Technical University of Denmark

DTU

Simple vertex correction improves GW band energies of bulk and two-dimensional crystals

Schmidt, Per Simmendefeldt; Patrick, Christopher; Thygesen, Kristian Sommer

DTU Library
Technical Information Center of Denmark

Simple vertex correction improves GW band energies of bulk and two-dimensional crystals

Per S. Schmidt, Christopher E. Patrick,[*] and Kristian S. Thygesen

Center for Atomic-scale Materials Design, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

The GW self-energy method has long been recognized as the gold standard for quasiparticle (QP) calculations of solids in spite of the fact that the neglect of vertex corrections and the use of a density-functional theory starting point lack rigorous justification. In this work we remedy this situation by including a simple vertex correction that is consistent with a local-density approximation starting point. We analyze the effect of the self-energy by splitting it into short-range and long-range terms which are shown to govern, respectively, the center and size of the band gap. The vertex mainly improves the short-range correlations and therefore has a small effect on the band gap, while it shifts the band gap *center* up in energy by around 0.5 eV, in good agreement with experiments. Our analysis also explains how the relative importance of short- and long-range interactions in structures of different dimensionality is reflected in their QP energies. Inclusion of the vertex comes at practically no extra computational cost and even improves the basis set convergence compared to GW. Taken together, the method provides an efficient and rigorous improvement over the GW approximation.

I. INTRODUCTION

The GW approximation [1–4], introduced by Hedin in 1965 [5], remains the most widely used method for quasiparticle (QP) calculations of semiconductors and insulators. Over the years it has been extensively applied to inorganic solids [6–8] and, more recently, also to molecules [9–12] and atomically thin two-dimensional (2D) materials [13–15].

The GW self-energy can be obtained by iterating Hedin's equations once starting from $\Sigma = 0$ (i.e., the Hartree approximation). This produces the trivial vertex function $\Gamma(1,2,3) = \delta(1,2)\delta(1,3)$, which corresponds to invoking the time-dependent Hartree approximation for the dynamical screening [i.e., the random phase approximation (RPA)]. For this approach to be consistent, the Green's function which should be used for the calculation of the self-energy is the Hartree G. This is known to be a poor approximation, and instead, practical GW calculations follow a "best G, best W" philosophy [3]. Most often, one uses a noninteracting G_0 from density-functional theory (DFT) and evaluates W within the RPA from the polarizability $\chi_0 = G_0 G_0$. This approximation is referred to as $G_0 W_0$ and has been shown to yield reasonably good, although somewhat underestimated, band gaps [6,16]. Carrying out self-consistency in only the Green's function, GW_0, has been found to improve the band gaps [17]. Iterating to full self-consistency in both the Green's function and screened interaction, GW, systematically overestimates the band gaps and worsens the agreement with experiments [17].

The most obvious way to go beyond the GW approximation is to perform another iteration of Hedin's equations starting from $\Sigma = iGW$. Neglecting derivatives of W, this produces the kernel $\delta\Sigma(1,2)/\delta G(3,4) = iW(1,2,3,4)$, which is known from the Bethe-Salpeter equation. The four-point nature of this kernel makes it difficult to invert the vertex equation,

$\Gamma = \delta + KGG\Gamma$, without loss of accuracy. Instead, one can perform a single iteration of the vertex equation to obtain $\Gamma = \delta + WGG$, which leads to a self-energy consisting of a second-order screened exchange term in addition to the usual iGW term. Grüneis *et al.* have shown, using a static approximation for W in the vertex, that this $GW\Gamma^1$ approximation, performed in a fully self-consistent manner, leads to significant improvements for band gaps and ionization potentials of solids [18]. From a theoretical point of view this is a highly satisfactory result. The drawback is the higher complexity of the formalism and the concomitant loss of physical transparency as well as the significant computational overhead compared to the GW method.

Time-dependent density-functional theory (TDDFT) in principle offers a framework for including exchange-correlation (xc) effects in the dynamical response via a two-point vertex function rather than the computationally challenging three-point vertex function that arises naturally in the diagrammatic many-body formalism. While using TDDFT-derived vertex functions for many-body calculations appears attractive, progress along these lines has been hindered by the poor quality of the local xc kernels derived from standard local xc potentials. However, recent work has shown that a simple renormalization of the adiabatic local-density approximation (LDA) xc kernel can overcome these problems and yield a dramatic improvement over the RPA for total-energy calculations based on the adiabatic-connection fluctuation-dissipation theorem (ACFDT) [19–21].

Here we show that the renormalized adiabatic LDA (rALDA) kernel, when introduced in Hedin's equations, produces a simple two-point vertex function that leads to systematically improved QP energies for a range of semiconductors and insulators. The most striking effect of the vertex is that it raises the absolute QP energies from $G_0 W_0$ by around 0.5 eV while the gaps are almost unaffected. These effects can be traced to an improved description of the short-range correlation hole and thus the (absolute) correlation energy of electrons in the ground state.

---

*Present address: Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom.

II. SELF-ENERGY AND XC KERNEL

As originally observed by Hybertsen and Louie [3], it is possible to start the iterative solution of Hedin's equation not with $\Sigma = 0$ (which leads to the GW approximation) but rather with a local xc potential: $\Sigma^0(1,2) = \delta(1,2)v_{xc}(1)$. As shown by Del Sole *et al.* [22], this leads to a self-energy of the form

$$\Sigma(1,2) = iG(1,2)\widetilde{W}(1,2), \qquad (1)$$

where

$$\widetilde{W} = v[1 - \chi^0(v + f_{xc})]^{-1} \qquad (2)$$

and $f_{xc}(1,2) = \delta v_{xc}(1)/\delta n(2)$ is the adiabatic xc kernel. Crucially, $\widetilde{W}(1,2)$ is the screened *effective* potential at 2 generated by a charge at 1. It consists of the bare potential plus the induced Hartree and xc potentials. It is thus the potential felt by a (Kohn-Sham) electron in the system. For comparison the potential felt by a classical test charge is the bare potential screened only by the induced Hartree potential:

$$\widehat{W} = v + v[1 - \chi^0(v + f_{xc})]^{-1}\chi^0 v. \qquad (3)$$

Using $\widehat{W}$ in Eq. (1) corresponds to including the vertex in the polarizability P (or irreducible response function) but neglecting it in the self-energy. We refer to the use of $\widetilde{W}$ and $\widehat{W}$ in Eq. (1) as $G_0W_0\Gamma$ and G_0W_0P, respectively. The subscripts indicate that the self-energies are evaluated non-self-consistently starting from DFT. Note that in contrast to the GW approximation, which strictly should be based on the Hartree G, the use of a DFT starting point is perfectly justified within the $G_0W_0\Gamma$ theory.

The adiabatic LDA kernel is given by

$$f_{xc}^{ALDA}[n](\mathbf{r},\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')f_{xc}^{ALDA}[n],$$

where

$$f_{xc}^{ALDA}[n] = \frac{d^2}{dn^2}\left(ne_{xc}^{HEG}\right)\Big|_{n=n(\mathbf{r})}.$$

While f_{xc}^{ALDA} equals the exact xc kernel of the homogeneous electron gas (HEG) in the $q \to 0$ and $\omega \to 0$ limits, it violates a number of other exact conditions. In particular, it does not incorporate the correct asymptotic $\propto q^{-2}$ behavior for large q. This deficiency leads to a divergent on-top correlation hole [23]. Moreover, the ALDA kernel diverges at small densities where $f_x^{ALDA} \sim n^{-2/3}$. We have observed that when applying the ALDA kernel to systems other than silicon (which was the system addressed by Del Sole *et al.* [22] and again by Shaltaf *et al.* [24]), these divergences make it impossible to obtain converged results in practice. This is exemplified in Fig. 5(b) below and emphasizes the critical importance of renormalizing the local kernel.

The rALDA kernel is defined for the HEG by setting $f_{xc}^{rALDA}[n](q) = f_{xc}^{ALDA}[n]$ for $q < 2k_F[n]$ and $-v(q)$ otherwise (this ensures continuity at $q = 2k_F$). This results in a nonlocal kernel with the (almost) exact asymptotic $q \to \infty$ behavior and without the divergences of the ALDA kernel [21]. In this work we have followed the wave-vector symmetrization scheme [see Eq. (38) in [21]] to generalize the rALDA to inhomogeneous densities. Furthermore, we include only the dominant exchange part of the kernel. We mention that a small inconsistency of our QP scheme is that we iterate Hedin's

TABLE I. The bulk crystal structures considered in this study. The lattice constants and **k**-point grids applied in the quasiparticle calculations are shown.

	Structure	Lattice constant (Å)	**k** points
MgO	rocksalt	4.212	$8 \times 8 \times 8$
SiC	zinc blende	4.350	$8 \times 8 \times 8$
LiF	rocksalt	4.024	$8 \times 8 \times 8$
CdS	zinc blende	5.818	$8 \times 8 \times 8$
Si	diamond	5.431	$8 \times 8 \times 8$
C	diamond	3.567	$8 \times 8 \times 8$
BN	zinc blende	3.615	$8 \times 8 \times 8$
AlP	zinc blende	5.451	$8 \times 8 \times 8$

equations from $\Sigma^0(1,2) = \delta(1,2)v_{xc}^{LDA}(1)$, while f_{xc}^{rALDA} does not exactly equal $\delta v_{xc}^{LDA}/\delta n$ due to the truncation at $q = 2k_F$.

The rALDA kernel provides an essentially exact description of the correlation hole of the HEG across a wide range of densities and has been shown to improve the RPA description of bond energies in molecules and solids [19–21]. However, more important for the present work is that the rALDA kernel provides a dramatic improvement of absolute correlation energies compared to RPA. For example, the RPA correlation energy of the HEG is 0.3–0.5 eV/electron too negative, while the rALDA error is below 0.03 eV/electron. Very similar trends are seen for small atoms and molecules [20] as well as for bulk silicon [21].

III. COMPUTATIONAL DETAILS

We have calculated the QP band gaps, ionization potentials (IPs), and electron affinities (EAs) for a range of semiconductors and insulators using five different approximations to the self-energy: (i) conventional G_0W_0, (ii) eigenvalue self-consistent GW_0, (iii) full eigenvalue self-consistent GW, (iv) $G_0W_0\Gamma$, and (v) G_0W_0P. The non-self-consistent calculations employed an LDA starting point, and the exchange-only rALDA kernel was used to obtain $\widetilde{W}$ from Eq. (2). The QP calculations for the bulk and 2D crystals in their experimental geometries were performed using the GPAW code [25]. See Tables I and II for lattice constants and thicknesses of the 2D materials. Response functions and screened interactions were evaluated along the real-frequency axis using a nonlinear frequency grid. The number of unoccupied bands included in χ_0 was set equal to the number of plane-wave-basis functions. An $8 \times 8 \times 8$ (18×18) **k**-point grid was used for all bulk (2D) materials. For the 2D materials we employed a recently developed method for treating the critical $\mathbf{q} \to \mathbf{0}$ limit of the screened interaction while avoiding spurious interactions with

TABLE II. The 2D crystal structures considered in this study. Lattice constant, layer thickness, and **k**-point grids are shown.

	Lattice constant (Å)	Thickness (Å)	**k** points
2H-MoS$_2$	3.160	3.170	$18 \times 18 \times 1$
2H-MoSe$_2$	3.289	3.340	$18 \times 18 \times 1$
2H-WS$_2$	3.153	3.360	$18 \times 18 \times 1$

neighboring supercells [14]. Fifteen angstroms of vacuum were used in the out-of-plane direction. The reported band positions and gaps were extrapolated to the infinite plane-wave-basis limit, and the results are estimated to be converged to within 0.02 eV. Band edge positions with respect to vacuum were determined by aligning the Hartree potential at the nuclei in the bulk calculations with that inside a slab with surface orientation and reconstruction as reported in available experimental studies. The considered surfaces are (111) 2×1 for Si in the diamond structure, (100) for MgO and LiF in the rocksalt structure, and (110) for the rest of the compounds in the zinc-blende structure. The slabs are represented by 10 layers (rocksalt), 14 layers (zinc blende), and 24 layers (diamond). The surfaces were relaxed with the Perdew-Burke-Ernzerhof xc functional [26], rescaled to the experimental lattice constant, and recalculated with LDA. All gaps and absolute band positions for the 2D materials are calculated at the K-point.

IV. RESULTS

Table III shows the band gaps obtained with the different methods and their deviations from experimental reference values for both bulk and 2D materials. The 2D materials are included because they are scientifically interesting and also because they offer a unique opportunity for obtaining accurate energy levels due to their well-defined surface

TABLE III. Band gaps obtained using different self-energy approximations (see text). The gaps for the 2D materials are calculated at the K-point. Experimental values are from [27] (and references therein) and corrected for zero-point motion (MgO: 0.15 eV, CdS: 0.06 eV, LiF: 0.46 eV, SiC: 0.11 eV, Si: 0.05 eV, C: 0.40 eV, BN: 0.26 eV, AlP: 0.02 eV) as found in [28,29] (and references therein). The experimental values for the 2D materials have not been corrected for zero-point motion since only a value for MoS_2 was found in the literature (75 meV) [30]. Calculated values for the 2D materials include spin-orbit coupling. The experimental values for the monolayer (ML) materials have not been corrected for zero-point motion. A zero-point correction of 75 meV was found for MoS_2[30] and it is expected to be of similar size for $MoSe_2$ and even smaller for WS_2. Calculated values for the ML materials include spin-orbit coupling. The last two rows show the mean absolute error (MAE) and mean signed error (MSE) with respect to the experimental values.

	LDA	G_0W_0	GW_0	$G_0W_0P_0$	$G_0W_0\Gamma_0$	Expt.
MgO	4.68	7.70	8.16	7.10	7.96	7.98
CdS	0.86	1.76	2.27	1.84	1.87	2.48
LiF	8.83	14.00	14.75	13.25	14.21	14.66
SiC	1.31	2.54	2.72	2.38	2.57	2.51
Si	0.52	1.23	1.34	1.16	1.29	1.22
C	4.10	5.74	5.97	5.62	5.69	5.88
BN	4.36	6.54	6.81	6.27	6.60	6.6
AlP	1.44	2.48	2.67	2.34	2.51	2.47
ML-MoS_2	1.71	2.47	2.61	2.28	2.47	2.50
ML-$MoSe_2$	1.43	2.08	2.23	1.99	2.07	2.31
ML-WS_2	1.33	2.75	3.07	2.56	2.81	2.72
MAE	1.89	0.20	0.17	0.41	0.16	
MSE	-1.89	-0.19	0.12	-0.41	-0.12	

structures. On the contrary, energy levels in bulk solids are greatly affected by variations and uncertainties in the surface orientation/termination, which makes a comparison between theoretical and experimental results problematic.

In agreement with previous findings G_0W_0 underestimates the experimental band gaps slightly, while GW_0 generally overestimates the gaps. The overestimation becomes even larger in GW (see Appendix B), which is therefore not considered further in this work. The experimental valence-band maxima (VBMs) are from Ref. [31]. The band gap of $MoSe_2$ is from [32], where a gap of 2.18 ± 0.04 eV is reported for $MoSe_2$ on top of bilayer graphene. The effect of the substrate is calculated in the same reference to be a lowering of the band gap of 0.13 eV, giving a band gap of 2.31 eV for free-standing $MoSe_2$.

The band gap of 2.5 eV for free-standing MoS_2 is from [33]. In [34] a band gap of 2.18 ± 0.05 eV for MoS_2 on top of quartz is reported. Comparing the two numbers, quartz is expected to lower the gap by 0.32 eV. In [34] the band gap of WS_2 on top of quartz is reported to be 2.40 ± 0.06 eV. Assuming the same substrate effect, the band gap of free-standing WS_2 is 2.72 eV.

The numbers in Table III include spin-orbit corrections. These are a splitting of the valence band by 0.15, 0.19, and 0.45 eV and of the conduction band by 0.00, 0.02, and 0.02 eV for MoS_2, $MoSe_2$, and WS_2, respectively [35].

From Table III we conclude that $G_0W_0\Gamma$ shows the best agreement with experiments, closely followed by GW_0, but the mean signed errors of the two methods come with opposite signs. Including the vertex only in the polarizability (G_0W_0P) leads to a closing of the gap (as previously reported in [27,28]) resulting in significantly underestimated gaps.

In Fig. 1 we show the absolute positions of the VBM and conduction-band minimum (CBM) relative to vacuum. The most striking effect of including the vertex is a systematic upshift of the band edges by around 0.5 eV. Remarkably, this upshift leads to a better overall agreement with experiments (dashed black lines). The upshift of band energies is not observed when the vertex is included only in the polarizability, i.e., when employing a test-charge-test-charge screened interaction (G_0W_0P). Moreover, no systematic upshift of the band edges is observed for the self-consistent GW flavors which also employ test-charge-test-charge screening. We conclude that the upshift of band energies originates from the presence of the vertex in the self-energy, i.e., the use of a test-charge-electron screened interaction.

V. DISCUSSION

In the following we analyze our results from a total-energy perspective focusing on the G_0W_0 and $G_0W_0\Gamma$ methods by a generalization of Koopmans' theorem. Subsequently, the effect of short- and long-range screening is discussed. We then exemplify how the vertex affects the results depending on whether it is included in the polarizability, self-energy, or both. Finally, the improved numerical convergence behavior upon inclusion of the kernel is shown.

A. Generalized Koopmans' theorem

From Koopmans' theorem it follows that the highest occupied and lowest unoccupied QP energies can be
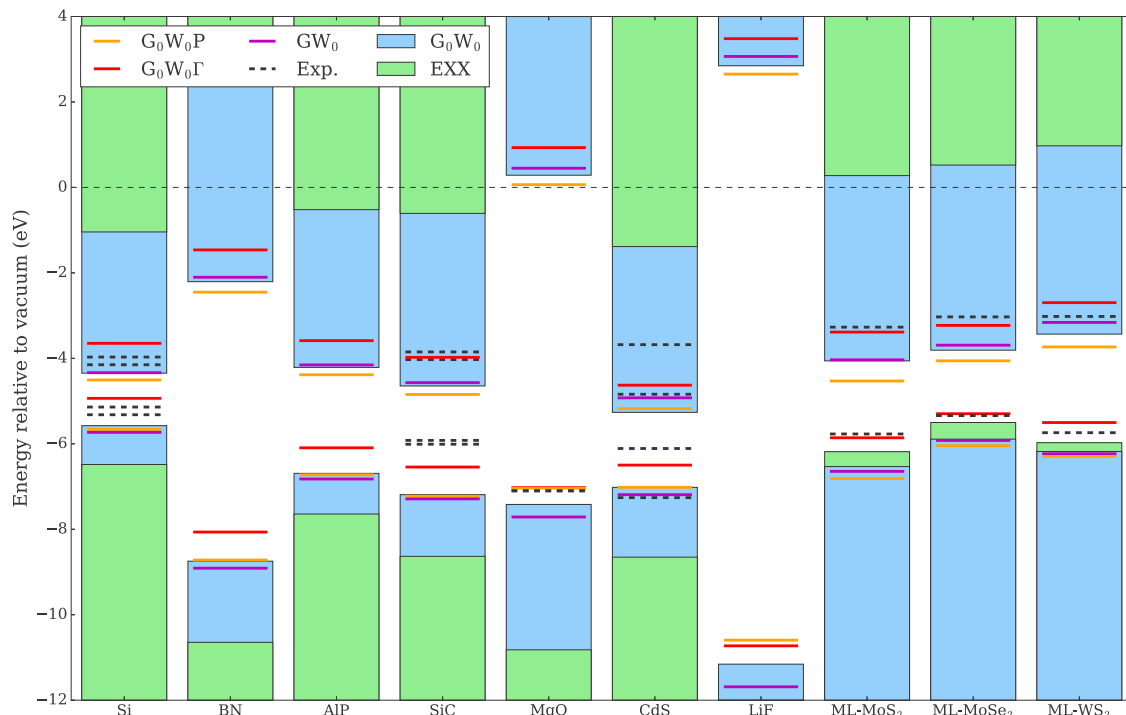
FIG. 1. IP and EA of a range of 3D and 2D semiconductors calculated with EXX (green), G_0W_0 (blue), $G_0W_0\Gamma$ (red), G_0W_0P (orange), and GW_0 (magenta) and compared with experimental values where available (black) [36].

expressed as

$$\varepsilon_N^{\mathrm{QP}} = \varepsilon_N^{\mathrm{HF}} + E_c[N] - E_c[N-1], \tag{4}$$

$$\varepsilon_{N+1}^{\mathrm{QP}} = \varepsilon_{N+1}^{\mathrm{HF}} + E_c[N+1] - E_c[N], \tag{5}$$

where $\varepsilon^{\mathrm{HF}}$ are the Hartree-Fock single-particle energies (evaluated on Kohn-Sham orbitals) and $E_c[N]$ is the correlation energy of the N-particle ground state. The latter can be calculated from the ACFDT, which can be cast in the form

$$E_c = -\int_0^1 d\lambda \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\{\chi^0(i\omega)(\widetilde{W}^\lambda(i\omega) - v)\}. \tag{6}$$

Here $\widetilde{W}^\lambda$ equals the screened test-charge-electron interaction of Eq. (2). Setting $f_{\mathrm{xc}} = 0$, we have $\widetilde{W} = W$, and E_c becomes the RPA correlation energy. Assuming no orbital relaxations (which is justified for an extended periodic crystal), Niquet and Gonze [37] have shown that the IP and EA calculated as total-energy differences with the ACFDT-RPA equal the highest occupied and lowest unoccupied QP energies from G_0W_0, respectively (when setting the renormalization factor Z to unity). In the same way it can be shown, at least for an exchange-only kernel, that the IP and EA obtained from the ACFDT with $\widetilde{W}$ from Eq. (2) equal the respective QP band edges obtained from $G_0W_0\Gamma$ when Γ is the vertex corresponding to f_x (see the Appendix A for a proof). These results represent a generalization of Koopmans' theorem of Hartree-Fock (HF) theory.

In general, HF is known to significantly overestimate the band gap of solids (see Fig. 1). Comparing HF with Eqs. (4) and (5), this means that the correlation energy in the $N \pm 1$ states must be larger (more negative) than the correlation

energy in the neutral N-particle ground state. It might seem surprising that $E_c[N-1] < E_c[N]$ since, naively, one would expect the correlation energy to be a monotonic decreasing function of N. However, the addition of an electron or hole to the system changes its character from insulating to metallic, and this entails an increase in the correlation energy. To make this idea more explicit we can split the change in correlation energy into two terms: the correlation energy per electron in the neutral ground state ($\varepsilon_c \equiv E_c[N]/N < 0$) and a remainder representing the extra correlation energy due to the insulator-metal transition [$\Delta_c^{+/-} \equiv E_c[N \pm 1] - (E_c[N] \pm \varepsilon_c)$]. With these definitions we can write

$$\varepsilon_N^{\mathrm{QP}} = \varepsilon_N^{\mathrm{HF}} + \varepsilon_c - \Delta_c^-, \tag{7}$$

$$\varepsilon_{N+1}^{\mathrm{QP}} = \varepsilon_{N+1}^{\mathrm{HF}} + \varepsilon_c + \Delta_c^+. \tag{8}$$

The relations are illustrated in Fig. 2(a). Clearly, the effect of ε_c is to downshift the band edges from their HF positions while $\Delta_c^\pm$ close the gap. In the quasiparticle picture, $\Delta_c^\pm$ represent the screening of the added electron or hole [see Fig. 2(b)], and we therefore refer to them as screening terms. By its stabilization of the final states (the $N \pm 1$ states) the effect of the screening terms is similar to that of orbital relaxations in finite systems, yet the underlying physics is completely different: orbital relaxations are vanishingly small in periodic crystals and occur even in noncorrelated theories like HF. In contrast $\Delta_c^\pm$ describe a pure correlation effect and do not vanish in infinite, periodic systems.

We find it useful to analyze the QP energies in terms of the band gap and the band gap center. These energies are related
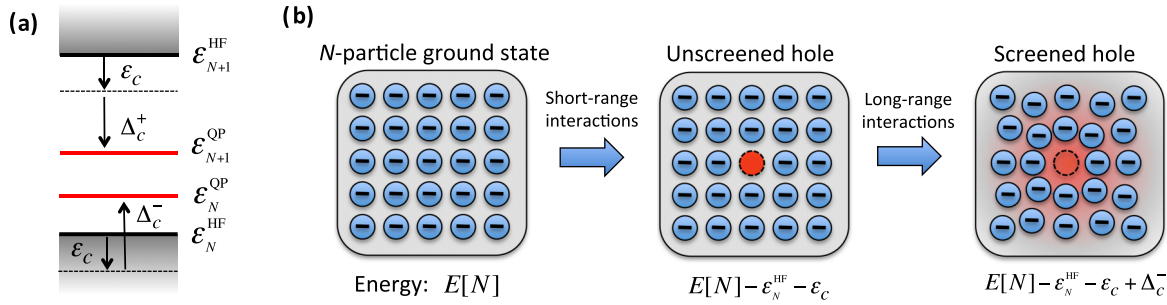
FIG. 2. (a) Schematic illustration of the different contributions to the highest occupied and lowest unoccupied QP levels of a semiconductor. (b) The energy cost of removing a valence electron consists of the Hartree-Fock energy $\varepsilon_N^{\mathrm{HF}}$, the correlation energy of an electron in the ground state ε_c, and a stabilizing screening contribution $\Delta_c^{\pm}$. The latter two are predominantly of short-range and long-range nature, respectively.

to ε_c and $\Delta_c^{\pm}$ by

$$E_{\mathrm{gap}}^{\mathrm{QP}} = E_{\mathrm{gap}}^{\mathrm{HF}} + (\Delta_c^- + \Delta_c^+), \qquad (9)$$

$$E_{\mathrm{cen}}^{\mathrm{QP}} = E_{\mathrm{cen}}^{\mathrm{HF}} + \varepsilon_c + (\Delta_c^+ - \Delta_c^-)/2. \qquad (10)$$

The correlation contribution to the gap is determined only by the screening terms $\Delta_c^{\pm}$. From the close agreement between the G_0W_0 and $G_0W_0\Gamma$ band gaps (red columns in Fig. 3) we conclude that the vertex has little effect on the (sum of the) screening terms. In contrast, the band gap center also depends on the ground-state correlation energy ε_c. We have calculated ε_c for all the investigated materials using the RPA and rALDA total-energy methods (see the Appendix C for the exact values). In Fig. 3 we compare the difference between the RPA- and rALDA-calculated ε_c (black line) with the difference between the G_0W_0- and $G_0W_0\Gamma$-calculated band gap centers (blue columns). The rather close agreement shows that the main difference in band gap center can be ascribed to ε_c. It is now clear that the upshift of QP energies obtained with $G_0W_0\Gamma$ originates from the smaller (less negative) correlation energy of electrons in the neutral ground state predicted by rALDA compared to RPA. The well-documented superiority of the rALDA over the RPA for the description of ground-state correlation energies, in combination with the improved agreement with experimental band energies (Fig. 1), constitutes strong evidence that our
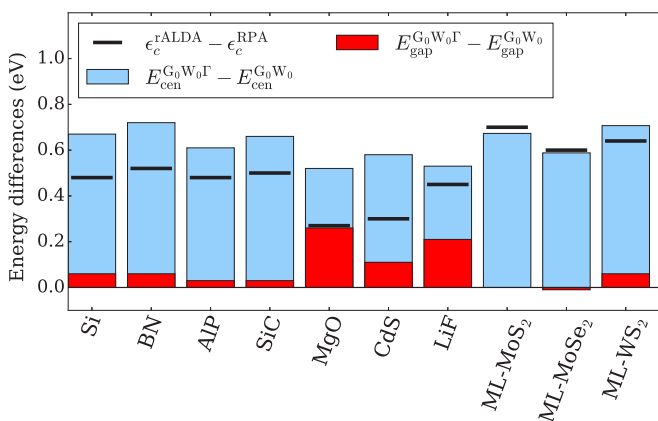
QP-rALDA scheme represents a genuine improvement over the *GW* approximation.

B. Short- versus long-range screening

We have seen that the dominant effect of the vertex correction is to shift the band gap center while the band gap itself is less affected. Physically, the main effect of the rALDA kernel is to modify the effective Coulomb interaction at short distances. More precisely, given a density variation δn, the corresponding induced electron potential $\delta v_{\mathrm{Hxc}} = (v + f_{\mathrm{xc}})\delta n$ is generally weaker than the bare Hartree potential $\delta v_H = v\delta n$ because $v(q) + f_{\mathrm{xc}}(q) < v(q)$. However, by definition of the rALDA kernel, the reduction is stronger for larger q, which translates to shorter distances in real space. From these observations we can conclude that the QP band gap is mainly determined by long-range interactions, while the band gap center is sensitive to the short-range interactions. This agrees well with the quasiparticle picture illustrated in Fig. 2: namely, adding a particle or hole without accounting for the screening represents a local (short-range) perturbation, while the screening of the added charge is a long-range process. While the rALDA kernel mainly reduces the short-range interactions, it also reduces the long-range components slightly. This leads to a slightly weaker long-range screening (smaller Δ_c) and slightly larger band gaps, as seen in Table III.

Returning to Fig. 1, we note that for the 2D materials Hartree-Fock theory predicts a lower IP than the *GW* methods, in clear contrast to the situation for bulk solids. This anomalous behavior is a result of the relatively more important effect of short- compared to long-range correlations in reduced dimensions. Indeed, the dielectric function of a 2D semiconductor approaches unity in the long-wavelength limit, which reduces the screening terms $\Delta_c^{\pm}$. At the same time we find that the 2D materials present the largest values for ε_c of all the materials (see Table V in Appendix C), showing that the absolute correlation energy per electron is larger for the 2D materials compared to the bulk materials.

C. Vertex in the polarizability and/or self-energy

As mentioned previously, the vertex enters Hedin's equations at two places, namely, the polarizability and the self-energy. For a consistent description the vertex should be included in both places (the $G_0W_0\Gamma_0$ approach). However, it is instructive to study the effect of including the vertex separately

FIG. 3. The difference in band gap, band gap center, and ϵ_c upon inclusion of the rALDA kernel.
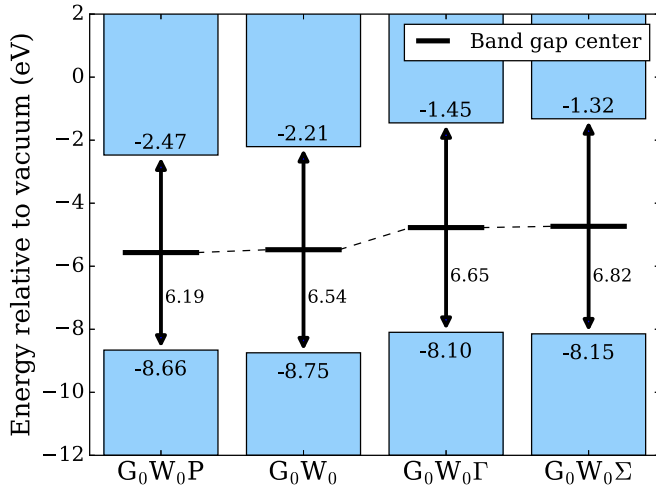
FIG. 4. Absolute position of the VBM and CBM relative to vacuum for BN calculated with the four different methods. The band gap center is shown with a dotted line.

in the polarizability and self-energy. To this end we note that the self-energy (excluding the exchange part) can be written

$$\Sigma_c = iGv\chi(v + f_{xc}), \qquad (11)$$

where $\chi = P + Pv\chi$ and $P = \chi^0 + \chi^0 f_{xc} P$. Including f_{xc} in P affects the description of screening, while including it in Σ_c affects the form of the potential created by the induced density when subject to a test charge (as explained above, it mainly reduces the short-range part of the potential). Based on this, we can obtain four different GW-like self-energies, where the only one not mentioned up to now includes the vertex in the self-energy but not in the polarizability, which we term $GW\Sigma$. Figure 4 shows the band gap size and center obtained for BN with these four self-energies. It is clear that the band gap center depends mainly on f_{xc} in the self-energy; that is, a correct description of the band gap center requires the inclusion of xc effects in the induced potential. The size of the band gap increases in the order G_0W_0P, G_0W_0, $G_0W_0\Gamma$, $G_0W_0\Sigma$. As previously argued, the size of the gap depends

mainly on the long-range screening (and less on the short-range form of the final induced potential). The f_{xc} reduces the long-range interactions somewhat. Thus the total induced potential will generally be smaller when xc effects are included in the final potential. This explains the larger band gaps found for $G_0W_0\Gamma$ and $G_0W_0\Sigma$. The remaining ordering comes from noting that $\chi^{rALDA} > \chi^{RPA}$ because the higher-order diagrams, which reduce the effect of χ^0, are larger in RPA.

D. Improved convergence

Finally, we mention that the reduction of large q components by the rALDA kernel not only improves the description of short-range correlations but also leads to faster convergence with respect to plane waves and the number of unoccupied states compared to standard GW calculations, as shown in Fig. 5(a) for the case of bulk BN. On the y axis we show the difference in the band gap compared to that from a calculation at a cutoff of 50 eV with the corresponding number of bands. The improved convergence also manifests itself in the QP corrections to the individual bands.

In Fig. 5(b) the band gap of bulk BN is shown versus the plane-wave cutoff applying the $G_0W_0\Gamma_0$ method with the ALDA and rALDA kernel as well as the standard G_0W_0 method. The need for the renormalization of the kernel in order to obtain converged results is very apparent.

VI. CONCLUSION

In conclusion, we have demonstrated that a more accurate description of short-range correlations in QP calculations can be obtained with a simple TDDFT-inspired vertex function. Inclusion of the vertex improves the agreement with experimental data for the absolute band energies of bulk and two-dimensional semiconductors. Moreover, it justifies the use of DFT as a starting point for non-self-consistent QP calculations and is thus formally more rigorous than the $G_0W_0@DFT$ approach. Importantly, these advantages come without increasing the numerical complexity or computational cost compared to G_0W_0 calculations.
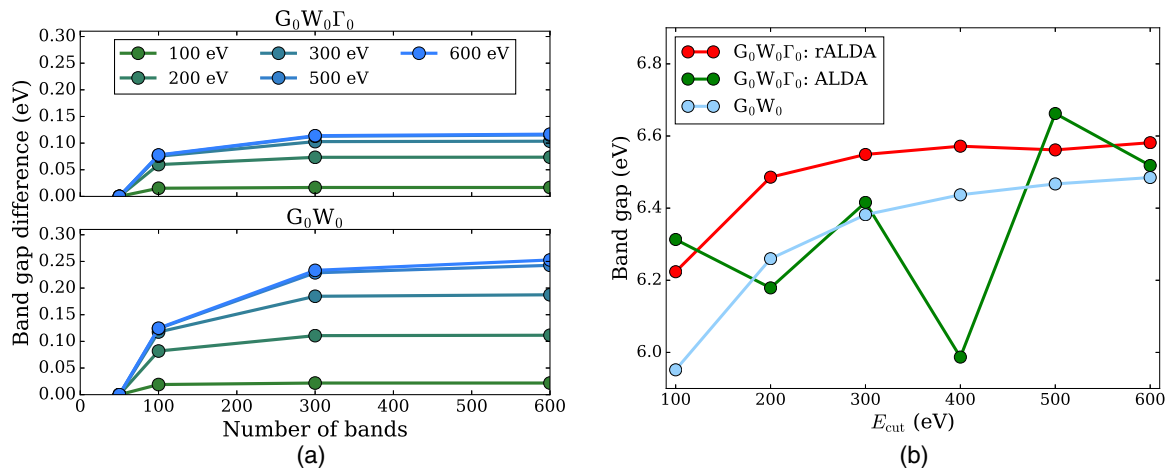

FIG. 5. (a) Convergence of the band gap in BN with respect to plane-wave cutoff and the number of bands included using the RPA (bottom) and rALDA (top) kernels. (b) Plane-wave convergence of the band gap of bulk BN using the ALDA and rALDA kernels in the $G_0W_0\Gamma_0$ method as well as the RPA kernel (G_0W_0 method).

TABLE IV. Band gaps calculated with eigenvalue self-consistency in G (GW_0) and in both G and W (GW).

	GW_0	GW_0P_0	$GW_0\Gamma_0$	GW	GWP_0	$GW\Gamma_0$	Expt.
MgO	8.16	7.52	8.21	9.21	8.56	9.47	7.98
CdS	2.27	2.03	2.34	2.73	2.44	2.83	2.48
LiF	14.75	14.02	14.90	16.29	15.59	16.46	14.66
SiC	2.72	2.52	2.73	3.06	2.80	3.11	2.51
Si	1.34	1.23	1.40	1.48	1.38	1.60	1.22
C	5.97	5.74	5.88	6.38	6.16	6.33	5.88
BN	6.81	6.50	6.84	7.44	7.08	7.49	6.6
AlP	2.67	2.48	2.67	3.00	2.74	3.01	2.47
MAE	0.16	0.23	0.18	0.72	0.38	0.81	
MSE	0.11	−0.22	0.15	0.72	0.37	0.81	

ACKNOWLEDGMENT

APPENDIX A: FROM RALDA TOTAL ENERGIES TO $GW\Gamma$

Using the adiabatic-connection fluctuation-dissipation theorem (ACFDT), the exact correlation energy of the system can be written in terms of the interacting response function $\chi^\lambda(i\omega)$ of a system with a scaled Coulomb interaction, $v \rightarrow \lambda v$:

$$E_c = -\int_0^1 d\lambda \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\{v[\chi^\lambda(i\omega) - \chi^{\mathrm{KS}}(i\omega)]\},$$

where $\chi^{\mathrm{KS}}(i\omega)$ is the response function of the noninteracting Kohn-Sham system. χ^λ can, in principle, be obtained from the Dyson equation,

$$\chi^\lambda(i\omega) = \chi^{\mathrm{KS}}(i\omega) + \chi^{\mathrm{KS}}(i\omega)\left[\lambda v + f_{\mathrm{xc}}^\lambda(i\omega)\right]\chi^\lambda(i\omega),$$

where all the complicated correlation effects have been transferred into $f_{\mathrm{xc}}^\lambda(i\omega)$, which needs to be approximated.

It can be shown that any pure exchange kernel must have the property $f_x^\lambda[n](\mathbf{r},\mathbf{r}',i\omega) = \lambda f_x[n](\mathbf{r},\mathbf{r}',i\omega)$, making it possible

to carry out the λ integration analytically:

$$E_c = \int_0^\infty \frac{d\omega}{2\pi}\mathrm{Tr}\{vf_{Hx}^{-1}(i\omega)\ln[1 - \chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)] + v\chi^{\mathrm{KS}}(i\omega)\},$$

where $f_{Hx}(i\omega) = v + f_x(i\omega)$.

Denoting the change in $\chi^{\mathrm{KS}}(i\omega)$ when adding one electron to the lowest unoccupied KS orbital by $\delta\chi^{\mathrm{KS}}(i\omega)$,

$$\delta\chi^{\mathrm{KS}}(\mathbf{r},\mathbf{r}',i\omega) = \phi_c^*(\mathbf{r})G_0(\mathbf{r},\mathbf{r}',\epsilon_c + i\omega)\phi_c(\mathbf{r}') + \mathrm{c.c.},$$

assuming that the density does not change with the addition of one electron [meaning $f_{Hx}(i\omega)$ does not change], we calculate the correlation contribution to the electron affinity (AE):

$$\mathrm{AE}_c = E_c[N] - E_c[N+1]$$
$$= \int_0^\infty \frac{d\omega}{2\pi}\mathrm{Tr}\{vf_{Hx}^{-1}(i\omega)\ln[1 - \chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)]$$
$$+ v\chi^{\mathrm{KS}}(i\omega) - \left(vf_{Hx}^{-1}(i\omega)\ln\{1 - [\chi^{\mathrm{KS}}(i\omega)$$
$$+ \delta\chi^{\mathrm{KS}}(i\omega)]f_{Hx}(i\omega)\} + v[\chi^{\mathrm{KS}}(i\omega) + \delta\chi^{\mathrm{KS}}(i\omega)]\right)\}$$
$$= -\int_0^\infty \frac{d\omega}{2\pi}\mathrm{Tr}\Big[vf_{Hx}^{-1}(i\omega)\big\{\ln\big[1 - \chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)$$
$$- \delta\chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)\big] - \ln\big[1 - \chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)\big]\big\}$$
$$+ v\delta\chi^{\mathrm{KS}}(i\omega)\Big]$$
$$= -\int_0^\infty \frac{d\omega}{2\pi}\mathrm{Tr}(vf_{Hx}^{-1}(i\omega)\ln\{1 - \delta\chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)$$
$$\times [1 - \chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)]^{-1}\} + v\delta\chi^{\mathrm{KS}}(i\omega)).$$

By adding and subtracting $v[1 - \chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)]^{-1}\delta\chi^{\mathrm{KS}}(i\omega)$, we split AE_c in two terms, $\mathrm{AE}_c = \mathrm{AE}_c^{\mathrm{QP}} + \mathrm{AE}_c'$, and following the arguments of Niquet and Gonze [37], AE_c' vanishes, and what is left is

$$\mathrm{AE}_c = \int_0^\infty \frac{d\omega}{2\pi}\mathrm{Tr}[\{v[1 - \chi^{\mathrm{KS}}(i\omega)f_{Hx}(i\omega)]^{-1} - v\}\delta\chi^{\mathrm{KS}}(i\omega)]$$
$$= \int_0^\infty \frac{d\omega}{2\pi}\mathrm{Tr}\{[W(i\omega) - v]\delta\chi^{\mathrm{KS}}(i\omega)\}.$$

This is exactly the correlation contribution to the conduction-band matrix element of the GW self-energy in the $GW\Gamma$

TABLE V. Absolute values of ϵ_c and $\Delta_c^\pm$ together with the sum and differences of $\Delta_c^\pm$ contributing to the band gaps and centers respectively. The Δ columns show the difference upon inclusion of the kernel.

	ϵ_c from total energies			Δ_c^+		Δ_c^-		$\Delta_c^- + \Delta_c^+$			$(\Delta_c^+ - \Delta_c^-)/2$		
	RPA	rALDA	Δ	G_0W_0	$G_0W_0\Gamma_0$	G_0W_0	$G_0W_0\Gamma_0$	G_0W_0	$G_0W_0\Gamma_0$	Δ	G_0W_0	$G_0W_0\Gamma_0$	Δ
Si	−1.53	−1.05	0.48	−1.77	−1.56	−2.44	−2.60	−4.21	−4.16	0.05	0.34	0.52	0.19
BN	−1.75	−1.23	0.52	−3.02	−2.79	−3.65	−3.81	−6.67	−6.60	0.07	0.32	0.51	0.20
AlP	−1.50	−1.02	0.48	−2.19	−2.05	−2.45	−2.57	−4.64	−4.62	0.02	0.13	0.26	0.13
SiC	−1.64	−1.14	0.50	−2.40	−2.23	−3.08	−3.23	−5.48	−5.46	0.02	0.34	0.50	0.16
MgO	−1.00	−0.73	0.27	−2.84	−2.46	−4.40	−4.53	−7.24	−6.99	0.25	0.78	1.04	0.26
CdS	−1.13	−0.83	0.30	−2.75	−2.41	−2.76	−2.98	−5.51	−5.39	0.12	0.01	0.29	0.28
LiF	−1.59	−1.14	0.45	−1.87	−1.69	−5.96	−5.94	−7.83	−7.63	0.2	2.05	2.13	0.08
ML-MoS$_2$	−1.97	−1.27	0.70	−2.37	−2.39	−1.62	−1.60	−3.99	−3.99	0.00	−0.38	−0.40	0.02
ML-MoSe$_2$	−1.87	−1.27	0.60	−2.36	−2.48	−1.58	−1.47	−3.94	−3.95	0.01	−0.39	−0.51	0.12
ML-WS$_2$	−1.73	−1.09	0.64	−2.44	−2.40	−1.76	−1.74	−4.20	−4.14	0.06	−0.34	−0.33	0.01

TABLE VI. Absolute values of the VBM, CBM and band gaps at the K-point of the three 2D semiconductors considered here calculated with six different methods. The experimental references are given in the results section.

	LDA	HSE	G_0W_0	G_0W_0P	$G_0W_0\Gamma$	GW_0	Expt.
				MoS_2			
VBM	−6.07	−6.19	−6.53	−6.81	−5.86	−6.65	−5.77
CBM	−4.35	−3.93	−4.06	−4.53	−3.39	−4.04	−3.27
Gap	1.71	2.25	2.47	2.28	2.47	2.61	2.50
				$MoSe_2$			
VBM	−5.48	−5.58	−5.89	−6.04	−5.30	−5.92	−5.34
CBM	−4.06	−3.64	−3.81	−4.06	−3.23	−3.69	−3.03
Gap	1.43	1.94	2.08	1.99	2.07	2.23	2.31
				WS_2			
VBM	−5.59	−5.78	−6.18	−6.29	−5.50	−6.23	−5.74
CBM	−4.26	−3.95	−3.43	−3.73	−2.70	−3.16	−3.02
Gap	1.33	1.83	2.75	2.56	2.81	3.07	2.72

method:

$$\mathrm{AE}_c = \langle \phi_c | \Sigma_c(\epsilon_c) | \phi_c \rangle,$$

where

$$\Sigma_c(\mathbf{r}, \mathbf{r}', \epsilon_c) = -\int \frac{d\omega}{2\pi} G_0(\mathbf{r}, \mathbf{r}', \epsilon_c + i\omega) W(\mathbf{r}, \mathbf{r}', i\omega)$$

and $W(i\omega) = v[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)]^{-1}$.

APPENDIX B: EIGENVALUE SELF-CONSISTENCY

The effect of eigenvalue self-consistency in G (GW_0) and in both G and W (GW) is shown in Table IV. The effect of the kernel is seen to be independent of eigenvalue self-consistency.

APPENDIX C: VALUES OF ϵ_c AND $\Delta_c^{\pm}$

The absolute values of ϵ_c and $\Delta_c^{\pm}$ as well as the sum and differences of $\Delta_c^{\pm}$ contributing to the band gaps and centers, respectively, are given in Table V. The RPA and rALDA total-energy calculations were done with the implementation described in [21]. $8 \times 8 \times 8$ **k** points were used, and the correlation energy was extrapolated to an infinite plane-wave cutoff from calculations up to 500 eV. The frequency integration was done using 16 frequency points on the imaginary axis together with a Gauss-Legendre quadrature.

APPENDIX D: TWO-DIMENSIONAL MATERIALS—IP, EA, AND GAPS

VBM, CBM, and band gaps at the K-point for three 2D semiconductors calculated with various methods, relative to vacuum, are shown in Table VI. The hybrid HSE functional [38] G_0W_0, and $G_0W_0\Gamma$ are all non-self-consistent calculations on top of LDA orbitals and eigenvalues. The procedure for obtaining the experimental values is described in the main text.

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

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

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

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

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

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

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

[8] A. Marini, C. Hogan, M. Grüning, and D. Varsano, Comput. Phys. Commun. **180**, 1392 (2009).

[9] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).

[10] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B **88**, 075105 (2013).

[11] F. Bruneval, Phys. Rev. Lett. **103**, 176403 (2009).

[12] X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).

[13] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **87**, 235132 (2013).

[14] F. A. Rasmussen, P. S. Schmidt, K. T. Winther, and K. S. Thygesen, Phys. Rev. B **94**, 155406 (2016).

[15] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **115**, 119901(E) (2015).

[16] P. Scherpelz, M. Govoni, I. Hamada, and G. Galli, J. Chem. Theory Comput. **12**, 3523 (2016).

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

[18] A. Grüneis, G. Kresse, Y. Hinuma, and F. Oba, Phys. Rev. Lett. **112**, 096401 (2014).

[19] T. Olsen and K. S. Thygesen, Phys. Rev. B **86**, 081103 (2012).

[20] T. Olsen and K. S. Thygesen, Phys. Rev. B **88**, 115131 (2013).

[21] C. E. Patrick and K. S. Thygesen, J. Chem. Phys. **143**, 102802 (2015).

[22] R. Del Sole, L. Reining, and R. W. Godby, Phys. Rev. B **49**, 8024 (1994).

[23] F. Furche and T. V. Voorhis, J. Chem. Phys. **122**, 164106 (2005).

[24] R. Shaltaf, G.-M. Rignanese, X. Gonze, F. Giustino, and A. Pasquarello, Phys. Rev. Lett. **100**, 186401 (2008).

[25] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, J. Phys. Condens. Matter **22**, 253202 (2010).

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

[27] M. Shishkin, M. Marsman, and G. Kresse, Phys. Rev. Lett. **99**, 246403 (2007).

[28] W. Chen and A. Pasquarello, Phys. Rev. B **92**, 041115 (2015).

[29] G. Antonius, S. Poncé, E. Lantagne-Hurtubise, G. Auclair, X. Gonze, and M. Côté, Phys. Rev. B **92**, 085137 (2015).

[30] A. Molina-Sánchez, M. Palummo, A. Marini, and L. Wirtz, Phys. Rev. B **93**, 155435 (2016).

[31] K. Keyshar, M. Berg, X. Zhang, R. Vajtai, G. Gupta, C. K. Chan, T. E. Beechem, P. M. Ajayan, A. D. Mohite, and T. Ohta, ACS Nano **11**, 8223 (2017).

[32] M. M. Ugeda, A. J. Bradley, S.-F. Shi, F. H. da Jornada, Y. Zhang, D. Y. Qiu, W. Ruan, S.-K. Mo, Z. Hussain, Z.-X. Shen, F. Wang, S. G. Louie, and M. F. Crommie, Nat. Mater. **13**, 1091 (2014).

[33] A. R. Klots, A. K. M. Newaz, B. Wang, D. Prasai, H. Krzyzanowska, J. Lin, D. Caudel, N. J. Ghimire, J. Yan, B. L. Ivanov, K. A. Velizhanin, A. Burger, D. G. Mandrus, N. H. Tolk, S. T. Pantelides, and K. I. Bolotin, Sci. Rep. **4**, 6608 (2014).

[34] H. M. Hill, A. F. Rigosi, K. T. Rim, G. W. Flynn, and T. F. Heinz, Nano Lett. **16**, 4831 (2016).

[35] F. A. Rasmussen and K. S. Thygesen, J. Phys. Chem. C **119**, 13169 (2015).

[36] W. Mönch, *Semiconductor Surfaces and Interfaces* (Springer, Berlin, 2001).

[37] Y. M. Niquet and X. Gonze, Phys. Rev. B **70**, 245115 (2004).

[38] J. Heyd, G. E. Scuseria, and M. Ernzerhof J. Chem. Phys. **118**, 8207 (2006).