

Accurate Quasiparticle Spectra from Self-Consistent GW Calculations with Vertex Corrections

M. Shishkin, M. Marsman, and G. Kresse

Faculty of Physics, Universität Wien and Center for Computational Materials Science, Sensengasse 8/12, A-1090 Wien, Austria

(Received 21 June 2007; published 12 December 2007)

Self-consistent GW calculations, maintaining only the quasiparticle part of the Green's function G , are reported for a wide class of materials, including small gap semiconductors and large gap insulators. We show that the inclusion of the attractive electron-hole interaction via an effective nonlocal exchange correlation kernel is required to obtain accurate band gaps in the framework of self-consistent GW calculations. If these are accounted for via vertex corrections in W , the band gaps are found to be within a few percent of the experimental values.

DOI: [10.1103/PhysRevLett.99.246403](https://doi.org/10.1103/PhysRevLett.99.246403)

PACS numbers: 71.10.-w, 71.20.Nr

The accurate prediction of band gaps is a long-standing challenge to computational materials science. This concerns particularly extended systems, such as bulk semiconductors and insulators, for which the widely successful density functional theory (DFT) method generally yields much too small band gaps, and in specific cases, even the wrong band order [1]. The most common approach to deal with this problem is the GW approximation [2–4], which is nowadays usually applied perturbatively on top of a computationally less demanding scheme relying typically on DFT wave functions and DFT one-electron energies. The corresponding approximation is referred to as G_0W_0 [3–5]. Although this approximation was originally vastly successful, recent calculations clearly demonstrate that part of the success was related to a fortuitous cancellation of errors between systematic shortcomings introduced by the G_0W_0 approximation itself and technical errors incurred by the pseudopotential approximation, on which many G_0W_0 calculations were based. This understanding originates mainly from accurate all-electron calculations relying on full potential methods that allow a consistent and accurate treatment of core and valence electrons on the same footing [6–10]. As shown recently the G_0W_0 approximation yields particularly large errors for systems with shallow d states such as GaAs, ZnO, and CdS [10,11]. The problem can be partly remedied by iterating the one-electron energies in the Green's function G until self-consistency is reached, but even this approach underestimates band gaps in the aforementioned materials by up to 15% (e.g. Ref. [10]). This has prompted Faleev, van Schilfgaarde, and Kotani to suggest a self-consistent GW method—subsequently called self-consistent quasiparticle GW (sc GW)—in which the GW Hamiltonian is approximated by a self-consistently determined Hermitian Hamiltonian [12]. In this way, the usual problems encountered in self-consistent GW calculations, such as loss of intensity from the quasiparticle peaks to satellites, are avoided. Although the method yields systematically improved band structures, the fundamental gaps are now overestimated with errors being largest for small gap ma-

terials like GaAs [13]. Van Schilfgaarde *et al.* pointed out that this might be related to the neglect of the attractive interaction between electrons and holes, which is responsible for the excitonic features in the adsorption spectra [14]. Unfortunately, including these effects, which is usually done via vertex corrections in the GW approximation, is not an easy matter. Here a recent development by Reining and co-workers comes in, who suggested to recast this computationally demanding term into an effective nonlocal exchange correlation kernel $f_{xc}(\mathbf{r}, \mathbf{r}')$ [15–18]. We demonstrate that the inclusion of these many-body correlation effects allows for the prediction of band gaps with very good precision.

Here we present sc GW calculations including a parameter-free f_{xc} for small gap and large gap materials, comprising typical semiconductors (Ge, Si, SiC, GaAs, GaN, ZnO, ZnS, CdS, and AlP), insulators (C, BN, MgO, LiF), and noble gas solids (Ar, Ne). The calculations are performed using the projector-augmented-wave method and the VASP code. The first step in our calculations consists of a hybrid Hartree-Fock density functional calculation using the HSE03 functional [19]. This functional mixes 25% of the nonlocal Hartree-Fock exchange to an otherwise semilocal functional and predicts reasonable band gaps for semiconductors, although the gaps remain seriously underestimated for ZnO and rare gas solids [20]. The final results are independent of the choice of the initial functional, but a good starting point allows us to cut down on the number of iterations required to achieve self-consistency in the subsequent GW calculations.

The sc GW calculations are performed using a variant of the method originally suggested by van Schilfgaarde *et al.* [12,14]. In order to determine the quasiparticle peaks in the GW method one needs to determine all solutions $n = 1, \dots, N_{\text{states}}$ (filled and unoccupied bands) of a nonlinear one-electron-like Schrödinger equation (the Brillouin zone index has been dropped for brevity):

$$[T + V + \Sigma(E_n)]|\psi_n\rangle = E_n|\psi_n\rangle, \quad (1)$$

where T is the kinetic energy operator, the local potential V

corresponds to the electrostatic potential of the nuclei and electrons, and the self-energy $\Sigma(\mathbf{r}, \mathbf{r}', E)$ is energy dependent and nonlocal. In principle, the solutions can be determined by iteration, and we proceed by a linearization around some reference energy E_n^i for state n and iteration i :

$$\left(T + V + \Sigma(E_n^i) + \frac{\partial \Sigma(E_n^i)}{\partial E_n^i} (E - E_n^i) \right) |\psi_n\rangle = E |\psi_n\rangle.$$

With the abbreviation $\xi(E_n^i) = \frac{\partial \Sigma(E_n^i)}{\partial E_n^i}$, the equation can be recast as a generalized and non-Hermitian (but linear) eigenvalue problem:

$$\underbrace{[T + V + \Sigma(E_n^i) - \xi(E_n^i)E_n^i]}_{\mathbf{H}(E_n^i)} |\psi_n\rangle = \underbrace{E[1 - \xi(E_n^i)]}_{\mathbf{S}(E_n^i)} |\psi_n\rangle. \quad (2)$$

The fact that the quasiparticle peaks are not orthonormalized is reflected by the presence of the overlap operator \mathbf{S} : as a consequence of \mathbf{S} , the quasiparticle peaks usually contain only 75% of the norm, whereas the rest of the intensity is found in satellites (lacking after linearization). Unfortunately, no unique method to map this problem onto a corresponding Hermitian eigenvalue problem exists, but one possible route is to determine the Hamiltonian and overlap operator in a suitable basis set $\{\phi_n^i | n = 1, \dots, N_{\text{states}}\}$ (e.g., DFT wave functions), take the Hermitian part of the Hamiltonian and overlap matrix in this basis $\bar{H}_{mn} = \text{Herm}[\langle \phi_m^i | \mathbf{H}(E_n^i) | \phi_n^i \rangle]$, $\bar{S}_{mn} = \text{Herm}[\langle \phi_m^i | \mathbf{S}(E_n^i) | \phi_n^i \rangle]$, and solve the corresponding simple Hermitian eigenvalue problem

$$\bar{S}^{-1/2} \bar{H} \bar{S}^{-1/2} \bar{U} = \bar{U} \Lambda, \quad (3)$$

where \bar{U} is a unitary matrix and Λ the diagonal eigenvalue matrix. The wave functions in the next iteration are then given by $\phi_n^{i+1} = \sum_m \bar{U}_{mn} \phi_m^i$ and the eigenvalues E_n^{i+1} by the corresponding eigenvalues Λ_{nn} . This approach yields band gaps within 1% of those obtained by calculating the right-hand eigenvectors of the generalized non-Hermitian problem [Eq. (2)] and subsequently reorthogonalizing the eigenstates. Furthermore, the method of Ref. [14] is recovered by setting the off-diagonal elements in \bar{S} to zero, which in practice also yields almost identical results. These observations indicate that the eigenfunctions of the original non-Hermitian problem (2) are almost orthogonal and that the off-diagonal elements in \bar{S} are negligible. The present method can be viewed as a generalization of the usual update equations in the diagonal GW method, where nondiagonal terms in \bar{H} and \bar{S} are neglected and the basis set thus remains identical to the DFT wave functions [5,8,10].

The second important issue is the inclusion of the electron-hole interaction in the dielectric matrices. These matrices are required in order to determine the screened Coulomb kernel $W = \varepsilon^{-1}v$, which enters in the previously used self-energy $\Sigma = iGW$. In time-dependent density functional theory, the full polarizability χ is given by the

Dyson-like equation

$$\chi = [1 - \chi_0(v + f_{xc})]^{-1} \chi_0, \quad (4)$$

and the frequency dependent test-charge test-charge dielectric matrix is defined by $\varepsilon^{-1} = 1 + v\chi$. The matrix χ_0 is the independent particle polarizability and v the Coulomb kernel $1/|\mathbf{r} - \mathbf{r}'|$. We note that χ_0 , χ , and ε are frequency dependent nonlocal matrices that are expanded in a plane wave basis set including all plane waves with kinetic energies of up to 200 eV in the present calculations (for details see Ref. [8]). For local density functionals, f_{xc} is simply the second variation of the exchange correlation energy functional with respect to the density:

$$f_{xc}(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \frac{\delta^2 E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r}) \delta n(\mathbf{r})}. \quad (5)$$

Because of the locality of the exchange correlation term, it lacks any long range interactions and inclusion or omission of f_{xc} in Eq. (4) changes the static screening properties only little in the long wave length limit. In a proper quasiparticle description, as applied in the present framework, the inclusion of a consistent exchange correlation kernel, however, modifies the dielectric properties significantly, as we will show below. To determine a consistent many-body $f_{xc}(\mathbf{r}, \mathbf{r}', \omega)$ in the GW framework, we follow the recent suggestions of Reining and Del Sole *et al.* [15–18]. Here f_{xc} is calculated only at zero frequency, since tests for the adsorption spectrum show that this is an excellent approximation; e.g., the exact time-dependent adsorption spectra can be recovered using a frequency independent $f_{xc}(\mathbf{r}, \mathbf{r}')$ (see for instance Ref. [17]). In the present case we have included the direct Coulomb-like electron-hole interaction, as well as the exchangelike contributions coupling the resonant and antiresonant part of the response. The present sc GW calculations are performed including f_{xc} right from the first iteration. Alongside the wave functions and eigenvalues [see Eq. (3)], f_{xc} is updated until self-consistency is reached. The kernel f_{xc} is calculated initially for the HSE03 functional and then using the static, screened Coulomb kernel $W(\omega \rightarrow 0)$, as commonly done in Bethe-Salpeter calculations [21]. In the tables below we will report calculations without f_{xc} (“RPA”—random phase approximation) and including f_{xc} (labeled “ $e-h$ ” for electron-hole).

Figure 1 reports the band gaps for a conventional semi-local DFT functional (here and in the following the acronym DFT is used for calculations using the Perdew Burke Ernzerhof [22] functional) and for the sc GW with and without f_{xc} in W . The vertex corrections are included via Eq. (4) only, but neglected in the self-energy (see below). All materials were considered in their zero temperature ground state structure, except for ZnO and BN, which were investigated in the zinc-blend structure. The Brillouin zone was sampled using $8 \times 8 \times 8$ points, although results are usually already converged using $6 \times 6 \times 6$ points (for test cases $10 \times 10 \times 10$ points were used). The total number of

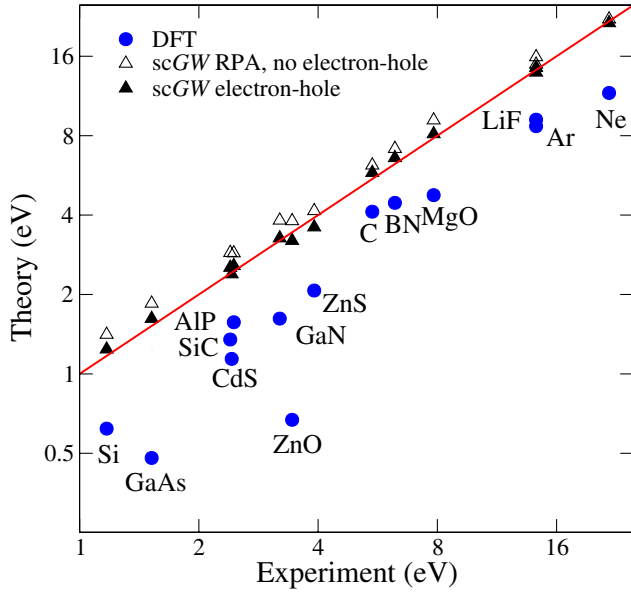


FIG. 1 (color online). DFT and scGW band gaps with and without attractive electron-hole interaction (vertex corrections in W).

valence and conduction band states was set to 150 for all materials. We estimate the technical accuracy of the calculations to be about 2% [10]. Finally, it is noted that the experimental temperature dependence is often significant, which test calculations indicate to be not related to thermal expansion, but more likely electron-phonon couplings not accounted for in the present framework. One thus has to compare with gaps measured at low temperature wherever possible (underlined experimental values in Table I).

As emphasized before, the scGW calculations overestimate the band gaps, with fractional errors being largest for GaAs and Ge, as already noticed by Chantis *et al.* [13]. To a large extent the inclusion of vertex corrections remedies this problem, with the mean absolute relative error becoming smaller than 4%. On average the band gaps are still slightly too large, even including the electron-hole interaction, with the notable exceptions of ZnS and ZnO, where the gaps are underestimated by about 6%. We furthermore observe that the d band position is roughly 1 eV too shallow for all materials with semicore $3d$ states (this includes GaAs and GaN; see exemplary results for ZnO and GaAs in Table I) and that self-consistency in the wave functions has only a small effect on the d band position. This might be the result of neglecting vertex corrections in the calculation of the self-energy (see Ref. [23]), and we will return to this point later. We furthermore believe that the incorrect $3d$ position is the main reason for too small band gaps in ZnS and ZnO.

The ion-clamped macroscopic dielectric constants ϵ_m ,

$$\epsilon_m^{-1} = \lim_{\omega \rightarrow 0, \mathbf{q} \rightarrow 0} \epsilon^{-1}(\mathbf{q}, \mathbf{q}, \omega),$$

are reported in Tables I and II. Agreement with experiment

TABLE I. Band gaps Δ and averaged d band positions ϵ_d for scGW calculations without (RPA) and with ($e-h$) attractive electron-hole interaction (vertex corrections in W), and for GW_0^{DFT} calculations, where W_0^{DFT} is calculated from DFT wave functions and eigenvalues, and wave functions and eigenvalues in G are updated until self-consistency is reached (values in parentheses are results for an update of the eigenvalues only in G from Ref. [10]). References for experimental values are collected in Ref. [10]; underlined values correspond to measurements at low temperature. Theoretical values are corrected for spin-orbit coupling (0.10 eV for GaAs and Ge). Also reported are the theoretical (scGW) and experimental ion-clamped (high frequency) dielectric constants ϵ_m .

	Δ (eV)			EXP	ϵ_m	
	scGW RPA	scGW $e-h$	GW_0^{DFT} RPA		scGW $e-h$	EXP
Ge	0.95	0.81	0.75	<u>0.74</u>	15.3	16.2
Si	1.41	1.24	1.28 (1.20)	<u>1.17</u>	11.4	11.9
GaAs	1.85	1.62	1.55 (1.42)	<u>1.52</u>	10.4	11.1
SiC	2.88	2.53	2.62 (2.43)	2.40	6.48	6.52
CdS	2.87	2.39	2.37 (2.28)	2.42	5.31	5.30
AIP	2.90	2.57	2.57 (2.59)	2.45	7.11	7.54
GaN	3.82	3.27	3.30 (3.00)	3.20	5.35	5.30
ZnO	3.8	3.2	3.0 (2.5)	<u>3.44</u>	3.78	3.74
ZnS	4.15	3.60	3.59 (3.50)	<u>3.91</u>	5.15	5.13
C	6.18	5.79	5.88 (5.68)	5.48	5.59	5.70
BN	7.14	6.59	6.66 (6.35)	\approx 6.25	4.43	4.50
MgO	9.16	8.12	8.25 (7.64)	7.83	2.96	3.00
LiF	15.9	14.5	14.8 (14.0)	14.20	1.98	1.90
Ar	14.9	13.9	14.0 (13.9)	14.20	1.69	
Ne	22.1	21.4	21.1 (20.5)	21.70	1.23	
	ϵ_d (eV)					
ZnO		6.7	6.4 (6.4)	7.5–8		
GaAs		17.6	17.3 (17.2)	18.9		

is again very good at the present level of theory. The present results allow for an assessment of why the commonly used random phase approximation for the calculation of W was so successful in combination with DFT wave functions (W_0^{DFT}). In DFT, the band gaps are way too small, but if f_{xc} is neglected in Eq. (4), the resultant dielectric constants agree well with the experimental values and the present high level theory (see Table II). If the wave functions and eigenvalues are updated in G (but not W) until self-consistency is reached, almost the same band gaps as in the scGW calculations are obtained. This indicates that W_0^{DFT} is very similar to the self-consistently determined W with vertex corrections. Only for ZnO and GaAs the final band gaps are smaller, which is clearly related to the strong overestimation of the static dielectric constant in DFT-RPA (see Table II). The success of the GW_0^{DFT} approximation has to result from a cancellation of errors—too small DFT band gaps and neglect of electron-hole interactions—but this cancellation seems to be rather universal for gradient corrected DFT functionals. On the other hand, the neglect of the attractive electron-hole term (f_{xc}) in the scGW

TABLE II. Theoretical and experimental ion-clamped (high frequency) dielectric constants ϵ_m . Values with 10% deviation from experiment are underlined. Projector-augmented-wave–DFT values in the independent particle approximation are presented in Ref. [24] alongside other all-electron values.

	scGW RPA	scGW <i>e-h</i>	DFT RPA	EXP
GaAs	<u>8.2</u>	10.4	<u>12.8</u>	11.1
Si	<u>9.2</u>	11.4	12.0	11.9
SiC	<u>5.22</u>	6.48	6.54	6.52
C	<u>5.00</u>	5.58	5.55	5.70
ZnO	<u>2.84</u>	3.78	<u>5.12</u>	3.74
MgO	<u>2.30</u>	2.96	2.99	3.00

calculations yields unreliable dielectric constants (scGW—RPA in Table II). The underestimation is typically 20%, which is the origin for the overestimated band gaps (5–10%) in the scGW approximation [14].

From a theoretical point of view, the present calculations leave one open question. We have included the vertex corrections via the electron-hole interaction f_{xc} in W only, but we have neglected them in the construction of the self-energy ($\Sigma = iGW\Gamma$), since the latter approach turned out to be numerically rather unstable and tended to bring the band gaps back to those obtained without vertex corrections (see also Ref. [18]). The approximation applied here is often termed $GW^{\text{TC-TC}}$, as opposed to the more concise $GW\Gamma$ approximation. Reining has recently shown that the present form for the nonlocal exchange correlation kernel f_{xc} is not optimal for approximating the vertices in the self-energy, but better kernels are not yet available [25]. Based on the good agreement of the present values with experiment, we believe the effect of appropriate vertex corrections in the self-energy should be rather small, and these corrections should mostly affect localized d states.

In summary, we have demonstrated that the inclusion of approximate many-electron vertex corrections in W yields excellent band gaps in the framework of the self-consistent GW method. This allows for a consistent and entirely parameterfree calculation of band gaps using the GW method with a precision approaching experimental methods. The approach opens a wide variety of applications, for instance band gap engineering in hypothetical compounds and prescreening of dopants, and it might be applicable to molecular systems as well. Since fully self-consistent GW

calculations with vertex corrections are exceedingly time-consuming (in average a single calculation takes 1 day on four Pentium Duo processors), the other result of the present study is equally important. The static screening properties calculated from gradient corrected functionals in the random phase approximation agree very well with the self-consistently determined screening properties in the scGW method. This offers a convenient shortcut and legitimates the often applied GW_0 approximation.

This work was supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung.

-
- [1] J. Furthmüller *et al.*, Phys. Rev. B **72**, 205106 (2005).
 - [2] L. Hedin, Phys. Rev. **139**, A796 (1965).
 - [3] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
 - [4] W. G. Aulbur, L. Jonsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (2000).
 - [5] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
 - [6] T. Kotani and M. van Schilfhaarde, Solid State Commun. **121**, 461 (2002).
 - [7] S. Lebegue *et al.*, Phys. Rev. B **67**, 155208 (2003).
 - [8] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
 - [9] C. Friedrich *et al.*, Phys. Rev. B **74**, 045104 (2006).
 - [10] M. Shishkin and G. Kresse, Phys. Rev. B **75**, 235102 (2007).
 - [11] P. Rinke *et al.*, New J. Phys. **7**, 126 (2005).
 - [12] S. V. Faleev, M. van Schilfhaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
 - [13] A. N. Chantis, M. van Schilfhaarde, and T. Kotani, Phys. Rev. Lett. **96**, 086405 (2006).
 - [14] M. van Schilfhaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
 - [15] L. Reining *et al.*, Phys. Rev. Lett. **88**, 066404 (2002).
 - [16] G. Adragna, R. Del Sole, and A. Marini, Phys. Rev. B **68**, 165108 (2003).
 - [17] F. Sottile, V. Olevano, and L. Reining, Phys. Rev. Lett. **91**, 056402 (2003).
 - [18] F. Bruneval *et al.*, Phys. Rev. Lett. **94**, 186402 (2005).
 - [19] J. Paier *et al.*, J. Chem. Phys. **124**, 154709 (2006).
 - [20] F. Fuchs *et al.*, Phys. Rev. B **76**, 115109 (2007).
 - [21] S. Albrecht *et al.*, Phys. Rev. Lett. **80**, 4510 (1998).
 - [22] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
 - [23] A. Fleszar and W. Hanke, Phys. Rev. B **71**, 045207 (2005).
 - [24] M. Gajdoš *et al.*, Phys. Rev. B **73**, 045112 (2006).
 - [25] L. Reining (private communication).