PHYSICAL REVIEW B **90**, 075125 (2014)

Predictive *GW* calculations using plane waves and pseudopotentials

Jiří Klimeš,* Merzuk Kaltak, and Georg Kresse
University of Vienna, Faculty of Physics and Center for Computational Materials Science, Sensengasse 8/12, A-1090 Vienna, Austria
(Received 14 April 2014; revised manuscript received 25 July 2014; published 14 August 2014)

We derive a finite-basis-set correction for quasiparticle (QP) energies in the *GW* approximation and many-body correlation energies in the random phase approximation. Since the correction requires only knowledge of the ground-state density distribution, it is straightforward to implement in any plane-wave code and significantly improves convergence at negligible computational cost. The expression also indicates that QP energies might converge to the wrong value using the projector augmented wave (PAW) method since the overlap densities of occupied orbitals and high-energy, plane-wave-like orbitals are inaccurately described. The error is shown to be related to the incompleteness of the partial waves inside the atomic spheres. It can be avoided by adopting norm-conserving partial waves. G_0W_0 and GW_0 results based on such norm-conserving PAW potentials are presented for a large set of semiconductors and insulators. Accurate extrapolation procedures to the infinite-basis-set limit and infinite-k-point limit are discussed in detail.

DOI: 10.1103/PhysRevB.90.075125 PACS number(s): 71.15.−m

I. INTRODUCTION

Density functional theory (DFT) does not allow us to calculate accurate quasiparticle (QP) energies as measured experimentally in x-ray photoelectron spectroscopy (XPS) or ultraviolet photoelectron spectroscopy (UPS), although the Kohn-Sham eigenvalues often show at least qualitative agreement with the corresponding measurements. In solid state physics, the standard approach to determine QP energies is currently the *GW* approximation often combined with orbitals from Kohn-Sham DFT calculations, for instance, the local density approximation (LDA). The *GW* method was originally proposed by Hedin [1] and later extensively applied by the group of Louie [2], following a procedure adopted by Sham and Schlüter for tight-binding models [3] and Strinati, Mattausch, and Hanke for local-basis-set calculations [4,5]. For recent reviews, we refer to Refs. [6–9], and for a recent rederivation we refer to Ref. [10]. With the advances in computer power over the last two decades, the *GW* method can now be applied to systems containing hundred atoms, and with some approximations, calculations for several hundred atoms are possible or will be possible in the near future.

Remarkably, even after several decades of research, publications with technically converged QP energies are rare and seem to pose a significant challenge to theory. The most problematic issue is the slow convergence of the QP energies with respect to the basis-set size [11–13]. That this slow convergence causes serious errors has not been realized for a long time, with most researchers using a few hundred unoccupied orbitals per atom in the calculation of the Green's function. However, the recent work of Shih *et al.* on ZnO suggests that convergence can be so slow that thousands of orbitals might be required for accurate predictions [14]. Furthermore, care must be taken to reach convergence with respect to both the basis set for the orbitals and the basis set for the response function [14]. Although the ZnO results were somewhat exaggerating the dependence on the number of orbitals and the final reported values for the band gap of ZnO were not very accurate, partly because of the use of a simplified plasmon-pole model [15], the calculations still mark a turning point: whenever highly accurate results are required for the QP energies, huge basis sets are needed to obtain them.

The situation is very similar to quantum chemistry (QC) calculations, where convergence of the total energies (and excitation energies) is also exceedingly slow with the basis-set size. In the QC community, it is well established that the interelectron cusp, the dependence of the many-electron wave function on the distance of two electrons, causes slow convergence [16–19]. As a consequence of a wave-function kink at zero distance, absolute correlation energies show a convergence proportional to the inverse of the number of basis functions [20]. A similar slow convergence was observed for the correlation energy in the random phase approximation (RPA). Based on calculations using the Lindhard dielectric function but without explicit derivations, Harl and Kresse suggested that the RPA correlation energy converges like $1/N$, where N is the plane-wave basis-set size [21]. For MP2 total energies, a similar behavior was derived by Shepherd *et al.*, closely paralleling the finding for quantum chemistry basis sets [22]. Gulans explicitly derived the same limiting behavior for the RPA correlation energy relying on the jellium model, and he furthermore showed that a similar $1/N$ convergence must be expected for the QP energies in solids [23,24]. Although his derivation suggests that the origin of the slow convergence of QP energies and total correlation energies is related, how exactly the cusp condition appears in methods based on one-electron Green's functions is somewhat "clouded." In fact, in one-electron Green's function approaches, the self-energy needs to account for the correlation hole or, equivalently, the screening charge, which relates at short distances to the cusp condition. Another way to understand the slow convergence of QP energies is that they essentially represent the difference between two total energy calculations, each with a $1/N$ convergence.

In this work, we derive a closed expression for the MP2 and RPA correlation energy and self-energy in the limit of high-energy states (plane waves with large momentum). The derivation does not rely on an explicit functional form for the dielectric function for the jellium electron gas. We only assume that for high energies, well above the vacuum level, the

*jiri.klimes@univie.ac.at

one-electron states become essentially plane waves [25,26]. The correction turns out to depend on the charge density distribution alone. Interestingly, the present derivation also shows that the static Coulomb-hole plus screened-exchange (COHSEX) approximation overestimates the high-energy contributions to the QP energies by a factor of 2, as observed recently by Kang and Hybertsen [27]. Furthermore, we show that in order to obtain converged values, one needs to increase both the basis set for the orbitals, as well as the basis set for the response function.

Unfortunately, the present derivation also highlights a severe problem for ultrasoft pseudopotentials and the related projector augmented wave potentials when they fail to conserve the norm in the plane-wave representation. For ground-state properties, violation of the norm is usually eliminated by adding appropriate augmentation or compensation charges centered at each atomic site. However, such a correction is missing for orbitals at high energies that are essentially plane waves with large momentum. As a result, the overlap densities of occupied orbitals and high-energy unoccupied orbitals are not properly described. A simple solution to this problem is to construct potentials that are norm conserving. In this work, this is achieved by determining projector augmented wave potentials with norm-conserving partial waves. We present tests for ZnO, GaAs, and AlAs that indicate that QP energies extrapolated to the infinite-basis-set limit are then independent of the detailed construction of the potential. The final gaps agree very well with previous calculations of Friedrich *et al.* for ZnO [28]. Motivated by this result, we apply the present procedure to 24 semiconductors and insulators and present G_0W_0 as well as GW_0 calculations based on LDA and PBE (Perdew-Burke-Ernzerhof) [29] orbitals and energies. We believe that these results are technically well converged and can serve as a stringent benchmark for future GW calculations.

II. THEORY

In the following sections, we study the asymptotic convergence of the total many-electron correlation energy (in Sec. II A) and self-energy (in Sec. II C). Specifically, we derive the contributions from plane waves with large momentum in the auxiliary basis set that is used to represent the overlap charge densities or the response function. By summing up all contributions between a finite and infinite cutoff, we obtain an estimate of the error made in practical calculations when a finite auxiliary basis set is used. In the derivation we assume that the orbital basis set is complete and all unoccupied states are included. Our crucial approximation is that the high-energy unoccupied states can be represented by plane waves. This is usually a good approximation since at high energies, the kinetic energy term $-\Delta/2$ will be the dominant part of the Hamiltonian so that the orbitals are well approximated by plane waves [25,26]. Our derivation first concentrates on the total correlation energies in the second-order approximation (direct term in MP2) since the essential results are easy to grasp. This derivation closely parallels the derivation for the jellium electron gas in Ref. [22]. Then, we discuss the shortcomings of the PAW method, and finally we move on to the self-energy.

A. Correlation energy in second-order perturbation theory (MP2)

We start by deriving the convergence behavior of the direct MP2 (dMP2) energy. The result, however, also applies to the convergence of the RPA energy, which reduces to the direct MP2 energy in the limit of large wave vectors. Specifically, for large wave vectors $\mathbf{G}$, the independent particle polarizability χ_0 becomes small and the RPA correlation energy (here v is the Coulomb operator)

$$E^{\mathrm{RPA}} = \mathrm{Tr}\left[\tfrac{1}{2}(\chi_0 v)^2 + \tfrac{1}{3}(\chi_0 v)^3 + \cdots\right] \approx \mathrm{Tr}\left[\tfrac{1}{2}(\chi_0 v)^2\right]$$

reduces to the leading-order dMP2 term. The relation between the dMP2 and RPA energy has also been discussed in Ref. [30]. We also note that even though the total dMP2 energy diverges for systems without a gap in the thermodynamic limit, the contribution of large wave vectors $\mathbf{G}$ to the energy is finite. Therefore, our derivation holds also for systems with vanishing gap.

The direct contribution to the MP2 correlation energy for a non-spin-polarized system can be written as (generalization to the spin-polarized case is straightforward)

$$E^{\mathrm{dMP2}} = -2\sum_{ij}^{\mathrm{occ}}\sum_{ab}^{\mathrm{unocc}}\sum_{\mathbf{G}\mathbf{G}'}\langle i|\mathbf{G}|a\rangle\langle a|-\mathbf{G}'|i\rangle\frac{4\pi}{\Omega\mathbf{G}'^2} \times \langle j|\mathbf{G}'|b\rangle\langle b|-\mathbf{G}|j\rangle\frac{4\pi}{\Omega\mathbf{G}^2}\frac{1}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j}. \quad (1)$$

Here and throughout the paper we use atomic units. The spin coordinates have been summed over, so that the summation over orbitals is over spatial orbitals only. The integers i,j and a,b are indices for occupied and unoccupied orbitals, respectively, and $\varepsilon_i, \varepsilon_j, \ldots$ denote the corresponding eigenenergies. The vectors $\mathbf{G}$ and $\mathbf{G}'$ are three-dimensional reciprocal lattice vectors of a cell with volume Ω. This auxiliary basis set is used to represent density related quantities, and in practical calculations this basis set is restricted by imposing a plane-wave cutoff E^{χ}_{cut} [ENCUTGW in the Vienna *ab initio* simulation package (VASP)] to improve the efficiency of the calculations. The quantity

$$\langle b|-\mathbf{G}|j\rangle = \int_{\Omega}\psi_b^*(\mathbf{r})\psi_j(\mathbf{r})\exp(-i\mathbf{G}\cdot\mathbf{r})\,d\mathbf{r} \quad (2)$$

is the overlap density of the orbitals $\langle\mathbf{r}|b\rangle = \psi_b(\mathbf{r})$ and $\langle\mathbf{r}|j\rangle = \psi_j(\mathbf{r})$ represented in this auxiliary plane-wave basis.

We now determine the contribution $E^{\mathrm{dMP2}}(\mathbf{G}',\mathbf{G})$ to the direct MP2 energy from auxiliary basis states $\mathbf{G}$ and $\mathbf{G}'$ that have no overlap with occupied orbitals. Specifically, we want to calculate the approximate correlation energy from plane waves $\mathbf{G}$ and $\mathbf{G}'$ that observe

$$\langle i|-\mathbf{G}|j\rangle \approx 0 \quad (3)$$

for all occupied orbitals i,j. Later, for QP calculations, we will extend this condition to a few unoccupied states of interest. These plane-wave components are obviously not relevant for the description of the density originating from occupied Kohn-Sham orbitals, but as we show below these plane waves still contribute to the many-body correlation energy and self-energy.

In a first step, we allow the indices a and b to run over all states instead of the unoccupied states in Eq. (1) [Eq. (3) allows for this "simplification"]:

$$E(\mathbf{G}',\mathbf{G}) = -2\sum_{ij}^{\text{occ}}\sum_{ab}^{\text{all}}\langle i|\mathbf{G}|a\rangle\langle a|-\mathbf{G}'|i\rangle\frac{4\pi}{\Omega\mathbf{G}'^2} \times\langle j|\mathbf{G}'|b\rangle\langle b|-\mathbf{G}|j\rangle\frac{4\pi}{\Omega\mathbf{G}^2}\frac{1}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j}. \tag{4}$$

For brevity, the superscript dMP2 will be dropped from now on. Second, we assume that high-energy unoccupied states a are essentially plane waves of the form $\psi_a(\mathbf{r}) = \frac{1}{\sqrt{\Omega}}\exp(i\mathbf{G}_a\cdot\mathbf{r})$ with the eigenenergies $\varepsilon_a = \mathbf{G}_a^2/2$. In this case, the quantity $\langle i|\mathbf{G}|a\rangle\langle a|-\mathbf{G}'|i\rangle$ is only sizable if $\mathbf{G}_a \approx -\mathbf{G}$ and $\mathbf{G}_a \approx -\mathbf{G}'$ and $\langle j|\mathbf{G}'|b\rangle\langle b|-\mathbf{G}|j\rangle$ will be non-negligible if $\mathbf{G}_b \approx -\mathbf{G}'$ and $\mathbf{G}_b \approx -\mathbf{G}$. This follows from the assumption that the occupied states have predominantly components at small wave vectors. Therefore, we can approximate $\varepsilon_a+\varepsilon_b \approx (\mathbf{G}^2+\mathbf{G}'^2)/2$. Moreover, to simplify the formula, we neglect the dependence on the energies of occupied states compared to those of the unoccupied states, so that

$$\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j \approx (\mathbf{G}^2+\mathbf{G}'^2)/2\,. \tag{5}$$

This approximation is also correct for the important diagonal components $\mathbf{G}=\mathbf{G}'$. Once rid of the eigenvalues ε, the resolution of identity $\sum_a |a\rangle\langle a| = 1$ can be used leading to a much simpler expression

$$E(\mathbf{G},\mathbf{G}') = -2\sum_{ij}^{\text{occ}}\langle i|\mathbf{G}-\mathbf{G}'|i\rangle\frac{4\pi}{\Omega\mathbf{G}'^2} \times\langle j|\mathbf{G}'-\mathbf{G}|j\rangle\frac{4\pi}{\Omega\mathbf{G}^2}\frac{2}{\mathbf{G}^2+\mathbf{G}'^2}. \tag{6}$$

We can also write this as

$$E(\mathbf{G},\mathbf{G}') = -2\sum_{ij}^{\text{occ}}\rho_i(\mathbf{G}-\mathbf{G}')\frac{4\pi}{\mathbf{G}'^2}\rho_j(\mathbf{G}'-\mathbf{G})\frac{4\pi}{\mathbf{G}^2}\frac{2}{\mathbf{G}^2+\mathbf{G}'^2}, \tag{7}$$

where the Fourier transformed density of the orbital i is defined as

$$\rho_i(\mathbf{G}) = \frac{1}{\Omega}\int_\Omega \psi_i^*(\mathbf{r})\psi_i(\mathbf{r})\exp(i\mathbf{G}\cdot\mathbf{r})\,d\mathbf{r}. \tag{8}$$

Since the total density can be calculated as $\rho(\mathbf{G}) = 2\sum_i \rho_i(\mathbf{G})$, we obtain the first central result of this work:

$$E(\mathbf{G},\mathbf{G}') = -\frac{1}{2}|\rho(\mathbf{G}-\mathbf{G}')|^2\frac{4\pi}{\mathbf{G}'^2}\frac{4\pi}{\mathbf{G}^2}\frac{2}{\mathbf{G}^2+\mathbf{G}'^2}. \tag{9}$$

This equation approximates the contribution to the correlation energy from two auxiliary plane-wave basis-set functions with wave vectors $\mathbf{G}$ and $\mathbf{G}'$. As before, "auxiliary" here refers to the basis set used for the overlap charge density. The equation can be evaluated easily and at modest cost in any plane-wave code and can be used to estimate the incompleteness error.

In practical calculations, the auxiliary basis set is restricted by a cutoff energy E_{cut}^{χ}, where basis functions with energies $|\mathbf{G}|^2/2 > E_{\text{cut}}^{\chi}$ are omitted. The resulting error can be estimated by summing up all the contributions $E(\mathbf{G},\mathbf{G}')$ for $\mathbf{G}$ and $\mathbf{G}'$ above the cutoff $G_{\text{cut}} = \sqrt{2E_{\text{cut}}^{\chi}}$:

$$\Delta E = \sum_{|\mathbf{G}|>G_{\text{cut}}}\sum_{|\mathbf{G}'|>G_{\text{cut}}} E(\mathbf{G},\mathbf{G}')\,. \tag{10}$$

We perform this summation by changing from the variables $\mathbf{G}$ and $\mathbf{G}'$ to their average $\mathbf{G}'' = (\mathbf{G}+\mathbf{G}')/2$ and their difference $\mathbf{g} = \mathbf{G}-\mathbf{G}'$. The relevant values of $\mathbf{g}$ are limited in the reciprocal space by the spread of the density and invoking Eq. (3) one can make the assumption that $|\mathbf{g}| \ll |\mathbf{G}''|$. Performing this substitution and neglecting $\mathbf{g}$ in the parts originating from the Coulomb interaction and the eigenenergies leads to

$$\Delta E \approx -\frac{1}{2}\sum_{\mathbf{g}}|\rho(\mathbf{g})|^2\sum_{|\mathbf{G}''|>G_{\text{cut}}}\frac{(4\pi)^2}{\mathbf{G}''^6}. \tag{11}$$

Summing over all plane waves with $|\mathbf{G}''| > G_{\text{cut}}$ is trivial and yields

$$\Delta E = \frac{4}{3}\frac{1}{G_{\text{cut}}^3}\sum_{\mathbf{g}}|\rho(\mathbf{g})|^2 \approx \frac{2}{9\pi^2}\frac{\Omega^2}{N_{\text{pw}}}\sum_{\mathbf{g}}|\rho(\mathbf{g})|^2, \tag{12}$$

where the last equation involves the number of plane waves N_{pw} in the cutoff sphere with radius G_{cut}. The formula is further simplified in the special case of the jellium electron gas (UEG), where $\rho(\mathbf{G}-\mathbf{G}') = \delta(\mathbf{G}-\mathbf{G}')\rho(0)$. In this case,

$$\Delta E^{\text{UEG}} \approx \frac{2}{9\pi^2}\frac{\rho^2\Omega^2}{N_{\text{pw}}} = \frac{2}{9\pi^2}\frac{N_{\text{el}}^2}{N_{\text{pw}}}, \tag{13}$$

where N_{el} is the total number of electrons in the cell (including spin).

The final two equations show that the leading-order error of the correlation energy decays asymptotically with the inverse of the number of plane waves N_{pw} in the auxiliary basis set. The rate is determined by the Fourier components of the ground-state density. It is important to note that this slow convergence of the MP2 energies is well established for quantum chemistry basis sets (Gaussian-type orbitals) with exactly the same one over basis-set size convergence [20]. Furthermore, it has been shown that the slow convergence is related to the description of the two-electron cusp condition [16–19,23]. This condition leads to a kink in the many-electron wave function $\Psi(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\ldots)$: when any two coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$ coincide, the many-electron wave function exhibits a discontinuity in the slope, which leads exactly to the basis-set convergence determined above.

The slow convergence originates from the fact that the density enters at wave vectors $\mathbf{g} = \mathbf{G}-\mathbf{G}'$ in the correlation energy since large wave vectors of the auxiliary basis set are folded back to wave vectors around $\mathbf{g} = 0$ by "interaction" with high-lying plane-wave-like orbitals. In other words, even auxiliary basis states not relevant for the description of ground-state densities [compare Eq. (3)] will contribute to the many-electron correlation energy. Since the correlation energy correction depends on the charge density distribution, one can also not assume that this contribution will cause a trivial shift of the total correlation energy. Obviously, when the density changes, the convergence rate will be altered as well.

Looking back at the derivation, a second related issue becomes obvious. We have performed the derivation assuming

a finite auxiliary basis set for the densities and a complete set of unoccupied orbitals. However, one can also perform the derivation by assuming an infinite cutoff on the auxiliary density basis set but restricting the number of unoccupied states. The final error expression is identical: if the basis set for the orbitals is truncated at $G_{\rm cut}^{\rm pw}$, or if the basis for the overlap densities is truncated at $G_{\rm cut}^{\chi}$, an error proportional to $1/G_{\rm cut}^3$ is introduced with $G_{\rm cut} = \min(G_{\rm cut}^{\rm pw}, G_{\rm cut}^{\chi})$. It is therefore important to increase the basis set for the one-electron orbitals, the number of unoccupied orbitals, and the auxiliary basis set for the overlap charge densities at a similar rate.

We also note that Eq. (12) gives the leading order of the error as $1/N_{\rm pw}$. In practical RPA or MP2 calculations, this error is approximately corrected for using extrapolation [21]. As discussed by Gulans, the next order correction falls off like $N_{\rm pw}^{-5/3}$ [23]. For UEG, this order could be recovered if the energies of the occupied states were taken into account. The next order follows $N_{\rm pw}^{-7/3}$ and such a dependence of the MP2 energy was found by Grüneis *et al.* after applying F12 corrections [31]. Presumably, the F12 corrections treat the error of the order $N_{\rm pw}^{-5/3}$ exactly by including both the energies and basis-set representation of the occupied states. In principle, one can improve the accuracy of the present scheme by keeping the energies of the occupied states in the final equation; we, however, leave the explicit derivation for future work.

B. Shortcomings of the PAW method: Completeness issues for MP2

In the PAW method, the all-electron orbitals $|i\rangle$ are related to the pseudopart $|\tilde{i}\rangle$ by a linear relation

$$|i\rangle = |\tilde{i}\rangle + \underbrace{\sum_{\alpha}(|\alpha\rangle - |\tilde{\alpha}\rangle)\langle p_\alpha|\tilde{i}\rangle}_{|i^{\rm aug}\rangle}, \tag{14}$$

where $\langle p_\alpha|$ is an atom-centered projector onto an atomic orbital α with corresponding all-electron and pseudopartial waves $\langle \mathbf{r}|\alpha\rangle = \phi_\alpha(\mathbf{r})$ and $\langle \mathbf{r}|\tilde{\alpha}\rangle = \tilde{\phi}_\alpha(\mathbf{r})$, respectively. One-electron properties such as the overlap charge density $\langle a|\mathbf{r}\rangle\langle \mathbf{r}|i\rangle = \psi_a^*(\mathbf{r})\psi_i(\mathbf{r})$ would, in principle, involve four terms

$$\begin{aligned}\langle a|\mathbf{r}\rangle\langle \mathbf{r}|i\rangle = {} & \langle \tilde{a}|\mathbf{r}\rangle\langle \mathbf{r}|\tilde{i}\rangle + \langle a^{\rm aug}|\mathbf{r}\rangle\langle \mathbf{r}|i^{\rm aug}\rangle \\ & + \langle a^{\rm aug}|\mathbf{r}\rangle\langle \mathbf{r}|\tilde{i}\rangle + \langle \tilde{a}|\mathbf{r}\rangle\langle \mathbf{r}|i^{\rm aug}\rangle.\end{aligned}$$

The crucial assumption in the PAW method is that the set of partial waves is complete. As discussed in Ref. [32], this allows us to drop the final two "mixed" terms and the overlap charge density simplifies to

$$\begin{aligned}\langle a|\mathbf{r}\rangle\langle \mathbf{r}|i\rangle \approx {} & \langle \tilde{a}|\mathbf{r}\rangle\langle \mathbf{r}|\tilde{i}\rangle \\ & + \sum_{\alpha\beta}\langle \tilde{a}|p_\alpha\rangle(\langle \alpha|\mathbf{r}\rangle\langle \mathbf{r}|\beta\rangle - \langle \tilde{\alpha}|\mathbf{r}\rangle\langle \mathbf{r}|\tilde{\beta}\rangle)\langle p_\beta|\tilde{i}\rangle.\end{aligned} \tag{15}$$

This simplification is crucial for the efficiency of the PAW method. It is easy to achieve completeness in the space of occupied orbitals, however, it is, in practice, impossible to achieve completeness for orbitals at very high energies. This is the source of an important error.

Returning to Eq. (4), the problem of the PAW method is easy to understand. If a is a plane wave with large momentum $\psi_a(\mathbf{r}) = \frac{1}{\sqrt{\Omega}}\exp(i\mathbf{G}_a\cdot\mathbf{r})$, the projection $\langle a|p_\alpha\rangle$ becomes very small since the projectors $|p_\alpha\rangle$ span only the occupied orbitals and are limited in Fourier space ($\langle \mathbf{G}_a|p_\alpha\rangle \approx 0$). The completeness relation is then violated, and, in principle, the mixed terms discussed above should be reintroduced to alleviate the issue. As long as the mixed terms are not included, the augmentation terms are effectively zero, so that $\langle i|\mathbf{G}|a\rangle\langle a| - \mathbf{G}'|i\rangle$ becomes identical to $\langle \tilde{i}|\mathbf{G}|\tilde{a}\rangle\langle \tilde{a}| - \mathbf{G}'|\tilde{i}\rangle$. In this case, the correlation energy in the limit of large wave vectors $\mathbf{G}$ and $\mathbf{G}'$ becomes

$$E(\mathbf{G},\mathbf{G}') = -|\tilde{\rho}(\mathbf{G}-\mathbf{G}')|^2\frac{4\pi}{\mathbf{G}'^2}\frac{4\pi}{\mathbf{G}^2}\frac{1}{\mathbf{G}^2+\mathbf{G}'^2}, \tag{16}$$

where $\tilde{\rho}(\mathbf{G}-\mathbf{G}')$ is the Fourier transformed ground-state density neglecting any charge augmentation. While the energy still converges as $1/N_{\rm pw}$, the convergence rate, as well as the converged correlation energy, possess an error. The error can be quite sizable, especially for d elements where the norm violation can be considerable: up to 80% for late $3d$ elements such as Zn, Ga, or Cu. The error will most significantly appear in the total value of the correlation energy. Its importance for relative energies should be smaller, although it is not negligible for $3d$ elements, as discussed in more detail in Appendix A.

C. Second-order self-energy contribution at large G vectors

We now derive the limiting behavior of the quasiparticle corrections within the GW approximation to the electron self-energy. We follow the same route as for the dMP2 correlation energy by approximating the high-energy unoccupied states by plane waves. Again, we assume a complete set of unoccupied states and a finite auxiliary basis set for the overlap charge density and related response function. However, similar results are obtained when a finite set of unoccupied states and a complete auxiliary basis set is used.

The usual definition for the self-energy in the GW approximation is

$$\Sigma(\omega) = \frac{i}{2\pi}\int d\omega'\exp(-i\delta\omega')G(\omega-\omega')W(\omega'), \tag{17}$$

with the Green's function

$$G(\mathbf{G},\mathbf{G}',\omega) = \sum_n^{\rm all}\frac{\langle \mathbf{G}|n\rangle\langle n|\mathbf{G}'\rangle}{\omega-\varepsilon_n \mp i\eta}, \tag{18}$$

where δ and η are positive infinitesimals, and the minus sign applies to occupied states and the plus sign to the unoccupied ones. As before, we use the second-order approximation, so that the screened interaction is $W = v + v\chi_0 v$, where the first term is the exchange contribution and the second one is the correlation part $W_{\rm c} = v\chi_0 v$. In the following, we consider only the correlation part, which can be written as (from now on we will drop the functional parameters $\mathbf{G},\mathbf{G}'$, and ω)

$$\begin{aligned}W_{\rm c} = \frac{2}{\Omega}\frac{4\pi}{\mathbf{G}^2}\frac{4\pi}{\mathbf{G}'^2}\sum_i^{\rm occ}\sum_a^{\rm unocc}\Bigg(&\frac{\langle i|-\mathbf{G}|a\rangle\langle a|\mathbf{G}'|i\rangle}{\omega+\varepsilon_i-\varepsilon_a+i\eta} \\ &- \frac{\langle a|-\mathbf{G}|i\rangle\langle i|\mathbf{G}'|a\rangle}{\omega+\varepsilon_a-\varepsilon_i-i\eta}\Bigg).\end{aligned} \tag{19}$$

Inserting the expressions for G and W_c into Eq. (17), one obtains

$$\Sigma = \frac{i}{2\pi}\frac{2}{\Omega^2}\int d\omega' \exp(-i\delta\omega') \times \sum_n \frac{\langle \mathbf{G}|n\rangle\langle n|\mathbf{G}'\rangle}{\omega-\omega'-\varepsilon_n \mp i\eta}\frac{4\pi}{\mathbf{G}^2}\frac{4\pi}{\mathbf{G}'^2} \times \sum_{i,a}\left(\frac{\langle i|-\mathbf{G}|a\rangle\langle a|\mathbf{G}'|i\rangle}{\omega'+\varepsilon_i-\varepsilon_a+i\eta} - \frac{\langle a|-\mathbf{G}|i\rangle\langle i|\mathbf{G}'|a\rangle}{\omega'+\varepsilon_a-\varepsilon_i-i\eta}\right). \tag{20}$$

The self-energy integral can now be calculated analytically via contour integration considering the poles of the integrand. As we choose to close the contour in the lower half, only the poles with a negative imaginary part will contribute. There are two distinct contributions. The first originates from the poles in the Green's function at the energies of the occupied states

$$\omega' = \omega - \varepsilon_n - i\eta \tag{21}$$

and gives the screened exchange (SEX) contribution (we drop η from the expressions after integration)

$$\Sigma^{\mathrm{SEX}(2)} = -\frac{2}{\Omega^2}\frac{4\pi}{\mathbf{G}^2}\frac{4\pi}{\mathbf{G}'^2}\sum_n^{\mathrm{occ}}\langle \mathbf{G}|n\rangle\langle n|\mathbf{G}'\rangle \times \sum_{i,a}\left(\frac{\langle i|-\mathbf{G}|a\rangle\langle a|\mathbf{G}'|i\rangle}{\omega-\varepsilon_n+\varepsilon_i-\varepsilon_a} - \frac{\langle a|-\mathbf{G}|i\rangle\langle i|\mathbf{G}'|a\rangle}{\omega-\varepsilon_n+\varepsilon_a-\varepsilon_i}\right). \tag{22}$$

The second contribution comes from the poles of W located at the excitation energies of the noninteracting response function

$$\omega' = -(\varepsilon_i - \varepsilon_a + i\eta). \tag{23}$$

This gives the Coulomb hole (COH) contribution to the self-energy

$$\Sigma^{\mathrm{COH}(2)} = \frac{2}{\Omega^2}\frac{4\pi}{\mathbf{G}^2}\frac{4\pi}{\mathbf{G}'^2}\sum_n^{\mathrm{all}}\langle \mathbf{G}|n\rangle\langle n|\mathbf{G}'\rangle \times \sum_{i,a}\frac{\langle i|-\mathbf{G}|a\rangle\langle a|\mathbf{G}'|i\rangle}{\omega+\varepsilon_i-\varepsilon_a-\varepsilon_n}. \tag{24}$$

These expressions are equivalent to the ones derived by Grüneis *et al.* for MP2 quasiparticle energies [33]. Note, however, that we have kept the first term in the SEX contribution and the part where $n \in$ occ in the COH contribution. These two terms are identical except for an opposite sign and would subtract out upon adding the SEX and COH contributions.

We now consider the contribution of the high momentum wave vectors $\mathbf{G}$ and $\mathbf{G}'$. When the SEX term is evaluated for an orbital m with eigenenergy ε_m, $\langle m|\Sigma^{\mathrm{SEX}(2)}|m\rangle$, the first term on the right-hand side of Eq. (22) contains

$$\sum_n^{\mathrm{occ}}\langle m|-\mathbf{G}|n\rangle\langle n|\mathbf{G}'|m\rangle.$$

Here, n corresponds to an occupied state, and m corresponds to an occupied state or one just above the Fermi level. According to our assumption (3), at large $\mathbf{G}$ and $\mathbf{G}'$ the corresponding contribution is zero. The SEX term, therefore, does converge rapidly with respect to the basis-set size.

The COH contribution, however, involves a summation over all states, and the matrix element of self-energy reads as

$$\langle m|\Sigma^{\mathrm{COH}(2)}(\varepsilon_m)|m\rangle = \frac{2}{\Omega^2}\frac{4\pi}{\mathbf{G}^2}\frac{4\pi}{\mathbf{G}'^2}\sum_n^{\mathrm{all}}\langle m|-\mathbf{G}|n\rangle\langle n|\mathbf{G}'|m\rangle \times \sum_{i,a}\frac{\langle i|-\mathbf{G}|a\rangle\langle a|\mathbf{G}'|i\rangle}{\varepsilon_m+\varepsilon_i-\varepsilon_a-\varepsilon_n}. \tag{25}$$

For large $\mathbf{G}$ vectors, we can modify this expression in the same way as in the MP2 total energy case, so that for orbital m, the approximate contribution to the QP correction can be written as

$$\Delta\varepsilon_m = -\frac{1}{2}\rho_m(\mathbf{G}-\mathbf{G}')\frac{4\pi}{\mathbf{G}'^2}\rho(\mathbf{G}'-\mathbf{G})\frac{4\pi}{\mathbf{G}^2}\frac{2}{\mathbf{G}^2+\mathbf{G}'^2}, \tag{26}$$

where $\rho_m(\mathbf{G}-\mathbf{G}')$ is the density of orbital m in reciprocal space, including a spin factor of 2. This is the second central result, and the formal equivalence to the MP2 total energy contribution is obvious. In fact, one could have derived this equation by applying Koopmans' theorem to the total energy expression (9).

After performing the summation over the omitted $\mathbf{G}$ vectors, the correction for a finite auxiliary basis set can be written as

$$\Delta\varepsilon_m^{\mathrm{tot}} = -\frac{2}{9\pi^2}\frac{\Omega^2}{N_{\mathrm{pw}}}\sum_{\mathbf{g}}\rho_m(\mathbf{g})\rho(-\mathbf{g}). \tag{27}$$

We note that we have again assumed that the density in reciprocal space has nonzero components only for $|\mathbf{g}| \ll G_{\mathrm{cut}}$. The validity of this approximation will, in principle, differ for each individual state. Indeed, states localized in real space are more spread in reciprocal space and for such states the terms beyond the leading order might be important, as we will show later in Sec. IV. In the case of the jellium electron gas, the equation simplifies to

$$\Delta\varepsilon_m^{\mathrm{tot,UEG}} = -\frac{4}{9\pi^2}\frac{N_{\mathrm{el}}}{N_{\mathrm{pw}}}. \tag{28}$$

Here and above, N_{pw} is the number of auxiliary basis-set functions used to describe overlap densities. A similar result is obtained for the case when the number of unoccupied states is limited and the auxiliary basis set is large, N_{pw} is then the number of unoccupied states.

The implications of Eq. (27) are analogous to the MP2 case, and here we only briefly reiterate the main points. The quasiparticle energies will converge with the inverse of N_{pw} and, in fact, such convergence behavior was observed in GW calculations before [13,28]. This leading order, as well as higher-order terms, were derived by Gulans for the jellium electron gas and by Schindlmayr for a model system of two electrons on a sphere [23,34]. As also discussed by Gulans and Schindlmayr, when calculating quasiparticle gaps, the leading order $1/N_{\mathrm{pw}}$ might accidentally subtract out and the calculated gap might exhibit a faster convergence. To observe the $1/N_{\mathrm{pw}}$ convergence and to converge to the correct value, both the number of unoccupied bands, the corresponding orbital basis set, and the size of the response function basis

set need to be increased. We also note that, when performing GW calculations, it is still quite common to fix the size of the response function matrix and converge only with respect to the unoccupied states. This, however, leads to a false convergence, as demonstrated by Shih *et al.* and Friedrich *et al.* [14,28]. Finally, we note that the derivation can be also used to derive the asymptotic convergence of the static COHSEX approximation, which is given in Appendix B. The derivation shows that static COHSEX overestimates the large **G** contributions by a factor of 2.

D. Shortcomings of the PAW method: Completeness issues for GW

There is very little to add compared to the previous discussion on the correlation energy. For large wave vectors **G** and **G**′, the exact contribution to the QP correction for the state m is given by Eq. (26). However, in the PAW method, we will observe only the pseudized contribution

$$\Delta\tilde{\varepsilon}_m = -\frac{1}{2}\tilde{\rho}_m(\mathbf{G}-\mathbf{G}')\frac{4\pi}{\mathbf{G}'^2}\tilde{\rho}(\mathbf{G}'-\mathbf{G})\frac{4\pi}{\mathbf{G}^2}\frac{2}{\mathbf{G}^2+\mathbf{G}'^2}. \quad (29)$$

Contributions from the augmentation charges are again missing since high-energy states are not picked up by the projectors. In principle, it is possible to correct for this error *a posteriori* by calculating the exact term (26) and subtracting the one effectively used in the PAW approximation. In the present case, this correction has not been applied.

Instead, Eq. (26) suggests a simple alternative procedure. If the pseudized density $\tilde{\rho}(\mathbf{g})$ has the correct norm, both $\rho(\mathbf{g})$ and $\tilde{\rho}(\mathbf{g})$ will exactly agree at $\mathbf{g}=0$, and approximately at larger wave vectors. Hence, norm conservation should help to obtain a reasonable approximation for the self-energy. In other words, if the norm of the pseudized partial waves and all-electron partial waves are identical, we expect smaller error than if the norm is not conserved. This conjecture will be validated by the tests shown in Sec. IV A.

III. COMPUTATIONAL DETAILS

For the GW calculations presented here, all PAW potentials apply approximately norm-conserving partial waves (a small norm violation of up to 10%–20% was allowed for some cases). As discussed in the previous section, such potentials should yield an accurate description of the correlation energy and self-energy even at very large wave vectors, corresponding to rapid variations in space. To construct these approximately norm-conserving partial waves, the following strategies were adopted. (i) First, non-norm-conserving pseudopartial waves were used, and it was attempted to set the core radius such that the pseudopartial wave $\tilde{\phi}_\alpha(\mathbf{r})$ possesses the same norm as the all-electron partial wave $\phi_\alpha(\mathbf{r})$ within a sphere around the atom [see Eq. (14)]. (ii) If this yielded too hard potentials or too large core radii, a norm-conserving pseudopartial wave $\tilde{\phi}_\alpha(\mathbf{r})$ was chosen. Table I reports the final core radii for all considered elements. Generally small core radii (below 1.6 a.u.) indicate that option (i) was chosen, whereas larger core radii indicate that option (ii) was used. The same potentials have been also applied in our recent work using vertex corrected GW calculations [35].

TABLE I. PAW potentials used in this work. The columns $\mathrm{r}_{s,p,d,f}$ specify the core radii for each angular quantum number in a.u. The "default" plane-wave cutoff energy $E_{\mathrm{cut}}^{\mathrm{pw}}$ for the orbitals is specified in eV. Column "local" specifies the chosen local potential. This is usually the all-electron potential replaced by a soft approximation inside the specified core radius or the f potential.

	r_s	r_p	r_d	r_f	local	$E_{\mathrm{cut}}^{\mathrm{pw}}$
B	1.10	1.10	1.10		d	700
C	1.00	1.10	1.10		0.8	740
N	0.90	1.10	1.10		0.9	755
O	1.00	1.10	1.10		0.9	765
Mg	1.15	1.65	1.65		1.2	821
Al	1.75	2.00	1.80	2.00	1.4	571
Si	1.70	1.95	1.70	2.00	1.4	609
P	1.70	1.95	1.70	2.00	1.4	616
S	1.23	1.35	1.70	1.80	f	486
Zn	1.27	1.70	1.90	1.90	1.2	802
Ga	1.23	1.70	1.90	1.90	1.3	801
Ge	1.18	1.70	1.90	1.90	1.3	807
As	1.14	1.54	2.20	1.90	1.3	613
Se	1.10	1.40	2.30	1.90	1.3	571
Cd	1.70	1.90	2.10	1.90	1.3	657
In	1.66	2.10	2.30	1.90	1.3	582
Sb	1.56	1.90	2.30	1.90	1.2	561
Te	1.51	1.82	2.40	1.90	1.4	617

The most shallow core states were treated as valence for all elements, except for boron, carbon, nitrogen, oxygen, fluorine, and sulfur. In these cases, the core states are well below 10 Ry, and they are very well localized. Unfreezing them yielded very small changes in test calculations. For 3d elements, as well as Ga, Ge, As, and Se, the 3s, 3p, and 3d states were treated as valence and for 4d elements, as well as In, Sb, and Te, the 4s, 4p, and 4d states were treated as valence. Usually, three partial waves were used for each angular quantum number l. One partial wave was placed at the uppermost core state, one at the binding energy of the valence state in the atom, and a third projector about 20 Ry above the vacuum level. This guarantees excellent high-energy scattering properties up to about 30 Ry. Beyond 30 Ry, however, no projectors and partial waves are available so that plane waves beyond 30 Ry are not properly picked up by the projectors p_α. These are the energies where norm conservation should help.

The plane-wave cutoff was chosen to be the maximum $E_{\mathrm{cut}}^{\mathrm{pw}}$ of all elements in the considered material. To determine basis-set converged GW values, the plane-wave cutoff $E_{\mathrm{cut}}^{\mathrm{pw}}$ was systematically increased by a factor of 1.25 and 1.587, corresponding to an increase of the number of plane waves N_{pw} by a factor 1.4 and 2, respectively. The results were then extrapolated to the infinite-basis-set limit, assuming that the QP energies converge like $(1/E_{\mathrm{cut}}^{\mathrm{pw}})^{3/2} = 1/N_{\mathrm{pw}}$. For sufficiently large plane-wave cutoffs, this behavior was strictly observed for all materials. However, as shown below for the example of ZnO, this behavior sets in only at very high plane-wave cutoffs for 3d states. This implies that the 3d states might have an error of about 100 meV (errors are smaller for

$4d$ and also for the somewhat problematic $2p$ states). The cutoff for the response function was set to half the plane-wave cutoff for the orbitals. Correct asymptotic convergence was only observed if both the plane-wave cutoffs for the orbitals and the response function were increased simultaneously.

The convergence behavior of the QP energies with respect to the basis-set size depends only weakly on the k-point set used. Therefore, to reduce the overall computational cost, we used $3 \times 3 \times 3$ k points to obtain a basis-set size correction as the difference of the basis-set converged values and data obtained with a smaller basis-set size. We then added this correction to the results calculated with a smaller basis-set size using $6 \times 6 \times 6$ k points, e.g.,

$$E_\infty(6\times6\times6) \approx E_{\rm red}(6\times6\times6) - E_{\rm red}(3\times3\times3) + E_\infty(3\times3\times3).$$

Here, E_∞ is a QP energy extrapolated to the infinite basis set, and $E_{\rm red}$ is the QP energy calculated with a reduced basis set. The extrapolation is based on the observation that the contributions from large wave vectors $\mathbf{G}$ converge rapidly with k points and are accurately modeled using coarse k-point grids. We, however, note that the calculations with dense k-point grids are performed at what one would usually consider to be well-converged GW calculations (half of the orbitals spanned by the orbital basis set, and cutoff of 240-eV for the response function basis set).

The GW calculations presented here have been performed using the fully frequency-dependent version as described in Refs. [36,37]. The QP energies were determined by linearization of the self-energy around the original LDA eigenvalue $E^0_{n\mathbf{k}}$:

$$E^{\rm QP}_{n\mathbf{k}} = E^0_{n\mathbf{k}} + Z_{n\mathbf{k}}{\rm Re}\big[\langle\psi_{n\mathbf{k}}|T + V_{n-e} + V_H + \Sigma\big(E^0_{n\mathbf{k}}\big)|\psi_{n\mathbf{k}}\rangle - E^0_{n\mathbf{k}}\big], \quad (30)$$

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

$$Z_{n\mathbf{k}} = \left(1 - {\rm Re}\langle\psi_{n\mathbf{k}}|\frac{\partial}{\partial\omega}\Sigma(\omega)\Big|_{E^0_{n\mathbf{k}}}|\psi_{n\mathbf{k}}\rangle\right)^{-1}. \quad (31)$$

For the GW_0 case, the independent particle screening χ_0 that is used to determine $W_0 = v + v\chi_0 v + \ldots$ is kept fixed at the level of DFT, but the eigenvalues in the Green's function are updated until convergence is reached. In practice, good self-consistency is reached after four iterations. Note that we also update the occupancies and Hartree potential in the course of the iterations, which changes the results slightly for those materials where the band gap is originally inverted in the DFT calculations (see following).

The final technical detail concerns the treatment of the augmentation term on the right-hand side of Eq. (15). It is approximated by pseudized augmentation functions leading to [38]

$$\langle a|\mathbf{r}\rangle\langle\mathbf{r}|i\rangle = \psi^*_a(\mathbf{r})\psi_i(\mathbf{r}) \approx \tilde\psi^*_a(\mathbf{r})\tilde\psi_i(\mathbf{r}) + \sum_{\alpha\beta}\langle\tilde a|p_\alpha\rangle\tilde Q_{\alpha\beta}(\mathbf{r})\langle p_\beta|\tilde i\rangle. \quad (32)$$

The pseudized augmentation density $\tilde Q_{\alpha\beta}(\mathbf{r})$ is evaluated in real space and constructed to closely reproduce the exact all-electron quantity. This is achieved by imposing two constraints: (i) $\tilde Q_{\alpha\beta}(\mathbf{r})$ must reproduce the multipoles of the exact augmentation charge density $Q_{\alpha\beta}(\mathbf{r}) = \phi^*_\alpha(\mathbf{r})\phi_\beta(\mathbf{r}) - \tilde\phi^*_\alpha(\mathbf{r})\tilde\phi_\beta(\mathbf{r})$, and (ii) the Fourier transform of $Q_{\alpha\beta}$ and $\tilde Q_{\alpha\beta}$ have to match up to a chosen plane-wave component. The second condition (ii) is not usually used in ground-state calculations, but employed in VASP in QP or any other correlated calculation.

IV. RESULTS

A. The example of ZnO, GaAs, and AlAs

As an illustrative example, we consider ZnO in the zinc-blende structure. In this section, the k-point set is set to $4 \times 4 \times 4$ points to simplify the computationally rather expensive convergence tests. Furthermore, the plane-wave cutoff for the response function was set to a kinetic energy of 1000 eV, corresponding to a basis set of about 1750 plane waves for the response function. We now focus on the convergence of the results with respect to the number of orbitals included in the calculation of the Green's function G and the response function χ_0. Figure 1 shows the QP corrections of the Kohn-Sham eigenvalues versus the inverse of the total number of bands.

We first concentrate on the results for the norm-conserving (NC) and ultrasoft (US) PAW potentials with a core radius of 1.7 a.u. It is clear that the results are indistinguishable for up to 150–200 bands, but beyond that the QP correction of the $3d$ orbitals deviates significantly for both calculations, with a sharp drop being visible for the NC PAW potential around 200 orbitals. The final converged d-band correction differs by almost 1 eV between the potentials, although both potentials are practically indistinguishable for ground-state calculations and GW calculations with few hundreds of orbitals. The error is related to the lack of projectors at high energies, as confirmed by inspection of $\langle p_\alpha|a\rangle$ for high-energy states a [compare

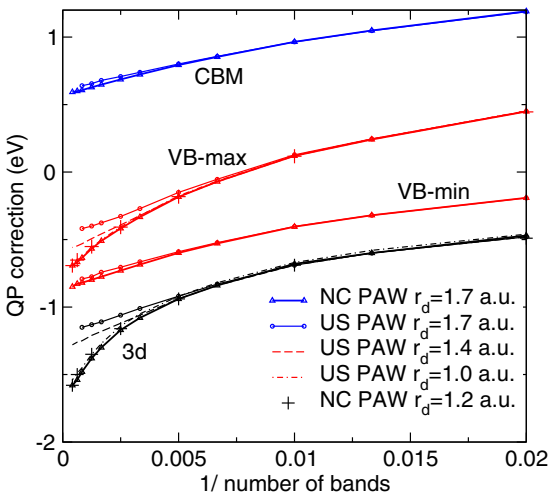

FIG. 1. (Color online) QP corrections to Kohn-Sham eigenvalues for ZnO G_0W_0 calculations. The number of bands varies between 50 (right most point) and 2400 (left most point). To improve the presentation, an upward shift of 1 eV was added to the VB and $3d$ QP corrections (i.e., the true corrections can be obtained by subtracting 1 eV, moving the lines down by 1 eV).

Eq. (14)]. A standard way to reduce the error of US PAWs is to decrease the core radius used upon creation of the PAW potential. Reducing the core radius for the d electrons r_d to 1.4 a.u. improves the results, but only slightly, and reasonable agreement with the NC PAW results is only obtained at a much smaller core radius of $r_d = 1.0$ a.u., where the violation of the norm is only about 20%. At this point, the ground-state calculations with the US PAW $r_d = 1.0$ a.u. potential, however, require already a larger basis set than the calculations with the NC PAW $r_d = 1.7$ a.u. potential. Specifically, the required plane-wave energy cutoffs for accurate ground-state calculations are 1150 eV for the NC PAW $r_d = 1.7$ a.u., and 500, 700, and 1400 eV for the US PAW potentials with core radii of $r_d = 1.7$, 1.4, and 1.0 a.u., respectively. Furthermore, within the line thickness, the results for NC PAW potentials are independent of the core radius for radii between 1.9 and 1.2 a.u. (see crosses in Fig. 1). Hence, the remaining Zn calculations were performed for a softer NC potential with the core radius increased to 1.9 a.u. (compare Table I).

The errors in the QP corrections are most pronounced for the 3d states. However, the valence band maximum (VBM), which is dominated by oxygen 2p states, follows the 3d states partially. This is of course related to the strong covalent Zn-3d O-2p interaction: at the Γ point the Zn-3d states hybridize with the O-2p states forming Zn-3d O-2p bonding and antibonding linear combinations. Since the antibonding linear combination, which forms the threefold-degenerated valence band maximum at Γ, has a strong Zn-3d character, the states partly follow the behavior of the 3d states. As a consequence, when the number of orbitals increases, the band gap opens.

This is, however, not always the case, as exemplified for GaAs and AlAs (Figs. 2 and 3). For these materials, there is no (AlAs) or little hybridization between the 3d states and the valence and conduction band states. Nevertheless, the 3d states still influence the valence band and conduction band states, as can be understood from Eq. (26). The d electrons will result in a spherically symmetric density at the As and Ga atoms,

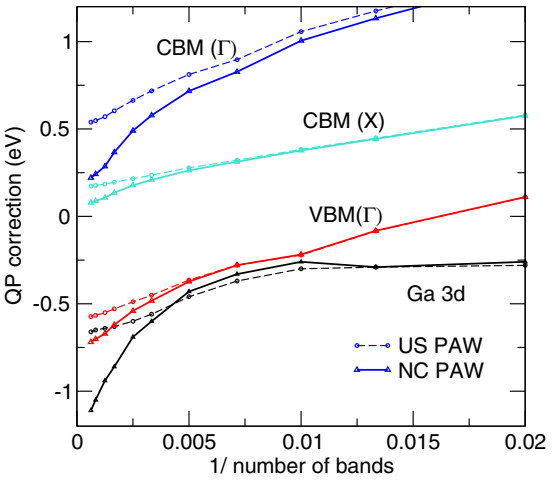

FIG. 2. (Color online) QP corrections to Kohn-Sham eigenvalues for GaAs G_0W_0 calculations. The number of bands varies between 50 (right most point) and 1600 (left most point). To the 3d QP corrections, 1.5 eV has been added.

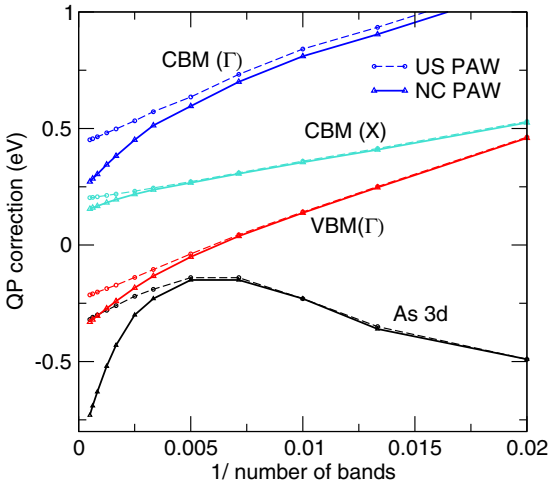

FIG. 3. (Color online) QP corrections to Kohn-Sham eigenvalues for AlAs G_0W_0 calculations. The number of bands varies between 50 (right most point) and 1600 (left most point). To the VB and 3d QP corrections, 0.5 and 3 eV have been added, respectively.

and this causes an attractive Coulomb hole at the atomic sites. The localized d states themselves and the s-like conduction band states are most strongly affected by this local potential and pulled down to lower energies.

For GaAs, the first thing to note is that the Ga 3d states first shift slightly upwards in energy and then suddenly drop in energy. This is a result of the strong spatial localization of the 3d electrons in Ga. The effect is even more pronounced for the As 3d states as shown in Fig. 3. The more important observation, however, is that the valence band states at the Γ point and the s-like conduction band minimum CBM(Γ) follow this drop. The CBM at the X point exhibits a much smaller slope than the valence band maximum VBM(Γ). This behavior reduces the Γ-Γ transition, but increases the Γ-X transition. Obviously, increasing the basis set can have quite a substantial impact on the relative position of the conduction band states and thus the direct and indirect gaps.

Remarkably, similar effects are observed even for AlAs as shown in Fig. 3. The changes of the QP corrections are about a factor of 2 smaller since the coupling to the strongly localized As 3d states is now rather small (note the different y scale in Fig. 3). Again, the QP correction of the CBM(Γ) state exhibits the largest slope, the CBM(X) the smallest one, and the VBM(Γ) is in-between. Consequently, the indirect gap of AlAs increases, when the basis-set size is increased. Clearly, whenever accurate direct and indirect gaps are needed, the d electrons must be taken into account, and this applies to both the anion as well as the cation (As, Se, Te). Furthermore, PAW potentials without norm-conserving partial waves yield too small QP corrections at Γ, in particular, for the CBM(Γ) state, as indicated by the thin dashed lines in Figs. 2 and 3, and consequently too large direct gaps and too small indirect gaps.

We performed extensive tests for ZnO, AlAs, CdS, GaN, and InP, all confirming these observations. Although up to about 100–200 orbitals per atom, the results are independent of the choice of the potential, once the number of orbitals is increased to several 1000, significant deviations between the

QP corrections become discernible. US PAW potentials always yield too small QP corrections for the d states and the results converge only slowly with decreasing core radii to a limiting value. In contrast, the results are almost independent of the specific choice of the core radius for NC PAW potentials.

A final word of caution is in place here. In the present implementation, the code restores an approximation of the exact all-electron density on the plane-wave grid. As briefly hinted at after Eq. (32), this is achieved by performing a Fourier transformation of the exact augmentation charge density $Q_{\alpha\beta}(\mathbf{r}) = \phi^*_\alpha(\mathbf{r})\phi_\beta(\mathbf{r}) - \tilde{\phi}^*_\alpha(\mathbf{r})\tilde{\phi}_\beta(\mathbf{r})$ to reciprocal space $Q_{\alpha\beta}(\mathbf{G})$ and then expanding the augmentation density in a set of orthogonal functions localized at each atomic site. In the present calculations, three functions for each quantum number l are used to expand the density difference between the all-electron and pseudopartial waves (NMAXFOCKAE = 2 in VASP). They are used to restore the proper norm ($l = 0$), dipoles ($l = 1$), and quadrupoles ($l = 2$) and higher multipoles ($l > 2$). The first function serves to restore the exact $Q_{\alpha\beta}(\mathbf{G})$ around $\mathbf{G} = 0$ and is not required if the potentials are norm conserving. The other two functions are chosen to restore the density at larger plane-wave vectors $\mathbf{G}$. Since the method guarantees that the density on the plane-wave grid is almost exactly equivalent to the exact all-electron density up to plane waves corresponding to a kinetic energy of 400 eV, the fairly complicated one-center terms in the screening matrix χ_0 and W do not need to be implemented. We believe that the present procedure and implementation is very accurate since it yields identical QP corrections for all potentials if the number of unoccupied orbitals is not too large (right-hand side in Figs. 1, 2, and 3). With norm-conserving partial waves, the results become even robust and stable for very large basis sets. Restoring just the norm of $Q_{\alpha\beta}(\mathbf{r})$ was found to be insufficient for GW calculations: if only the norm is restored, differences between different potentials can be around 1 eV for the QP energies of localized d states, even when only a small number of unoccupied orbitals is included in the GW calculations.

Finally, the interaction between the core and valence electrons is always evaluated exactly at the level of HF without any shape approximation, which is not always the case for codes using standard norm-conserving potentials. Based on our experience, we would not expect that such standard potentials without the PAW information (all-electron orbitals) can yield similar accuracy.

Technically, we believe that the procedures adopted here allow us to get accurate and converged results for virtually any material, although the construction of PAW potentials can be tedious at times. For instance, ghost states at energies below bound states or in the conduction band need to be avoided and a sufficient number of projectors at high energies need to be included.

B. Basis-set incompleteness correction

Converging GW results with respect to the basis-set size requires tedious extrapolation procedures, as performed here or in other work [28,39]. An alternative to the extrapolation is to use the simple correction for the QP energies [Eq. (27)], correcting the incompleteness of the auxiliary basis set. Here, we test how accurate this procedure is. To avoid any contamination of the results, convergence with respect to all other parameters, for instance, the number of bands and the plane-wave cutoff for the orbitals $E^{\mathrm{pw}}_{\mathrm{cut}}$, was achieved.

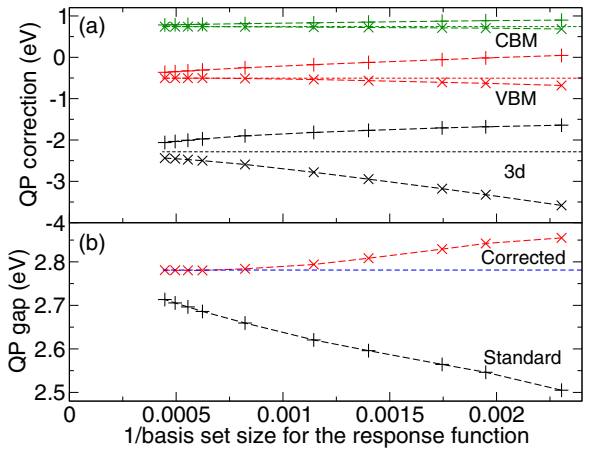

FIG. 4. (Color online) Dependence of (a) the unscaled QP correction (without Z) and (b) of the quasiparticle band gap of wurzite ZnO on the basis set used for the response function. Bare data (+ sign) and data corrected for the leading error according to Eq. (27), marked with × sign, are shown. Horizontal lines are extrapolated values. In (a) the data for the 3d and VBM states were shifted to a higher energy by 1.5 eV.

To test the correction, we again consider the well-investigated ZnO here in the wurzite structure (see, e.g., Refs. [14,28]). Figure 4(a) shows the dependence of the QP corrections on the inverse of the number of auxiliary basis functions in the response function both without and with the correction according to Eq. (27). The thin dashed lines show values extrapolated from the last three (noncorrected) data points assuming a linear convergence with the inverse of the basis-set size, i.e., $\varepsilon(N^{\chi}_{\mathrm{pw}}) = \varepsilon(\infty) + A/N^{\chi}_{\mathrm{pw}}$. The renormalization factor Z changes slightly when the cutoff is increased and we did not apply it here to remove this effect. The plane-wave cutoff for orbitals was taken to be three times the cutoff for the response properties, which was between 250 and 750 eV. Clearly, the QP corrections of the VB maximum and the CB minimum converge much faster when the correction is applied. The correction is larger for the VB maximum, which contains a contribution of the 3d states of Zn. The correction formula is less accurate for the 3d states: the corrected shift of the 3d states deviates initially more from the extrapolated value, and the difference is still pronounced for the largest basis set used. However, the extrapolated value for the 3d states might still possess some error because of the very slow convergence with N^{χ}_{pw}. Importantly, the coefficient for the correction calculated using Eq. (27) can be directly compared to the value of the coefficient A obtained from the fit. For the VB maximum and CB minimum, the agreement is excellent, the fit yields $A = 317$ and 89 eV for the VB maximum and CB minimum, respectively, while Eq. (27) gives $A = 315$ and 93 eV. For the 3d state, we obtain 502 eV from the fit and 842 eV from Eq. (27). Clearly, even larger cutoffs would have to be used to observe the $1/N^{\chi}_{\mathrm{pw}}$ behavior. However, this is hardly possible, not least because the largest calculation used already over 12 000 unoccupied bands.

The faster convergence of the QP corrections leads to a much faster convergence of the band gap, shown in Fig. 4(b). When the correction is used, a rather small cutoff of $E_{\text{cut}}^{\chi} = 300$ eV is sufficient to obtain a band gap converged to within 50 meV, whereas the error in the uncorrected band gap is still more than 200 meV. For similar errors, the corrected calculations are about 30 times cheaper, indicating that the correction is a promising tool to improve the convergence of QP energies in GW calculations. We also note that this correction is complementary to convergence accelerations with respect to the number of included orbitals [40,41]. Here, we correct for the error caused by the finite basis set used for the response function, whereas those address errors incurred by truncating the virtual orbital set.

C. G_0W_0 band gaps for semiconductors and insulators

Tables II, III, and V collect the band gaps for the materials considered in this work for LDA, G_0W_0@LDA, and GW_0@LDA calculations starting from LDA orbitals. The results for G_0W_0 starting from PBE orbitals are shown in Table IV. The lattice parameters of the materials are the same as in our previous work and are listed therein [35].

TABLE II. Position of valence band (VB) maximum at X, L (with respect to Γ), and conduction band (CB) minimum at Γ, X, and L with respect to VB maximum at Γ for the local density approximation. For materials with an inverted gap, the VB maximum is set to the triple degenerated state that normally corresponds to the VB maximum. All materials were considered in the diamond or zinc-blende structure, except for GaN and ZnO where the lines "wz" report the results for the wurtzite structures.

	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c	d
C	−21.32	5.54	−2.79	8.38	−6.29	4.70	
SiC	−15.33	6.28	−1.06	5.38	−3.20	1.30	
Si	−11.96	2.52	−1.20	1.42	−2.85	0.60	
Ge	−12.81	−0.15	−1.41	0.06	−3.08	0.66	
BN	−20.07	8.68	−1.94	10.19	−4.91	4.34	
AlP	−11.50	3.10	−0.77	2.66	−2.12	1.44	
AlAs	−11.90	1.86	−0.82	2.02	−2.17	1.35	
AlSb	−10.77	1.46	−0.91	1.26	−2.19	1.15	
GaN	−15.64	1.63	−0.95	4.42	−2.66	3.24	−13.50
wz		1.94					−13.30
GaP	−12.51	1.63	−1.13	1.51	−2.71	1.49	−14.68
GaAs	−12.76	0.32	−1.15	0.86	−2.69	1.34	−14.91
GaSb	−11.59	−0.06	−1.17	0.32	−2.61	0.83	−15.14
InP	−11.52	0.48	−0.98	1.31	−2.35	1.60	−14.13
InAs	−11.88	−0.43	−0.99	0.79	−2.33	1.43	−14.30
InSb	−10.77	−0.38	−1.01	0.42	−2.27	1.25	−14.51
MgO	−17.00	4.68	−0.67	7.75	−1.37	8.89	
ZnO	−17.68	0.62	−0.80	5.32	−2.21	5.13	−5.30
wz		0.75					−5.24
ZnS	−13.08	1.87	−0.87	3.10	−2.23	3.19	−6.31
ZnSe	−13.28	1.05	−0.87	2.36	−2.20	2.81	−6.55
ZnTe	−11.82	1.06	−0.90	1.65	−2.18	2.17	−6.94
CdO	−15.73	0.92	1.42	5.68	−0.98	5.10	−5.14
CdS	−12.40	0.90	−0.78	2.79	−1.95	3.30	−7.60
CdSe	−12.65	0.36	−0.77	2.19	−1.89	2.94	−7.80
CdTe	−11.23	0.55	−0.79	1.66	−1.89	2.45	−8.18

TABLE III. G_0W_0@LDA results for absolute shift of the valence band (VB) at Γ compared to LDA calculations (Δ IP), position of the VB minimum (Γ_{VBmin}), position of VB maximum at X, L (with respect to Γ), and conduction band (CB) minimum at Γ, X, and L. For Ga, In, Zn, and Cd, the position of the d band (average at the Γ point) is also indicated. Calculations are for $6 \times 6 \times 6$ k points. Results for materials that are metallic within LDA and where this k-point set might thus be insufficient for convergence are in italics.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c	d
C	−1.07	−22.06	7.43	−2.92	10.37	−6.55	6.21	
SiC	−0.92	−15.61	7.30	−1.09	6.58	−3.27	2.43	
Si	−0.60	−11.82	3.21	−1.21	2.06	−2.86	1.22	
Ge	*−0.54*	*−12.69*	*0.43*	*−1.42*	*0.56*	*−3.09*	*1.10*	
BN	−1.35	−21.09	11.29	−2.08	12.25	−5.19	6.37	
AlP	−0.83	−11.36	4.11	−0.78	3.70	−2.13	2.42	
AlAs	−0.82	−11.81	2.82	−0.83	2.96	−2.17	2.23	
AlSb	−0.65	−10.65	2.27	−0.91	1.99	−2.20	1.82	
GaN	−1.04	−15.86	2.88	−0.97	5.95	−2.68	4.59	−15.87
wz	−1.08		3.23					−15.66
GaP	−0.70	−12.34	2.50	−1.14	2.36	−2.71	2.25	−16.86
GaAs	−0.63	−12.59	1.08	−1.15	1.57	−2.68	1.96	−17.10
GaSb	*−0.54*	*−11.45*	*0.54*	*−1.17*	*0.84*	*−2.61*	*1.29*	*−17.32*
InP	−0.61	−11.32	1.13	−0.99	2.00	−2.35	2.21	−15.73
InAs	*−0.60*	*−11.69*	*0.13*	*−1.00*	*1.43*	*−2.33*	*2.02*	*−15.84*
InSb	*−0.54*	*−10.61*	*0.13*	*−1.03*	*0.93*	*−2.28*	*1.72*	*−16.06*
MgO	−2.01	−17.47	7.55	−0.71	10.86	−1.44	11.91	
ZnO	−1.53	−17.75	2.46	−0.79	7.43	−2.15	7.00	−6.22
wz	−1.78		2.83					−6.09
ZnS	−1.15	−12.62	3.36	−0.86	4.70	−2.18	4.62	−7.55
ZnSe	−1.09	−12.97	2.38	−0.85	3.75	−2.12	4.06	−7.82
ZnTe	−0.84	−11.59	2.17	−0.88	2.71	−2.13	3.15	−8.31
CdO	*−0.81*	*−16.16*	*1.57*	*1.14*	*6.83*	*−0.98*	*6.46*	*−6.78*
CdS	−0.96	−11.98	2.05	−0.77	4.09	−1.91	4.48	−8.67
CdSe	−0.90	−12.37	1.33	−0.75	3.27	−1.83	3.93	−8.84
CdTe	−0.75	−10.99	1.46	−0.78	2.58	−1.85	3.28	−9.34

The first important issue to note is that some of the materials show band inversion in LDA and PBE, namely, Ge, GaSb, InAs, and InSb: for these materials, the threefold-degenerated (cation) p orbitals are incorrectly above the anion s orbital at the Γ point. In Table II, the gap is then given as a negative value. From Table III, we see that the band inversion is already abolished in the first G_0W_0 iteration. Note, however, that the orbitals are kept fixed at the LDA level and that any incorrect hybridization can not be corrected for by updating the eigenvalues only [42,43]. Furthermore, the band inversion causes sometimes convergence problems in the GW_0 iteration. Specifically, this happened for CdO, where subsequent iterations did not converge.

A further technical problem is that both in the GW_0, as well as G_0W_0 calculations, the screening of the system is practically metallic for systems with improper band ordering in the DFT. Hence, we expect that these results are possibly not well converged with respect to the k-point mesh, and the final reported numbers should be considered with some caution. We report those values in italics in the tables. Another reason for errors is the frequency integration. We represent the polarizability on a discretized frequency grid [36] with 200 frequency points, and double checked whether a reduction

TABLE IV. Same as Table III but for G_0W_0@PBE. The IPs are generally 0.15 eV more negative than for G_0W_0@LDA.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c	d
C	−1.22	−22.12	7.43	−2.93	10.38	−6.58	6.23	
SiC	−1.05	−15.69	7.35	−1.10	6.62	−3.30	2.42	
Si	−0.72	−11.83	3.25	−1.21	2.14	−2.86	1.28	
Ge	*−0.82*	*−12.70*	*0.63*	*−1.43*	*0.67*	*−3.08*	*1.16*	
BN	−1.53	−21.05	11.33	−2.06	12.29	−5.18	6.40	
AlP	−0.96	−11.37	4.23	−0.77	3.79	−2.13	2.48	
AlAs	−1.01	−11.82	2.99	−0.82	3.08	−2.17	2.31	
AlSb	−0.84	−10.67	2.40	−0.91	2.07	−2.20	1.87	
GaN	−1.20	−15.93	2.85	−0.98	5.93	−2.70	4.50	−15.82
wz	−1.23		3.20					−15.61
GaP	−0.86	−12.34	2.62	−1.14	2.45	−2.71	2.30	−16.77
GaAs	−0.91	−12.56	1.23	−1.14	1.68	−2.66	2.04	−16.97
GaSb	*−0.77*	*−11.41*	*0.68*	*−1.17*	*0.92*	*−2.60*	*1.34*	*−17.22*
InP	−0.78	−11.29	1.23	−0.99	2.10	−2.35	2.28	−15.66
InAs	*−0.83*	*−11.65*	*0.23*	*−1.00*	*1.48*	*−2.32*	*2.04*	*−15.73*
InSb	*−0.75*	*−10.57*	*0.25*	*−1.03*	*0.99*	*−2.28*	*1.76*	*−15.94*
MgO	−2.16	−17.61	7.55	−0.73	10.80	−1.45	11.82	
ZnO	−1.73	−17.84	2.42	−0.79	7.41	−2.17	6.93	−6.16
wz	−1.87		2.76					−5.98
ZnS	−1.33	−12.64	3.46	−0.87	4.79	−2.19	4.68	−7.45
ZnSe	−1.34	−12.95	2.55	−0.85	3.88	−2.13	4.15	−7.64
ZnTe	−1.07	−11.56	2.27	−0.89	2.78	−2.14	3.21	−8.10
CdO	*−0.95*	*−16.29*	*1.50*	*1.29*	*6.74*	*−0.99*	*6.45*	*−6.20*
CdS	−1.11	−11.97	2.15	−0.78	4.19	−1.93	4.55	−8.62
CdSe	−1.15	−12.34	1.52	−0.75	3.44	−1.84	4.06	−8.71
CdTe	−0.99	−10.96	1.57	−0.79	2.67	−1.86	3.35	−9.19

of the frequency points from 200 to 100 points changes the results. For all systems, except ZnO, the changes in the QP energies and gap are well below 50 meV, with most changes being only 10–20 meV. For ZnO, the gap increases by 100 meV, when the number of frequency points is reduced to 100. The origin for this change is the one single sharp *p*-*s* transition occurring at the Γ point, which is not accurately represented with the coarser frequency grid. Although increasing the frequency grid to 400 points changed the gap by less than 50 meV for ZnO, error bars for ZnO are possibly somewhat larger than for other systems. All in all, we expect the final errors to be around 50 meV for *s* and *p* dominated states and 100 meV for *d* states.

The effect of the starting functional (LDA versus PBE, Table III versus Table IV) is small for all materials without *d* electrons (see, for example Si). However, if *d* electrons are present, the *d* electrons shift slightly towards the Fermi level by 0.1 eV in the G_0W_0, as well as in the preceding PBE calculations. More notable is the increase of the band gap by about 0.2 eV, in particular, for the Γ-Γ transition for both PBE (not shown) and G_0W_0@PBE. This implies that G_0W_0@LDA calculations can not be straightforwardly compared to G_0W_0@PBE calculations for systems with *d* electrons. Although the PBE starting point seems better suited since the final gaps are larger and always in better agreement with experiment, we will mainly concentrate on the G_0W_0@LDA calculations in the following discussion since it is common practice to start from LDA orbitals.

Let us start with a discussion of those materials that have been most widely discussed in literature, Si and ZnO. For Si, Friedrich *et al.* [44] reported valence band and conduction band positions of 3.2 eV (Γ-Γ), −1.22, 2.12 (Γ-*L*), −2.92, 1.19 (Γ-*X*) compared to our values of 3.22 (Γ-Γ), −1.21, 2.06 (Γ-*L*) −2.87, 1.23 (Γ-*X*). However, Friedrich's calculations were performed using only 250 bands and $4 \times 4 \times 4$ *k* points, and they note that "the results are lowered by 0.02 eV if the *k*-point mesh is fully converged, and by another 0.02 eV if screening due to the $2p$ electrons is included in the correlation self-energy." These effects might further improve agreement with our results.

ZnO is certainly the most controversial material, with reported band gaps varying between 2.1 and 3.4 eV for G_0W_0 calculations [14,37,41]. The larger value of 3.4 has been shown to be related to the plasmon pole model [15], whereas the smaller value of 2.1 eV has been reported for VASP, albeit for zinc-blende and PBE orbitals [37]. This decreases the gap by 0.2 and 0.1 eV, respectively; the G_0W_0@LDA band gap for wurtzite ZnO is 2.4 eV with a similar setup and potential as in Ref. [37]. In the present calculations, we predict a basis-set converged band gap of 2.83 eV using $6 \times 6 \times 6$ *k* points. Increasing the grid to $8 \times 8 \times 8$ *k* points, the value increases to 2.87 eV. These values are only slightly larger then the value of 2.83 eV reported by Friedrich *et al.* [28] for $8 \times 8 \times 8$ *k* points. Due to the very slow convergence with the number of orbitals and slow *k*-point convergence, we, however, emphasize that our ZnO values are possibly less accurate than for other materials: errors around 50 meV are technically very difficult to achieve for this material.

Concerning the *d* levels and the other materials considered in this work, we start with a comparison with the most recent WIEN2K full potential linearized augmented-plane-wave (FLAPW) values by Li *et al.* [45]. For the *d* states, they report values of −6.84, −7.12, −7.55 eV (ZnS, ZnSe, ZnTe) and −8.16, −8.45, −8.74 eV (CdS, CdSe, CdTe), typically 0.5–0.7 eV higher than our present values. This is likely to be a result of insufficient band and basis-set convergence in the FLAPW calculations. For instance, in the FLAPW calculations the number of bands was nowhere close to the values required to see the drop in the *d*-level energies (compare Fig. 1). The reported apparent convergence might be an indication that the one-center basis sets, which play a similar role as the projectors in the PAW method, were not sufficiently large. Compared to our own previous values reported in Ref. [37], we note that these early calculations were neither basis-set converged, nor did we very accurately restore the all-electron density distribution on the plane-wave grid. Both errors might or might not accidentally cancel; certainly the present values are more accurate and supersede the previous values.

For the band gaps, the WIEN2K FLAPW calculations reported in Ref. [46] are, except for GaAs (see below), generally smaller than our G_0W_0 band gaps: 1.00 (Si), 5.42 (C), BN (6.03), AlP (2.18), GaAs (1.29), compared to the present values of 1.10 (Si), 5.61 (C), BN (6.37), AlP (2.43), GaAs(1.08). We suspect that this is again a result of insufficient convergence with respect to the plane-wave or one-center FLAPW basis sets. More accurate values have been published by Friedrich *et al.* using the FLAPW method [47]. In this case, the one-center basis sets were improved using local orbitals.

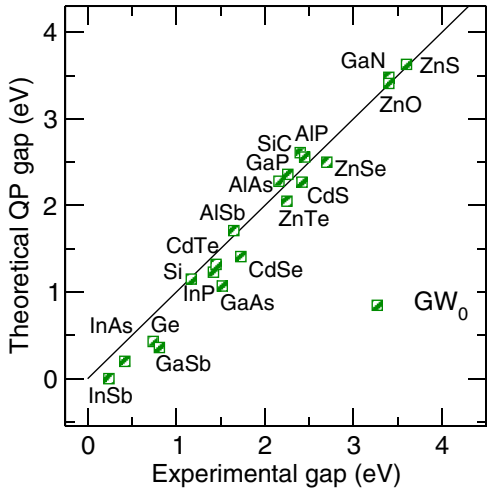

FIG. 5. (Color online) Band gaps predicted using GW_0@LDA compared to experimental values. The theoretical band gaps have been corrected for spin-orbit coupling given in the Supplemental Material of Ref. [35]. Experimental data are also collected in Ref. [35].

They find band gaps of 1.11 (Si), GaAs (1.31), CdS (2.18), GaN (2.83), 5.62 (C), BN (6.20), MgO (7.17), most in good agreement with our values 1.10 (Si), GaAs (1.08), CdS (2.10), GaN (2.90), 5.61 (C), BN (6.37), MgO (7.54). As to why our band gaps are generally larger for oxides and nitrides compared to the FLAPW calculations and our previous calculations in Ref. [37], we note that the $2p$ levels need to be treated with similar care as the d states, and we suspect that the one-center basis sets for those materials were still not fully converged in the FLAPW calculations.

For GaAs, our present band gap is smaller than the two FLAPW values (both around 1.30) and our previous calculations. As already discussed above and shown in Fig. 2, as opposed to other materials, in GaAs the band gap decreases when the basis-set size is increased. After communicating our values, Friedrich carefully recalculated the FLAPW gaps with improved settings. For similar setups and PBE starting orbitals, his new calculations yielded only a 50 meV larger band gap than our calculations, which is within our estimated error bars [48]. This confirms our suspicion that most of the previously reported values (including our own publications) were not fully basis-set converged. In this work, we have done everything to converge the results carefully, for instance, by checking every single PAW potential painstakingly, and comparing them with even smaller core PAW potentials.

Figure 5 compares the calculated GW_0@LDA band gaps tabulated in Table V with experimental values collected in Ref. [35]. We observe that the band gaps for the "metallic" LDA systems (InSb, InAs, GaSb, and Ge) are noticeably too small. For sp bonded systems, the agreement is otherwise seemingly rather good, although it should be noted that electron-phonon coupling can significantly reduce the experimentally measured gap with the effect being about 0.4 eV for C and 0.1 eV for Si [49–51]. We expect similar contributions of the order of 0.4 eV for other oxides, nitrides, and carbides, and of the order of 0.1–0.2 eV for sulfides and materials containing aluminium. If these corrections are considered, the comparison with experiment would be less favorable. However, to address this issue properly requires one to perform the actual calculations with the phononic contributions accounted for.

TABLE V. Same as Table III for GW_0@LDA. GW_0@LDA calculations did not converge for CdO because of the incorrect band order for the LDA starting orbitals. The GW_0@PBE are reported in the Supplemental Material of Ref. [35].

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c	d
C	−1.21	−22.38	7.57	−3.02	10.60	−6.71	6.36	
SiC	−1.10	−15.80	7.50	−1.10	6.79	−3.32	2.61	
Si	−0.72	−11.93	3.32	−1.22	2.15	−2.90	1.31	
Ge	*−0.56*	*−12.78*	*0.53*	*−1.44*	*0.62*	*−3.12*	*1.15*	
BN	−1.56	−21.30	11.66	−2.10	12.58	−5.25	6.65	
AlP	−0.99	−11.50	4.27	−0.79	3.86	−2.17	2.58	
AlAs	−1.00	−11.96	2.96	−0.83	3.12	−2.20	2.39	
AlSb	−0.78	−10.79	2.38	−0.93	2.10	−2.23	1.93	
GaN	−1.30	−16.41	3.11	−0.98	6.25	−2.71	4.88	−16.82
wz	−1.33		3.48					−16.57
GaP	−0.85	−12.51	2.63	−1.15	2.50	−2.74	2.39	−17.98
GaAs	−0.78	−12.75	1.19	−1.16	1.68	−2.71	2.07	−18.30
GaSb	*−0.66*	*−11.57*	*0.62*	*−1.18*	*0.92*	*−2.63*	*1.37*	*−18.66*
InP	−0.76	−11.47	1.23	−1.01	2.12	−2.38	2.34	−16.48
InAs	*−0.86*	*−11.72*	*0.32*	*−0.91*	*1.65*	*−2.24*	*2.24*	*−16.47*
InSb	*−0.68*	*−10.71*	*0.26*	*−1.02*	*1.03*	*−2.29*	*1.82*	*−16.87*
MgO	−2.42	−18.21	8.04	−0.73	11.42	−1.49	12.48	
ZnO	−2.17	−18.11	3.10	−0.81	8.20	−2.21	7.71	−6.68
wz	−2.30		3.41					−6.47
ZnS	−1.43	−12.82	3.66	−0.86	5.01	−2.21	4.92	−8.27
ZnSe	−1.37	−13.13	2.64	−0.85	4.03	−2.14	4.34	−8.62
ZnTe	−1.03	−11.75	2.35	−0.89	2.89	−2.16	3.32	−9.14
CdS	−1.22	−12.16	2.30	−0.79	4.37	−1.94	4.77	−9.18
CdSe	−1.16	−12.53	1.55	−0.76	3.52	−1.85	4.19	−9.39
CdTe	−0.95	−11.17	1.62	−0.79	2.76	−1.88	3.46	−9.87

V. DISCUSSION AND CONCLUSION

One central result of this work is a simple asymptotic estimate of the self-energy contribution from high-energy plane-wave-like states [see Eq. (26)]. Remarkably, this estimate allows us to determine the error in QP corrections using only the total charge density and the charge density of the QP state in question. The estimate is exceedingly easy to calculate and can be readily implemented in any plane-wave or all-electron code. A similar error estimate is obtained for the correlation energies, given by Eq. (9). That both errors are related is fairly obvious since QP removal and addition energies can be regarded as the energy differences between systems with N and $N-1$ and $N+1$ electrons, respectively. The important observation is that in both cases, plane-wave-like states $\mathbf{G}$ and $\mathbf{G}'$, that are not relevant for ground-state calculations, contribute to the correlation energy since the density at a wave vector $\mathbf{G}-\mathbf{G}'$ is involved. Due to this folding, even plane waves with very large momentum contribute to the correlation energy. We have discussed briefly that the origin of this problem is related to the so-called cusp condition. The addition of an electron at position $\mathbf{r}$ induces a charge depletion around $\mathbf{r}$, which is described by the exchange-correlation hole $h^{\text{xc}}(\mathbf{r},\mathbf{r}')$. The

correlation contribution to the hole is related to the screening charge density, and these quantities have a cusp for $\mathbf{r}' \to \mathbf{r}$, which can be accounted for only when an infinitely large basis set is used.

The present correction is complementary to resolution of identity methods [40] or effective energy denominator techniques [41]. These techniques allow one to obtain converged results for a specific plane-wave basis set using only a small number of orbitals. They, however, do not allow one to determine the error incurred by disregarding plane-wave components in the auxiliary basis set for the overlap charge density resulting from two orbitals. Our present correction does exactly this. Specifically, it allows us to estimate the residual error resulting from the neglect of high momentum plane-wave components in the density response function. As observed before (and re-derived here), this error is proportional to the inverse of the number of basis functions included in the auxiliary basis set for the density or response function. The correction has been tested for the case of ZnO and seems to be quite promising, much reducing the computational effort required to obtain an accurate value for the QP gap. However, for localized states, such as the Zn-d states, the approximations made in deriving the leading-order correction become less accurate and terms beyond the leading order should be included.

Alternatively, one can extrapolate to the infinite-basis set-limit by performing a series of calculations with an increasing basis set for the orbitals and the response function. According to the derived asymptotic behavior, the error should fall off like one over the number of basis functions. This procedure has already been applied by Garcia and Marini in RPA calculations [52]. To avoid any bias, this procedure was adopted in this work to predict QP energies for 24 materials.

Unfortunately, the derivation of Eq. (26) indicates that standard PAW potentials with *non*-norm-conserving "ultrasoft" partial waves will not describe QP levels accurately since the norm is not correctly restored when the overlap density between ground-state orbitals and high-energy plane-wave-like orbitals is calculated. We observe that the resulting error can be substantial. For instance, for ZnO using US PAW potentials, the d levels are almost 1 eV too high in energy compared to accurate reference calculations, and the band-gap error is about 400 meV. Although ZnO is an extreme example, since the d levels are rather shallow (7 eV below the Fermi level) and the violation of the norm is particularly large for $3d$ elements, we found that similar errors are observed for all materials containing elements with $3d$, as well as $4d$ and $5d$ electrons. Even for AlAs and GaAs, the inclusion of the As $3d$ levels changes the indirect gap by about 200 meV. Equation (26) suggests two strategies to reduce the error: (i) either more projectors at higher energies need to be included, or alternatively (ii) the partial waves are made norm conserving. Although we have not discussed option (i) in detail, we found that it is, in practice, difficult to increase the number of projectors beyond three for a given quantum number l. Solution (ii), however, works reliably and yields PAW potentials that are robust and accurate. Furthermore, the specific choice for the core/pseudization radius influences the QP energies only very little for NC PAW potentials, as demonstrated for ZnO (and observed for all materials considered here). The disadvantage of NC PAW potentials is that they require about 50% larger plane-wave cutoff energies than standard ultrasoft PAW potentials. Although this makes the potentials impractical for ground-state calculations, fairly large plane-wave cutoffs are anyway required to obtain converged QP energies, so that the increased cutoff has, in practice, little consequences for GW calculations.

In this work, we have calculated G_0W_0 and GW_0 band gaps for prototypical semiconductors and insulators using the newly constructed NC PAW potentials. Overall, we observe what has now been known for quite some time. The QP gaps for medium gap materials are very accurate using the G_0W_0 approximation, slightly outperformed by the GW_0 approximation. Errors are even acceptable for ZnO, that has long been considered to be problematic. For small gap systems, however, our results are clearly unsatisfactory, and this concerns, in particular, materials that show incorrect band order at the Γ point in the local density approximation: in Ge, GaSb, InAs, InSb, and CdO, the anion p states are above the Fermi level, whereas the cation s state is below the Fermi level for LDA. This results in sizable errors in the standard perturbative GW approach. Better starting points than DFT orbitals and DFT screening are required in these cases. Presently, the most successful approaches are the self-consistent procedures of Kotani and Schilfgaarde [13,53] or somewhat cheaper hybrid functionals [35,54,55] or self-consistent calculations based on optimized effective potentials in the RPA [39].

We believe that our reported values are as accurate as we can possibly make them within the limits of the GW@DFT approximation, and we hope that they can serve as a benchmark for other GW codes. It is in fact uttermost time for the *ab initio* community to establish such benchmarks to make G_0W_0 a truly validated tool (as DFT already is). After all, validation has made quantum chemistry approaches using Gaussian-type orbitals and DFT so successful. Unfortunately, GW results with that kind of reference quality are still rare. Specifically, we are only aware of accurate calculations for Si and ZnO, for which we have made a comparison and found very good agreement [28,44]. In this context, we want to emphasize that methods employing compact basis sets, such as the linearized muffin tin orbital methods and even the standard full-potential linearized plane-wave method, will experience similar problems to the ones we experience with non-norm-conserving PAW potentials: to obtain highly accurate QP corrections and many-electron correlation energies, plane waves and local basis functions that are irrelevant for density functional ground-state calculations need to be included in excited-state calculations. This fact has been known in the quantum chemistry community for some time and certainly imposes a computational challenge that needs to be overcome when reference quality data are sought.

ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF) within the SFB ViCoM (Grant No. F 41) and I597-N16 (research unit FOR 1346 of the Deutsche Forschungsgemeinschaft and FWF). Supercomputing time on the Vienna Scientific cluster (VSC) is gratefully acknowledged. We thank C. Friedrich for useful comments on the manuscript.

APPENDIX A: COMPARISON OF RPA RESULTS WITH AND WITHOUT NORM-CONSERVING PAW POTENTIALS

The incorrect asymptotic convergence affects also the correlation energies obtained with the MP2 and RPA methods. To assess the magnitude and significance of the errors, we have recalculated the lattice constants of d metals using the NC PAW potentials and compared them to previous results published in Ref. [56]. We find that the absolute correlation energies indeed increase in magnitude when NC PAWs are used, typically by 10% to 20%. The larger correlation energy leads in almost all cases to a reduction of the equilibrium volume. The effect depends on the magnitude of the norm violation in the original US PAW potentials, which has been up to 90% for some $3d$ metals. Concomitantly, for the $3d$ metals Sc, Ti, Co, and Ni, the equilibrium volume is reduced by 2.5%–3%, significantly improving agreement between the RPA and experiment. The norm was better preserved for $4d$ and $5d$ metals and here the changes of the equilibrium volumes are only around -0.5%. Changes for s, p bonded materials were found to be negligible since PAW potentials for those elements violate the norm very little. A more detailed discussion will be published elsewhere.

APPENDIX B: ASYMPTOTIC CONVERGENCE OF STATIC COHSEX

It is interesting to relate the second-order quasiparticle correction [Eqs. (22) and (24)] to the second order of the commonly employed static COHSEX. The static approximation is obtained by assuming $\omega - \varepsilon_n = 0$ (or $\varepsilon_m - \varepsilon_n = 0$), i.e., assuming that the largest contributions to the self-energy come from states close in energy to the state m. For the second-order COH contribution [Eq. (24)], this approximation yields the following formula:

$$\langle m|\Sigma^{\text{stat}-\text{COH}(2)}|m\rangle = \frac{2}{\Omega^2}\frac{4\pi}{\mathbf{G}^2}\frac{4\pi}{\mathbf{G}'^2}\sum_{n}^{\text{all}}\langle m|-\mathbf{G}|n\rangle\langle n|\mathbf{G}'|m\rangle \times \sum_{ia}\frac{\langle i|-\mathbf{G}|a\rangle\langle a|\mathbf{G}'|i\rangle}{\varepsilon_i-\varepsilon_a}. \quad \text{(B1)}$$

For large $\mathbf{G}$ and $\mathbf{G}'$, a sizable contribution is only expected to arise if the orbital a corresponds to a plane wave $\mathbf{G}_a \approx -\mathbf{G}$ and $\mathbf{G}_a \approx -\mathbf{G}'$. Neglecting the energy of the occupied state ε_i, one obtains $\varepsilon_i - \varepsilon_a \approx \mathbf{G}^2/2$. Comparing this to the approximation for the energy denominator used in the full equation [Eq. (5)], we see that the static COH leads to exactly twice the proper value. Since the SEX part converges faster, the static COHSEX overestimates the high $\mathbf{G}$ contributions by a factor 2, as has been recently observed in actual calculations (without proof) by Kang and Hybertsen [27]. We note that the factor $\frac{1}{2}$ can be also derived using the plasmon pole approximation as shown by Deslippe *et al.* following a route similar to our current derivation [57]. The present derivation, however, is exact in second order and does not involve any specific model (except the assumption that at high energy, plane waves do not contribute to the expansion of the occupied orbitals).

[1] L. Hedin, Phys. Rev. **139**, A796 (1965).
[2] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[3] M. Lannoo, M. Schlüter, and L. J. Sham, Phys. Rev. B **32**, 3890 (1985).
[4] G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. Lett. **45**, 290 (1980).
[5] G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. B **25**, 2867 (1982).
[6] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[7] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[8] G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, 2005).
[9] F. Bechstedt, F. Fuchs, and G. Kresse, Phys. Status Solidi B **246**, 1877 (2009).
[10] R. Starke and G. Kresse, Phys. Rev. B **85**, 075119 (2012).
[11] A. Fleszar and W. Hanke, Phys. Rev. B **56**, 10228 (1997).
[12] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).
[13] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
[14] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **105**, 146401 (2010).
[15] M. Stankovski, G. Antonius, D. Waroquiers, A. Miglio, H. Dixit, K. Sankaran, M. Giantomassi, X. Gonze, M. Côté, and G.-M. Rignanese, Phys. Rev. B **84**, 241201 (2011).
[16] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).
[17] C. Schwartz, Phys. Rev. **126**, 1015 (1962).
[18] W. Kutzelnigg and J. D. Morgan, J. Chem. Phys. **96**, 4484 (1992).
[19] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[20] A. Halkier, T. Helgaker, T. Jorgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).
[21] J. Harl and G. Kresse, Phys. Rev. B **77**, 045136 (2008).
[22] J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi, Phys. Rev. B **86**, 035111 (2012).
[23] A. Gulans (unpublished).
[24] T. Björkman, A. Gulans, A. V. Krasheninnikov, and R. M. Nieminen, Phys. Rev. Lett. **108**, 235502 (2012).
[25] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. D. White, and R. W. Godby, Comput. Phys. Commun. **125**, 105 (2000).
[26] E. Luppi, H.-C. Weissker, S. Bottaro, F. Sottile, V. Veniard, L. Reining, and G. Onida, Phys. Rev. B **78**, 245124 (2008).
[27] W. Kang and M. S. Hybertsen, Phys. Rev. B **82**, 195108 (2010).
[28] C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B **83**, 081101 (2011); **84**, 039906(E) (2011).
[29] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); **78**, 1396 (1997).
[30] M. Marsman, A. Grüneis, J. Paier, and G. Kresse, J. Chem. Phys. **130**, 184103 (2009).
[31] A. Grüneis, J. J. Shepherd, A. Alavi, D. P. Tew, and G. H. Booth, J. Chem. Phys. **139**, 084112 (2013).
[32] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[33] A. Grüneis, M. Marsman, and G. Kresse, J. Chem. Phys. **133**, 074107 (2010).
[34] A. Schindlmayr, Phys. Rev. B **87**, 075104 (2013).
[35] A. Grüneis, G. Kresse, Y. Hinuma, and F. Oba, Phys. Rev. Lett. **112**, 096401 (2014).
[36] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
[37] M. Shishkin and G. Kresse, Phys. Rev. B **75**, 235102 (2007).
[38] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[39] J. Klimeš and G. Kresse, J. Chem. Phys. **140**, 054516 (2014).
[40] F. Bruneval and X. Gonze, Phys. Rev. B **78**, 085125 (2008).
[41] J. A. Berger, L. Reining, and F. Sottile, Phys. Rev. B **85**, 085126 (2012).
[42] M. van Schilfgaarde, T. Kotani, and S. V. Faleev, Phys. Rev. B **74**, 245125 (2006).
[43] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B **76**, 165106 (2007).
[44] C. Friedrich, A. Schindlmayr, S. Blügel, and T. Kotani, Phys. Rev. B **74**, 045104 (2006).
[45] X.-Z. Li, R. Gomez-Abal, H. Jiang, C. Ambrosch-Draxl, and M. Scheffler, New J. Phys. **14**, 023006 (2012).
[46] R. Gómez-Abal, X. Li, M. Scheffler, and C. Ambrosch-Draxl, Phys. Rev. Lett. **101**, 106404 (2008).
[47] C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B **81**, 125102 (2010).
[48] C. Friedrich (private communication).
[49] F. Giustino, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **105**, 265501 (2010).
[50] E. Cannuccia and A. Marini, Phys. Rev. Lett. **107**, 255501 (2011).
[51] S. Poncé, G. Antonius, P. Boulanger, E. Cannuccia, A. Marini, M. Côté, and X. Gonze, Comput. Mater. Sci. **83**, 341 (2014).
[52] P. García-González, J. J. Fernández, A. Marini, and A. Rubio, J. Phys. Chem. A **111**, 12458 (2007).
[53] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
[54] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, Phys. Rev. B **76**, 115109 (2007).
[55] J. Paier, M. Marsman, and G. Kresse, Phys. Rev. B **78**, 121201 (2008).
[56] L. Schimka, R. Gaudoin, J. Klimeš, M. Marsman, and G. Kresse, Phys. Rev. B **87**, 214102 (2013).
[57] J. Deslippe, G. Samsonidze, M. Jain, M. L. Cohen, and S. G. Louie, Phys. Rev. B **87**, 165124 (2013).