

Self-consistent GW calculations for semiconductors and insulators

M. Shishkin* and G. Kresse

Institut für Materialphysik and Centre for Computational Materials Science, Universität Wien, A 1090 Wien, Austria

(Received 22 November 2006; revised manuscript received 5 April 2007; published 4 June 2007)

We present GW calculations for small and large gap systems comprising typical semiconductors (Si, SiC, GaAs, GaN, ZnO, ZnS, CdS, and AlP), small gap semiconductors (PbS, PbSe, and PbTe), insulators (C, BN, MgO, and LiF), and noble gas solids (Ar and Ne). It is shown that the G_0W_0 approximation always yields too small band gaps. To improve agreement with experiment, the eigenvalues in the Green's function G (GW_0) and in the Green's function and the dielectric matrix (GW) are updated until self-consistency is reached. The first approximation leads to excellent agreement with experiment, whereas an update of the eigenvalues in G and W gives too large band gaps for virtually all materials. From a pragmatic point of view, the GW_0 approximation thus seems to be an accurate and still reasonably fast method for predicting quasiparticle energies in simple sp -bonded systems. We furthermore observe that the band gaps in materials with shallow d states (GaAs, GaN, and ZnO) are systematically underestimated. We propose that an inaccurate description of the static dielectric properties of these materials is responsible for the underestimation of the band gaps in GW_0 , which is itself a result of the incomplete cancellation of the Hartree self-energy within the d shell by local or gradient corrected density functionals.

DOI: [10.1103/PhysRevB.75.235102](https://doi.org/10.1103/PhysRevB.75.235102)

PACS number(s): 71.10.-w, 71.15.Mb, 71.20.Nr

I. INTRODUCTION

The GW approximation¹ is a widely used method to predict quasiparticle band gaps, as opposed to density-functional theory (DFT), which is only applicable to ground-state properties.^{2,3} The large computational effort associated with this method limits GW calculations to rather small systems, and various approximations, such as the plasmon-pole model⁴ or model GW ,⁵ have been used to make GW calculations more tractable. The most common approximation is the non-self-consistent evaluation of the quasiparticle self-energy on top of some computationally less demanding scheme, usually the local-density approximation (LDA) or generalized gradient approximation (GGA) to density-functional theory.⁴ Although the calculated values are considerably improved compared to LDA/GGA results, single shot GW (usually called G_0W_0) has a tendency toward too small gaps compared to experiment.⁶⁻⁹ For many materials (e.g., Si or diamond), the discrepancy of the LDA based G_0W_0 gaps with experiment is relatively small, although for others, it may be quite significant: e.g., various reports on GW calculations for ZnO and ZnS yield gaps that are either overestimated^{10,11} or, more commonly, seriously underestimated for LDA based G_0W_0 calculations (e.g., typically 1–2 eV).¹²⁻¹⁵ It is believed that calculations beyond the usual single shot LDA+ G_0W_0 approximation will remedy most of these problems.

However, GW calculations with a full update of the Green's function G and screened potential W carried out for the free-electron gas¹⁶ and more recently for metals (K) and semiconductors (Si) (Ref. 17) show a significant overestimation of the bandwidth and band gaps, equal in absolute magnitude to the underestimation in LDA calculations. Such poor performance is usually explained by a shift of intensity from quasiparticle (QP) peaks into satellites, with a concomitant reduction of the screening. Although in a subsequent paper claims were made that the poor accuracy of the self-

consistent GW for Si can be cured by a more accurate treatment of the semicore electrons,¹⁸ it remains a controversial issue whether fully self-consistent GW calculations without vertex corrections can yield accurate band gaps (see also Refs. 19 and 20).

Technically, fully self-consistent GW calculations are exceedingly demanding. To obtain a more tractable approach and to avoid that intensity is moved from the QP peaks into satellites, calculations are often performed by updating the eigenvalues in the one-electron Green's function and/or the dielectric function but keeping the one-electron wave functions identical to the LDA/GGA solutions. This method was already suggested in the pioneering work of Hybertsen and Louie⁴ and has been pursued by several groups. Most of these calculations rely on the plasmon-pole model^{21,22} often used in combination with model functions for the static dielectric matrix.²³⁻²⁷ Overall, partially self-consistent GW calculations yield larger gaps than the single shot G_0W_0 method for a wide range of materials, at least within the applied approximations (model dielectric function and plasmon pole). Other calculations with a more explicit evaluation of the frequency-dependent dielectric matrix have also been carried out.^{15,28,29} They confirm the trend to increase the band gaps, although general conclusions on the absolute accuracy cannot be drawn, as only few materials have been considered and the applied approximations vary a lot. For instance, Louie and co-workers often updated only the eigenvalues in G and kept W fixed,^{21,23-27} whereas for II–V compounds, both G and W were updated.²² Fleszar and Hanke,¹⁵ also for the II–V compounds, obtained excellent results when updating the eigenvalues in G and W as well. Given the numerous approximations used in the past (model dielectric functions, plasmon-pole approximation, pseudopotential approximation, and approximate treatment of core-valence interaction), there is a need for a careful assessment of the effect of partial self-consistency, ideally pursued using a full-potential method. This is the main goal of the present work.

Admittedly, justifying the applied approximations is not an entirely easy matter, except by observing, *a posteriori*, good agreement with experiment. The most profound shortcoming of the applied method is that it still entirely depends on the wave functions supplied as input to the *GW* calculations, and although LDA/GGA wave functions are most likely reasonable for many semiconductors, more difficult materials—such as transition-metal oxides or rare-earth oxides—are not well described using LDA/GGA. We therefore avoid calculations on systems where LDA/GGA is believed to give unreliable wave functions but note that Faleev *et al.* have recently proposed a method to construct an approximate Hermitian Hamiltonian, which permits one to include the off-diagonal elements of Σ .³⁰ This method is particularly useful for materials such as NiO, MnO, or ceria, where the LDA/GGA wave functions are qualitatively wrong, but for simple *sp*-bonded systems, results are suggested to be rather close to the simpler diagonal approximation used in the present work (1%–2%). Systematic tests on the difference between LDA/GGA wave functions and wave functions by this method are discussed in Ref. 31 for Si and Ar. We finally note that this self-consistent quasiparticle *GW* technique (scQP*GW*) seems to overestimate band gaps typically by 10%–15%,^{9,32} and as we will show, this also holds for the much simpler diagonal approximation if the eigenvalues in *G* and *W* are updated.

To address the issue of self-consistency, we perform *GW* calculations using two different degrees of self-consistency on a wide range of materials in the framework of the full-potential projector augmented wave (PAW) method.⁸ Starting from standard G_0W_0 calculations, we perform GW_0 and *GW* calculations. In both cases, the wave functions are kept fixed (i.e., equal to GGA wave functions), whereas the eigenvalues are updated in the Green's function only for the case of GW_0 and in the dielectric matrix of the screened potential *W* and Green's function *G* in the case of *GW*. These methods do not pose significant computational burdens compared to G_0W_0 calculations and can be used routinely for the prediction of band gaps even for fairly large systems.⁸

II. COMPUTATIONAL METHOD

A. *GW* method

The quasiparticle energies in the *GW* approximation are calculated as follows (for details on the notation and implementation, we refer to Ref. 8):

$$E_{n\mathbf{k}} = \text{Re}[\langle \psi_{n\mathbf{k}} | T + V_{n-e} + V_H + \Sigma(E_{n\mathbf{k}}) | \psi_{n\mathbf{k}} \rangle], \quad (1)$$

where *T* is the kinetic-energy operator, V_{n-e} the potential of the nuclei, V_H the Hartree potential, and *n* and \mathbf{k} the band and *k*-point indices, respectively. The diagonal elements of the self-energy matrix are given by

$$\begin{aligned} \langle \psi_{n\mathbf{k}} | \Sigma(\omega) | \psi_{n\mathbf{k}} \rangle &= \frac{1}{\Omega} \sum_{\mathbf{q}\mathbf{G}\mathbf{G}'} \sum_{n'} \frac{i}{2\pi} \int_{-\infty}^{\infty} d\omega' W(\mathbf{G} + \mathbf{q}, \mathbf{G}' + \mathbf{q}, \omega') \langle \psi_{n\mathbf{k}} | e^{i(\mathbf{q}+\mathbf{G})\mathbf{r}} | \psi_{n'\mathbf{k}-\mathbf{q}} \rangle \\ &\quad \times \langle \psi_{n'\mathbf{k}-\mathbf{q}} | e^{-i(\mathbf{q}+\mathbf{G}')\mathbf{r}'} | \psi_{n\mathbf{k}} \rangle \frac{1}{\omega + \omega' - E_{n'\mathbf{k}-\mathbf{q}} + i\eta \text{sgn}[\epsilon_{n'\mathbf{k}-\mathbf{q}} - \mu]}. \end{aligned} \quad (2)$$

In Eq. (2), *W* is the dynamically screened potential calculated in the common random-phase approximation, and μ is the Fermi energy. In our present calculations, we set the nondiagonal elements of the self-energy matrix to zero, which is justified for materials where GGA wave functions are close to the *GW* ones. The updated quasiparticle energy ($E_{n\mathbf{k}}^{N+1}$) is then obtained from the quasiparticle energy at the previous iteration ($E_{n\mathbf{k}}^N$) by linearization of Eq. (1) as follows:⁴

$$E_{n\mathbf{k}}^{N+1} = E_{n\mathbf{k}}^N + Z_{n\mathbf{k}} \text{Re}[\langle \psi_{n\mathbf{k}} | T + V_{n-e} + V_H + \Sigma(E_{n\mathbf{k}}^N) | \psi_{n\mathbf{k}} \rangle - E_{n\mathbf{k}}^N], \quad (3)$$

where $Z_{n\mathbf{k}}$ is the renormalization factor,

$$Z_{n\mathbf{k}} = \left(1 - \text{Re} \langle \psi_{n\mathbf{k}} | \frac{\partial}{\partial \omega} \Sigma(\omega) \Big|_{E_{n\mathbf{k}}^N} | \psi_{n\mathbf{k}} \rangle \right)^{-1}. \quad (4)$$

Initially, we perform single shot G_0W_0 calculations using Eq. (3) and setting $E_{n\mathbf{k}}^1 = E_{n\mathbf{k}}^{GGA}$ (i.e., we used GGA eigenvalues as initial input to the *GW* calculations). For the GGA calcula-

tions, the gradient corrected Perdew-Burke-Ernzerhof (PBE) functional is applied.³³ Then, the single electron energies $E_{n'\mathbf{k}-\mathbf{q}}$ are updated in the denominator of Eq. (2) using the values obtained after the first iteration. *W* is kept fixed for the GW_0 case, whereas in the *GW* case, the dielectric matrix ϵ and $W(\mathbf{r}, \mathbf{r}', \omega) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}'', \omega) / (\mathbf{r}'' - \mathbf{r}')$ are re-evaluated in each iteration using the new QP energies. The iterative procedure is carried out until self-consistency is reached. In our calculations, four updates proved to yield well converged results, so in the following, we provide results for the fourth iteration. In the G_0W_0 and GW_0 calculations, we used an $8 \times 8 \times 8$ *k*-point grid centered at the Γ point, whereas the *GW* calculations were performed using $6 \times 6 \times 6$ *k* points. With respect to the *k*-point mesh, the results are converged to within roughly 10 meV, except for ZnO (discussed below). For simplicity, the total number of bands is always 150. Residual errors due to the number of unoccupied bands are expected to be of the order of 1% and 20–30 meV for materials with small band gaps. To demonstrate this, we show the QP convergence for the direct band gap in Si and ZnS

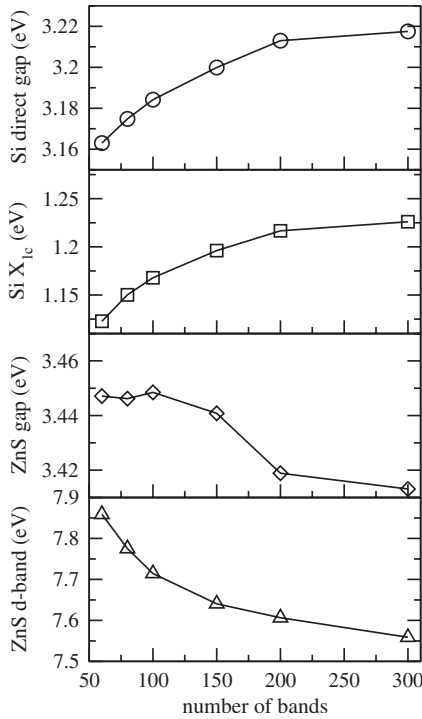


FIG. 1. G_0W_0 quasiparticle energies determined with respect to the valence-band maximum for Si and ZnS.

and the indirect band gap in Si at the X point in Fig. 1. For both materials, convergence to the previously specified threshold is reached around 150 bands. It is, however, clear that convergence for the d -band position is slower, and we estimate that they are converged to only roughly 100 meV, which we believe to be sufficient in view of the large errors for the d -band position even on the GW level (see below).

The GGA calculations are performed using a consistent GGA treatment of valence and core electrons, but in the GW calculation, the core-valence interaction is re-evaluated on the Hartree-Fock level. We have shown in our previous work that this allows us to keep most of the electrons in the core.⁸ For materials that contain semicore d electrons (ZnO, ZnS, GaAs, GaN, CdS, and PbX), the quasiparticle energies have been calculated by treating these d electrons as valence. Treating the p states with the same main quantum number as valence states as well had little influence on the gaps as discussed in Sec. II B and in Ref. 8. For the GW_0 and GW calculations, the energies of all occupied and eight unoccupied states were updated, whereas the remaining eigenvalues were kept fixed at the initial GGA energies. Updating 32 unoccupied states (for a $6 \times 6 \times 6$ k -point grid) yielded values within 10 meV, justifying the previous approximation. Finally, we note that the PAW potentials have been constructed with considerable care in order to guarantee an accurate description of the scattering properties up to 100 eV above the vacuum level.⁸ Overall, we expect that our calculations are thus converged to around 2% or 30 meV for small gap materials, with the dominant error stemming from the finite number of unoccupied bands.

The values of the calculated gaps are corrected for spin-orbit (SO) coupling effects. This is crucial for PbS, PbSe,

and PbTe, as for these materials, the corrections due to spin-orbit coupling are rather large (Table I). Spin-orbit coupling is significantly smaller for GaAs (only 0.1 eV) and negligibly small for other materials. Since our present GW code does not allow us to include spin-orbit coupling fully consistently, we have determined the reduction of the gap using self-consistent Hartree-Fock hybrid density-functional calculations with 25% exact exchange.³⁴ These hybrid functionals already yield band gaps close to experiment, and from comparison to full-potential calculations and experiment, we conclude that the calculated SO corrections are accurate to within 20 meV.⁴⁶

The calculations were performed at $T=0$ K lattice constants, if available, or at the room-temperature lattice constants otherwise. The lattice constants are compiled in Table I, with underlined values corresponding to low-temperature values. In general, we found that, with the exception of the lead chalcogenides, the GW results depend very little on the employed lattice constants, with typical changes between $T=0$ K and $T=300$ K lattice constants being only 1%–2%. The dependence of the measured gaps on temperature is often much larger, pointing toward strong anharmonic electron-phonon couplings, and we have generally tried to compare with gaps measured at $T=0$ K wherever available (experimental low-temperature gaps are again underlined). If not available, we compare with the band gaps measured at room temperature. The calculations were performed for the diamond structure for C and Si, the NaCl structure for the lead chalcogenides, MgO, and LiF, the zinc-blende structure for other semiconductors, and the simple fcc structure for rare-gas solids.

B. Potential details

In the present implementation, the GW Hamiltonian is only evaluated on the plane-wave grid, whereas the one-center terms are calculated using the Hartree-Fock approximation. This allows for a fairly simple implementation, and usually the results are quite insensitive to the chosen core radius for the construction of the PAW potentials. Also, results change little upon moving electrons from the core to the valence: in pseudopotential GW codes, it is, for instance, often found that, for ZnO, ZnS, GaAs, and GaN, the $3s$ and $3p$ shells must be treated as valence electrons in order to obtain reasonable values for the position of the $3d$ states and the band gap. Approximating the core-valence interaction by the Hartree-Fock Hamiltonian usually allows us to keep these $3s$ and $3p$ states in the core, as discussed in our previous work.⁸ Upon careful convergence studies with respect to the PAW core radius r_c , we found that it is essential that the shape of the pseudized $3d$ wave functions follows closely the shape of the all-electron $3d$ wave functions, since the exchange interaction within the $3d$ shell is partially screened by the other valence electrons. In the present work, all potentials involving $3d$ electrons have been carefully optimized to observe these criteria. We note that this was not the case in Ref. 8, placing the $3d$ states at too deep energies for GaAs (see Table III in Ref. 8). The present calculations are more precise in this respect, although we note that band gaps for GaAs are hardly changed (typically 30 meV).

TABLE I. Results of DFT-PBE and quasiparticle (G_0W_0 , GW_0 , and GW) calculations. An $8 \times 8 \times 8$ k -point mesh is used for all calculations except for the GW case (see text). Experimental values for gaps (Expt.), lattice constants (a), and the calculated values for spin-orbit coupling (SO) are also provided. Underlined values correspond to zero-temperature values. The mean absolute relative error (MARE) and the mean relative error (MRE) are also reported; lead chalcogenides are excluded in the MARE and MRE.

	PBE	G_0W_0	GW_0	GW	Expt.	a	SO
PbSe	-0.17	0.10	0.15	0.19	<u>0.15</u> ^a	<u>6.098</u> ^b	0.40
PbTe	-0.05	0.20	0.24	0.26	<u>0.19</u> ^c	<u>6.428</u> ^b	0.73
PbS	-0.06	0.28	0.35	0.39	<u>0.29</u> ^d	<u>5.909</u> ^b	0.36
Si	0.62	1.12	1.20	1.28	<u>1.17</u> ^e	<u>5.430</u> ^f	
GaAs	0.49	1.30	1.42	1.52	<u>1.52</u> ^e	<u>5.648</u> ^f	0.10
SiC	1.35	2.27	2.43	2.64	2.40 ^g	4.350 ^g	
CdS	1.14	2.06	2.26	2.55	2.42 ^h	5.832 ^h	0.02
AlP	1.57	2.44	2.59	2.77	2.45 ^h	5.451 ^h	
GaN	1.62	2.80	3.00	3.32	3.20 ⁱ	4.520 ⁱ	0.00
ZnO	0.67	2.12	2.54	3.20	<u>3.44</u> ^c	4.580 ^h	0.01
ZnS	2.07	3.29	3.54	3.86	<u>3.91</u> ^e	5.420 ^h	0.02
C	4.12	5.50	5.68	5.99	5.48 ^g	3.567 ^g	
BN	4.45	6.10	6.35	6.73	6.1–6.4 ^j	3.615 ^h	
MgO	4.76	7.25	7.72	8.47	7.83 ^k	4.213 ^l	
LiF	9.20	13.27	13.96	15.10	14.20 ^m	<u>4.010</u> ⁿ	
Ar	8.69	13.28	13.87	14.65	14.20 ^o	5.260 ^p	
Ne	11.61	19.59	20.45	21.44	21.70 ^o	4.430 ^p	
MARE	45%	9.9%	5.7%	6.1%			
MRE	45%	-9.8%	-3.6%	4.7%			

^aReference 35.

^bReference 46.

^cReference 36.

^dReference 37.

^eReference 38.

^fReference 47.

^gReference 39.

^hReference 40.

ⁱReference 41.

^jReference 42.

^kReference 43.

^lReference 48.

^mReference 44.

ⁿReference 47.

^oReference 45.

^pReference 49.

III. RESULTS

The calculated G_0W_0 and GW_0 quasiparticle energies for a wide range of materials are illustrated in Fig. 2 and presented in Table I. Note that the lead chalcogenides are excluded

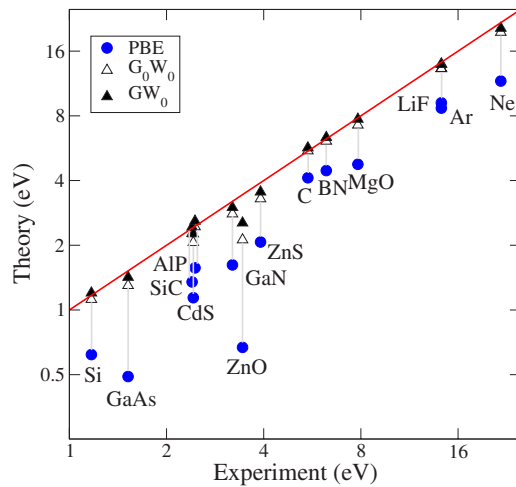


FIG. 2. (Color online) PBE and quasiparticle energies G_0W_0 and GW_0 . Logarithmic scale is used for both axes.

from Fig. 2 to allow a better presentation of the other materials. Furthermore, in Table II, the columns $r_{G_0W_0}$ and r_{GW_0} show the relative error of the predicted band gaps with respect to experiment. These relative errors are defined as $r_{G_0W_0} = (E_{G_0W_0} - E_{exp})/E_{exp}$ and $r_{GW_0} = (E_{GW_0} - E_{exp})/E_{exp}$, where $E_{G_0W_0}$, E_{GW_0} , and E_{exp} are the G_0W_0 , GW_0 , and experimental gaps, respectively.

We first note that these calculations are very well converged with respect to the number of k points. In fact, if a less accurate $6 \times 6 \times 6$ k -point mesh is used, the eigenvalues for all materials, except ZnO, change by ± 20 meV for all calculations (PBE, G_0W_0 , and GW_0). The band gap of ZnO, however, converges exceedingly slowly, and even the $8 \times 8 \times 8$ k -point grid results in errors of roughly 0.1 eV, as indicated by G_0W_0 calculations using $12 \times 12 \times 12$ k points. We have thus corrected all reported values for ZnO by this estimated convergence error.

As already amply demonstrated in the literature, the G_0W_0 band gaps are significantly larger than the GGA ones (Fig. 2). However, with the single exception of C, the G_0W_0 calculations still yield consistently underestimated values (see also Table II). As we will discuss below, the slight overestimation for C is most likely related to the random-phase approximation (RPA) yielding a too weak screening. The

TABLE II. Differences between calculated band gaps (G_0W_0 and GW_0) and calculated dielectric constants (in RPA) and experimental results, normalized by experimental values ($r_{G_0W_0}$, r_{GW_0} , and r_ϵ , respectively). For BN, two values of $r_{G_0W_0}$ and r_{GW_0} are provided for minimum and maximum experimental gaps (Table I). The values of the dielectric constants, calculated within the RPA and the experimental values, used to obtain r_ϵ are given for reference.

	$r_{G_0W_0}$	r_{GW_0}	ϵ_{RPA}	ϵ_{expt}	r_ϵ
Si	-0.04	0.03	12.09	11.90 ^a	-0.02
GaAs	-0.15	-0.07	12.84	11.10 ^a	-0.16
SiC	-0.05	0.01	6.56	6.52 ^a	-0.01
CdS	-0.15	-0.07	5.80	5.30 ^b	-0.09
AlP	-0.00	0.06	7.53	7.54 ^a	0.00
GaN	-0.13	-0.06	5.68	5.30 ^c	-0.07
ZnO	-0.38	-0.26	5.12	3.74 ^d	-0.37
ZnS	-0.16	-0.09	5.62	5.13 ^a	-0.10
C	0.00	0.04	5.54	5.70 ^a	0.03
BN	0.0/-0.05	0.04/-0.01	4.30	4.50 ^e	0.04
MgO	-0.07	-0.01	2.99	3.00 ^b	0.00
LiF	-0.07	-0.02	1.96	1.90 ^b	-0.03
Ar	-0.07	-0.02	1.66		
Ne	-0.10	-0.06	1.25		

^aReference 39.

^bReference 50.

^cReference 51.

^dReference 52.

^eReference 53.

other extreme cases are ZnO, ZnS, GaAs, GaN, and CdS, where the G_0W_0 gaps are more than 10% smaller than the experimental ones. For the cases of GaAs and ZnO, this might be explained by the rather small PBE gaps, i.e., less than half of the experimental value, necessitating a treatment beyond first-order perturbation theory.

Partially self-consistent GW_0 calculations rectify this deficiency by further increasing the gaps (Fig. 2). The band gaps are enlarged by a roughly constant 7% shift across most of the materials. There are two important exceptions to this rule, namely C, where GW_0 increases the band gap only by an additional 4%, and ZnO, where the band gap increases by almost 16%. This is satisfactory, since the already overestimated G_0W_0 band gap of C becomes hardly worse in GW_0 , whereas for ZnO, the agreement with experiment is improved. For the three materials Si, AlP, and the already mentioned C, the band gaps are now slightly too large, but for the other materials, the band gaps remain underestimated, with fractional errors being largest for ZnO, ZnS, GaAs, and CdS (Table II).

To analyze the origin for this rather large scatter of the predicted quasiparticle band gaps around the experimental values, we have calculated the static dielectric constant in the random-phase approximation, i.e., on the same level of theory as applied in the construction of W_0 . Although we compare only the head of the dielectric matrix with experiment and not the entire matrix, we believe that this comparison provides a general idea on the adequacy of the entire static dielectric matrix. Table II reports the relative difference between the experimental and the theoretically predicted di-

electric constants, defined as $r_\epsilon = (\epsilon_{\text{exp}} - \epsilon_{RPA}) / \epsilon_{\text{exp}}$, and Fig. 3 graphically shows the relative error versus the error in the calculated dielectric constants.

Figure 3 clearly shows that a linear correlation between r_{GW_0} and r_ϵ exists, demonstrating that an accurate prediction of band gaps requires an accurate prediction of the static screening properties of the considered material. If the RPA overestimates the screening, the band gaps tend to be too small (GaAs, ZnO, and ZnS), whereas if the screening is underestimated, the band gaps tend to be too large (C).

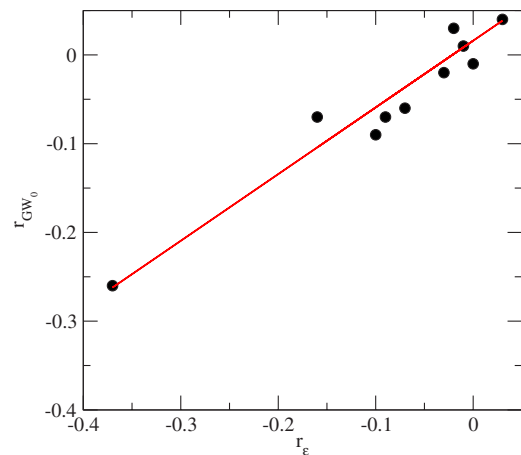


FIG. 3. (Color online) Band-gap error r_{GW_0} versus error in the calculated dielectric constants r_ϵ . The line shows a linear fit $r_{GW_0} = 0.01 + 0.75r_\epsilon$ to all data points.

PbS, PbSe, and PbTe have not been included in Table II. Actually, we generally find that our present calculations without inclusion of SO coupling underestimate the static dielectric constant in these materials, with errors sometimes even exceeding 20%. However, if spin-orbit coupling is included on the DFT level, the band order in these materials is no longer properly predicted as demonstrated elsewhere,⁴⁶ and the static dielectric constants are now too large. Also, GW calculations including a full treatment of SO coupling are presently not possible using our code. Altogether, we tend to believe that the agreement on the GW_0 level is partly fortuitous, but it is gratifying that the GW_0 approximation yields the correct band order $E_{\text{PbS}} > E_{\text{PbTe}} > E_{\text{PbSe}}$. For details, we refer to Ref. 46.

Although the GW_0 calculations widen the gap as compared to G_0W_0 , they still have a tendency to yield underestimated values. Therefore, we investigated how well self-consistent GW calculations perform. Again, we have limited the present investigations to an update of the eigenvalues only, but now in the Green's functions G and the screened potential W , neglecting nondiagonal elements in the GW Hamiltonian. The gaps are indeed larger than the ones we obtained using GW_0 , but with few exceptions (ZnO, ZnS, and Ar), they are now too large. This observation agrees with the results published by Kotani and co-workers,^{9,32} although we note that an inclusion of the nondiagonal elements of the GW Hamiltonian has the tendency to open the gaps even further. In particular, for GaAs, Chantis *et al.*³² reported a larger gap of 1.80 eV using the scQPGW. Using a similar method as suggested in Refs. 9 and 32, we found that the GaAs band gap indeed increases to 1.79 eV using our code (including SO corrections), implying that the inclusion of nondiagonal elements will worsen the agreement with experiment even further. The “best” results—in the sense that they are closest to experiment—are thus obtained on the level of the GW_0 approximation, with the self-consistent quasiparticle GW calculations yielding generally too large gaps. The reason for the failure of the self-consistent GW calculations lies in an underestimation of the static dielectric constants: although a precise determination of the static dielectric constants using the self-consistent GW approximation is beyond the scope of this work and would require much denser k -point grids, we find in exemplary studies for Si and SiC that the static dielectric constants in the GW -RPA become 9.9 and 5.5, respectively, i.e., significantly underestimating the experimental values. We note that the corresponding data points for Si and SiC are lying almost exactly on the straight line shown in Fig. 3. Thus, we conclude that the self-consistent GW approximation generally worsens the agreement with experiment, because it underestimates the screening if it is used in combination with the random-phase approximation.

A further important aspect is the position of the d bands. It is emphasized again that in the present calculations, not all orbitals with the main quantum number equal to the main quantum number of the semicore d shell are included in the valence. Contrary to conventional pseudopotential calculations, the core-valence interaction is evaluated on the Hartree-Fock level in the present PAW implementation (Sec. II B), which allows us to keep these orbitals in the core with-

TABLE III. The calculated positions and experimental values of the d bands for GaN, GaAs, ZnO, and ZnS. The results for calculations with the $3s$ and $3p$ electrons treated as valence are referred to as G_{3sp} and Zn_{3sp} .

	PBE	G_0W_0	GW_0	Expt.
GaAs	-14.8	-16.9	-18.0	-18.9 ^a
G_{3sp} As	-14.8	-16.9	-18.1	
GaN	-13.3	-15.8	-16.7	-17.0 ^b
G_{3sp} N	-13.3	-16.0	-16.9	
ZnO	-5.2	-6.1	-6.6	-(7.5–8.81) ^c
Zn_{3sp} O	-5.2	-6.2	-6.6	
ZnS	-6.1	-7.0	-7.5	-9.03 ^c
Zn_{3sp} S	-6.1	-7.0	-7.5	

^aReference 54.

^bReference 55.

^cReference 56.

out sacrificing the precision of the quasiparticle band gaps. Nevertheless, we have performed additional calculations with the $3s$ and $3p$ shells in the valence for GaAs, GaN, ZnO, and ZnS. Since these calculations are significantly more time consuming, only a $6 \times 6 \times 6$ k -point grid was applied, which is sufficient to converge the position of the d band. On passing, we note that unfreezing the $3s$ and $3p$ shells had a negligible impact on the band gaps, with values changing by typically 1% (using an identical k -point set). The calculated positions of the d bands are presented in Table III. They have been determined by calculating the mean value of the d band over the entire Brillouin zone. The results indicate that the inclusion of the $3s$ and $3p$ electrons hardly influences the position of the $3d$ states, and thus the Hartree-Fock approximation for the $3sp$ - $3d$ interaction seems to be fairly reliable.

In comparison to experiment, the PBE d -band positions are much too shallow, whereas the G_0W_0 and GW_0 calculations place these bands at deeper binding energies. Comparison to previously published values is generally very good. For instance, for GaN, the present d -band position is in good agreement with the G_0W_0 calculations by Rohlffing *et al.*, who obtained a value of -15.7 eV.¹⁴ The remaining difference is most likely related to their pseudopotential approximation. For ZnS, Fleszar and Hanke obtained a binding energy of 6.87 eV at the G_0W_0 level, in good agreement with the present work,¹⁵ and Miyake *et al.* reported a value of -6.95 eV, also in excellent agreement with the present value. Using an all-electron code, Kotani obtained $3d$ -band positions of -18.1 eV for GaAs, -16.4 eV for GaN, and -7.1 eV for ZnS, also in reasonable agreement with our results (deviations of typically 7%), in particular, in view of the rather slow band convergence of the $3d$ -band position with the number of unoccupied bands and the small number of bands often used in all-electron calculations (see Fig. 1).

Concerning comparison to experiment, one observes that the G_0W_0 and GW_0 approximations both yield too shallow d states. The error is approximately constant and amounts to roughly 1 eV on the GW_0 level across all considered materials. Most likely, this is related to an unsatisfactory descrip-

tion of the hybridization between valence states and the d states using the semilocal-density functionals applied in the present work and errors for the screening within the localized d shell. The random-phase approximation in combination with wave functions from local-density functionals is unreliable (see below). Fleszar and Hanke, for instance, found that LDA vertex corrections place the $3d$ shell of ZnS at deeper binding energies, bringing GW results in better agreement with experiment.¹⁵

Returning now to the predicted band gaps, we note that the errors are largest for materials with semicore d states. In combination with the unsatisfactory description of the $3d$ states, it seems likely that the origin for the error in the band gaps is related to the LDA/GGA error in the description of these $3d$ states. It is well known that the local-density approximation cancels the Coulomb self-interaction within the localized d shell not completely. Is that the main reason why the subsequent GW_0 calculations fail?

To test for this hypothesis, we have used a simple DFT+ U approach to shift the semicore d states to lower energies in the DFT calculations.^{57,58} Specifically, we have used the more accurate Ga_{3sp} and Zn_{3sp} potentials (compare Table III) and set $U-J$ to 8.0 eV for Ga and Zn. In the DFT+ U calculations, the d electrons are now located at -18.9 , -17.0 , -7.5 , and -8.4 eV for GaAs, GaN, ZnO, and ZnS, which are in good agreement with experiment (see Table III). The dielectric constants using DFT+ U are in better agreement with the experimental values (GaAs 12.01, GaN 5.44, ZnO 4.17, and ZnS 5.2), with the errors now being within the range of other simple sp -bonded materials. The subsequent GW_0 calculations also gave GW_0 band gaps in better agreement with experiment (GaAs 1.49, GaN 3.05, ZnO ≈ 2.70 , and ZnS 3.70), whereas with the exception of ZnO, the GW band gaps became too large as previously observed for all other materials. The positions of the d bands remain in good agreement with experiment for GaAs and GaN but are still much too shallow for ZnO and ZnS (GaAs -18.3 , GaN -17.2 , ZnO -6.6 , and ZnS -7.0). We conclude that although we do observe an improvement for the band gaps, the overall description is certainly still far from being satisfactory, in particular, for the position of the d bands in ZnO and ZnS. Recently, Miyake *et al.* performed similar studies for ZnS.⁵⁹ We note that their LDA+ U implementation is based on Wannier orbitals, as opposed to our one-center PAW implementation,⁵⁸ but similar to our case, they observed a slight increase of the band gap and no significant improvement of the d -band position.

On passing, we note that placing the d electrons in the core, and thus removing any s - d hybridization, also yields better agreement with experiment, as long as the interaction between the core and the valence is treated on the Hartree-Fock level (as done in our implementation). This approximation is, however, only applicable to systems where the d shell is located well below the valence band and thus restricted to GaAs and GaN.

IV. DISCUSSION

The effect of different degrees of self-consistency (G_0W_0 , GW_0 , and GW) has been investigated for a representative

selection of materials. Although it is long established that the G_0W_0 approximation yields band gaps in good agreement with experiment, it is not generally accepted that the G_0W_0 approximation based on LDA/GGA wave functions and one-electron energies gives underestimated gaps for practically all materials. Only a recent work, relying on the all-electron full-potential linear-muffin-tin-orbital method,^{6,9} has clearly established this trend. Given the large amount of pseudopotential calculations yielding excellent band gaps already on the G_0W_0 level, some doubt prevailed on whether this underestimation might be related to basis-set problems in the linear-muffin-tin-orbital method. Careful convergence studies for Si, however, showed that even converged basis sets (full-potential linearized-augmented plane waves) yield too small band gaps for Si.⁶⁰ Our recent study using the PAW method showed a similar underestimation for Si, GaAs, and CdS.⁸ Here, we have extended our studies to a representative number of insulators and semiconductors confirming the observation of van Schilfgaarde's group: generally, the predicted G_0W_0 band gaps are within a few percent of those reported using the linear-muffin-tin-orbital code, and, with the single exception of diamond, the band gaps are significantly underestimated in the G_0W_0 approximation.

To cure this failure, we have applied two well established strategies. In the first case, the eigenvalues in the Green's function G were updated, whereas the screening properties were calculated in the random-phase approximation using GGA wave functions and GGA eigenvalues (GW_0). The incentive for such an approximation is that gradient corrected functionals yield reliable static screening properties and thus possibly an accurate approximation for W . Indeed, using this approach, the predicted band gaps are in excellent agreement with experiment, with typical errors being around 3%–5%. Nevertheless, for materials with shallow d states, the discrepancies to experiment remain larger, and the band gaps are underestimated by typically 10%, with the largest deviations observed for ZnO (25%). We have shown that a quantitative relationship between the calculated static (ion-clamped) dielectric constant and the predicted GW_0 quasiparticle band gaps exists (see Fig. 3). If the dielectric constants, calculated within the random-phase approximation, are too large, quasiparticle band gaps tend to be too small, and vice versa, if the predicted dielectric constants are too small, the quasiparticle band gaps tend to be too large. This is certainly not astonishing, but due to the lack of systematic studies, the trend has not yet been established rigorously.

In the final approach, we have updated the eigenvalues in the Green's function and in the dielectric matrix. Except for materials with shallow d states, this approximation yields consistently too large band gaps. van Schilfgaarde *et al.* have already argued that this is related to the neglect of vertex corrections in W ,⁹ which would result in additional screening from the attractive electron-hole interaction. Although the self-consistent quasiparticle GW technique (scQP GW) of Faliev *et al.*³⁰ is probably yet the most concise method to perform GW calculations, it fails short in an accurate prediction of band gaps. From a *pragmatic* point of view, GW_0 calculations are an excellent, may often be even preferable, alternative. This method is at least a factor 2–4 faster than a fully self-consistent quasiparticle calculation and readily available

in most GW codes. Our conclusion might well change once vertex corrections, i.e., electrostatic electron-hole interactions, are properly accounted for in the dielectric matrix, but the effort to do so will be rather large.

The last issue is whether further improvements on the density-functional side are conceivable. Since we have shown that a one-to-one correspondence between the predicted static screening properties and the predicted quasiparticle band gaps exists, any effort to devise density functionals yielding better static screening properties should give improvements for the GW_0 band gaps as well. For instance, all calculations presented here are based on gradient corrected functionals which yield slightly larger band gaps and smaller static dielectric constants already on the DFT level (compared to LDA) and, therefore, more accurate band gaps in the GW_0 approximation. We have also discussed that the remaining errors in the screening properties are most likely related to an incorrect description of the semicore d states in III–V and II–VI semiconductors within LDA/GGA. These too shallow d states are a result of the only partial cancellation of the Coulomb self-interaction within the d shell. Promising systematic functionals to overcome these shortcomings are the self-interaction free exact-exchange optimized effective-potential method, which has been applied successfully to ZnS and GaN, yielding excellent G_0W_0 band gaps.¹³ However, we also note that EXX-OEP+ G_0W_0 calculations give too large band gaps for Si, Ge, and other semiconductors,⁶¹ implying that the suitability of EXX-OEP+ G_0W_0 calculations is not yet settled. Other promising approaches to overcome the deficiencies of local functionals are hybrid Hartree-Fock density functionals, on

which we will report in future work.⁶² Here, we have used a simpler approach shifting the d states toward stronger binding energies using an DFT+ U approach in the calculations preceding the GW calculations. This gave more accurate static dielectric properties and slightly better quasiparticle band gaps. However, the d -band position remained much too shallow in the successive GW_0 calculations. More work is certainly required to understand the failure of the GW approximation in describing localized electrons.

V. CONCLUSION

Our present work shows that an update of the quasiparticle energies in the one-electron Green's function yields generally quasiparticle energies in excellent agreement with experiment, if and only if the screening properties and W are calculated using density-functional one-electron energies. With a mean absolute relative error of 5%, the approximation is superior to the more common single shot G_0W_0 approximation (10% error). The scheme is computationally efficient, and we firmly believe that this GW_0 approximation is and will remain a valuable tool for the prediction of band gaps. The success and failure of this approximation are clearly linked to the accuracy of the predicted static screening properties in the local-density-functional approximation.

ACKNOWLEDGMENT

This work has been supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung within the START project.

*Present address: Department of Chemistry, University of Calgary, University Drive 2500, Calgary, AB, T2N 1N4 Canada.

¹L. Hedin, Phys. Rev. **139**, A796 (1965).

²F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).

³W. G. Aulbur, L. Jonsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (2000).

⁴M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

⁵F. Bechstedt, R. Del Sole, G. Cappellini, and L. Reining, Solid State Commun. **84**, 765 (1992).

⁶T. Kotani and M. van Schilfhaarde, Solid State Commun. **121**, 461 (2002).

⁷S. Lebegue, B. Arnaud, M. Alouani, and P. E. Blochl, Phys. Rev. B **67**, 155208 (2003).

⁸M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).

⁹M. van Schilfhaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).

¹⁰S. Massidda, R. Resta, M. Posternak, and A. Baldereschi, Phys. Rev. B **52**, R16977 (1995).

¹¹M. Oshikiri and F. Aryasetiawan, J. Phys. Soc. Jpn. **69**, 2123 (2000).

¹²M. Usuda, N. Hamada, T. Kotani, and M. van Schilfhaarde, Phys. Rev. B **66**, 125101 (2002).

¹³P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheff-

ler, New J. Phys. **7**, 126 (2005).

¹⁴M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **57**, 6485 (1998).

¹⁵A. Fleszar and W. Hanke, Phys. Rev. B **71**, 045207 (2005).

¹⁶B. Holm and U. von Barth, Phys. Rev. B **57**, 2108 (1998).

¹⁷W. D. Schöne and A. G. Eguiluz, Phys. Rev. Lett. **81**, 1662 (1998).

¹⁸W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002).

¹⁹K. Delaney, P. Garcia-Gonzalez, A. Rubio, P. Rinke, and R. W. Godby, Phys. Rev. Lett. **93**, 249701 (2004).

²⁰M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).

²¹M. P. Surh, J. E. Northrup, and S. G. Louie, Phys. Rev. B **38**, 5976 (1988).

²²W. Luo, S. Ismail-Beigi, M. L. Cohen, and S. G. Louie, Phys. Rev. B **66**, 195215 (2002).

²³M. S. Hybertsen and S. G. Louie, Phys. Rev. B **37**, 2733 (1988).

²⁴M. P. Surh, S. G. Louie, and M. L. Cohen, Phys. Rev. B **43**, 9126 (1991).

²⁵X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14142 (1991).

²⁶O. Zakharov, A. Rubio, X. Blase, M. L. Cohen, and S. G. Louie, Phys. Rev. B **50**, 10780 (1994).

²⁷M. P. Surh, H. Chacham, and S. G. Louie, Phys. Rev. B **51**, 7464 (1995).

- ²⁸M. Rohlfling, Peter Kruger, and J. Pollmann, Phys. Rev. B **56**, R7065 (1997).
- ²⁹M. M. Rieger, L. Steinbeck, I. D. White, H. N. Rojas, and R. W. Godby, Comput. Phys. Commun. **117**, 211 (1999).
- ³⁰S. V. Faleev, M. van Schilfhaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
- ³¹F. Bruneval, N. Vast, and L. Reining, Phys. Rev. B **74**, 045102 (2006).
- ³²A. N. Chantis, M. van Schilfhaarde, and T. Kotani, Phys. Rev. Lett. **96**, 086405 (2006).
- ³³J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
- ³⁴J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán, J. Chem. Phys. **124**, 154709 (2006).
- ³⁵T. C. Harman and I. Maingailis, Appl. Solid State Sci. **4**, 1 (1974).
- ³⁶C. K. Hewes, M. S. Adler, and S. D. Senturia, Phys. Rev. B **7**, 5195 (1973).
- ³⁷D. L. Mitchell, E. D. Palik, and J. N. Zemel, in *Proceedings of the Seventh International Conference of the Physics of Semiconductors*, edited by M. Hulin (Dunod, Paris, 1964), p. 325.
- ³⁸C. Kittel, *Introduction to Solid State Physics*, 6th ed. (Wiley, New York, 1986), p. 185.
- ³⁹P. Y. Yu and M. Cardona, *Fundamentals of Semiconductions* (Springer-Verlag, Berlin, 2001).
- ⁴⁰S. M. Sze, *Physics of Semiconductor Devices* (Wiley Interscience, New York, 1981).
- ⁴¹T. Lei, T. D. Moustakas, R. J. Graham, Y. He, and S. J. Berkowitz, J. Appl. Phys. **71**, 4933 (1992).
- ⁴²*Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, SiC, and SiGe*, edited by M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur (Wiley, New York, 2001).
- ⁴³R. C. Whited, C. J. Flaten, and W. C. Walker, Solid State Commun. **13**, 1903 (1973).
- ⁴⁴M. Piacentini, D. W. Lynch, and C. G. Olson, Phys. Rev. B **13**, 5330 (1976).
- ⁴⁵N. Schwentner, F. J. Himpsel, V. Saile, M. Skibowski, W. Steinmann, and E. E. Koch, Phys. Rev. Lett. **34**, 528 (1975).
- ⁴⁶K. Hummer, A. Grüneis, and G. Kresse, Phys. Rev. B **75**, 195211 (2007).
- ⁴⁷V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, Phys. Rev. B **69**, 075102 (2004).
- ⁴⁸Z. Zhang, R. A. Lukaszew, C. Cionca, X. Pan, R. Clarke, M. Yeadon, A. Zambano, D. Walko, E. Dufresne, and S. te Velthuis, J. Vac. Sci. Technol. A **22**, 1868 (2004).
- ⁴⁹R. W. G. Wyckoff, *Crystal Structures* (Interscience New York, 1963).
- ⁵⁰F. Bechstedt and R. Del Sole, Phys. Rev. B **38**, 7710 (1988).
- ⁵¹M. Gehler, M. Ramsteiner, O. Brandt, H. Yang, and K. H. Ploog, Appl. Phys. Lett. **67**, 733 (1995).
- ⁵²N. Ashkenov, B. N. Mbenkum, C. Bundesmann, V. Riede, M. Lorenz, D. Spemann, E. M. Kaidashev, A. Kasic, M. Schubert, M. Grundmann, G. Wanger, H. Newmann, V. Darakchieva, H. Arwin, and B. Monema, J. Appl. Phys. **93**, 126 (2003).
- ⁵³J. Chen, Z. H. Levine, and J. W. Wilkins, Appl. Phys. Lett. **66**, 1129 (1995).
- ⁵⁴N. J. Shevchik, J. Tajeda, and M. Cardona, Phys. Rev. B **9**, 2627 (1974).
- ⁵⁵G. Martin, A. Botchkarev, A. Rockett, and H. Morkoc, Appl. Phys. Lett. **68**, 2541 (1996).
- ⁵⁶L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, and D. A. Shirley, Phys. Rev. B **9**, 600 (1974).
- ⁵⁷S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).
- ⁵⁸A. Rohrbach, J. Hafner, and G. Kresse, Phys. Rev. B **69**, 075413 (2004).
- ⁵⁹T. Miyake, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. B **74**, 245213 (2006).
- ⁶⁰C. Friedrich, A. Schindlmayr, S. Blügel, and T. Kotani, Phys. Rev. B **74**, 045104 (2006).
- ⁶¹A. Fleszar, Phys. Rev. B **64**, 245204 (2001).
- ⁶²F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse (private communication).