More efficient dielectric matrix calculations using the Lanczos algorithm for faster many-body $\mathcal{G}_0\mathcal{W}_0$ implementations

Jonathan Laflamme Janssen, Bruno Rousseau, and Michel Côté

Département de physique et Regroupement Québécois sur les Matériaux de Pointe (RQMP), Université de Montréal, C. P. 6128 Succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada

(Dated: August 14, 2014)

We present a $\mathcal{G}_0\mathcal{W}_0$ implementation that assesses the two major bottlenecks of traditional plane waves implementations, the summations over conduction states and the inversion of the dielectric matrix, without introducing new approximations in the formalism. The first bottleneck is circumvented by converting the summations into Sternheimer equations. Then, the novel avenue of expressing the dielectric matrix in a Lanczos basis is developed, which reduces the matrix size by orders of magnitude while being computationally efficient. We also develop a model dielectric operator, which allows to further reduce the size of the dielectric matrix without accuracy loss. Furthermore, we develop a scheme that reduces the numerical cost of the contour deformation technique to the level of the lightest plasmon pole model. Finally, the use of the Simplified Quasi-Minimal Residual scheme in replacement of the conjugate gradients algorithm allows a direct evaluation of the $\mathcal{G}_0\mathcal{W}_0$ corrections at the desired real frequencies, without need for analytical continuation. The performance of the resulting $\mathcal{G}_0\mathcal{W}_0$ implementation is demonstrated by comparison with a traditional plane wave implementation, which reveals a 500-fold speedup. Finally, the accuracy our $\mathcal{G}_0\mathcal{W}_0$ implementation is demonstrated by comparison with other $\mathcal{G}_0\mathcal{W}_0$ calculations and experimental results.

PACS numbers: 71.15.Qe, 31.15.ag, 33.60.+q

I. INTRODUCTION

Density functional theory (DFT)\textsuperscript{1–3} is currently the most popular approach for electronic structure simulation of periodic materials, molecules and nanostructures. However, its predictive power is formally limited to ground state properties. Consequently, while in practice DFT is widely used to calculate band structures, the precision of the results is limited. A formally sound and more precise\textsuperscript{4,5} method is provided by the GW framework\textsuperscript{5,7}. However, these calculations are computationally more demanding than their DFT counterparts. Indeed, GW calculations are typically limited to a few tens of atoms while DFT codes can handle a few hundreds. In a conventional plane wave implementation\textsuperscript{8}, two bottlenecks account for this limitation: the sums to be carried out over all conduction states\textsuperscript{8–11} and the inversion of the dielectric matrix\textsuperscript{12–14} that describes the screening of an external potential by the simulated system. Many different approaches have been explored to assess the summation over conduction states: it can be converted into a linear equation problem\textsuperscript{12–14}, the conduction states can be replaced by simple approximate orbitals\textsuperscript{15}, a so-called extrapolar trick can be used to reduce the number of conduction states required for a given level of convergence\textsuperscript{16}, the summations can be eliminated using the effective-energy technique\textsuperscript{17} or the size of the basis in which the Hamiltonian is expressed can be reduced through the use of localized basis sets\textsuperscript{18}. Substantial attention has also been devoted to the assessment of the inversion of the dielectric matrix: it can be avoided by reformulating the problem into a self-consistent Sternheimer equation\textsuperscript{19} or the size of the basis in which the dielectric matrix is expressed can be reduced either using Wannier orbitals\textsuperscript{19} or eigenvectors of the static dielectric matrix\textsuperscript{14}.

In this article, we present a novel implementation of the $\mathcal{G}_0\mathcal{W}_0$ method that circumvents both bottlenecks. This implementation is designed to be suitable for periodic systems as well as systems where the charge density may not be centered on the atoms, such as heterojunctions in organic photovoltaic materials. It is coded within the ABINIT project\textsuperscript{20} and uses a plane wave basis framework. To assess the summations over conduction states, we adopt the strategy to convert them into linear equation problems, since it is suitable for our choice of basis, efficient and well established\textsuperscript{12–14}. To assess the bottleneck of the inversion of the dielectric matrix, we elaborate a novel approach where the matrix is expressed in a Lanczos basis\textsuperscript{21}. This reduces the size of the matrix as effectively as the traditional spectral decomposition method\textsuperscript{14,22} while being computationally an order of magnitude more efficient. We also develop a model dielectric operator, which allows to further reduce the size of the dielectric matrix.

Furthermore, in the present $\mathcal{G}_0\mathcal{W}_0$ implementation, the use of the contour deformation technique\textsuperscript{23,24} was preferred over plasmon pole models\textsuperscript{8,23,25–27} to avoid considerations on the range of systems which can be accurately described\textsuperscript{28,29}. Traditionally, this choice implies a greater computational cost. In the present method, we explore two different directions to reduce this cost to a level of the simplest plasmon pole model\textsuperscript{8}. First, we use a Lorentzian to model the frequency dependence of the dielectric matrix and only treat the difference between this model and the exact dielectric matrix with the contour deformation technique, which alleviates the computational work required by the numerical integration. This
idea is inspired by previous work involving a Gaussian model \(^{30}\), with the distinction of being compatible with the conversion of the summations over conduction states into linear equation problems and allowing a direct theoretical analogy with the plasmon pole technique. Also, we elaborate a scheme to recycle the information computed in the construction of the static dielectric matrix and obtain the dynamical dielectric matrix at all relevant non-zero frequencies at a negligible computational cost, thus effectively reducing the numerical cost of the contour deformation technique to the level of a simple plasmon pole model.

The evaluation of the \( G_0 W_0 \) corrections at the desired real frequency is usually unstable, due to the presence of poles on the real axis in the inverse dielectric matrix. This difficulty is traditionally avoided by analytic continuation of the self-energy from the imaginary axis to the real axis\(^{14,31}\). In the present implementation, we solve this problem using a recently developed numerical method, the Simplified Quasi-Minimal Residual (SQMR) algorithm\(^{32}\). It is as efficient as the traditional conjugate gradients method, but stable when indefinite or nearly singular linear equations are solved, such as those involved in the calculation of the dielectric matrix at real frequencies. To our knowledge, it is the first time that such an algorithm is used in an electronic structure implementation.

This article is organized as follows. First, a summary of the \( G_0 W_0 \) method is given in section II. Then, the bottleneck of the sums over conduction states is assessed in section III. In particular, the Lorentzian model is developed in subsection III B. Next, the bottleneck of the dielectric matrix inversion is assessed in section IV. Then, the model dielectric operator is developed in section V. A strategy to use the information generated in the construction of the static dielectric matrix to accelerate its computation at imaginary frequencies is devised in section VI. A theoretical analysis of the numerical cost of the present \( G_0 W_0 \) implementation is given in section VII. Then, we assess the accuracy of our implementation by comparing our results with previously published ones in section VIII. Finally, we assess its performance with respect to traditional implementations in section IX. Atomic units are used throughout unless otherwise specified.

**II. THE \( G_0 W_0 \) METHOD**

In conventional implementations of the \( G_0 W_0 \) method, corrections to DFT eigenenergies are obtained using first order perturbation theory, where the perturbation is the difference between the \( G_0 W_0 \) self-energy \( \Sigma(\omega) \) and the DFT exchange-correlation potential \( V^{xc} \):

\[
\Delta \varepsilon_e = \langle e | \Sigma(\varepsilon_e + \Delta \varepsilon_e) - V^{xc} | e \rangle, \tag{1}
\]

where \( \varepsilon_e \) is a DFT eigenenergy, \( | e \rangle \) is the associated eigenstate and \( \Delta \varepsilon_e \) is the \( G_0 W_0 \) correction to \( \varepsilon_e \). The self-consistency with respect to \( \Delta \varepsilon_e \) is easily avoided by making a Taylor expansion of Eq. (1) to first order with respect to \( \Delta \varepsilon_e \) around zero\(^{33}\). Also, the DFT exchange-correlation energy \( \langle e | V^{xc} | e \rangle \) can easily be extracted from the DFT calculation. The only non-trivial part of the calculation is therefore the evaluation of the \( G_0 W_0 \) exchange-correlation energy \( \langle e | \Sigma(\omega) | e \rangle \), where \( \Sigma(\omega) \) is defined as

\[
\langle r | \Sigma(\omega) | r' \rangle = \frac{i}{2\pi} \int_{-\infty}^{\infty} d\omega' e^{i\eta \omega'} \langle r | G_0(\omega' + \omega) | r' \rangle \langle r | \hat{W}_0(\omega') | r' \rangle, \tag{2}
\]

where \( G_0(\omega) \) is the Green’s function, \( \hat{W}_0(\omega) \) is the screened Coulomb potential, \(| r \rangle \) is an eigenfunction of the position operator and \( \omega \) is the angular frequency (eventually \( \in \mathbb{C} \)).

In the following, we will restrict ourselves to the non-spin-polarized and non-periodic (molecular) case for simplicity. The Green’s function is easily expressed in terms of DFT eigenstates \(| n \rangle \) and eigenvalues \( \varepsilon_n \).

\[
\hat{G}_0(\omega) = \sum_{n=1}^{\infty} \frac{| n \rangle \langle n |}{\omega - \varepsilon_n - i\eta \text{sgn}(\varepsilon_n - \mu)}, \tag{3}
\]

where \( \eta \rightarrow 0^+ \) is a positive infinitesimal and \( \mu \) is the chemical potential. The screened Coulomb potential is related to the bare Coulomb potential \( \hat{v}_C \) by the inverse dielectric matrix \( \varepsilon^{-1}(\omega) \),

\[
\hat{W}_0(\omega) = \hat{v}_C^{1/2} \varepsilon^{-1}(\omega) \hat{v}_C^{1/2}, \tag{4}
\]

where \( \hat{v}_C^{1/2} \) is the square root of the Coulomb potential. We choose to work here with the symmetric form of the dielectric matrix, because of its computational advantages\(^{34}\). These definitions allow us to express the \( G_0 W_0 \) exchange-correlation energy as

\[
\Sigma_e(\delta) = \langle e | \hat{\Sigma}(\varepsilon_e + \delta) | e \rangle = \frac{i}{2\pi} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} d\omega e^{i\eta \omega} \frac{\langle en^* \hat{v}_C^{1/2} \varepsilon^{-1}(\omega) \hat{v}_C^{1/2} | n^* e \rangle}{\omega - \varepsilon_n + i\eta \text{sgn}(\varepsilon_n - \mu)}, \tag{5}
\]

where \( \omega_n e \equiv \varepsilon_n - \varepsilon_e - \delta \), \( \langle r | n^* e \rangle \equiv \langle r | n^* \rangle \) and \( \langle r | n^* e \rangle \equiv \langle r | n^* \rangle \). It is customary at this stage to split the self-energy matrix element into an exchange \( \Sigma_e^x(\delta) \) and a correlation \( \Sigma_e^c(\delta) \) part,

\[
\Sigma_e^x(\delta) = \langle e | \hat{\Sigma}^x(\varepsilon_e + \delta) | e \rangle = \frac{i}{2\pi} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} d\omega e^{i\eta \omega} \frac{\langle en^* \hat{v}_C^{1/2} (\varepsilon^{-1}(\omega) - 1) \hat{v}_C^{1/2} | n^* e \rangle}{\omega - \varepsilon_n + i\eta \text{sgn}(\varepsilon_n - \mu)}, \tag{6}
\]

and

\[
\Sigma_e^c(\delta) = \langle e | \hat{\Sigma}^c(\varepsilon_e + \delta) | e \rangle = \frac{i}{2\pi} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} d\omega e^{i\eta \omega} \frac{\langle en^* \hat{v}_C^{1/2} \varepsilon^{-1}(\omega) \hat{v}_C^{1/2} | n^* e \rangle}{\omega - \varepsilon_n + i\eta \text{sgn}(\varepsilon_n - \mu)}. \tag{7}
\]
The integral in the exchange part can be evaluated by closing the integration contour with a half-circle of infinite radius in the upper complex plane. The factor $e^{i\eta\omega}$ reduces the integral over this half-circle to 0 and the residues of the poles included in the contour become the only contributions to the exchange term $\Sigma^e$. Since the presence of a pole above or below the real axis is determined by the presence of its eigenenergy below or above the chemical potential $\mu$ (see Eq. (6)), the poles included in the contour are those associated to the valence states. The exchange term thus takes the familiar form

$$
\Sigma^e = \langle e | \hat{\Sigma}^e | e \rangle = -\sum_n \langle ev^* | \hat{v}_C | v^* e \rangle,
$$

where the index $v$ labels the valence states.

A popular way to calculate the integration over $\omega$ in $\Sigma^e(\delta)$ is the plasmon-pole approximation, which reduces the associated computational cost by only requiring the explicit calculation of the dielectric matrix at one or two frequencies. However, the range of systems where this approximation is robust is subject to some debate.

To preserve the precision and the wide applicability (in terms of physical systems) of the plane wave basis set, we instead carry out the integration numerically. However, the high number of poles along the real axis makes it unwieldy to integrate numerically along this direction. Therefore, we use the residue theorem to reformulate the problem into an integration along the imaginary axis, using the contour illustrated in Fig. 1.

The correlation part then becomes

$$
\Sigma^c(\delta) = -\frac{i}{2\pi} \sum_{n=1}^{\infty} \left( \int_{C_1} + \int_{-\infty}^{\infty} + \int_{C_3} \right) d\omega \langle en^* | \hat{v}^{1/2}_C (\epsilon^{-1}(\omega) - \hat{1}) \hat{v}^{1/2}_C | n^* e \rangle \frac{\omega_n e}{\omega^2 + \omega^2_n} + \Sigma^p(\delta),
$$

where we took the limit $\eta \to 0^+$ after selecting the poles lying inside the contour and where the index $c$ labels the conduction states. It can be shown that $\lim_{|\omega| \to \infty} (\epsilon^{-1}(\omega) - \hat{1}) \to 0$ as $1/\omega^2$. Consequently, the integrals over both quarters of circle ($\int_{C_1}$ and $\int_{C_3}$) vanish. By substituting $\omega \to i\omega$, the domain of integration of the second term of Eq. (9) can be made real,

$$
\Sigma^c(\delta) = \frac{1}{\pi} \sum_{n=1}^{\infty} \int_0^{\infty} d\omega \langle en^* | \hat{v}^{1/2}_C (\epsilon^{-1}(i\omega) - \hat{1}) \hat{v}^{1/2}_C | n^* e \rangle \frac{\omega_n e}{\omega^2 + \omega^2_n} + \Sigma^p(\delta),
$$

where we have defined the residue term

$$
\Sigma^p(\delta) \equiv -\sum_v \langle ev^* | \hat{v}^{1/2}_C (\epsilon^{-1}(\omega_{ve}) - \hat{1}) \hat{v}^{1/2}_C | v^* e \rangle \Theta(\omega_{ve}) + \sum_v \langle ev^* | \hat{v}^{1/2}_C (\epsilon^{-1}(\omega_{ce}) - \hat{1}) \hat{v}^{1/2}_C | v^* e \rangle \Theta(-\omega_{ce}).
$$

The resulting expression is

$$
\Sigma^c(\delta) = \frac{1}{\pi} \int_0^{\infty} d\omega \left( \sigma^e_N(i\omega, \delta) - \sigma^e_{N0}(i\omega, \delta) \right) + \Sigma^p(\delta) + \Sigma^c(\delta),
$$

where

$$
\sigma^e_N(i\omega, \delta) \equiv \sum_n \langle en^* | \hat{v}^{1/2}_C (\epsilon^{-1}(i\omega) - \hat{1}) \hat{v}^{1/2}_C | n^* e \rangle \times \frac{\omega_n e}{\omega^2 + \omega^2_n},
$$

$$
\sigma^e_{N0}(i\omega, \delta) \equiv \sum_n \langle en^* | \hat{v}^{1/2}_C (\epsilon^{-1}(0) - \hat{1}) \hat{v}^{1/2}_C | n^* e \rangle \times f(\omega) \frac{\omega_n e}{\omega^2 + \omega^2_n},
$$

where $f(\omega)$ is a function that models the frequency dependence of the dielectric matrix, which makes the integrand smaller and thus easier to sample numerically.
 involve calculating a sufficient number of conduction states to converge the summations in Eqs. (13), (14), (15) and (18) as well as inverting $\epsilon(\omega)$ in a plane waves basis.

### III. AVOIDING SUMMATIONS OVER CONDUCTION STATES

The bottleneck of the summation over conduction states can be avoided without introducing further approximations to the preceding scheme at the expense of introducing a linear equation problem to be solved iteratively\textsuperscript{12–14}. This strategy is commonly used in density functional perturbation theory\textsuperscript{39,40}, where this type of linear equation is frequently referred to as a Sternheimer equation\textsuperscript{41}. First, we will apply this idea to the polarizability. Then, we will eliminate the summation over states present in Eqs. (13), (14) and (15) as well.

#### A. The polarizability

We start from the action of the polarizability on some vector $|\psi_j\rangle$, labeled by the index $j$,

$$\hat{P}(\omega)|\psi_j\rangle = -2 \sum_v |v\rangle \left( \sum_c |c\rangle \frac{1}{\varepsilon_c - \varepsilon_v - \omega} \langle vc^*|\psi_j\rangle + \sum_c |c\rangle \frac{1}{\varepsilon_c - \varepsilon_v + \omega} \langle vc^*|\psi_j\rangle \right),$$  

where $|v\rangle |f_{jv\pm}(\omega)\rangle = |vf_{jv\pm}(\omega)\rangle$ and where we have introduced the new vector

$$|f_{jv\pm}(\omega)\rangle = \sum_c |c\rangle \frac{1}{\varepsilon_c - \varepsilon_v \pm \omega} \langle c|\psi_j^*\rangle.$$  

The idea is simply to use the completeness rule $\sum_c |c\rangle \langle c| = 1 - \sum_v |v\rangle \langle v| \equiv \hat{P}_c$, which can readily be done if $\varepsilon_c$ is substituted by $\hat{H}$ in the denominator of Eq. (20). We thus obtain

$$|f_{jv\pm}(\omega)\rangle = \frac{\hat{P}_c}{\hat{H} - \varepsilon_v \pm \omega} |v\psi_j^*\rangle.$$  

However, inverting the Hamiltonian is a problem similar in size to its full diagonalization, e.g. the calculation of all its eigenvalues and eigenvectors, which is exactly what we set out to avoid in the first place. A numerically less expensive alternative is to turn the problem into a linear equation,

$$\left(\hat{H} - \varepsilon_v \pm \omega\right) |f_{jv\pm}(\omega)\rangle = \hat{P}_c |v\psi_j^*\rangle.$$  

This process can be done if $\varepsilon_c$ is substituted by $\hat{H}$ in the denominator of the linear equation. Once we have the correct vectors $|f_{jv\pm}(\omega)\rangle$, we can calculate the polarizability by using the completeness rule.

$$\epsilon(\omega) = 1 - \hat{P}(\omega) \hat{\epsilon}^{-1}(\omega).$$  

Using the completeness rule, the polarizability can be calculated as

$$\epsilon(\omega) = 1 - \hat{P}(\omega) \hat{\epsilon}^{-1}(\omega).$$  

The preceding equations suffice to describe conventional $G_0W_0$ calculations in a plane wave basis set, which
This equation becomes problematic to solve when \( \omega \to 0 \), since the left-hand side operator becomes singular. The customary way to avoid this instability is to add a term \( \beta P \) to the Hamiltonian \( \hat{H} \), where \( \beta \) is larger than the valence bandwidth. Thus, when \( \omega \to 0 \), the operator acting on \( |f_{j+i}(\omega)\rangle \) does not become singular. Since the right-hand side of Eq. (22) is outside the valence subspace, this algebraic trick leaves the final answer unaffected. However, when the argument of the dielectric operator \( \omega \) is real, as it is the case in Eq. (11), this trick does not prevent the left-hand side from being singular. Indeed, \( \beta \) could be chosen so that \( \omega \), for some \( \varepsilon \) (or \( \omega_{\text{ce}} \)), for some \( \varepsilon \) (or \( \omega_{\text{ce}} \)), equals some valence-conduction transition \( \omega_{\text{vc}} \) (or \( \omega_{\text{vc}} \)). Then, the operator \( \hat{H} - \varepsilon_{v} - \omega_{\text{vc}} \) (or \( \hat{H} - \varepsilon_{v} + \omega_{\text{vc}} \)) would become singular in some part of the conduction subspace. Since the right-hand side of Eq. (22) can have non-zero components in this subspace, the previous trick cannot be applied to eliminate the singularity. Indeed, there will be an associated subspace \( C \) where \( \hat{1} \) (or \( \hat{1} \)) will be infinite and where \( 1 - \hat{1} \) will be 1, so that the kernel of \( \hat{H} - \varepsilon_{v} \pm \omega \) will contribute to \( \Sigma_{c}^{\text{p}}(\delta) \). However, it is still possible to stabilise Eq. (22) without altering this physical contribution by using the Simplified Quasi-Minimal Residual (SQMR) algorithm instead of conjugate gradients. Indeed, the former is stable for indefinite matrices close to singularity \((\langle n | \hat{H} - \varepsilon_{v} | n \rangle) \approx 10^{-3} \) Ha \( \forall |n\rangle \) while the latter is stable only for positive definite matrices. It thus becomes easy to choose a suitable value of \( \beta \) such that Eq. (22) is stable, without adding substantially to the computation time. Still, for the important case \( \omega = 0 \), the operator will be singular in the valence subspace up to machine precision. Therefore, the addition of \( \beta P \) to \( \hat{H} - \varepsilon_{v} \pm \omega \) remains necessary. In our implementation, we adopted the equivalent strategy to orthogonalize the solution vector with respect to the valence subspace at each SQMR step. This SQMR-based scheme has the advantage of allowing the direct calculation of \( \Sigma_{c}^{\text{p}}(\delta) \) at the (real) desired frequency, which eliminates the need for analytic continuation and related stability considerations.

We note that for imaginary frequencies \( i \omega \), \( \hat{H} - \varepsilon_{v} \pm i \omega \) is not hermitian, which prevents the use of SQMR or conjugate gradients. However, in this case, we can instead solve the hermitian linear equation

\[
\left( (\hat{H} - \varepsilon_{v})^2 - (i \omega)^2 \right) |f_{j+i}(\omega)\rangle = 2(\hat{H} - \varepsilon_{v})\hat{P}_{c} |\psi_{j}^{\text{c}}\rangle,
\]

where the solution obtained is the sum of the vectors defined in Eq. (21),

\[
|f_{j+i}(\omega)\rangle \equiv |f_{j+i}(\omega)\rangle + |f_{j-i}(\omega)\rangle.
\]

Typically, solving this equation to a precision where the residual \( \left| \left( (\hat{H} - \varepsilon_{v})^2 - (i \omega)^2 \right) |f_{j+i}(\omega)\rangle - 2(\hat{H} - \varepsilon_{v})\hat{P}_{c} |\psi_{j}^{\text{c}}\rangle \right| \) is less than \( 10^{-10} \) Ha² converges the calculated \( \Sigma_{c}^{\text{p}}(\delta) \) to 5 meV.

**B. The Self-Energy**

Similarly to the case of the polarizability, we use the completeness rule \( \sum_{n} |n\rangle \langle n| = 1 \) to eliminate the summation over all states in \( \Sigma_{c}^{\text{p}}(\delta) \). However, it is now necessary to introduce an intermediate basis that we will refer to as \( \{|l\rangle\} \). We only require that this basis be complete, up to a convergence criteria on \( \Sigma_{c}^{n}(\delta) \). For \( \sigma_{c}^{n}(i \omega, \delta) \), we obtain from Eq. (13)

\[
\sigma_{c}^{n}(i \omega, \delta) = \sum_{n,l,l'} (e_{n} | v_{C}^{1/2} | l \langle l | \hat{\epsilon}^{-1}(i \omega) - \hat{1} | l' \rangle \langle l' | v_{C}^{1/2} | n^{*} \rangle c_{l} \times \frac{\omega_{ne}}{\omega^{2} + \omega_{ne}^{2}} \langle n^{*} | v_{C}^{1/2} | l^{*} \rangle.
\]

Let \( \Phi_{e} \) be an operator so that

\[
\langle \phi | \Phi_{e} | r \rangle = \phi^{*}(r) \phi(r)
\]

Then, \( \sigma_{c}^{n}(i \omega, \delta) \) becomes

\[
\sigma_{c}^{n}(i \omega, \delta) = \sum_{n,l,l'} \langle l | \hat{\epsilon}^{-1}(i \omega) - \hat{1} | l' \rangle \times \langle l' | v_{C}^{1/2} \Phi_{e}^{*} | n \rangle \frac{\omega_{ne}}{\omega^{2} + \omega_{ne}^{2}} \langle n | v_{C}^{1/2} | l^{*} \rangle.
\]

Similarly to the case of the polarizability, we need to replace the eigenvalues \( \varepsilon_{n} \) by the Hamiltonian \( \hat{H} \) to use the completeness relation,

\[
\sigma_{c}^{n}(i \omega, \delta) = \sum_{l,l'} \langle l | \hat{\epsilon}^{-1}(i \omega) - \hat{1} | l' \rangle \times \langle l' | v_{C}^{1/2} \Phi_{e}^{*} \hat{H} - \varepsilon_{v} - \delta \frac{\omega_{ne}}{\omega^{2} + (\hat{H} - \varepsilon_{v} - \delta)^{2}} \Phi_{e} v_{C}^{1/2} | l^{*} \rangle.
\]

To avoid inverting the Hamiltonian, we need to transform the problem into a linear equation. We can do this easily by defining the vector \( |\sigma_{c}^{n,l}(i \omega, \delta)\rangle \),

\[
|\sigma_{c}^{n,l}(i \omega, \delta)\rangle = \sum_{l,l'} \langle l | \hat{\epsilon}^{-1}(i \omega) - \hat{1} | l' \rangle \times \langle l' | v_{C}^{1/2} \Phi_{e}^{*} | \sigma_{c}^{n,l}(i \omega, \delta) \rangle,
\]

which is given by the following linear equation,

\[
(\omega^{2} + (\hat{H} - \varepsilon_{v} - \delta)^{2}) |\sigma_{c}^{n,l}(i \omega, \delta)\rangle = (\hat{H} - \varepsilon_{v} - \delta) \Phi_{e} v_{C}^{1/2} | l^{*} \rangle.
\]
Similarly, for $\sigma_c^{\text{NO}}(i\omega, \delta)$, we obtain

$$
\sigma_c^{\text{NO}}(i\omega, \delta) = \sum_{l,l'} \langle l | (\hat{e}^{-1}(0) - \hat{1}) | l' \rangle f(\omega) 
\times \langle l' | \hat{v}_c^{1/2} \Phi_c^{\dagger} [\sigma_c^{\text{NO}}(i\omega, \delta)] \rangle.
$$

(31)

The same strategy can also be applied to eliminate the summation over all states in $\Sigma_c^\Delta(\delta)$

$$
\Sigma_c^\Delta(\delta) = \sum_{n,l,l'} \langle l | (\hat{e}^{-1}(0) - \hat{1}) | l' \rangle
\times \langle l' | \hat{v}_c^{1/2} \Phi_c^{\dagger} | n \rangle F_{n,e}(\delta) \langle n | \Phi_c \hat{v}_c^{1/2} | l'' \rangle.
$$

To go further, we now need to know the explicit expression for $F_{n,e}(\delta)$ and, therefore, to choose the model function $f(\omega)$. The obvious minimalist case is to not attempt any modelization of the frequency dependence of $\hat{e}^{-1}(i\omega) - \hat{1}$. We then have

$$
f(\omega) = 1,
F_{n,e}(\delta) = \frac{1}{2} \left\langle \Theta(\omega_{ne}) - \Theta(-\omega_{ne}) \right\rangle,
$$

$$
\Sigma_c^\Delta(\delta) = \frac{1}{2} \sum_{l,l'} \langle l | (\hat{e}^{-1}(0) - \hat{1}) | l' \rangle
\times \langle l' | \hat{v}_c^{1/2} \Phi_c^{\dagger} (\hat{Q}_c(\delta) - \hat{P}_c(\delta)) \hat{P}_c \hat{v}_c^{1/2} | l'' \rangle,
$$

(32)

where

$$
\hat{Q}_c(\delta) = \sum_{\epsilon_n > \epsilon_c + \delta} | n \rangle \langle n |,
\hat{P}_c(\delta) = \sum_{\epsilon_n < \epsilon_c + \delta} | n \rangle \langle n |.
$$

(33)

We note the presence of discontinuities with respect to $\delta$ in $\Sigma_c^\Delta(\delta)$ and $\Sigma_e^\Delta(\delta)$ (see Eq. (11)), which must be treated carefully so that $\Sigma_e^\Delta(\delta)$ remains continuous.

For the highest occupied molecular orbital (HOMO) of silane, choosing $f(\omega) = 1$ allows to converge $\Sigma_c^\Delta(\epsilon) = \Sigma^\Delta - V^\text{occ} | \epsilon \rangle$ to 1 meV with only 8 frequency samplings of the integrand of Eq. (12) in contrast to 24 when $f(\omega) = 1$ is chosen to be 0.

The only non-trivial model function used in previous works was of Gaussian form. This choice yields

$$
f(\omega, \alpha) = e^{-\omega^2/\alpha^2},
F_{n,e}(\delta, \alpha) = \frac{1}{2} \text{sgn}(\omega_{ne}) e^{\omega_{ne}^2/\alpha^2} \text{erfc}(|\omega_{ne}|/\alpha),
$$

$$
\Sigma_c^\Delta(\delta, \alpha) = \frac{1}{2} \sum_{l,l'} \langle l | (\hat{e}^{-1}(0) - \hat{1}) | l' \rangle
\times \langle l' | \hat{v}_c^{1/2} \Phi_c^{\dagger} \text{sgn}(\hat{H} - \epsilon_c - \delta e^{(\hat{H} - \epsilon_c - \delta)^2/\alpha^2}) \text{erfc}(|(\hat{H} - \epsilon_c - \delta)/\alpha|) \hat{P}_c \hat{v}_c^{1/2} | l'' \rangle,
$$

where $\alpha$ is a model parameter that characterizes the width of the Gaussian. However, the Taylor expansions required to calculate explicit values for the functions of the Hamiltonian $\hat{H}$ make this choice numerically too cumbersome to be practical in $G_0W_0$ implementations where summations over conduction states are avoided, like the present method. Therefore, in this work, we explore a novel choice of model function, which is both compatible with the elimination of summations over conduction states and physically motivated. The chosen form for $f(\omega)$ is a Lorentzian,

$$
f(\omega, \alpha) = \frac{\alpha^2}{\omega^2 + \alpha^2}, \quad (\alpha > 0),
$$

$$
F_{n,e}(\delta, \alpha) = \frac{\pi}{\omega_{ne} + \alpha \text{sgn}(\omega_{ne})},
$$

$$
\Sigma_c^\Delta(\delta, \alpha) = \frac{1}{2} \sum_{l,l'} \langle l | (\hat{e}^{-1}(0) - \hat{1}) | l' \rangle
\times \langle l' | \hat{v}_c^{1/2} \Phi_c^{\dagger} (\hat{Q}_c(\delta) | h_{e,c}^+(\delta, \alpha) \rangle
+ \langle l' | \hat{v}_c^{1/2} \Phi_c^{\dagger} \hat{P}_c(\delta) | h_{e,c}^-(\delta, \alpha) \rangle \rangle,
$$

(34)

where

$$
(h_{e} - \epsilon_e - \delta + \alpha) h_{e,c}^+(\delta, \alpha) \equiv \alpha \hat{Q}_c(\delta) \hat{P}_c \hat{v}_c^{1/2} | l'' \rangle,
$$

$$
(h_{e} - \epsilon_e - \delta - \alpha) h_{e,c}^-(\delta, \alpha) \equiv \alpha \hat{P}_c(\delta) \hat{P}_c \hat{v}_c^{1/2} | l'' \rangle,
$$

and where no complicated functions of $\hat{H}$ are involved. Choosing $\alpha = 1.0$ Ha in this form of model function allows to converge $\Sigma_c^\Delta(\epsilon) = \Sigma^\Delta - V^\text{occ} | \epsilon \rangle$ for the highest occupied molecular orbital (HOMO) of silane to 1 meV with only 4 frequency samplings of the integrand of Eq. (12) in contrast to 8 when $f(\omega) = 1$.

Also, as it can be seen from the following expression,

$$
\hat{e}^{-1}(i\omega) - 1 \approx \frac{\alpha^2}{\omega^2 + \alpha^2} (\hat{e}^{-1}(0) - 1),
$$

$$
\approx \frac{\alpha}{2} \left( \frac{1}{\alpha + i\omega} + \frac{1}{\alpha - i\omega} \right) (\hat{e}^{-1}(0) - 1),
$$

(36)

approximating the dynamical character of the inverse dielectric matrix $\hat{e}^{-1}(i\omega) - 1$ by a Lorentzian is equivalent to replacing all its poles on the positive real axis by a single one. This choice of model function can therefore be physically interpreted as a scalar version of the plasmon pole model. It is interesting to note that such a model function has the correct high frequency behavior

$$
\lim_{\omega \to \infty} \frac{\alpha^2}{\omega^2 + \alpha^2} \propto \frac{1}{\omega^2} \propto \lim_{\omega \to \infty} \hat{e}^{-1}(i\omega) - 1
$$

while this is not the case for the Gaussian model. A generalization of this model to the level of conventional plasmon pole approximations is currently under way.

IV. AN EFFICIENT BASIS \{l\} FOR $\hat{e}^{-1}(i\omega)$

Eliminating the sums over conduction states that were present in Eqs. (12), (13), (14) and (15) required us to introduce a complete basis \{l\} in the terms $\sigma_c^\Delta(i\omega, \delta)$, $\sigma_c^{\text{NO}}(i\omega, \delta, \alpha)$ and $\Sigma_c^\Delta(\delta, \alpha)$. 
These terms can be rewritten in the form of a trace in the basis \( \{|l\} \), starting from Eqs. (29), (30), (31), (34) and (35):

\[
\sigma_e^N(i\omega, \delta) = \sum_l \langle l | (\hat{\epsilon}^{-1}(i\omega) - \hat{1}) \hat{v}^{1/2}_e \hat{\Phi}_e \frac{\hat{H}^* - \epsilon_e - \delta}{\omega^2 + (\hat{H}^* - \epsilon_e - \delta)^2} \hat{\Phi}_e^\dagger \hat{v}^{1/2}_e | l \rangle = \sum_l \langle l | \sigma_e^N(i\omega, \delta) | l \rangle \\
\sigma_e^{N0}(i\omega, \delta, \alpha) = \sum_l \langle l | (\hat{\epsilon}^{-1}(0) - \hat{1}) \hat{v}^{1/2}_e \hat{\Phi}_e \frac{\hat{H}^* - \epsilon_e - \delta}{\omega^2 + (\hat{H}^* - \epsilon_e - \delta)^2} \hat{\Phi}_e^\dagger \hat{v}^{1/2}_e | l \rangle \frac{\alpha^2}{\omega^2 + \alpha^2} = \sum_l \langle l | \sigma_e^{N0}(i\omega, \delta, \alpha) | l \rangle \\
\Sigma_e^A(\delta, \alpha) = \frac{1}{2} \sum_l \langle l | (\hat{\epsilon}^{-1}(0) - \hat{1}) \hat{v}^{1/2}_e \hat{\Phi}_e \hat{\Phi}_e^\dagger \frac{\alpha}{\hat{H}^* - \epsilon_e - \delta + \alpha} \hat{\Phi}_e^\dagger \hat{v}^{1/2}_e | l \rangle = \sum_l \langle l | \Sigma_e^A(\delta, \alpha) | l \rangle.
\]

(38)

The smallest orthonormal basis \( \{|l\} \) such that the above traces are converged must contain the subspace associated with the highest eigenvalues of \( \sigma_e^N(i\omega, \delta) \), \( \sigma_e^{N0}(i\omega, \delta) \) and \( \Sigma_e^A(\delta, \alpha) \). This subspace corresponds qualitatively to the intersection of the subspaces associated with the highest eigenvalues (in absolute value) of \( \hat{\epsilon}^{1/2}_e \), \( \hat{\Phi}_e \), \( \hat{H}^* - \epsilon_e - \delta \) and \( \hat{\epsilon}^{-1}(i\omega) - \hat{1} \). Since \( \hat{\epsilon}^{1/2}_e \) is diagonal in the basis of plane waves, \( \hat{\Phi}_e \) is diagonal in real space and \( \hat{H}^* - \epsilon_e - \delta \) is diagonal in the basis of complex-conjugated DFT states \( |n^*\rangle \), the subspace generated by the eigenvectors associated with their highest eigenvalues is readily available in the present formalism. However, \( \hat{\epsilon}^{-1}(i\omega) - \hat{1} \) is not known in any basis yet and, therefore, the subspace associated to its highest eigenvalues remains to be found.

In conventional plane wave implementations of \( G_0W_0 \), the dielectric matrix \( \hat{\epsilon} \) is obtained in a plane wave basis (using Eq. (18) and (17)) and directly inverted. This inversion then becomes a bottleneck, since the size of the basis in which wavefunctions are expanded is quite large (~115,000 plane waves for a molecule as simple as silane). Usually, this bottleneck is mitigated by expressing the dielectric matrix in a plane wave basis smaller than the one used for the wavefunctions. However, this practice leads to increased uncertainties and the calculation size remains limited by this factor. It is therefore desirable to obtain a basis spanning the subspace associated with the highest eigenvalues of \( \hat{\epsilon}^{-1}(i\omega) - \hat{1} \) directly, without explicitly expressing \( \hat{\epsilon} \) in a plane wave basis.

It is useful at this stage to have some insight into the spectrum of \( \hat{\epsilon}(i\omega) - \hat{1} \) for \( \omega \in \mathbb{R} \). From Eqs. (17) and (18), this operator can be written as

\[
\hat{\epsilon}(i\omega) - \hat{1} = 4 \sum_{cv} \hat{v}^{1/2}_c | e^* v \rangle \langle e - e^* v | \hat{\epsilon}^{1/2}_c,
\]

where \( (\epsilon_e - e^* v)/(\epsilon_e - e^* v)^2 + \omega^2) > 0 \forall \omega \in \mathbb{R} \) and where \( \hat{v}^{1/2}_c \) has only positive eigenvalues. It results that \( \hat{\epsilon}(i\omega) - \hat{1} \) has only positive eigenvalues. In particular, for isolated systems, its eigenvalue spectrum is formed by a few large discrete eigenvalues associated with transitions from valence states to bound conduction states and a continuous spectrum of smaller eigenvalues associated to transitions from valence states to a continuum of conduction states, with an integrable divergence at the origin.

Moreover, since the eigenvalues of \( \hat{\epsilon}(i\omega) - \hat{1} \) are in the range \([0, +\infty[ \), those of \( 1 - \hat{\epsilon}^{-1} \) are located in the range \([0,1[ \). Also, if the eigenvalues of both operators are sorted in decreasing order, corresponding eigenvalues will occupy the same rank. Consequently, the eigenvalues of \( 1 - \hat{\epsilon}^{-1} \) that contribute most to \( \Sigma_e^A(\delta) \) correspond to the largest eigenvalues of \( \hat{\epsilon} - \hat{1} \). However, in contrast to \( 1 - \hat{\epsilon}^{-1} \), it is possible to apply \( \hat{\epsilon} - \hat{1} \) on an arbitrary vector without having to explicitly construct its matrix representation, using Eqs. (17), (19) and (22).

Therefore, to approximately find the subspace associated to its largest eigenvalues, one can apply \( \hat{\epsilon} - \hat{1} \) repeatedly on some random vector \( |\psi\rangle \) to construct a Krylov subspace:

\[
\{ |\psi\rangle, (\hat{\epsilon} - \hat{1}) |\psi\rangle, (\hat{\epsilon} - \hat{1})^2 |\psi\rangle, \ldots, (\hat{\epsilon} - \hat{1})^N |\psi\rangle \}.
\]

(40)

Applying \( \hat{\epsilon} - \hat{1} \) on a random vector will cause the directions associated to the largest eigenvalues to grow faster than the others with respect to the number of applications. Thus, orthonormalizing the vectors of Eq. (40) would yield a basis approximately generating the desired subspace.

In practice, our implementation uses the vector \( |\psi\rangle = \hat{v}^{1/2}_c | e^* v \rangle \) as a starting point, but we observed that using any other vector (or all other vectors) of the form \( |\psi\rangle = \hat{v}^{1/2}_c | w^* v \rangle \) does not change significantly the number of dimensions \( N \) required to achieve convergence. Also, in practice, we use the Lanczos procedure to obtain an orthogonal basis that spans the subspace of Eq. (40) and tridiagonalizes \( \hat{\epsilon} - \hat{1} \). This procedure also provides the associated matrix elements and its cost is only marginally higher than the successive applications of \( \hat{\epsilon} - \hat{1} \) on \( |\psi\rangle \).

Theoretically, the Lanczos procedure should not require explicit orthogonalization of each basis vector with all previously generated ones. However, in practice, the
vectors generated by a direct implementation of the Lanczos procedure rapidly lose their orthogonality with the number of steps due to numerical error\textsuperscript{21}. For example, in our implementation, orthogonality is typically lost in about 10 steps while 100’s of steps are required to achieve convergence. Therefore, in our implementation of the Lanczos procedure, we added a Gram-Schmidt orthogonalisation of each residual vector with respect to all previously generated Lanczos vectors, so that orthogonality is achieved to machine precision. A fully converged calculation of $\Sigma_\epsilon^c(\delta)$ for the HOMO of silane and an identical one except for the omission of the orthogonalization of Lanczos vectors have shown a difference in computation time of less than 0.05\%. Thus, the necessity of introducing this orthogonalization has no impact on the performance of the present $G_0W_0$ implementation.

At this stage, it is possible to invert the dielectric matrix $\hat{\epsilon}$ at a very low computational cost, since the Lanczos basis is much smaller than the plane waves basis and since a tridiagonal matrix can be inverted at a cost $\propto N^2$, in contrast to $N^3$ for a full matrix. We thus approximately obtain $1 - \hat{\epsilon}^{-1}$ in the subspace of its $N$ largest eigenvalues, which contribute most per dimension to the quasiparticle energy $\Sigma_\epsilon^c(\delta)$. An alternative way to obtain $\hat{\epsilon}$ in approximately the same subspace is to iteratively diagonalize it\textsuperscript{22}, as was implemented in another plane waves $G_0W_0$ code\textsuperscript{14}. However, the latter scheme costs about 10 applications of $\hat{\epsilon}$ per dimension while our scheme costs a single application per dimension. This efficiency gain is made possible by the fact that adding an extra dimension to the Lanczos basis does not only converge the traces of Eq. (38) by adding a term to the sum. It also increases the agreement between all the eigenvalues of the tridiagonal matrix and the corresponding eigenvalues of the exact operator. Moreover, according to the Kage-Pangel theory\textsuperscript{21}, the quickest convergence is achieved for the eigenvalues most separated from the others, which are the largest ones in the case of the dielectric matrix, e.g. the ones contributing most to the final result $\Sigma_\epsilon^c(\delta)$. Thus, constructing the dielectric operator in a whole subspace at once using the Lanczos algorithm allows for a substantial efficiency gain with respect to iterative diagonalization techniques, mostly due to the fact that the latter cannot use the information obtained in the construction of an eigenvector/eigenvalue pair for the refinement of another pair.

There remains only one step to obtain a basis $\{|l]\}$ that approximately spans the smallest possible subspace where the traces of Eq. (38) are converged. It is to intersect the Lanczos basis obtained for $\hat{\epsilon}$ with the subspaces associated with the largest eigenvalues of the operators $\tilde{\epsilon}_c^{1/2}$, $\Phi_0$, and $(\hat{H}^* - \epsilon_c - \delta)^{-1}$. However, the cost of calculating the traces in Eq. (38) is of order $N^2 \log(N)$ for a plane wave implementation, which is lower than the $N^3$ order associated to the projection of the Lanczos vectors on the relevant subspaces. Moreover, leaving extra dimensions in the basis $\{|l]\}$ just increases the precision of the calculated trace. Therefore, in plane waves implementations, the most efficient choice of basis $\{|l]\}$ for the calculation of the traces in Eq. (38) is to pick the smallest of the four aforementioned subspaces, that is, the Lanczos basis for the dielectric matrix $\hat{\epsilon}$. As an example, calculations of $\Sigma_\epsilon^c(\delta)$ for the HOMO of silane require a Lanczos basis of dimension $\sim 500$ to be converged to $\sim 20$ meV, while conventional plane waves $G_0W_0$ calculations require a dielectric matrix of dimension $\sim 15,000$ to achieve a similar convergence.

\section{Dielectric Model}

We mentioned in section IV that, when the Lanczos basis $\{|l]\}$ of the dielectric operator $\hat{\epsilon} - 1$ is iteratively constructed, the fastest converging eigenvalues are the largest ones. Therefore, as the construction of $\{|l]\}$ progresses, an increasing proportion of the computational work becomes aimed at sampling the integrable divergence in the spectrum of eigenvalues of $1 - \hat{\epsilon}^{-1}$, while the large eigenvalues are already converged. We have also mentioned in section IV that the small, continuous eigenvalues making up this integrable divergence are associated with transitions from valence bands to free conduction states. Thus, at some point in the construction of $\{|l]\}$, when applying $\hat{\epsilon} - 1$ to the current basis vector $|l\rangle$, the largest terms in the sum of Eq. (39) will be associated to eigenenergies $\epsilon_c$ of free conduction states, dominated by kinetic energy. We will then have $\hat{H} |c\rangle = \epsilon_c |c\rangle \approx \hat{T} |c\rangle$, where $\hat{T}$ is the kinetic energy operator. Therefore, substituting $\hat{H} \rightarrow \hat{T}$ in Eqs. (21), (22) and (23) should become an accurate approximation at some point in the construction of $\{|l]\}$.

Since $\hat{T}$ is diagonal in the plane wave basis, in contrast to $\hat{H}$, the conversion of Eq. (21) into a linear equation problem (Eq. (22)) is no longer required. Indeed, the former equation can then be directly solved with a single application of $\hat{T}^{-1}$ on a vector at a cost $\propto N$, where $N$ is the number of plane waves in the basis in which the wavefunctions are expressed. In contrast, the solution of Eq. (22) or (23) using SQMR requires typically $\sim 15$ application of the hamiltonian $\hat{H}$ on a vector at a cost $\propto N \log(N)$. Since these successive applications of $\hat{H}$ are the bottleneck in the construction of $\{|l]\}$, it becomes interesting to use the approximation $\hat{H} \rightarrow \hat{T}$ in Eq. (21) as soon as it becomes accurate.

To implement this idea, we first define the approximate dielectric operator from Eqs. (17), (19) and (21), where we substitute $\hat{H} \rightarrow \hat{T}$

\[ \hat{\tilde{\epsilon}}(\omega) \equiv 1 - \tilde{\epsilon}_c^{1/2} \hat{P}(\omega) \tilde{\epsilon}_c^{1/2}, \]
\[ \hat{\tilde{P}}(\omega) |\psi_j\rangle \equiv -2 \sum_v |v\rangle \left( |f_{jv}^+(\omega)\rangle + |f_{jv}^-(\omega)\rangle \right), \]
\[ |\tilde{f}_{jv\pm}(\omega)\rangle \equiv \tilde{P}_c \sum_{v'} \frac{1}{G^2 \pm \epsilon_v \pm \omega} \langle G| \hat{P}_c |v\psi_j^*\rangle, \]

(41)
where $|G\rangle$ is a plane wave of the basis used to express the wavefunctions and $G$ is the corresponding wavevector. Then, we add and subtract the model operator $\hat{\epsilon}^{-1}$ from the exact operator $\hat{\epsilon}^{-1}$

$$
\hat{\epsilon}^{-1} - 1 = (\hat{\epsilon}^{-1} - \hat{\epsilon}^{-1}) + (\hat{\epsilon}^{-1} - 1)
$$

(42)

in $\sigma_{\epsilon}^N$, $\sigma_{\epsilon}^0$ and $\Sigma_{\epsilon}^A$ as expressed in Eq. (38). Since $\hat{\epsilon}^{-1}$ should accurately describes the integrable divergence in the spectrum of eigenvalues of $\hat{\epsilon}^{-1}$ near 1, the operator $(\hat{\epsilon}^{-1} - \hat{\epsilon}^{-1})$ should be devoid of such a divergence. Therefore, the calculation of its trace should require a smaller Lanczos basis than the $(\hat{\epsilon}^{-1} - 1)$ operator or the $\hat{\epsilon}^{-1}$ operator.

We exploit this by splitting each trace of Eq. (38) in two others. In the first traces, we substitute $(\hat{\epsilon}^{-1} - 1)$ by $(\hat{\epsilon}^{-1} - \hat{\epsilon}^{-1})$ and can thus use a smaller Lanczos basis. In the second traces, we substitute $(\hat{\epsilon}^{-1} - 1)$ by $(\hat{\epsilon}^{-1} - 1)$ and the basis size will remain similar. Thus, we use the exact dielectric operator $\hat{\epsilon}$ to generate the first basis $\{|l\rangle\}$ at a reduced cost, thanks to its smaller size. Then, we use the approximate dielectric operator $\hat{\epsilon}^{-1}$ to generate the second basis $\{|l\rangle\}$, also at a reduced cost, since the operator is simpler to apply.

Once those bases are available, we split each of $\sigma_{\epsilon}^N$, $\sigma_{\epsilon}^0$ and $\Sigma_{\epsilon}^A$ as expressed in Eqs. (29), (31) and (34) in a sum of two contributions. In the first contributions, we substitute $(\hat{\epsilon}^{-1} - 1)$ by $(\hat{\epsilon}^{-1} - \hat{\epsilon}^{-1})$ and evaluate the resulting expression using the basis $\{|l\rangle\}$. In the second contributions, we substitute $(\hat{\epsilon}^{-1} - 1)$ by $(\hat{\epsilon}^{-1} - 1)$ and evaluate the resulting expression using the basis $\{|l\rangle\}$. Then, we sum the two results and obtain $\sigma_{\epsilon}^N$, $\sigma_{\epsilon}^0$ and $\Sigma_{\epsilon}^A$ at a reduced computational cost.

It is interesting to note that, since the dielectric model is subtracted and added to the exact dielectric operator, the scheme described in this section does not introduce new approximations in the $G_0W_0$ formalism, provided that the size of the Lanczos bases $\{|l\rangle\}$ and $\{|l\rangle\}$ are sufficient to obtain converged results.

VI. CALCULATING THE DIELECTRIC MATRIX AND THE INTEGRAND AT DIFFERENT FREQUENCIES

In principle, the integration over frequencies $\omega$ in Eq. (12) could be carried out by using Eqs. (29), (31), and (34) and by building a new Lanczos basis $\{|l\rangle\}$ at each different value of $\omega$ (we consider the case where the dielectric model is not used for simplicity). However, in practice, the Lanczos bases for $\epsilon(\omega)$ at different imaginary $\omega$ span subspaces that do not differ significantly. This can be understood from Eq. (18) for the polarizability, which can be rewritten as

$$
P(\omega) = -4 \sum_{cv} |e^c v\rangle \frac{\epsilon_c - \epsilon_v}{\omega^2 + (\epsilon_c - \epsilon_v)^2} \langle v e^c|.
$$

(43)

When $\omega$ is displaced along the positive direction of the imaginary axis, each term of the sum decreases monotonically. In contrast, making $\omega$ real and displacing it along the real axis in Eq. (18) would cause strong changes in the terms of the summation, since they each contain one pole located on the real axis. Thus, we found that constructing a basis for the static dielectric matrix $\hat{\epsilon}(0)$ and using it to express the dynamic dielectric matrix $\hat{\epsilon}(\omega)$ at all other frequencies $\omega \in \{0, i\omega\}$ is a sound approximation, in agreement with previous work.14,26

Still, a new Lanczos basis must be constructed to calculate the dielectric matrix at each real frequency required by Eq. (11). However, in contrast to the other contributions to $\Sigma_{\epsilon}(\delta)$ as expressed in Eqs. (12) and (38), only a single matrix element of $\hat{\epsilon}^{-1} - 1$ per frequency is required in Eq. (11) instead of some related trace. This causes Eq. (11) to converge dramatically faster than Eq. (38) with respect to the size of the Lanczos basis, provided the seed vector is chosen appropriately. Indeed, to calculate a matrix element of the inverse dielectric matrix $\langle \epsilon_p | \hat{\epsilon}_{\omega}^{-1} | \epsilon_b \rangle$ (where $|\epsilon\rangle$ is the state generating the pole whose residual is being calculated, knowledge of the dielectric matrix in a subspace formed by the eigenvectors that both correspond to its largest eigenvalues and overlap substantially with the vector $\hat{\epsilon}_{\omega}^{1/2} |p e\rangle$ is required. The latter will automatically be satisfied if the seed vector is chosen to be $\hat{\epsilon}_{\omega}^{1/2} |p e\rangle$ and the former is a feature of the Lanczos procedure. Thus, with this choice of seed vector, we found that 4 Lanczos iterations converge $\Sigma_{\epsilon}(\delta)$ to 1 meV for all the systems studied. Together with the small number of terms involved by Eq. (11) when only states close to the band gap are corrected, the preceding observation keeps the computational time spent on $\Sigma_{\epsilon}(\delta)$ small with respect to the remainder of $\Sigma_{\epsilon}(\delta)$. Thus, the necessity of building a separate Lanczos basis for each real frequencies present in Eq. (11) has a small impact on the performance of the implementation.

Also, keeping the basis in which $\hat{\epsilon}$ is expressed fixed for all $\omega$ enables some tricks to speed up the calculation. For instance, we adopt the shift Lanczos technique to solve Eq. (30) simultaneously at all frequencies14,46 (a few tens of iterations is typically enough to converge $\Sigma_{\epsilon}$ to 5 meV). This technique requires that the linear equation have the general form $(\hat{M} + \hat{I}\omega) |\epsilon(\omega)\rangle = |b\rangle$ where both the operator $\hat{M}$ and the right-hand side $|b\rangle$ are independent of $\omega$. However, here $|b\rangle = (\hat{\epsilon} - \delta)^{1/2} |\epsilon_b\rangle$. Therefore, if $\{|l\rangle\}$ was dependent on $\omega$, Eq. (30) would have to be solved individually for each frequency $\omega$. Keeping the basis $\{|l\rangle\}$ fixed is thus required to allow Eq. (30) to be solved at all frequencies simultaneously.

It would also be possible to use the same technique to solve Eq. (23) (where we substitute $|\epsilon_p\rangle \rightarrow \hat{\epsilon}_{\omega}^{1/2} |l\rangle$ and $j \rightarrow l$ simultaneously at all frequencies, at the cost of applying the Hamiltonian $\hat{H}$ a few tens of times per dimension of the dielectric matrix14,46. However, it is possible to avoid any iterative solution of Eq. (23) in a
plane wave basis beside those that were already solved in the construction of the static dielectric matrix \( \hat{\epsilon}(0) \). To do this, we construct one basis per valence state \( v \), \( \{ \gamma_{i,v} \} \), much smaller than the plane wave basis, in which the Hamiltonian \( \hat{H} \) and the right-hand side of Eq. (23) are expressed, which allows to solve the latter equation by direct inversion of the resulting matrix at a negligible computational cost. To suitably choose the basis \( \{ \gamma_{i,v} \} \), it is useful to rewrite the quantity we wish to calculate with the solution of Eq. (23), that is, \( \langle l | \hat{\epsilon}(i\omega) - 1 | l' \rangle \), from Eqs. (17), (19), (23) and (24)

\[
\langle l | \hat{\epsilon}(i\omega) - 1 | l' \rangle = 4 \sum_v \langle l | \hat{\Phi}_v | l' \rangle \frac{(\hat{H}^* - \epsilon_v)^2 + \omega^2}{(\hat{H} - \epsilon_v)^2 + \omega^2} \hat{P}_v^{*} \hat{\Phi}_v \hat{\phi}_v^{1/2} \langle l' |\rangle ,
\]

or

\[
\langle l | \hat{\epsilon}(i\omega) - 1 | l' \rangle = 4 \sum_v \langle l | \hat{A}_v^{-1}(i\omega) | b_{l,v} \rangle ,
\]

\[
\langle l | \hat{\epsilon}(i\omega) - 1 | l' \rangle = 4 \sum_v \langle l | \hat{b}_{l,v} \hat{A}_v^{-1}(i\omega) \rangle \langle \hat{b}_{l,v} | l' \rangle ,
\]

where \( | b_{l,v} \rangle \equiv \hat{P}_v^{*} \hat{\Phi}_v \hat{\phi}_v^{1/2} | l \rangle \), \( \hat{A}_v(i\omega) \equiv ((\hat{H} - \epsilon_v)^2 + \omega^2)/((\hat{H} - \epsilon_v)^2 + \omega^2) \) and \( | x_{l,v}(i\omega) \rangle \equiv \hat{A}_v^{-1}(i\omega) | b_{l,v} \rangle \). The optimal basis to pick as \( \{ | \gamma_{i,v} \rangle \} \) would be the one spanning the same subspace as the eigenvectors \( \{ | \lambda_{i,v}(i\omega) \rangle \} \) of \( \hat{A}_v(i\omega) \) that contribute most to the desired quality \( \langle l | \hat{\epsilon}(i\omega) - 1 | l' \rangle = 4 \sum_{v} \lambda_{i,v}(i\omega) \langle \lambda_{i,v}(i\omega) | b_{l,v} \rangle \) that is, those that are both associated to small eigenvalues \( \lambda_{i,v}(i\omega) \) of \( \hat{A}_v(i\omega) \) and overlapping substantially with the basis \( | b_{l,v} \rangle \). The solutions \( | x_{l,v}(i\omega) \rangle \) are naturally dominated by these directeds, so that \( \{ | x_{l,v}(i\omega) \rangle \} \) should be a proper choice of basis in which to express Eq. (23).

There only remains to pick the frequency \( \omega \) at which the solutions \( \{ | x_{l,v}(i\omega) \rangle \} \) will be used to build the basis. At this point, it is interesting to note that \( \hat{A}_v(i\omega) \) is simply the shifted conjugated Hamiltonian at the beginning of the domain of integration \( \langle \hat{A}_v(0) = (\hat{H}^* - \epsilon_v) \rangle \) and monotonically evolves toward the inverse of this operator as the frequency increases \( \lim_{\omega \to +\infty} \hat{A}_v(i\omega) \to (\hat{H}^* - \epsilon_v)^{-1} \times \omega^2 \). The two most natural frequencies to select would therefore be 0 and \( i\omega \). Since \( \lim_{\omega \to +\infty} | x_{l,v}(i\omega) \rangle \to (\hat{H}^* - \epsilon_v) \times \omega^{-2} | b_{l,v} \rangle \), the second set of solutions is available at a negligible computational cost. Also, Eq. (23) has already been solved at \( i\omega = 0 \) for all \( l \) and \( v \) in the process of building the static dielectric matrix. Therefore, the \( | x_{l,v}(0) \rangle \) are available at no additional computational cost.

The resulting solutions can be made into a basis using singular value decomposition and orthonormalization. Finally, \( H_{l,v} = \langle \gamma_{i,v} | (\hat{H} - \epsilon_v) | \gamma_{j,v} \rangle \) and \( b_{l,t,v} = 2 \langle \gamma_{i,v} | \hat{P}_v^{*} \hat{\Phi}_v \hat{\phi}_v^{1/2} | l \rangle \) are computed once and Eq. (23) can be solved at all required frequencies by direct inversion of \( \hat{A}_v(i\omega) \) in the \( \{ | \gamma_{i,v} \rangle \} \) basis at a very low computational cost. Indeed, calculation times for the \( G_0W_0 \) correction to the HOMO of silane show only random fluctuations of about 1% when the number of frequencies considered for the numerical integration of Eq. (12) is varied between 1 and 12.

It is possible to control the accuracy of this choice of basis by adding a set of solutions \( | \gamma_{i,v}(i\omega_0) \rangle \) to those we have already selected and observing that the result is negligibly affected. We have selected for \( \omega_0 \) a value of 1.0 Ha, which is approximately the frequency at which the integrand in Eq. (12) is maximal for the HOMO of silane. The resulting value of \( \Sigma_{\epsilon}(\delta) \) was affected by about 7 meV. Further testing with values of 0.1 Ha and 0.01 Ha for \( \omega_0 \) has shown similar behavior.

VII. THEORETICAL ANALYSIS OF COMPUTATIONAL COST

Provided that DFT eigenstates \( | n \rangle \) and eigenvalues \( \epsilon_n \) have already been calculated, our implementation starts with the construction of the Lanczos bases \( \{ | l \rangle \} \) and \( \{ | l' \rangle \) as described in section IV and V. The bottlenecks of this step are the applications of \( \hat{\epsilon}(0) \) on the \( N_L \) Lanczos vectors \( | l \rangle \) as described by Eqs. (17), (19) and (22) and the applications of \( \hat{\epsilon}(0) \) on the \( N_L \) Lanczos vectors \( | l' \rangle \) as described by Eq. (41). By analysing the first bottleneck, we find that for each of the \( N_v \) terms of the sum in Eq. (19), Eq. (22) needs to be solved iteratively, which involves \( N_{SQMR} \) applications of \( \hat{H} - \epsilon_v \) on a vector and the same amount of orthogonalization with the valence wavefunctions. Applying the Hamiltonian costs \( N_{FFT} \log(N_{FFT}) \) operations in a plane waves implementation, where \( N_{FFT} \) is the number of components describing a wavefunction in real space, while the orthogonalization costs \( N_v N_{PW} \), where \( N_{PW} \) is the number of components describing a wavefunction in reciprocal space. Thus, we conclude that the worst scaling in this bottleneck comes from the orthogonalization and is \( \propto N_L N_v N_{SQMR} \), which is \( \propto N^4 \), since \( N_{SQMR} \) is independent of the system size. Similarly, the worst scaling in the construction of \( \{ | l \rangle \} \) is \( \propto N_L N_v N_{PW} \propto N^4 \).

It turns out that this \( N^4 \) scaling is also the worst scaling found in the remainder of the code, as can be expected in general for plane waves \( G_0W_0 \) codes. We proceed below with an exhaustive list of the operations in our implementation having this scaling. We limit ourselves to the case where no dielectric model is used, since the scalings involved in the latter are all identical to the corresponding scalings for the exact dielectric operator.

Once the Lanczos basis \( \{ | l \rangle \} \) is available, our implementation starts the calculations of Eq. (12) by the numerical integration of \( \sigma_{\epsilon}^{N_L}(i\omega, \delta) - \sigma_{\epsilon}^{N_L}(i\omega, \delta) \), whose integrand can be calculated using Eqs. (17), (19), (23), (24), (29), (30) and (31). The process starts with the construction, for each valence state \( v \), of a basis \( \{ | \gamma_{i,v} \rangle \} \) of dimension \( \propto N_L \) for the Sternheimer equation (Eq. (23)), as described in section VI. This involves \( N_v \) singular value decompositions that scale as \( \propto N_L^{2} N_{PW} \), which results in a total cost \( N_v N_L^{2} N_{PW} \propto N^4 \). Then, the projection
of Eq. (23) in these bases also cost $N_c N_p^2 N_{PW} \propto N^4$. Finally, the solution of Eq. (23) by direct inversion of $(\hat{H} - \varepsilon_n)|n\rangle^2$ in these bases scales as $N_c N_n N_L^3$. Since $N_c$ is the number of frequencies used for the numerical integration and is not dependent on system size, we obtain $N_c N_n N_L^3 \propto N^4$. All other operations in this section of the code scale at worst as $N^3 \log(N)$.

Next, our implementation calculate $\Sigma^A(\delta)$ using Eqs. (34) and (35). However, this part of the code scales as $N^3$ and is not a bottleneck of the calculation.

Finally, our implementation proceeds with the calculation of $\Sigma^L(\delta)$ using Eqs. (11), (17), (19) and (22) as well as the Lanczos algorithm described in section IV. However, again, this part of the code scales as $N^3$ and is not a bottleneck of the calculation.

VIII. RESULTS

The ABINIT software package\textsuperscript{29} is used to produce the DFT eigenstates and eigenvalues used as input in our $G_0 W_0$ calculations. To ease reproducibility, we choose to simulate all molecules with their experimental geometry: silane\textsuperscript{47}, thiophene\textsuperscript{48}, benzene\textsuperscript{49}, naphthalene\textsuperscript{48}, anthracene\textsuperscript{48}, tetracene\textsuperscript{48} and C\textsubscript{60}\textsuperscript{49,50}. We use LDA (Teter Pade parametrization\textsuperscript{51}) and/or PBE\textsuperscript{52} functional for DFT calculations, as specified in each result table. The corresponding Hartwigsen-Goedecker-Hutter LDA\textsuperscript{53} or PBE\textsuperscript{54} pseudopotentials are used throughout. The reference $G_0 W_0$ implementation in section IX is the conventional $G_0 W_0$ implementation present in ABINIT.

Since this study focuses on isolated systems, a spherical truncation of the Coulomb potential\textsuperscript{55} is used in both the current implementation and the reference implementation of $G_0 W_0$. Its radius is set to at least the length of the system plus 1 Å on each side, so that the wavefunctions are coupled together by the Coulomb operator over their full spatial extent. Also, the periodic replicas of the simulated system are kept apart by a distance equal to the radius of truncation of the Coulomb potential, so that the wavefunctions of the system are not coupled to those of one of its periodic replicas by the Coulomb operator. However, the spherical truncation of the Coulomb potential is not implemented in the DFT part of ABINIT. Moreover, in periodic simulations, eigenenergies are difficult to position with respect to vacuum. To simultaneously correct the eigenenergies for the spurious Coulomb interactions between periodic replicas of the molecule and position them with respect to the energy level of the vacuum, we also run DFT calculations using the BigDFT project\textsuperscript{56} as a plug-in for ABINIT. This code uses a localized basis set (Daubechies wavelets\textsuperscript{57}) and a Poisson solver with free boundary conditions\textsuperscript{58}, which allow for non-periodic simulations, free of spurious Coulombic interactions, and a physical determination of the energy level of the vacuum. A spacing of 0.25 bohr for the wavelet grid is used to converge the eigenenergies to 1 meV. The difference between the ionization energy obtained using BigDFT and the DFT part of ABINIT can then be added to the DFT eigenenergies to correctly position them with respect to the vacuum and eliminate the part of the spurious Coulombic interactions that causes a uniform shift of the energy levels. The part of the spurious Coulombic interaction that causes a non-uniform shift of the eigenvalues and a deformation of the eigenstates with respect to those obtained with free boundaries condition is assessed by increasing the side of the box in which the molecule is placed for both the DFT and $G_0 W_0$ calculations until the $G_0 W_0$ correction is converged. The radius of truncation of the Coulomb potential is kept to half the length of the side of the box at each step of this convergence.

Also, Gaussian quadratures with 8 points are used to integrate Eq. (12) (or Eq. (10) for the reference $G_0 W_0$ implementation). DFT eigenvalues and eigenstates are converged until the squared residual of the wavefunctions $(|\langle \hat{H} - \varepsilon_n | n \rangle|^2$ is less than $10^{-12}$ Ha$^2$. The solutions to Eq. (22) or Eq. (23) are iteratively refined until the squared residual $|\langle A(x) - b | b \rangle|^2$ is smaller than $10^{-20}$ Ha$^2$ or Ha$^4$, respectively. Eq. (11) is calculated using 4 iterations of the Lanczos scheme described in section VI. Eqs. (30) and (35) are solved with the shift-Lanczos method\textsuperscript{12,59}, where 8 and 16 iterations are used respectively. The parameter $\alpha$ of the Lorentzian model to the dielectric matrix (Eq. (34)) is kept fixed at 1.0 Ha. The frequencies $\delta$ chosen for the evaluation of $\Sigma^E(x)$ are $\varepsilon_n + \langle e | \Sigma^E - \Sigma^E \langle e |$ and their average. The cutoff energy is converged separately for the calculation of $\Sigma^E$ and $\Sigma^E(\delta)$ until the former is converged to 10 meV and the latter to 50 meV. The number of Lanczos vectors $N_L$ describing the exact part of the dielectric matrix $\hat{\varepsilon} - \hat{\varepsilon}$ (as described in Section V) and the number of Lanczos vectors $N_{Lc}$ describing the approximate dielectric matrix $\hat{\varepsilon} - 1$ are also selected so that $\Sigma^E(\delta)$ is converged to 50 meV, unless specified otherwise. For some molecules, it is mentioned that the dielectric model is not used. In these cases, the exact dielectric matrix $\hat{\varepsilon} - 1$ is constructed using $N_L$ Lanczos vectors and $N_{Lc}$ is not used.

A. Silane

A 30 bohr unit cell is used for the calculations on the silane molecule. A cutoff energy of 50 Ha is used for the calculation of $\Sigma^E$ while 20 Ha is used for $\Sigma^E(\delta)$. $N_L = 512$ and no dielectric model are used to describe the dielectric matrix, which converges $\Sigma^E(\delta)$ to 20 meV. The results are given in Table I.

Good agreement is observed among the DFT results. Indeed, while our DFT results are converged to 1 meV, the other DFT studies were converged to 0.1 eV, which is comparable to the agreement between the results. Also, excellent agreement is observed between our final $G_0 W_0$ result and those of Ref. 65. Incidentally, this high agreement is associated to an identical level of convergence for both studies ($\sim$0.05 eV). Also, the $G_0 W_0$ correc-
results are in eV. This is also given to ease comparison with future calculations. All occupations to the DFT results are 

|            | LDA G₀W₀ | Σₑ | Expt. |
|------------|----------|----|-------|
| Ref. 60    | −8.4     | −12.7 |       |
| Ref. 61    | −8.4     | −12.7 |       |
| Ref. 62    | −8.42    | −12.41 |       |
| This work  | −8.51    | −12.41 | 0.72  |
| Ref. 63    |          | −12.3 |       |
| Ref. 64    |          | −12.36 |       |

|            | PBE G₀W₀ | Σₑ |
|------------|----------|----|
| Ref. 65ᵃ   | −12.31   |    |
| Ref. 65ᵇ   | −12.31   |    |
| This work  | −8.51    | −12.30 | 0.74 |

ᵃ Result obtained with the FHI-aims code with a 16-parameter Padé analytic continuation for the self-energy.
ᵇ Result obtained with the TURBOMOLE code with no resolution-of-identity approximation.

TABLE II. Results (in eV) for the HOMO energy of the thiophene molecule.

|            | LDA G₀W₀ | Σₑ | Expt. |
|------------|----------|----|-------|
| Ref. 18    | −6.15    | −8.37 |       |
| Ref. 66ᵃ   | −6.06    | −8.55 |       |
|            | −6.06    | −8.69 |       |
| This work  | −6.04    | −8.93 | −0.17 |
| Ref. 67    |          | −8.85 |       |
| Ref. 68    |          | −8.86±0.02 |       |
| Ref. 69    |          | −9.0 |       |
| Ref. 14    | −5.70    | −8.49 |       |
| This work  | −5.86    | −8.73 | −0.16 |

ᵃ Ref. 66 performs all-electron calculations with the Dunning aug-cc-pVTZ and aug-cc-pVQZ basis, instead of the much reduced double-zeta plus polarization (DZP) basis of Ref. 18, but with the same Gaussian-basis contour-deformation methodology, and finds a G₀W₀ ionization energy for thiophene of 8.55 eV and 8.69 eV, respectively, with a starting LDA value of 6.06 eV with both basis. They conclude these results indicate the slow convergence with Gaussian basis in the case of small molecules with unbound virtual states.

A 28 bohr unit cell is used for the calculations on the benzene molecule. A cutoff energy of 30 Ha is used for the calculation of Σₑ while 20 Ha is used for Σₑ(δ). Nₐ = 240 and Nₐ = 1920 are used to describe the dielectric matrix. The results are given in Table III.

Again, the agreement between DFT results is good. However, the comparison of the G₀W₀ corrections becomes more difficult, since the DFT HOMO energies are not available in many cases. Since 0.1 eV is the convergence criteria typically used in other works for DFT eigenvalues and G₀W₀ corrections, it should be a reasonable estimate for the expected accuracy of the resulting G₀W₀ eigenenergies. In this regard, the majority of G₀W₀ results (e.g. Ref. 72, Ref. 15, Ref. 14 and Ref. 65) are in good agreement with our results, which supports the accuracy of our implementation. Moreover, our agreement with the experiment is also good.

TABLE III. Comparison between the present implementation and previously published G₀W₀ results for the highest occupied molecular orbital (HOMO) energy of the silane molecule. The underlying DFT HOMO energies are given to assess the agreement of the starting point of the G₀W₀ calculations. The correlation part of the G₀W₀ correction Σₑ is also given to ease comparison with future calculations. All results are in eV.

|            | LDA G₀W₀ | Σₑ | Expt. |
|------------|----------|----|-------|
| Ref. 60    | −8.4     | −12.7 |       |
| Ref. 61    | −8.4     | −12.7 |       |
| Ref. 62    | −8.42    | −12.41 |       |
| This work  | −8.51    | −12.41 | 0.72  |
| Ref. 63    |          | −12.3 |       |
| Ref. 64    |          | −12.36 |       |

A 30 bohr unit cell is used for all calculations on the napththalene molecule. A cutoff energy of 30 Ha is used for calculations on the benzene molecule. A cutoff energy of 30 Ha is used for the calculation of Σₑ while 20 Ha is used for Σₑ(δ). Nₐ = 240 and Nₐ = 1920 are used to describe the dielectric matrix. The results are given in Table III.

A 30 bohr unit cell is used for all calculations on the napththalene molecule. A cutoff energy of 30 Ha is used for the calculation of Σₑ while 20 Ha is used for Σₑ(δ). Nₐ = 240 and Nₐ = 1920 are used to describe the dielectric matrix. The results are given in Table III.
for the calculation of \( \Sigma_e^c \) while 20 Ha is used for \( \Sigma_e^c(\delta) \). \( N_L = 384 \) and \( N_L = 3072 \) are used to describe the dielectric matrix. The results are given in Table IV.

In the case of naphthalene, the agreement with the DFT result of Ref. 74 is excellent. The \( G_0W_0 \) corrections to the PBE results (−2.23 eV and −2.36 eV for Ref. 74 and the present work, respectively) also show a very good agreement, considered that Ref. 74 uses a LDA exchange-correlation kernel to enhance the description of the polarisability (see their Eq. (69)), while the present implementation do not (see Eqs. (17) and (18)). The others \( G_0W_0 \) results show reasonable but somewhat lesser agreement. Finally, our results are in good agreement with (and, incidentally, closest to) the experiment. This, together with the agreement with Ref. 74, seems to validate, again, the accuracy of our implementation.

### E. Anthracene

A 40 bohr unit cell is used for the calculations on the anthracene molecule. A cutoff energy of 30 Ha is used for the calculation of \( \Sigma_e^c \) while 20 Ha is used for \( \Sigma_e^c(\delta) \). \( N_L = 528 \) and \( N_L = 3168 \) are used to describe the dielectric matrix. The results are given in Table V.

Both at the DFT and the \( G_0W_0 \) level, we observe good agreements with previously published results. Interestingly, we observe better agreement with Ref. 14 and Ref. 74 than with Ref. 18, similarly to the case of the thiophene molecule. We also observe good agreement between our \( G_0W_0 \) results and the experiment.

### F. Tetracene

A 51 bohr unit cell is used for the calculations on the tetracene molecule. A cutoff energy of 30 Ha is used for the calculation of \( \Sigma_e^c \) while 20 Ha is used for \( \Sigma_e^c(\delta) \). \( N_L = 504 \) and \( N_L = 4032 \) are used to describe the dielectric matrix. The results are given in Table VI.

Again, we observe our results to be in good agreement with previously published results at the DFT level. At the \( G_0W_0 \) level, the agreement with Ref. 74 is good, as in the cases of naphthalene and anthracene. Also, the agreement with Ref. 18 is similar to the one observed in the case of thiophene, which is coherent with the uniform basis size they used to simulate all their molecules (see footnote of Table II). However, the agreement with Ref. 14 is somewhat lesser than expected. Given that our agreement with Ref. 14 is good for all other molecules investigated in this study, it is plausible that the difference stems from the choice of simulation parameters and not the implementation. Finally, we observe rea-

**TABLE III.** Results (in eV) for the HOMO energy of the benzene molecule.

|          | LDA | \( G_0W_0 \) | \( \Sigma_e^c \) | Expt. |
|----------|-----|--------------|-----------------|-------|
| Ref. 12  | -   | −9.8         |                 |       |
| Ref. 71  | -   | −9.88        |                 |       |
| Ref. 72  | -   | −9.06        |                 |       |
| Ref. 15  | −6.49| −9.05        |                 |       |
| Ref. 14  | -   | −9.22        |                 |       |
| This work | −6.50| −9.23 | −0.16          |       |
| Ref. 68  | -   | −9.24378±0.00007 |           |       |

**TABLE IV.** Results (in eV) for the HOMO energy of the naphthalene molecule.

|          | LDA | \( G_0W_0 \) | \( \Sigma_e^c \) | Expt. |
|----------|-----|--------------|-----------------|-------|
| Ref. 71  | -   | −8.69        |                 |       |
| Ref. 73  | -   | −7.61        |                 |       |
| This work | −5.67| −8.05 | −0.36          |       |
| Ref. 68  | -   | −8.144±0.001 |                 |       |

**TABLE V.** Results (in eV) for the HOMO energy of the anthracene molecule.

|          | LDA | \( G_0W_0 \) | \( \Sigma_e^c \) | Expt. |
|----------|-----|--------------|-----------------|-------|
| Ref. 18  | −5.47| −6.89        |                 |       |
| Ref. 14  | −5.18| −7.25        |                 |       |
| This work | −5.18| −7.31 | −0.52          |       |
| Ref. 68  | -   | −7.439±0.006 |                 |       |

**TABLE VI.** Results (in eV) for the HOMO energy of the tetracene molecule.

|          | LDA | \( G_0W_0 \) | \( \Sigma_e^c \) | Expt. |
|----------|-----|--------------|-----------------|-------|
| Ref. 18  | −5.15| −6.37        |                 |       |
| Ref. 14  | −4.85| −7.04        |                 |       |
| This work | −4.86| −6.79 | −0.61          |       |
| Ref. 68  | -   | −6.97±0.05   |                 |       |

**TABLE VI.** Results (in eV) for the HOMO energy of the tetracene molecule.

|          | LDA | \( G_0W_0 \) | \( \Sigma_e^c \) | Expt. |
|----------|-----|--------------|-----------------|-------|
| Ref. 74  | −4.65| −6.67        |                 |       |
| This work | −4.66| −6.57 | −0.60          |       |
TABLE VII. Results (in eV) for the HOMO energy of the C_{60} molecule.

|        | PBE | G_{0}W_{0} | \Sigma_{c}^{e} (eV) | Expt. |
|--------|-----|------------|---------------------|-------|
| Ref. 15 | −5.84 | −7.21      |                     |       |
| Ref. 14 | −5.81 | −7.31      |                     |       |
| This work | −5.84 | −7.25      | −0.03               |       |
| Ref. 75 |       | −7.6±0.2   |                     |       |

reasonable agreement between our G_{0}W_{0} results and the experiment.

G. C_{60}

A 40 bohr unit cell is used for the calculations on the C_{60} molecule. A cutoff energy of 50 Ha is used for the calculation of \Sigma_{c}^{e} while 20 Ha is used for \Sigma_{c}^{e}(\delta). N_{L} = 1024 and N_{PW} = 4096 are used to describe the dielectric matrix, which converges \Sigma_{c}^{e}(\delta) to 0.1 eV. The results are given in Table VII.

All DFT and G_{0}W_{0} results show excellent agreement. We also observe a reasonable agreement with experiment.

H. Common trends

In Tables I to VI, values of \Sigma_{c}^{e}(\Delta \varepsilon_{c}) (the correlation part of the G_{0}W_{0} correction, see Eq. (1)) are provided for both the LDA and the PBE functionals. It is interesting to note that when the functional is changed, the relative change of \Sigma_{c}^{e}(\Delta \varepsilon_{c}) is roughly of the same order of magnitude as the relative change in \varepsilon_{c} + \Delta \varepsilon_{c}, the G_{0}W_{0} eigenvalue. This similar relative change with respect to the choice of functional, combined with the much smaller size of the correlation part of the G_{0}W_{0} correction \Sigma_{c}^{e}(\Delta \varepsilon_{c}) with respect to the total G_{0}W_{0} eigenvalue, makes the former quite insensitive to the choice of functional for the molecules selected in our study.

Also, for molecules containing only first row elements, changing LDA for PBE had the same effect on all the DFT and G_{0}W_{0} ionization potentials: it lowered them by about 0.2 eV. Therefore, the G_{0}W_{0} correction \Delta \varepsilon_{c} = \langle \varepsilon| \Sigma(\varepsilon_{c} + \Delta \varepsilon_{c}) - V^{xc} | \varepsilon \rangle is quite insensitive to the choice of functional for these molecules. However, since their HOMOs are qualitatively similar (in that they are formed of delocalized \pi orbitals), the full range of molecules to which this trend applies is difficult to guess.

IX. PERFORMANCE

Since the scaling of our implementation is the same as the traditional G_{0}W_{0} implementation (\propto N^4 in both cases), it becomes interesting to assess the prefactor by direct comparison of the computation times for both methods. To this end, we calculate the expectation value of the correlation part of the self-energy for the HOMO orbital of the silane molecule \Sigma_{c}^{e}(0) using the present and the conventional G_{0}W_{0} implementation of the ABINIT project. The convergence studies are carried out with respect to the size of the dielectric matrix, which is controlled by number of Lanczos vectors N_{L} in the first implementation and the number of plane waves used to describe the dielectric matrix N_{PW}, in the second. The horizontal lines show the energy zone considered to be converged and the vertical lines show the approximate CPU time required to reach this level of convergence, based on a linear interpolation between the first data point to be converged and the preceding one.

FIG. 2. Convergence study on the expectation value of the correlation part of the self-energy for the HOMO orbital of the silane molecule \Sigma_{c}^{e}(0) using the present and the conventional G_{0}W_{0} methods. The convergence studies are carried out with respect to the size of the dielectric matrix, which is controlled by number of Lanczos vectors N_{L} in the first implementation and the number of plane waves used to describe the dielectric matrix N_{PW}, in the second. The unit cell size and the cutoff energy are common to both implementations to reach a given level of convergence, based on a linear interpolation between the first data point to be converged and the preceding one.
The results obtained for the convergence studies are shown in Fig. 2. This analysis shows that our implementation is about 500 times faster than the conventional one for this calculation size.

X. CONCLUSION

The $G_0W_0$ implementation presented here successfully circumvents the two bottlenecks present in conventional plane wave implementations. The conversion of the summations over conduction states into Sternheimer equations effectively eliminates the first bottleneck. The second one is solved by expressing the dielectric matrix in a Lanczos basis. This effectively reduces its size by orders of magnitude, to the level of spectral decomposition methods, while being computationally an order of magnitude cheaper than the latter. Also, we developed a model dielectric operator, which further reduces the size of the dielectric matrix without accuracy loss. Furthermore, we explored two ways to alleviate the computational cost of the integration over frequencies without resorting to approximations such as the plasmon pole model. First, a scalar Lorentzian model for the frequency dependence of the dielectric matrix on the imaginary axis is used to reduce the frequency sampling required to evaluate the integral. This particular model has the advantages of having a simple physical interpretation, presenting the right high-frequency behaviour and being compatible with the conversion of summations over conduction states into linear equations. Also, we use a scheme that provides the dielectric matrix at any imaginary frequency for a negligible computational cost, based on the recycling of the information computed in the construction of the static dielectric matrix. The latter two concepts reduce the computational cost of the integration over frequencies by an order of magnitude. This, combined with the elimination of the bottlenecks mentioned previously, effectively reduces the computation time required to achieve convergence by orders of magnitude with respect to conventional plane wave $G_0W_0$ implementations. Small tests for the silane molecule revealed a 500-fold speedup.

This reduction in computational cost is achieved while preserving the high numerical precision provided by plane wave basis sets. Indeed, this implementation uses the contour deformation technique at almost no additional cost with respect to plasmon pole approximations, thanks to the use of the Lorentzian model and the recycling scheme. Also, the conversion of the summations over conduction states into linear equation problems eliminates the need to converge the results with respect to the number of states included in these notoriously slow converging summations, thus eliminating a numerical source of uncertainty. Moreover, the natural ability of the Lanczos method to first explore the biggest contributions of the dielectric matrix to the $G_0W_0$ results smooths the convergence behaviour with respect to the matrix size. Finally, the use of the SQMR algorithm for the iterative solution of linear equation problems allows the $G_0W_0$ corrections to be computed directly at the desired real frequency, thus avoiding the need for analytic continuations and related stability considerations.

For all molecules considered in this work, the computed quasiparticle energies show good agreement with previously published $G_0W_0$ results, thus validating the accuracy of our implementation. Moreover, our agreements with experimental ionization energies are similar or better than those of previously published $G_0W_0$ studies, suggesting that this implementation is effective in preserving the full precision of the $G_0W_0$ method. Also, though the results presented here are for molecules, the extension of our method to crystals is straightforward and under way.

XI. ACKNOWLEDGMENTS

We would like to thank Xavier Blase and Steven Louie for helpful discussions. Also, we acknowledge the Natural Sciences and Engineering Research Council of Canada, the Fonds de recherche du Québec - Nature et technologies and the Regroupement québécois sur les matériaux de pointe for funding. Finally, we acknowledge Calcul Québec and Compute Canada for computational resources.

1 W. Kohn and L. J. Sham, Physical Review 140, A1133 (1965).
2 K. Capelle, Brazilian Journal of Physics 36, 1318 (2006).
3 R. M. Martin, Electronic Structure, Basic Theory and Practical Methods (Cambridge University Press, 2004).
4 G. Onida, L. Reining, and A. Rubio, Reviews of Modern Physics 74, 601 (2002).
5 W. Aulbur, L. Jonsson, and J. Wilkins, Solid State Physics: Advances in Research and Applications 54, 1 (2000).
6 L. Hedin, Physical Review 139, 796 (1965).
7 L. Hedin and S. Lundqvist, Solid State Physics 23, 1 (1970).
8 M. S. Hybertsen and S. G. Louie, Physical Review B 34, 5390 (1986).
9 B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Physical Review Letters 105, 146401 (2010).
10 C. Friedrich, M. C. Müller, and S. Blügel, Physical Review B 83, 081101 (2011).
11 I. Tamblyn, P. Darancet, S. Y. Quek, S. A. Bonev, and J. B. Neaton, Physical Review B 84, 201402 (2011).
12 P. Umari, G. Stenuit, and S. Baroni, Physical Review B 81, 115104 (2010).
Once the Taylor expansion of Eq. (1) is done, one can solve for $\Delta \varepsilon_C$. The evaluation of $\langle c | \Sigma(\omega) | c \rangle$ at 2 frequencies then allows $\Delta \varepsilon_C$ to be obtained. In the present implementation, we evaluate $\langle c | \Sigma(\omega) | c \rangle$ at 3 different energies to control the accuracy of the linear approximation for Eq. (1). We choose for these energies the values $\varepsilon_C$ (the DFT eigenvalue to be corrected), $\varepsilon_C + |c| \hat{v}_C - \hat{V}_{xc} |c\rangle$ (the associated Hartree-Fock eigenvalue, evaluated using the bare Coulomb potential $\hat{v}_C$ and the DFT orbital) and their average. This choice usually covers a range of energies nearly enclosing the corrected $G_0W_0$ eigenvalue and thus effectively controls the accuracy of the linearization of Eq. (1).

The symmetric form of the dielectric matrix \[ \tilde{\varepsilon}(\omega) = 1 - \hat{v}_C \hat{P}(\omega) \hat{v}_C^{-1/2} \] where $\hat{P}(\omega)$ is the irreducible polarizability, is hermitian, in contrast to its usual form $\varepsilon(\omega) = 1 - \varepsilon_C \hat{P}(\omega)$. This makes it possible to apply the matrix on a vector only one time per iteration instead of two in Lanczos based linear algebra methods. Indeed, in the present $G_0W_0$ implementation, a substantial part of the computation time is spent on the application of the dielectric matrix within the Lanczos algorithm. Using this form of the matrix thus halves this part of the computation time, which provides a substantial numerical advantage.

A simple change of variable $\omega = (1 + t)/(1 - t)$ allows $\varepsilon$ to remap the semi-infinite interval $[0, +\infty]$ to $[-1, 1]$. The integration can now be carried out numerically using methods such as Gaussian quadrature. We found that a sampling of about 10 frequencies converges the integral to 1 meV, in agreement with previous work.

Both SQMR and conjugate gradients algorithms solve approximately the linear equation:

\[ \hat{A} |x\rangle = |b\rangle, \]

where $\hat{A}$ is an hermitian operator, $|b\rangle$ is a vector and $|x\rangle$ is the solution to the linear equation. More precisely, at the $N$th iteration, they have constructed an approximate solution $|x_N\rangle$, which is the exact solution of the linear equation projected in the $N$-dimension Krylov subspace

\[ \{ |b\rangle, \hat{A} |b\rangle, \hat{A}^2 |b\rangle, \ldots, \hat{A}^{N-1} |b\rangle \}. \]

The key advantage of these methods is that they do not require the whole basis describing the Krylov subspace to be stored in memory. Instead, they use recursion relations where the solution $|x_{N+1}\rangle$ is expressed in term of $|x_N\rangle$ and a small constant number of other vectors. In the specific case of conjugate gradients, a $LDL^T$ factorization is used to obtain such a recursion relation, which requires the operator $\hat{A}$ to be positive definite. The SQMR method circumvent this requirement by replacing the $LDL^T$ factorization by a $QR$ factorization, which only requires the operator $\hat{A}$ to be nonsingular but increases slightly the computational cost. Since the SQMR and the MINRES methods are mathematically equivalent when no preconditioning is used, the interested reader may consult the original work on the MINRES method for a more detailed explanation of the method.

Discontinuities in $\Sigma_C^S(\delta)$ (see Eq. (32)) and $\Sigma_C^U(\delta)$ (see Eq. (11)) appear whenever there exists a $m$ such that $\delta \rightarrow \varepsilon_m - \varepsilon_\delta$ and graphically correspond to poles of the Green function crossing the imaginary axis in Fig. 1 as $\delta$ is varied. The choice of convention for $\Theta(0)$ must be made
so that $\Sigma^c(\delta)$ remains continuous in these situations. To proceed, we begin by observing that the sum of the discontinuous terms in Eq. (11) and Eq. (15) (those for which $n = m$) when $\delta \rightarrow (\varepsilon_m - \varepsilon_c)^{\pm}$ is

\[
-\frac{1}{2} |m| \langle \varepsilon^{1/2} (\varepsilon - 1) \varepsilon^{1/2} |m| v \rangle \quad \text{if } |m\rangle \in \{|v\}\},
\]

\[
+\frac{1}{2} |m| \langle \varepsilon^{1/2} (\varepsilon - 1) \varepsilon^{1/2} |m| v \rangle \quad \text{if } |m\rangle \in \{|c\}\}.
\]

The appropriate choice of convention for $\Theta(0)$ is therefore

\[
\Theta(x) = \begin{cases} 
1 & \text{if } x > 0, \\
1/2 & \text{if } x = 0, \\
0 & \text{if } x < 0.
\end{cases}
\] (45)

Graphically, this convention can be seen as including half of the pole of the Green function inside the contour when it lies precisely on the imaginary axis.

The choice $\delta = \langle \varepsilon \Sigma^{\text{xc}} - \hat{V}^{\text{xc}} | \varepsilon \rangle$ is motivated by the fact that it is the value of $\delta$ closest to $\Delta \varepsilon_m$ among those we use for our G$_0$W$_0$ simulations.

H.-V. Nguyen, T. A. Pham, D. Rocca, and G. Galli, Physical Review B 85, 081101 (2012).

National Institute of Standards, “Computational Chemistry Comparison and Benchmark DataBase (Version 15b),” NIST Standard Reference Database 101, http://ccdb.nist.gov (2013).

National Center for Biotechnology Information, “PubChem Compound Database,” http://pubchem.ncbi.nlm.nih.gov (2013).

O. Gunnarsson, Alkali-Doped Fullerenes, Narrow-Band Solids with Unusual Properties (World Scientific Publishing Company, 2004).

W. I. David, R. M. Ibberson, J. C. Matthewman, K. Prasides, T. J. S. Dennis, J. P. Hare, H. W. Kroto, R. Taylor, and D. R. Walton, Nature 353, 147 (1991).

S. Goedecker, M. Teter, and J. Hutter, Physical Review B 54, 1703 (1996).

J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters 77, 3865 (1996).

X. Gonze, B. Amadon, P. M. Anglade, J. M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Cote, T. Deutsch, L. Genovese, P. Ghosez, M. Giannoni, S. Goedecker, D. R. Hamann, P. Hermet, F. Jollet, G. Jonard, S. Leroux, M. Mancini, S. Mazevet, M. J. T. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G. M. Rignanese, D. Sangalli, R. Shaltuf, M. Torrent, M. J. Verstraete, G. Zerah, and J. W. Zwan zigter, “The ABINIT project,” http://www.abinit.org/ (2013).

M. Krack, Theoretical Chemistry Accounts 114, 145 (2005).

J. Spencer and A. Alavi, Physical Review B 77, 193110 (2008).

L. Genovese, A. Nee, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, and R. Schneider, Journal Of Chemical Physics 129, 014109 (2008).

I. Daubechies, Ten Lectures On Wavelets (SIAM, 1992).

L. Genovese, T. Deutsch, A. Neev, S. Goedecker, and G. Beylkin, Journal Of Chemical Physics 125, 074105 (2006).

D. Rocca, R. Gebauer, Y. Saad, and S. Baroni, Journal Of Chemical Physics 128, 154105 (2008).

M. Rohlfing and S. G. Louie, Physical Review Letters 80, 3320 (1998).

J. C. Grossman, M. Rohlfing, L. Mitas, S. G. Louie, and M. L. Cohen, Physical Review Letters 86, 472 (2001).

P. H. Hahn, W. G. Schmidt, and F. Bechstedt, Physical Review B 72, 245425 (2005).

R. Roberge, C. Sandorfy, J. I. Matthews, and O. P. Straus, Journal Of Chemical Physics 69, 5105 (1978).

U. Itoh, Y. Toshoshima, H. Onuki, N. Washida, and T. Ikuki, The Journal of Chemical Physics 85, 4867 (1986).

M. J. van Setten et al., “GW 100: Benchmarking G0W0 for molecular systems,” (2014), unpublished.

P. Boulanger and X. Blase, unpublished.

L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, The Journal Of Chemical Physics 106, 1063 (1997).

National Institute of Standards, “NIST Chemistry WebBook,” NIST Standard Reference Database 69, http://webbook.nist.gov/chemistry/ (2013).

S. Sharifzadeh, I. Tamblyn, P. Doak, P. T. Darancet, and J. B. Neaton, The European Physical Journal B 85, 323 (2012).

We suppose our convergence criteria contribute to the final uncertainty in an uncorrelated manner. Therefore, we estimate the uncertainty on the final results due to the convergence criteria as:

\[
\Delta \approx \sqrt{(0.05 \text{ eV})^2 + (0.05 \text{ eV})^2} \approx 0.07 \text{ eV},
\] (46)

where the convergence criteria on the Lanczos basis and the cutoff energy of the plane waves basis have both contributed 0.05 eV to the estimated error.

M. L. Tiago and J. R. Chelikowsky, Physical Review B 73, 205334 (2006).

X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, New Journal Of Chemical Physics 14, 053020 (2012).

D. Försster, P. Koval, and D. Sánchez-Portal, Journal Of Chemical Physics 135, 074105 (2011).

M. J. van Setten, F. Weigend, and F. Evers, Journal Of Chemical Theory And Computation 9, 232 (2013).

D. L. Lichtenberger, K. W. Nebesny, C. D. Ray, D. R. Huffman, and L. D. Lamb, Chemical Physics Letters 176, 203 (1991).

C. C. Paige and M. A. Saunders, SIAM Journal on Numerical Analysis 12, 617 (1975).