Convergence of quasiparticle band structures of Si and Ge nanowires in the GW approximation and the validity of scissor shifts

H. Peelaers, B. Partoens, M. Giantomassi, T. Rangel, E. Goossens, G.-M. Rignanese, X. Gonze, and F. M. Peeters

1 Universiteit Antwerpen, Departement Fysica, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium
2 Université Catholique de Louvain, Institute of Condensed Matter and Nanosciences, Place Croix du Sud 1 bte 3, B-1348 Louvain-la Neuve, Belgium

(Received 8 September 2010; revised manuscript received 15 December 2010; published 21 January 2011)

Starting from fully converged density-functional theory calculations, the quasiparticle corrections are calculated for different sized Si and Ge nanowires using the GW approximation. The effectiveness of recently developed techniques in speeding up the convergence of the quasiparticle calculations is demonstrated. The complete quasiparticle band structures are also obtained using an interpolation technique based on maximally localized Wannier functions. From the quasiparticle results, we assess the correctness of the commonly applied scissor-shift correction. Dispersion changes are observed, which are also reflected in changes in the effective band masses calculated taking into account quasiparticle corrections.

DOI: 10.1103/PhysRevB.83.045306
PACS number(s): 73.21.Hb, 71.15.Qe, 71.45.Gm

I. INTRODUCTION

Nanowires are one-dimensional structures that have shown a promising prospective to be used as building blocks for future applications. Wires consisting of the semiconductor species Si and Ge are particularly interesting due to their compatibility with current semiconductor industry. Prototype applications such as field-effect transistors, p-n diodes, sensors for chemical and biological substances, and solar cells have already been realized. These nanowires can be grown using standard vapor-liquid-solid growth methods. The growth of thin wires has also been reported. For Si, these are mainly oriented along the [110] direction, while for Ge both the [111] and [110] directions are reported. In this work we will focus on the [110] direction.

Motivated by the possibilities of these nanowires, several theoretical ab initio studies have been already performed in order to investigate the wire properties, such as the electronic structure of bare and passivated nanowires, the properties of doped nanowires, and the phonon-related properties. A thorough overview of the performed theoretical studies can be found in the review paper by Rurali.

The knowledge of how the size of the band gap depends on the specific structure is very important, as this determines in part the properties of the wires. The standard ab initio density-functional theory (DFT) calculations using approximate functionals to model the exchange-correlation energy, such as the local-density approximation (LDA) or the generalized gradient approximation (GGA), tends to underestimate the size of this band gap by as much as 50%. This is a well-known shortcoming of these functionals, which do not properly take into account the many-body effects due to electron-electron interactions. In fact, there is no formal justification that the one-particle DFT eigenenergies can be used as quasiparticle (QP) energies and their differences cannot be interpreted as photoemission gaps or optical excitation energies. In contrast, many-body perturbation theory (MBPT) provides a formal ground for evaluating the experimentally observed QP band structures. Here, we use Hedin’s GW approximation for the electron self-energy. We compute the QP corrections from first-order perturbation theory \( G_0 W_0 \) applied on top of DFT-LDA results, which is currently a method of choice for computing QP band structures in solids. Due to the importance of these QP corrections, this method has already been applied to study Si and Ge nanowires (NWs) and Si chains. A recent overview of the QP corrections and the excitonic effects, including some of the technical aspects, is given by Palummo et al. In general, the QP corrections to the band structure are calculated first in order to be able to study the optical properties, including excitonic effects. The MBPT was also used to include interface-induced correlation effects in the calculation of transport properties in SiNW metal-oxide-semiconductor field-effect transistors. It is often assumed that the QP corrections cause a uniform shift of all the conduction bands, so that a global constant energy change (the so-called scissor shift) is applied to these bands using the QP corrections obtained only in one \( k \) point. In this paper, we (i) calculate the QP corrections for Si- and Ge- [110] oriented nanowires at the \( \Gamma \) point (as these nanowires have a direct band gap), (ii) calculate the complete QP band structures, (iii) assess the correctness of applying a scissor shift, and (iv) study the effective electron and hole masses, calculated both within standard DFT and including QP corrections.

The first part of this paper addresses the convergence issues related to the one-dimensional nature of the nanowires, as illustrated by the test case of a 0.5 nm Ge nanowire (consisting of six Ge and eight passivating H atoms). The main parameters that need to be converged are the number of bands (including a large amount of empty bands), the number of \( k \) points along the periodic direction, and the size of the unit cell in the directions perpendicular to the wire, i.e., the amount of vacuum separating the wire from its periodically repeated images. We aim at a global accuracy of 0.01 eV on the band energies, with a maximal tolerance of 0.002 eV for the convergence with respect to the number of \( k \) points, the number of bands, and the cell size separately. Several techniques recently introduced to solve the aforementioned convergence problems arising in GW calculations are discussed and applied to the test system.

In the second part, the calculated QP corrections are used to obtain the complete band structures employing an interpolation technique based on maximally localized Wannier functions (MLWFs). The resulting band structures are compared with
the corresponding DFT-LDA results and the reliability of a
scissor-shift operator is discussed. Finally, the electron and
hole masses are calculated and compared both at the DFT-LDA
and MBPT levels. Our results are obtained with the ABINIT
code, using a plane-wave basis set and norm-conserving
 pseudopotentials.

II. SPEEDING UP THE CONVERGENCE

The first parameter leading to convergence problems is the
number of \( k \) points employed for sampling the Brillouin zone.
As can be seen from Fig. 1 (black curve), the value of the
QP corrected band gap for the 0.5 nm Ge nanowire increases
when increasing the number of \( k \) points in the irreducible
part of the Brillouin zone. There is no sign of convergence (even
with 100 \( k \) points the curve still shows a linear slope).
This lack of convergence is due to the presence of a long-range
Coulomb interaction between the periodically repeated images
of the wire. Indeed, in contrast with DFT-LDA calculations
in which the wire is neutral (and thus, the convergence is
not problematic), \( GW \) calculations involve the addition of an
electron (or a hole) to the wire. The resulting charge is not
screened completely by the passivating hydrogen layer, hence,
there remains a long-range Coulombian interaction between
neighboring cells. In fact, the size of the unit cell that would
be required to render the spurious electrostatic interaction
negligible would be extremely large and the analysis could
not be handled with the present computing resources.

A possible approach for overcoming this problem consists
in introducing a truncation in real space of the Coulomb
interaction beyond a certain radius, so that the number of
interacting replicas is considerably reduced. The modified
expression equals the bare Coulomb term in the region of
interaction of the wire. Indeed, in contrast with DFT-LDA calculations
in which the wire is neutral (and thus, the convergence is
not problematic), \( W_{kk'}(q) = v_{kk'}(q + \mathbf{G}) \),
\[ v_{kk'}(r) = \frac{\theta(x,y)}{|r|}, \]
where the truncated Coulomb interaction for a wire oriented
along the \( z \) axis is given by
in which \( \theta(x,y) \) is one when \( x \) and \( y \) are inside the Wigner-
Seitz unit cell, and 0 otherwise (the interested reader is referred
to Ref. 36 for a more complete discussion of this cutoff
 technique). The interaction in real space has infinite extent
along the \( z \) axis and its Fourier transform is given by
\[ v_{kk'}(k) = \int dx dy \theta(x,y) 2K_0(|k|\rho) \cos(k_xx + k_yy), \]
where \( \rho = (x^2 + y^2)^{1/2} \) and \( K_0(z) \) is the modified Bessel
function. Since the integral is of finite extent, the only diverging
term in the limit \( k_z \to 0 \) originates from \( K_0 \) and behaves as
\[ -2 \ln(|k_z|) \int dx dy \theta(x,y) \cos(k_xx + k_yy). \]
When \( k_x^2 + k_y^2 \neq 0 \), the divergence vanishes, as the projection
of \( k \) in the \( xy \) plane is a reciprocal lattice vector. The
expression is singular only if the \( \Gamma \) point is approached along
the periodic dimension. This singularity is, however, milder
than the one presented by the bare Coulomb term and can be
accurately and efficiently treated via standard quadrature
 techniques. It is worth noting that the cutoff radius depends
only on the geometrical setup of the supercell, hence no
additional convergence parameters are needed.

The effect of the Ismail-Beigi cutoff technique on the
convergence can be clearly seen in Fig. 1 (red curve). A
detailed view of the convergence graph in the case of the
Ismail-Beigi approach is shown in the inset of Fig. 1. Here,
we also fitted the values of the band gap as a function of the
number of \( k \) points used, with the form
\[ E_{\text{gap}}^{GW}(N_{\text{gpt}}) = E_{\text{gap}}^{GW}(\infty) + \frac{A}{N_{\text{gpt}}}, \]
where \( N_{\text{gpt}} \) is the number of irreducible \( k \) points, \( E_{\text{gap}}(\infty) \) is
the value of the band gap when an infinite number of \( k \) points
is taken into account, and \( A \) is a fitting parameter. We have not
tried other fitting functional forms.

For what concerns the convergence with respect to the
number of bands, the extrapolate method\(^{18}\) allows one to
decrease (by around ten times) the number of unoccupied
bands that have to be used to converge the \( GW \) calculations.

Reducing the number of empty states is of fundamental
importance since it leads to an important decrease both in
CPU time and memory requirements. Strictly speaking, a
well-converged \( GW \) calculation should take into account a
number of states similar to the Hilbert space dimensionality

FIG. 1. (Color online) Convergence of the QP band gap \( E_{\text{gap}}^{GW} \)
with respect to the number of \( k \) points with (in red) and without
(in black) Ismail-Beigi cutoff for the 0.5 nm Ge nanowire. Inset:
Zoom of the convergence with Ismail-Beigi cutoff. The squares and
circles represent the calculated values, while the lines show the best
fit with Eq. (5).
that equals the total number of basis functions. In a plane-wave-based approach, the number of basis functions is very large, thus rendering $GW$ calculations extremely CPU demanding. To improve the convergence, the eigenenergies of states not explicitly treated are replaced by a common energy, determined from the highest computed state and the single parameter $\Delta E$ (which is the energy added to the energy of the highest treated band). The contribution of the states that are not explicitly considered in the calculation are taken into account by means of the closure relation

$$\sum_{i>N_b} |i\rangle \langle i| = 1 - \sum_{i\leq N_b} |i\rangle \langle i|,$$

where $N_b$ is the number of treated bands. The value of the parameter $\Delta E$ might be adjusted to provide the fastest convergence. It is worth stressing, however, that for the calculation of the polarizability, it is possible to obtain an optimal value for $\Delta$ by monitoring the fulfillment of a particular sum rule (see Ref. 38 for details).

The convergence of the $GW$ results with respect to the number of states at fixed $\Delta E$ is shown in Fig. 2(a). In this figure, the same number of bands and the same value of $\Delta E$ is used both for the screening and the subsequent self-energy calculation. When the extrapolar technique is used, we find that the band gap is converged with only 180 bands. Such a level of convergence is not obtained when the extrapolar technique is not used. For a small number of bands, it seems that the calculation using the extrapolar technique is converging toward another value for the band gap, compared to the calculations without using the technique. However, when the number of bands is increased further, both approaches converge toward the same value. The reduction from 2000 bands to 180 bands influences both memory requirements and CPU time substantially.

In the present case, the best value to take for $\Delta E$ is 1 Ha, leading to the fastest convergence, with the smallest slope in the convergence graphs. From Fig. 2, it is also clear that the convergence of the band gap is faster than the convergence of the individual parts (band energies of the highest valence $E_{hvb}$ and lowest conduction band $E_{hcb}$), which is an effect of the cancellation of errors and which confirms that energy differences converge faster than the individual components.

This technique is not limited to one-dimensional (or more generally, low dimensional) systems, but it can also be used in bulk systems.

Finally, the convergence with respect to the interwire distance and thus of the lateral size of the unit cell, is performed. Here, we use the previously discussed extrapolar technique with $\Delta E = 1$ Ha. This convergence for the value of the band gap, with (red dots) and without (black squares) the Ismail-Beigi cutoff, is shown in Fig. 3. Without cutoff, a huge unit cell would be required to converge the band gap because of the Coulomb interaction. In contrast, convergence is already achieved for a lateral unit cell size of 26 bohrs with the Ismail-Beigi cutoff. Please note that, in the inset, the scale of the $y$ axis is such that the differences between two successive labels is 0.001 eV, which is smaller than the desired accuracy.

With this combination of numerical techniques, we were able to compute efficiently the $QP$ band gap of 0.5 and 1.2 nm Si and Ge nanowires, and of the 1.6 nm Si nanowire. The obtained DFT-LDA values, the $QP$ results, and the size of

FIG. 2. (Color online) Convergence of (a) the band gap $E_{GW}^g$, (b) the highest valence band $E_{hvb}^GW$, and (c) the lowest conduction band $E_{lcb}^GW$ with respect to the number of bands in the calculation for the 0.5 nm Ge nanowire. The convergence is shown for different values of the parameter $\Delta E$ (see text).

FIG. 3. (Color online) Convergence of the band gap $E_{GW}^g$ with respect to the size of the unit cell in the directions perpendicular to the wire with (in red) and without (in black) Ismail-Beigi cutoff for the 0.5 nm Ge nanowire. Inset: Zoom of the convergence with Ismail-Beigi cutoff.
the QP corrections are shown in Table I. These corrections are larger for the 0.5 nm nanowire than for the 1.2 and 1.6 nm nanowires. These values are comparable to the values previously reported in literature, which are also reported in Table I. Experimental results for these wire sizes are not available, as the experimental wires have larger diameters.

### III. BAND STRUCTURES AND EFFECTIVE MASSES

The ability to compute the electronic energies for arbitrary wave vectors in the Brillouin zone is a natural prerequisite for obtaining the full band structure. Within the DFT formalism, the eigenenergies at an arbitrary \( k \) point are easily obtained by solving a non-self-consistent problem in which only the self-consistent density is needed. A similar approach, on the other hand, cannot be used within the GW formalism, Evaluating the QP corrections at an arbitrary wave vector \( k \), indeed requires the knowledge of the Kohn-Sham eigenenergies and wave functions on a homogeneous grid of points containing the wave vector of interest. This leads to very time-consuming self-energy calculations, which cannot be afforded, especially when a fine sampling along high symmetry lines is wanted. A solution to this practical problem is to use an efficient and accurate interpolation based on maximally localized Wannier functions (MLWFs)\(^{30,40}\) In this scheme the corrections to any arbitrary \( k \) point are obtained from a reduced \( k \)-point set at which the QP eigenenergies were calculated explicitly.

In this work, the MLWFs are determined using the wannier90 code,\(^{41}\) which only requires the overlap matrices and an initial guess of the projection of the Bloch states onto trial localized orbitals. Through a minimization procedure, a transformation matrix is computed which can be used to express the Hamiltonian of the system in the basis of the MLWFs. This representation is the starting point for a Slater-Koster interpolation\(^{32}\) scheme that is employed to compute the Hamiltonian, and therefore its eigenvalues and eigenstates, on a finer \( k \)-point mesh.

Here, we first search the smallest initial \( k \)-point mesh needed to obtain good band structures at the DFT-LDA level. For this mesh, the QP corrections are subsequently calculated. These are then used to interpolate the band structures based on the MLWFs. The latter are the same as in DFT-LDA, since the \( G_0 W_0 \) approximation uses the DFT-LDA wave functions and only corrects the eigenenergies.

Figure 4 shows the DFT-LDA band structure of the 0.5 nm Ge nanowire interpolated using MLWFs. The latter are obtained using a disentanglement procedure\(^{40}\) setting an inner window from the bottom of the lowest valence band to a value of 5 eV and an outer window to 6 eV. In Fig. 4(a) the MLWFs were obtained using an initial (unshifted) \( 1 \times 5 \times 5 \) \( k \)-point grid in the full Brillouin zone, corresponding to three symmetry inequivalent \( k \) points. The \( k \)-point grid is unshifted so that it contains the \( \Gamma \) point needed to calculate the band gap in the GW calculations. The eigenenergies at these three points are indicated by red dots. It is clear that the interpolation is wrong in this case. In Fig. 4(b), a finer \( k \)-point grid was used, containing six inequivalent \( k \) points corresponding to an unshifted \( 1 \times 1 \times 10 \) grid. The agreement with the eigenenergies obtained directly from non-self-consistent calculations (blue dots) is excellent, as can be observed from Fig. 4(c).

If the QP corrected eigenenergies are available for the same \( k \)-point mesh, the corresponding band structure can also be obtained by interpolation, as shown in Fig. 5(a) (black lines). The obtained QP band structures can be used to study the validity of the commonly applied scissor shift. The latter aims to correct the DFT-LDA band structures by rigidly shifting all conduction bands in such a way that the band gap reaches the calculated GW value (or the experimental one if available). Such an approach clearly neglects all curvature differences between the DFT-LDA and QP band structures. As we now have the full QP band structure, we can test the validity of this approach. This is done in Fig. 5(a), where the red lines are the interpolated scissor-shifted DFT-LDA band structures. As can be seen from the zoom around the region close to the band gap shown in Fig. 5(b), this approach is only valid

---

**TABLE I.** The band gap (in eV) for different wires as calculated within DFT-LDA (\( E_{\text{gap}}^{\text{LDA}} \)) and MBPT (\( E_{\text{gap}}^{\text{LDA}} \)). The QP corrections (\( \Delta E_{\text{gap}}^{\text{GW}} = E_{\text{gap}}^{\text{GW}} - E_{\text{gap}}^{\text{LDA}} \)) are also given. The last column shows previously reported values for the QP band gap.

<table>
<thead>
<tr>
<th>Wire diameter (nm)</th>
<th>( E_{\text{gap}}^{\text{LDA}} )</th>
<th>( \Delta E_{\text{gap}}^{\text{GW}} )</th>
<th>( E_{\text{gap}}^{\text{GW}} )</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ge 0.5</td>
<td>2.76</td>
<td>2.11</td>
<td>4.87</td>
<td>4.5(^a)</td>
</tr>
<tr>
<td>Ge 1.2</td>
<td>1.57</td>
<td>1.41</td>
<td>2.98</td>
<td>3.01(^b)</td>
</tr>
<tr>
<td>Si 0.5</td>
<td>3.20</td>
<td>2.35</td>
<td>5.55</td>
<td>5(^c)</td>
</tr>
<tr>
<td>Si 1.2</td>
<td>1.70</td>
<td>1.62</td>
<td>3.32</td>
<td>3.12,3.22,3.4(^d)</td>
</tr>
<tr>
<td>Si 1.6</td>
<td>1.14</td>
<td>1.18</td>
<td>2.31</td>
<td>2.2,2.32,2.33(^e)</td>
</tr>
</tbody>
</table>

\(^a\)Reference 22.  \(^b\)Reference 29.  \(^c\)Reference 23.  \(^d\)Reference 24.  \(^e\)Reference 25.
for $k$ points close to the $\Gamma$ point and for bands close to the highest valence and lowest conduction band. One also observes that the gap between the highest valence and lowest conduction band in other $k$ points is underestimated in spite of the applied scissor shift: the highest valence band is too high in energy, while the lowest conductance band is too low in energy. The QP corrections are clearly not uniform, in contrast with what is implicitly assumed in the scissor-shift approach. To check if the curvatures near the $\Gamma$ point, i.e., the hole and electron effective masses, are different in the QP corrected case as compared to the DFT-LDA case, these were calculated by fitting a quadratic curve for small $k$ points.

From these calculations, shown in Table II, one can observe that the curvature is not always the same, e.g., the curvature of the highest valence band (thus corresponding to the hole effective mass) for the Ge nanowires or for the 1.6 nm diameter nanowires. Panels (b) and (c) show a zoom of the region around the band gap. This is visible in Fig. 5(c), where a detailed zoom of the band structure near the band gap is shown for the 1.2 nm Si nanowire. The curvature of the highest valence band (thus corresponding to the hole effective mass) is different for the interpolated QP (black) and scissor-shifted DFT-LDA (red) band structures. The values for the electron effective masses in the DFT-LDA and QP corrected cases are very similar, indicating that for these thin nanowires the lowest conduction band is shifted rigidly close to the $\Gamma$ point. The fact that the corrections to the band masses are larger for the holes than for the electrons has also been observed in bulk systems.43

From these calculations, shown in Table II, one can observe that the gap between the highest valence and lowest conduction band. One also observes that the gap between the highest valence and lowest conduction band in other $k$ points is underestimated in spite of the applied scissor shift: the highest valence band is too high in energy, while the lowest conductance band is too low in energy. The QP corrections are clearly not uniform, in contrast with what is implicitly assumed in the scissor-shift approach. To check if the curvatures near the $\Gamma$ point, i.e., the hole and electron effective masses, are different in the QP corrected case as compared to the DFT-LDA case, these were calculated by fitting a quadratic curve for small $k$ points.

From these calculations, shown in Table II, one can observe that the curvature is not always the same, e.g., the curvature of the highest valence band (thus corresponding to the hole effective mass) for the Ge nanowires or for the 1.6 nm Si nanowire differs significantly. This is visible in Fig. 5(c), where a detailed zoom of the band structure near the band gap is shown for the 1.2 nm Si nanowire. The curvature of the highest valence band (thus corresponding to the hole effective mass) is different for the interpolated QP (black) and scissor-shifted DFT-LDA (red) band structures. The values for the electron effective masses in the DFT-LDA and QP corrected cases are very similar, indicating that for these thin nanowires the lowest conduction band is shifted rigidly close to the $\Gamma$ point. The fact that the corrections to the band masses are larger for the holes than for the electrons has also been observed in bulk systems.43

TABLE II. The calculated effective hole and electron masses (expressed in units of the electron mass) for Si and Ge nanowires, calculated within DFT-LDA and using QP corrections.

<table>
<thead>
<tr>
<th>Diameter (nm)</th>
<th>$m_e^{\text{LDA}}$</th>
<th>$m_h^{\text{LDA}}$</th>
<th>$m_e^{\text{QP}}$</th>
<th>$m_h^{\text{QP}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.10</td>
<td>-0.12</td>
<td>0.11</td>
<td>-0.29</td>
</tr>
<tr>
<td>1.2</td>
<td>0.12</td>
<td>-0.33</td>
<td>0.09</td>
<td>-0.18</td>
</tr>
<tr>
<td>Si</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.27</td>
<td>-0.16</td>
<td>0.22</td>
<td>-0.16</td>
</tr>
<tr>
<td>1.2</td>
<td>0.13</td>
<td>-0.54</td>
<td>0.13</td>
<td>-0.19</td>
</tr>
</tbody>
</table>

IV. SUMMARY

In this paper, we have presented a systematic convergence study of the $GW$ calculations for one-dimensional nanowires. It was shown that it is necessary to use a cutoff of the Coulomb interactions to achieve convergence. The application of the extrapolar technique leads to an important reduction in the number of empty states by a factor of 10. The rule of thumb to take an amount of empty bands equal to ten times the number of occupied bands is also valid in the case of nanowires. By using an interpolation scheme based on maximally localized Wannier functions, we found that accurate band structures can be obtained. From these, it was demonstrated that applying a scissor shift to the DFT-LDA band structure leads to substantial differences with respect to the full one-shot $GW$ results, when one compares the band structures at $k$ points away from the $\Gamma$ point or for bands with lower (higher) energies than the highest valence (lowest conduction) band. This can lead to noticeable differences in the optical spectra. Significant differences can also be observed close to $\Gamma$, depending on the diameter and composition of the wire, as is evident from the calculated effective masses.

ACKNOWLEDGMENTS

We are grateful to Yann Pouillon for valuable technical support with the build system of ABINIT, related to the WANNIER90 library. This work was supported by the Flemish Science Foundation (FWO-V1) and by the Interuniversity Attraction Poles Program (P6/42)—Belgian State—Belgian Science Policy. X.G. and G.-M.R. acknowledge funding from the EU’s 7th Framework Programme through the ETSF I3 e-Infrastructure project (Grant No. 211956), the Communauté française de Belgique through the Action de Recherche Concertée 07/12-003 “Nanosystèmes hybrides métal-organiques,” and the Wallon Region Project No. 816849 “European Theoretical Spectroscopy Facility” (WALL-ETSF). M.G. acknowledges funding from the FRFC Project No. 2.4.589.09.F.

1D. Hartwin.Peelaers@ua.ac.be
2I.Bart.Partoens@ua.ac.be
3Francois.Peeters@ua.ac.be


R. Rurali, Rev. Mod. Phys. 82, 427 (2010).


