we obtain that the ionization potential calculated from a difference in total energies, $E(\dots,0,\dots) - E(\dots,1,\dots)$ is just minus the area above the curves in the figure. As we are integrating from 0 to 1 functions that are almost linear, the eigenvalue at half-occupation, $\varepsilon_i(\dots,0.5,\dots) \approx E(\dots,1,\dots) - E(\dots,0,\dots)$ is a good estimate of minus the ionization energy (this is Slater's transition state method). From Table 4 we can see that Slater's transition state predicts quite accurately the differences in total energies.

Figure 2 – The 2s and the 2p eigenvalues for C and the 3s and 3p eigenvalues for Si calculated in the LDA are shown as a function of fractional occupation.



The error in atomic eigenvalues shown for atoms in Table 4 should be propagated when we assemble a crystal out of atoms. In the case of Si the occupied valence bands are formed from the atomic s and p states with different degrees of hybridization. As the lines in figure 2 for Si are almost parallel, the difference between ε_s and ε_p eigenvalues of the neutral atom is almost the same as the difference in ionization potentials. The main error of LDA is just a rigid shift of the eigenvalues, which is irrelevant in an infinite solid. The valence bands of Si are well described in LDA. The conduction band minimum is mainly an interstitial state, quite different from a superposition of atomic s and p states, and the prediction of LDA for the band gap is inaccurate. In the case of carbon the lines in Fig. 2 are not parallel. In graphite the separation of s and p states should reflect the separation of the parent s and p states in the atom. As the difference in atomic energies is under-estimated by the orbital eigenvalues (see Fig. 2 or Table 4), it is not surprising that the separation between occupied s and p bands in graphite is also under-estimated.
The basic rule is that the smaller the atomic orbital size, the larger the self-interaction, and the higher the associated orbital eigenvalues. If one valence orbital has a smaller extension compared to another (like the 2p orbital in carbon compared to the 2s or the 4d of Ag compared with the 5s) then it will

appear higher in energy relatively to that other orbital in LDA calculations than in experiment. If the orbitals have the same extension, then differences in eigenvalues should predict differences in energies.

4. References

1. M. T. Yin and M. L. Cohen, Phys. Rev. Lett. 45, 1004 (1980).
2. A. Zunger, Phys. Rev. B 21, 4785 (1980).
3. K. J. Chang and M. L. Cohen, Phys. Rev. B 31, 7819 (1985).
4. R. J. Neeads and A. Mujica, Phys. Rev. B 51, 9652 (1995).
5. C. Filippi, D. J. Singh, and C. J. Umrigar, Phys. Rev. B 50, 14947 (1994).
6. N. Moll, M. Bockstedte, M. Fuchs, E. Pelkhe, and M. Scheffler, Phys. Rev. B 52, 2550 (1995).
7. A. Dal Corso, A. Pasquarello, A. Baldereschi and R. Car, Phys. Rev. B 53, 1180 (1996).
8. R. M. Wentzcovitch, J. L.Martins, and G. D.Price, Phys. Rev. Lett. 70, 3947 (1993).
9. I. Souza and J. L. Martins, Phys. Rev. B 55, 8733 (1997).
10. P. R. Vansant, P. E. Van Camp, V. E. Van Doren, and J. L. Martins, Phys. Rev. B 57, 7615 (1998).
11. M. Y. Chern, D. A. Vennos and F. J. Disalvo, J. Sol. Stat. Chem., 96, 415 (1992).
12. M. Y. Chern, F. J. Disalvo, J. B. Parise and J. A. Goldstone, J. Sol. Stat. Chem. 96, 426 (1992).
13. M. Kaminska, M. Skowronski and W. Kuszko, Phys. Rev. Lett. 55, 2204 (1985).
14. D. J. Chadi and K. J. Chang, Phys. Rev. Lett. 60, 2187 (1988).
15. D. J. Chadi, Phys. Rev. B 46, 15053 (1992).
16. R. M. Feenstra, J. M. Woodall, and G. D. Pettit, Phys. Rev. Lett. 71, 1176 (1993).
17. R. B. Capaz, K. Cho and J. D. Joannopoulos, Phys. Rev. Lett. 75, 1811 (1995).
18. M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. 55, 1418 (1985); Phys. Rev. B 34, 5390 (1986); E. L. Shirley, X.-J. Zhu, and S. G. Louie, Phys. Rev. Lett. 69, 2955 (1992).
19. J. Perdew and A. Zunger, Phys. Rev. B 23, 5048 (1981)
20 J. Perdew and M. Levy, Phys. Rev. Lett 51, 1884 (1983)
21. L. Sham and M. Schlüter, Phys. Rev. B 32, 3883 (1985)
22. M. C. Schabel and J. L. Martins, Phys. Rev. B 46, 7185 (1992).
23. J. L. Martins, N. Troullier, and S.-H. Wei, Phys. Rev. B 43, 2213 (1991).
24. G. S. Nunes, P. B. Allen, and J. L. Martins, Solid State Commun. 105, 377 (1998).