SCREENING MODELS AND SIMPLIFIED GW APPROACHES: Si & GaN AS TEST CASES

M. Palummo and R. Del Sole
Dipartimento di Fisica, Università di Roma "Tor Vergata", via della Ricerca Scientifica 1, 00133 Roma, Italy
Lucia Reining
Laboratoire des Solides Irradiés, CNRS URA No 1380-CEA-DTA/DECM, École Polytechnique, 91128 Palaiseau, France
F. Bechstedt
Friedrich-Schiller-Universität, Physikalisch-Astronomische Fakultät, Max-Wien-Platz 1, D-07743 Jena, Germany
and
G. Cappellini
Istituto di Fisica, Facoltà di Medicina e Chirurgia, Università di Cagliari, Via della Pineta 77, I-09125 Cagliari, Italy

(Received 20 February 1995; accepted 8 March 1995 by F. Yndurain)

We compare models of diagonal and off-diagonal screening with LDA-RPA full calculations in Si and cubic GaN. Simplified GW calculations relying on these models are also compared with full GW calculations for the same materials. A recipe for simplified GW calculations is obtained which works well for small, moderate and wide gap semiconductors.

1. INTRODUCTION

AN ACCURATE calculation of the electronic band structure of semiconductors and insulators is possible within Green's function theory, according to the GW approximation of the exchange-correlation self-energy [1]. The application of this method [2] to complex systems, as surfaces, interfaces, clusters etc. is however hampered by the large computer time required; therefore GW calculations have up to now been mostly limited to simple systems (e.g., elemental or binary semiconductors). On the other hand, simplified GW schemes have been developed in the last years, in order to substantially reduce the computation time [3–5]; all these schemes involve model dielectric matrices, since the determination of the latter is the most time-consuming step of a GW calculation. In particular models of diagonal [6] and off-diagonal [7] screening and a simplified, yet accurate, GW method [5], which reduces the CPU time by a factor of 100 with respect to full GW calculations, have been developed by some of the present authors. They have been successfully tested in the case of silicon and of other semiconductors with moderate band gaps.

Since these methods involve approximate treatments of local-field (LF) effects, their reliability may decrease in the case of wide-gap or ionic semiconductors, where LF effects are more important. The purpose of the present communication is to test these methods in the case of ionic wide-gap semiconductors, among which we consider cubic GaN as a prototype. Furthermore, we repeat the calculations of [7] for Si, but using nonlocal pseudopotentials as we do for GaN, in order to eliminate the possibility that different results might be obtained in the two cases simply as a consequence of using different ingredients.

2. DIAGONAL SCREENING

Cappellini et al. [6] have introduced the following model for the diagonal dielectric function of a
Fig. 1. Dielectric function for GaN. Empty boxes are obtained after the RPA-DFT-LDA calculation, while the two lines are obtained from the model described here, using the theoretical value $\epsilon_0 = 6.14$ and two different values of $\alpha$ (thin solid line $\alpha = 1$; thick solid line $\alpha = 1.56$).

semiconductor:

$$e(q) = 1 + \left[\epsilon(0) - 1\right]^{-1} + \alpha(q/q_{TF})^2 + q^4/(4k_Fq_{TF}^2)$$

(1)

where the Fermi wavevector $k_F$ and the Thomas–Fermi wavevector $q_{TF}$ are calculated at the mean electron density. This expression interpolates between the correct behaviors at high and low $q$ vectors, and is forced to reproduce the static dielectric constant of the semiconductors at $q = 0$. The model has already been shown [6] to reproduce very well the RPA dielectric function, calculated using semiempirical pseudopotentials which yield the correct experimental gap [8], in the cases of Si, GaAs, AlAs and ZnSe, by choosing $\alpha = 1.56$. The accuracy is comparable with that of the Levine-Louie model [10], which is however described by a more complicated formula, not allowing analytical calculations of quantities of interest in the GW method.

In the case of GaN, empirical-pseudopotential calculations are not available, so that we carry out ab initio RPA calculations within the DFT-LDA approach. These calculations have already been used in obtaining the full-GW band structure of GaN [11]. A well-known shortcoming of these calculations is the overestimate of the static dielectric constant: in the case of cubic GaN, we get 6.14, to be compared with the available experimental value for the wurtzite phase, ranging between 5.4 and 5.8 [12]. The comparison of the model of equation (1) with the full calculation is shown in Fig. 1 for two values of $\alpha$, 1.56 and 1. Both values yield a reasonable description of the full calculation; however, $\alpha = 1$ yields a slightly better fit than the other value, at variance with the results obtained in the cases of Si, Ge, GaAs and ZnSe using empirical pseudopotentials. The same occurs in the case of our LDA–RPA calculations for Si. We believe that this is a consequence of fitting the LDA calculations, with a larger value at $q = 0$, rather than empirical-pseudopotential calculations. In the following we will use $\alpha = 1$, since we will compare our results for off-diagonal screening and simplified-GW methods with full calculations employing LDA–RPA screening.

3. OFF-DIAGONAL SCREENING

Because of the microscopic structure of the crystal, the effective interaction potential between two point charges at $r$ and $r'$, $W(r, r')$, depends separately on the two positions, and not only on their distance $r - r'$, as in free space. This is called the “local-field effect” on screening. The space Fourier transform of $W(r, r')$, and therefore of the inverse dielectric function $e^{-1}(r, r')$, depends on two wave vectors, $q + G$ and $q + G'$: $W(q + G, q + G')$. The model considered in Section 2 disregards this effect, namely it approximates the screened potential as $W(r, r')$, and neglects its Fourier transform for $G$ different from $G'$. In this section, we will generalize the model of Section 2 to account for local-field effects. We will consider two approximate methods of doing this, and will compare them with ab initio calculations.

The first method, proposed by Hybertsen and Louie in 1988 [13], is based on the insertion of the local density $\rho(r)$, instead of the average electronic density, in $k_F$ and $q_{TF}$ appearing in equation (1). The effective interaction $W(r, r')$ takes the form

$$W(r, r') = \frac{1}{2} \{W^h(r - r', n(r)) + W^h[r - r', n(r')]\}$$

(2)

from which the Fourier transform of the inverse dielectric function can be calculated:

$$c_{G,G'}^{-1}(q, 0) = \frac{1}{2} \left\{e^{-1}(q + G', \rho(r)) + e^{-1}(q + G', \rho(r)) e^{-i(G - G')^2} d^3 r\right\}$$

(3)

In the present work $e^{-1}(q + G', \rho(r))$ are the diagonal elements obtained with the model previously discussed, but evaluating $q$ at the local density $\rho(r)$. This will be called method A in the following.

The second method has been introduced by Bechstedt et al. in [7], based on a generalization of the $f$-sum rule to $G$ different from $G'$ [9]. The inverse
The dielectric matrix is obtained as

$$\epsilon^{-1}(q + G, q + G') = \delta_{q, q'} - \frac{\omega_p^2}{\Omega^2(q + G, q + G')} \times \frac{(q + G) \cdot (q + G') n(G - G')}{|q + G||q + G'| n(0)}$$

(4)

where $n(G)$ is the Fourier component of the electron density and $\omega_p = [4\pi e^2 n(0)/m]^{1/2}$. $\Omega(q + G, q + G')$ can be calculated from first principles. In order to reduce the computational work, an ansatz was proposed in [7] for $\Omega(q + G, q + G')$, starting from equation (1): in the case of off-diagonal matrix elements, $q$ in equation (1) is replaced by the average of $q + G$ and $q + G'$, obtaining

$$\Omega^2(q + G, q + G') = \omega_p^2 \left[ \frac{\epsilon_0}{\epsilon_0 - 1} + \frac{|q + \frac{1}{2}(G + G')|^2}{q_{TF}^2} \right] + \left( \frac{n^2}{2m} |q + \frac{1}{2}(G + G')|^2 \right)^2$$

(5)

where $q_{TF}$ is calculated at the average electron density. This will be called in the following model B.

The comparison between the two models and full LDA–RPA calculations for Si and GaN is shown in Fig. 2. Model B seems to be always less accurate than model A, especially in the case of the ionic compound. In fact, while in the case of Si, the results obtained with model B are qualitatively correct, although the two components $G = (200)$, $G' = (111)$ and $G = (020)$, $G' = (111)$ are not well reproduced like the other ones, in the case of GaN, where larger off-diagonal elements occur, only model A is a good approximation of the full calculation. These findings are not contradictory with the fact that [6] can claim overall good results for model B in silicon. In fact, the deviations found in that reference are of the same order of magnitude as ours, but simply not visible due to the reduced scale of the figure. The difference between our calculations and those of [6] arises from the fact that, in [6], contrary to us, the comparison was made with a semi-empirical pseudopotential and the Bloch states were expanded in a Gaussian basis set. Let us emphasize that, in our case, the comparison is unbiased, since all models and the full calculation are based on the same band-structure and on the same Fourier components of the electron density. We conclude therefore that model A of off-diagonal screening is always better than model B and it can be used in covalent and ionic, small- and wide-gap semiconductors, while the validity of model B is limited to covalent compounds.

**4. GW CALCULATIONS**

Before describing the simplified GW methods used and discussing their results vs full GW calculations, we give some computational details of our LDA, RPA and GW calculations. We use plane wave basis sets and norm-conserving pseudopotentials [14]. All calculations are performed at the experimental lattice parameters ($a = 10.26$ a.u. for Si, $a = 8.5$ a.u. for GaN). In the case of Si, a kinetic energy cut-off of $14$ Ry, $10$ special $k$ points in the irreducible BZ and the Von Barth Hedin exchange-correlation potential [15] have been used in the DFT-LDA calculation, while for the GW part we used $169$ plane waves for the dielectric matrix and for $\Sigma$, $10$ special $k$ points for $\epsilon(q)$ and $19$ for $\Sigma$. In the case of GaN, the same ingredients of [1] have been used: a kinetic energy cut-off of $100$ Ry, $2$ special $k$ points in the irreducible BZ, the Ceperley–Alder exchange-correlation potential [16], and nonlinear core corrections for gallium. We used $169$ plane waves for the
Table 1. Quasiparticle shifts for GaN calculated within method I. $A_{nk} = \Sigma_{\text{dyn}}^e(E_{nk}^0)$ and $\beta_{nk}$ are the dynamical terms. The last column gives the full GW results.

<table>
<thead>
<tr>
<th>State</th>
<th>Static shift (eV)</th>
<th>$A_{nk}$ (eV)</th>
<th>$\beta_{nk}$</th>
<th>Shift (eV)</th>
<th>GW shift (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaN ($\epsilon_0 = 6.14$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Gamma_{15\nu}$</td>
<td>-2.72</td>
<td>2.26</td>
<td>0.14</td>
<td>-0.41</td>
<td>-0.17</td>
</tr>
<tr>
<td>$\Gamma_{1c}$</td>
<td>-1.33</td>
<td>2.06</td>
<td>0.16</td>
<td>0.62</td>
<td>0.17</td>
</tr>
<tr>
<td>$X_{3\nu}$</td>
<td>-2.84</td>
<td>2.14</td>
<td>0.15</td>
<td>-0.61</td>
<td>-0.46</td>
</tr>
<tr>
<td>$X_{1c}$</td>
<td>+0.29</td>
<td>1.69</td>
<td>0.21</td>
<td>1.16</td>
<td>0.60</td>
</tr>
<tr>
<td>GaN ($\epsilon_0 = 5.35$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 1.5635$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Gamma_{15\nu}$</td>
<td>-2.43</td>
<td>1.85</td>
<td>0.13</td>
<td>-0.52</td>
<td></td>
</tr>
<tr>
<td>$\Gamma_{1c}$</td>
<td>-0.90</td>
<td>1.70</td>
<td>0.15</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>$X_{3\nu}$</td>
<td>-2.60</td>
<td>1.77</td>
<td>0.14</td>
<td>-0.73</td>
<td></td>
</tr>
<tr>
<td>$X_{1c}$</td>
<td>-0.21</td>
<td>1.43</td>
<td>0.20</td>
<td>1.01</td>
<td></td>
</tr>
</tbody>
</table>

The quasiparticle energies in the GW scheme are

$$E_{nk} = E_{nk}^0 + \frac{1}{1 + \beta_{nk}} \left\{ \Sigma_{nk} + \Sigma_{nk}^{\text{dyn}}(E_{nk}^0) - V_{xc}^{\text{LDA}} \right\}, \quad (6)$$

where the self-energy operator $\Sigma$ has been divided in a sum of a static and a dynamical contribution and $\beta_{nk}$ is the linear term of the expansion of $\Sigma_{\text{dyn}}^e$ in $E$ around the DFT–LDA eigenvalue $E_{nk}^0$.

The main approximation in [5] was introduced in the calculation of $\Sigma_{nk}^{\text{dyn}}(E_{nk}^0)$ and $\beta_{nk}$, which turn out to be given by simple integrals of the dielectric function. The remaining part, the static term, can be divided in two parts:

$$\Sigma''(r, r') = \frac{1}{2} \sum_{nk} \psi_{nk}^0(r) \psi_{nk}^0(r') [W(r, r'; 0) - v(r - r')]$$

$$- \sum_p \gamma^0(r) \psi_{nk}^0(r) \psi_{nk}^0(r') W_p(r, r'; 0). \quad (7)$$

The first term is the Coulomb hole $\Sigma_{\text{COH}}$ and the second is the screened exchange $\Sigma_{\text{SEX}}$, $\psi_{nk}^0(r)$ is the DFT–LDA wavefunction and $W(r, r'; 0)$ and $v(r - r')$ are the statically screened and bare Coulomb potentials, respectively. Using the dielectric function (1), and accounting for LF effects according to equation (2), the static Coulomb hole turns out to be a local potential given in analytical form [5, 6]. Therefore it does not require more computational effort than the exchange-correlation potential of the LDA, yielding no further computational problem.

Within the same approximations, the screened exchange is

$$\Sigma_{nk}^{\text{SEX}} = - \sum_{n'k'} \left\{ \psi_{nk'}^0(r') \psi_{nk}^0(r') W_{nk'}^k(r - r'; n) \right\} \times \psi_{nk'}^0(r) \psi_{nk}^0(r) d^3r d^3r'. \quad (8)$$

Bechstedt et al. [5] have taken local fields into account (in the dynamical terms and in $\Sigma_{nk}^{\text{SEX}}$) by using a suitable state-dependent density in the calculation of $k_F$ and $q_{TF}$: given a state $nk$, they define the electron density $\rho(r)$ averaged on that state as

$$\rho_{nk} = \int d^3r (n(r)) |\psi_{nk}(r)|^2. \quad (9)$$

This value is inserted in place of the density in the screened interaction (2), mimicking the local density $\rho(r)$, with the advantage that the screened interaction is still a function of $r - r'$ only; after going to Fourier space, this allows to consider only the terms $G = G'$ in the calculation of the screened exchange, with a considerable reduction of computer time. This approximation works well in the case of small and moderate gap semiconductors, as Si, GaAs, AlAs and ZnSe [5].

The quasiparticle shifts obtained using this
Table 2. Self-energy corrections for top valence and bottom conduction bands at high symmetry points of the Brillouin zone for cubic gallium nitride and silicon. The first column gives the results of the full GW calculations, while the second and third columns show the results obtained with the two different approximated GW methods I and II, discussed in the text. We used $\alpha = 1$ and $\epsilon_0 = 6.14$ for GaN, $\alpha = 1$ and $\epsilon_0 = 11.3$ for silicon.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta E_{nk}^{GW}(F)$</th>
<th>$\Delta E_{nk}^{GW}(I)$</th>
<th>$\Delta E_{nk}^{GW}(II)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Gamma_5$</td>
<td>-0.17</td>
<td>0.41</td>
<td>0.18</td>
</tr>
<tr>
<td>$\Gamma_1$</td>
<td>0.71</td>
<td>0.62</td>
<td>0.91</td>
</tr>
<tr>
<td>$I_3^u$</td>
<td>-0.29</td>
<td>-0.49</td>
<td>-0.26</td>
</tr>
<tr>
<td>$L_1^c$</td>
<td>0.82</td>
<td>0.95</td>
<td>1.06</td>
</tr>
<tr>
<td>$X_3^u$</td>
<td>-0.46</td>
<td>-0.61</td>
<td>-0.40</td>
</tr>
<tr>
<td>$X_1^c$</td>
<td>0.60</td>
<td>1.16</td>
<td>0.91</td>
</tr>
<tr>
<td>$\Gamma_5^u - \Gamma_1^c$</td>
<td>2.76</td>
<td>2.91</td>
<td>2.98</td>
</tr>
<tr>
<td>$\Gamma_5^u - L_1^c$</td>
<td>5.66</td>
<td>6.03</td>
<td>5.91</td>
</tr>
<tr>
<td>$\Gamma_5^u - X_1^c$</td>
<td>3.98</td>
<td>4.78</td>
<td>4.30</td>
</tr>
<tr>
<td>Si</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Gamma_5^u$</td>
<td>0.03</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>$\Gamma_1^c$</td>
<td>0.66</td>
<td>0.99</td>
<td>0.96</td>
</tr>
<tr>
<td>$X_3^u$</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.07</td>
</tr>
<tr>
<td>$X_1^c$</td>
<td>0.67</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td>$\Gamma_5^u - \Gamma_1^c$</td>
<td>3.21</td>
<td>3.42</td>
<td>3.44</td>
</tr>
<tr>
<td>$\Gamma_5^u - X_1^c$</td>
<td>1.30</td>
<td>1.44</td>
<td>1.45</td>
</tr>
</tbody>
</table>

method (method I) are compared with their full GW counterparts in Table 1. It is clear that some of the gaps of GaN are by no means well described by this model. Analyzing the various contributions to the self energy and comparing them with their full-calculation counterparts, it turns out that the largest error is made in the calculation of the screened exchange.

In order to improve the method, we produce a modified version of it (method II), where the statically screened exchange is calculated using the form of the dielectric matrix given in equation (3). This slows down the calculation by about 10 times, with respect to the previous method, but substantially improves the results, especially for GaN as it is apparent in Table 2. In the case of Si, methods I and II yield very similar results in good agreement with full GW calculations (see Table 2).

5. CONCLUSIONS

We have evaluated models of screening and an approximate GW method in the case of Si, taken as the prototype of covalent small-gap semiconductors, and of GaN, representative of ionic wide-gap semiconductors. We have shown that the model of diagonal screening given in equation (1) works well in both cases. The same conclusion is reached in the case of Hybertsen and Louie's LDA-like treatment of local-field effects [13]. The model of off-diagonal screening introduced by Bechstedt et al. [7] on the basis of the generalized $f$-sum rule and ansatz (5) is instead restricted to covalent semiconductors. The simplified GW approach of Bechstedt et al. [5], which reduces the CPU time by a factor 100 with respect to full GW calculations, is very good in the case of Si and similar semiconductors, but yields errors of up to 0.8 eV for some gaps of GaN. It can be improved, by taking into account local-field effects in the statically-screened exchange within Hybertsen and Louie's approach [13], reaching an accuracy of 0.2–0.3 eV, at the expense however of increasing the computation time. In the latter version, this method reduces the CPU time by (only) a factor of 10 with respect to full GW calculations. We believe that this method (according to either versions, depending on the importance of local-field effects) can now be safely applied to bulk semiconductors with more complex structures than those of Si or GaN, as, for instance, SiC polytypes.

Acknowledgements – This work has been supported in part by the European Community programme Human Capital and Mobility through contract No. CHRX-CT93-0337.

REFERENCES


