PHYSICAL REVIEW B **107**, 125105 (2023)

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

(Received 2 June 2022; revised 12 December 2022; accepted 26 January 2023; published 2 March 2023)

The self-screening error in the random-phase approximation 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. Here, we 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 a local form of the 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, while we additionally observe that it suffers from causality violations in the strongly correlated regime. The self-screening correction, on the other hand, significantly improves the description in more delocalized states. It is found to provide 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 with experiment, while leaving the localized semicore *d* states mostly unaffected.

DOI: 10.1103/PhysRevB.107.125105

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 calculations of 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) [1], which is also the basic approximation used in the well-known *GW* approximation (GWA) [2] to compute the screened interaction W starting from the Kohn-Sham states and eigenvalues obtained from a density functional theory (DFT) [3,4] calculation. The GWA has become routine in the calculation of the quasiparticle dispersion in materials and has proven highly successful in the prediction of band gaps in semiconductors and insulators [5–8].

A well-known and longstanding 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 [9]; with only one electron in the system, it is not possible to screen the interaction between the electron and an external field upon electron removal since there are no other electrons present, and yet the RPA erroneously yields a nonzero response function. This leads to a nonzero contribution from the correlation part of the self-energy for the electron removal spectrum when calculated within the GWA, while it evidently should be zero. The source of this error arises from a self-screening process inherent in the RPA. As illustrated in Fig. 1(a), the only electron in the system, represented by a Green's function line, should not participate in the polarization bubble, yet within the GWA, this self-energy diagram is nonzero. In general, however, even with the exact response of the system, the GWA would still provide a nonzero contribution to the correlation self-energy coming from the electron addition part of the spectrum.

The self-screening error has received attention in recent years, and a few different approaches to treat it have been proposed in the literature. Romaniello *et al.* [10] used vertex corrections to the self-energy beyond the RPA and applied it to simple model systems. Wetherell *et al.* [11] proposed the use of a local potential based on DFT added to the self-energy and tested it successfully for simple one-dimensional models. In an earlier work [12], 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 (highest occupied molecular orbital–lowest unoccupied molecular orbital) gap of a hydrogen dimer was correctly reproduced in the weakly to moderately correlated regime.

In this paper, we will employ the schemes proposed in Ref. [12] to consider a correction to the self-screening error. The idea is that, in a given self-energy diagram, the Green's function line representing the propagation of an electron or a hole in a particular orbital is removed from the Green's function when calculating the polarization, schematically shown in Fig. 1. In this way, the screening processes associated with this electron or hole are removed from the total polarization, at the cost of introducing an orbital- and spin-dependent interaction.

---

[*] viktor.christiansson@unifr.ch

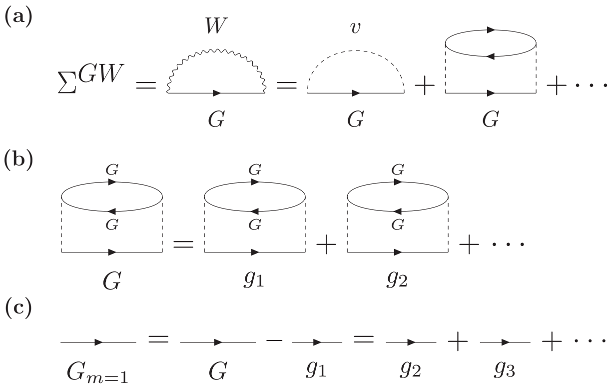

FIG. 1. (a) Contributions to the self-energy diagram within the *GW* approximation (GWA). (b) Expansion of Green's function in orbital Green's functions g_m for one diagram. The presence of the full G in the polarization bubble reveals the screening on the propagating electron coming from itself, as discussed in the text. (c) The resulting Green's function with orbital Green's function g_1 removed. In the self-screening correction scheme, it replaces G in the polarization bubble in the first term of (b).

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 what is termed *self-polarization*.

The aim of this paper is twofold. 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 in 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 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 band gaps and semicore quasiparticle energies are compared with conventional *GW* calculations and experimental values.

This paper is organized as follows. In Sec. II, we introduce the self-screening and self-polarization corrections following the derivation given in Ref. [12] and describe a method for making the self-screening correction tractable for the calculation of real materials. In Secs. 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 with 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 σ is calculated as

$$\begin{aligned}\Sigma_\sigma^{GW}(\mathbf{r}t,\mathbf{r}'t') &= iG_\sigma^0(\mathbf{r}t,\mathbf{r}'t')W(\mathbf{r}'t',\mathbf{r}t)\\ &= i\sum_m g_{m\sigma}(\mathbf{r}t,\mathbf{r}'t')W(\mathbf{r}'t',\mathbf{r}t),\end{aligned} \tag{1}$$

where W is the screened electron-electron interaction, and the noninteracting Green's function G_σ^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 $\varepsilon_{m\sigma}$:

$$g_{m\sigma}(\mathbf{r},\mathbf{r}',\omega) = \frac{\varphi_{m\sigma}(\mathbf{r})\varphi_{m\sigma}^*(\mathbf{r}')}{\omega-\varepsilon_{m\sigma}\pm i\delta}, \tag{2}$$

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

The self-screening correction scheme for the GWA [12] is based on recognizing 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 describes the propagating electron is present also in the polarization bubble, as demonstrated in Fig. 1(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 (hole) 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 noninteracting Green's function:

$$G_{m\sigma}^0 = G_\sigma^0 - g_{m\sigma}. \tag{3}$$

A new polarization function for each state $m\sigma$ can then be calculated within RPA with the bubble diagram excluding the screening from the $g_{m\sigma}$ line:

$$P_{m\sigma}^0 = -i\big[G_{m\sigma}^0G_{m\sigma}^0 + G_{-\sigma}^0G_{-\sigma}^0\big]. \tag{4}$$

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

$$R_{m\sigma} = P_{m\sigma}^0 + P_{m\sigma}^0vR_{m\sigma}, \tag{5}$$

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

$$W_{m\sigma} = v + vR_{m\sigma}v, \tag{6}$$

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) $g_{m\sigma}$ with 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 has been removed. The corrected self-energy then finally takes the form:

$$\Sigma_\sigma^{GW\text{-ss}} = i\sum_m g_{m\sigma}W_{m\sigma}. \tag{7}$$

This self-screening correction scheme has been shown to partially correspond to including exchange diagram vertex corrections to the self-energy [12].

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

A further self-screening error arises from an electron-hole excitation (dipole) screening itself [12]. This so-called self-polarization (to distinguish it from the self-screening described in the previous section) will be outlined next.

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

$$P^0(\mathbf{r},\mathbf{r}',\omega)=\sum_\alpha p_\alpha(\mathbf{r},\mathbf{r}',\omega), \tag{8}$$

$$p_\alpha(\mathbf{r},\mathbf{r}',\omega)=\frac{\varphi_{m\sigma}(\mathbf{r})\varphi^*_{n\sigma}(\mathbf{r})\varphi^*_{m\sigma}(\mathbf{r}')\varphi_{n\sigma}(\mathbf{r}')}{\omega-\varepsilon_m+\varepsilon_n+i\delta} -\frac{\varphi_{m\sigma}(\mathbf{r}')\varphi^*_{n\sigma}(\mathbf{r}')\varphi^*_{m\sigma}(\mathbf{r})\varphi_{n\sigma}(\mathbf{r})}{\omega+\varepsilon_m-\varepsilon_n-i\delta}, \tag{9}$$

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

$$R=[1-P^0v]^{-1}P^0=\sum_\alpha\left[1-\sum_\beta p_\beta v\right]^{-1}p_\alpha. \tag{10}$$

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

$$P_\alpha=P^0-p_\alpha, \tag{11}$$

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

$$R^{\text{RPA-sp}}=\sum_\alpha[1-P_\alpha v]^{-1}p_\alpha. \tag{12}$$

Using this corrected response function, the calculation of the screened interaction and self-energy then follows as in the usual GWA. Also, this form of self-screening correction can be regarded as an approximate inclusion of higher-order exchange diagrams, in this case to the polarization function [12].

Although both the self-screening and self-polarization corrections are seemingly applied to the polarization (response) function, the self-screening correction is instead to be regarded as a correction directly to the self-energy; the $P_{m\sigma}$'s are merely auxiliary quantities for each state in this case, used only to identify a physical origin for the self-screening, and in this sense, there is no longer a single polarization function within the self-screening scheme. The self-polarization correction, on the other hand, is a correction directly to the RPA and therefore yields a self-screening-corrected response function, which replaces the common RPA response when calculating the screened interaction used in the GWA.

C. Self-screening active-space approximation

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 extended states in a periodic system, some level of approximation is necessary to make the method numerically feasible. This becomes clear as many unoccupied bands can be required for convergence (as is the case for, e.g., wurtzite ZnO, see Refs. [13–15]). Calculating a screened interaction for each band would quickly become highly impractical from a computational perspective. We therefore propose using an approximate form of the method where the self-screening correction is applied only to states in a smaller subspace depending on the region of interest. We denote this by $\mathcal{S}$.

The self-energy in Eq. (7) is then modified as, separating out the unaffected exchange contribution Σ^{x}_σ explicitly,

$$\Sigma_\sigma=\Sigma^{\text{x}}_\sigma+\sum_m g_{m\sigma}\tilde{W}^{\text{c}}_{m\sigma}, \tag{13}$$

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

$$\tilde{W}^{\text{c}}_{m\sigma}=\begin{cases}W^{\text{c}}_{m\sigma} & \text{if } m\in\mathcal{S}\\ W^{\text{c}} & \text{if } m\notin\mathcal{S}.\end{cases} \tag{14}$$

Here, W^c indicates that the screened interaction is otherwise calculated using RPA (as in the conventional GWA) when m does not belong to the subspace $\mathcal{S}$. This approach effectively turns the size of the chosen subspace into a controllable convergence parameter.

To treat the extended states, we write the Green's function in terms of the Bloch states $\psi_{\mathbf{k}n}$ as

$$G^0_\sigma(\mathbf{r},\mathbf{r}',\omega)=\sum_{\mathbf{k}n}\frac{\psi_{\mathbf{k}n}(\mathbf{r})\psi^*_{\mathbf{k}n}(\mathbf{r}')}{\omega-\varepsilon_{\mathbf{k}n}}=\sum_{\mathbf{k}n}g_{n\mathbf{k}}(\mathbf{r},\mathbf{r}',\omega), \tag{15}$$

where the infinitesimal factor $\pm i\delta$ in the denominator has been dropped for simplicity. Following Ref. [12], we can expand the Bloch states in a Wannier basis $\chi_{m\mathbf{R}}$, leading to the expression for the self-screening-corrected Green's function:

$$G^0_n(\mathbf{r},\mathbf{r}',\omega)=G^0(\mathbf{r},\mathbf{r}',\omega)-\sum_{\mathbf{k}}\sum_{mm'}\sum_{\mathbf{R}}\frac{U_{nm}(\mathbf{k})\chi_{m\mathbf{R}}(\mathbf{r})\chi^*_{m'\mathbf{R}}(\mathbf{r}')U^\dagger_{m'n}(\mathbf{k})}{\omega-\varepsilon_{\mathbf{k}n}}. \tag{16}$$

One can then make the approximation, assuming a sufficiently narrow band:

$$\begin{aligned}G^0_n(\mathbf{r},\mathbf{r}',\omega)&\approx G^0(\mathbf{r},\mathbf{r}',\omega)-\sum_{\mathbf{R}}\frac{\chi_{n\mathbf{R}}(\mathbf{r})\chi^*_{n\mathbf{R}}(\mathbf{r}')}{\omega-\sum_{\mathbf{k}}\varepsilon_{\mathbf{k}n}}\\&=G^0(\mathbf{r},\mathbf{r}',\omega)-\sum_{\mathbf{R}}g_{n\mathbf{R}}(\mathbf{r},\mathbf{r}',\omega).\end{aligned} \tag{17}$$

A proper treatment of the localized basis using Wannier functions, as in Eq. (16), would be desirable. Particularly for the more delocalized *sp* states, where the above approximation is less justified, the Wannier approach would be needed to fully validate the scheme and would be an interesting future direction of work.

As a first test of the self-screening correction for realistic materials, however, we have in this paper instead made use of a simplified local version of the self-screening correction. Physically, the screening at a site $\mathbf{R}$ from an electron on another site $\mathbf{R}_1$ would be expected to be less important if the spacing is sufficiently large, and, as a consequence, also the self-screening should be less relevant. We therefore further approximate the self-screening-corrected Green's function as

$$\begin{aligned} G^0_n(\mathbf{r}, \mathbf{r}', \omega) &\approx G^0(\mathbf{r}, \mathbf{r}', \omega) - g^{\text{loc}}_n(\mathbf{r}, \mathbf{r}', \omega) \\ &= G^0(\mathbf{r}, \mathbf{r}', \omega) - \sum_{\mathbf{k}} g_{n\mathbf{k}}(\mathbf{r}, \mathbf{r}', \omega). \end{aligned} \tag{18}$$

Although, here, we have used an approximate treatment that only captures a local part of the self-screening correction, we believe that this is a reasonable compromise due to the large computational simplification offered, especially in light of the phenomenological origin of the correction itself.

We have implemented the scheme described in this section in the *GW* code SPEX [16]. The preceding DFT calculations in the local density approximation (LDA) [3] were performed using the full-potential linearized augmented plane-wave (FLAPW) code FLEUR [17]. The resulting Kohn-Sham eigenfunctions and eigenvalues are 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 $\mathcal{S}$ was systematically increased as described in Sec. III C.

III. RESULTS AND DISCUSSION

A. Model calculations

1. The Hubbard dimer

In the previous work [12], the self-screening correction was compared with the GWA and exact results for a simple one-orbital Hubbard dimer model calculation solved analytically, 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 two-times-too-large contribution from the correlation part of the self-energy. This result indicates that the self-screening correction ($GW^{\text{-ss}}$) corrects the *GW* deficiency in the weakly to moderately correlated regime for this simple model. Performing the equivalent derivation for the self-polarization correction ($GW^{\text{-sp}}$) gives the gap:

$$\Delta E^{GW\text{-sp}} = 2t + U_1 + \frac{2t}{\Delta_{sp}} \frac{(U_0 - U_1)^2}{2t + \Delta_{sp}}, \tag{19}$$

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

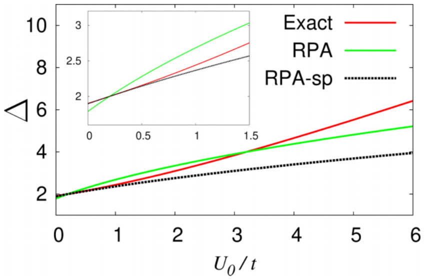

FIG. 2. Exact, random-phase approximation (RPA), and self-polarization-corrected RPA (RPA-sp) excitation energies for the one-orbital Hubbard dimer in units of $t = 1$ for a fixed $U_1 = 0.2$. The inset shows the small interaction limit.

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 toward an isolated atomic picture, the exact gap reduces to $\lim_{t\to 0} \Delta E^{\text{exact}} = U_0$. We note that this becomes independent of the intersite Coulomb interaction U_1. Both the *GW* and $GW^{\text{-ss}}$ results, however, reduce to $(U_0 + U_1)/2$, whereas intriguingly the $GW^{\text{-sp}}$ correction correctly follows the exact result as $\lim_{t\to 0} \Delta E^{GW\text{-sp}} = U_0$. Even when the intersite interaction is negligible, the gap predicted by *GW* and $GW^{\text{-ss}}$ is still found to be half that of the exact and $GW^{\text{-sp}}$ 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 an indication 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 [18] 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 $U_0/t \gtrsim 1.5$, the energies within RPA are in better agreement than the correction schemes, although with the same qualitatively wrong large interaction behavior. 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

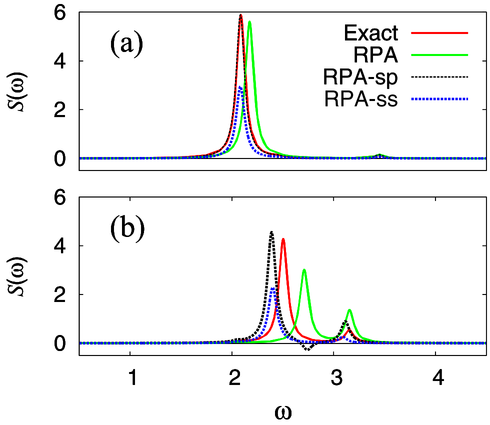

FIG. 3. Diagonal components of the exact, random-phase approximation (RPA), and self-polarization-corrected RPA (RPA-sp) spectral response functions for the two-orbital Hubbard dimer. Also shown is one auxiliary response function calculated using the self-screening correction (RPA-ss). (a) is calculated for an interaction $U_0/t = 0.2$ and (b) for $U_0/t = 1.0$.

simultaneously both $GW^{\text{-ss}}$ and $GW^{\text{-sp}}$ 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:

$$H = -\sum_{ij\alpha\beta} t_{i\alpha,j\beta}\hat{c}^{\dagger}_{i\alpha\sigma}\hat{c}_{j\beta\sigma} + \sum_{i\alpha} U_{\alpha}\hat{c}^{\dagger}_{i\alpha\uparrow}\hat{c}_{i\alpha\uparrow}\hat{c}^{\dagger}_{i\alpha\downarrow}\hat{c}_{i\alpha\downarrow}, \quad (20)$$

where $\hat{c}^{\dagger}_{i\alpha\sigma}$ ($\hat{c}_{i\alpha\sigma}$) is the creation (annihilation) operator of an electron with spin σ at site i in orbital α, and $t_{i\alpha,j\beta}$ the hopping parameters. We have restricted the interaction $U_\alpha = U_0$ to be nonzero only for the lower orbital $\alpha = 1$. For simplicity, here, we only vary the interaction strength while fixing the hopping parameters in units of the intersite hopping $t_{11,21} = t_{21,11} = t = 1$, 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 $t_{i1,i1} - t_{i2,i2} = 2t$ used for the figures in this section (the intrasite hopping $t_{i\alpha,i\beta}$, $\alpha \neq \beta$, 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 good agreement with the exact result for the excitation energies obtained from the response function, as indicated in Fig. 3 (note that within the self-screening correction, the response function is an auxiliary quantity, and a single representative one is shown). At larger interaction strengths, as would be expected, all the RPA-based schemes fail and deviate significantly from the exact results. We can at this point also note another important difference between the two correction schemes: For moderate interaction strengths ($U_0/t \gtrsim 1$), causality violations are seen to emerge from the self-polarization correction, as indicated in Fig. 3(b), where a partially negative spectral response function is found. These were not present in the single-orbital model, and the origin will be further discussed in Sec. III B.

To conclude the model calculations, we will briefly discuss the renormalized Green's function. The quasiparticle peaks for smaller interaction strengths are only marginally renormalized, while the satellite features in the spectral functions also for the two-orbital model are found to be significantly improved in both correction schemes over those calculated within GW. The errors in the satellite positions are again approximately halved for small to intermediate U_0/t, whereas all GW-based methods are worsened significantly for larger interactions. The causality violations appearing in the self-polarization correction scheme are also observed for the spectral function.

B. Causality considerations

As we noted in the previous section, the self-polarization correction can produce noncausal features in the response function, specifically in parameter regimes where the electrons are more localized, leading to similar issues also in the renormalized Green's function. In contrast, we have not observed any noncausality in any of the calculations in this paper based on the self-screening correction. As the corrections in spirit are very similar, we can propose a simple physical picture for the origin by focusing on the effects of the corrections on the screening processes from the perspective of a single occupied state $\varphi_{m\sigma}$.

Within the self-screening scheme, we can think of the correction as completely removing the state $\varphi_{m\sigma}$ from being involved in the screening of itself, as discussed previously, by virtue of it being fully removed from the polarization propagator $P^0_{m\sigma}$ in Eq. (4). On the other hand, in the self-polarization scheme, it is instead an electron-hole excitation that is removed from the screening, involving the occupied state $\varphi_{m\sigma}$ and an unoccupied state $\varphi_{n_1\sigma}$, which we label $\alpha_1 = (m, n_1)$. In this way, it becomes clear that what is prevented from screening itself when calculating R in Eq. (12) is not the physical states involved but rather the single excitation. This in turn leads to that an excitation labeled $\alpha_2 = (m, n_2)$, involving the same $\varphi_{m\sigma}$ as before but another unoccupied state $\varphi_{n_2\sigma}$, is still involved in the screening of α_1. When calculating the self-energy using the self-polarization corrected screened interaction:

$$\Sigma^{GW\text{-sp}} = \sum_m g_m W^{\text{RPA-sp}}, \quad (21)$$

it is now evident that the state $\varphi_{m\sigma}$ is still partially screening itself.

To make the physical picture clearer, we can draw a comparison with 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 difference between the schemes is apparent; while in the self-screening-corrected (auxiliary) polarization $P^{\text{-ss}}_{m\sigma}$ all instances of the occupied state $\varphi_{m\sigma}$ have been removed, the self-polarization scheme gives three different P_α, where all still contain excitations from the occupied $\varphi_{m\sigma}$ to unoccupied states. In the summation over excitations in Eq. (12), the occupied $\varphi_{m\sigma}$ is then only partially removed for each term, as all the other excitations in P_α retain its presence. We therefore believe we can attribute the noncausality observed for the self-polarization correction to an only partial removal of a physical state from screening itself.

For future calculations using the self-polarization correction, a more rigorous understanding is, however, needed to avoid this noncausality, particularly in realistic materials calculations. The above observations could therefore help in the development of a causal self-polarization scheme. A more rigorous route for avoiding noncausality has been proposed in the work of Stefanucci *et al.* [19] and Uimonen *et al.*, [20] which could be very relevant to the issue of noncausality encountered in this paper and could also serve as a starting point for work in this direction.

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 performed the *ab initio* calculations for several semiconductors and focused on two well-characterized problems for a one-shot *GW* (G^0W^0) calculation: the band gap and the position of the more localized semicore states. For these calculations, we used the approximate 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 with DFT calculations, although there is still a slight underestimation for many materials [7,8,21]. 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* (QSGW) [8]. Additionally, semicore states calculated within the GWA, albeit improved over those obtained from LDA calculations, are found to still be located too high in energy [22].

The quasiparticle energies are obtained by finding the solution to the quasiparticle equation [7]:

$$E_{n\mathbf{k}}^{\mathrm{QP}} = \varepsilon_{n\mathbf{k}}^{\mathrm{DFT}} + \Sigma_{n\mathbf{k}}\left(E_{n\mathbf{k}}^{\mathrm{QP}}\right) - V_{n\mathbf{k}}^{\mathrm{xc}}, \tag{22}$$

for a state $n\mathbf{k}$ where the self-energy corrects the exchange-correlation potential V^{xc} contained in the one-particle energies $\varepsilon_{n\mathbf{k}}^{\mathrm{DFT}}$. 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. We use a one-shot calculation for both *GW* and $GW^{\text{-ss}}$, as is common practice when using the GWA, where the noninteracting Green's function is obtained directly from a preceding DFT calculation and is not subsequently updated in further iterations. It should therefore be noted that, as the calculation is perturbative in nature and is not performed self-consistently, the choice of starting point will have an effect on the final result. However, as our aim is to study the effects coming from the self-screening error in real materials (which would be present also in a fully self-consistent *GW* calculation), this choice should not have a large effect on our conclusions. A $8 \times 8 \times 8$ **k**-grid was used for all calculations, except in the case of Si where we instead used a $12 \times 12 \times 12$ grid to more accurately estimate the location of the indirect band gap.

The band gaps calculated for the semiconductors within LDA, *GW*, and $GW^{\text{-ss}}$ 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 $\mathcal{S}$. 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 $\mathcal{S}$ is further doubled in size to the 40 lowest bands. This encouraging result we believe helps justify 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 in this case instead leads to a slight overestimation of the gap, with the error being comparable with the one in conventional *GW*. The observed trend of increasing gaps is in agreement with the expectation that a reduction in screening (removal of the self-screening) would bring the situation more toward a Hartree-Fock picture (here, the self-energy contains only the exchange part Σ^{x}, neglecting screening effects), where it is well known that too-large band gaps are found [21,26,27].

It has previously been shown [13–15] 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 $\mathcal{S}$ subset of bands, an increase in the number of (untreated) unoccupied bands should lead to a similar situation as for a usual *GW* calculation. We therefore follow Friedrich *et al.* in Ref. [14] and use a hyperbolical fit for *N* bands of the form:

$$E_{\mathrm{gap}}(N) = \frac{a}{N - N_0} + b, \tag{23}$$

with parameters *a*, *b*, and N_0 extracted from the corrected fit in Ref. [15]. 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 good agreement between our *GW* asymptote at 2.79 eV and the previously reported value of the *GW* band gap at 2.83 eV [15]. The similarly extrapolated gap obtained from the self-screening-corrected results brings our final gap to better agreement at 2.95 eV. The band gaps resulting from this fit are indicated in Table I.

As briefly mentioned earlier, the results of a one-shot calculation will depend on the choice of initial starting point. To further test our scheme, we therefore perform a self-screening correction calculation also using an initial G^0 from a generalized gradient approximation (GGA) [28] calculation. In the case of Si, both the indirect and direct band gaps using conventional *GW* (1.08 and 3.19 eV) and with the correction (1.19 and 3.28 eV) are in better agreement with

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 $\mathcal{S}$. The experimental values are taken from Refs. [23–25] as indicated. The values in parentheses 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	$GW^{\text{-ss}}$ (20)	$GW^{\text{-ss}}$ (40)	Exp.
Ge	–	0.70 (0.80)	0.79 (0.89)	0.79 (0.89)	0.74 (0.90) [23]
Si	0.47 (2.52)	1.00 (3.16)	1.12 (3.25)	1.12 (3.25)	1.17 (3.37) [24]
InP	0.47	1.24	1.36	1.36	1.42 [23]
GaAs	0.30	1.38	1.50	1.50	1.52 [23]
CdTe	0.51	1.43	1.55	1.55	1.61 [25]
ZnSe	1.04	2.42	2.60	2.60	2.82 [25]
ZnO	0.76	2.41 [2.79]	2.56 [2.94]	2.57 [2.95]	3.44 [25]

the experimental values. The important point to note here is that, although the gaps calculated based on the GGA are improved compared with LDA, there is still a relatively significant self-screening error of 0.11 (0.09) eV in the calculated indirect (direct) gaps. Indeed, it is of the same magnitude as that found when starting from LDA, where the error is 0.12 (0.09) eV. Such an error would be present even when a more optimal starting point, e.g., QSGW, which gives an excellent agreement for calculated gaps, was used. This is because the error is inherent in the GWA itself, irrespective of the starting Green's function used. It is, of course, possible that the choice of a different starting G^0 would provide a negligible self-screening error in the end, although this seems unlikely to be the case. Further systematic studies with other density functionals and calculations using (partial) self-consistency schemes, e.g., QSGW, would therefore be of large interest to verify this conjecture.

In Table II, we finally present our calculated center-of-weight values of the 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 with the GW values for some materials when only treating the semicore states with the self-screening correction. This small change is canceled 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 I. 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 d states remain mostly unaffected by the self-screening correction.

TABLE II. Position of the center of weight of the d semicore states in eV, relative to the valence band maximum calculated within LDA, GW, and self-screening-corrected GW. The self-screening correction was applied with 20 and 40 bands, respectively, and additionally with only the semicore states included in the subspace $\mathcal{S}$. The experimental positions of the semicore states are taken from Ref. [24], averaged over the reported $d_{3/2}$ and $d_{5/2}$ energies.

	LDA	GW	$GW^{\text{-ss}}$ (d)	$GW^{\text{-ss}}$ (20)	$GW^{\text{-ss}}$ (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

IV. CONCLUSIONS

We have applied two recently proposed self-screening corrections to the RPA (self-polarization) and the GWA (self-screening) in model systems to investigate the effects and regions of validity of the schemes compared with the conventional methods. We have furthermore employed one of the schemes, a local form of the self-screening correction, in *ab initio* calculations for quantitative predictions of several semiconductors, specifically studying its effect on the band gaps and the position of the semicore d states. 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. [12].

(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, for which 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. Furthermore, 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) regimes.

The *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. As the self-screening error is inherent in the GWA, this relatively significant error would be present even when different schemes, which yield closer agreement with experiment, are used. Although, at this point, further studies using additional methods are needed to elucidate potential differences in the self-screening error due to starting point effects.

That the more delocalized *sp* states involved around the band gap have a large part of the remaining error in the GWA compared with experiment removed agrees with what we would expect based on the model calculations. Similarly, the more localized *d* 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. While self-screening can be expected to also be significant for these localized states, our model calculations and *ab initio* predictions of the *d* semicore state positions indicate that some form of the so-called self-polarization correction could prove more suitable in this regime. Based on these findings, we therefore believe that it would be of interest to apply also the self-polarization scheme in realistic calculations of materials; in particular, the more localized *d* semicore states in the studied semiconductors could be expected to show an improvement. However, the issue of noncausality would first have to be addressed. The proposed origin of this issue, an incomplete removal of the self-screening from a physical state, combined with a more rigorous approach [19,20], could provide a starting point for an improved scheme.

To summarize, it is well-known that RPA- and GWA-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. To further verify the usefulness of the self-screening correction and thereby also examine the self-screening error inherent in *GW* calculations, investigations of different 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 Foundation Grant No. 2017.0061 and the Swedish Research Council, Vetenskapsrådet 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] G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. Lett. **45**, 290 (1980).

[6] G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. B **25**, 2867 (1982).

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

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

[9] W. Nelson, P. Bokes, P. Rinke, and R. W. Godby, Phys. Rev. A **75**, 032505 (2007).

[10] P. Romaniello, S. Guyot, and L. Reining, J. Chem. Phys. **131**, 154111 (2009).

[11] J. Wetherell, M. J. P. Hodgson, and R. W. Godby, Phys. Rev. B **97**, 121102(R) (2018).

[12] F. Aryasetiawan, R. Sakuma, and K. Karlsson, Phys. Rev. B **85**, 035106 (2012).

[13] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **105**, 146401 (2010).

[14] C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B **83**, 081101(R) (2011).

[15] C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B **84**, 039906(E) (2011).

[16] C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B **81**, 125102 (2010).

[17] The FLEUR group, "The FLEUR project," http://www.flapw.de.

[18] 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 $R^{\mathrm{ss}}_{m\sigma} = R^{\mathrm{sp}}/2$. The predicted N-particle excitation energy is therefore the same.

[19] G. Stefanucci, Y. Pavlyukh, A.-M. Uimonen, and R. van Leeuwen, Phys. Rev. B **90**, 115134 (2014).

[20] A.-M. Uimonen, G. Stefanucci, Y. Pavlyukh, and R. van Leeuwen, Phys. Rev. B **91**, 115104 (2015).

[21] R. Hott, Phys. Rev. B **44**, 1057 (1991).

[22] F. Aryasetiawan and O. Gunnarsson, Phys. Rev. B **54**, 17564 (1996).

[23] 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 (Springer-Verlag, Berlin, Heidelberg, 2002), Vol. 41A1b.

[24] T. C. Chiang and F. J. Himpsel, in *Electronic Structure of Solids: Photoemission Spectra and Related Data*, Landolt-Börnstein—

Group III Condensed Matter, edited by A. Goldmann and E.-E. Koch (Springer-Verlag, Berlin, Heidelberg, 1989), Vol. 23A.

[25] O. Madelung, U. Rössler, and M. Schulz, eds., *II-VI and I-VII Compounds; Semimagnetic Compounds*, Landolt-Börnstein—Group III Condensed Matter (Springer-Verlag, Berlin, Heidelberg, 1999), Vol. 41B.

[26] A. Svane, Phys. Rev. B **35**, 5496 (1987).

[27] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[28] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).