Exploring the Statically Screened G3W2 Correction to the GW Self-Energy: Charged Excitations and Total Energies of Finite Systems

Arno Förster and Lucas Visscher

Theoretical Chemistry, Vrije Universiteit, De Boelelaan 1083, NL-1081 HV, Amsterdam, The Netherlands
(Dated: March 18, 2022)

Electron correlation in finite and extended systems is often described in an effective single-particle framework within the GW approximation. Here, we use the statically screened second-order exchange (SOX) contribution to the self-energy (G3W2) to calculate a perturbative correction to the GW self-energy. We use this correction to calculate total correlation energies of atoms, relative energies, and to increase the lowest unoccupied molecular orbitals (LUMO) QP energies. We show here that the statically screened G3W2 correction consistently increases the LUMO QP energies, while no consistent trend can be observed for the HOMO levels. Also, confirming previous results, the magnitude of the correction is much smaller with the statically screened interaction, than with the dynamically screened one. qsGW by itself is shown to be an excellent method for the calculation of charged excitation of finite systems, and it can not consistently be improved upon by the G3W2 correction. For range-separated hybrid starting points, the description of fundamental gaps and HOMO QP energies is slightly worsened. However, tremendous improvements upon the GW LUMO energies leading to almost perfect agreement with high-level coupled cluster reference methods are observed. The evaluation of the statically screened G3W2 correction only comes with small additional computational cost compared to G0W0 for systems with up to 100 atoms and should therefore be suitable for practical applications.

I. INTRODUCTION

Electron densities, total energies and spectral functions of an interacting many-electron system can be extracted from its single-particle Green’s function. Instead of expanding the Green’s function directly in powers of the electron-electron interaction, it is usually easier to expand the irreducible self-energy and to calculate the interacting Green’s function via Dyson’s equation. The difficulty is then to find an approximation which captures the most important correlation effects and remains computationally tractable.

In Hedin’s GW approximation (GWA), the self-energy is obtained as the first term of an expansion in terms of a screened electron-electron interaction where screening is (usually) calculated in the random-phase approximation (RPA). Compared to the Hartree-Fock (HF) approximation, the first-order term in the expansion of the self-energy in terms of the bare electron-electron interaction, it takes into account that the electron-electron interaction at large distances is screened by the presence of other electrons. For properties which are mainly dominated by long-range correlation effects, such as fundamental gaps or ionization potentials in organic molecules, the GWA is therefore often very accurate. This is especially true for the most common computational approach, where the full solution of Dyson’s equation is bypassed and the quasiparticle (QP) energies are calculated from a non-interacting propagator. In the static long-range limit, the errors introduced by the neglect of self-consistency and the absence of higher order terms in the polarizability and self-energy cancel to a large extent by virtue of the Ward identity. When short-range correlation becomes important, the GWA is less accurate.

From the GW self-energy one can calculate correlation energies using the Klein functional or evaluated with a Kohn-Sham (KS) Green’s function, one obtains (particle-hole) RPA or often derived in the framework of the adiabatic connection (AC) fluctuation-dissipation theorem or as a subset of terms in the coupled cluster (CC) singles and doubles (CCD) expansion. Since the effect of charge screening dominates in this limit, the RPA describes long-range electron correlation very accurately. At short electron-electron distances, charge screening is less important and the RPA drastically overestimates the magnitude of electron correlation. In this limit it is crucial to take into account higher-order contributions (vertex corrections) to the self-energy which can then dominate. These contributions are fundamentally short-ranged and become less and less important for large electron-electron distances.

Besides range-separation based approaches, the next-to-leading-order term, the second-order exchange (SOX), is therefore the most obvious starting point to treat electron correlation beyond the RPA. The most common correction is the second-order screened exchange (SOSEX) correction. SOSEX can be derived from the CC doubles (CCD) equations by neglecting ladder and “crossed-ring” diagrams. SOSEX has been applied to total energies of atoms, and molecules and to the uniform electron gas and several extensions...
to it have been proposed as well.\cite{50,51} In another variant the SOX term is renormalized by resummation of hole-hole ladder diagrams.\cite{51,52} Typically, these approaches improve total correlation energies,\cite{27,28,44,45,49} but relative energies are not necessarily improved.\cite{16}

The close connection between RPA correlation energies and the GW method suggests that, beyond-RPA methods might also be used to improve over the GW method for charged excitations. For example, Ren et al. applied SOSEX to molecule and Maggio and Kresse,\cite{51} used a similar method to calculate HOMO QP energies for small molecules. It is, however, more common to formulate beyond-GW schemes by taking into account vertex corrections in Hedin’s equations.\cite{49} Kutepov has solved Hedin’s equations in a fully self-consistent fashion including several vertex corrections.\cite{3} He has applied the second-order self-energy variant (coined GW + G3W2) to a wide range of metals, semiconductors, and insulators\cite{55,56} and observed major improvements over fully self-consistent GW for spectral properties using the GW + G3W2 approximation to the self-energy. According to his work, the vertex corrected calculations should be performed without any constraining approximations. This means, Hedin’s equations should be solved fully self-consistently and the frequency dependence of all quantities should be properly accounted for. As many other authors\cite{55,56} he also emphasized the importance of vertex corrections in the polarizability.

On the other hand, the GW + G3W2 self-energy has also been applied by Grünèis et al.\cite{21} to calculate QP energies in solids in a more approximate fashion. In contrast to Kutepov, Grünèis et al.\cite{21} used a statically screened interaction and calculated perturbative corrections to quasiparticle self-consistent GW (qsGW)\cite{57,58} QP energies. They also combined this method with a beyond RPA screening of the Coulomb interaction using a static exchange-correlation kernel. As Kutepov, they reported major improvements over the GW method for the band structures of solids. Comparing their results to Kutepov’s, the logical conclusion would be, that the errors introduced by the static approximation, the use of a non-interacting Green’s function, and the perturbative treatment of the G3W2 cancel to a large extent. A recent study by Kutepov has confirmed this conjecture.\cite{55}

From a pragmatic point of view this is of course convenient, since all of these approximations come with drastically reduced computational cost compared to the rigorous, self-consistent formalism.\cite{29} In a recent study, the GW + G3W2 self-energy has also been employed by Rinke, Ren and coworkers\cite{25} to ionization potentials and electron affinities of molecules. Rinke, Ren and coworkers used the dynamically (RPA) screened interaction and did not consider any self-consistency in solving Dyson’s equation. Improvements over GW were found to be substantial, especially for electron affinities and the dependence on the starting point was reduced.\cite{25}

Encouraged by these results, we herein follow Grünèis et al.\cite{21} and calculate perturbative G3W2 corrections to the GW self-energy. We follow their approach in using non-interacting Green’s functions throughout and exclusively consider the statically screened G3W2 self-energy. However, following Rinke, Ren and coworkers,\cite{25} we do not consider any beyond-RPA screening in our work. One should expect an expansion of the polarizability in terms of W to converge much faster for finite, than for extended systems, since charge screening will be much weaker.\cite{25} It is known for finite systems, that inclusion of the vertex in the polarizability alone does not improve\cite{25,55} and might even deteriorate the accuracy of charged excitations compared to GW.\cite{25} While Grünèis et al.\cite{21} investigated periodic systems, we focus here on the calculation of correlation energies as well as QP energies for a wide range of finite systems: We show that massively improved absolute correlation energies are obtained compared to the RPA and also that relative energies are improved for the majority of the considered systems. For QP energies we show, that the G3W2 correction provides improvements over G0W0 evaluated with KS orbitals and energies from range-separated hybrid functionals, especially for electron affinities. However, no consistent improvements over qsGW are found which we show to be very accurate by itself.

The remainder of this work is organized as follows: In the next section, we introduce the G3W2 self-energy and discuss how it can be used efficiently to calculate corrections to GW QP energies and RPA correlation energies. In section II, we first benchmark the performance of the GW + G3W2 self-energy for absolute correlation energies as well as relative energies of several organic molecules and then proceed to scrutinize the effect of the G3W2 self-energy correction to G0W0 on ionization potentials, electron affinities and fundamental gaps of molecular systems. Finally, we summarize and conclude this work in section ref. V.

II. THEORY

Dyson’s equation\cite{29}

\begin{equation}
G(1, 2) = G^0(1, 2) + G^0(1, 3)\Sigma(3, 4)G(4, 2),
\end{equation}

relates the interacting Green’s function G of a many-electron system to the Green’s function of a non-interacting reference system G0, via the non-Hermitian self-energy operator Σ. We have used the notation 1 = (r1, σ1, ω1) and integration over indices with upper bars is implied. Following Hubbard, Phillips, and Hedin\cite{29} the self-energy can be expanded in terms of the screened electron-electron interaction. The first two terms in this expansion are

\begin{equation}
\Sigma_{GW+G3W2}(1, 2) = \Sigma_{GW}(1, 2) + \Sigma_{G3W2}(1, 2),
\end{equation}

with

\begin{equation}
\Sigma_{GW}(1, 2) = iG(1, 2)W(1, 2),
\end{equation}

\begin{equation}
\Sigma_{G3W2}(1, 2) = iG(1, 2)W(1, 2).\end{equation}
approach will fail. As Σ, P can be expanded in powers of the screened interaction where to zeroth order in W, the RPA is obtained,

\[ P^{RPA}(1, 2) = -iG(1, 2)G(2, 1). \]  

Solving these Equations fully self-consistently is cumbersome due to the complicated frequency dependence of the interacting Green’s function and the second-order self-energy. Since we are interested in a computationally efficient approach, we introduce a few approximations:

First, all quantities above are exclusively evaluated with the QP approximation. Since we are interested in a computationally efficient approach, we introduce a few approximations. We first introduce the non-interacting Green’s functions, which implies that we need to approximate (1). Second, the range-separated hybrid calculation as input. This is usually an excellent starting point since the screened Coulomb potential in the RPA reduces to

\[ W(r \to \infty, \omega = 0) = e^{-\lambda r}/r, \]  

with λ proportional to the Fermi-momentum. This is the basic idea behind the construction of range-separated hybrid (RSH) functionals. Solving the KS equations using such a potential emphasizes the importance of screening of the Coulomb interaction for large electron-electron distances and a perturbative correction with eq. (8) is usually rather small. The range-separation parameter can then also be adjusted in a way that the eigenvalues from the RSH calculation gives very accurate QP energies by itself.

The second approach we employ here is QP self-consistent GW (qsGW). For a given approximation to the self-energy an optimized KS potential can be obtained by solving the [linearized] Sham-Schlüter equations (SSE) and QP self-energy. With Σ = GW, the SSE can be solved as in the exchange-only case if the self-energy is approximated as static. One then obtains a statically screened exchange (SEX) correction depending on ReΣ(ωP). The derivation via the SSE leaves some ambiguity in the choice of the off-diagonal elements. As shown by Ismail-Beigi one can also derive an optimized potential via minimization of the gradient of the Klein functional leading to the following Hermitian form which has already been suggested by van Schilfgaarde et al.

\[ \left[ v^{OP}_{pq} \right] = \frac{1}{2} \left\{ \text{Re} \left[ \Sigma_{xc,p} (\epsilon_p) \right] \right\} \]  

Using this potential, the KS equations are then solved self-consistently. It should be stressed that in fully self-consistent GW, an interacting Green’s function is constructed in each iteration by solving (1) while in qsGW a non-interacting Green’s function is constructed in each iteration. qsGW should thus rather be seen as an approach to optimize a non-interacting Green’s function. It can also be combined with (COHSEX) approximation.
interacting Green’s function. Since we exclusively work with non-interacting Green’s functions we will suppress the superscript \((0)\) in the following. The time-ordered free propagator \(G\) is diagonal in the MO basis and can be expressed in terms of greater \(G^{\geq}\) (particle) and lesser \(G^{\leq}\) (hole) propagators,

\[
G_p(t_1,t_2) = \Theta(t_1 - t_2)G^{\geq}_p(t_1,t_2) + \Theta(t_2 - t_1)G^{\leq}_p(t_1,t_2),
\]

where

\[
G^{\geq}_p(t_1,t_2) = G^{\geq}_p(t_1 - t_2) = -i\tilde{f}(\epsilon_p) e^{-i\epsilon_p(t_1 - t_2)}
\]

\[
G^{\leq}_p(t_1,t_2) = G^{\leq}_p(t_1 - t_2) = if(\epsilon_p) e^{-i\epsilon_p(t_1 - t_2)}.
\]

Here, \(f\) denotes the zero-temperature Fermi-function and \(\tilde{f} = 1 - f\), and we have indicated that the Green’s function only depends on the time-difference at equilibrium. The single-particle energies in (13) and (12) are understood to be relative to the chemical potential which we place in the middle of the HOMO-LUMO gap.

The structure of the self-energy terms has been scrutinized by Stefanucci, Pavlyukh, van Leeuwen and coworkers.\cite{101,102,103} We follow their work and use the framework of the Keldysh formalism to discuss the \(G^{\geq}\) (anti-time-ordered) Green’s function in the middle of the HOMO-LUMO gap.

\[
\Sigma^{GW,\leq} = \text{Diagram}
\]

\[
\Sigma^{GW2,\leq} = \text{Diagram}
\]

FIG. 1. Diagrammatic representation of the lesser components of the GW (top) and \(G^{\leq}\) (bottom) self-energy parts. Time-ordering from left to right is implied. The signs at the vertices denote which branches of the Keldysh contour are connected by \(W\). Greater components are obtained by exchanging plus and minus signs. Straight lines denote propagators and the double wiggly line is the screened Coulomb interaction.

The structure of the self-energy terms has been scrutinized by Stefanucci, Pavlyukh, van Leeuwen and coworkers.\cite{101,102,103} We follow their work and use the framework of the Keldysh formalism to discuss the \(G^{\leq}\) self-energy.\cite{102} In the Keldysh formalism, we work on the contour \(C = C^+ \cup C^-\), with \(C^+\) being the backward branch and \(C^-\) being the forward branch. The time-ordered (anti-time-ordered) Green’s function \(G^{\geq} (G^{\leq})\) is built from field-operators evolving on \(C^- (C^+)\) while lesser \((G^{\leq} = G^{\geq})\) and greater \((G^{\geq} = G^{\leq})\) Green’s function involve both branches and describe propagation of holes and particles, respectively. In the same way, the dynamically screened interaction can either connect both different branches on the Keldysh contour \((W^{++} = W^+ + W^-)\) or not \((W^{--} = W^- + W^+)\).

The lesser components of both contributions to the self-energy in (12) are given diagrammatically in figure 1. For the GW self-energy each component only consists of a single term (upper part of figure 1). For finite systems, the GW self-energy is most conveniently evaluated in a basis of localized atomic orbitals and imaginary time,

\[
\Sigma^{GW,\leq}_{\mu\nu}(i\tau) = -i \sum_{\mu'\nu'} G^{\leq}_{\mu'\nu'}(i\tau) W^{\leq}_{\mu\nu'}(i\tau),
\]

with \(i\tau = i\tau_1 - i\tau_2\). In our GW implementation\cite{103} we transform the self-energy to the imaginary frequency axis and analytically continue (either the diagonal elements in the MO basis in \(G_0W_0\) or the full self-energy matrix in \(q_0GW\)) to the complex plane. Greater and lesser component of the second-order contribution consist of four terms each, since both intermediate vertices can connect both branches of the Keldysh contour. Together, the four diagrams describe three distinct scattering processes.\cite{102,103} Among others, they are responsible for spectral features which do not appear in fully self-consistent GW, like the excitation of two plasmons and two particle hole pairs. In this work, these terms should be of minor relevance only since we are interested in improving QP energies and correlation energies. The last diagram in fig. 1 however, only contributes to the \(2h1p\) (1h2p) space and describes the exchange of two final particles/holes.\cite{102} In the same way as the static SOX implementation it ensures the antisymmetry of the 4-point vertex to first order in the electron-electron interaction.

Static screening comes with two distinct advantages. Obviously, it drastically reduces the computational cost. Second, only the last diagram in the expression in figure 1 remains. As an important side effect, the loss of the positive definiteness of the spectral function arising from straightforward inclusion of all four terms is avoided. The first three diagrams vanish, since a static interaction can only connect points on the same branch of the Keldysh contour.\cite{102} To see this clearly we first write down \(\Sigma^{GW2}_{\mu\nu}(\tau)\) explicitly,

\[
\Sigma^{GW2}_{\mu\nu}(\tau_{12}) = -\sum_{\lambda\lambda',\kappa\kappa'} \int d\tau_3 d\tau_4 G_{\mu'\lambda'}(\tau_{14}) W_{\mu\nu'}(\tau_{31}) G_{\lambda'\kappa'}(\tau_{34}) W_{\lambda\nu'}(\tau_{42}),
\]

Using a time-independent interaction, the integrals over the internal times can be evaluated easily and transformation to frequency space using

\[
F(i\omega) = -i \int dt F(i\tau) e^{-i\omega\tau}
\]

gives

\[
\Sigma^{GW2}_{\mu\nu}(i\omega) = i \sum_{\lambda\lambda',\kappa\kappa'} \int d\tau_{12} e^{-i\omega\tau_{12}} G_{\mu'\lambda'}(\tau_{12}) W_{\mu\nu'} W_{\lambda\nu'},
\]

which corresponds to the last diagram in figure 1 and can be expressed in terms of greater and lesser components of \(G\) only (plus and minus signs alternate). Using (11), the
integral has a simple analytic solution and transforming to the MO basis gives

\[ \left[ \Sigma_{G3W^2} \right]_{pq} (\omega) = \sum_{iab} W_{iabp} W_{ibqa} \epsilon_a + \epsilon_b - \epsilon_i - \omega \\
- \sum_{ija} W_{ija} \epsilon_a - \epsilon_i - \epsilon_j + \omega , \]

which is the same expression already employed by Grünert et al.\cite{21} The computational effort to evaluate these terms only grows as \( N^3 \) with systems size if one is only interested in a small number of diagonal elements. Calculation of the matrix elements of \( W \) will generally scale as \( N^4 \), but these will be available in some form when a GW calculation is performed. The asymptotic scaling can be further reduced by exploiting sparsity in the AO basis or in real space\cite{105,107} but we do not explore these possibilities here.

For the self-energy beyond GW, we again rely on the assumption that GW already gives rather accurate QP energies. We thus expand \( \Sigma_{G3W^2} \) around the GW QP energies. At zeroth order, we obtain

\[ \epsilon_p^{GW+G3W^2} = \epsilon_p^{GW} + \Sigma_{G3W^2}(\epsilon_p^{GW}) . \]

This is the most economical way to calculate the QP energy correction due to \( \Sigma_{G3W^2} \). Note, that on the GWA level, the self-energy is often treated to first order, which results in multiplication with a renormalization factor \( Z \). This is commonly done for solids, but it has been argued that the zeroth-order treatment is to be preferred on the GWA level\cite{110} Of course, evaluating \( \Sigma_{G3W^2} \) directly with the \( \Sigma_{GW+G3W^2} \) self-energy is possible as well as has been done in ref.\cite{73} Due to its much smaller computational cost, all values presented in this work have been obtained using equation \ref{eq:19}.

\[ \Omega_{\alpha} = \Omega_{\alpha}^{(c)} + \Omega_{\alpha}^{(e)} = \frac{1}{2} \text{Tr} \{ \ln(1 - P \rho \bar{A} V) + P \rho \bar{A} V \} , \]

where \( \Omega_{\alpha}^{(c)} \) and \( \Omega_{\alpha}^{(e)} \) denote even and odd parts of \( F \), respectively. The transformation from imaginary frequency to imaginary time only requires the (pseudo)inversion of \( \Omega_{\alpha}^{(c)} \) and \( \Omega_{\alpha}^{(e)} \), respectively. Our procedure to calculate \( \Omega_{\alpha}^{(c)} \) and \( \Omega_{\alpha}^{(e)} \) as well as \( \mathcal{T} \) and \( \mathcal{W} \) follows Kresse and coworkers\cite{111,113} The technical specifications of our implementation have been described in the appendix of ref.\cite{109} All 4-point correlation functions are expressed in auxiliary basis sets of Slater type functions which are usually 3 to 5 times larger than the primary bases\cite{103,113} The transformation between both bases is implemented using the pair-atomic density fitting (PADF) method\cite{115} For an outline of the implementation of this method, we refer to ref.\cite{114}.

We have implemented RPA and RPA+SOSSX correlation energies using \ref{eq:20}. In analogy to \ref{eq:18} and in the same way as in typical MP2 implementations in quantum chemistry codes\cite{114,115}, the last term on the r.h.s of \ref{eq:20} is evaluated directly in the MO basis. The second term on the r.h.s of \ref{eq:20} is evaluated in the space of auxiliary basis functions, following our GW implementation described in ref.\cite{103}.

\[ A. \text{ Energies} \]

The correlation energies can be calculated with different orbitals. In this work, we use orbitals from HF, PBE\cite{117,118}, and the exact exchange (EXX) only OEP\cite{93,94} in the implementation of Scuseria and
1. Relative energies

For relative energies, we use imaginary frequency and time grids of 20 points. Note, that 20 grid points are typically sufficient to converge total energies to within 1e−6 Hartree. Correlation energies are calculated using

\[ E_{CBS} = E_{QZ} - E_{QZ}^c + \frac{E_{QZ}^c * 3^3 - E_{TZ}^c}{4^3 - 3^3}, \quad (23) \]

where \( E_{QZ}^c \) (\( E_{TZ}^c \)) denotes the correlation energy at the QZ6P (TZ3P) level and \( E_{QZ} \) is the total energy at the QZ6P level. The extrapolation scheme has been shown to be suitable for correlation consistent basis sets but can not be used for KS or HF contributions. For the purpose of calculating relative energies, we assume these contributions to be converged at the QZ6P level. We use the VeryGood numerical quality for integrals over real space and distance cut-offs and the corresponding auxiliary basis set. Dependency thresholds have been set to 5e−4.

2. Total Correlation Energies

We use 32 points for imaginary time and imaginary frequency each. We use here the TZ3P and QZ6P basis sets with additional diffuse functions as described in ref. 109 and additional tight 1s and 2p functions as will be described below. Eq. (23) is used for extrapolation and numerical quality and auxiliary fit quality are set to excellent. No dependency thresholds have been set.

B. Quasiparticle Energies

Charged excitations in this work are calculated using \( \Sigma^{GW2} \). For qSGW, we set the dependency threshold to 5e−3 and perform a maximum of 15 iterations of the self-consistency cycle. Following the recommendations given in ref. 109 we use VeryGood numerical quality and the corresponding auxiliary basis set. We perform \( G_0W_0 \) calculations using PBE, PBE0, LR-CωPBE and aB97X-D orbitals and eigenvalues. We use Good numerical quality and the corresponding auxiliary basis set for all \( G_0W_0 \) calculations and set the dependency threshold to 5e−4. In ref. 103 these settings have been shown to be appropriate. All QP energies are calculated using

\[ \epsilon_{n}^{CBS} = \epsilon_{n}^{QZ} - \frac{1}{N_{QZ}^{bas}} \frac{\epsilon_{n}^{QZ} - \epsilon_{n}^{TZ}}{N_{TZ}^{bas} - N_{QZ}^{bas}}, \quad (24) \]

where \( \epsilon_{n}^{QZ} \) (\( \epsilon_{n}^{TZ} \)) denotes the value of the QP energy using QZ6P (TZ3P) and \( N_{QZ}^{bas} \) and \( N_{TZ}^{bas} \) denote the respective numbers of basis functions (in spherical harmonics so that there are 5 d and 7 f functions). This expression is commonly used for the extrapolation of GW QP energies to the complete basis set limit for localized basis functions. In ref. 109 we demonstrated that using this extrapolation scheme with Slater type basis sets good agreement with other codes is obtained for GW QP energies.

IV. RESULTS

A. Total Correlation Energies

To see how the inclusion of the \( \Sigma^{GW2} \) self-energy influences the description of electron correlation effects, we first calculate the total correlation energies of 16 atoms with in between 2 and 36 electrons. For all systems with less or equal than 18 electrons, we compare the RPA and RPA+SOSSX correlation energies to almost exact values by Froese-Fischer and coworkers and for the heavier elements we use the CCSD(T) values by McCarthy and Thakkar as references. For Argon, their CCSD(T) energy deviates from the value from ref. by only 0.01 %. We also compare them to different beyond-RPA approaches by Jiang and Engel (RPA+SOX and RPA+SOX) and Gould et al. (gRPA+). To be consistent with ref. and ref. 46, we evaluate the correlation energies with EXX only OEP orbitals (EXX for short), implemented within the KLI approximation.

To obtain an idea about the numerical quality of our RPA correlation energies, we compare them against the ones from Engel and coworkers which are free of basis set errors. We find deviations between 3 and 15 % for neutral atoms and much larger ones for cations. Clearly, our standard basis sets are not compact enough for these
systems (especially for the cations) and do not capture the full correlation energy. Therefore, we augment them with four tight 1s and two 2p functions each for TZ3P and QZ6P. Using these basis sets, total energies for all atoms deviate to the ones from ref. [27] by about 7% on average. This is not perfect but accurate enough for a qualitative comparison of the different beyond-RPA approaches.

The relative errors of correlation energies with respect to the reference values are shown in figure 2 [29]. Simple RPA@EXX overestimates the correlation energies by typically between 25% and 100%. In accordance with the expectation that the correlation energy is more and more dominated by charge screening with increasing electron number, the agreement with the exact values becomes better for larger atomic numbers. RPA+SOSSX reduced the RPA correlation energy by 29% on average, improving agreement with the reference considerably. Especially for the systems with 2 or 4 electrons, the total correlation energies are still too high. For the last three systems with 18 electrons or more, the agreement becomes much better, but the correlation energy is still overestimated by a small amount. As one can expect, RPA+SOX considerably underestimates the correlation energies. RPA+RSOX shows a tendency to underestimate the correlation energies and out of all assessed methods, the deviations to the reference energies are clearly the smallest. For gRPA+, no clear trend in any direction can be identified.

In summary, the results presented here show that the GW+G3W2 self-energy improves the description of electron correlation over GW. This, however, does not necessarily imply improvements for properties like relative energies and charged excitations in realistic systems. For example, RPA is often very accurate for relative energies since errors in the correlation energies tend to cancel [10]. On the other hand, beyond-RPA methods, while improving total correlation energies, often do not yield improvements for relative energies. For this reason, we shortly assess the performance of RPA+SOSSX in comparison to RPA for relative energies, before we move on to charged excitations.

### B. Energy Differences

| dataset | RPA | RPA+SOSSX |
|---------|-----|-----------|
| ISO | 21.5 | 1.4 (14.3) | 21.1 (15.1) | 1.0 (9.1) | 1.9 (14.1) |
| FH51 | 31.0 | 1.6 (23.6) | 1.6 (21.2) | 1.6 (23.7) | 1.4 (24.0) |

TABLE I. Mean absolute deviations (MAD) and mean absolute percentage deviations (MAPD) in parentheses for RPA and RPA+SOSSX using PBE0 and HF orbitals for the ISO and FH51 datasets. The second column shows the average value of the datapoints in each set. All values are in kcal/mol.

We now compare the accuracy of RPA+SOSSX against RPA for relative energies [29]. We consider HF and PBE0 starting points. Since we do not evaluate the single contribution to the total energy, the former is formally more rigorous due to Brillouin’s theorem. On the other hand, RPA calculations with PBE0 starting points (RPA@PBE0) are usually more accurate than for HF (RPA@HF). Our benchmark is based on two datasets. The first one consists of 84 relative energies of different conformers of small and medium organic molecules, including transition states of inversion and perycyclic reactions. It is a compilation of the ISO34 [130–133], INV24 [131], and BHPERI [131,132,136] datasets, with reference values obtained using the W1-F12 [137] and W2-F12 protocol [138] except for the largest systems in INV24, where reference values have been calculated on the DLPNO-CCSD(T)/vide [143,144] CBS level of theory. The second one consists of 51 reaction energies of small organic and inorganic molecules [131,136,137] and the reference values have been obtained on the CCSD(T)-F12/CBS limit of theory.

For RPA@HF and RPA+SOSSX@HF, we report MADs and mean percentage deviations in table I. The results are promising. PBE0 reference orbitals generally leads to better agreement with experiment, for both, RPA and RPA+SOSSX. Also RPA+SOSSX@HF does not lead to improvements over RPA@HF but RPA+SOSSX@PBE0 leads to significant improvements over RPA@PBE0 for isomerization energies. Relative errors are reduced from 14% to 9%. For FH51, no improvement is found.

### C. Quasiparticle Energies

#### 1. Organic Acceptor Molecules

We first assess the performance of the G3W2 QP energies in comparison to qsGW for a set of 24 organic acceptor molecules [135]. The reference values are of CC singles doubles and perturbative triples [CCSD(T)] quality and have been extrapolated to the CBS limit. Comparing these values [29] to the experimental data shown in figure 5, we see that there are sizable differences to the CC values, especially for the fundamental gaps. Among the factors which might contribute to the discrepancies are errors in the optimized geometries, missing zero point vibrational energy corrections, and geometry relaxation after oxidation/reduction. For benzonitrile, the authors of ref. [145] calculated the values of the latter two corrections to be of the order of 0.18 and 0.14 eV, respectively. The errors in geometry or thermodynamical contributions are more difficult to assess but can be sizable as well: For example, the structure used in the calculations might not correspond to a global minimum on the potential energy surface. Finally, we note that the basis set extrapolation can also introduce some errors, especially for the larger systems where no basis sets larger than QZ were used [53]. We estimate the error of our own CBS limit extrapolation to be of the order of 50 meV for IPs and EAs of medium

---

**Note:** The text is a continuation of the previous discussion and is not a complete reflection of the original document's content. It includes a table and a few equations that are not fully transcribed here due to the limitations of the text representation. The discussion focuses on comparing the performance of different correlation methods (RPA, RPA+SOX, etc.) for relative energies and quasiparticle energies, emphasizing the challenges and factors affecting the accuracy of these calculations.

---

**Table I:**

| dataset | RPA | RPA+SOSSX |
|---------|-----|-----------|
| ISO | 21.5 | 1.4 (14.3) | 21.1 (15.1) | 1.0 (9.1) | 1.9 (14.1) |
| FH51 | 31.0 | 1.6 (23.6) | 1.6 (21.2) | 1.6 (23.7) | 1.4 (24.0) |

**Note:** The table shows mean absolute deviations (MAD) and mean absolute percentage deviations (MAPD) for RPA and RPA+SOSSX using PBE0 and HF orbitals for the ISO and FH51 datasets. The second column shows the average value of the datapoints in each set. All values are in kcal/mol.
a) Ionization Potential

b) Electron affinity

c) Fundamental Gap

FIG. 3. Ionization potentials (top), electron affinities (middle) and fundamental gaps (bottom) from qsGW and qsGW + G3W2 for a dataset of 24 organic acceptor molecules. The dashed black diagonal lines are CCSD(T) reference values. Experimental results are given for comparison as well. The values in parentheses denote mean absolute deviations (MAD) with respect to CCSD(T). All values are in eV.

organic molecules. Due to all these factors that affect the direct comparison to experiment, we exclusively use the CCSD(T) reference values for the following quantitative discussion and only show the experimental values for comparison.

a. Performance of qsGW

For a variety of solids and metals it has been found that qsGW commonly overestimates band gaps and IPs by about 15-20 % when screening is calculated within the RPA. In contrast, the qsGW IPs shown in figure 3a are in excellent agreement with the CC reference values, with no systematic overestimation. The fundamental gaps in figure 3c are even (a few exceptions aside) systematically underestimated by qsGW. Overestimation of band gaps with qsGW issue is usually explained by missing electron-hole interaction via vertex corrections in the polarizability. Inclusion of an effective two-point kernel from time-dependent (TD) DFT or the Bethe-Salpeter equation (BSE) has been demonstrated to significantly improve the agreement of band gaps and IPs with experiment, demonstrating the importance of beyond-RPA screening. For polar materials, i.e. materials with strong longitudinal-optical (LO) and transverse-optical (TO) phonon splitting, electron-phonon coupling and phonon contributions to the frequency-dependent screening can have a sizable effect on the QP spectrum as well. For example, qsGW overestimates the experimental band gap of V$_2$O$_5$ by about 100 %, which to a large extent is due to LO-TO splitting. The systems we consider here are rather small and have a planar geometry. The reduction of charge screening in low-dimensional materials has often been emphasized, for example in comparative studies on bulk and layered V$_2$O$_5$, MoS$_2$ or polythiophene. Anti-screening has been observed in a spin chain and also in finite conjugated systems. Finally, LO-TO splitting will be absent entirely. These qualitative differences most likely explain the much higher accuracy of qsGW for the systems studied herein as compared to periodic systems.

The data clearly demonstrates qsGW to be an excellent first-principle method for the description of charged excitations for these weakly correlated, organic molecules. It is worthwhile to compare the performance of this method for to previous benchmark results of different GW methods. In ref. the accuracy of a large number of GW methods has been assessed for the same dataset. Out of all GW based methods, the authors found $G_0W_0@LRC$-ωPBE to perform best (with optimized range separation parameter), with a MAD of 0.13 eV for IPs, and 0.18 eV for EAs. qsGW seems to be superior.

b. The effect of the statically screened G3W2 correction

Now we look at the effect of adding the G3W2 correction. As can be seen in figure 3, on average it lowers the qsGW IPs (see also the MDs in table II) by a small amount and decreases the EAs by a relatively larger amount, implying increasing fundamental gaps. For IPs, this slightly worsens the agreement with the
CC reference values, increasing the mean absolute deviation (MAD) from 0.09 eV to 0.12 eV. However, given the errors from the basis set limit extrapolation, this difference is not significant. Also, the G3W2 correction does not alter the MAD for the fundamental gaps. Since the statically screened G3W2 correction tends to increase the QP energies, a method which systematically overestimates IPs and EAs will be systematically improved by the G3W2 correction. This is demonstrated in figure 4 for \( W_0 \) @ ωB97-X and \( W_0 \) @ LRCωPBEh. Here, the G3W2 correction slightly improves the MAD with respect to the CC reference values for the IPs from 0.26 to 0.16 eV, and from 0.16 to 0.13 eV. For the EAs, the improvements are tremendous, and the inclusion of the G3W2 term leads to almost perfect agreement with the reference values. In table [III] these results are summarized. Despite the great performance of the G3W2 correction for the range-separated hybrids, the description of fundamental gaps is actually deteriorated, which can be considered as a serious drawback of this method.

### c. Comparison of dynamically and statically screened SOX

In table [III] we compare a few beyond-GW approaches for the same dataset. These are GW + SOSEX from ref. [47] and GW + dynamically screened G3W2 from ref. [73] (denoted as \( G_0 W_0 \Gamma^{(1)}_0 \)) by the authors of ref. [73]. These methods have all been implemented perturbatively and only differ in the way the electron-electron interaction in the SOX-term is screened (see for example figure 1 in ref. [73]). The results from refs. [47] and [73] clearly demonstrate that especially methods with bad or only mediocre performance, like \( W_0 \) @ PBE or \( W_0 @ PBE 0 \), profit immensely from vertex corrections with fully dynamical self-energy: \( G_0 W_0 \Gamma^{(1)}_0 @ PBE 0 \) is very accurate, with a MAD of 0.16 eV for the IPs and 0.09 eV for EAs. Especially \( G_0 W_0 \Gamma^{(1)}_0 @ PBE \) performs much better than \( G_0 W_0 \) PBE. For EAs, the MAD improve from 0.60 eV to 0.06 meV, however, with 0.28 eV for IPs, \( G_0 W_0 \Gamma^{(1)}_0 @ PBE \) is not very accurate. For EAs, with a MAD of 0.06 eV, \( G_0 W_0 @ PBE + SOSEX \) performs excellent, but it is considerably less accurate for IPs (MAD = 0.33 eV). In summary, \( G_0 W_0 \Gamma^{(1)}_0 @ PBE 0 \) is the most accurate of these methods. However, it can no beat qgSW and also not \( G_0 W_0 @ ω B-97-X + G3W2 \) and \( G_0 W_0 @ LR Cω PBEh + G3W2. \)

For the sake of a direct comparison of dynamically and statically screened G3W2 corrections, we also calculated \( G_0 W_0 @ P B E + G3W2 \) or \( G_0 W_0 @ PBE 0 + G3W2 \) with our implementation. First, we note the reasonable agreement of the results obtained with our implementation and FHI-AIMS on the GW level, which allows for a qualitative comparison. The differences mainly result from the different basis sets used in the calculations. The MADs in table [III] clearly show, that the statically screened G3W2 correction does not give good results for these starting points. For EAs, the performance of the statically screened correction is comparable to the dynamical one. However, the description of IPs is even worsened. Here, the methods with dynamical screening are significantly better.

In ref. [73] it was found that the magnitude of the G3W2 correction was much smaller, when the statically screened interaction instead of the dynamically screened one was used. This is in line with our results. For the HOMO QP energies, the correction often changes sign, i.e. the HOMO QP energy is increased when the interaction is statically screened. Also in this case, the correction for the LUMO level is typically much larger than the one for the HOMO level, increasing the fundamental gaps. Due to the small magnitude of the correction, the statically screened G3W2 correction works well if QP energies are already well described on the GW level. The dynamically screened G3W2 term leads to a correction of larger magnitude and works best for GW methods which severely underestimate IPs and overestimate EAs. This is the case for \( W_0 @ PBE \) and (to a smaller extent) also for \( W_0 @ PBE 0 \) [93] and consequently, the addition of SOSEX or dynamically screened G3W2 leads to large improvements. On the other hand, \( W_0 @ HF \) underestimates the HOMO QP levels and overestimates LUMO QP levels. Therefore, the addition of dynamically screened SOX deteriorates the results for this starting point.

### 2. Ionization Potentials of Small Molecules

So far, the performance of the G3W2 self-energy correction has only been assessed for a very specific type of molecules. We now also consider a second database curated by Bartlett and coworkers [159]. They calculated the IPs of 40 small organic and inorganic molecules, using equation of motion (EOM) CCSD (EOM-CCSD)/cc-pVTZ & cc-pVQZ and EOM-CCSDT/cc-pVTZ. The reference values we use here are obtained as follows: From the EOM-CCSD results obtained with the cc-pVTZ and
FIG. 4. Ionization potentials (left) and electron affinities (right) from $G_0W_0$@ωB97-X (+ $\Sigma_{G3W2}$) (top) and $G_0W_0$@LRCω-PBEh (+ $\Sigma_{G3W2}$) (bottom) and $G_0W_0$@ωB97-X + $\Sigma_{G3W2}$, for a dataset of 24 organic acceptor molecules. The dashed black diagonal lines are CCSD(T) reference values and the values in parantheses denote mean absolute deviations (MAD) with respect to CCSD(T). All values are in eV.

The starting points are $G_0W_0$, $G_0W_0$@ωB97-X + SOSEX, and $G_0W_0$@ωB97-X + $G_3W2$ for different starting points. All values are in eV. The subscript (s) denote that the screened interaction is static, whereas (d) denotes a static interaction. For $G_0W_0$, the two numbers denote MAD with ADF/FHI-AIMS.

Table III. Comparison of MADs with respect to the CC reference values for different beyond-$GW$ method in comparison to $G_0W_0$ for different starting points. All values are in eV. The subscript (d) denote that the screened interaction is dynamic, whereas (s) denotes a static interaction. For $G_0W_0$, the two numbers denote MAD with ADF/FHI-AIMS.

| Starting Point | $G_0W_0$ | $G_0W_0$@ωB97-X + SOSEX | $G_0W_0$@ωB97-X + $G_3W2^{(d)}$ | $G_0W_0$@ωB97-X + $G_3W2^{(s)}$ |
|----------------|---------|--------------------------|-------------------------------|-------------------------------|
|                | IP      | IP                       | IP                            | IP                            |
| PBE            | 0.56/0.66 | 0.44/0.60 | 0.08                         | 0.08                          |
| PBE0           | 0.19/0.22 | 0.39/0.39 | 0.16                         | 0.09                          |

cc-pVQZ basis sets, we extrapolate the IPs to the CBS limit with the formula by Helgaker et al.\textsuperscript{121}, eq. \textsuperscript{23}. Note, that we used eq.\textsuperscript{24} to obtain our results. Subsequently, we add the difference between the EOM-CCSD and EOM-CCSDT IPs to the CBS limit extrapolated EOM-CCSD IPs. Thus, the reference values should be close to EOM-CCSDT quality at the CBS limit. The MADs of the considered methods with respect to the
TABLE IV. MADs with respect to the EOM-CCSDT reference for the first ionization potentials of a set of 40 small molecules. All values are in eV.

| Starting Point | GW | GW + G3W2 |
|----------------|----|-----------|
| qG1W          | 0.21 | 0.27 |
| wB97-X        | 0.20 | 0.27 |
| LRC-ωPBeH     | 0.12 | 0.21 |

TABLE V. Computational Timings (in core hours) for two tested molecules. All values are in eV.

EOM-CCSDT reference values are shown in table [LV]. The qG1W IPs are with a MAD of 0.21 eV still in reasonable agreement with the reference values, but the agreement is worse than for the acceptor molecules. For all tested GW starting points, the G3W2 term worsens the IPs.

D. Timings

|                     | Azulene | Borrelidin |
|---------------------|---------|------------|
| QZ6P                | 7.3     | 7.3        |
| TZ3P                | 17      | 17         |
| QZ6P                | 74      | 74         |
| \(N_{\text{atoms}}\) | 18      | 78         |
| \(N_{\text{freq}}\) | 18      | 18         |
| \(N_{\text{AO}}\)  | 782     | 1687       |
| \(N_{\text{aux}}\) | 2296    | 9612       |
| G0W0 calc. [Core h]| 0.015   | 0.40       |

V. CONCLUSION

In the GWA, the electron-electron self-energy in Dyson’s equation is expanded in terms of a screened Coulomb interaction and truncated after first order. In this work, we have analysed different aspects of the statically screened second-order contribution to the self-energy (G3W2) and applied it in a perturbative fashion to calculate correlation energies of atoms, relative energies of chemical reactions, and IPs and EAs of a wide range of molecules. The results we have presented herein shed some light on many interesting aspects of the second-order correction and also raise many questions which we hope to be able to address in future work.

For correlation energies, the second-order correction (SOSSX) reduces the RPA@EXX correlation energies by around 25% and brings them into much better agreement with nearly exact reference values, but the correlation energies are still too large. Generally, relative energies are also improved over RPA, but more detailed investigations are needed to obtain a better understanding in what situations RPA+SOSSX yields improvements over the RPA.

It would also be interesting to investigate the relation between the RSOX and SOSSX terms further. SOSSX and RSOX are two different strategies to renormalizes the SOX term. RPA+RSOX is based on the resummation of the Epstein-Nesbet series of hole-hole ladder diagrams while SOSSX results from the resummation of ring diagrams, with the additional approximation of static screening. Both classes of diagrams are contained in CCD which also includes particle-particle ladder diagrams and a third class of diagrams, coupling ladder and ring diagrams. Thus, CCD can be seen as a first-principle method to couple both renormalization strategies. Unfortunately, CCD scales as \(N^6\), as opposed to \(N^5\). Since RPA+SOSSX overestimates, and RSOX underestimates the correlation energies, it seems worthwhile to look for alternative ways to combine SOSSX with RSOX while retaining the \(N^5\) scaling.

For charged excitations, our work reveals qualitative differences between the dynamically screened G3W2 correction as recently tested in ref. and the statically screened one, benchmarked herein. The magnitude of the G3W2 correction becomes much smaller when the electron-electron interaction is statically screened. The addition of the dynamically screened SOX term consistently lowers EAs and increases IPs, while the statically screened one often decreases the IPs.

Popular methods like \(G_0W_0\) based on PBE and PBE0 starting points predict too low ionization potentials and too high electron affinities and the addition of dynamically screened G3W2 correction results in major im-

TABLE V. Computational Timings (in core hours) for two selected molecules for different basis sets. \(N_{\text{AO}}\) and \(N_{\text{aux}}\) denote the sizes of primary and auxiliary basis, respectively.

*Timing per state
provements for these methods. The statically screened G3W2 correction, on the other hand, gives improvements over GW calculations which consistently underestimate QP energies. This is the case for GW₀LRCωPBEh and GW₀A97-X. Especially for EAs, improvements are tremendous. For IPs and fundamental gaps, the improvements are not consistent and seem to be system-specific. For molecules with nearly 100 atoms, evaluating the perturbative G3W2 correction for a few states does not come with significantly increased computational costs compared to a GW calculation. This is a big advantage over the dynamically screened SOX correction.

Also for qsGW, the statically screened G3W2 correction does not lead to systematic improvements, even though both corrections improve fundamental gaps. As a result, charge screening can be expected to be stronger in larger molecules and it would be interesting to see whether qsGW also performs well for these or if beyond-RPA screening is then necessary.

Even though the SOX correction to the self-energy leads to improvements over GW in certain situations, our results confirm recent studies which show that QP approximations to GW are difficult to improve upon diagrammatically, at least for molecular systems. Systematic and reliable improvements over GW are most likely only possible starting from the fully self-consistent solution of the GW equations.

**ACKNOWLEDGMENTS**

This research received funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the Innovation Fund for Chemistry and from the Ministry of Economic Affairs in the framework of the TKI/PPS-Toeslagregeling.
154. C. Bhandari, W. Lambrecht, and M. van Schilfgaarde, Phys. Rev. B - Condens. Matter Mater. Phys. 91, 125116 (2015).

155. C. Bhandari and W. Lambrecht, Phys. Rev. B - Condens. Matter Mater. Phys. 89, 045109 (2014).

156. A. Molina-Sánchez and L. Wirtz, Phys. Rev. B - Condens. Matter Mater. Phys. 84, 155413 (2011).

157. J. van der Horst, P. A. Bobbert, P. de Jong, M. Michels, G. Brocks, and P. Kelly, Phys. Rev. B - Condens. Matter Mater. Phys. 61, 15817 (2000).

158. H. J. De Groot, P. A. Bobbert, and W. van Haeringen, Phys. Rev. B 52, 11000 (1995).

159. D. S. Ransinghe, J. T. Margraf, A. Perera, and R. J. Bartlett, J. Chem. Phys. 150, 074108 (2019).

160. Q. Ma and H. J. Werner, J. Chem. Theory Comput. 16, 3135 (2020).

161. N. S. Szabo, Attila, and Ostlund, Modern quantum chemistry: introduction to advanced electronic structure theory (Courier Corporation, 2012).

162. F. Bruneval, N. Dattani, and M. J. van Setten, Front. Chem. 9, 749779 (2021).