Efficient dielectric matrix calculations using the Lanczos algorithm for fast many-body G_0W_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*

We present a G_0W_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 that allows us 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 quasiminimal residual scheme in replacement of the conjugate gradients algorithm allows a direct evaluation of the G_0W_0 corrections at the desired real frequencies, without need for analytical continuation. The performance of the resulting G_0W_0 implementation is demonstrated by comparison with a traditional plane-waves implementation, which reveals a 500-fold speedup for the silane molecule. Finally, the accuracy of our G_0W_0 implementation is demonstrated by comparison with other G_0W_0 calculations and experimental results.

I. INTRODUCTION

Density functional theory (DFT) [1–3] is currently the most popular approach for electronic structure simulations 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 [4,5] method is provided by the GW framework [6,7]. 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 [8], two bottlenecks account for this limitation: the sums to be carried out over all conduction states [9–12] and the inversion of the dielectric matrix [13–15] 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 [13,15–17], the conduction states can be replaced by simple approximate orbitals [18], a so-called extrapolar trick can be used to reduce the number of conduction states required for a given level of convergence [19,20], the summations can be eliminated using the effective-energy technique [21,22] or the size of the basis in which the Hamiltonian is expressed can be reduced through the use of localized basis sets [23]. 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 [14,17] or the size of the basis in which the dielectric matrix is expressed can be reduced either using Wannier orbitals [24] or eigenvectors of the static dielectric matrix [15].

In this paper, we present a plane-wave implementation of the G_0W_0 method within the ABINIT project [25–29] that circumvents both bottlenecks. The choice of plane waves is motivated by its suitability for extended systems as well as its systematic convergence controlled by a single parameter. In this implementation, 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 [13,15–17]. To assess the bottleneck of the inversion of the dielectric matrix, we elaborate an approach where the matrix is expressed in a Lanczos basis [30]. This reduces the size of the matrix as effectively as the traditional spectral decomposition method [15,31,32] while being computationally an order of magnitude more efficient. We also develop a model dielectric operator that allows to further reduce the size of the dielectric matrix.

Furthermore, in the present G_0W_0 implementation, the use of the contour deformation technique [33,34] was preferred over plasmon pole models [8,33,35–37] to avoid considerations on the range of systems that can be accurately described [9,38]. Traditionally, this choice implies a greater computational cost. In the present method, we explore two different directions to reduce this cost to the level of the simplest plasmon pole model [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 [39], 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 nonzero

─────────

[*]Corresponding author: michel.cote@umontreal.ca

frequencies at a small computational cost, independent of the number of frequencies, thus reducing the numerical cost of the contour deformation technique to a level close to the simplest plasmon pole model [8].

The evaluation of the G_0W_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 [13,15,17,40]. In the present implementation, we solve this problem using a recently developed numerical method, the simplified quasiminimal residual (SQMR) algorithm [41]. 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.

This article is organized as follows. First, a summary of the G_0W_0 method is given in Sec. II. Then, the bottleneck of the sums over conduction states is assessed in Sec. III. In particular, the Lorentzian model is developed in Sec. III B. Next, the bottleneck of the dielectric matrix inversion is assessed in Sec. IV. Then, the model dielectric operator is developed in Sec. 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 Sec. VI. A theoretical analysis of the numerical cost of the present G_0W_0 implementation is given in Sec. VII. We compare our implementation with existing G_0W_0 schemes that also convert the summations over conduction states into Sternheimer equations in Sec. VIII. Then, we assess the accuracy of our implementation by comparing our results with previously published ones in Sec. IX. Finally, we assess its performance with respect to traditional implementations in Sec. X. Atomic units are used throughout unless otherwise specified.

II. THE G_0W_0 METHOD

In conventional implementations of the G_0W_0 method, corrections to DFT eigenenergies are obtained using first-order perturbation theory, where the perturbation is the difference between the G_0W_0 self-energy $\hat{\Sigma}(\omega)$ and the DFT exchange-correlation potential $\hat{V}^{xc}$:

$$\Delta\varepsilon_e = \langle e| \hat{\Sigma}(\varepsilon_e + \Delta\varepsilon_e) - \hat{V}^{xc} |e\rangle, \quad (1)$$

where ε_e is a DFT eigenenergy, $|e\rangle$ is the associated eigenstate, and $\Delta\varepsilon_e$ is the G_0W_0 correction to ε_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 [42]. Also, the DFT exchange-correlation energy $\langle e| \hat{V}^{xc} |e\rangle$ can easily be extracted from the DFT calculation. The only nontrivial part of the calculation is therefore the evaluation of the G_0W_0 exchange-correlation energy $\langle e| \hat{\Sigma}(\omega) |e\rangle$, where $\hat{\Sigma}(\omega)$ is defined as

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

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

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

$$\hat{G}_0(\omega) = \sum_{n=1}^{\infty} \frac{|n\rangle \langle n|}{\omega - \varepsilon_n + i\eta \, \mathrm{sgn}(\varepsilon_n - \mu)}, \quad (3)$$

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

$$\hat{W}_0(\omega) = \hat{v}_C^{1/2}\hat{\epsilon}^{-1}(\omega)\hat{v}_C^{1/2}, \quad (4)$$

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 [43]. These definitions allow us to express the G_0W_0 exchange-correlation energy as

$$\Sigma_e(\delta) \equiv \langle e| \hat{\Sigma}(\varepsilon_e + \delta) |e\rangle$$
$$= \frac{i}{2\pi} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} d\omega e^{i\eta\omega} \frac{\langle en^*| \hat{v}_C^{1/2}\hat{\epsilon}^{-1}(\omega)\hat{v}_C^{1/2} |n^*e\rangle}{\omega - \omega_{ne} + i\eta \, \mathrm{sgn}(\varepsilon_n - \mu)}, \quad (5)$$

where $\omega_{ne} \equiv \varepsilon_n - \varepsilon_e - \delta$, $\langle r|n^*\rangle \equiv \langle r|n\rangle^*$, and $\langle r|n^*e\rangle \equiv \langle r|n\rangle^* \langle r|e\rangle$. 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,

$$\Sigma_e^x(\delta) \equiv \langle e| \hat{\Sigma}^x(\varepsilon_e + \delta) |e\rangle$$
$$= \frac{i}{2\pi} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} d\omega e^{i\eta\omega} \frac{\langle en^*| \hat{v}_C^{1/2} \hat{1} \, \hat{v}_C^{1/2} |n^*e\rangle}{\omega - \omega_{ne} + i\eta \, \mathrm{sgn}(\varepsilon_n - \mu)} \quad (6)$$

and

$$\Sigma_e^c(\delta) \equiv \langle e| \hat{\Sigma}^c(\varepsilon_e + \delta) |e\rangle$$
$$= \frac{i}{2\pi} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} d\omega e^{i\eta\omega}$$
$$\times \frac{\langle en^*| \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(\omega) - \hat{1})\hat{v}_C^{1/2} |n^*e\rangle}{\omega - \omega_{ne} + i\eta \, \mathrm{sgn}(\varepsilon_n - \mu)}. \quad (7)$$

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 Σ_e^x. 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 μ [see Eq. (6)], the poles included in the contour are those associated to the valence states. The exchange term thus takes the familiar form

$$\Sigma_e^x = \langle e| \hat{\Sigma}^x |e\rangle = -\sum_v \langle ev^*| \hat{v}_C |v^*e\rangle, \quad (8)$$

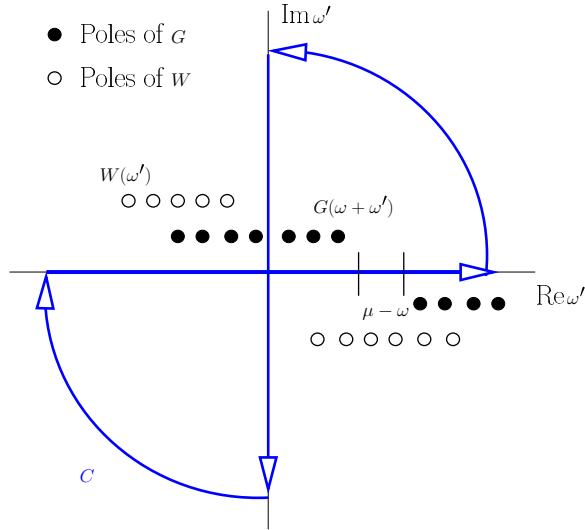where the index v labels the valence states.

FIG. 1. (Color online) The path used in the contour deformation technique. The poles of the screened Coulomb interaction $\hat{W}_0(\omega)$ lie outside the contour, only some poles of the Green's function $\hat{G}_0(\omega)$ lie inside. This figure is reproduced with permission from Ref. [33].

A popular way to calculate the integration over ω in $\Sigma_e^c(\delta)$ is the plasmon-pole approximation [8,33,35–37], 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 [9,38]. 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 [33,34,39,44].

The correlation part then becomes

$$\Sigma_e^c(\delta) = \frac{-i}{2\pi} \sum_{n=1}^{\infty} \left(\int_{C1} + \int_{-i\infty}^{i\infty} + \int_{C3} \right)$$

$$\times \, d\omega \frac{\langle en^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(\omega) - \hat{1})\hat{v}_C^{1/2} \, |n^*e\rangle}{\omega - \omega_{ne}}$$

$$- \sum_v \langle ev^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(\omega_{ve}) - \hat{1})\hat{v}_C^{1/2} \, |v^*e\rangle \, \Theta(\omega_{ve})$$

$$+ \sum_c \langle ec^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(\omega_{ce}) - \hat{1})\hat{v}_C^{1/2} \, |c^*e\rangle \, \Theta(-\omega_{ce}),$$

$$(9)$$

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 $\lim_{|\omega| \to \infty}(\hat{\epsilon}^{-1}(\omega) - \hat{1}) \to 0$ as $1/\omega^2$. Consequently, the integrals over both quarters of circle ($\int_{C1}$ and $\int_{C3}$) vanish. By substituting $\omega \to i\omega$, the domain of integration of the second term of Eq. (9) can be

made real,

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

$$\times \frac{\omega_{ne}}{\omega^2 + \omega_{ne}^2} + \Sigma_e^P(\delta), \qquad (10)$$

where we have defined the residue term

$$\Sigma_e^P(\delta) \equiv - \sum_v \langle ev^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(\omega_{ve}) - \hat{1})\hat{v}_C^{1/2} \, |v^*e\rangle \, \Theta(\omega_{ve})$$

$$+ \sum_c \langle ec^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(\omega_{ce}) - \hat{1})\hat{v}_C^{1/2} \, |c^*e\rangle \, \Theta(-\omega_{ce}).$$

$$(11)$$

The matrix element $\langle en^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(i\omega) - \hat{1})\hat{v}_C^{1/2} \, |n^*e\rangle$ in the integral over frequencies varies slowly as a function of ω with respect to the Lorentzian $\omega_{ne}/(\omega^2 + \omega_{ne}^2)$. Therefore the integral can be made smoother by subtracting a properly normalized Lorentzian from the integrand and carrying out its integration analytically. This step turns out to be necessary when $\omega_{ne} \to 0$, e.g., when a pole of the Green's function lies on the imaginary axis, since the Lorentzian then becomes a Dirac delta and its numerical integration then becomes problematic. This idea can be refined by multiplying the Lorentzian by a scalar function $f(\omega)$ that models approximately the frequency dependence of the dielectric matrix, which makes the integrand smaller and thus easier to sample numerically [39]. The resulting expression is

$$\Sigma_e^c(\delta)$$

$$= \frac{1}{\pi} \int_0^{\infty} d\omega \big(\sigma_e^N(i\omega,\delta) - \sigma_e^{N0}(i\omega,\delta) \big) + \Sigma_e^A(\delta) + \Sigma_e^P(\delta),$$

$$(12)$$

where

$$\sigma_e^N(i\omega,\delta) \equiv \sum_n \langle en^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(i\omega) - \hat{1})\hat{v}_C^{1/2} \, |n^*e\rangle$$

$$\times \frac{\omega_{ne}}{\omega^2 + \omega_{ne}^2}, \qquad (13)$$

$$\sigma_e^{N0}(i\omega,\delta) \equiv \sum_n \langle en^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(0) - \hat{1})\hat{v}_C^{1/2} \, |n^*e\rangle$$

$$\times f(\omega) \frac{\omega_{ne}}{\omega^2 + \omega_{ne}^2}, \qquad (14)$$

$$\Sigma_e^A(\delta) \equiv \frac{1}{\pi} \int_0^{\infty} d\omega \, \sigma_e^{N0}(i\omega,\delta)$$

$$= \sum_n \langle en^*| \, \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(0) - \hat{1})\hat{v}_C^{1/2} \, |n^*e\rangle \, F_{ne}(\delta), \quad (15)$$

and where

$$F_{ne}(\delta) \equiv \frac{1}{\pi} \int_0^{\infty} d\omega f(\omega) \frac{\omega_{ne}}{\omega^2 + \omega_{ne}^2}. \qquad (16)$$

In this form, the numerical integration required by Eq. (12) becomes straightforward [45].

Equations (11)–(16) contain quantities that can all be obtained from DFT calculations (eigenvalues ε_n and eigenfunctions $|n\rangle$), except for the inverse dielectric matrix $\hat{\epsilon}^{-1}(\omega)$. The latter is usually obtained by direct inversion of the dielectric matrix in some basis. The required dielectric matrix elements can be obtained from DFT eigenvalues and eigenfunctions using the Adler-Wiser expression for the random phase approximation to the irreducible polarizability [7,46,47]:

$$\hat{\epsilon}(\omega) = 1 - \hat{v}_C^{1/2} \hat{P}(\omega) \hat{v}_C^{1/2}, \tag{17}$$

$$\hat{P}(\omega) = 2 \sum_{cv} |c^*v\rangle \left[\frac{1}{\omega - (\varepsilon_c - \varepsilon_v)} - \frac{1}{\omega + (\varepsilon_c - \varepsilon_v)} \right] \langle vc^*|. \tag{18}$$

The preceding equations suffice to describe conventional G_0W_0 calculations in a plane-wave basis set, which involve calculating a sufficient number of conduction states to converge the summations in Eqs. (13), (14), (15), and (18) as well as inverting $\hat{\epsilon}(\omega)$ 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 [13,15–17]. This strategy is commonly used in density functional perturbation theory [48,49], where this type of linear equation is frequently referred to as a *Sternheimer equation* [50]. First, we will apply this idea to the polarizability. Then, we will eliminate the summation over states present in Eqs. (13)–(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,

$$\hat{P}(\omega) |\psi_j\rangle = -2 \sum_v |v\rangle \left(\sum_c |c^*\rangle \frac{1}{\varepsilon_c - \varepsilon_v - \omega} \langle vc^*|\psi_j\rangle \right.$$
$$\left. + \sum_c |c^*\rangle \frac{1}{\varepsilon_c - \varepsilon_v + \omega} \langle vc^*|\psi_j\rangle \right),$$
$$= -2 \sum_v |v\rangle (|f_{jv-}^*(\omega)\rangle + |f_{jv+}^*(\omega)\rangle), \tag{19}$$

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

$$|f_{jv\pm}(\omega)\rangle = \sum_c |c\rangle \frac{1}{\varepsilon_c - \varepsilon_v \pm \omega} \langle c|v\psi_j^*\rangle. \tag{20}$$

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

$$|f_{jv\pm}(\omega)\rangle = \frac{\hat{\mathcal{P}}_c}{\hat{H} - \varepsilon_v \pm \omega} |v\psi_j^*\rangle. \tag{21}$$

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,

$$(\hat{H} - \varepsilon_v \pm \omega) |f_{jv\pm}(\omega)\rangle = \hat{\mathcal{P}}_c |v\psi_j^*\rangle. \tag{22}$$

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 $\beta \mathcal{P}_v$ to the Hamiltonian $\hat{H}$, where β is larger than the valence bandwidth. Thus, when $\omega \to 0$, the operator acting on $|f_{jv\pm}(\omega)\rangle$ does not become singular [48]. 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 ω is real, as it is the case in Eq. (11), this trick does not prevent the left-hand side from being singular. Indeed, δ could be chosen so that ω_{ve}, for some $\varepsilon_v > \varepsilon_e + \delta$ (or ω_{ce}, for some $\varepsilon_c < \varepsilon_e + \delta$) equals some valence-conduction transition $\omega_{c'v'}$ (or $-\omega_{c'v'}$). Then, the operator $\hat{H} - \varepsilon_{v'} - \omega_{ve}$ (or $\hat{H} - \varepsilon_{v'} + \omega_{ce}$) would become singular in some part of the conduction subspace. Since the right-hand side of Eq. (22) can have nonzero components in this subspace, the previous trick cannot be applied to eliminate the singularity. Indeed, there will be an associated subspace where $\hat{\epsilon}(\omega_{ce})$ (or $\hat{\epsilon}(\omega_{ve})$) will be infinite and where $1 - \hat{\epsilon}^{-1}$ will be 1, so that the kernel of $\hat{H} - \varepsilon_{v'} \pm \omega$ will contribute to $\Sigma_e^P(\delta)$ as per Eq. (11). However, it is still possible to stabilize Eq. (22) without altering this physical contribution by using the SQMR algorithm [41] instead of conjugate gradients [30,51]. Indeed, the former is stable for indefinite matrices close to singularity ($\langle n| \hat{H} - \varepsilon_v \pm \omega |n\rangle \gtrsim 10^{-3}$ Ha $\forall |n\rangle$) while the latter is stable only for positive definite matrices [52]. It thus becomes easy to choose a suitable value of δ 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 $\beta \mathcal{P}_v$ 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 $\Sigma_e^c(\delta)$ at the (real) desired frequency, without requiring the addition of an imaginary infinitesimal to the Hamiltonian in Eq. (22) and the associated convergence study [53] nor the use of analytic continuation [15] and related stability considerations [40].

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

$$((\hat{H} - \varepsilon_v)^2 - (i\omega)^2) |f_{jv}(i\omega)\rangle = 2(\hat{H} - \varepsilon_v) \hat{\mathcal{P}}_c |v\psi_j^*\rangle, \tag{23}$$

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

$$|f_{jv}(\omega)\rangle \equiv |f_{jv+}(\omega)\rangle + |f_{jv-}(\omega)\rangle. \tag{24}$$

Typically, solving this equation to a precision where the residual $|((\hat{H} - \varepsilon_v)^2 - (i\omega)^2) |f_{jv}(i\omega)\rangle - 2(\hat{H} - \varepsilon_v) \hat{\mathcal{P}}_c |v\psi_j^*\rangle|$ is

less than 10^{-10} Ha^2 converges the calculated $\Sigma_e^c(\delta)$ 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 $\Sigma_e^c(\delta)$. 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 $\Sigma_e^c(\delta)$. For $\sigma_e^N(i\omega,\delta)$, we obtain from Eq. (13),

$$
\sigma_e^N(i\omega,\delta) = \sum_{n,l,l'} \langle en^*| \hat{v}_C^{1/2} |l\rangle \langle l| (\hat{\epsilon}^{-1}(i\omega) - \hat{1})|l'\rangle
$$
$$
\times \langle l'| \hat{v}_C^{1/2} |n^*e\rangle \frac{\omega_{ne}}{\omega^2 + \omega_{ne}^2},
$$
$$
= \sum_{n,l,l'} \langle l| (\hat{\epsilon}^{-1}(i\omega) - \hat{1}) |l'\rangle
$$
$$
\times \langle l^*| \hat{v}_C^{1/2} |e^*n\rangle \frac{\omega_{ne}}{\omega^2 + \omega_{ne}^2} \langle ne^*| \hat{v}_C^{1/2} |l'^*\rangle . \quad (25)
$$

Let $\hat{\Phi}_e$ be an operator so that

$$
\langle \phi| \hat{\Phi}_e |r\rangle \equiv \phi^*(r)\phi_e(r). \quad (26)
$$

Then, $\sigma_e^N(i\omega,\delta)$ becomes

$$
\sigma_e^N(i\omega,\delta) = \sum_{n,l,l'} \langle l| (\hat{\epsilon}^{-1}(i\omega) - \hat{1}) |l'\rangle
$$
$$
\times \langle l^*| \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger |n\rangle \frac{\omega_{ne}}{\omega^2 + \omega_{ne}^2} \langle n| \hat{\Phi}_e \hat{v}_C^{1/2} |l'^*\rangle . \quad (27)
$$

Similarly to the case of the polarizability, we need to replace the eigenvalues ε_n by the Hamiltonian $\hat{H}$ to use the completeness relation,

$$
\sigma_e^N(i\omega,\delta) = \sum_{l,l'} \langle l| (\hat{\epsilon}^{-1}(i\omega) - \hat{1}) |l'\rangle \langle l^*| \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger
$$
$$
\times \frac{\hat{H} - \varepsilon_e - \delta}{\omega^2 + (\hat{H} - \varepsilon_e - \delta)^2} \hat{\Phi}_e \hat{v}_C^{1/2} |l'^*\rangle . \quad (28)
$$

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

$$
\sigma_e^N(i\omega,\delta) = \sum_{l,l'} \langle l| (\hat{\epsilon}^{-1}(i\omega) - \hat{1}) |l'\rangle
$$
$$
\times \langle l^*| \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger |\sigma_{e,l'}^N(i\omega,\delta)\rangle , \quad (29)
$$

which is given by the following linear equation,

$$
(\omega^2 + (\hat{H} - \varepsilon_e - \delta)^2) |\sigma_{e,l'}^N(i\omega,\delta)\rangle
$$
$$
= (\hat{H} - \varepsilon_e - \delta) \hat{\Phi}_e \hat{v}_C^{1/2} |l'^*\rangle . \quad (30)
$$

Similarly, for $\sigma_e^{N0}(i\omega,\delta)$, we obtain

$$
\sigma_e^{N0}(i\omega,\delta) = \sum_{l,l'} \langle l| (\hat{\epsilon}^{-1}(0) - \hat{1}) |l'\rangle f(\omega)
$$
$$
\times \langle l^*| \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger |\sigma_{e,l'}^N(i\omega,\delta)\rangle . \quad (31)
$$

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

$$
\Sigma_e^A(\delta) = \sum_{n,l,l'} \langle l| (\hat{\epsilon}^{-1}(0) - \hat{1}) |l'\rangle
$$
$$
\times \langle l^*| \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger |n\rangle F_{ne}(\delta) \langle n| \hat{\Phi}_e \hat{v}_C^{1/2} |l'^*\rangle .
$$

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

$$
f(\omega) = 1,
$$
$$
F_{ne}(\delta) = \frac{1}{2}(\Theta(\omega_{ne}) - \Theta(-\omega_{ne})),
$$
$$
\Sigma_e^A(\delta) = \frac{1}{2} \sum_{l,l'} \langle l| (\hat{\epsilon}^{-1}(0) - \hat{1}) |l'\rangle
$$
$$
\times \langle l^*| \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger (\hat{\mathcal{Q}}_e(\delta) - \hat{\mathcal{P}}_e(\delta)) \hat{\Phi}_e \hat{v}_C^{1/2} |l'^*\rangle , \quad (32)
$$

where

$$
\hat{\mathcal{Q}}_e(\delta) \equiv \sum_{\varepsilon_n > \varepsilon_e + \delta} |n\rangle \langle n| ,
$$
$$
\hat{\mathcal{P}}_e(\delta) \equiv \sum_{\varepsilon_n < \varepsilon_e + \delta} |n\rangle \langle n| . \quad (33)
$$

We note the presence of discontinuities with respect to δ in $\Sigma_e^A(\delta)$ and $\Sigma_e^P(\delta)$ [see Eq. (11)], which must be treated carefully so that $\Sigma_e^c(\delta)$ remains continuous [54]. For the highest occupied molecular orbital (HOMO) of silane, choosing $f(\omega) = 1$ allows to converge $\Sigma_e^c(\langle e| \hat{\Sigma}^x - \hat{V}^{xc} |e\rangle)$ [55] to 1 meV with only eight frequency samplings of the integrand of Eq. (12) in contrast to 24 when $f(\omega)$ is chosen to be 0.

The only nontrivial model function used in previous works was of Gaussian form [39]. This choice yields

$$
f(\omega,\alpha) = e^{-\omega^2/\alpha^2},
$$
$$
F_{ne}(\delta,\alpha) = \frac{1}{2} \text{sgn}(\omega_{ne}) e^{\omega_{ne}^2/\alpha^2} \text{erfc}(|\omega_{ne}/\alpha|),
$$
$$
\Sigma_e^A(\delta,\alpha) = \frac{1}{2} \sum_{l,l'} \langle l| (\hat{\epsilon}^{-1}(0) - \hat{1}) |l'\rangle
$$
$$
\times \langle l^*| \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger \text{sgn}(\hat{H} - \varepsilon_e - \delta) e^{(\hat{H} - \varepsilon_e - \delta)^2/\alpha^2}
$$
$$
\times \text{erfc}(|(\hat{H} - \varepsilon_e - \delta)/\alpha|) \hat{\Phi}_e \hat{v}_C^{1/2} |l'^*\rangle ,
$$

where α 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,

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

$$F_{ne}(\delta,\alpha) = \frac{\pi}{2} \frac{\alpha}{\omega_{ne} + \alpha \, \text{sgn}(\omega_{ne})},$$

$$\Sigma_e^A(\delta,\alpha) = \frac{1}{2} \sum_{l,l'} \langle l| \, (\hat{\epsilon}^{-1}(0) - \hat{1}) \, |l'\rangle$$

$$\times \left(\langle l^*| \, \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger \hat{\mathcal{Q}}_e(\delta) \, |h_{e,l'}^+(\delta,\alpha)\rangle \right.$$

$$\left. + \langle l^*| \, \hat{v}_C^{1/2} \hat{\Phi}_e^\dagger \hat{\mathcal{P}}_e(\delta) \, |h_{e,l'}^-(\delta,\alpha)\rangle \right), \quad (34)$$

where

$$(\hat{H} - \varepsilon_e - \delta + \alpha) \, |h_{e,l'}^+(\delta,\alpha)\rangle \equiv \alpha \hat{\mathcal{Q}}_e(\delta) \hat{\Phi}_e \hat{v}_C^{1/2} \, |l'^*\rangle \,,$$

$$(\hat{H} - \varepsilon_e - \delta - \alpha) \, |h_{e,l'}^-(\delta,\alpha)\rangle \equiv \alpha \hat{\mathcal{P}}_e(\delta) \hat{\Phi}_e \hat{v}_C^{1/2} \, |l'^*\rangle \,, \quad (35)$$

and where no complicated functions of $\hat{H}$ are involved. Choosing $\alpha = 1.0$ Ha in this form of model function allows to converge $\Sigma_e^c(\langle e| \, \Sigma^x - V^{xc} \, |e\rangle)$ for the highest occupied molecular orbital (HOMO) of silane to 1 meV with only four frequency samplings of the integrand of Eq. (12) in contrast to eight when $f(\omega) = 1$.

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

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

$$= \frac{\alpha}{2} \left(\frac{1}{\alpha + i\omega} + \frac{1}{\alpha - i\omega} \right) (\hat{\epsilon}^{-1}(0) - 1), \quad (36)$$

approximating the dynamical character of the inverse dielectric matrix $\hat{\epsilon}^{-1}(i\omega) - 1$ 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,

$$\lim_{\omega \to \infty} \frac{\alpha^2}{\omega^2 + \alpha^2} \propto \frac{1}{\omega^2} \propto \lim_{\omega \to \infty} \hat{\epsilon}^{-1}(i\omega) - 1, \quad (37)$$

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 $\{|l\rangle\}$ FOR $\hat{\epsilon}^{-1}(i\omega)$

Eliminating the sums over conduction states that were present in Eqs. (12)–(15) required us to introduce a complete basis $\{|l\rangle\}$ in the terms $\sigma_e^N(i\omega,\delta)$, $\sigma_e^{N0}(i\omega,\delta,\alpha)$, and $\Sigma_e^A(\delta,\alpha)$. These terms can be rewritten in the form of a trace in the basis $\{|l\rangle\}$, starting from Eqs. (29), (30), (31), (34), and (35):

$$\sigma_e^N(i\omega,\delta) = \sum_l \langle l| \, (\hat{\epsilon}^{-1}(i\omega) - \hat{1}) \hat{v}_C^{1/2} \hat{\Phi}_e \frac{\hat{H}^* - \varepsilon_e - \delta}{\omega^2 + (\hat{H}^* - \varepsilon_e - \delta)^2} \hat{\Phi}_e^\dagger \hat{v}_C^{1/2} \, |l\rangle \equiv \sum_l \langle l| \, \hat{\sigma}_e^N(i\omega,\delta) \, |l\rangle \,,$$

$$\sigma_e^{N0}(i\omega,\delta,\alpha) = \sum_l \langle l| \, (\hat{\epsilon}^{-1}(0) - \hat{1}) \hat{v}_C^{1/2} \hat{\Phi}_e \frac{\hat{H}^* - \varepsilon_e - \delta}{\omega^2 + (\hat{H}^* - \varepsilon_e - \delta)^2} \hat{\Phi}_e^\dagger \hat{v}_C^{1/2} \, |l\rangle \frac{\alpha^2}{\omega^2 + \alpha^2} \equiv \sum_l \langle l| \, \hat{\sigma}_e^{N0}(i\omega,\delta) \, |l\rangle \frac{\alpha^2}{\omega^2 + \alpha^2},$$

$$\Sigma_e^A(\delta,\alpha) = \frac{1}{2} \sum_l \langle l| \, (\hat{\epsilon}^{-1}(0) - \hat{1}) \hat{v}_C^{1/2} \hat{\Phi}_e \hat{\mathcal{Q}}_e^* \frac{\alpha}{\hat{H}^* - \varepsilon_e - \delta + \alpha} \hat{\mathcal{Q}}_e^* \hat{\Phi}_e^\dagger \hat{v}_C^{1/2} \, |l\rangle$$

$$+ \frac{1}{2} \sum_l \langle l| \, (\hat{\epsilon}^{-1}(0) - \hat{1}) \hat{v}_C^{1/2} \hat{\Phi}_e \hat{\mathcal{P}}_e^* \frac{\alpha}{\hat{H}^* - \varepsilon_e - \delta - \alpha} \hat{\mathcal{P}}_e^* \hat{\Phi}_e^\dagger \hat{v}_C^{1/2} \, |l\rangle \equiv \sum_l \langle l| \, \hat{\Sigma}_e^A(\delta,\alpha) \, |l\rangle \,. \quad (38)$$

The smallest orthonormal basis $\{|l\rangle\}$ such that the above traces are converged must contain the subspace associated with the highest eigenvalues of $\hat{\sigma}_e^N(i\omega,\delta)$, $\hat{\sigma}_e^{N0}(i\omega,\delta)$, and $\hat{\Sigma}_e^A(\delta,\alpha)$. This subspace corresponds qualitatively to the intersection of the subspaces associated with the highest eigenvalues (in absolute value) of $\hat{v}_C^{1/2}$, $\hat{\Phi}_e$, $(\hat{H}^* - \varepsilon_e - \delta)^{-1}$, and $\hat{\epsilon}^{-1}(i\omega) - \hat{1}$. Since $\hat{v}_C^{1/2}$ is diagonal in the basis of plane waves, $\hat{\Phi}_e$ is diagonal in real space and $(\hat{H}^* - \varepsilon_e - \delta)^{-1}$ 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-waves implementations of $G_0 W_0$, the dielectric matrix $\hat{\epsilon}$ is obtained in a plane-wave basis [using Eqs. (18) and (17)] and directly inverted. This inversion then

becomes a bottleneck, since the size of the basis in which wave functions are expanded is quite large ($\sim$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 wave functions [8]. However, this practice leads to increased uncertainties [10] 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

$$\hat{\epsilon}(i\omega) - \hat{1} = 4 \sum_{cv} \hat{v}_C^{1/2} \, |c^*v\rangle \frac{\varepsilon_c - \varepsilon_v}{(\varepsilon_c - \varepsilon_v)^2 + \omega^2} \langle vc^*| \, \hat{v}_C^{1/2}, \quad (39)$$

where $(\varepsilon_c - \varepsilon_v)/((\varepsilon_c - \varepsilon_v)^2 + \omega^2) > 0 \ \forall \ \omega \in \mathbb{R}$ and where $\hat{v}_C^{1/2}$ has only positives eigenvalues. It results that $\hat{\epsilon}(i\omega) - \hat{1}$ has only positives 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 [31].

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, the corresponding eigenvalues will occupy the same rank. Consequently, the eigenvalues of $1 - \hat{\epsilon}^{-1}$ that contribute most to $\Sigma_e^c(\delta)$ 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 [30]:

$$\{|\psi\rangle, (\hat{\epsilon} - \hat{1})|\psi\rangle, (\hat{\epsilon} - \hat{1})^2|\psi\rangle, \ldots, (\hat{\epsilon} - \hat{1})^N|\psi\rangle\}. \quad (40)$$

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 $|\psi\rangle = \hat{v}_C^{1/2}|e^*e\rangle$ as a starting point, but we observed that using any other vector (or all other vectors) of the form $|\psi\rangle = \hat{v}_C^{1/2}|v^*e\rangle$ does not change significantly the number of dimensions N required to achieve convergence. This also implies that the basis generated from $|\psi\rangle = \hat{v}_C^{1/2}|e^*e\rangle$ can be used to obtain G_0W_0 corrections for all desired DFT eigenstates and not only $|e\rangle$.

Also, in practice, we use the Lanczos procedure [30] 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 loose their orthogonality with the number of steps due to numerical error [30]. For example, in our implementation, orthogonality is typically lost in about ten steps, while 100's of steps are required to achieve convergence. Therefore, in our implementation of the Lanczos procedure, we added a Gram-Schmidt orthogonalization 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 $\Sigma_e^c(\delta)$ 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 then 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 $\Sigma_e^c(\delta)$. An alternative way to obtain $\hat{\epsilon}$ in approximately the same subspace is to iteratively diagonalize it [31,32], as was implemented in another plane waves G_0W_0 code [15]. However, the latter scheme costs about ten 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 Kaige-Pangel theory [30], 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 $\Sigma_e^c(\delta)$. 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 $\{|l\rangle\}$ 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 $\hat{v}_C^{1/2}$, $\hat{\Phi}_e$ and $(\hat{H}^* - \varepsilon_e - \delta)^{-1}$. However, the cost of calculating the traces in Eq. (38) is of order $N^2 \ln(N)$ 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 $\{|l\rangle\}$ just increases the precision of the calculated trace. Therefore, in plane waves implementations, the most efficient choice of basis $\{|l\rangle\}$ 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 $\Sigma_e^c(\delta)$ 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.

V. DIELECTRIC MODEL

We mentioned in Sec. IV that, when the Lanczos basis $\{|l\rangle\}$ of the dielectric operator $\hat{\epsilon} - \hat{1}$ is iteratively constructed, the fastest converging eigenvalues are the largest ones. Therefore, as the construction of $\{|l\rangle\}$ 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 Sec. 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 $\{|l\rangle\}$, when applying $\hat{\epsilon} - \hat{1}$ to the current basis vector $|l\rangle$, the largest terms in the sum of Eq. (39) will be associated to eigenenergies ε_c of free conduction states, dominated by kinetic energy. We will then have $\hat{H}|c\rangle = \varepsilon_c |c\rangle \approx \hat{T}|c\rangle$, where $\hat{T}$ is the kinetic energy operator. Therefore substituting $\hat{H} \rightarrow \hat{T}$ in Eqs. (21)–(23) should become an accurate approximation at some point in the construction of $\{|l\rangle\}$.

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 wave functions are expressed. In contrast, the solution of Eqs. (22) or (23) using SQMR requires typically ~ 15 application of the Hamiltonian $\hat{H}$ on a vector at a cost $\propto N \ln(N)$. Since these successive applications of $\hat{H}$ are the bottleneck in the construction of $\{|l\rangle\}$, 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 $\hat{H} \rightarrow \hat{T}$:

$$\hat{\tilde{\epsilon}}(\omega) \equiv 1 - \hat{v}_C^{1/2} \hat{\tilde{P}}(\omega) \hat{v}_C^{1/2},$$

$$\hat{\tilde{P}}(\omega)|\psi_j\rangle \equiv -2 \sum_v |v\rangle \left(|\tilde{f}_{jv-}^*(\omega)\rangle + |\tilde{f}_{jv+}^*(\omega)\rangle \right), \quad (41)$$

$$|\tilde{f}_{jv\pm}(\omega)\rangle \equiv \hat{\mathcal{P}}_c \sum_G |G\rangle \frac{1}{G^2/2 - \varepsilon_v \pm \omega} \langle G| \hat{\mathcal{P}}_c |v\psi_j^*\rangle,$$

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

$$\hat{\epsilon}^{-1} - 1 = (\hat{\epsilon}^{-1} - \hat{\tilde{\epsilon}}^{-1}) + (\hat{\tilde{\epsilon}}^{-1} - 1), \quad (42)$$

in σ_e^N, σ_e^{N0}, and Σ_e^A as expressed in Eq. (38). Since $\hat{\tilde{\epsilon}}^{-1}$ should accurately describes the integrable divergence in the spectrum of eigenvalues of $\hat{\epsilon}^{-1}$ near 1, the operator $(\hat{\epsilon}^{-1} - \hat{\tilde{\epsilon}}^{-1})$ 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 $(\hat{\tilde{\epsilon}}^{-1} - 1)$ 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 $(\hat{\epsilon}^{-1} - \hat{\tilde{\epsilon}}^{-1})$ and can thus use a smaller Lanczos basis. In the second traces, we substitute $(\hat{\epsilon}^{-1} - 1)$ by $(\hat{\tilde{\epsilon}}^{-1} - 1)$ and the basis size will remain similar. Thus we use the exact dielectric operator $\hat{\epsilon} - 1$ to generate the first basis $\{|l\rangle\}$ at a reduced cost, thanks to its smaller size. Then, we use the approximate dielectric operator $\hat{\tilde{\epsilon}} - 1$ to generate the second basis $\{|\tilde{l}\rangle\}$, also at a reduced cost, since the operator is simpler to apply.

Once those bases are available, we split each of σ_e^N, σ_e^{N0}, and Σ_e^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 $(\hat{\epsilon}^{-1} - \hat{\tilde{\epsilon}}^{-1})$ and evaluate the resulting expression using the basis $\{|l\rangle\}$. In the second contributions, we substitute $(\hat{\epsilon}^{-1} - 1)$ by $(\hat{\tilde{\epsilon}}^{-1} - 1)$ and evaluate the resulting expression

using the basis $\{|\tilde{l}\rangle\}$. Then, we sum the two results and obtain σ_e^N, σ_e^{N0}, and Σ_e^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 $\{|\tilde{l}\rangle\}$ are sufficient to obtain converged results.

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

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

$$P(i\omega) = -4 \sum_{cv} |c^*v\rangle \frac{\varepsilon_c - \varepsilon_v}{\omega^2 + (\varepsilon_c - \varepsilon_v)^2} \langle vc^*|. \quad (43)$$

When $i\omega$ is displaced along the positive direction of the imaginary axis, each term of the sum decreases monotonically. In contrast, making $i\omega$ 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 $\hat{\epsilon}(i\omega)$ at all other frequencies $i\omega \in [0, i\infty]$ is a sound approximation, in agreement with previous work [15,36].

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 $\Sigma_e^c(\delta)$ as expressed in Eqs. (12) and (38), only a single matrix element of $\hat{\epsilon}^{-1} - \hat{1}$ 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 $\langle ep^*| \hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(\omega_{pe}) - \hat{1}) \hat{v}_C^{1/2} |p^*e\rangle$, where $|p\rangle$ 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 $\hat{v}_C^{1/2} |p^*e\rangle$ is required. The latter will automatically be satisfied if the seed vector is chosen to be $\hat{v}_C^{1/2} |p^*e\rangle$ and the former is a feature of the Lanczos procedure. Thus, with this choice of seed vector, we found that four Lanczos iterations converge $\Sigma_e^P(\delta)$ 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 $\Sigma_e^P(\delta)$ small with respect to the remainder of $\Sigma_e^c(\delta)$. 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 $i\omega$ enables some tricks to speed up the calculation. For instance, we adopt the shift Lanczos technique to solve Eq. (30) simultaneously at all frequencies [15,56] (a few tens of iterations is typically enough to converge Σ_e^c to 5 meV). This technique requires that the linear equation have the general form $(\hat{M} + \hat{I}\omega)|x(\omega)\rangle = |b\rangle$ where both the operator $\hat{M}$ and the right-hand side $|b\rangle$ are independent of ω. However, here $|b\rangle = (\hat{H} - \varepsilon_e - \delta)\hat{\Phi}_e\hat{v}_C^{1/2}|l^*\rangle$. Therefore, if $\{|l\rangle\}$ was dependent on ω, Eq. (30) would have to be solved individually for each frequency ω. 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 $|\psi_j\rangle \rightarrow \hat{v}_C^{1/2}|l\rangle$ 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 matrix [15,56]. 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, $\{|\gamma_{i,v}\rangle\}$, 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 $\{|\gamma_{i,v}\rangle\}$, it is useful to rewrite the quantity we wish to calculate with the solution of Eq. (23), that is, $\langle l| \hat{\epsilon}(i\omega) - \hat{1} |l'\rangle$, from Eqs. (17), (19), (23), and (24):

$$\langle l| \hat{\epsilon}(i\omega) - \hat{1} |l'\rangle = 4 \sum_v \langle l| \hat{v}_C^{1/2}\hat{\Phi}_v\hat{\mathcal{P}}_c^*$$

$$\times \frac{\hat{H}^* - \varepsilon_v}{(\hat{H}^* - \varepsilon_v)^2 + \omega^2} \hat{\mathcal{P}}_c^* \hat{\Phi}_v^\dagger \hat{v}_C^{1/2} |l'\rangle$$

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

$$= 4 \sum_v \langle b_{l,v}|x_{l',v}(i\omega)\rangle , \qquad (44)$$

where $|b_{l,v}\rangle \equiv \hat{\mathcal{P}}_c^*\hat{\Phi}_v^\dagger\hat{v}_C^{1/2}|l\rangle$, $\hat{A}_v(i\omega) \equiv ((\hat{H}^* - \varepsilon_v)^2 + \omega^2)/(\hat{H}^* - \varepsilon_v)$, 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 $|\lambda_v(i\omega)\rangle$ of $\hat{A}_v(i\omega)$ that contribute most to the desired quantity ($\langle l| \hat{\epsilon}(i\omega) - \hat{1} |l'\rangle = 4 \sum_{v,\lambda} \langle b_{l,v}|\lambda_v(i\omega)\rangle 1/\lambda_v(i\omega) \langle \lambda_v(i\omega)|b_{l,v}\rangle$), that is, those that are both associated with small eigenvalues $\lambda_v(i\omega)$ of $\hat{A}_v(i\omega)$ and overlapping substantially with the states $|b_{l,v}\rangle$. The solutions $|x_{l,v}(i\omega)\rangle = \sum_\lambda |\lambda_v(i\omega)\rangle 1/\lambda_v(i\omega) \langle \lambda_v(i\omega)|b_{l,v}\rangle$ are naturally dominated by these directions, 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 $i\omega$ 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 [$\hat{A}_v(0) = (\hat{H}^* - \varepsilon_v)$] and monotonically evolves toward the inverse of this operator as the frequency increases [$\lim_{i\omega\to+i\infty} \hat{A}_v(i\omega) \rightarrow (\hat{H}^* - \varepsilon_v)^{-1} \times \omega^2$]. The two most natural frequencies to select would therefore be 0 and $i\infty$. Since $\lim_{i\omega\to+i\infty} |x_{l,v}(i\omega)\rangle \rightarrow (\hat{H}^* - \varepsilon_v) \times \omega^{-2} |b_{l,v}\rangle$, 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 [30] and orthonormalization. Finally, $H_{ij,v} = \langle \gamma_{i,v}| (\hat{H} - \varepsilon_v) |\gamma_{j,v}\rangle$ and $b_{i,l,v} = 2 \langle \gamma_{i,v}| \hat{\mathcal{P}}_c\hat{\Phi}_v\hat{v}_C^{1/2} |l^*\rangle$ 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 $|x_{l,v}(i\omega_0)\rangle$ to those we have already selected and observing that the result is negligibly affected. We have selected for ω_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 $\Sigma_e^c(\delta)$ was affected by about 7 meV. Further testing with values of 0.1 Ha and 0.01 Ha for ω_0 has shown similar behavior.

VII. THEORETICAL ANALYSIS OF COMPUTATIONAL COST

Provided that DFT eigenstates $|n\rangle$ and eigenvalues ε_n have already been calculated, our implementation starts with the construction of the Lanczos bases $\{|l\rangle\}$ and $\{|\tilde{l}\rangle\}$ as described in Secs. 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 $\hat{\tilde{\epsilon}}(0)$ on the $N_{\tilde{L}}$ Lanczos vectors $|\tilde{l}\rangle$ as described by Eq. (41). By analyzing 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 $\hat{H} - \varepsilon_v$ on a vector and the same amount of orthogonalization with the valence wave functions. Applying the Hamiltonian costs $N_{FFT} \ln(N_{FFT})$ operations in a plane-waves implementation, where N_{FFT} is the number of components describing a wave function in real space, while the orthogonalization costs $N_v N_{PW}$, where N_{PW} is the number of components describing a wave function in reciprocal space. Thus we conclude that the worst scaling in this bottleneck comes from the orthogonalization and is $\propto N_L N_v N_{SQMR} N_v N_{PW}$, which is $\propto N^4$, since N_{SQMR} is independent of the system size. Similarly, the worst scaling in the construction of $\{|\tilde{l}\rangle\}$ is $\propto N_{\tilde{L}} N_v N_v N_{PW} \propto N^4$.

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 [5]. 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 $\sigma_e^N(i\omega,\delta) - \sigma_e^{N0}(i\omega,\delta)$, 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 [see Eq. (23)], as described in Sec. VI. This involves N_v singular value decompositions that scale as $N_L^2 N_{PW}$, 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 $N_v N_L^2 N_{PW} \propto N^4$. Finally, the solution of Eq. (23) by direct inversion of $(\hat{H} - \varepsilon_v)^2 - (i\omega)^2$ in these bases scales as $N_\omega N_v N_L^3$. Since N_ω is the number of frequencies used for the numerical integration and is not dependent on system size, we obtain $N_\omega N_v N_L^3 \propto N^4$. All other operations in this section of the code scale, at worst, as $N^3 \ln(N)$. Next, our implementation calculate $\Sigma_e^A(\delta)$ 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 $\Sigma_e^P(\delta)$ using Eqs. (11), (17), (19), and (22) as well as the Lanczos algorithm described in Sec. IV. However, again, this part of the code scales as N^3 and is not a bottleneck of the calculation.

VIII. RELATION WITH EXISTING IMPLEMENTATIONS

As mentioned in the introduction, the conversion of summations over conduction states into Sternheimer equations is a well established technique both in density functional perturbation theory [48,49] and G_0W_0. It was first applied to the latter more than 15 years ago by Ref. [16] and has been successfully used in other implementations [13,15,17] since then. In all cases, the sums over conduction states present in Eqs. (3) and (18) are converted into linear equations analogous to Eqs. (23) and (30). It is also common to solve these equations at imaginary frequencies, to avoid the poles on the real axis [13,15,17]. However, the use of the SQMR algorithm [41] allows us to solve these equations at real frequencies, thus enabling the use of the contour deformation technique [33,34], in contrast to the aforementioned implementations, which either use analytic continuation [40] (Refs. [13,15,17]) or plasmon pole models [8] (Ref. [16]). Our implementation is thus distinctively devoid of approximations on the frequency dependence of the self-energy or the dielectric matrix.

It is also common to solve the Sternheimer equations simultaneously at all relevant frequencies using a multishift linear equation solver, either Lanczos-based [57,58] (present work, Refs. [13,15]), Frommer's multishift solver [59] (Ref. [17]), or the Taylor expansion of $(\hat{H} - \omega)^{-1}$ (Ref. [16]). However, our strategy to reduce the computational cost of the solution of Eq. (23) below the level of a multishift linear equation solver, as presented in the second half of Sec. VI, has not previously been described in the literature to our knowledge. Indeed, while Ref. [13] uses a singular value decomposition (SVD)-like procedure on the $\{|b_{l,v}\rangle\}$ vectors of Eq. (44) to reduce the number of solutions $|x_{l,v}(\omega)\rangle$ to be obtained using multishift Lanczos, our implementation improves on this procedure by performing the SVD on the (already available) solutions $\{|x_{l,v}(0)\rangle, |x_{l,v}(\infty)\rangle\}$ and directly solving Eq. (44) in the resulting basis. This divides the number of applications of the Hamiltonian involved in the procedure by, approximately, the number of multishift Lanczos steps.

The expression of the dielectric matrix in a Lanczos basis is the core improvement proposed in this article to

the G_0W_0 method. Indeed, it should outperform the iterative diagonalization of the dielectric matrix used in Refs. [13,15], since the latter requires $\sim$10 applications of the dielectric matrix per dimension, while our method requires a single application per dimension, with a comparable number of dimensions for both methods at a given convergence criterion (see Sec. IV). It also outperforms the self-consistent evaluation of the dielectric matrix (under the form of the screen coulomb potential) proposed in Ref. [17], since the latter does not outperform the calculation of the dielectric matrix in a plane-wave basis at small system size, while our method does.

Finally, it is common to keep the basis of the dielectric matrix fixed with respect to imaginary frequencies, since it is prerequisite for the use of multishift linear equation solvers, as described earlier in this section as well as in Sec. VI. Indeed, this approximation has successfully been used in all implementations that generate an optimal basis to represent the dielectric matrix, that is, Refs. [13,15] and the present implementation.

IX. RESULTS

The ABINIT software package [25] is used to produce the DFT eigenstates and eigenvalues used as input in our G_0W_0 calculations. To ease reproducibility, we choose to simulate all molecules with their experimental geometry: silane [60], thiophene [61], benzene [61], naphthalene [61], anthracene [61], tetracene [61], and C_{60} [62,63]. We use LDA (Teter Pade parametrization [64]) and/or PBE [65] functional for DFT calculations, as specified in each result table. The corresponding Hartwigsen-Goedeker-Hutter LDA [66] or PBE [67] pseudopotentials are used throughout. The reference G_0W_0 implementation in Sec. X is the conventional G_0W_0 implementation present in ABINIT.

For all systems studied, values for the highest occupied molecular orbital (HOMO) are provided. However, for the lowest unoccupied molecular orbital (LUMO), we concentrate on molecules where the orbital is bound (anthracene, tetracene, and C_{60}). Indeed, for unbound cases, both G_0W_0 and experimental results are scarcer. Due to this diminished interest, and because plane-wave G_0W_0 calculations of molecules are cumbersome even with our implementation, we only provide results for the computationally simplest of these cases, the LUMO of silane.

Since this study focuses on isolated systems, a spherical truncation of the Coulomb potential [68] is used in both the current implementation and the reference implementation of G_0W_0. Its radius is set to half the size of the cubic unit cell used to simulate the molecule. Thus converging the G_0W_0 correction [see Eq. (1)] with respect to the size of the unit cell ensures both that the wave functions of a given molecule are coupled together over their full spatial extent and that the wave functions belonging to different periodic replicas of the molecule do not interact. However, the spherical truncation of the Coulomb potential is not implemented in the DFT part of ABINIT, so that this strategy cannot be applied to the DFT eigenenergies. Moreover, in periodic simulations, the latter are difficult to position with respect to vacuum. To simultaneously correct them 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 [69]. This code uses a localized basis set (Daubechies wavelets [70]) and a Poisson solver with free boundary conditions [71], which allow for nonperiodic simulations, free of spurious Coulombic interactions, and a physical determination of the energy level of the vacuum. A spacing of 0.4 bohr for the wavelet grid is used to converge the eigenenergies to a few meV. The difference between the eigenenergies obtained using BigDFT and ABINIT can then be added to the G_0W_0 eigenenergies to ensure that they are correctly positioned with respect to the vacuum and devoid of spurious Coulombic interactions, provided that the G_0W_0 correction is converged with respect to the unit cell size.

Also, Gauss-Legendre quadratures with eight points are used to integrate Eq. (12) (or Eq. (10) for the reference G_0W_0 implementation). DFT eigenvalues and eigenstates are converged until the squared residual of the wave functions $|(\hat{H} - \varepsilon_n)|n\rangle|^2$ is less than 10^{-12} Ha^2. The solutions to Eqs. (22) or (23) are iteratively refined until the squared residual $|\hat{A}|x\rangle - |b\rangle|^2$ is smaller than 10^{-20} Ha^2 or Ha^4, respectively. Equation (11) is calculated using four iterations of the Lanczos scheme described in Sec. VI. Equations (30) and (35) are solved with the shift-Lanczos method [13,57], where eight and 16 iterations are used, respectively. The parameter α of the Lorentzian model to the dielectric matrix [Eq. (34)] is kept fixed at 1.0 Ha. The cutoff energy is converged separately for the calculation of Σ_e^x and $\Sigma_e^c(\delta)$ 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 $\hat{\epsilon} - \hat{\tilde{\epsilon}}$ (as described in Sec. V) and the number of Lanczos vectors $N_{\tilde{L}}$ describing the approximate dielectric matrix $\hat{\tilde{\epsilon}} - 1$ are also selected so that $\Sigma_e^c(\delta)$ is converged to 50 meV, unless specified otherwise. When no value is given for $N_{\tilde{L}}$, the exact dielectric matrix $\hat{\epsilon} - 1$ is constructed using N_L Lanczos vectors and the dielectric model is not used. The converged values for the box size, the cutoff energy and the number of Lanczos vectors (N_L and $N_{\tilde{L}}$) are given for each molecule in Table I.

A. Silane

Results for silane are given in Table II. Good agreement is observed among the DFT results. Indeed, while our DFT results are converged to a few meV, the other DFT studies were converged to 0.1 eV, which is comparable to the agreement between the results. Similarly, good agreement is observed between the G_0W_0 results for the HOMO, except

TABLE I. Converged values for molecule dependent parameters. Σ_e^x cutoff and Σ_e^c cutoff are the cutoff energies used for Σ_e^x and $\Sigma_e^c(\delta)$, respectively. All quantities are given in atomic units.

molecule	box size	Σ_e^x cutoff	Σ_e^c cutoff	N_L	$N_{\tilde{L}}$
silane	50	20	20	512	–
thiophene	26	40	40	512	2048
benzene	28	30	20	240	1920
naphthalene	30	30	20	384	3072
anthracene	40	30	20	528	3168
tetracene	51	30	20	504	4032
C$_{60}$	40	40	20	1024	4096

TABLE II. Comparison between the present implementation and previously published G_0W_0 results for the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) energies of the silane molecule. The underlying DFT energies are given to assess the agreement of the starting point of the G_0W_0 calculations. All results are in eV.

	HOMO		LUMO	
	LDA	G_0W_0	LDA	G_0W_0
Ref. [72]	−8.4	−12.7	–	–
Ref. [73]	−8.4	−12.7	−0.6	0.3
Ref. [74]	−8.42	−12.41	−0.50	0.50
Ref. [75]	–	−12.43	–	–
This work	−8.51	−12.43	−0.53	0.79
	Exp.		Exp.	
Ref. [76]	−12.3		–	
Ref. [77]	−12.36		–	
	PBE	G_0W_0	PBE	G_0W_0
Ref. [78]	−8.47	−12.11	–	–
Ref. [75]	–	−12.40	–	–
Ref. [79]	–	−12.29	–	–
Ref. [80]^a	–	−12.31	–	2.51
Ref. [80]^b	–	−12.31	–	2.51
This work	−8.51	−12.33	−0.51	0.77

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

for Ref. [72,73]. Moreover, the comparison with experimental HOMO energies is also good. These results thus support the accuracy of our implementation.

Interestingly, G_0W_0 results for the LUMO do not show such an agreement. Given the scatter of available G_0W_0 results and the absence of experiments, it becomes hard to gather information on the validity of our implementation from silane LUMO energies.

B. Thiophene

The results for thiophene are given in Table III. Again, the difference between DFT results is reasonable, and gives an estimate of the expected agreement between G_0W_0 results, given the differences in boundary conditions, geometry, pseudopotentials, and convergence criteria between studies. In that respect, the agreement between G_0W_0 results is good excepted for Ref. [23]. However, Ref. [81] repeated the calculations of Ref. [23] with larger localized basis sets and obtained a G_0W_0 result in good agreement with ours along with a slow convergence rate, as explained in the footnote of Table III. Together, these observations resolve this discrepancy. Finally, the agreement with experiments is good.

C. Benzene

Results for benzene are given in Table IV. Again, the agreement between DFT results is good. We also find good agreement between G_0W_0 results (except for Ref. [85]), which supports the accuracy of our implementation. Moreover, our agreement with the experiment is also good.

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

	HOMO	
	LDA	G_0W_0
Ref. [23]	−6.15	−8.37
Ref. [81][a]	−6.06	−8.55
	−6.06	−8.69
This work	−6.04	−8.93
	Exp.	
Ref. [82]	−8.85	
Ref. [83]	−8.86 ± 0.02	
	PBE	G_0W_0
Ref. [84]	−	−9.0
Ref. [15]	−5.70	−8.49
This work	−5.86	−8.73

[a]Reference [81] 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. [23], but with the same Gaussian-basis contour-deformation methodology, and finds a G_0W_0 ionization energy for thiophene of 8.55 and 8.69 eV, respectively, with a starting LDA value of 6.06 eV with both basis. They conclude that these results indicate a slow convergence with respect to the size of the Gaussian basis in the case of small molecules with unbound virtual states.

D. Naphthalene

Results for naphthalene are given in Table V. In this case, the agreement with the DFT result of Ref. [78] is excellent.

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

	HOMO	
	LDA	G_0W_0
Ref. [85]	−	−9.88
Ref. [86]	−6.67	−8.78
Ref. [13]	−	−9.40
Ref. [79]	−	−9.05
Ref. [18]	−6.49	−9.03
Ref. [15]	−	−9.22
This work	−6.50	−9.23
	Exp.	
Ref. [83]	−9.24378 ± 0.00007	
	PBE	G_0W_0
Ref. [84]	−	−9.4
Ref. [78]	−6.39	−8.87
Ref. [80][a]	−	−8.99
Ref. [80][b]	−	−8.97
Ref. [79][c]	−	−9.00
Ref. [15]	−6.18	−9.04
This work	−6.31	−9.03

[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 resolution-of-identity approximation.
[c]Result obtained with the numeric atom-centered orbitals set 'tier 4 + a5Z-*d*'.

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

	HOMO	
	LDA	G_0W_0
Ref. [85]	−	−8.69
Ref. [86]	−	−7.67
This work	−5.67	−8.05
	Exp.	
Ref. [83]	−8.144 ± 0.001	
	PBE	G_0W_0
Ref. [78]	−5.50	−7.73
This work	−5.48	−7.84

For the G_0W_0 results, the agreement with Ref. [78] is also very good. The other G_0W_0 results show reasonable but somewhat lesser agreement. However, our results are in good agreement with the experiment. This, together with the agreement with Ref. [78], supports the accuracy of our implementation.

E. Anthracene

Results for anthracene are given in Table VI. Both at the DFT and the G_0W_0 level, we observe good agreements with previously published results for the HOMO energies. Interestingly, we observe better agreement with Refs. [15,78] than with Ref. [23], similarly to the case of thiophene, and Ref. [86], similarly to the case of benzene. We also observe good agreement between our HOMO G_0W_0 results and the experiment.

For LUMO DFT results, we also observe a good agreement with previously published results. Still, the agreement between the LUMO G_0W_0 results is slightly diminished with respect to the HOMO, but this is due to the larger scatter of published data. Finally, the agreement with experiments is not very good for LUMO energies. This is relatively unsurprising, since the accuracy of G_0W_0 results depends on the quality of the DFT starting point and LDA/PBE can produce particularly poor results for orbitals close to the vacuum level.

TABLE VI. Results (in eV) for the HOMO and LUMO energies of the anthracene molecule.

	HOMO		LUMO	
	LDA	G_0W_0	LDA	G_0W_0
Ref. [23]	−5.47	−6.89	−3.22	−0.74
Ref. [86]	−	−6.89	−	−0.77
Ref. [15]	−5.18	−7.25	−2.81	−1.05
This work	−5.18	−7.31	−2.89	−1.24
	Exp.		Exp.	
Ref. [83]	−7.439 ± 0.006		−0.66 to −0.42	
	PBE	G_0W_0	PBE	G_0W_0
Ref. [78]	−4.96	−7.12	−2.68	−1.17
This work	−4.98	−7.09	−2.67	−1.01

TABLE VII. Results (in eV) for the HOMO and LUMO energies of the tetracene molecule.

	HOMO		LUMO	
	LDA	G_0W_0	LDA	G_0W_0
Ref. [23]	−5.15	−6.37	−3.58	−1.34
Ref. [15]	−4.85	−7.04	−3.19	−1.41
This work	−4.86	−6.79	−3.26	−1.80
	Exp.		Exp.	
Ref. [83]	−6.97 ± 0.05		−1.06 to −0.88	
	PBE	G_0W_0	PBE	G_0W_0
Ref. [78]	−4.65	−6.70	−3.05	−1.84
This work	−4.66	−6.57	−3.04	−1.56

F. Tetracene

Results for tetracene are given in Table VII. Again, we observe our results to be in good agreement with previously published results at the DFT level. At the G_0W_0 level, for HOMO energies, the agreement with Ref. [78] is good, as in the cases of naphthalene and anthracene. Also, the agreement with Ref. [23] 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 III). However, the agreement with Ref. [15] is somewhat lesser than expected. Given that our agreement with Ref. [15] 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. Moreover, we observe reasonable agreement between our HOMO G_0W_0 results and the experiment.

Available G_0W_0 data is more scattered for the LUMO than the HOMO, but the agreement remain otherwise similar, i.e., we agree best with Ref. [78], then Ref. [15], and finally Ref. [23]. Also, due to the poor description of the weakly bound LUMO orbital by LDA/PBE, the LUMO G_0W_0 results are not in very good agreement with experiment.

G. C$_{60}$

Results for C$_{60}$ are given in Table VIII. All DFT and G_0W_0 results show good agreement. We also observe a reasonable agreement with experiment.

TABLE VIII. Results (in eV) for the HOMO and LUMO energies of the **C$_{60}$** molecule.

	HOMO		LUMO	
	PBE	G_0W_0	PBE	G_0W_0
Ref. [18]	−5.84	−7.21	−4.19	−2.62
Ref. [15]	−5.81	−7.31	−4.13	−2.74
This work	−5.84	−7.41	−4.15	−2.92
	Exp.		Exp.	
Ref. [87]	−7.6 ± 0.2		–	
Ref. [88]	–		−2.684 ± 0.007	

TABLE IX. Contributions to G_0W_0 eigenenergies $\varepsilon_e^{G_0W_0}$. The DFT eigenenergy obtained using BIGDFT is noted $\varepsilon_e^{\text{DFT}}$, while V_e^{xc}, Σ_e^{x}, and Σ_e^{c} stand for the expectation value of the corresponding operator in the state considered $|e\rangle$. Σ_e^{c} is evaluated at the energy solving Eq. (1) when the ABINIT DFT eigenenergy is used.

molecule	orbital	funct.	$\varepsilon_e^{\text{DFT}}$	V_e^{xc}	Σ_e^{x}	Σ_e^{c}	$\varepsilon_e^{G_0W_0}$
sil.	HOMO	LDA	−8.51	−10.98	−15.61	0.71	−12.43
		PBE	−8.51	−11.29	−15.82	0.71	−12.33
	LUMO	LDA	−0.53	−2.90	−0.92	−0.67	0.79
		PBE	−0.51	−2.68	−0.81	−0.59	0.77
thio.	HOMO	LDA	−6.04	−13.03	−15.75	−0.17	−8.93
		PBE	−5.86	−13.09	−15.79	−0.16	−8.73
benz.	HOMO	LDA	−6.50	−12.97	−15.54	−0.16	−9.23
		PBE	−6.31	−13.01	−15.58	−0.15	−9.03
napht.	HOMO	LDA	−5.67	−13.13	−15.14	−0.36	−8.05
		PBE	−5.48	−13.17	−15.18	−0.35	−7.84
anthr.	HOMO	LDA	−5.18	−13.23	−14.84	−0.52	−7.31
		PBE	−4.98	−13.26	−14.86	−0.51	−7.09
	LUMO	LDA	−2.89	−13.03	−8.80	−2.59	−1.24
		PBE	−2.67	−13.05	−8.81	−2.58	−1.01
tetr.	HOMO	LDA	−4.86	−13.31	−14.62	−0.61	−6.79
		PBE	−4.66	−13.36	−14.66	−0.60	−6.57
	LUMO	LDA	−3.26	−13.08	−9.12	−2.49	−1.80
		PBE	−3.04	−13.11	−9.15	−2.48	−1.56
C$_{60}$	HOMO	PBE	−5.84	−13.85	−15.23	−0.18	−7.41
	LUMO	PBE	−4.15	−13.56	−9.81	−2.52	−2.92

H. Analysis

Globally, our G_0W_0 results for the HOMO of all molecules as well as the LUMO of anthracene, tetracene, and C$_{60}$ support the validity and accuracy of the present implementation. The G_0W_0 results for the LUMO of silane are, however, harder to analyze, due to the large scatter in available data. Still, this latter point has no consequence in the assessment of our implementation and, therefore, this study clearly support the implementation's accuracy.

To enable further analysis, the decomposition of G_0W_0 eigenenergies into their different contributions is given in Table IX. It is interesting to note that changing the functional has little effect on the value of Σ_e^{c} for all molecules studied. Thus the choice of functional affects $\varepsilon_e^{G_0W_0}$ through the computationally simpler contributions $\varepsilon_e^{\text{DFT}}$, V_e^{xc}, and Σ_e^{x}.

Also, for molecules containing only the first row elements, changing LDA for PBE had the same effect on all the DFT and G_0W_0 eigenenergies: it raised them by about 0.2 eV. Therefore the G_0W_0 correction $\Delta\varepsilon_e = \langle e | \hat{\Sigma}(\varepsilon_e + \Delta\varepsilon_e) - \hat{V}^{\text{xc}} | e \rangle$ 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 π orbitals), the full range of molecules to which this trend applies is difficult to guess.

X. PERFORMANCE

Since the scaling of our implementation is the same as the traditional G_0W_0 implementation [8] ($\propto N^4$ 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 $\Sigma_e^c(0)$ [see Eq. (7)] with both our implementation and the traditional G_0W_0 implementation found in the ABINIT project [25]. To allow a direct comparison of the computation times, we keep all parameters to their converged value, except for the unit cell size, the cutoff energy, the number of Lanczos vectors N_L and the number of plane waves used to describe the dielectric matrix $N_{PW\epsilon}$. The unit cell size and the cutoff energy are common to both calculations and are kept to an underconverged value of 18 bohr and 4 Ha, respectively, to allow the conventional G_0W_0 calculations to fit on the memory available on our computers. Since the number of Lanczos vectors N_L and the number of plane waves used to describe the dielectric matrix $N_{PW\epsilon}$ are analogous parameters, i.e., they control the precision of the dielectric operator in the present and the conventional G_0W_0 implementation, respectively, they are chosen as convergence parameters. No dielectric model was used in the present implementation, so that the dimension of the dielectric matrix remains controlled by one unique parameter. The ratio of the prefactors is then obtained from the ratio of the computation times required for both implementations to reach a given level of convergence. Here, a level of convergence of ± 50 meV is chosen, since it is the typical expected accuracy of G_0W_0 calculations.

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.
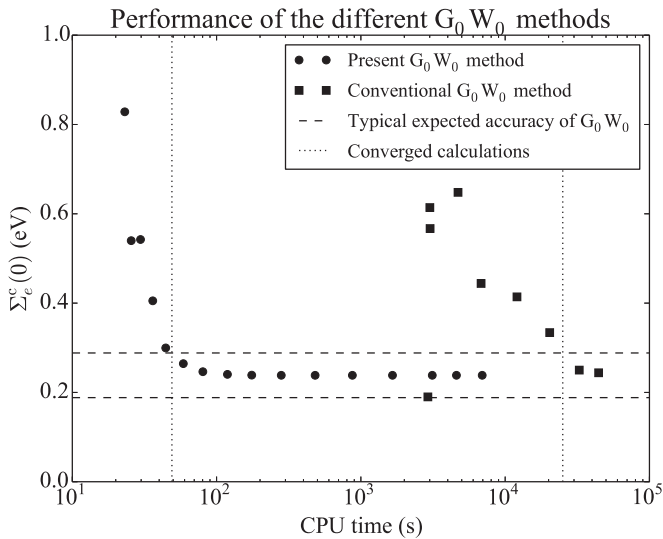

FIG. 2. Convergence study on the expectation value of the correlation part of the self-energy for the HOMO orbital of the silane molecule $\Sigma_e^c(0)$ using the present and the conventional G_0W_0 implementation of the ABINIT project. The convergence studies are carried out with respect to the size of the dielectric matrix, which is controlled by the number of Lanczos vectors N_L in the first implementation and the number of plane waves used to describe the dielectric matrix $N_{PW\epsilon}$ 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.

Although the present implementation does not yet support periodic systems, it is still possible to extend further the performance analysis by doing (unphysical) simulations of crystals where only one k point is used to sample the Brillouin zone (Γ). We carry out such simulations for silicon, diamond, and graphite. All parameters are kept to a common, converged value, except for the number of Lanczos vectors N_L in the present implementation as well as the number of conduction states N_c and the number of plane waves used to describe the dielectric matrix $N_{PW\epsilon}$ in the conventional implementation. For the present implementation, we choose N_L so that the result is converged to 50 meV, while for the conventional implementation, we choose N_c and $N_{PW\epsilon}$ so that they each contribute 25 meV to the final error, the latter choice being generally the most computationally efficient way to reach a final error of 50 meV. These tests show our implementation to be five times, six times, and 20 times faster than the conventional implementation for silicon, diamond, and graphite, respectively. Thus the speedup of the present implementation increases as the number of plane waves per valence electron required to accurately describe the dielectric matrix increases, silicon likely being one of the lowest possible speedup.

Furthermore, this speedup is for an unphysical simulation where only one k point is used to sample the Brillouin zone. Indeed, the generalization of our implementation to periodic systems should profit from reuse of computationally expensive information (such as the Lanczos basis $\{|l\rangle\}$ or the bases $\{|\gamma_{i,v}\rangle\}$ used to solve Eq. (23) at finite frequency) from one k point to another, unlike the conventional implementation, which requires equally large computational effort for all k points. The speedup obtained here for extended systems should therefore be substantially lower than those to be reached with the extension of our implementation to periodic systems, by a factor of up to the number of k points used to sample the Brillouin zone.

XI. 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 [15], 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 behavior 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 a 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 behavior 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 (except for the LUMO of silane, where the scatter of available data makes the analysis difficult), thus validating the accuracy of our implementation. Moreover, our agreements with experimental ionization energies is 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 (using k-point grids) is straightforward and under way.

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[2] K. Capelle, Braz. J. Phys. **36**, 1318 (2006).

[3] R. M. Martin, *Electronic Structure, Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, UK, 2004).

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

[5] W. Aulbur, L. Jonsson, and J. Wilkins, in *Solid State Physics: Advances in Research and Applications* (Academic Press, New York, 1999), Vol. 54, p. 1.

[6] L. Hedin, Physical Review **139**, 796 (1965).

[7] L. Hedin and S. Lundqvist, in *Solid State Physics: Advances in Research and Applications* (Academic Press, New York, 1969), Vol. 23, p. 1.

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

[9] M. Stankovski, G. Antonius, D. Waroquiers, A. Miglio, H. Dixit, K. Sankaran, M. Giantomassi, X. Gonze, M. Cote, and G. M. Rignanese, Phys. Rev. B **84**, 241201 (2011).

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

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

[12] I. Tamblyn, P. Darancet, S. Y. Quek, S. A. Bonev, and J. B. Neaton, Phys. Rev. B **84**, 201402 (2011).

[13] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **81**, 115104 (2010).

[14] F. Giustino, M. L. Cohen, and S. G. Louie, Phys. Rev. B **81**, 115105 (2010).

[15] T. A. Pham, H.-V. Nguyen, D. Rocca, and G. Galli, Phys. Rev. B **87**, 155148 (2013).

[16] L. Reining, G. Onida, and R. W. Godby, Phys. Rev. B **56**, R4301 (1997).

[17] H. Lambert and F. Giustino, Phys. Rev. B **88**, 075117 (2013).

[18] G. Samsonidze, M. Jain, J. Deslippe, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **107**, 186404 (2011).

[19] F. Bruneval and X. Gonze, Phys. Rev. B **78**, 085125 (2008).

[20] J. Deslippe, G. Samsonidze, M. Jain, M. L. Cohen, and S. G. Louie, Phys. Rev. B **87**, 165124 (2013).

[21] J. A. Berger, L. Reining, and F. Sottile, Phys. Rev. B **82**, 041103 (2010).

[22] J. A. Berger, L. Reining, and F. Sottile, Eur. Phys. J. B **85**, 326 (2012).

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

[24] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **79**, 201104 (2009).

[25] 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. Giantomassi, S. Goedecker, D. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. J. T. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G. M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. J. Verstraete, G. Zerah, and J. W. Zwanziger, Comp. Phys. Commun. **180**, 2582 (2009).

[26] X. Gonze, G. M. Rignanese, M. Verstraete, J. M. Beuken, Y. Pouillon, R. Caracas, F. Jollet, M. Torrent, G. Zerah, M. Mikami, P. Ghosez, M. Veithen, J. Y. Raty, V. Olevano, F. Bruneval, L. Reining, R. Godby, G. Onida, D. Hamann, and D. Allan, Zeitschrift für Kristallographie **220**, 558 (2005).

[27] X. Gonze, J. M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J. Y. Raty, and D. Allan, Comp. Mater. Sci. **25**, 478 (2002).

[28] F. Bottin, S. Leroux, A. Knyazev, and G. Zerah, Comp. Mater. Sci. **42**, 329 (2008).

[29] S. Goedecker, SIAM J. Sci. Comp. **18**, 1605 (1997).

[30] G. H. Golub and C. F. Van Loan, *Matrix Computations* (The Johns Hopkins University Press, London, UK, 2012).

[31] H. F. Wilson, D. Lu, F. Gygi, and G. Galli, Phys. Rev. B **79**, 245106 (2009).

[32] H. F. Wilson, F. Gygi, and G. Galli, Phys. Rev. B **78**, 113303 (2008).

[33] M. Giantomassi, M. Stankovski, R. Shaltaf, M. Gruning, F. Bruneval, P. Rinke, and G. M. Rignanese, Physica Status Solidi B **248**, 275 (2011).

[34] B. I. Lundqvist, Physik der Kondensierten Materie **7**, 117 (1968).

[35] R. W. Godby and R. J. Needs, Phys. Rev. Lett. **62**, 1169 (1989).

[36] W. von der Linden and P. Horsch, Phys. Rev. B **37**, 8351 (1988).

[37] G. E. Engel and B. Farid, Phys. Rev. B **47**, 15931 (1993).

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

[39] V. I. Anisimov, *Strong Coulomb Correlations in Electronic Structure Calculations*, Advances in Condensed Matter Science Vol. 1 (Gordon and Breach Science Publishers, Amsterdam, 2000).

[40] M. M. Rieger, L. Steinbeck, I. D. White, H. N. Rojas, and R. W. Godby, Comp. Phys. Commun. **117**, 211 (1999).

[41] R. W. Freund and N. M. Nachtigal, Appl. Num. Math. **19**, 319 (1995).

[42] Once the Taylor expansion of Eq. (1) is done, one can solve for $\Delta\varepsilon_e$. The evaluation of $\langle e|\hat{\Sigma}(\omega)|e\rangle$ at two frequencies then allows $\Delta\varepsilon_e$ to be obtained. In the present implementation, we evaluate $\langle e|\hat{\Sigma}(\omega)|e\rangle$ at three different energies to control the accuracy of the linear approximation for Eq. (1). We choose for these energies the values ε_e (the DFT eigenvalue to be corrected), $\varepsilon_e + \langle e|\hat{v}_C - \hat{V}^{xc}|e\rangle$ (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).

[43] The symmetric form of the dielectric matrix [33,14,15] $\hat{\epsilon}(\omega) = 1 - \hat{v}_C^{1/2}\hat{P}(\omega)\hat{v}_C^{1/2}$, where $\hat{P}(\omega)$ is the irreducible polarizability, is Hermitian, in contrast to its usual form $\hat{\epsilon}(\omega) = 1 - \hat{v}_C\hat{P}(\omega)$. 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 [41]. 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.

[44] R. W. Godby, M. Schluter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[45] A simple change of variable $\omega = (1+t)/(1-t)$ allows to remap the semi-infinite interval $[0, +\infty[$ to $]-1,1]$. The integration can now be carried out numerically using methods such as Gauss-Legendre quadrature. We found that a sampling

[46] S. L. Adler, Phys. Rev. **126**, 413 (1962).

[47] N. Wiser, Phys. Rev. **129**, 62 (1963).

[48] S. Baroni, S. de Gironcoli, A. dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).

[49] X. Gonze, Phys. Rev. A **52**, 1096 (1995).

[50] R. M. Sternheimer, Phys. Rev. **96**, 951 (1954).

[51] Y. Saad, *Iterative Methods for Sparse Linear Systems* (SIAM, Philadelphia, 2003).

[52] 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 Nth 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 [30] 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 [30], 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 [41], the interested reader may consult the original work on the MINRES method [89] for a more detailed explanation of the method.

[53] See the *zcut* input variable documentation on the ABINIT website [66].

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

$$-\frac{1}{2}\langle em^*|\hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(0) - \hat{1})\hat{v}_C^{1/2}|m^*e\rangle \quad \text{if } |m\rangle \in \{|v\rangle\},$$

$$+\frac{1}{2}\langle em^*|\hat{v}_C^{1/2}(\hat{\epsilon}^{-1}(0) - \hat{1})\hat{v}_C^{1/2}|m^*e\rangle \quad \text{if } |m\rangle \in \{|c\rangle\}.$$

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

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

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. Also, since $\Theta(x)$ takes eigenenergies in argument (in contrast to band indices) due to the form of Eq. (3), degenerate states are uniformly treated

by construction and need no special treatment in the present formalism. Furthermore, in a future generalization of the present implementation to extended systems, the use of k-point grids to sample the Brillouin zone will automatically discretize the continuous spectrum of eigenvalues, so that the present treatment of the discontinuities will still apply.

[55] The choice $\delta = \langle e|\hat{\Sigma}^{x} - \hat{V}^{xc}|e\rangle$ is motivated by the fact that it is the value of δ closest to $\Delta\varepsilon_e$ among those we use for our G_0W_0 simulations.

[56] H.-V. Nguyen, T. A. Pham, D. Rocca, and G. Galli, Phys. Rev. B **85**, 081101 (2012).

[57] D. Rocca, R. Gebauer, Y. Saad, and S. Baroni, J. Chem. Phys. **128**, 154105 (2008).

[58] B. Walker, A. M. Saitta, R. Gebauer, and S. Baroni, Phys. Rev. Lett. **96**, 113001 (2006).

[59] A. Frommer, Computing **70**, 87 (2003).

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

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

[62] O. Gunnarsson, *Alkali-Doped Fullerides*, Narrow-Band Solids with Unusual Properties (World Scientific, Singapore, 2004).

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

[64] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).

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

[66] 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. Giantomassi, S. Goedecker, D. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. J. T. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G. M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. J. Verstraete, G. Zerah, and J. W. Zwanziger, The ABINIT project, http://www.abinit.org/ (2013).

[67] M. Krack, Theor. Chem. Acc. **114**, 145 (2005).

[68] J. Spencer and A. Alavi, Phys. Rev. B **77**, 193110 (2008).

[69] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson,

A. Bergman, and R. Schneider, J. Chem. Phys. **129**, 014109 (2008).

[70] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).

[71] L. Genovese, T. Deutsch, A. Neelov, S. Goedecker, and G. Beylkin, J. Chem. Phys. **125**, 074105 (2006).

[72] M. Rohlfing and S. G. Louie, Phys. Rev. Lett. **80**, 3320 (1998).

[73] J. C. Grossman, M. Rohlfing, L. Mitas, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **86**, 472 (2001).

[74] P. H. Hahn, W. G. Schmidt, and F. Bechstedt, Phys. Rev. B **72**, 245425 (2005).

[75] F. Bruneval and M. A. L. Marques, J. Chem. Theor. Comp. **9**, 324 (2012).

[76] R. Roberge, C. Sandorfy, J. I. Matthews, and O. P. Strausz, J. Chemi. Phys. **69**, 5105 (1978).

[77] U. Itoh, Y. Toyoshima, H. Onuki, N. Washida, and T. Ibuki, J. Chem. Phys. **85**, 4867 (1986).

[78] M. J. van Setten, F. Weigend, and F. Evers, J. Chem. Theor. Comp. **9**, 232 (2013). Erratum: *The GW-Method for Quantum Chemistry Applications: Theory and Implementation* [JCTC 9, 232 (2013)], to be published in J. Chem. Theor. Comp.

[79] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, New J. Phys. **14**, 053020 (2012).

[80] M. J. van Setten *et al.* (unpublished).

[81] P. Boulanger and X. Blase (unpublished).

[82] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997).

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

[84] S. Sharifzadeh, I. Tamblyn, P. Doak, P. T. Darancet, and J. B. Neaton, Eur. Phys. J. B **85**, 323 (2012).

[85] M. L. Tiago and J. R. Chelikowsky, Phys. Rev. B **73**, 205334 (2006).

[86] D. Foerster, P. Koval, and D. Sánchez-Portal, J. Chem. Phys. **135**, 074105 (2011).

[87] D. L. Lichtenberger, K. W. Nebesny, C. D. Ray, D. R. Huffman, and L. D. Lamb, Chem. Phys. Lett. **176**, 203 (1991).

[88] D.-L. Huang, P. D. Dau, H.-T. Liu, and L.-S. Wang, J. Chem. Phys. **140**, 224315 (2014).

[89] C. C. Paige and M. A. Saunders, SIAM J. Num. Anal. **12**, 617 (1975).