Quasiparticle band structure based on a generalized Kohn-Sham scheme

F. Fuchs,[*] J. Furthmüller, and F. Bechstedt

Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität, Max-Wien-Platz 1, D-07743 Jena, Germany

M. Shishkin and G. Kresse

Faculty of Physics and Center for Computational Materials Science, Universität Wien, A-1090 Wien, Austria
(Received 2 August 2007; published 11 September 2007)

We present a comparative full-potential study of generalized Kohn-Sham (gKS) schemes with explicit focus on their suitability as starting point for the solution of the quasiparticle equation. We compare G_0W_0 quasiparticle band structures calculated upon local-density approximation (LDA), screened-exchange, HSE03, PBE0, and Hartree-Fock functionals for exchange and correlation (XC) for Si, InN, and ZnO. Furthermore, the HSE03 functional is studied and compared to the generalized gradient approximation (GGA) for 15 nonmetallic materials for its use as a starting point in the calculation of quasiparticle excitation energies. For this case, the effects of self-consistency in the GW self-energy are also analyzed. It is shown that the use of a gKS scheme as a starting point for a perturbative quasiparticle correction can improve upon the deficiencies found for LDA or GGA starting points for compounds with shallow d bands. For these solids, the order of the valence and conduction bands is often inverted using local or semilocal approximations for XC, which makes perturbative G_0W_0 calculations unreliable. The use of a gKS starting point allows for the calculation of fairly accurate band gaps even in these difficult cases, and generally single-shot G_0W_0 calculations following calculations using the HSE03 functional are very close to experiment.

PACS number(s): 71.15.Mb, 71.15.Qe, 71.20.Nr

I. INTRODUCTION

Density functional theory (DFT) has become the most successful method for condensed matter calculations. This success is largely rooted in the simplicity of the exchange and correlation (XC) energy in the local-density (LDA) or generalized gradient (GGA) approximation. However, the underlying Kohn-Sham (KS) formalism fails in the prediction of electronic excitation energies of semiconductors and insulators.[1] A significant step forward to correct excitation energies was achieved when the first *ab initio* calculations of quasiparticle (QP) states were performed.[2,3] Their description is based on the quasiparticle equation with the XC self-energy Σ for one excited electron or hole.[1–4] In general, its solution is based on Hedin's GW approximation (GWA) for the self-energy[4] and a perturbative treatment of the difference to the XC potential used in the KS equation. The central quantity is the dynamically screened Coulomb potential W, which characterizes the reaction of the electronic system after excitation. For many nonmetals, such as the semiconductor silicon (Si), the method works well with an accuracy of $0.1-0.3$ eV for their QP gaps,[1] if the Green's function G is described by one pole at the KS (i.e., G_0), or better, at the QP (i.e., G) energy. However, for systems with a wrong energetic ordering of the KS bands, first-order perturbation theory is not applicable.[5] Examples are semiconductors with a negative fundamental gap in DFT-LDA or -GGA, e.g., InN (Ref. 6 and references therein), or with shallow d bands, e.g., ZnO.[7]

Besides the KS approach itself, the origin of the band gap problem is related to the semilocal approximation (LDA and/or GGA) for XC, which introduces an unphysical self-interaction and lacks a derivative discontinuity.[8,9] These deficiencies can be partially overcome using self-interaction-free exact exchange (EXX) potentials,[10] which are special

realizations of an optimized effective potential (OEP) method.[7] A, by conception, different way to address the band gap problem is the use of a generalized Kohn-Sham (gKS) scheme, which means starting from a scheme with a spatially nonlocal XC potential.[11] In this framework, the screened-exchange (sX) approximation uses a statically screened Coulomb kernel instead of the bare kernel in the Hartree-Fock (HF) exchange[11] and, with it, resembles the screened-exchange (SEX) contribution to the XC self-energy in the GWA.[3,4] Other hybrid functionals such as those following the suggestions of Adamo and Barone (PBE0)[12] or Heyd, Scuseria, and Ernzerhof (HSE03)[13] combine parts of bare or screened exchange with an explicit density functional. The gKS eigenvalues are usually in much better agreement with the experiment than the LDA and/or GGA ones. Therefore, the gKS solutions are supposed to be superior starting points for a QP correction, since first-order perturbation theory should be justified. Hence, the replacement of G by G_0 calculated from solutions of a gKS scheme may be interpreted as a first step toward a self-consistent determination of the self-energy operator.[14,15]

Here, we report a systematic study of QP energies calculated from GW corrections to the results of gKS schemes. First, in Sec. III A, we evaluate the performance of G_0W_0 corrections to different gKS starting points for Si, InN, and ZnO using the functionals sX, HSE03, PBE0, and HF. The results are compared to those of the standard KS approach based on an LDA functional. In Sec. III B, QP gaps are calculated for a benchmark set of 15 nonmetals utilizing the HSE03 starting point. The effects of self-consistency in G and W are discussed in comparison to the results based on a GGA starting point used in Ref. 16.

II. METHOD

All calculations are performed at the experimental lattice constants. We use the projector augmented-wave method as implemented in the Vienna *ab initio* simulation package.[17] For the details of the *GW* implementation, we refer to Refs. 16 and 18. The *GW* calculations are carried out using a total number of 150 bands for all materials. For the Brillouin-zone integrations, $8 \times 8 \times 8$ **k**-point meshes including the Γ point were used, except in the case of ZnO (LDA) and InN (sX, HSE03) where the **k**-point convergence of W was found to be critical for meshes containing Γ. In these cases, $8 \times 8 \times 8$ Monkhorst-Pack **k**-point grids avoiding Γ were used for the evaluation of W.

One problem of the presence of shallow d levels is the strong core-valence XC interaction.[14,19] It can be estimated within the LDA or HF approximation, where the latter is expected to be more reliable since the *GW* self-energy approaches the bare Fock exchange operator in the short wave length regime (i.e., at large electron binding energies). Therefore, we apply the HF approximation to the core-valence XC self-energy for all *GW* calculations reported here (see Ref. 16 for a validation of this approach).

Here, we do not update the QP wave functions corresponding to the neglect of nondiagonal matrix elements of the self-energy represented in terms of the gKS wave functions ψ_λ^{gKS}. The QP excitation energy ϵ_λ^{N+1} of a state λ in the $(N+1)$th iteration is related to the Nth iteration through the linearized equation

$$\varepsilon_\lambda^{N+1} = \varepsilon_\lambda^N + Z_\lambda^N \operatorname{Re}[\langle \psi_\lambda^{gKS} | T + V_{n-e} + V_H + \Sigma(\varepsilon_\lambda^N) | \psi_\lambda^{gKS} \rangle - \varepsilon_\lambda^N], \tag{1}$$

where T is the kinetic energy operator, V_{n-e} the nuclei potential, V_H the Hartree potential, and Z_λ^N the renormalization factor given by

$$Z_\lambda^N = \left(1 - \operatorname{Re}\langle \psi_\lambda^{gKS} | \left. \frac{\delta}{\delta \varepsilon} \Sigma(\varepsilon) \right|_{\varepsilon_\lambda^N} | \psi_\lambda^{gKS} \rangle \right)^{-1}. \tag{2}$$

The first iteration, usually denoted by G_0W_0,[1,14] is based on the gKS eigenvalues $\epsilon_\lambda^0 = \epsilon_\lambda^{gKS}$ as initial input to the *GW* calculation. Within this approximation, the perturbation operator in Eq. (1) becomes $\Sigma(\epsilon_\lambda^{gKS}) - V_{XC}^{gKS}$ corresponding to the difference between the *GW* self-energy and the nonlocal XC

TABLE I. Parameters of the gKS exchange functionals used.

Functional	α	sr Coulomb kernel	μ (Å^{-1})
LDA	0.00	$1/|\mathbf{x}|$	
sX	1.00	$\exp(-\mu|\mathbf{x}|)/|\mathbf{x}|$	1.55
HSE03	0.25	$\operatorname{erfc}(\mu|\mathbf{x}|)/|\mathbf{x}|$	0.3
PBE0	0.25	$1/|\mathbf{x}|$	
HF	1.00	$1/|\mathbf{x}|$	

potential used in the gKS equation, and the G_0W_0-QP shift for a certain gKS state is given by

$$\Delta_{\lambda\lambda} = Z_\lambda^0 \operatorname{Re}\langle \psi_\lambda^{gKS} | \Sigma(\varepsilon_\lambda^0 = \varepsilon_\lambda^{gKS}) - V_{XC}^{gKS} | \psi_\lambda^{gKS} \rangle. \tag{3}$$

In the actual implementation of the gKS schemes, we split the gKS XC energy into the form

$$E_{XC}^{gKS} = E_{XC}^{DFT} + \alpha[E_X^{sr}(\mu) - E_X^{DFT,sr}(\mu)], \tag{4}$$

i.e., a short-range nonlocal exchange term is added and treated exactly resulting in a nonlocal (screened) exchange potential. The superscript *DFT* indicates that the respective quantity is evaluated in some (quasi)local approximation, while E_X^{sr} corresponds to one of the (screened) Coulomb kernels given in Table I. The weight α of the short-range part and the inverse screening length μ are also listed in this table. For simplicity, in the case of sX, the inverse screening length μ corresponding to the Thomas-Fermi wave vector was chosen to be material independent, $k_{TF} = 1.55$ Å^{-1}. Gradient corrections were used for the HSE03 and PBE0 functionals.

III. RESULTS

A. G_0W_0 quasiparticle band structure of Si, ZnO, and InN

1. Generalized Kohn-Sham bands

Results of the gKS calculations are summarized in Table II together with experimental results. We notice an increase of the computed gaps when going from the LDA to a truly nonlocal XC functional. This can be attributed to the reduction of the spurious self-interaction found in LDA, the inclu-

TABLE II. Direct and indirect generalized KS band gaps $E_{g(d,i)}^{gKS}$ and average d-band binding energies E_d^{gKS} calculated for cubic Si, InN, and ZnO. Experimental values are from data collections in Refs. 6, 16, and 20–24. In the case of ZnO they refer to the wurtzite polytype. All values are given in eV.

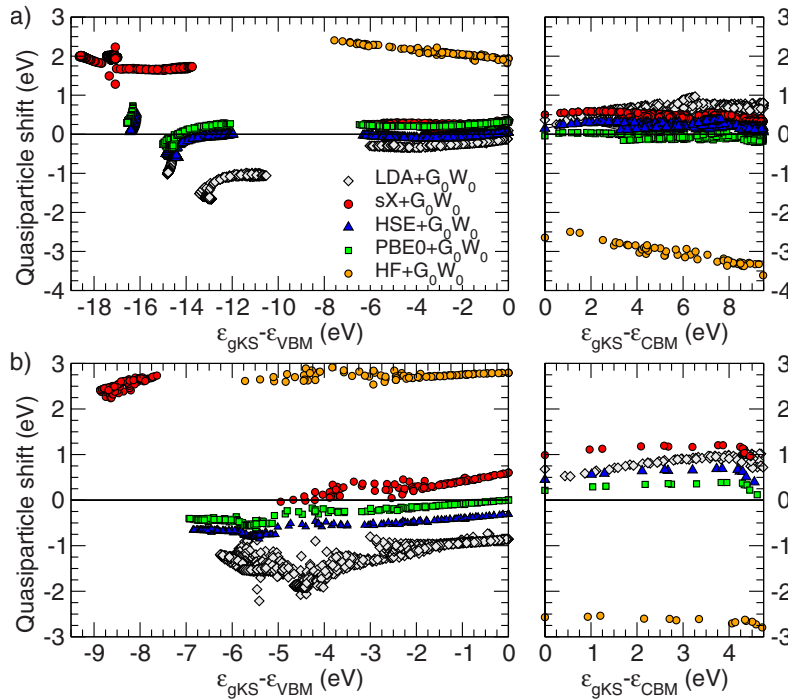|	Energy	LDA	sX	HSE03	PBE0	HF	Expt.
Si	$E_{g,i}^{gKS}$	0.51	0.98	1.04	1.85	6.57	1.17
	$E_{g,d}^{gKS}$	2.53	3.23	3.15	3.98	9.08	3.40
InN	E_g^{gKS}	−0.38	0.39	0.37	1.14	7.15	0.61
	E_d^{gKS}	13.1	17.2	14.6	14.7	18.6	16.0–16.9
ZnO	E_g^{gKS}	0.6	2.97	2.11	3.03	11.07	3.44
	E_d^{gKS}	4.6	8.2	5.7	5.8	9.3	7.5–8.8

FIG. 1. (Color online) First-order quasiparticle shifts $\Delta_{\lambda\lambda}$ versus gKS eigenvalues for (a) InN and (b) ZnO. The valence-band maximum (VBM) and the conduction-band minimum (CBM) are taken as energy zeros for occupied and empty states, respectively. Results for five different starting gKS band structures are shown.

sion of a potential discontinuity between filled and empty states in the XC functional,[8,9,11] and the—in comparison to LDA—enhanced core-valence exchange. Furthermore, for direct gaps we observe the tendency to increase from LDA over HSE03, sX, and PBE0 to the HF values.

For ZnO and InN, the d-band binding energies increase with respect to the LDA, approaching the experimental values. This is an important fact to note, since the difficulties of the LDA+G_0W_0 approach for these compounds partly result from too shallow d electrons in LDA, which is itself a result of the spuriously contained self-interaction in local and semilocal functionals. The too shallow d electrons hybridize too strongly with the p bands at the valence-band maximum (VBM), pushing them upward, in turn decreasing the gap (pd repulsion)[6] beyond the common LDA gap underestimation found, for example, in Si. Contrary to the LDA, the gKS starting points yield stronger bound d electrons (see Table II) with binding energies closer to the experimental values. Hence, the gKS functionals can be expected to give a more reasonable estimate for the influence of the pd repulsion on the fundamental gap and to provide better starting wave functions for the GW calculations. For instance, in ZnO the Zn d character of the wave functions at the VBM is decreased from 0.3 for LDA to 0.25 for HSE03.

In comparison to experiment, the gKS functionals, with exception of HF, generally perform better than standard DFT-LDA for the gaps and d-electron binding energies. For InN, there is another important fact to note. All gKS functionals are found to yield the correct ordering of the Γ_{1c} and Γ_{15v} states at the zone center, in contrast to LDA findings which give a negative sp gap.[6] This is essential for the QP

description, since a correct energetic ordering of the single-particle states is an inevitable prerequisite for a perturbative treatment of the GW corrections.[1,15]

2. Quasiparticle shifts

Figure 1 shows the calculated QP shifts [Eq. (3)] for InN and ZnO plotted versus the gKS eigenvalues. In the extreme limit of the HF starting point, the QP shifts are given by the correlation self-energy Σ_C scaled with the dynamical renormalization factor and consequently undergo the sign change of Σ_C at the Fermi level. For the LDA, the exact opposite sign change is observed, i.e., valence bands acquire a negative shift and conduction bands are shifted upward. It is remarkable that the two hybrid functionals HSE03 and PBE0, which essentially mix 25% HF and 75% DFT exchange, yield very small QP shifts. This highlights that the 1/4 recipe is indeed a remarkable good and robust choice not only for total energies but also for one-electron energies in semiconductors.

The performance of GW upon sX is somewhat disappointing. The d-band positions were very well described using the sX approach (see Table II), but applying GW corrections, the localized d states are shifted upward by 1.5–2 eV, significantly deteriorating agreement with experiment. We will return to this issue in the next section and in Sec. IV. Since the SEX term in the GW approximation is very similar to the sX term in gKS schemes, the differences must be mainly related to a different behavior of the Coulomb hole term in the GW approximation and the local-density part of the sX functional.

TABLE III. Direct (d) and indirect (i) G_0W_0 QP band gaps, average d-band binding energies, and static electronic macroscopic dielectric constants ε_∞ calculated upon the respective gKS band structures for Si, ZnO, and InN. All energy values are given in eV. Experimental results are given for comparison.

	Energy	LDA	sX	HSE03	PBE0	HF	Expt.
Si	$E_{g,i}^{QP}$	1.08	1.31	1.32	1.65	2.93	1.17
	$E_{g,d}^{QP}$	3.18	3.49	3.48	3.72	5.21	3.40
	ε_∞	13.9	10.8	9.8	7.8	3.4	11.90
InN	E_g^{QP}	0.00	0.55	0.47	0.78	2.56	0.61
	E_d^{QP}	15.1	15.6	15.2	15.3	16.6	16.0–16.9
	ε_∞	12.2	6.6	6.8	4.9	2.4	7.96
ZnO	E_g^{QP}	2.14	3.36	2.87	3.24	5.71	3.44
	E_d^{QP}	5.6	6.2	6.1	6.2	7.0	7.5–8.8
	ε_∞	5.3	3.0	3.4	3.0	1.8	3.74

Turning now to a detailed discussion of the InN results [Fig. 1(a)], we note that the QP shifts for the upper valence bands and the conduction bands show a rather weak dispersion with the eigenstate for all gKS schemes. In this energy range, they may be approximated by scissors operators for occupied and empty states, whose absolute values depend on the used gKS functional. Only for the lowest In $4d$- and N $2s$-derived bands, at gKS energies below -10 eV, do noticeable deviations occur in the form of a downward or an upward shift for the LDA and sX starting points, respectively. In the considered energy range, the QP shifts calculated for the HSE03 or PBE0 starting point are remarkably small with respect to the dispersion and amplitude, and they remain well below 1 eV. For ZnO [Fig. 1(b)], similar observations can be made; however, as discussed before, the smaller binding energy of the Zn $3d$ states causes a stronger hybridization of p and d states than in InN. This might explain the upward bending of the sX shifts close to the VBM and the stronger bending of the LDA conduction-band shifts.

3. Non-self-consistent quasiparticle bands

Table III shows the results of the G_0W_0 calculations for the fundamental gaps and d-electron binding energies. In contrast to the LDA or gKS energies, these QP energies should have a physical meaning as measurable quantities and thus can be compared directly to the experimental gap values. However, it is necessary to note that the experimental value for ZnO corresponds to the wurtzite instead of the zincblende polytype used in the calculations. The zincblende gap is expected to be about 0.2 eV smaller than the wurtzite one.[6,25] For silicon, the QP gaps calculated upon the LDA, sX, or HSE03 starting point bracket the experimental values closely, with a maximum deviation of 0.2 eV. The LDA $+G_0W_0$ values slightly underestimate the gaps. In fact, this corresponds to the general trend of gap underestimation in the LDA$+G_0W_0$ and/or GGA$+G_0W_0$ approach using full-potential methods recently established by different groups.[15,16] For the gKS starting points, the calculated gaps are larger than the experimental ones. In more detail, starting from sX or HSE03, we find virtually identical results very

close to the experiment, while the PBE0$+G_0W_0$ approach overestimates the gaps significantly. For the HF starting point, the largest deviations are found, which could be expected already from the large differences between the QP eigenvalues ε_λ^{QP} and the HF ones ε_λ^{HF} and the consequent breakdown of the perturbative G_0W_0 treatment in Eq. (1). For the PBE0 starting point, however, the deviations demand for a more subtle interpretation. To understand this deviation, the static random phase approximation (RPA) dielectric constants calculated from the gKS eigenvalues and wave functions are included in Table III. Indeed, a comparison with the experimental values suggests significant underscreening for the PBE0 starting point, which certainly contributes to the overestimation of the gaps. The values calculated upon the sX and HSE03 functional are in much better agreement with the experiment, which is bracketed closely by them (sX/HSE03) and the LDA value.

For InN and ZnO, the trends just discussed for Si still hold, but the actual benefit of a gKS starting point becomes more apparent. In contrast to the LDA, the gKS starting points reproduce the correct ordering of gap states for InN and yield more meaningful dielectric constants in the case of HSE03 and, with some restrictions, in the case of sX. Taking into account that different polytypes are compared, the agreement between the experimental gap values and the sX $+G_0W_0$ or HSE03$+G_0W_0$ results is fair.

Similar to the gaps, the QP d-band binding energies calculated for InN and ZnO show a much weaker variation with the starting XC functional than the original gKS one-electron d-band energies. The agreement of the calculated energies with experimental values (Table III) is good for InN. For ZnO with more shallow d levels, the picture is less clear, although here the comparison with experiment is also hampered by the remarkable scatter in the measured values. In general, we observe that the d states become more strongly bound for the gKS$+G_0W_0$ approach than for LDA$+G_0W_0$, therefore certainly moving in the right direction. Hence, the better treatment of exchange and correlation in the gKS starting functionals improves the prediction of the semicore d bands, although a general tendency toward too shallow theoretical d states clearly remains. Note, in particular, that the

TABLE IV. Results for the fundamental gaps of the HSE03 and quasiparticle (G_0W_0, GW_0, and GW) calculations and static electronic macroscopic dielectric constants as used in W_0 (RPA). The calculated values for the spin-orbit (SO) coupling induced gap closing given in the last column have been included in the gaps. Also reported is the mean absolute relative error (MARE) and the mean relative error (MRE) for the gaps. Experimental data for the gaps and dielectric constants are given for comparison (for references, see Ref. 16); underlined values indicate zero temperature values.

	HSE03	G_0W_0	GW_0	GW	Expt.	ε	ε^{expt}	SO
Ge	0.54	0.79	0.82	0.83	0.74	14.0	16.00	0.08
Si	1.04	1.32	1.35	1.37	1.17	9.8	11.90	
GaAs	1.12	1.66	1.71	1.75	1.52	9.5	11.10	0.10
SiC	2.03	2.60	2.68	2.76	2.40	5.6	6.52	
CdS	1.97	2.55	2.65	2.80	2.42	4.6	5.30	0.02
AlP	2.09	2.69	2.77	2.86	2.45	6.3	7.54	
GaN	2.65	3.29	3.38	3.53	3.20	4.6	5.30	0.00
ZnO	2.11	2.86	3.02	3.33	3.44	3.4	3.74	0.01
ZnS	3.05	3.69	3.79	3.95	3.91	4.5	5.13	0.02
C	5.08	5.84	5.92	6.03	5.48	4.9	5.70	
BN	5.54	6.54	6.66	6.85	6.1–6.4	3.9	4.50	
MgO	6.22	7.94	8.20	8.66	7.83	2.6	3.00	
LiF	11.2	14.1	14.5	15.2	14.20	1.8	1.90	
Ar	10.1	13.7	14.1	14.7	14.20	1.6		
Ne	14.1	20.2	20.7	21.4	21.70	1.2		
MARE	21%	6.8%	8.0%	10.0%				
MRE	−21%	2.3%	5.3%	9.3%				

d bands shift upward—away from the experimental values—starting from the sX functional. Similar observations have been made for an LDA+U starting point in Refs. 16 and 26. This clearly points to a deficiency of the GW approximation.

B. Self-consistent quasiparticle calculations starting from the HSE03 functional

One important question concerns the influence of self-consistency and the resulting QP corrections in the case of a gKS starting point, especially in comparison to local or semilocal DFT. For that reason, we have calculated the QP gaps for 15 materials, without and with partial (only in G) or full (in both G and W) self-consistency with respect to the eigenvalues, starting from the eigenvalues and wave functions of the HSE03 functional. In Secs. III A 2 and III A 3, this starting point was found to give the best results for the fundamental gaps and the smallest quasiparticle shifts. The materials addressed are nonmetals, spanning the range from small-gap semiconductors to insulators. They are chosen as a subset of those considered in Ref. 16. With the technical details kept largely identical, except for the different starting point, the data collected in Table IV allow for a direct and unbiased comparison of the GGA(PBE) starting point[16] and the HSE03 starting point used in this work. The only important difference from Ref. 16 is that we now restore the all-electron charge density *exactly* on the plane wave grid for the calculation of the correlation energy. This yields technically more accurate d-band binding energies. Details of the applied procedure will be published elsewhere.[27]

Our results show that the HSE03 gaps, even though they are generally closer to the experiment than the DFT-LDA and/or DFT/GGA ones, still underestimate the experimental gaps on average by 21%. This underestimation is cured and turned into a slight overestimation of about 2.3% upon the inclusion of G_0W_0 quasiparticle corrections. The mean absolute relative error (MARE) is reduced to 6.8%, which is a significant improvement compared to the 9.9% MARE obtained for the GGA+G_0W_0 gaps.[16] Basically, this improvement results from the good performance of the HSE03 starting point for materials that comprise d electrons such as GaAs, CdS, GaN, ZnO, and ZnS, for which the HSE03 +G_0W_0 gap MARE is calculated to be 7.9%, while it is about 19.2% in the GGA+G_0W_0 approach. We attribute the better agreement to the improved description of the pd repulsion on the HSE03 level, which impacts the energy levels and wave functions (see Sec. III A 1). For the rest of the materials, both approaches GGA and HSE03 perform on par with a MARE of 4.8/ and 6.1%. While HSE03+G_0W_0 usually slightly overestimates the band gaps, the band gaps for ZnO, ZnS, LiF, Ar, and Ne remain underestimated compared to experiment. This may be related to the large errors of the HSE03 gap exceeding 30% for these materials. Thus, the inaccuracy of perturbation theory prevails for these systems, which share a relatively weakly screened exchange (static dielectric constant smaller than 4). Another interesting point to note is the different performance of both approaches for indirect semiconductors such as Si, SiC, AlP, C, and BN. For them, the GGA+G_0W_0 approach performs unexpectedly well, resulting in a MARE of only 2.6%, while the perfor-

TABLE V. Results for the *d*-band binding energies of GaAs, GaN, ZnO, and ZnS on different levels of quasiparticle self-consistency compared to experimental values starting from HSE03 and PBE wave functions and eigenvalues, respectively.

	HSE03	G_0W_0	GW_0	GW	Expt.
GaAs	17.2	17.5	17.6	17.6	18.9
GaN	15.4	16.1	16.3	16.5	17.0
ZnO	5.7	6.1	6.3	6.4	7.5–8.8
ZnS	7.5	7.2	7.2	7.3	9.0
	PBE	G_0W_0	GW_0		Expt.
GaAs	14.8	16.8	17.2		18.9
GaN	13.3	15.4	16.1		17.0
ZnO	5.2	6.1	6.4		7.5–8.8
ZnS	6.1	6.8	7.2		9.0

mance of the HSE03 starting point is a little worse on average with a MARE of 8.4%.

Partial self-consistency following the GW_0 scheme is found to increase the gaps further by about 0.1–0.2 eV (0.4 eV for LiF, Ar, and Ne). Since the HSE03+G_0W_0 gaps already showed the tendency to overestimate the experimental values, the MARE increases to 8.0%. This is in contrast to the findings for the GGA+GW_0 approach, where an update of the eigenvalues in G reduces the MARE to 5.7%.[16] Following the arguments given in Ref. [16], this can be related to the electronic macroscopic dielectric constants ε_∞ calculated within RPA, which are indicative of the screening involved in W_0. Obviously, the ε_∞ calculated using the HSE03 functional (see Table IV) are underestimated with respect to the experiment and, furthermore, they are generally lower than those calculated using the GGA XC functional, due to the generally larger HSE03 gaps. Only inclusion of excitonic effects, i.e., electron-hole binding, allows for the calculation of accurate electronic dielectric constants for the HSE03 functional.[28,29]

Further self-consistent calculations, updating the eigenvalues in both G and W, according to the GW scheme, were performed. Parallel to the findings for a GGA starting point, the GW gaps are larger than the GW_0 gaps, also for the HSE03 starting point. Consequently, they overestimate the experimental gaps with a MARE of 10%, which exceeds the GGA+GW MARE of 6.1%. The continued gap increase is found to be due to a further reduction of the screening upon updating the eigenvalues in W. The difference between the GGA+GW and HSE03+GW schemes (updating eigenvalues in both G and W) can be related only to different starting wave functions. Obviously, the wave functions influence the resulting gap and increase it on average by 4%. This is a fairly small change confirming the common conjecture that wave functions have only a small effect on the band gaps. Similar changes were observed in the self-consistent quasiparticle GW (scQPGW) suggested by Shishkin and Kresse[28] and Faleev and co-workers.[30,31]

Finally, we address the *d*-band binding energies calculated upon the HSE03 and PBE starting point for different levels of quasiparticle self-consistency, as shown in Table V. We

have to note that the present calculations do not include the *s* and *p* orbitals with the same main quantum number as the semicore *d* shell. However, since here the core-valence interaction is approximated by HF exchange rather than by LDA, as inherent to conventional pseudopotential calculations, the values given in Table V already provide a reasonable estimate for the *d*-band binding energies (see Refs. [16] and [18]). Updated values for the PBE case are also supplied; the present values supersede those in Ref. [16] and are more accurate, since the all-electron charge density is now accurately restored on the plane wave grid.[27] In general, the QP binding energies increase over the HSE03 one-electron values due to the quasiparticle corrections, which can be understood mainly from the effects of the enhanced core-valence interaction in the GW calculations. It is only in the case of ZnS where this trend does not hold for yet unknown reasons. However, compared to the experimental values, all three quasiparticle schemes studied here underestimate the *d*-band binding energies. With increasing self-consistency along the row (G_0W_0, GW_0, and GW), the *d* bands shift to larger binding energies. Thereby, for the Ga 3*d* levels, the calculated values approach the experimental ones. In the case of the Zn 3*d* levels, which are just below the *p*-like upper valence-band complex, the situation is different with a more pronounced underestimation of the binding energies. Analogue observations have been made for a GGA starting point,[16,26,32] which gives a smaller *d*-band binding energy than the HSE03 starting point for each level of self-consistency. We note that Fleszar and Hanke observed that the inclusion of vertex corrections in the self-energy shifts the *d* states to stronger binding energies,[32] suggesting that the neglect of such corrections is responsible for the erroneous behavior of the GW approximation for *d* states.

IV. SUMMARY AND CONCLUSIONS

We have presented G_0W_0 QP calculations starting from a variety of XC functionals: LDA, sX, HSE03, PBE0, and HF for Si, InN, and ZnO. We have shown that the gKS schemes, which take into account a screened-exchange potential or part of it (sX, HSE03), give rise to eigenvalues close to the

QP excitation energies. The resulting G_0W_0 corrections were found to yield QP energies in good agreement with the experimental data, and the QP shifts are less dispersive than those calculated upon LDA. Overall, the HSE03 and PBE0 functionals gave one-electron energies very close to the successive GW calculations, resulting in small QP gap corrections across the considered energy range. For the HSE03 $+G_0W_0$ case, the final QP energies were in very good agreement with experiment, whereas the PBE0 functional was found to yield too large QP gaps. We traced this back to a significant underestimation of the screening for the PBE0 functional, when the random phase approximation is used (also applied to determine W).

Furthermore, the QP gaps for 15 materials comprising small and large gap systems were calculated in order to provide a benchmark of the HSE03 starting point against the GGA one. It was shown that the HSE03+G_0W_0 approach yields an almost halved overall error for the fundamental gaps compared to the GGA starting point. The largest improvement over the LDA and/or GGA starting points was found for materials with shallow d states such as ZnO, ZnS, InN, GaAs, and GaN, where the LDA and/or GGA starting point suffers from a significant underestimation of the d-band binding energies and a consequently overestimated repulsion between p- and d-like states. Since the d-band binding energies calculated using one of the gKS schemes (e.g., HSE03) are closer to the experiment, the influence of the pd repulsion on the gap is described more accurately.

Furthermore, the effects of different degrees of self-consistency were investigated. It was found that both self-consistency in G (GW_0) and self-consistency in G and W (GW) impair the agreement with experimental data. For GW_0, this is in contrast to the findings for a GGA starting point.[16] This could be traced back to the poorer description of the dielectric screening for the HSE03 starting point. Self-consistency according to the GW scheme with an update of eigenvalues in both G and W further diminishes the agreement with measurements due to a further reduction of the already underestimated screening. This is analogous to the findings for GGA based GW calculations. In general, concerning self-consistent QP schemes, the present work confirms the observation already made in previous work: To obtain accurate QP gaps, it is essential to use an electronic response function and a screened interaction W that agree closely with experiment,[16,28] and to combine this screened interaction with an accurate Green's function G. For the GW_0 case based upon HSE03 wave functions, the overestimation of the gaps clearly relates to an underestimation of the static electronic screening employing HSE03 and the random phase approximation. In this light, the success of the perturbative single-shot HSE03+G_0W_0 approach is a little bit fortuitous, since the overestimation of the screening is partially canceled by too small gap corrections obtained using the single-shot perturbative G_0W_0 approach. Probably, the same is true for some other single-shot approaches, such as EXX-OEP+G_0W_0. Nevertheless, if computational efficiency is an important issue—and it more often is than not—then the HSE03+G_0W_0 approach is indeed an excellent balance between accuracy and speed. The calculations are as efficient as for the commonly used LDA+G_0W_0 method and, with very few exceptions, the errors are smaller than 10%. If better accuracy is required, it can be achieved but only by updating the wave functions and including excitonic effects in the calculation of the screening properties, i.e., vertex corrections in W.[28] For most matter, such calculations are currently not feasible due to the large computational requirements.

A similar accuracy as for the HSE+G_0W_0 approach can be achieved by starting from GGA wave functions and eigenvalues, and updating the eigenvalues in G until convergence is reached (see Ref. 16). This approach yields comparable errors as HSE03+G_0W_0, but it is computationally more demanding, since several iterations are required to converge G. The latter method is also problematic for materials with an inverted band order in LDA and/or GGA. Which approach to choose (HSE03+G_0W_0 or GGA+GW_0) is to some extent a matter of taste, and the final results are usually very close and often bracket the experiment. If efficiency and robustness (band order) are issues, the HSE03+G_0W_0 approach seems to be preferable, and it is certainly much more accurate than the traditional LDA+G_0W_0 method.

Concerning the position of the d levels, the HSE03 $+G_0W_0$ method shows an underestimation of the d-band binding energies by about 1 eV for almost all materials. Similar observations were made for the LDA and/or GGA case,[16,26,32] LDA+U based GW calculations,[16,26] and self-consistent quasiparticle GW (scQPGW) calculations.[28] The underestimation of the d-band binding energy is thus universal to the GW approximation and not related to the starting wave functions. The origin for this underestimation is yet unknown, but it is most likely related to the fact that the GW approximation is not entirely free of self-interaction errors, and only inclusion of vertex corrections in the self-energy might remedy this deficiency.[32]

*fuchs@ifto.physik.uni-jena.de

[1] W. Aulbur, L. Jönsson, and J. Wilkins, Solid State Phys. **54**, 1 (2000).

[2] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).

[3] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).

[4] L. Hedin, Phys. Rev. **139**, A796 (1965).

[5] O. Pulci, F. Bechstedt, G. Onida, R. Del Sole, and L. Reining, Phys. Rev. B **60**, 16758 (1999).

[6] J. Furthmüller, P. H. Hahn, F. Fuchs, and F. Bechstedt, Phys. Rev. B **72**, 205106 (2005).

[7] P. Rinke, A. Qteish, Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).

[8] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[9] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

[10] M. Städele, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. Lett. **79**, 2089 (1997).

[11] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, Phys. Rev. B **53**, 3764 (1996).

[12] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[13] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).

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

[15] T. Kotani and M. van Schilfgaarde, Solid State Commun. **121**, 461 (2002).

[16] M. Shishkin and G. Kresse, Phys. Rev. B **75**, 235102 (2007).

[17] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).

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

[19] S. Sharma, J. K. Dewhurst, and C. Ambrosch-Draxl, Phys. Rev. Lett. **95**, 136402 (2005).

[20] J. Schörmann, D. J. As, K. Lischka, P. Schley, R. Goldhahn, S. F. Li, W. Löffler, M. Hetterich, and H. Kalt, Appl. Phys. Lett. **89**, 261903 (2006).

[21] L. F. J. Piper, T. D. Veal, P. H. Jefferson, C. F. McConville, F. Fuchs, J. Furthmüller, F. Bechstedt, Hai Lu, and W. J. Schaff, Phys. Rev. B **72**, 245319 (2005).

[22] L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, and D. A. Shirley, Phys. Rev. B **9**, 600 (1974).

[23] R. T. Girard, O. Tjernberg, G. Chiaia, S. Soderholm, U. O. Karlsson, C. Wigren, H. Nylen, and I. Lindau, Surf. Sci. **373**, 409 (1997).

[24] P. Schley and R. Goldhahn (private communication).

[25] A. Schleife, F. Fuchs, J. Furthmüller, and F. Bechstedt, Phys. Rev. B **73**, 245212 (2006).

[26] T. Miyake, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. B **74**, 245213 (2006).

[27] J. Harl and G. Kresse (private communication).

[28] M. Shishkin and G. Kresse (unpublished).

[29] J. Paier, M. Marsman, and G. Kresse (private communication).

[30] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).

[31] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).

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