Ren, X.; Marom, N.; Caruso, F.; Scheffler, M.; Rinke, Patrick

Beyond the GW approximation: A second-order screened exchange correction

Published in:
Physical Review B

DOI:
10.1103/PhysRevB.92.081104

Published: 01/01/2015

Please cite the original version:
https://doi.org/10.1103/PhysRevB.92.081104
Beyond the GW approximation: A second-order screened exchange correction

Xinguo Ren,1,2,3 Noa Marom,4 Fabio Caruso,2,5 Matthias Scheffler,2 and Patrick Rinke2,6
1Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China
2Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany
3Synergistic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, 230026 Hefei, China
4Department of Physics and Engineering Physics, Tulane University, New Orleans, Louisiana 70118, USA
5Department of Materials, University of Oxford, Parks Road, OX1 3PH, United Kingdom
6COMP/Department of Applied Physics, Aalto University, P.O. Box 11100, Aalto FI-00076, Finland

Motivated by the recently developed renormalized second-order perturbation theory for ground-state energy calculations, we propose a second-order screened exchange correction (SOSEX) to the GW self-energy. This correction follows the spirit of the SOSEX correction to the random-phase approximation for the electron correlation energy and can be clearly represented in terms of Feynman diagrams. We benchmark the performance of the perturbative $G_0W_0$+SOSEX scheme for a set of molecular systems, including the G2 test set from quantum chemistry as well as benzene and tetracyanoethylene. We find that $G_0W_0$+SOSEX improves over $G_0W_0$ for the energy levels of the highest occupied and lowest unoccupied molecular orbitals. In addition, it can resolve some of the difficulties encountered by the GW method for relative energy positions as exemplified by benzene where the energy spacing between certain valence orbitals is severely underestimated.

DOI: 10.1103/PhysRevB.92.081104 PACS number(s): 71.15.Mb, 31.15.ae, 31.15.vq, 82.37.—j

The energy gain from adding an electron to, or the cost of removing an electron from, a molecule, a nanostructure, or a solid is a fundamental property, which can be measured by experimental techniques such as photoemission or inverse photoemission spectroscopy. The first-principles theory of choice for describing such charged excitations in solid-state physics has been many-body perturbation theory in the GW approximation [1], due to its balance between accuracy and computational expense [2–4]. Recently, the GW approach has also increasingly been applied to molecules and nanosystems [5–12].

An accurate determination of single-particle excitation energies is indispensable in many areas of chemistry, physics, materials science, and nanoscience. However, despite the success of the GW approach, several challenges remain. While some of GW’s shortcomings can be attributed to the starting-point dependence [11,13–15] of the common perturbative ($G_0W_0$) scheme, others persist also in self-consistent schemes [10,16]. For example, in solids, the binding energies of semiconductors states tend to be underestimated [17,18], whereas band gaps of polar materials are often severely underestimated [19]. Also an assessment of $d$- and $f$-electron compounds is only just emerging with partial success [20–23]. For finite systems, $G_0W_0$ based on a Perdew-Burke-Ernzerhof hybrid functional (PBE0) [24] reference yields excellent vertical ionization energies (IEs) for molecules with an average deviation from experiment of only 3% [9,11,12,15]. However, relative energy differences in the full spectrum are not always this accurate. A good example is the benzene molecule, for which the energy separation of the two states just below the highest-occupied-molecular-orbital (HOMO) is considerably underestimated in $GW$ ($\sim 0.1$ eV compared to $\sim 0.6$ eV in experiment) [15]. For molecules with lone pair orbitals, the energy ordering of the first few valence orbitals can be incorrect [15], whereas for some acceptor molecules with positive electron affinities (EAs), e.g., tetracyanoethylene (TCNE), $GW$ tends to overestimate the EAs substantially, irrespective of the starting point or the self-consistent scheme. These are just a few examples to illustrate the need to go beyond the GW approximation.

Beyond-GW schemes, so-called “vertex corrections,” have a long history [25–30]. However, simple, computationally efficient vertex corrections [25] have almost no effect on the GW description, while more elaborate schemes are computationally so involved that they have only been applied to the homogeneous electron gas [26,29], the Hubbard molecule [31], or simple solids such as silicon or argon [25,27,28,32]. Thus, there is currently no generally accepted way to go beyond the GW approach for real solids, molecules, or nanostructures. Here, we focus on finite systems and present a self-energy correction to GW that derives rigorously from many-body perturbation theory, is computationally tractable, and improves on $GW$ in almost all the cases mentioned above.

We propose to go beyond $GW$ by adding a subset of higher-order exchange-type diagrams. This additional self-energy term is inspired by the renormalized second-order perturbation theory (rPT2) [33] for the electron correlation energy that some of us have developed recently. rPT2 improves on the random-phase approximation (RPA) [34–36] with second-order screened exchange (SOSEX) [37,38] and a renormalized single-excitation term (rSE) [39]. This yields a much more balanced description than RPA alone or RPA combined with either SOSEX or rSE can achieve [33]. The exchange-correlation diagrams in RPA are topologically identical to those of the (perturbative) $G_0W_0$ approach [40,41]. $GW$ thus provides the corresponding self-energy to RPA. Here we demonstrate that, in a similar fashion, we can derive a self-energy that corresponds to the SOSEX term. The rSE term is automatically included and thus does not appear explicitly in the self-energy.
The RPA correlation energy can be computed as \[ E_c^{\text{RPA}} = \frac{1}{2} \sum_{ij,ab} \langle ij|ab \rangle T_{jb,ia}. \] (1)

\( T_{jb,ia} \) are the ring coupled-cluster doubles amplitudes and \( \langle ij|ab \rangle \) the bare two-electron Coulomb integrals,

\[
\langle ij|ab \rangle = \int dx_1 dx_2 \frac{\psi_i^*(x_1)\psi_j^*(x_2)\psi_a(x_1)\psi_b(x_2)}{|r_1 - r_2|}.
\] (2)

Here \( x = (r, \sigma) \) is a combined spatial-spin variable, and \( i,j \) and \( a,b \) refer to occupied and unoccupied Kohn-Sham single-particle spin orbitals, respectively. The RPA+SOSEX correlation energy can then be obtained by antisymmetrizing the Coulomb integrals in Eq. (1) [37], i.e., \( \langle ij|ab \rangle \rightarrow \langle ij||ab \rangle = (ij|ab) - (ij|ba) \).

To apply the same strategy to the self-energy, we separate the \( GW \) self-energy into its exact exchange and the remaining correlation part \( \Sigma^{GW} = \Sigma_c + \Sigma^{GW}_c \). The \( \Sigma^{GW}_c \) term is then given by

\[
\Sigma^{GW}_c(1,2) = iG(1,2)[W(1,2) - v(1,2)],
\] (3)

where \( G \) is the Green’s function of the interacting electron system, \( v \) is the bare Coulomb interaction, and \( W \) is the screened Coulomb interaction. The numbers are a shorthand notation for combined spatial, spin, and time variables [i.e., \( 1 \equiv (r_1, \sigma_1, t_1) \)]. Here, as usual, \( W \) is evaluated at the RPA level, i.e.,

\[
W(1,2) = v(1,2) + \int d3d4v(1,3)\chi_0(3,4)W(4,2),
\] (4)

where \( \chi_0 \) is the irreducible polarizability,

\[
\chi_0(1,2) = -iG(1,2)G(2,1).
\] (5)

\( \Sigma^{GW}_c \) can thus be rewritten as

\[
\Sigma^{GW}_c(1,2) = \int d3d4 G(1,2)v(1,3)G(3,4)G(4,3)W(4,2),
\] (6)

which is illustrated by Feynman diagrams in Fig. 1(a). The diagrams reveal that \( \Sigma^{GW}_c \) arises from screened second-order direct (Coulomb) interactions.

The key step comes next: By exchanging the connection between the Green’s function and the (bare and screened) Coulomb interaction lines, we arrive at a second-order screened exchange diagram, which we call the SOSEX self-energy

\[
\Sigma^{\text{SOSEX}}(1,2) = -\int d3d4 G(1,4)v(1,3)G(4,3)G(3,2)W(4,2).
\] (7)

The procedure is illustrated diagrammatically in Fig. 1(b) [43].

Traditionally, perturbation theories have been mostly carried out by considering \( \Sigma^{\text{SOSEX}} \) to be a functional of either \( G \) and \( W \) or a functional of \( G \) and \( v \). The diagrammatic representation of \( \Sigma^{GW}_c \) in Fig. 1(a) suggests that one may combine both options, which makes the construction of diagrams more flexible, but carries the danger of double counting. By further expanding \( W \) in terms of \( v \), we ensured that no term in our theory is counted more than once. The \( \Sigma^{GW} \) and \( \Sigma^{\text{SOSEX}} \) diagrams are the only two nontrivial self-energy diagrams which can be constructed with one \( v \) line and one \( W \) line, in analogy to the conventional second-order self-energy that is given in terms of the two bare Coulomb lines [9,44,45].

The construction of different diagrammatic series is guided by different principles, e.g., conservation of certain properties [46,47], positivity of the spectral function [48] or size consistency [45]. Our SOSEX self-energy in Eq. (7) can be viewed as a functional derivative of a SOSEX-type correlation energy with respect to the Green’s function, while keeping the screened Coulomb interaction fixed. This SOSEX-type correlation energy is, however, different from the coupled-cluster SOSEX [37] and the adiabatic-connection SOSEX [49] discussed in the literature. In future work, we will put our double expansion in \( v \) and \( W \) on a more rigorous footing [50].

It should be noted that the SOSEX correction to \( GW \) proposed here is different from a straightforward second-order expansion in terms of \( W \), as originally formulated by Hedin [1], for which the SOSEX self-energy diagram [cf. Fig. 1(b)] contains two screened Coulomb lines. If one expands \( W \) in terms of \( v \) for both self-energies, one will find the two approaches are identical to leading order (second order in \( v \)), but differ for higher orders. For example, at third order in \( v \), the SOSEX self-energy is only half of Hedin’s second-order self-energy in \( W \), because the latter contains two topologically equivalent diagrams while the former only picks up one of them. The SOSEX self-energy diagrams are a subset of those included in Hedin’s self-energy, and hence can be considered as an approximation to the latter. Compared to the SOSEX self-energy proposed here, a full treatment of the second-order self-energy in \( W \) is numerically much more challenging because of the presence of a double frequency integration. Historically, it was first examined numerically for silicon by Bobbert and van Haeringen [27] with a plasmon-pole approximation for \( W \), but no appreciable change to \( G_0 W_0 \) was observed. Very recently, Grünweis et al. [30] implemented an approximate version of this scheme using static \( W \)’s and found...
that the ionization energies and d-electron binding energies of solids were improved.

Returning to the rPT2 analogy, rPT2 contains a third term that arises from single excitations [33]. Close inspection reveals that the single-excitation contributions as included in rSE only leads to improper self-energy diagrams [51], i.e., merely trivial repetitions of the irreducible self-energy part. As demonstrated in the Supplemental Material [51], the irreducible part of the rSE self-energy is nothing but the difference rSE only leads to improper self-energy diagrams [51], i.e., reveals that the single-excitation contributions as included in term that arises from single excitations [33]. Close inspection solids were improved.

We use the notation @DF to denote the density functional (DF) starting point that the perturbative calculation is based on [e.g., G0W0@PBE refers to a Perdew-Burke-Ernzerhof (PBE)

\[ G_0(x_1, x_2; i \omega) = \sum_p \frac{\langle \psi_p(x_1) \rangle \langle \psi_p(x_2) \rangle}{i \omega - \epsilon_n}. \]  

(8)

For numerical simplicity we work on the imaginary frequency axis, and the results will be analytically continued to the real axis at the end. Using (6), (7), and (8), the diagonal matrix elements of \( \Sigma_{c}^{GW} \) within the set of orbitals \( \{\psi_p(x)\} \) are given by

\[ \langle \psi_p \rangle \Sigma_{c}^{GW} + \Sigma_{c}^{SOSEX} \langle \psi_p \rangle |\psi_p\rangle = -\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \sum_{q,r} \sum_{\rho,\sigma} \langle \rho | W(i\omega') | \sigma \rangle \langle \rho | (i\omega + i\omega' - \epsilon_r) (i\omega' + \epsilon_q - \epsilon_r) | \sigma \rangle. \]

(9)

where \( \langle \rho | W(i\omega') | \sigma \rangle \) are the two-electron integrals for the screened Coulomb interaction. In Eq. (9), the summation over spin-orbital indices \( p,q,r \) runs over both occupied and unoccupied states, and \( f_q \) and \( f_r \) are Fermi occupation factors. In this work we only consider closed-shell molecules, and hence \( f_q = 1 \) for occupied states, and 0 for unoccupied states.

Equation (9) has been implemented in the local-orbital based, all-electron Fritz Haber Institute ab initio molecular simulations (FHI-aims) code package [9,52]. In FHI-aims two-electron Coulomb integrals are evaluated using the resolution-of-identity (RI) technique. The implementation details for correlated methods (including RPA, RPA+SOSEX, and GW) have been documented in Refs. [9,33], which our GW+SOSEX implementation follows closely. Computationally, a standard RI implementation of Eq. (9) using canonical molecular orbitals scales as \( O(N^2) \) with respect to the system size \( N \). Ways to reduce the scaling are foreseeable if one exploits the locality of the atomic orbitals and the Green’s function.

Next, we demonstrate the performance of \( G_0W_0+SOSEX \). We use the notation @DF to denote the density functional (DF) starting point that the perturbative calculation is based on [e.g., \( G_0W_0@PBE \) refers to a Perdew-Burke-Ernzerhof (PBE) [53] based \( G_0W_0 \) calculation]. Unless otherwise stated, the high-quality tier 4 FHI-aims-2009 basis sets [52] are used in the calculations below [54].

We first check how well GW+SOSEX performs for ionization energies, considering a subset of 50 molecules from the G2 test set [56]. The vertical IEs (vIEs) used here as reference are those obtained by the G2 theory itself [56], corrected by the difference between adiabatic ionization energies (aIEs) and vIEs given by experiment:

\[ E^{\text{vIE(Ref.)}} = E^{\text{aIE(G2)}} + E^{\text{IE(Expt.)}} - E^{\text{IE(Expt.)}}. \]

(10)

The error distribution of the vIEs is presented in Fig. 2 for \( G_0W_0 \) and \( G_0W_0+SOSEX \), based on both PBE and PBE0. The \( G_0W_0 \) results were already presented in Ref. [9] where it was shown that \( G_0W_0@PBE \) systematically underestimates vIEs, whereas \( G_0W_0@PBE0 \) yields results that are in good agreement with experiment on average. This behavior does not change when comparing to the reference values obtained from the G2 theory, as can be seen from Fig. 2. Adding the SOSEX correction, the aIEs are systematically increased, and now \( (G_0W_0+SOSEX)@PBE \) yields a vanishingly small mean error (ME) of only \(-0.01\) eV, and a mean absolute error (MAE) of \( 0.28 \) eV that slightly overestimates vIEs, in contrast to \( (G_0W_0+SOSEX)@PBE0 \) that slightly underestimates vIEs. Not surprisingly, as a perturbative treatment, \( G_0W_0+SOSEX \) still shows noticeable dependence on the reference orbitals. However, on average this dependence is not as significant as the \( G_0W_0 \) case, as measured by the respective ME differences. In the Supplemental Material [51] we further present \( G_0W_0+SOSEX \) benchmark results for two other subsets of G2, and compare them to reference results obtained with coupled-cluster theory with single, double, and perturbative triple truncations [CCSD(T)]. The message from these additional tests is the same as conveyed in Fig. 2.
The benzene molecule is our next example. The HOMO is a twofold degenerate \( \pi \) state (\( e_{1g} \) representation), whose \( G_0W_0 \) energy has been calculated by various groups [9,57–59]. The general agreement of \( G_0W_0 \) and the experimental vIE is satisfactory (cf. Fig. 3). However, our real interest lies in the energy spacing of the next two states below the HOMO (denoted as HOMO-1 and HOMO-2, respectively). These are the twofold degenerate \( e_{2g}(\sigma) \) and the nondegenerate \( a_{1u}(\pi) \) states [60,61]. The exact energy spacing between these two peaks varies slightly between different experiments [61–63], but is larger than 0.5 eV in all cases. However, in all \( G_0W_0 \) as well as in self-consistent \( GW \) (sc\( GW \)) calculations [10] this separation is vanishingly small (\( \sim 0.1 \) eV). The two peaks merge into one when a Gaussian broadening of 0.3 eV is applied as in Fig. 3. Most importantly, the splitting in \( G_0W_0 \) is independent of the starting point and therefore does not depend on the self-interaction error that might be present in the preceding mean-field calculation. We thus conclude that the significant underestimation of the \( e_{2g}(\sigma) - a_{2u}(\pi) \) splitting in benzene is an intrinsic \( GW \) error that requires a correlation treatment beyond \( GW \).

Figure 3 shows that the addition of a bare second-order exchange term (\( G_0W_0+20X \)) increases the splitting of the two peaks [15]. Unfortunately, \( G_0W_0+20X \) overshoots, leading to an energy spacing of \( \sim 2 \) eV. Screening the second-order exchange term, as in the \( G_0W_0+\)SOSEX scheme, moves the \( a_{2u} \) peak back. Now the \( e_{2g}(\sigma) - a_{2u}(\pi) \) separation is \( \sim 1 \) eV, which although not in perfect agreement with experiment, is a significant improvement upon \( GW \).

The difficulty that \( GW \) encounters for benzene can be understood in terms of the delicate balance between the \( \pi-\pi \) (i.e., HOMO and HOMO-2) and the \( \pi-\sigma \) splitting (i.e., HOMO and HOMO-1). The HOMO and the HOMO-2 are both bonding \( \pi \) states that derive from the \( p_z \) orbital of the C atoms. However, their nodal structures are different: the lower \( a_{2u}(\pi) \) state has no nodes, whereas the higher \( e_{1g}(\pi) \) states have two nodes. The HOMO-1 state, on the other hand, exhibits a completely different \( \sigma \) bonding character arising from the \( p_{x/y} \) orbitals of the C atoms. Describing the relative energy positions of these different orbitals is a challenging task. Although \( GW \) improves the KS spectrum in general, the self-screening problem of \( GW \) [65,66] affects the relative energy positions for orbitals that have drastically different bonding characters. Figure 3 illustrates that the higher-order exchange contribution controls the magnitude of the \( e_{2g}(\sigma) - a_{1u}(\pi) \) splitting in benzene and that \( GW+\)SOSEX captures the essential physics that is missing in \( GW \). Predicting relative energies accurately is of paramount importance for the level alignment at heterointerfaces as found in microelectronics, organic electronics, and at hybrid interfaces. We thus expect \( GW+\)SOSEX to play a significant role for such systems in the future.

The third example we consider here is the TCNE molecule—an excellent electron acceptor because of its large EA value, which is a highly desired property in certain electronic devices. In Fig. 4 we present several \( GW \) spectra...
for TCNE in comparison to the experimental photoemission data [67]. The vertical EA (vEA) values of 3.05 eV obtained from CCSD(T) [69] and 2.91 eV from the complete active space method with a second-order perturbation (CASPT2) [70], as well as the experimental adiabatic EA value (3.16) are indicated by vertical dashed lines. We observe that all GW methods significantly overestimate the EA, even $G_0W_0$@HF, which usually yields the lowest unoccupied molecular orbital with the highest energy (smallest vEA). $(G_0W_0+SOSEX)@PBE$ improves on $G_0W_0$@HF by more than 0.3 eV, reducing vEA from 3.61 to 3.30 eV. Moreover, we note that the improved description of $(G_0W_0+SOSEX)$ for the unoccupied states does not deteriorate the occupied states. $(G_0W_0+SOSEX)@PBE$ also yields the best valence spectrum.

To summarize, we have proposed a GW+SOSEX scheme for self-energy calculations. It goes systematically beyond the GW approximation and has a clear diagrammatic representation. We note that a perturbative expansion may appear intuitively clear and systematic. However, it always also carries a subjective component. Thus, the final proof of the quality of such expansion is only provided by comparing to (essentially) exact results. This is the strategy taken in the present work. Extensive benchmark calculations were carried out, which show that $G_0W_0+SOSEX$ gives vIEs of atoms and small molecules in excellent agreement with the best theoretical reference values. $G_0W_0+SOSEX$ also improves on GW for the relative energy position of molecular orbitals in difficult cases such as benzene and for electron affinities of strong acceptors such as TCNE, which $G_0W_0$ consistently overestimates.

Work at University of Science and Technology of China was supported by the National Natural Science Foundation of China Award No. 11374276. Work at Tulane University was supported by the Louisiana Alliance for Simulation-Guided Materials Applications (LA-SiGMA), funded by the National Science Foundation (NSF) Award No. EPS-1003897. This research has also been supported by the Academy of Finland through its Centres of Excellence Program (Project No. 251748).

[43] The sign difference between Eq. (6) and Eq. (7) arises from the fact that the Feynman diagram for $\Sigma_{\text{SOSEX}}$ contains no closed fermionic loop, while $\Sigma_{\text{GW}}$ has one.
[50] It appears that the SOSEX self-energy is not a $\Phi$- or $\Psi$-derivable theory [47], and one may then question its conserving property which is a relevant issue when running self-consistent calculations.
[54] The convergence behavior of FHI-aims-2009 basis sets for $G_0W_0$-type calculations has been thoroughly investigated in Ref. [9].
[64] The PBE-relaxed (with tier 2 basis) geometry is used in the calculation.