Self-screening corrections beyond the random-phase approximation: Applications to band gaps of semiconductors

Viktor Christiansson  
Department of Physics, University of Fribourg, 1700 Fribourg, Switzerland

Ferdi Aryasetiawan  
Department of Physics, Division of Mathematical Physics, Lund University, Professorsgatan 1, 223 63 Lund, Sweden

The self-screening error in the random-phase approximation (RPA) and the GW approximation (GWA) is a well-known issue and has received attention in recent years with several methods for a correction being proposed. We here apply two of these, a self-screening and a so-called “self-polarization” correction scheme, to model calculations to examine their applicability. We also apply an explicit self-screening correction to ab-initio calculations of real materials. We find indications for the self-polarization scheme to be the more appropriate choice of correction for localized states, and additionally we observe that it suffers from causality violations in the strongly correlated regime. The self-screening correction used in this work on the other hand significantly improves the description in more delocalized states. It provides a notable reduction in the remaining GWA error when calculating the band gaps of several semiconductors, indicating a physical explanation for a part of the remaining discrepancy in one-shot GW compared to experiment, while leaving the localized semicore $d$ states mostly unaffected.

I. INTRODUCTION

One of the most important quantities in many-electron theory is the linear density response function, henceforth referred to as response function, which describes how the charge density in a system of electrons is modified upon application of a time-dependent potential. Although the formal expression for the response function is known, in practice one must resort to approximations in order to compute it for real materials, as is the usual case in many-electron theory. Perhaps the most successful and widely used approximation is the random-phase approximation (RPA)\cite{1} which is also the basic approximation used in the well-known GW approximation (GWA)\cite{2} nowadays routinely used to calculate the quasiparticle dispersion in materials. In the GWA the screened interaction $W$ is computed within RPA starting from the Kohn-Sham states and eigenvalues obtained from a DFT\cite{3,4} calculation.

A well-known and long-standing problem associated with the RPA is the problem of self-screening. The deficiency can be illustrated most clearly for the case of the hydrogen atom\cite{5} with only one electron in the system, it is not possible to screen the interaction between the electron and an external field since there are no other electrons present, and yet the RPA erroneously yields a non-zero response function. Hence, when the GWA is applied to hydrogen, it yields a correlation self-energy which should be zero. The source of this error arises from a self-screening process inherent in the RPA. As illustrated in Fig.\ref{fig1}a, the only electron in the system, represented by a Green’s function line, should not participate in the polarization bubble; however, in the GWA this self-energy diagram is non-zero.

The self-screening error has received attention in recent years and a few different approaches to treat it have been proposed in the literature. Romanelli et al\cite{6} used vertex corrections to the self-energy beyond the RPA and applied it to simple model systems. Wetherell et al\cite{7} proposed the use of a local potential, based on density functional theory, added to the self-energy, and tested it successfully for simple one-dimensional models. In an earlier work by Aryasetiawan et al\cite{8}, it was instead suggested to use a correction based on the introduction of an orbital- and spin-dependent screened interaction, for which it was found that the HOMO-LUMO gap of a hydrogen dimer was correctly reproduced in the weakly to moderately correlated regime.

In the present work, we will employ the schemes proposed in Ref.\cite{8} to take the self-screening correction into account. The idea is that, in a given self-energy diagram, the Green’s function line representing the propagation of an electron or a hole of a given orbital is removed from the Green’s function appearing in the polarization. In this way, the screening processes associated with the electron or hole of the orbital are eliminated from the total polarization, at the cost of introducing an orbital- and spin-dependent interaction. A similar idea can also be applied directly to the response function; instead of removing an orbital line in the Green’s function, individual contributions to the polarization associated with an electron-hole excitation are eliminated from the polarization diagram to avoid “self-polarization”.

The aim of this paper is two-fold. First, by using model calculations on a one- and two-orbital Hubbard dimer we gain further insight into the strengths and applicability regimes of the two correction schemes, when applied to response function and self-energy calculations. Causality violations observed for the self-polarization correction are investigated, and an explanation of the origin is proposed based on the differences and similarities between the two.
correction schemes. Secondly, we move beyond previous studies of the self-screening error which have only dealt with model calculations, by applying the self-screening correction in fully ab initio calculations for a number of semiconductors. The calculated bandgaps and semicore quasiparticle energies are compared to conventional GW calculations and experimental values.

The paper is organized as follows. In Sec. II we introduce the self-screening and self-polarization corrections following the derivation given in Ref. [8] and describe a method for making the self-screening correction tractable for the calculation of real materials. In Sec. III A and III B we describe our result for the model calculations and discuss the causality considerations, respectively, for the two correction schemes. In Sec. III C we present the results from the ab initio calculations and compare them to experiments, and finally we discuss and summarize our results in Sec. IV.

II. THEORY AND IMPLEMENTATION

A. Self-screening corrected GW

In the GWA[2] the self-energy for a spin \( \sigma \) is calculated as

\[
\Sigma^{GW}_\sigma (r,t') = i G^{0}_\sigma (r,t') W (r',t, r) = i \sum_m g_{m\sigma} (r,t') W (r',t, r),
\]

where \( W \) is the screened interaction, and the non-interacting Green’s function \( G^0 \) is decomposed in components of orbital Green’s functions \( g_{m\sigma} \). These can, after a transformation to real frequencies, be written in terms of the one-particle orbitals \( \varphi_{m\sigma} \) and energies \( \varepsilon_{m\sigma} \),

\[
g_{m\sigma} (r,t',r') = \varphi_{m\sigma} (r,t') \varphi^*_{m\sigma} (r'), \omega - \varepsilon_{m\sigma} \pm i\delta,
\]

with \(- (+)\) for an occupied (unoccupied) state.

The self-screening correction scheme for the GW[8] is based on identifying that an electron in a state \( \varphi_{m\sigma} \) takes part in the screening process of itself in Eq. (1). This occurs when the screened interaction \( W \) is calculated within RPA through the inclusion of all propagators \( g_{m\sigma} \) in the bubble diagram. When calculating the self-energy contribution from an orbital Green’s function, the same \( g_{m\sigma} \) which is describing the propagating electron is present also in the polarization bubble, as demonstrated in Fig. (b). This leads to an unphysical self-screening of the electron. To remove this self-screening, spin-dependent auxiliary functions \( W_{m\sigma} \), calculated within RPA, are defined for each orbital \( m \) where the screening from the electron associated with orbital \( m \) has been removed. These auxiliary screened interactions are subsequently connected with the corresponding orbital Green’s function components in the calculation of the self-energy.

Schematically, the procedure is formulated as follows. First a Green’s function component from orbital \( m\sigma \) is removed from the full non-interacting Green’s function, \( G^0_{m\sigma} = G^0_{\sigma} - g_{m\sigma} \). A new polarization function for each state \( m\sigma \) can then be calculated within RPA with the bubble diagram not including the screening from the \( g_{m\sigma} \) line

\[
P^0_{m\sigma} = -i \left[ \varepsilon^0_{m\sigma} G^0_{m\sigma} + G^0_{-\sigma} G^-_{m\sigma} \right],
\]

Note that the full contribution to the polarization from the other spin channel is kept, as an electron with opposite spin in orbital \( m \) can participate in the screening process. Thereafter, corresponding auxiliary response functions can be obtained as usual through

\[
R_{m\sigma} = P^0_{m\sigma} + P^0_{m\sigma} v R_{m\sigma},
\]

and finally the orbital- and spin-dependent screened interactions with the self-screening removed are obtained from

\[
W_{m\sigma} = v + v R_{m\sigma} v,
\]

with \( v \) being the bare interaction. The final step to eliminate the unphysical screening from \( g_{m\sigma} \) of itself in the self-energy diagram is by associating in Eq. (1) a \( g_{m\sigma} \) to its corresponding \( W_{m\sigma} \). This reflects the fact that an electron associated with orbital \( m\sigma \) experiences a screened interaction \( W_{m\sigma} \) in which the contribution to the screening coming from an electron in the same state now has been removed. The corrected self-energy then finally takes the form:

\[
\Sigma^{GW-ss}_\sigma = i \sum m g_{m\sigma} W_{m\sigma}.
\]
It has been shown that this self-screening correction scheme partially corresponds to including exchange diagram vertex corrections to the self-energy.\textsuperscript{3}

**B. Self-polarization corrected RPA and GW**

Another view on the self-screening error proposed in Ref.\textsuperscript{3} is to consider it as arising from an electron-hole excitation, or a dipole. This so-called self-polarization (to distinguish it from the approach in the previous section), will be outlined next.

To see where the self-polarization error would occur, we decompose the polarization in terms of the individual electron-hole excitations of the system

\[ P^0(\mathbf{r}, \mathbf{r}', \omega) = \sum_{\alpha} p_{\alpha}(\mathbf{r}, \mathbf{r}', \omega). \]  

(8)

\[ p_{\alpha}(\mathbf{r}, \mathbf{r}', \omega) = \frac{\varphi_{m\sigma}(\mathbf{r})\varphi_{n\sigma}^*(\mathbf{r}')\varphi_{m\sigma}(\mathbf{r}')\varphi_{n\sigma}(\mathbf{r})}{\omega - \varepsilon_m + \varepsilon_n + i\delta} - \frac{\varphi_{m\sigma}(\mathbf{r}')\varphi_{n\sigma}^*(\mathbf{r}')\varphi_{m\sigma}(\mathbf{r})\varphi_{n\sigma}(\mathbf{r})}{\omega + \varepsilon_m - \varepsilon_n - i\delta}, \]  

(9)

with the total polarization being the sum over all such excitations. $\alpha$ is here a combined index of the occupied state $n$ and unoccupied state $m$, as well as the spin $\sigma$. Similar to the discussion in the previous section where an electron screened itself through $R$ (and hence $W$), the given excitation $p_{\alpha}$ is now involved in the screening of itself. This can be seen by recognizing that $P^0$ appearing in the screening function $[1 - P^0 v]^{-1}$ includes the excitation $\alpha$ and hence self-screens the polarization $p_{\alpha}$:

\[ R = [1 - P^0 v]^{-1} P^0 = \sum_{\alpha} [1 - \sum_{\beta} P_{\beta} v]^{-1} p_{\alpha}. \]  

(10)

As before, a new polarization function can be introduced to remedy this,

\[ \tilde{P}_\alpha = P^0 - p_{\alpha}, \]  

(11)

with the excitation $\alpha$ removed. By replacing the full polarization $P^0$ in the screening function $[1 - P^0 v]^{-1}$ in Eq. (10) with $\tilde{P}_\alpha$ for each excitation, a new response function corrected for the self-polarization is obtained:

\[ R^{\text{RPA-SP}} = \sum_{\alpha} [1 - \tilde{P}_{\alpha} v]^{-1} p_{\alpha}. \]  

(12)

With this corrected response function, the calculation of the screened interaction and self-energy then follows as in the usual GWA. Also this form of a self-screening correction can be regarded as an approximate inclusion of higher order exchange diagrams.\textsuperscript{3}

**C. Self-screening active-space approximation**

In order to apply the self-screening correction in Sec. II A to ab initio calculations for real materials, where in principle the knowledge of an auxiliary screened interaction would be needed for all states $m\sigma$, some level of approximation is necessary to make the method numerically advantageous. This becomes clear as a large number of unoccupied bands can be required for convergence (as is the case for, e.g., wurtzite $\text{ZnO}$, see Refs.\textsuperscript{9,11}). Calculating a screened interaction for each band would quickly become highly impractical from a computational perspective. We therefore propose to use an approximate form of the method where the self-screening correction is applied only to states in a smaller subspace of the full band structure around the region of interest. We denote this by $S$. The self-energy in Eq. (7) is then modified as, separating out the unaffected exchange contribution $\Sigma^e$ explicitly,

\[ \Sigma^e = \Sigma^e + \sum_m g_{m\sigma} \tilde{W}^e_{m\sigma} \]  

(13)

where the self-screening correction is applied in the calculation of the correlation part of the screened interaction $\tilde{W}^e_{m\sigma}$, only if the band $m$ belongs to $S$. Formally, the screened interaction in the sum in Eq. (13) then becomes

\[ \tilde{W}^e_{m\sigma} = \left\{ \begin{array}{ll} W^e_{m\sigma} & \text{if } m \in S \\ W^c & \text{if } m \notin S \end{array} \right. \]  

(14)

($W^c$ here indicates that the correlation part of the screened interaction is otherwise calculated using RPA within the conventional GWA when $m$ does not belong to the subspace $S$.) This approach effectively turns the size of the chosen subspace into a controllable convergence parameter.

We have implemented the scheme described in this section in the GW code $\text{SPEX}$.\textsuperscript{12} The preceding DFT calculations in the local density approximation (LDA) were performed using the full-potential linearized augmented plane-wave (FLAPW) code $\text{FLEUR}$. The resulting Kohn-Sham eigenfunctions and eigenvalues are then used to construct the starting $G^0$ for the one-shot GW calculations. To check for convergence in the studied properties within our scheme, the subspace $S$ was systematically increased as described in Sec. III C.

**III. RESULTS AND DISCUSSION**

**A. Model calculations**

1. The Hubbard dimer

In the previous work\textsuperscript{3} the self-screening correction was compared to the GWA and exact results for a simple one-orbital Hubbard dimer model calculation, whereas
the effect of the self-polarization correction has not been investigated previously. There it was found that the self-screening correction correctly describes the HOMO-LUMO gap for the Hubbard dimer with one orbital per site and two electrons (half-filled) in the limit of small \((U_0 - U_1)/2t\), while GW predicts a twice too large contribution from the correlation part of the self-energy. This result indicates that the self-screening correction \((GW^{ss})\) corrects the GW deficiency in the weakly-to-moderately correlated regime for this simple model. For the self-polarization correction \((GW^{sp})\) an equivalent derivation gives the gap

\[
\Delta E^{GW-sp} = 2t + U_1 + \frac{2t}{\Delta_{sp}} \left( \frac{(U_0 - U_1)^2}{2t + \Delta_{sp}} \right),
\]

which, in the limit of small \((U_0 - U_1)/2t\), yields the same (incorrect) overestimation as in conventional GW. \(U_0\) and \(U_1\) are the on- and intersite interactions respectively, \(t\) the hopping parameter, and the excitation energy is \(\Delta_{sp} = \sqrt{4t^2 + 2(U_0 - U_1)}\). For the Hubbard dimer, the intersite interaction can actually be absorbed into the onsite one but we have kept it to see its effects more explicitly.

Taking now instead the other limiting case of small hopping \(t\), which physically can be interpreted as increasing the separation between the two sites moving towards an isolated atomic picture, the exact gap reduces to \(\lim_{t \to 0} \Delta E^{exact} = U_0\). We note that this becomes independent of the intersite Coulomb interaction \(U_1\). Both the GW and GW-ss results, however, reduce to \((U_0 + U_1)/2\), whereas intriguingly the GW-sp correction correctly follows the exact result as \(\lim_{t \to 0} \Delta E^{GW-sp} = U_0\). Even when the intersite interaction is negligible, the gap predicted by GW and GW-ss is still found to be half that of the exact and GW-sp results. Similarly, in the regime of large \(U_0\) the self-polarization correction predicts the gap closest to the exact result. Taken together, it is indicative for that the two correction schemes are complementary and valid in different regimes.

Two further effects coming from the correction schemes not previously discussed are how they change the excitation energies of the \(N\)-electron system, obtained from the response function \(R\) and the \((N \pm 1)\)-electron excitation energies obtained from the renormalized Green’s function. The two-electron excitation energy is found to be in almost perfect agreement with the exact result for both of the correction schemes \([2]\) in the regime \((U_0 - U_1)/t < 1\), while the excitation energy found from the conventional RPA is both qualitatively and quantitatively incorrect. This is indicated in the inset of Fig. 2.

With increasing interaction strength all methods deviate from the exact excitation energies. At \(U_0/t > 1.5\) the energies within RPA is in a better agreement compared to the correction schemes, although with the same qualitatively wrong large interaction behaviour. This more “accidental” improvement in RPA at large \(U_0/t\) is to be compared with the good agreement in the low-to-moderate interaction strength regime for the two correction schemes.

The positions of the quasiparticle peaks in the renormalized Green’s functions (not shown) are close to the exact result for all methods at low-to-moderate interaction strengths, however, both correction schemes are found to approximately halve the incorrect overestimation of the satellite positions in GW. In the large interaction regime the self-polarization correction predicts the peak closest to the exact result, while simultaneously both GW-ss and GW-sp maintain their improved description of the satellites.

2. Two-orbital Hubbard dimer

We next extend the model calculations to a dimer model with two orbitals per site, with a Hamiltonian of the form

\[
H = -\sum_{ij \alpha \beta} t_{ij \alpha \beta} \hat{c}_{i \alpha}^\dagger \hat{c}_{j \beta} + \sum_{\alpha} U_{\alpha} \hat{c}_{i \alpha \uparrow}^\dagger \hat{c}_{i \alpha \downarrow} \hat{c}_{j \alpha \uparrow} \hat{c}_{j \alpha \downarrow}.
\]

where \(\hat{c}_{i \alpha}^\dagger\) (\(\hat{c}_{i \alpha}\)) is the creation (annihilation) operator of an electron with spin \(\sigma\) at site \(i\) in orbital \(\alpha\), and \(t_{ij \alpha \beta}\) the hopping parameters. We have restricted the interaction \(U_{\alpha} = U_0\) to be non-zero only for the lower orbital \(\alpha = 1\). For simplicity, we here only vary the interaction strength while fixing the hopping parameters in units of the inter-site hopping \(t_{11,11} = t_{21,11} = t = 1\), with the specific choice of \(t_{12,22} = t_{22,12} = 0.5t\), \(t_{11,22} = t_{22,11} = 0.2t\), and an orbital energy separation \(t_{13,11} - t_{23,12} = 2t\) used for the figures in this section (the intra-site hopping \(t_{13,i, \beta} = \alpha \neq \beta\), is absorbed as a shift in the orbital energy separation). We have checked that other choices of parameters do not affect the discussions.

In the regime of lower interaction strengths \((U_0/t < 1)\) the correction schemes provide an excellent agreement with the exact result for the excitation energies obtained.
In contrast, we have not observed any non-causality in any of the calculations in this work based on the self-screening correction. As the corrections in spirit are very similar, we propose a simple picture for the origin by focusing on the effect of the corrections from the perspective of a single occupied state \( \varphi_{m\sigma} \).

The correction in the self-screening scheme amounts to the state \( \varphi_{m\sigma} \) being completely removed in the screening of itself, as discussed previously, by virtue of it being completely removed from the polarization propagator \( P_{m\sigma}^{0} \) in Eq. (4). On the other hand, in the self-polarization scheme what is removed from the screening through the polarization \( P_{\alpha} \) is an electron-hole excitation with the occupied state \( \varphi_{m\sigma} \) and an unoccupied state \( \varphi_{n\sigma} \). Through the calculation of \( R \) in Eq. (12) it is then clear that it is this excitation which is effectively prohibited from screening itself, not the physical states involved. This in turn leads to only a partial removal of the physical self-screening coming from the electron in state \( \varphi_{m\sigma} \), as it is still involved in the screening of itself through another excitation involving the unoccupied state \( \varphi_{n'\sigma} \).

To make the physical picture clearer, we can draw a comparison to the model dimer calculations. In the case of the single-orbital dimer all results remain causal, since there exists only one unoccupied (antibonding) state that the bonding state, occupied by two electrons of opposite spin, can be excited to. When this excitation is removed, the only state involved in the screening process is for the electron with opposite spin. If the system instead has three unoccupied orbitals into which the electrons can be excited, as is the case for the two-orbital dimer, the issue becomes clear; whereas in the self-screening corrected (auxiliary) polarization \( P_{m\sigma}^{ss} \) all instances of the occupied state \( \varphi_{m\sigma} \) have been removed, the self-polarization scheme provides three different \( P_{\alpha} \), all being excitations from the occupied \( \varphi_{m\sigma} \) to the three unoccupied states. In the summation over excitations in Eq. (12),

\[
R^{\text{RPA-sp}} = \sum_{\alpha} [1 - P_{\alpha} v]^{-1} p_{\alpha}
\]

the occupied \( \varphi_{m\sigma} \) is then only partially removed for each term, as the other excitations retain its presence in \( P_{\alpha} \). We therefore believe we can attribute the non-causality observed for the self-polarization correction to be due to a partial removal of the physical state from screening itself. This observation could be used as a starting point to develop a causal self-polarization scheme, and we plan to investigate this in a future work.

C. Ab initio calculations

Next we move beyond the model calculations of the previous sections to test the reliability of the self-screening correction also for predicting properties of real materials. We have carried out \textit{ab initio} calculations for several semiconductors and focused on two well-characterized problems for a “one-shot” \( GW \) \( (G^{0}W^{0}) \)
Table I. Band gaps in eV calculated within LDA, GW, and self-screening corrected GW. The self-screening correction was applied with 20 and 40 bands respectively included in the subspace $S$. The experimental values are taken from Refs. [15–17] as indicated. The values in parenthesis for Ge and Si give the direct gap at the Γ-point, and the values in square brackets for ZnO show the extrapolated band gaps discussed in the text.

|       | LDA | GW   | GW−ss (20) | GW−ss (40) | Exp. |
|-------|-----|------|------------|------------|------|
| Ge    | -   | 0.70 (0.80) | 0.79 (0.89) | 0.79 (0.89) | 0.74 (0.90) |
| Si    | 0.47 (2.52) | 1.00 (3.16) | 1.12 (3.25) | 1.12 (3.25) | 1.17 (3.37) |
| InP   | 0.47 | 1.24 | 1.36 | 1.36 | 1.42 |
| GaAs  | 0.30 | 1.38 | 1.50 | 1.50 | 1.5 |
| CdTe  | 0.51 | 1.43 | 1.55 | 1.55 | 1.6 |
| ZnSe  | 1.04 | 2.42 | 2.60 | 2.60 | 2.8 |
| ZnO   | 0.76 | 2.41 [2.79] | 2.56 [2.94] | 2.57 [2.95] | 3.44 |

calculation: the band gap and the position of the more localized semicore states. For these calculations we used the self-screening correction with the implementation as described in Sec. II C. The band gap is known to be significantly improved within the GWA compared to DFT calculations, although there is still a slight underestimation for many materials.[15–20] For several semiconductors and insulators this can be improved by introducing a limited form of self-consistency in the so-called quasiparticle self-consistent GW.[20] Semicore states calculated within the GWA, albeit improved over those obtained from LDA calculations, are found to be still located too high in energy.[20]

The quasiparticle energies were obtained by finding the solution to the quasiparticle equation[15]

$$E_{nk}^{QP} = \varepsilon_{nk}^{DFT} + \Sigma_{nk}(E_{nk}^{QP}) - V_{nk}^{XC}$$  \hspace{1cm} (17)

for a state $nk$ where the self-energy corrects the exchange-correlation potential $V^{XC}$ contained in the one-particle energies $\varepsilon_{nk}^{DFT}$. To benchmark the correction we have applied it to several semiconductors with varying band gaps: Si, GaAs, Ge, CdTe, InP, ZnSe, and wurtzite ZnO, where the six latter ones are known to also have high-lying $d$ semicore states. A $8 \times 8 \times 8$ k-grid was used for the calculations (12 $\times$ 12 $\times$ 12 in the case of Si to more accurately estimate the location of the indirect bandgap). Both the GW and GW−ss calculations are performed within $G^{\omega}W^{0}$, where the non-interacting Green’s function is obtained directly from a preceding DFT calculation, and the Green’s function is not subsequently updated in further iterations. The band gaps calculated for the semiconductors within LDA, GW, and GW−ss are listed in Table I and compared with the experimental values. To test the convergence of the approximate treatment of the self-screening correction, we carried out convergence tests by systematically increasing the active region $S$. We find rapidly converging results already with the 20 lowest bands included in the subset treated with the correction, counting from the high-lying semicore $d$ states. No change is observed when $S$ is further doubled in size to the 40 lowest bands. This important result we believe justifies the approximate treatment, and shows the method to be numerically tractable for realistic material calculations.

For all semiconductors, we find the band gaps which are underestimated within GW to be increased when the correction scheme is applied, bringing them to a closer agreement with the experimental values. The only exception is Ge which is predicted to be a metal in the LDA calculation, where the band gap is already relatively well described by GW. The use of the correction instead leads to a slight overestimation of the gap, with the error being comparable to the one in conventional GW. The observed trend of increasing gaps is in agreement with the expectation that a reduction in screening would bring the situation more towards a Hartree-Fock picture (here the self-energy contains only the exchange part, $\Sigma^x$, neglecting screening effects), where it is well-known that too large band gaps are found.[19–22]

It has previously been shown[10,23] that a very large number of unoccupied bands is required to converge the band gap for wurtzite ZnO. Assuming our approximate treatment of the self-screening correction is fully converged within the chosen $S$ subset of bands, by increasing the number of untreated unoccupied bands a similar situation as for a usual GW calculation should occur. We therefore follow Friedrich et al. in Ref. [10] and use a hyperbolical fit for $N$ bands of the form

$$E_{gap}(N) = \frac{a}{N - N_0} + b,$$  \hspace{1cm} (18)

with parameters $a$, $b$, and $N_0$ extracted from the corrected fit in Ref. [11]. As they similarly used the FLAPW method in their work, we believe our results should be comparable. Indeed, by extrapolating the hyperbolical fit shifted to match our values calculated at $N = 500$ bands, we find a good agreement between our GW asymptote at 2.79 eV and the previously reported value of the band gap at 2.83 eV [11]. The similarly extrapolated gap obtained from the self-screening corrected results brings our final
conclusions regarding the differences between the correction schemes:

1. The self-polarization scheme corrects the HOMO-LUMO gap in the localized (strongly correlated) regime, whereas for low interaction strengths it reduces to the GW result. This can be compared with the self-screening correction, which instead reproduces the exact result in the low-to-moderate interaction (delocalized) regime as shown earlier in Ref. [8].

2. The self-polarization scheme suffers from causality violations, not observed for the self-screening correction. We have proposed an origin of this issue from a picture based on an occupied physical state, where the self-screening is only partly removed from the different excitations in the system.

The two schemes also display some similarities. Specifically, we find that the excitation energies are well described for both the self-screening and self-polarization schemes in the low-to-moderate interaction strength regime. Further, we observe that both correction schemes predict an improved position of the satellite features in the spectral function, as compared with the conventional GWA. Taken together, this indicates that the two formulations could be considered complementary in more delocalized (self-screening) and localized (self-polarization) cases.

The \textit{ab-initio} results demonstrate that the error in the band gap is decreased notably when the self-screening correction is applied, predicting values within only a few percent from experiments in most cases. That the more delocalized \textit{sp} states involved around the band gap have a large part of the remaining error in the GWA compared to experiment removed agrees with what we would expect based on the model calculations. Similarly, the more localized semicore states show no noticeable changes from experiment removed agrees with what we would expect based on the model calculations. Based on these findings, we believe that it would be of interest to apply

| d states | LDA | GW | GW\textsuperscript{-ss} \((d)\) | GW\textsuperscript{-ss} \((20)\) | GW\textsuperscript{-ss} \((40)\) | Exp. |
|----------|-----|----|-----------------|-----------------|-----------------|-----|
| Ge       | 24.7| 27.0| 27.0            | 27.0            | 27.0            | 29.6|
| InP      | 14.2| 15.5| 15.5            | 15.5            | 15.5            | 17.3|
| GaAs     | 14.8| 16.5| 16.7            | 16.6            | 16.6            | 18.8|
| CdTe     | 8.0 | 8.9 | 9.00            | 8.9             | 8.9             | 10.5|
| ZnSe     | 6.4 | 7.4 | 7.5             | 7.4             | 7.4             | 9.3 |
| ZnO      | 5.1 | 5.8 | 5.8             | 5.7             | 5.7             | 8.9 |

In Table II we finally present our calculated center-of-weight values of the \textit{d} semicore states, averaged over bands and the Brillouin zone, in GaAs, Ge, InP, CdTe, ZnSe, and ZnO relative to the valence band maximum. We observe a very minor improvement compared to the GW values for some materials when only treating the semicore states with the self-screening correction. This small change is cancelled and even reversed when increasing the number of bands treated. On closer inspection, however, this can be explained by the reported widening of the band gap in Table II. Instead, comparing the position with the conduction band minimum, or the center of the gap, conversely shows a slight further improvement in the energy position. This change is, however, negligible, and shows that the localized semicore \textit{d} states remain mostly unaffected by the self-screening correction.

IV. CONCLUSIONS

We have applied two recently proposed self-screening corrections to RPA (self-polarization) and the GWA (self-screening) in model systems to further investigate the effects and regions of validity of the schemes compared with the conventional methods. We have furthermore employed one of the schemes, the self-screening correction, in \textit{ab-initio} calculations for quantitative predictions of a number of semiconductors, specifically studying its effect on the band gaps and the position of the semicore \textit{d} states. In order to apply the correction to realistic materials calculations we proposed an approximate scheme, where we introduce an “active-space” with only a subset of the full underlying DFT band structure used in the construction of \(G^0\) being treated with the correction. We have shown that this approximate treatment quickly converges in the number of required bands for the properties studied, justifying it as a numerically advantageous method.

From our model calculations we can draw two clear conclusions regarding the differences between the correction schemes:

1. The self-polarization scheme corrects the HOMO-LUMO gap in the localized (strongly correlated) regime, whereas for low interaction strengths it reduces to the GW result. This can be compared with the self-screening correction, which instead reproduces the exact result in the low-to-moderate interaction (delocalized) regime as shown earlier in Ref. [8].

2. The self-polarization scheme suffers from causality violations, not observed for the self-screening correction. We have proposed an origin of this issue from a picture based on an occupied physical state, where the self-screening is only partly removed from the different excitations in the system.

The two schemes also display some similarities. Specifically, we find that the excitation energies are well described for both the self-screening and self-polarization schemes in the low-to-moderate interaction strength regime. Further, we observe that both correction schemes predict an improved position of the satellite features in the spectral function, as compared with the conventional GWA. Taken together, this indicates that the two formulations could be considered complementary in more delocalized (self-screening) and localized (self-polarization) cases.

The \textit{ab-initio} results demonstrate that the error in the band gap is decreased notably when the self-screening correction is applied, predicting values within only a few percent from experiments in most cases. That the more delocalized \textit{sp} states involved around the band gap have a large part of the remaining error in the GWA compared to experiment removed agrees with what we would expect based on the model calculations. Similarly, the more localized semicore states show no noticeable changes from the correction, strengthening our supposition that the self-screening correction is more suitable for improving the description of delocalized states. Based on these findings, we believe that it would be of interest to apply
also the self-polarization scheme in realistic calculations of materials; specifically, the more localized $d$ semicore states could be expected to show an improvement. However, the issue of non-causality would first have to be addressed. The proposed origin of this issue, an incomplete removal of the self-screening from a given physical state, could provide a starting point for an improved scheme.

To summarize, it is well-known that RPA and GW based methods overall work well for more delocalized systems. Our calculations for both models and real materials indicate that a significant part of the remaining error in a conventional GWA calculation for such delocalized states can be reduced by considering a correction of the self-screening error coming from the RPA. While self-screening can be expected to also be significant for a localized state, our model calculations and ab-initio predictions of $d$ semicore state positions indicate that some form of the so-called self-polarization correction could prove more suitable in this regime. To further verify the usefulness of the self-screening correction, investigations of additional classes of materials would be of interest.

ACKNOWLEDGMENTS

V.C. acknowledges support from the European Research Council through ERC Consolidator Grant No. 724103. F. A. gratefully acknowledges financial support from the Knut and Alice Wallenberg (KAW) Foundation Grant No. 2017.0061 and the Swedish Research Council, Vetenskapsrådet (VR) Grant. No. 2021-04498_3. The calculations were performed on the Beo05 cluster at the University of Fribourg.

1. D. Pines, *Elementary Excitations in Solids* (W. A. Benjamin, New York, 1963).
2. L. Hedin, Phys. Rev. 139, A796 (1965).
3. P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
4. W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
5. W. Nelson, P. Bokes, P. Rinke, and R. W. Godby, Phys. Rev. A 75, 032505 (2007).
6. P. Romaniello, S. Guyot, and L. Reining, J. Chem. Phys. 131, 15411 (2009).
7. J. Wetherell, M. J. P. Hodgson, and R. W. Godby, Phys. Rev. B 97, 121102 (2018).
8. F. Aryasetiawan, R. Sakuma, and K. Karlsson, Phys. Rev. B 85, 035106 (2012).
9. B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. 105, 146401 (2010).
10. C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B 83, 081101 (2011).
11. C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B 84, 039906 (2011).
12. C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B 81, 125102 (2010).
13. The FLEUR group, “The FLEUR project,” https://www.flapw.de
14. The response functions in the self-screening correction are merely auxiliary quantities, and for the one-orbital dimer they are simply related to the self-polarization response as $R_{\text{ss}}^{\text{RPA}} = R^{\text{sp}}/2$. The predicted $N$-particle excitation energy is therefore the same.
15. O. Madelung, U. Rössler, and M. Schulz, eds., *Group IV Elements, IV-IV and III-V Compounds. Part b - Electronic, Transport, Optical and Other Properties, Landolt-Börnstein - Group III Condensed Matter, Vol. 41A1β* (Springer-Verlag, Berlin Heidelberg, 2002).
16. T. C. Chiang and F. J. Himpsel, in *Electronic Structure of Solids: Photoemission Spectra and Related Data, Landolt-Börnstein - Group III Condensed Matter, Vol. 23A*, edited by A. Goldmann and E.-E. Koch (Springer-Verlag, Berlin Heidelberg, 1989).
17. O. Madelung, U. Rössler, and M. Schulz, eds., *II-VI and I-VII Compounds; Semimagnetic Compounds, Landolt-Börnstein - Group III Condensed Matter, Vol. 41B* (Springer-Verlag, Berlin Heidelberg, 1999).
18. M. S. Hybertsen and S. G. Louie, Phys. Rev. B 34, 5390 (1986).
19. R. Hott, Phys. Rev. B 44, 1057 (1991).
20. M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. 96, 226402 (2006).
21. F. Aryasetiawan and O. Gunnarsson, Phys. Rev. B 54, 17564 (1996).
22. A. Svane, Phys. Rev. B 35, 5496 (1987).
23. G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. 74, 601 (2002).