GW method with the self-consistent Sternheimer equation

Feliciano Giustino,[1,2,3,*] Marvin L. Cohen,[2,3] and Steven G. Louie[2,3]
[1]*Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom*
[2]*Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA*
[3]*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
(Received 17 December 2009; published 3 March 2010)

We propose an approach to quasiparticle GW calculations which does not require the computation of unoccupied electronic states. In our approach the screened Coulomb interaction is evaluated by solving self-consistent linear-response Sternheimer equations and the noninteracting Green's function is evaluated by solving inhomogeneous linear systems. The frequency dependence of the screened Coulomb interaction is explicitly taken into account. In order to avoid the singularities of the screened Coulomb interaction the calculations are performed along the imaginary axis, and the results are analytically continued to the real axis through Padé approximants. As a proof of concept we implemented the proposed methodology within the empirical pseudopotential formalism and we validated our implementation using silicon as a test case. We examine the advantages and limitations of our method and describe promising future directions.

DOI: 10.1103/PhysRevB.81.115105 PACS number(s): 71.15.−m

I. INTRODUCTION

During the past two and a half decades the GW method[1,2] for the study of electron quasiparticle excitations has had a number of successes and witnessed significant growth of interest within the computational electronic-structure community. The GW method is currently being used for predicting electron quasiparticle excitation spectra as well as optical spectra in a variety of materials ranging from bulk solids to nanostructures and organic systems. The GW method is also of widespread use as a starting point for Bethe-Salpeter calculations of two-particle neutral excitations.[3–8] Current implementations find many diverse applications, including among others the calculation of the optical response of nanostructures,[9] quantum transport in nanoscale junctions,[10] pump-probe spectroscopy,[11] angle-resolved photoemission spectroscopy,[12] and strongly correlated systems.[13]

Current trends in the development of improved computational approaches for quasiparticle excitations based on the GW method include the refinement of the initial guess for the noninteracting Green's function and for the polarization operator,[14,15] the inclusion of approximate vertex corrections or higher-order self-energy diagrams,[16] and the description of the frequency-dependent dielectric response beyond the original generalized plasmon-pole approximation.[17,18] Detailed reviews of past and current developments in GW techniques can be found in Refs. 8, 14, and 19–21.

The majority of current GW implementations obtain the screened Coulomb interaction W and the noninteracting Green's function G using a perturbative expansion over the Kohn-Sham eigenstates (cf. Sec. II A below). Such expansion requires the calculation of both occupied and unoccupied electronic states, as well as their associated optical matrix elements.[2] A common bottleneck of this approach is that the convergence of the quasiparticle excitation energies with the number of unoccupied states is rather slow.[22] This difficulty is particularly relevant when calculating the absolute values of the quasiparticle excitation energies.[23] Several avenues have been explored so far in order to circumvent this bottleneck and to perform GW calculations by employing only occupied electronic states,[24–27] or a small number of unoccupied states.[23]

The main aim of the present work is to demonstrate the feasibility of GW calculations entirely based on occupied states only.[28] In practice we adopt the principles of density-functional perturbation theory (DFPT) to determine (i) the frequency-dependent screened Coulomb interaction by directly solving self-consistent linear-response Sternheimer equations and (ii) the noninteracting Green's function by solving inhomogeneous linear systems. The main advantage of the proposed method is that it does not require the computation of unoccupied electronic states. In addition, we demonstrate the possibility of fast evaluations of the frequency dependence of the screened Coulomb interaction based on *multishift* linear-system solvers.[29] As a proof of concept we have implemented our method within a plane-waves empirical pseudopotential scheme[30] and validated it by comparing with previous work for the prototypical test case of silicon.

The use of the Sternheimer equation for calculating the polarizability in the random-phase approximation (RPA) or the inverse dielectric matrix has already been discussed in Refs. 31 and 32, respectively, within the framework of a nonperturbative supercell approach. After the introduction of DFPT in the context of lattice-dynamical calculations,[33] the authors of Ref. 24 proposed the use of the non-self-consistent Sternheimer method for the calculation of the dielectric matrix. The elimination of unoccupied electronic states in the evaluation of the screened Coulomb interaction has also been proposed recently within the framework of a Wannier-type representation of the polarization propagator and the Lanczos recursion method.[25,26]

This paper is organized as follows. In Sec. II we describe how the self-consistent Sternheimer formalism can be adapted to perform GW calculations. In particular, we outline the procedure to obtain the screened Coulomb interaction in Sec. II A, the noninteracting Green's function in Sec. II B, and the GW self-energy in Sec. II C. In Sec. III we specialize to a plane-wave basis set representation and derive the key

equations for the case of Bloch electrons. Sections III A–III C parallel the corresponding sections in the general theory part, respectively. In Sec. IV we critically analyze the advantages and limitations of the present approach with an emphasis on the scaling of the calculations with system size. In Sec. V we report the results of our test calculations for silicon and compare with previous calculations in the literature. Specifically, we presents results for the direct and inverse dielectric matrix (Sec. V A), for the analytic continuation of the dielectric matrix using Padé approximants (Sec. V B), for the self-energy (Sec. V C), and for the spectral function (Sec. V D). In Sec. VI we discuss possible future developments of our method and discuss our conclusions. The appendices provide technical details on some numerical algorithms adopted in this work, in particular, the preconditioned complex biconjugate gradient method (Appendix A), the analysis of the conditioning of the Sternheimer equations (Appendix B), the analytic continuation using Padé approximants (Appendix C), and the use of multishift methods for the simultaneous calculation of the polarization at multiple frequencies (Appendix D).

II. GENERAL THEORY

A. Screened Coulomb interaction

In this section we describe how to exploit the Sternheimer scheme within density-functional perturbation theory in order to calculate the screened Coulomb interaction $W(\mathbf{r},\mathbf{r}';\omega)$ (where $\mathbf{r}$, $\mathbf{r}'$ are the space variables and ω is the excitation frequency). While the use of the Sternheimer approach in DFPT was originally developed bearing in mind the Kohn-Sham effective Hamiltonian, we note that the present procedure applies without restrictions also to post-DFT methods such as the local-density approximation (LDA) plus U method,[34] hybrid functionals,[35] and exact exchange.[36] We assume Rydberg atomic units throughout this paper. The Hedin's equation which defines the screened Coulomb interaction reads,[20]

$$W(\mathbf{r},\mathbf{r}';\omega) = v(\mathbf{r},\mathbf{r}') + \int d\mathbf{r}'' W(\mathbf{r},\mathbf{r}'';\omega) \times \int d\mathbf{r}''' P(\mathbf{r}'',\mathbf{r}''';\omega) v(\mathbf{r}''',\mathbf{r}'), \tag{1}$$

where $v(\mathbf{r},\mathbf{r}')=e^2/|\mathbf{r}-\mathbf{r}'|$ denotes the bare Coulomb interaction and $P(\mathbf{r},\mathbf{r}';\omega)$ the irreducible polarization propagator. As Eq. (1) is a self-consistent Dyson equation for the screened Coulomb interaction, it should be possible to solve it recursively in the spirit of density-functional perturbation theory. For simplicity, we here specialize to the case of the RPA for the polarization propagator. The generalization of this procedure to include exchange and correlation effects can be performed without difficulties (cf. Sec. II A 1). Within the random-phase approximation the polarization propagator can be written as,[20]

$$P(\mathbf{r},\mathbf{r}';\omega) = 2\sum_{nm} \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m - \omega} \psi_n(\mathbf{r})\psi_m^\star(\mathbf{r})\psi_n^\star(\mathbf{r}')\psi_m(\mathbf{r}'), \tag{2}$$

where $\psi_n(\mathbf{r})$ indicates an electronic eigenstate of the single-particle Hamiltonian with energy eigenvalue ε_n and occupation number f_n. In the following we assume that the $\psi_n(\mathbf{r})$ are Kohn-Sham eigenstates for definiteness. In Eq. (2) the summation indices m and n run over both occupied and unoccupied electronic states, and the factor of 2 accounts for the spin degeneracy.[20] Although the expression for the RPA polarization Eq. (2) has been derived for real frequencies in Ref. 20, it is possible to continue the polarization throughout the complex plane by using Eq. (2) as a definition outside of the real axis.

Our goal is to rewrite Eqs. (1) and (2) by avoiding explicit summations over the unoccupied electronic states. For this purpose it is convenient to regard the screened Coulomb interaction $W(\mathbf{r},\mathbf{r}';\omega)$ as a function of the second space variable $\mathbf{r}'$ while the first space variable $\mathbf{r}$ and the frequency ω are kept as parameters: $\Delta V_{[\mathbf{r},\omega]}(\mathbf{r}')=W(\mathbf{r},\mathbf{r}';\omega)$. If the system under consideration is subject to the perturbation $\Delta V_{[\mathbf{r},\omega]}(\mathbf{r}')$, then within the RPA the first-order variation in the single-particle density matrix $\Delta n_{[\mathbf{r},\omega]}$ reads

$$\Delta n_{[\mathbf{r},\omega]} = 2\sum_{v\sigma} \psi_v^\star \Delta\psi^\sigma_{v[\mathbf{r},\omega]}. \tag{3}$$

In Eq. (3) the index v stands for "valence" and runs over the occupied electronic states only, the factor of 2 is for the spin degeneracy, and the superscript $\sigma=\pm$ refer to the positive and negative frequency components of the induced charge. The first-order variations in the occupied wave functions $\Delta\psi^\sigma_{v[\mathbf{r},\omega]}$ can be determined by solving the following two Sternheimer equations:

$$(\hat{H} - \varepsilon_v \pm \omega)\Delta\psi^\pm_{v[\mathbf{r},\omega]} = -(1-\hat{P}_{\text{occ}})\Delta V_{[\mathbf{r},\omega]}\psi_v, \tag{4}$$

where $\hat{H}$ is the effective single-particle Hamiltonian and $\hat{P}_{\text{occ}}=\Sigma_v|\psi_v\rangle\langle\psi_v|$ is the projector on the occupied states manifold. In the particular case of vanishing frequency ($\omega=0$) the $\sigma=\pm$ variations in the wave functions do coincide, and the standard DFPT equations are recovered. The screening Hartree potential associated with the induced charge $\Delta n_{[\mathbf{r},\omega]}$ is calculated as usual through

$$\Delta V^{\text{H}}_{[\mathbf{r},\omega]}(\mathbf{r}') = \int d\mathbf{r}'' \Delta n_{[\mathbf{r},\omega]}(\mathbf{r}'') v(\mathbf{r}'',\mathbf{r}'), \tag{5}$$

and finally the screened Coulomb interaction in the RPA is obtained as

$$W(\mathbf{r},\mathbf{r}';\omega) = \Delta V_{[\mathbf{r},\omega]}(\mathbf{r}') = v(\mathbf{r},\mathbf{r}') + \Delta V^{\text{H}}_{[\mathbf{r},\omega]}(\mathbf{r}'). \tag{6}$$

It is tedious but otherwise straightforward to verify that Eqs. (3)–(6) are *equivalent* to the original Eqs. (1) and (2). The only assumptions made in our derivation are that time-reversal symmetry applies, and that the system under consideration has a finite-energy gap for electronic excitations. The assumption of time-reversal symmetry is not essential and is mainly used to obtain a compact expression for the $\sigma=\pm$

wave-function perturbations. The assumption of finite-energy gap can be relaxed by using the extension of DFPT to metallic systems developed in Ref. 37.

There is a simple and intuitive physical meaning associated with the calculation scheme outlined above. To see this we consider an external *test charge* introduced in the system at the point $\mathbf{r}$. This charge generates a bare Coulomb potential $v(\mathbf{r},\mathbf{r}')$, and the system responds to such perturbation by generating the induced charge $\Delta n_{[\mathbf{r},\omega]}(\mathbf{r}')$ and the associated screening potential $\Delta V^{\mathrm{H}}_{[\mathbf{r},\omega]}(\mathbf{r}')$. The sum of the external perturbation $v(\mathbf{r},\mathbf{r}')$ and the screening potential $\Delta V^{\mathrm{H}}_{[\mathbf{r},\omega]}(\mathbf{r}')$ yields the screened Coulomb interaction $W(\mathbf{r},\mathbf{r}';\omega)$ at the point $\mathbf{r}'$ within the RPA.

The linear systems in Eq. (4) must be solved self-consistently. For this purpose we begin by initializing the screened Coulomb interaction W using the bare interaction v. We then calculate the linear variations in the wave functions $\Delta\psi^{\sigma}_{v}$. Using the calculated linear variations we update the induced charge density Δn and the associated screening potential ΔV^{H}. This allow us to generate an improved estimate of the screened Coulomb interaction W. We cycle through these steps until convergence of the screened Coulomb interaction is achieved. Equations (3) and (4) can be regarded as the generalization of the self-consistent Sternheimer equations used for lattice-dynamical calculations[28] to finite-frequency test-charge perturbations.

In practical calculations we solve Eq. (4) along the imaginary frequency axis in order to avoid the null eigenvalues of the operator $\hat{H}-\varepsilon_v \pm \omega$, and then we perform the analytic continuation of the screened Coulomb interaction to real frequencies (cf. Appendices A and C). In the special case of $\omega=0$ it is convenient to modify the linear operator on the left-hand side of Eq. (4) by adding the projector on the occupied states manifold $\hat{P}_{\mathrm{occ}}$,

$$(\hat{H}-\varepsilon_v+\alpha\hat{P}_{\mathrm{occ}})\Delta\psi_{v[\mathbf{r},0]}=-(1-\hat{P}_{\mathrm{occ}})\Delta V_{[\mathbf{r},0]}\psi_v, \qquad (7)$$

with α set to twice the occupied bandwidth. This extra term does not affect the solutions $\Delta\psi_{v[\mathbf{r},0]}$ which are linear combinations of unoccupied electronic states. At the same time, the extra term has the effect of shifting away from zero the null eigenvalues of the linear operator $\hat{H}-\varepsilon_v$ thereby making it nonsingular. This strategy is common practice in DFPT implementations,[28,38] and is discussed in greater detail in Appendix B.

1. Vertex correction

Within the scheme outlined here it is rather straightforward to introduce an approximate vertex correction to the GW self-energy along the lines of Refs. 2 and 39. This correction results from setting the self-energy in the first iteration of Hedin's equations to the DFT exchange-correlation (XC) potential, $\Sigma_0(\mathbf{r},\mathbf{r}';\omega)=\delta(\mathbf{r},\mathbf{r}')V_{\mathrm{xc}}(\mathbf{r})$. Within the present scheme this correction is simply obtained by including the variation in the exchange-correlation potential in the self-consistent potential used in Eq. (4),

$$(\hat{H}-\varepsilon_v\pm\omega)\Delta\psi^{\pm}_{v[\mathbf{r},\omega]}=-(1-\hat{P}_{\mathrm{occ}})[\Delta V_{[\mathbf{r},\omega]}+K_{\mathrm{xc}}\Delta n_{[\mathbf{r},\omega]}]\psi_v, \qquad (8)$$

$K_{\mathrm{xc}}=\delta V_{\mathrm{xc}}/\delta n$ being the functional derivative of the XC potential with respect to the density. The screened Coulomb interaction is still to be calculated through Eq. (6). This approach has been called "GW+K_{xc}" approximation in Ref. 39 due to the inclusion of the XC contribution in the screening of the test charge. That the inclusion of the XC term in the self-consistent induced potential leads to the GW+K_{xc} approximation can easily be seen as follows. We combine Eqs. (3) and (8) to yield the induced charge density (we use symbolic operator notation for clarity),

$$\Delta n=v[1-P(v+K_{\mathrm{xc}})]^{-1}P. \qquad (9)$$

Then, we substitute this result in the definition of the screened Coulomb interaction Eqs. (5) and (6) to find

$$W=v\{1+v[1-P(v+K_{\mathrm{xc}})]^{-1}P\}. \qquad (10)$$

The last equation yields precisely the screened Coulomb interaction in the GW+K_{xc} approximation.[2,39] The difference between this approach and the standard GW approximation is that in this case the screening charge is calculated for an *electron* while in the GW-RPA approximation the screening is calculated for a *test charge*. It is worth pointing out that in standard implementations of DFPT the XC term is already included in the variation in the self-consistent potential,[28] therefore the use of the GW+K_{xc} approximation would not require any additional computational developments if the present approach was to be implemented on top of existing DFT software.

2. Non-self-consistent calculation of the dielectric matrix

An alternative approach to the calculation of the screened Coulomb interaction using the self-consistent Sternheimer equation consists in solving Eq. (4) non-self-consistently. For this purpose we can replace the self-consistent perturbation $\Delta V_{[\mathbf{r},\omega]}(\mathbf{r}')$ in the right-hand side (rhs) of Eq. (4) by the bare Coulomb potential $v_{[\mathbf{r}]}(\mathbf{r}')=v(\mathbf{r},\mathbf{r}')$ as follows:

$$(\hat{H}-\varepsilon_v\pm\omega)\Delta\psi^{\mathrm{NS},\pm}_{v[\mathbf{r},\omega]}=-(1-\hat{P}_{\mathrm{occ}})v_{[\mathbf{r}]}\psi_v, \qquad (11)$$

and we can solve this Sternheimer equation with the known term on the right-hand side kept fixed. By constructing the non-self-consistent induced charge density $\Delta n^{\mathrm{NS}}_{[\mathbf{r},\omega]}$ as in Eq. (3) we then obtain the dielectric matrix $\epsilon(\mathbf{r},\mathbf{r}';\omega)$,

$$\epsilon(\mathbf{r},\mathbf{r}';\omega)=\delta(\mathbf{r},\mathbf{r}')-\Delta n^{\mathrm{NS}}_{[\mathbf{r},\omega]}(\mathbf{r}'). \qquad (12)$$

It is straightforward to check that this procedure correctly leads to the RPA dielectric matrix.[40] The difference between this approach and the self-consistent calculation described in Sec. II A is that here we also need to invert the dielectric matrix obtained through Eq. (12) in order to calculate the screened Coulomb interaction.

This non-self-consistent procedure was first proposed in Ref. 24. One additional step that we make in the present work is to notice that Eq. (11) constitutes a *shifted* linear system, i.e., a system where the linear operator $\hat{H}-\varepsilon_v\pm\omega$

differs from the "seed" operator $\hat{H}-\varepsilon_v$ only by a constant shift $\pm\omega I$ (I being the identity operator). In this case we can take advantage of the *multishift* linear system solver of Ref. 29 to determine $\Delta\psi^{\mathrm{NS},\pm}_{v[\mathbf{r},\omega]}$ for every frequency ω at the computational cost of one single calculation for the *seed* system $(\hat{H}-\varepsilon_v)\Delta\psi^{\mathrm{NS}}_{v[\mathbf{r},0]}=-(1-\hat{P}_{\mathrm{occ}})v_{[\mathbf{r}]}\psi_v$. This procedure is extremely advantageous as it makes it possible to calculate the entire frequency dependence by performing one single iterative minimization. The technical implementation of this procedure is described in Appendix D.

B. Green's function

The calculation of the Green's function can efficiently be performed by adopting a strategy similar to the Sternheimer approach described in Sec. II A. We introduce the noninteracting Green's function following Ref. 20,

$$G(\mathbf{r},\mathbf{r}';\omega)=\sum_n \frac{\psi_n(\mathbf{r})\psi_n^\star(\mathbf{r}')}{\omega-\varepsilon_n-i\eta_n}, \tag{13}$$

where the sum extends over occupied as well as unoccupied electronic states. The real infinitesimal η_n is positive ($\eta_n=\eta$) for occupied states and negative ($\eta_n=-\eta$) for unoccupied states.[2,20,41] We now split the sum in Eq. (13) into occupied (v) and unoccupied (c) states,

$$G(\mathbf{r},\mathbf{r}';\omega)=\sum_v \frac{\psi_v(\mathbf{r})\psi_v^\star(\mathbf{r}')}{\omega-\varepsilon_v-i\eta}+\sum_c \frac{\psi_c(\mathbf{r})\psi_c^\star(\mathbf{r}')}{\omega-\varepsilon_c+i\eta}, \tag{14}$$

and we add and subtract $\Sigma_v\psi_v\psi_v^\star/(\omega-\varepsilon_v+i\eta)$ to obtain,

$$G(\mathbf{r},\mathbf{r}';\omega)=G^{\mathrm{A}}(\mathbf{r},\mathbf{r}';\omega)+G^{\mathrm{N}}(\mathbf{r},\mathbf{r}';\omega), \tag{15}$$

with

$$G^{\mathrm{A}}(\mathbf{r},\mathbf{r}';\omega)=\sum_n \frac{\psi_n^\star(\mathbf{r})\psi_n(\mathbf{r}')}{\omega-\varepsilon_n^-}, \tag{16}$$

$$G^{\mathrm{N}}(\mathbf{r},\mathbf{r}';\omega)=2\pi i\sum_v \delta(\omega-\varepsilon_v)\psi_v^\star(\mathbf{r})\psi_v(\mathbf{r}'). \tag{17}$$

In the above derivation we assumed again time-reversal symmetry, we used the Lorentzian representation of the Dirac's delta function for small η [$\pi\delta(x)=\eta/(x^2+\eta^2)$], and we defined $\varepsilon_n^-=\varepsilon_n-i\eta$. The component G^{A} of the Green's function is obviously analytic in the upper half of the complex energy plane as its poles lie below the real axis. The nonanalytic component G^{N} vanishes whenever the frequency ω is above the chemical potential. For frequencies ω below the chemical potential, the nonanalytic component introduces the poles associated with the occupied electronic states. The partitioning of Eqs. (16) and (17) closely reflects the analytic structure of the noninteracting Green's function. A detailed discussion of this aspect can be found in Ref. 20. The two components of the Green's function in Eq. (16) are associated with the Coulomb hole and the screened exchange terms of the self-energy, $\Sigma^{\mathrm{COH}}=G^{\mathrm{A}}W$ and $\Sigma^{\mathrm{SEX}}=G^{\mathrm{N}}W$, respectively.[2]

The computation of the nonanalytic component G^{N} of the Green's function in Eq. (17) is straightforward once the occupied electronic eigenstates have been determined. In order to calculate the analytic component G^{A} it is convenient to proceed as in the case of the screened Coulomb interaction, by regarding $G^{\mathrm{A}}(\mathbf{r},\mathbf{r}';\omega)$ as a parametric function of the first space variable and of the frequency: $G^{\mathrm{A}}_{[\mathbf{r},\omega]}(\mathbf{r}')=G^{\mathrm{A}}(\mathbf{r},\mathbf{r}';\omega)$. If we apply the operator $\hat{H}-\omega^+$ to both sides of Eq. (16), with $\omega^+=\omega+i\eta$, and we use the completeness relation $\delta_{[\mathbf{r}]}(\mathbf{r}')=\delta(\mathbf{r},\mathbf{r}')=\Sigma_n\psi_n(\mathbf{r})\psi_n^\star(\mathbf{r}')$, then we find immediately,

$$(\hat{H}-\omega^+)G^{\mathrm{A}}_{[\mathbf{r},\omega]}=-\delta_{[\mathbf{r}]}. \tag{18}$$

As expected, we can determine the analytic part of the Green's function by directly solving a linear system. As Eq. (18) does not explicitly require unoccupied electronic states, this procedure mimics the Sternheimer approach for the screened Coulomb interaction outlined in Sec. II A, albeit without the self-consistency requirement.

The procedure described here is especially advantageous because Eq. (18) constitutes a shifted linear system in the same way as Eq. (11). Also in this case we exploit the multishift method of Ref. 29 to determine $G^{\mathrm{A}}_{[\mathbf{r},\omega]}$ for every frequency ω at the computational cost of one single calculation for the seed system $\hat{H}G^{\mathrm{A}}_{[\mathbf{r},0]}=-\delta_{[\mathbf{r}]}$ (cf. Appendix D).

The presence of the infinitesimal $i\eta$ in $\omega^+=\omega+i\eta$ guarantees that the linear operator $\hat{H}-\omega^+$ in Eq. (41) is never singular. This operator can nonetheless become ill conditioned, hence the use of appropriate preconditioners may become necessary. We discuss this aspect in Appendix B.

C. Self-energy

The electron self-energy in the GW approximation is,[2]

$$\Sigma(\mathbf{r},\mathbf{r}';\omega)=\frac{i}{2\pi}\int_{-\infty}^{+\infty} d\omega' G(\mathbf{r},\mathbf{r}',\omega+\omega')W(\mathbf{r},\mathbf{r}',\omega')e^{-i\delta\omega'}, \tag{19}$$

where δ is a positive infinitesimal. At large frequencies the Green's function decays as ω^{-1} and the screened Coulomb interaction tends to the frequency-independent bare Coulomb interaction v. As a consequence, the integrand in Eq. (19) decays as ω^{-1} and the integration requires some care.

It is convenient to split the self-energy into an exchange contribution $\Sigma^{\mathrm{ex}}=Gv$ and a Coulomb term $\Sigma^{\mathrm{c}}=G(W-v)$.[18] It is easy to check that the integrand in the Coulomb term decays as ω^{-2} at large frequencies, therefore the integral is well behaved and the integration can be performed by using a numerical cutoff $|\omega'|<\omega_{\mathrm{C}}$ in Eq. (19). A detailed analysis of the analytic properties of the Coulomb term $\Sigma^{\mathrm{c}}(\omega)$ shows that it must decay as ω^{-1} at large frequencies, and that the use of the cutoff ω_{C} in the integration introduces an error on the order of $\omega_{\mathrm{p}}/\omega_{\mathrm{C}}$, where ω_{p} denotes the characteristic plasmon frequency of the system.

In order to integrate the exchange term we observe that $\Sigma^{\mathrm{ex}}=G^{\mathrm{A}}v+G^{\mathrm{N}}v$ and that the poles of G^{A} lie in the lower half of the complex plane, hence the integration of the term $G^{\mathrm{A}}v$ yields a vanishing contribution. On the other hand, the integration of $G^{\mathrm{N}}v$ yields a constant (frequency-independent)

term.[42] In summary, we perform the frequency integration in Eq. (19) by evaluating numerically the Coulomb term using an energy cutoff, and by integrating analytically the exchange term,

$$\Sigma(\mathbf{r},\mathbf{r}';\omega) = \Sigma^{c}(\mathbf{r},\mathbf{r}';\omega) + \Sigma^{ex}(\mathbf{r},\mathbf{r}'), \tag{20}$$

with

$$\Sigma^{c}(\mathbf{r},\mathbf{r}';\omega) = \frac{i}{2\pi}\int_{-\omega_C}^{\omega_C} d\omega' G(\mathbf{r},\mathbf{r}';\omega+\omega')[W(\mathbf{r},\mathbf{r}',\omega') - v(\mathbf{r},\mathbf{r}')] \tag{21}$$

and

$$\Sigma^{ex}(\mathbf{r},\mathbf{r}') = -\sum_{v} \psi_v^{\star}(\mathbf{r})\psi_v(\mathbf{r}')v(\mathbf{r},\mathbf{r}'). \tag{22}$$

III. IMPLEMENTATION IN A BASIS OF PLANE WAVES

We here describe our plane-waves implementation of the scheme developed in Sec. II. The choice of a plane-waves representation was motivated by the need for making contact with existing literature on dielectric matrices,[40,43–45] and by the availability of DFPT software for lattice-dynamical calculations[38] which was used as a reference for our implementation.

We adopt the following conventions for the transformation from real to reciprocal space. The wave functions transform as usual according to

$$\psi_{n\mathbf{k}}(\mathbf{r}') = \frac{1}{\sqrt{\Omega}}\sum_{\mathbf{G}'} e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}'}u_{n\mathbf{k}}(\mathbf{G}') \tag{23}$$

with Ω the volume of the unit cell and $\mathbf{k}$ the Bloch wave vector. The bare Coulomb interaction transforms according to

$$v(\mathbf{r},\mathbf{r}') = \frac{1}{N_{\mathbf{q}}\Omega}\sum_{\mathbf{qG}} v(\mathbf{q}+\mathbf{G})e^{i(\mathbf{q}+\mathbf{G})\cdot(\mathbf{r}'-\mathbf{r})}, \tag{24}$$

where $\mathbf{q}$ is also a Bloch wave vector and $N_{\mathbf{q}}$ is the number of such wave vectors in our discretized Brillouin zone. In Eq. (24) we have $v(\mathbf{q}+\mathbf{G})=4\pi e^2/|\mathbf{q}+\mathbf{G}|^2$. The latter expression for $v(\mathbf{q}+\mathbf{G})$ is arrived at by replacing the integration $\int d\mathbf{r}\, \exp(i\mathbf{q}\cdot\mathbf{r})/|\mathbf{r}|$ over the crystal volume by an integration over all space. This choice corresponds to assuming that we can rely on a very fine sampling of the Brillouin zone. Had we performed the integration on a sphere with a radius R_c defined by the crystal volume $(4/3\pi R_c^3 = N_{\mathbf{q}}\Omega)$, then we would have obtained

$$v_t(\mathbf{q}+\mathbf{G}) = \frac{4\pi e^2}{|\mathbf{q}+\mathbf{G}|^2}(1-\cos|\mathbf{q}+\mathbf{G}|R_c), \tag{25}$$

which corresponds to the truncated Coulomb potential introduced in Ref. 46. We will come back to this aspect in Sec. III A. The screened Coulomb interaction and the noninteracting Green's function transform according to

$$W(\mathbf{r},\mathbf{r}';\omega) = \frac{1}{N_{\mathbf{q}}\Omega}\sum_{\mathbf{qGG}'} e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}W_{\mathbf{GG}'}(\mathbf{q};\omega)e^{i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} \tag{26}$$

and

$$G(\mathbf{r},\mathbf{r}';\omega) = \frac{1}{N_{\mathbf{k}}\Omega}\sum_{\mathbf{kGG}'} e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}G_{\mathbf{GG}'}(\mathbf{k};\omega)e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}'}, \tag{27}$$

with similar expressions for G^A and G^N. We note that the sign convention adopted here in the Fourier transforms [e.g., $\exp(+i\mathbf{q}\cdot\mathbf{r}')$ in the rhs of Eq. (26)] is necessary to obtain the compact expression (35) below for the induced charge and is opposite to the convention adopted in Ref. 2. Before proceeding it is also convenient to introduce the "right-sided" inverse dielectric matrix through

$$W(\mathbf{r},\mathbf{r}';\omega) = \int d\mathbf{r}'' v(\mathbf{r},\mathbf{r}'')\epsilon^{-1}(\mathbf{r}'',\mathbf{r}';\omega). \tag{28}$$

By adopting the same convention for the inverse dielectric matrix as for the screened Coulomb interaction the above equation can be rewritten as,

$$W_{\mathbf{GG}'}(\mathbf{q};\omega) = v(\mathbf{q}+\mathbf{G})\epsilon^{-1}_{\mathbf{GG}'}(\mathbf{q};\omega). \tag{29}$$

We note that our Eq. (29) is slightly different from the standard expression [e.g., Eq. (22) of Ref. 2], due to our choice of using the right-sided inverse dielectric matrix.

A. Screened Coulomb interaction

In order to rewrite Eqs. (3)–(6) in the Bloch representation and in reciprocal space we proceed as follows. We first write the linear systems Eq. (4) by relabeling the wave functions ψ_v as Bloch states $\psi_{v\mathbf{k}}$,

$$(\hat{H}-\varepsilon_{v\mathbf{k}} \pm \omega)\Delta\psi^{\pm}_{v\mathbf{k}[\mathbf{r},\omega]} = -(1-\hat{P}_{occ})\Delta V_{[\mathbf{r},\omega]}\psi_{v\mathbf{k}}. \tag{30}$$

The linear variations in the wave functions can be expanded in terms of Bloch waves as follows:

$$\Delta\psi^{\sigma}_{v\mathbf{k}[\mathbf{r},\omega]} = \frac{1}{N_{\mathbf{q}}\Omega}\sum_{\mathbf{qG}}\Delta u^{\sigma}_{v\mathbf{k}[\mathbf{q},\mathbf{G},\omega]}e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}'}e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}, \tag{31}$$

where $\Delta u^{\sigma}_{v\mathbf{k}[\mathbf{q},\mathbf{G},\omega]}$ is cell periodic in $\mathbf{r}'$. From the linear variations in the wave functions we construct the induced charge density using Eq. (3),

$$\Delta n_{[\mathbf{r},\omega]} = \frac{2}{N_{\mathbf{k}}}\sum_{v\mathbf{k}\sigma}\psi^{\star}_{v\mathbf{k}}\Delta\psi^{\sigma}_{v\mathbf{k}[\mathbf{r},\omega]}. \tag{32}$$

Here the factor $N_{\mathbf{k}}$ takes into account the normalization of the Bloch states in the unit cell [the wave functions ψ_v in Eq. (4) are normalized in the whole crystal]. Next we expand the screened Coulomb interaction in terms of Bloch waves,

$$\Delta V_{[\mathbf{r},\omega]}(\mathbf{r}') = \frac{1}{N_{\mathbf{q}}\Omega}\sum_{\mathbf{qG}}\Delta v_{[\mathbf{q},\mathbf{G},\omega]}(\mathbf{r}')e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}e^{i\mathbf{q}\cdot\mathbf{r}'}, \tag{33}$$

where $\Delta v_{[\mathbf{q},\mathbf{G},\omega]}$ is cell periodic in $\mathbf{r}'$. If we now place Eqs. (31) and (33) into Eq. (30) we discover that the component

$\Delta v_{[\mathbf{q},\mathbf{G},\omega]}$ of the perturbing potential corresponding to the Bloch wave $\exp(-i\mathbf{q}\cdot\mathbf{r})$ couples only to the variations in the wave functions corresponding to the Bloch wave $\exp[i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}']$. As a result the linear system Eq. (4) becomes

$$(\hat{H}_{\mathbf{k}+\mathbf{q}} - \varepsilon_{v\mathbf{k}} \pm \omega)\Delta u^{\pm}_{v\mathbf{k}[\mathbf{q},\mathbf{G},\omega]} = -(1-\hat{P}^{\mathbf{k}+\mathbf{q}}_{\mathrm{occ}})\Delta v_{[\mathbf{q},\mathbf{G},\omega]}u_{v\mathbf{k}}, \tag{34}$$

where $\hat{H}_{\mathbf{k}}=e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{H}e^{i\mathbf{k}\cdot\mathbf{r}}$ and $\hat{P}^{\mathbf{k}}_{\mathrm{occ}}=\Sigma_v|u_{v\mathbf{k}}\rangle\langle u_{v\mathbf{k}}|$. The induced charge density associated with the Bloch wave $\exp(-i\mathbf{q}\cdot\mathbf{r})$ now reads

$$\Delta n_{[\mathbf{q},\mathbf{G},\omega]} = \frac{2}{N_{\mathbf{k}}}\sum_{v\mathbf{k}\sigma} u^{\star}_{v\mathbf{k}}\Delta u^{\sigma}_{v\mathbf{k}[\mathbf{q},\mathbf{G},\omega]}. \tag{35}$$

This result is very similar to the case of standard DFPT.[28] The main difference is that in the present case the translational invariance of the screened Coulomb interaction induces a coupling between the perturbation with Bloch wave $\exp(-i\mathbf{q}\cdot\mathbf{r})$ in the variable $\mathbf{r}$ and its induced response with Bloch wave $\exp(+i\mathbf{q}\cdot\mathbf{r}')$ in the variable $\mathbf{r}'$. To conclude our derivation, we rewrite the screened Coulomb interaction Eq. (6) after expanding the cell-periodic function $\Delta n_{[\mathbf{q},\mathbf{G},\omega]}(\mathbf{r}')$ in plane waves,

$$W_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega) = [\delta_{\mathbf{G}\mathbf{G}'} + \Delta n_{[\mathbf{q},\mathbf{G},\omega]}(\mathbf{G}')]v(\mathbf{q}+\mathbf{G}'). \tag{36}$$

In practical calculations we proceed as follows: we first initialize the perturbation in the linear systems using $\Delta v^{\mathrm{bare}}_{[\mathbf{q},\mathbf{G},\omega]}(\mathbf{r}')=v(\mathbf{q}+\mathbf{G})\exp(i\mathbf{G}\cdot\mathbf{r}')$. The solution of the linear systems yields the change in the wave functions, which are then used to construct the induced charge, the induced Hartree potential, and the updated screened Coulomb interaction. We repeat this procedure by starting with the updated screened Coulomb interaction until convergence is achieved. At convergence the self-consistent perturbing potential yields the screened Coulomb interaction $W_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)$. The calculation must be repeated for every perturbation, i.e., for each set of parameters $[\mathbf{q},\mathbf{G},\omega]$. At the end of this procedure it is straightforward to obtain the inverse dielectric matrix $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)$ using Eqs. (29) and (36). Alternatively, it is also possible to scale the initial perturbation and use $\Delta v^{\mathrm{bare}}_{[\mathbf{q},\mathbf{G},\omega]}=\exp(i\mathbf{G}\cdot\mathbf{r}')$ to obtain the inverse dielectric matrix at the end of the self-consistent procedure [indeed Eq. (34) is a linear system].

The scheme developed here allows us to calculate one row (in $\mathbf{G}'$) of the inverse dielectric matrix $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)$ by determining the linear response to the perturbation $\exp[-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}]$. This idea has been discussed already in Ref. 32 in the framework of nonperturbative methods based on supercell calculations.

Singularities in the inverse dielectric matrix and the screened Coulomb interaction

In order to avoid the singular behavior of the wings of the inverse dielectric matrix in the long-wavelength limit ($|\mathbf{q}|\to 0$) it is convenient to work with the symmetrized inverse dielectric matrix defined as follows:[44]

$$\tilde{\epsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega) = \epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)\frac{|\mathbf{q}+\mathbf{G}'|}{|\mathbf{q}+\mathbf{G}|}. \tag{37}$$

Unlike its unsymmetrized counterpart $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)$, the wings of $\tilde{\epsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)$ have finite limits at long wavelengths. The screened Coulomb interaction of Eq. (29) is now rewritten in symmetrized form as

$$W_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega) = \frac{4\pi e^2}{|\mathbf{q}+\mathbf{G}||\mathbf{q}+\mathbf{G}'|}\tilde{\epsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega). \tag{38}$$

While the symmetrized inverse dielectric matrix has finite limits at long wavelengths, the screened Coulomb interaction still presents a divergence corresponding to the long-range tail of the Coulomb potential in real space. This divergence requires special handling when performing the Brillouin-zone integration to calculate the GW self-energy.[2] We here overcome this difficulty following the prescription of Ref. 46. For this purpose we replace the bare Coulomb potential $v(\mathbf{r},\mathbf{r}')$ by the truncated potential $v_{\mathrm{t}}(\mathbf{r},\mathbf{r}')=v(\mathbf{r},\mathbf{r}')[1-\Theta(|\mathbf{r}-\mathbf{r}'|-R_{\mathrm{c}})]$, $\Theta(x)$ being the Heaviside step function. The truncation radius is defined as in Sec. III A. Using this truncated Coulomb potential, the final expression for the screened Coulomb interaction in reciprocal space becomes

$$W^{\mathrm{t}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega) = 4\pi e^2\frac{1-\cos R_{\mathrm{c}}|\mathbf{q}+\mathbf{G}|}{|\mathbf{q}+\mathbf{G}||\mathbf{q}+\mathbf{G}'|}\tilde{\epsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega). \tag{39}$$

In the long-wavelength limit $\mathbf{q}\to 0$ the head of the truncated screened Coulomb interaction ($\mathbf{G}=\mathbf{G}'=0$) tends to the finite limit $2\pi e^2 R_{\mathrm{c}}^2\tilde{\epsilon}^{-1}_{00}(\mathbf{q}\to 0;\omega)$ and the singular behavior is removed. Optimized truncation strategies have been developed for nonisotropic materials and systems with reduced dimensionality.[22]

B. Green's function

We now specialize Eqs. (13)–(18) to the case of a plane-waves basis and the Bloch representation. We start by rewriting Eq. (13) after relabeling the electronic states ψ_n as Bloch states $\psi_{n\mathbf{k}}$ and taking into account the normalization, as already done in Sec. III A. Next we expand the Green's function in terms of the Bloch waves $\exp[-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}]$ and $\exp(i\mathbf{k}\cdot\mathbf{r}')$,

$$G^{\mathrm{A}}_{[\mathbf{r},\omega]}(\mathbf{r}') = \frac{1}{N_{\mathbf{k}}\Omega}\sum_{\mathbf{k}\mathbf{G}} g^{\mathrm{A}}_{[\mathbf{k},\mathbf{G},\omega]}(\mathbf{r}')e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}e^{i\mathbf{k}\cdot\mathbf{r}'}, \tag{40}$$

with $g^{\mathrm{A}}_{[\mathbf{k},\mathbf{G},\omega]}(\mathbf{r}')$ cell periodic in $\mathbf{r}'$. An analogous expansion holds for the nonanalytic component G^{N}. Equations (16) and (17) are now rewritten as

$$(\hat{H}_{\mathbf{k}} - \omega^{+})g^{\mathrm{A}}_{[\mathbf{k},\mathbf{G},\omega]}(\mathbf{G}') = -\delta_{\mathbf{G}\mathbf{G}'}, \tag{41}$$

$$g^{\mathrm{N}}_{[\mathbf{k},\mathbf{G},\omega]}(\mathbf{G}') = 2\pi i\sum_{v}\delta(\omega-\varepsilon_{v\mathbf{k}})u^{\star}_{v\mathbf{k}}(\mathbf{G})u_{v\mathbf{k}}(\mathbf{G}'). \tag{42}$$

In deriving Eqs. (41) and (42) we made use once again of time-reversal symmetry, yielding $u^{\star}_{v\mathbf{k}}(\mathbf{G})=u_{v,-\mathbf{k}}(-\mathbf{G})$. Similarly to the case of the screened Coulomb interaction, by

solving the linear system in Eq. (41) for a set of parameters $[\mathbf{k},\mathbf{G},\omega]$ we obtain an entire row $\mathbf{G}'$ of the analytic component of the Green's function $g^{\mathrm{A}}_{[\mathbf{k},\mathbf{G},\omega]}(\mathbf{G}')$.

C. Self-energy

The GW self-energy in Eq. (19) is calculated in real space after performing the Fourier transforms of $W_{\mathbf{GG}'}(\mathbf{q};\omega)$ and $G_{\mathbf{GG}'}(\mathbf{k};\omega)$. The result is then transformed back in reciprocal space to obtain $\Sigma_{\mathbf{GG}'}(\mathbf{k};\omega)$. The evaluation of the matrix elements of the self-energy in the basis of Kohn-Sham eigenstates is performed in reciprocal space. Since the plane-waves cutoff required to describe the inverse dielectric matrix and the self-energy is typically much smaller than the cutoff used in density-functional calculations,[2] the procedure described here does not require an excessive computational effort and accounts for only a fraction of the total computation time.

IV. SCALING PROPERTIES

In this section we analyze the computational complexity of the algorithms proposed in Sec. III by focusing on our plane-waves implementation. Without loss of generality we consider a Γ-point sampling of the Brillouin zone and we leave aside the frequency dependence. We assume that the Kohn-Sham electronic wave functions are expanded in a basis of plane waves with a kinetic-energy cutoff $E^{\mathrm{wf}}_{\mathrm{cut}}$, corresponding to $N^{\mathrm{wf}}_{\mathbf{G}}$ plane waves. In the simplest case of norm-conserving pseudopotential approaches the electronic charge density is described using a basis set with a cutoff $E^{\mathrm{den}}_{\mathrm{cut}}=4E^{\mathrm{wf}}_{\mathrm{cut}}$, and the corresponding numbers of basis functions and real-space grid points are $N^{\mathrm{den}}_{\mathbf{G}}$ and $N^{\mathrm{den}}_{\mathbf{r}}$, respectively. The screened Coulomb interaction and the Green's function are described by a smaller cutoff $E^{\mathrm{s}}_{\mathrm{cut}}$ and $N^{\mathrm{s}}_{\mathbf{G}}$ plane waves. The self-energy is expanded in a plane-waves basis with cutoff $E^{\mathrm{SE}}_{\mathrm{cut}}=4E^{\mathrm{s}}_{\mathrm{cut}}$, and we denote by $N^{\mathrm{SE}}_{\mathbf{r}}$ the number of real-space grid points associated with this basis.

A. Screened Coulomb interaction

Equation (34) needs to be solved for each one of the $N^{\mathrm{s}}_{\mathbf{G}}$ plane-wave perturbations and the N_v occupied electronic states. For the solution of Eq. (34) we adopt the complex biorthogonal conjugate gradient (cBiCG) method of Ref. 47, as described in Appendix A. Each solution of Eq. (34) requires two cBiCG minimizations (for $\pm\omega$), and each cBiCG minimization consists of two CG sequences. The most time-consuming operation in each CG step is the application of the Hamiltonian to the previous search direction, and, in particular, the Fourier transform of the wave functions to real space and back for evaluating the product with the local potential. Fast Fourier-transform (FFT) algorithms allow us to perform these calculation in $N^{\mathrm{den}}_{\mathrm{FFT}}=4N^{\mathrm{den}}_{\mathbf{r}}\log N^{\mathrm{den}}_{\mathbf{r}}$ floating point operations.[48] If in average the CG minimization requires N_{CG} steps and the self-consistency loop requires N_{SCF} iterations, then the total cost of the entire calculation corresponds to a number of floating point operations

$$N^{\mathrm{SGW}}_{\mathrm{flops}}=8N_{\mathrm{CG}}N_{\mathrm{SCF}}N^{\mathrm{s}}_{\mathbf{G}}N_v N^{\mathrm{den}}_{\mathrm{FFT}}, \tag{43}$$

where SGW stands for "Sternheimer GW." As $N^{\mathrm{s}}_{\mathbf{G}}$, N_v, and $N^{\mathrm{den}}_{\mathbf{r}}$ scale linearly with the size of the system as measured by the number of atoms N_{at}, the overall scaling of this procedure is $N^3_{\mathrm{at}}\log N_{\mathrm{at}}$.

For comparison it is useful to consider the scaling of standard GW calculations based on the expansion over unoccupied states (hereafter referred to as the "HL" method).[2] The calculation of the irreducible RPA polarization requires the evaluation of the optical matrix elements between each of the N_v occupied states and each of the N_c unoccupied states. These matrix elements are typically computed by using Fourier transforms of $\psi^{\star}_c(\mathbf{r})\psi_v(\mathbf{r})$, therefore this procedure requires essentially N_vN_c Fourier transforms from real to reciprocal space. Each Fourier transform is performed on the real-space grid for the density with $N^{\mathrm{den}}_{\mathbf{r}}$ grid points, therefore the total cost of the standard method corresponds to a number of floating point operations

$$N^{\mathrm{HL}}_{\mathrm{flops}}=N_cN_vN^{\mathrm{den}}_{\mathrm{FFT}}. \tag{44}$$

Even in this case therefore the overall scaling is $N^3_{\mathrm{at}}\log N_{\mathrm{at}}$.

Since the method of Ref. 2 calculates the dielectric matrix and then performs a matrix inversion, in order to compare the prefactors in Eqs. (43) and (44) we consider the non-self-consistent calculation of the dielectric matrix as described in Sec. II A 2 [$N_{\mathrm{SCF}}=1$ in Eq. (43)], and we restrict ourselves to the calculation of the static dielectric matrix. In this case only one calculation of Eq. (4) is required instead of two for $\pm\omega$, and the two CG sequences of the cBiCG algorithm do coincide. As a result, a factor 4 drops out of the prefactor in Eq. (43). If we assume for definiteness a perfectly well-conditioned linear system (condition number $\kappa=1$), and express the number of CG iterations required to achieve convergence through Eq. (B1), then the ratio between the complexity of the SGW approach in a plane-waves implementation and the standard approach becomes

$$N^{\mathrm{SGW}}_{\mathrm{flops}}/N^{\mathrm{HL}}_{\mathrm{flops}}=N^{\mathrm{s}}_{\mathbf{G}}/N_c\,\log(2/\varepsilon), \tag{45}$$

where ε is the relative accuracy of the results. As an example, for a relative accuracy of $\varepsilon=10^{-5}$ we find this ratio to be $\simeq 12N^{\mathrm{s}}_{\mathbf{G}}/N_c$. In the case of silicon, using a typical cutoff $E^{\mathrm{s}}_{\mathrm{cut}}=10$ Ry we obtain $N^{\mathrm{s}}_{\mathbf{G}}=137$, therefore the SGW approach becomes convenient when more than ~ 1650 unoccupied states are used in the standard approach. This is rarely the case as most calculations reported to date use only a few hundreds of unoccupied electronic states. Of course the accuracy of the standard sum-over-states expression is difficult to quantify, and probably a convergence on five significant digits is not warranted by a few hundreds of electronic states.

Our estimate suggests that the plane-waves implementation of our method can be as expensive as the standard approach. It should be noted, however, that our scheme has the advantage of providing the whole self-energy $\Sigma(\mathbf{r},\mathbf{r}';\omega)$, while the standard approach typically provides the matrix elements of the self-energy on a small subset of states on the order of N_v. Therefore if we were to perform a comparison

based on the same amount of output information, we should use $N_c N_{\mathbf{G}}^{\mathrm{s}}$ in Eq. (44) instead of $N_c N_v$. In this case Eq. (45) would change into

$$N_{\mathrm{flops}}^{\mathrm{SGW}}/N_{\mathrm{flops}}^{\mathrm{HL}} = N_v/N_c \log(2/\varepsilon), \tag{46}$$

and for $\varepsilon = 10^{-5}$ we would have $N_{\mathrm{flops}}^{\mathrm{SGW}}/N_{\mathrm{flops}}^{\mathrm{HL}} \simeq 12 N_v/N_c$. This clearly shows that, if the entire self-energy was needed (as opposed to a few matrix elements), then our proposed SGW approach would be more convenient that the standard sum-over-states approach.

The above analysis shows that the main bottlenecks of our method are (i) the Fourier transform for the application of the Hamiltonian and (ii) the large basis sets adopted. In order to make the approach proposed here more efficient we could either move to real-space methods where the application of the Hamiltonian scales linearly with system size[49] or reduce the size of the basis set by using local orbitals.[50] Fast evaluations of the operation $\hat{H}\psi$ in order-N operations should make it feasible GW calculations with N_{at}^3 scaling and with a very favorable prefactor. We will come back to this aspect in Sec. VI.

B. Green's function

The complexity of the procedure for calculating the Green's function proposed in Sec. II B can be analyzed along the same lines of Sec. IV A. The main differences in this case are that (i) the linear system Eq. (18) does not depend on the occupied states, (ii) the calculation is non-self-consistent, and (iii) the calculation is performed for one single frequency ω^+, while the entire frequency dependence is generated through the multishift method. As a result, a factor $2N_{\mathrm{SCF}}N_v$ drops out of Eq. (43) and the computational cost of the Green's function calculation reads

$$N_{\mathrm{flops}}^{\mathrm{GF}} = 4 N_{\mathrm{CG}} N_{\mathbf{G}}^{\mathrm{s}} N_{\mathrm{FFT}}^{\mathrm{den}}. \tag{47}$$

The complexity of this calculation is significantly smaller than the complexity of the algorithm for the screened Coulomb interaction. In particular, the calculation of the Green's function scales as $N_{\mathrm{at}}^2 \log N_{\mathrm{at}}$. This procedure for calculating the Green's function is advantageous with respect to an expansion over empty states, as the orthogonalization of the unoccupied states would require a number of floating point operations scaling as $\sim N_{\mathrm{at}}^3$.

C. Scaling of the self-energy calculation

The self-energy is computed in real space after obtaining $G(\mathbf{r},\mathbf{r}';\omega)$ and $W(\mathbf{r},\mathbf{r}';\omega)$ from $G(\mathbf{G},\mathbf{G}';\omega)$ and $W(\mathbf{G},\mathbf{G}';\omega)$, respectively, and then is transformed back into reciprocal space. The six-dimensional FFT transforms require $(N_{\mathbf{G}}^{\mathrm{s}}+N_{\mathbf{r}}^{\mathrm{SE}})N_{\mathrm{FFT}}^{\mathrm{SE}}$ operations for each frequency of the screened Coulomb interaction, having defined $N_{\mathrm{FFT}}^{\mathrm{SE}} = 4 N_{\mathbf{r}}^{\mathrm{SE}} \log N_{\mathbf{r}}^{\mathrm{SE}}$. The computational cost of this procedure scales as $N_{\mathrm{at}}^2 \log N_{\mathrm{at}}$, and is small with respect to the cost of calculating the screened Coulomb interaction.

TABLE I. Long-wavelength limit of the static symmetrized inverse dielectric matrix of silicon $\tilde{\epsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)$ [$\mathbf{q}=(0.01,0,0)2\pi/a$ and $\omega=0$]. We compare our calculations performed within the self-consistent Sternheimer approach with the results obtained in Ref. 44 using the expansion over empty states and the inversion of the dielectric matrix. For the calculations we sampled the Brillouin zone with 29 irreducible $\mathbf{k}$ points, corresponding to a $8\times8\times8$ grid (Refs. 44 and 45), and a plane-wave cutoff of 5 Ry (Ref. 44). Following Ref. 44 we employed the empirical pseudopotential parameters from Ref. 30. The reciprocal-lattice vectors are in units of $2\pi/a$, a being the lattice parameter.

		$\tilde{\epsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega)$	
$\mathbf{G}$	$\mathbf{G}'$	Ref. 44	Present work
(0,0,0)	(0,0,0)	0.083	0.0866
(1,1,1)	(1,1,1)	0.605	0.6055
$(\bar{1},1,1)$	(1,1,1)	0.008	0.0076
$(1,\bar{1},1)$	$(\bar{1},1,1)$	0.010	0.0102
$(1,\bar{1},\bar{1})$	$(\bar{1},1,1)$	0.045	0.0463
(2,0,0)	(1,1,1)	−0.038	−0.0382
(2,0,0)	$(\bar{1},1,1)$	−0.005	−0.0049
(2,0,0)	(2,0,0)	0.667	0.6671
$(\bar{2},0,0)$	(2,0,0)	0.006	0.0063
(0,2,0)	(2,0,0)	0.016	0.0166

V. RESULTS

In order to demonstrate the approach proposed in Secs. II and III we have realized a prototype implementation within the empirical pseudopotential method (EPM) of Ref. 30, and we have validated our implementation for the test case of silicon.

A. Dielectric matrix

Table I contains some of the components of the inverse dielectric matrix calculated using the self-consistent Sternheimer method described in Sec. III. In all our calculations we used inverse dielectric matrices of size 59×59, corresponding to a kinetic-energy cutoff of 5 Ry for the screened Coulomb interaction. For the purpose of comparison with Ref. 44 we calculated the static and long-wavelength limit ($\omega=0$, $\mathbf{q}\rightarrow0$) of the inverse dielectric matrix for the first few reciprocal-lattice vectors. The authors of Ref. 44 adopted the standard approach based on the expansion over the unoccupied electronic states of the dielectric matrix and obtained the inverse dielectric matrix by performing matrix inversions. In our calculations we used the self-consistent method of Sec. III and no matrix inversion was necessary. The excellent agreement which can be seen in Table I between our calculations and Ref. 44 supports the validity of our approach.

Next we consider the wave-vector dependence of the head of the dielectric matrix $\varepsilon_{00}(\mathbf{q},\omega=0)$. We performed the calculation by using the non-self-consistent method described in Sec. II A 2 in order to compare our results with Ref. 51.

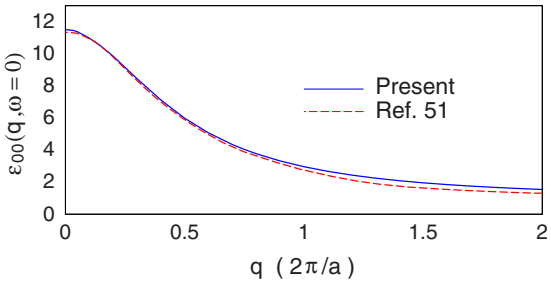

FIG. 1. (Color online) Dielectric function of silicon calculated using the empirical pseudopotential method and the non-self-consistent Sternheimer method of Sec. II A 2: calculated head of the static dielectric function as a function of wave vector $\epsilon_{00}(\mathbf{q},\omega=0)$ (blue solid line) and results from Ref. 51 (red dashed line). We used a plane-wave kinetic-energy cutoff of 5 Ry and sampled the Brillouin zone through a uniform $8\times8\times8$ grid. The wave vectors are in units of $2\pi/a$, a being the lattice parameter.

Figure 1 shows that our calculations are in very good agreement with the results of Ref. 51. The slight differences between our results and those of Ref. 51 at large wave vectors can likely be ascribed to the use of a limited number of empty states in the perturbative expansion over unoccupied states in the latter work.

Figure 2 compares our results for the frequency dependence of the dielectric matrix at long wavelength with the results reported in Ref. 2. We focused, in particular, on the cases illustrated in Fig. 3 of Ref. 2. Apart from some small differences possibly arising from the use of the expansion over unoccupied states in Ref. 2, even in this case the agreement between our calculations and those of Ref. 2 is very good throughout the entire frequency range. The agreement is consistently good for the head of the dielectric matrix and for diagonal and off-diagonal matrix elements.

B. Padé approximants and convergence

Figure 3 shows the quality of the analytic continuation from the imaginary to the real frequency axis using Padé approximants (cf. Appendix C).[52,53] We found that this procedure based on Padé approximants is generally very stable and requires minimal manual intervention. Approximants of order 5 and higher are able to reproduce the location, the strength, and the width of the main plasmon-pole structure of the dielectric matrix. Head, wings, and body of the dielectric matrix are all described consistently (cf. Fig. 3). Although the singularity corresponding to the absorption onset in Fig. 3(a) is smoothed out by Padé approximants of low order, this effect is washed out when calculating the frequency convolution of the Green's function with the screened Coulomb interaction for the GW self-energy. The advantages of performing calculations along the imaginary axis are that (i) the linear system in Eq. (4) becomes increasingly more well conditioned when approaching large imaginary frequencies, and (ii) a moderate Brillouin-zone sampling is required to perform calculations along the imaginary axis unlike the case of real-axis calculations. As a result, the worst case scenario for the solution of the linear system Eq. (4) corresponds to the

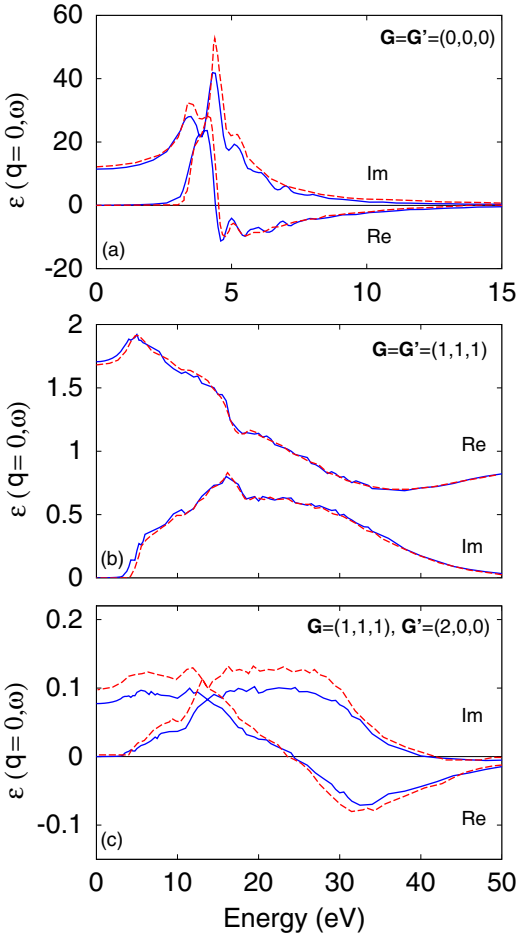

FIG. 2. (Color online) Frequency-dependent dielectric matrix of silicon at long wavelength $[\mathbf{q}=(0.01,0,0)2\pi/a]$. The calculations were performed using the empirical pseudopotential method of Ref. 30 and the non-self-consistent Sternheimer method of Sec. II A 2. We used a plane-wave kinetic-energy cutoff of 5 Ry and sampled the Brillouin zone through a uniform $24\times24\times24$ grid. Such a dense Brillouin-zone sampling was necessary to correctly describe the absorption onset. In order to avoid null eigenvalues in the linear system, Eq. (11), we performed the calculations by including a small imaginary component of 0.1 eV in the frequency ω. The panels (a)–(c) correspond to the cases illustrated in Fig. 3 of Ref. 2 and show $\epsilon_{\mathbf{GG}'}(\mathbf{q}\to0,\omega)$ for (a) $\mathbf{G}=\mathbf{G}'=0$, (b) $\mathbf{G}=\mathbf{G}'=(1,1,1)2\pi/a$, and (c) $\mathbf{G}=(1,1,1)2\pi/a$, $\mathbf{G}'=(2,0,0)2\pi/a$. The blue solid lines are our calculations, the red dashed lines are from Ref. 2. The real and imaginary parts of the dielectric matrix are indicated in each panel. We note that the scales on the vertical axes correspond to three different orders of magnitude.

static case $\omega=0$. These technical aspects are described in detail in Appendix B.

The typical number of non-self-consistent iterations required to solve Eq. (4) for a fixed $\Delta V_{[\mathbf{r},\omega]}$ with a relative accuracy of $\varepsilon_{\mathrm{NSCF}}=10^{-10}$ using the cBiCG algorithm described in Appendix A is $N_{\mathrm{CG}}\simeq21$ (using the preconditioner of Ref. 54). This estimate has been obtained by averaging over all the $\mathbf{G},\mathbf{G}'$ reciprocal-lattice vectors, $\mathbf{q}$ vectors, and imaginary frequencies. The typical number of self-consistent cycles required to obtain the screened Coulomb interaction through Eq. (4) with a relative accuracy of $\varepsilon_{\mathrm{SCF}}=10^{-5}$ is

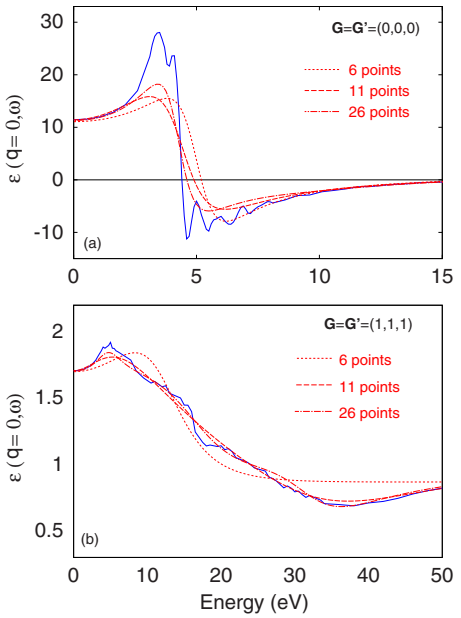

FIG. 3. (Color online) Real part of the long-wavelength dielectric matrix of silicon as a function of frequency. Panels (a) and (b) of this figure correspond to panels (a) and (b) of Fig. 2, respectively. The technical details of the calculations are the same as those described in the caption of Fig. 2. Solid blue lines: dielectric matrices calculated directly along the real frequency axis, from Fig. 2. Dotted, dashed, and dash-dotted red lines: dielectric matrices obtained from the analytic continuation on the real axis using Padé approximants of order 6, 11, and 26, respectively. The Padé approximants were generated using dielectric matrices calculated along the imaginary frequency axis on uniform frequency grids in the range 0–50 eV. For instance, the six-points approximant corresponds to calculations at the imaginary frequencies of $0, 10, \ldots, 50$ eV.

$N_{\mathrm{SCF}} \simeq 5$. Charge-sloshing effects are attenuated by using the potential mixing method proposed in Ref. 55, appropriately modified to deal with complex potentials.

For completeness we report here the corresponding figures for the calculation of the Green's function using Eq. (18). The average number of non-self-consistent iterations required to obtain the analytic part of the Green's function is $N_{\mathrm{GF}} \simeq 25$ when preconditioning is adopted (for this purpose we used a straightforward adaptation of the method of Ref. 54). However, multishift minimizations as described in Appendix D do not allow for the use of preconditioning, and in the latter case the number of iterations required to achieve convergence (within a relative accuracy $\varepsilon_{\mathrm{NSCF}} = 10^{-10}$) can be as high as $N_{\mathrm{GF}} \simeq 100$.

C. Self-energy

Figures 4 and 5 show the real part $\mathrm{Re}\langle n\mathbf{k}|\Sigma|n\mathbf{k}\rangle$ and the imaginary part $\mathrm{Im}\langle n\mathbf{k}|\Sigma|n\mathbf{k}\rangle$ of the GW self-energy calculated for the first few silicon eigenstates at Γ using our SGW method within the EPM scheme. Our results are compared with the calculations of Ref. 18 performed within DFT/LDA and the projector-augmented wave method (PAW). We calculated the screened Coulomb interaction by using a uniform $6 \times 6 \times 6$ grid to sample the Brillouin zone and Padé approxi-

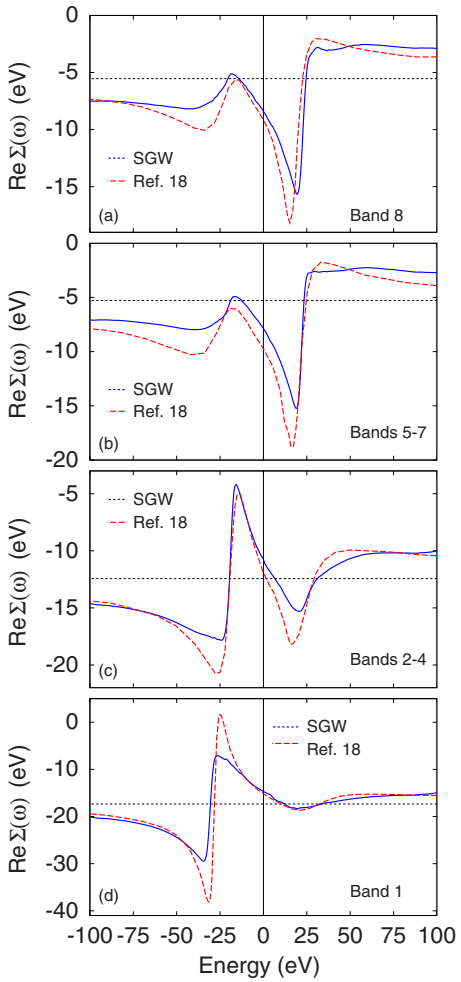

FIG. 4. (Color online) Real part $\mathrm{Re}\langle n\mathbf{k}|\Sigma(\mathbf{r},\mathbf{r}';\omega)|n\mathbf{k}\rangle$ of the expectation value of the GW self-energy for the first eight silicon eigenstates at $\mathbf{k}=0$. The solid blue lines are our SGW results using the empirical pseudopotential method. The dashed red lines are from the first-principles calculation of Ref. 18 using the LDA and the PAW method. Panels (a)–(d) correspond to the states Γ'_{2c} (Band 1), Γ_{15c} (Bands 2–4), Γ'_{25v} (Bands 5–7), and Γ_{1v} (Band 8), respectively. The energy axis is aligned with the top of the valence band. The horizontal dotted lines indicate the calculated bare exchange contribution to the self-energy $\mathrm{Re}\langle n\mathbf{k}|\Sigma^{\mathrm{ex}}(\mathbf{r},\mathbf{r}';\omega)|n\mathbf{k}\rangle$.

mants of order 7 along the imaginary frequency axis. The frequency integration of the Coulomb term Σ^{c} of the GW self-energy in Eq. (21) was performed by using a Coulomb cutoff $\omega_{\mathrm{C}} = 100$ eV and a grid spacing of 0.5 eV. The Green's function was calculated using an imaginary component $\eta = 0.3$ eV in Eq. (18). Apart from some differences in the damping of the plasmaron peaks, the agreement between our calculated self-energy and the results of Ref. 18 is rather good throughout the entire frequency range ± 100 eV. This finding is quite surprising since we are comparing our empirical pseudopotential calculations with low kinetic-energy cutoff (5 Ry) with the *ab initio* LDA calculations including PAW core reconstruction of Ref. 18. Such agreement probably reflects the ability of the EPM method to provide not only a good description of the band structure of silicon but also a reasonable description of the electronic wave functions.

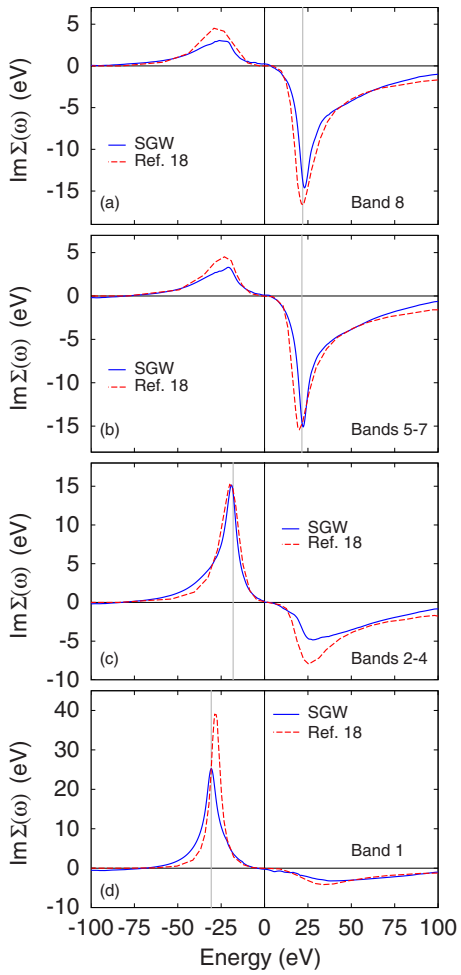

FIG. 5. (Color online) Imaginary part $\mathrm{Im}\langle n\mathbf{k}|\Sigma(\mathbf{r},\mathbf{r}';\omega)|n\mathbf{k}\rangle$ of the expectation value of the GW self-energy for the first eight silicon eigenstates at $\mathbf{k}=0$. The solid blue lines are our SGW results using the empirical pseudopotential method. The dashed red lines are from the first-principles calculation of Ref. 18. Panels (a)–(d) correspond to the states Γ'_{2c}, Γ_{15c}, Γ'_{25v}, and Γ_{1v}, respectively. The energy axis is aligned with the top of the valence band. For comparison, the vertical gray lines indicate the locations of the logarithmic singularities at $\varepsilon_{n\mathbf{k}} \pm \omega_p$ (ω_p being the plasmon energy) that would arise using a model plasmon-pole dielectric function (Ref. 20).

D. Quasiparticle excitations and spectral function

Within the GW method the values of the quasiparticle excitation energies are typically calculated by using first-order perturbation theory on the DFT eigenvalues.[2] The perturbation operator $\Delta\Sigma(\mathbf{r},\mathbf{r}';\omega)$ corresponds to the difference between the GW self-energy $\Sigma(\mathbf{r},\mathbf{r}';\omega)$ and the DFT exchange and correlation potential $V^{\mathrm{xc}}(\mathbf{r})$. This approach is sensible because the complete quasiparticle equations are similar to the ordinary Kohn-Sham equations if we replace the self-energy Σ by the DFT XC potential V^{xc}.[20]

Within the EPM scheme the total potential V^{EPM} acting on the electrons is specified[30] but the electronic charge density is not connected to this potential through a self-consistent procedure.[56] This limitation makes it difficult to identify an XC contribution within the empirical pseudopotential. However, since the charge density obtained within the EPM method can be regarded as an approximation to the actual charge density,[57] it appears sensible to obtain the effective XC potential as a functional of the EPM charge density using the local-density approximation.[58,59] This procedure is formally equivalent to assuming that the unscreened ionic pseudopotential V^{ion} is given by $V^{\mathrm{EPM}}-V^{\mathrm{Ha}}-V^{\mathrm{xc}}$, where the Hartree potential V^{Ha} and the XC potential V^{xc} are calculated using the EPM charge density. *This uncertainty on the* XC *potential renders the calculation of the quasiparticle excitation energies somewhat arbitrary, therefore the results presented in the following should be regarded as qualitative and are presented only for the purpose of demonstrating a complete* GW calculation within our SGW methodology.

Despite the above limitations, the expectation values of the XC potential and of the exchange term of the GW self-energy calculated here are surprisingly close to those obtained in Ref. 2 using *ab initio* pseudopotentials at the DFT/LDA level. Indeed, for the valence-band top Γ'_{25v} state and the conduction-band bottom close to the X_{1c} state our calculated XC expectation values are −11.27 and −8.97 eV, respectively, while Ref. 2 gives −11.80 and −9.61 eV for the corresponding states at the DFT/LDA level. The agreement is even better when comparing the expectation values of the bare exchange part of the GW self-energy. In this case we find −12.43 and −5.07 eV for the Γ'_{25v} state and the X_{1c} state, respectively, to be compared to the corresponding values of −12.54 and −5.28 eV of Ref. 2. These results provide an *a posteriori* justification to our choice of calculating the XC potential using the EPM charge density and the LDA functional.

Table II compares our calculated quasiparticle excitation energies for electronic states at the Γ point and for the conduction-band edge of silicon with the results of Refs. 2 and 18. We find a good overall agreement between these different sets of calculations. Taking into account that we are comparing our SGW scheme within the EPM implementation with more sophisticated *ab initio* calculations, such agreement is rather encouraging.

Figure 6 shows the calculated quasiparticle spectral function

$$\langle n\mathbf{k}|A|n\mathbf{k}\rangle = \frac{|\mathrm{Im}\,\Delta\Sigma_{n\mathbf{k}}|}{|\omega-\varepsilon_{n\mathbf{k}}-\mathrm{Re}\,\Delta\Sigma_{n\mathbf{k}}|^2+|\mathrm{Im}\,\Delta\Sigma_{n\mathbf{k}}|^2} \quad (48)$$

with $A(\mathbf{r},\mathbf{r}';\omega)$ denoting the GW spectral function and $\Delta\Sigma_{n\mathbf{k}}=\langle n\mathbf{k}|\Sigma-V_{\mathrm{xc}}|n\mathbf{k}\rangle$. We note that Eq. (48) is obtained by assuming that the off-diagonal matrix elements of the self-energy in the basis of the unperturbed wave functions are negligible. A diagonal approximation is not always justified, nonetheless we decided to adopt Eq. (48) to be consistent with Ref. 18. Figure 6 shows good overall agreement between our calculations and the LDA/PAW results of Ref. 18. This comparison demonstrates once again the validity of our methodology.

TABLE II. GW quasiparticle excitation energies for the Γ'_{2c}, Γ_{15c}, Γ'_{25v}, Γ_{1v} states of silicon, as well as for the conduction band bottom of silicon close to the X_{1c} state. We obtained the quasiparticle energies $E_{n\mathbf{k}}$ by solving the nonlinear equations $E_{n\mathbf{k}}=\varepsilon_{n\mathbf{k}}+\mathrm{Re}[\Delta\Sigma_{n\mathbf{k}}(E_{n\mathbf{k}})]$ with $\Delta\Sigma_{n\mathbf{k}}=\langle n\mathbf{k}|\Sigma-V_{\mathrm{xc}}|n\mathbf{k}\rangle$. For completeness we also show the unperturbed eigenvalues calculated here within the EPM scheme, and those of Ref. 18 within the DFT/LDA scheme for comparison. It is interesting to observe that the value of the minimum band gap is not increased from the EPM value of 0.82 eV when applying the GW correction. In absolute terms the X_{1c} state shifts upwards by 0.41 eV upon the GW correction but this shift is compensated by the concurrent upward shift of the Γ'_{25v} state by 0.44 eV. This problem relates to the uncertainty on the XC potential within the EPM formalism, as discussed in Sec. V D.

	EPM/LDA eigenvalues		Quasiparticle energies			
	Present	Ref. 18	Present	Ref. 2	Ref. 18	Expt.
Γ'_{2c}	3.89	3.23	4.21	4.08	4.05	4.23[a],4.1[b]
Γ_{15c}	3.42	2.54	3.53	3.35	3.09	3.40[a],3.05[b]
Γ'_{25v}	0.00	0.00	0.00	0.00	0.00	0
Γ_{1v}	−12.62	−11.97	−13.23	−12.04	−11.85	-12.5 ± 0.6 [a]
X_{1c}	0.82	0.55	0.79	1.29	0.92	1.17[a]

[a]Reference 60.
[b]Reference 61.

VI. CONCLUSIONS AND OUTLOOK

The results presented in Sec. V demonstrate the feasibility of the self-consistent Sternheimer approach to GW calculations in the simple case of a prototype implementation based on the empirical pseudopotential method. The extension of the present methodology to *ab initio* approaches based on norm-conserving[62] or ultrasoft[63] pseudopotentials should not present any difficulties as the crucial issues in the calculation have already been addressed in this work. The main advantage of the present methodology consists of the definitive elimination of the unoccupied electronic states from the calculations of *both* the screened Coulomb interaction and the noninteracting Green's function. Another appealing aspect is that our methodology constitutes a generalization to frequency-dependent perturbations of density-functional perturbation theory,[28] which is a well-established technique with a long history of successes.

As discussed in Sec. IV the present approach is comparable in performance to standard[2] GW techniques. The question remains on whether it is possible to make significant improvements over the methodology proposed here *without compromising on the numerical accuracy*. The most time-consuming step of the entire procedure is the application of the single-particle Hamiltonian $\hat{H}$ to a search direction ψ in the Hilbert space spanned by the wave-function basis set during the iterative solution of the linear systems in Eq. (4). In order to accelerate this operation there are possibly three ways ahead: (i) the improvement of the minimization algorithms adopted, (ii) the use of sparse representations of the Hamiltonian, and (iii) the use of smaller basis sets.

(i) The iterative solution of Eq. (4) is here performed by first solving for $\Delta\psi_v^{+}$ at fixed ΔV, and then by updating ΔV in the self-consistency cycle. It should be possible, at least in principle, to combine these two operations in a single minimization step. This could be achieved, for instance, by using the variational formulation of density-functional perturbation

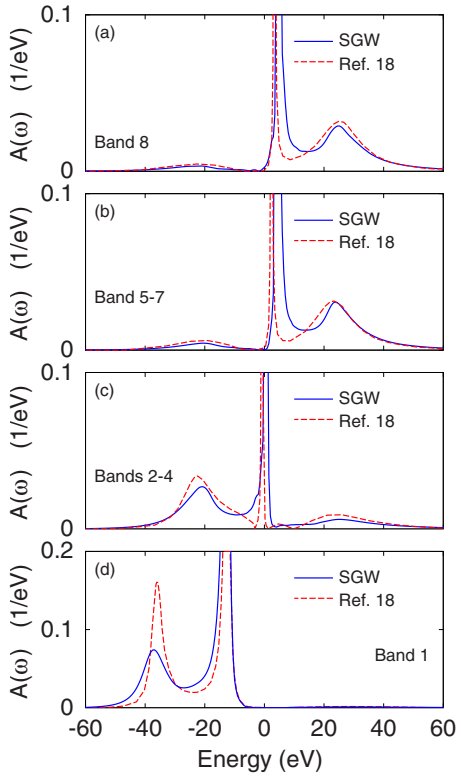

FIG. 6. (Color online) Expectation values of the quasiparticle spectral function $\langle n\mathbf{k}|A(\mathbf{r},\mathbf{r}';\omega)|n\mathbf{k}\rangle$ for the first eight silicon eigenstates at $\mathbf{k}=0$. The solid blue lines are our SGW results using the EPM, the dashed red lines are from the first-principles calculation of Ref. 18 using the LDA and the PAW method. The expectation values have been calculated within the diagonal approximation of Eq. (48). In each panel the sharp peak near the band extrema corresponds to a well-defined quasiparticle while the two broad peaks corresponds to plasmarons, i.e., electrons or holes coupled to a cloud of real plasmons (Ref. 20). The suppression of one of the plasmaron peaks reflects the large imaginary parts of the self-energy in the corresponding panels of Fig. 5. We point out the different vertical scale in panel (d).

theory developed in Ref. 64, or by using simulated annealing techniques such as the Car-Parrinello method.[65] Both of these approaches were developed for Hermitian systems, hence appropriate generalizations to non-Hermitian systems would be required to solve Eq. (4). As mentioned in Sec. V B the total number of cBiCG iterations required to solve Eq. (4) is typically $N_{\mathrm{CG}}N_{\mathrm{SCF}}\sim 100$, therefore the variational formulation of DFPT or the Car-Parrinello minimization would be convenient if they resulted in a significant reduction in the number of iterations over this figure.

(ii) Another possibility for improving the procedure presented in this work is to resort to a real-space representation of the kinetic-energy operator in the single-particle Hamiltonian. This can be achieved by calculating the kinetic energy on a real-space mesh using finite-differences methods.[49] The main advantage of this approach would be to have GW calculations scaling with the cube of the system size N_{at}^3 instead of $N_{\mathrm{at}}^3 \log N_{\mathrm{at}}$. However, in actual calculations the numerical prefactor associated with this scaling could be unfavorable. Indeed, the fast Fourier transforms used in a plane-waves representation to calculate the product of the potential and the wave functions require $2\times 4N_{\mathbf{r}}^{\mathrm{den}} \log N_{\mathbf{r}}^{\mathrm{den}}$ floating-point operations while the cost of the real-space calculation of the Laplacian operator to sixth order is $37N_{\mathbf{r}}^{\mathrm{den}}$ (for a supercell with orthogonal axes).[49] While these estimates seem to speak in favor of the real-space method, it is advisable to consider that the preconditioner used here[54] for the cBiCG minimization cannot be simply adapted to real-space calculations. In absence of effective preconditioners the plane-wave method remains advantageous for essentially any relevant system size.

(iii) Another interesting option for improving our methodology is to drastically reduce the size of the basis set adopted. This could be achieved, for instance, by adapting our implementation to electronic-structure packages exploiting local-orbitals basis sets.[50] Interestingly the three possibilities here outlined are not mutually exclusive, and probably an appropriate combination of all of these would eventually open the way to the study of electronic excitations in very large systems using the GW method.

In summary, we propose a methodology for performing GW calculations using the self-consistent Sternheimer equation (SGW). We show how to calculate the screened Coulomb interaction and the noninteracting Green's function without resorting to unoccupied electronic states. We successfully demonstrate our method within a plane-waves empirical pseudopotential implementation and compare with previous studies for the prototypical test case of silicon. In our method the standard generalized plasmon-pole approximation for the frequency dependence of the screened Coulomb interaction has been replaced by a direct calculation along the imaginary frequency axis, followed by an analytic continuation to the real axis. In addition, we introduce the use of multishift linear system solvers for the simultaneous calculation of multiple frequency responses at the cost of one single iterative minimization.

It is our plan to adapt the present approach to deal with first-principles pseudopotentials and to explore the performance of our procedure in a local orbital real-space implementation. We hope that this work will stimulate further effort to develop improved methodologies for excited-states calculations in large systems.

ACKNOWLEDGMENTS

F.G. is grateful to Manish Jain for drawing Ref. 46 to his attention. This work was supported by National Science Foundation under Grant No. DMR07-05941 and by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant Agreement No. 239578. Computational resources were provided by the Oxford Supercomputing Centre.

APPENDIX A: PRECONDITIONED COMPLEX BICONJUGATE GRADIENT METHOD

In this work we solve the linear systems in Eq. (4) using the complex biorthogonal variant (cBiCG) of the conjugate gradient method (CG) following Ref. 47. This method is an extension of the standard conjugate gradients algorithm to the case of general complex matrices. As the cBiCG algorithm introduced in Ref. 47 did not include preconditioning, in this appendix we describe the preconditioned version of the algorithm which we derived following Ref. 66. We are interested in solving the linear system

$$Ax = b \tag{A1}$$

with A a complex linear operator (not necessarily Hermitian), b a complex vector, and x the unknown solution vector. In the cBiCG algorithm of Ref. 47 two sequences of residuals r_n and $\tilde{r}_n$ are generated in such a way that successive residuals are biorthogonal [i.e., $\langle r_{n+1}|\tilde{r}_n\rangle=0$ and $\langle \tilde{r}_{n+1}|r_n\rangle=0$]. Two sequences of search directions p_n and $\tilde{p}_n$ are generated so that successive directions are biconjugate [i.e., $\langle Ap_{n+1}|\tilde{p}_n\rangle=0$ and $\langle A^\dagger \tilde{p}_{n+1}|p_n\rangle=0$]. The algorithm starts by setting the initial residuals to $r_0=b-Ax_0$ (x_0 being the initial guess for the solution vector x) and $\tilde{r}_0=r_0^\star$, and the initial search directions to $p_0=r_0$ and $\tilde{p}_0=p_0^\star$. Next the solution vector, the search directions, and the residuals are updated at each iteration as follows:

$$\alpha_n = \langle \tilde{r}_n|r_n\rangle/\langle \tilde{p}_n|Ap_n\rangle, \tag{A2}$$

$$x_{n+1} = x_n + \alpha_n p_n, \tag{A3}$$

$$r_{n+1} = r_n - \alpha_n A p_n, \tag{A4}$$

$$\tilde{r}_{n+1} = \tilde{r}_n - \alpha_n^\star A^\dagger \tilde{p}_n, \tag{A5}$$

$$\beta_n = -\langle A^\dagger \tilde{p}_n|r_{n+1}\rangle/\langle \tilde{p}_n|Ap_n\rangle, \tag{A6}$$

$$p_{n+1} = r_{n+1} + \beta_n p_n, \tag{A7}$$

$$\tilde{p}_{n+1} = \tilde{r}_{n+1} + \beta_n^\star \tilde{p}_n. \tag{A8}$$

The time-consuming step in this algorithm corresponds to the application of the operators A and $A^\dagger$ to the search directions p_n and $\tilde{p}_n$. As there are two such operations per iteration, the computational complexity is twice that of the standard conjugate gradient algorithm.

The preconditioning of the linear operator can be achieved by left multiplying the linear system in Eq. (A1) by $M^{-1}: M^{-1}Ax = M^{-1}b$. If we assume that the preconditioner M can be written as $M = E^{\mathrm{T}}E$, then we can rewrite the system as follows:

$$E^{-1}AE^{-\mathrm{T}}E^{\mathrm{T}}x = E^{-1}b. \tag{A9}$$

By defining $A' = E^{-1}AE^{-\mathrm{T}}$, $x' = E^{\mathrm{T}}x$, and $b' = E^{-1}b$ we obtain the transformed system $A'x' = b'$, for which the standard cBiCG method applies. While this procedure is formally correct, it is not convenient to explicitly transform the linear operator. It is convenient instead to rewrite the procedure in terms of A, b, and x by performing a few formal manipulations. For this purpose we make the substitutions $r' = E^{-1}r$ and $p' = E^{\mathrm{T}}p$. Some algebra leads straightforwardly to the preconditioned version of the cBiCG algorithm,

$$\alpha_n = \langle \tilde{r}_n | M^{-1} r_n \rangle / \langle \tilde{p}_n | A p_n \rangle, \tag{A10}$$

$$x_{n+1} = x_n + \alpha_n p_n, \tag{A11}$$

$$r_{n+1} = r_n - \alpha_n A p_n, \tag{A12}$$

$$\tilde{r}_{n+1} = \tilde{r}_n - \alpha_n^\star A^\dagger \tilde{p}_n, \tag{A13}$$

$$\beta_n = -\langle A^\dagger \tilde{p}_n | M^{-1} r_{n+1} \rangle / \langle \tilde{p}_n | A p_n \rangle, \tag{A14}$$

$$p_{n+1} = M^{-1} r_{n+1} + \beta_n p_n, \tag{A15}$$

$$\tilde{p}_{n+1} = M^{-1} \tilde{r}_{n+1} + \beta_n^\star \tilde{p}_n. \tag{A16}$$

The preconditioned cBiCG algorithm needs to be initialized with $r_0 = b - Ax_0$, $p_0 = M^{-1}r_0$, $\tilde{r}_0 = r_0^\star$, and $\tilde{p}_0 = p_0^\star$. In this work we have used the Teter-Payne-Allan function as the preconditioner M^{-1}.[54]

APPENDIX B: CONDITION NUMBER OF THE LINEAR SYSTEM

1. Screened Coulomb interaction

The iterative calculation of the screened Coulomb interaction through Eq. (4) at finite real frequencies ω can be considerably more time consuming than in the static ($\omega = 0$) case. Simple tests indicate that the number of iterations required to achieve convergence increases with increasing frequency ω. This behavior suggests that the linear operator in Eq. (4) becomes progressively more ill conditioned as the frequency ω increases.

In order to rationalize this observation, we here examine the condition number of the linear operator $\hat{H} - \varepsilon_v \pm \omega$ in Eq. (4). The minimum number of iterations $N_{\min}$ required for the solution of a linear system using the conjugate-gradients algorithm is given by

$$N_{\min} = \frac{1}{2}\sqrt{\kappa}\log(2/\varepsilon), \tag{B1}$$

κ being the condition number of the linear operator and ε the desired relative accuracy.[66] While the estimate Eq. (B1) has been derived for the original CG algorithm, we found empirically that it also provides a reasonable description of the convergence rate of the complex cBiCG version. The condition number κ of a linear operator can be calculated as the ratio of its largest to smallest eigenvalues. For a given valence state $|v'\rangle$ the linear operator $\hat{H} - \varepsilon_{v'} + \alpha \hat{P}_{\mathrm{occ}} - \omega$ in Eq. (7) has the eigenvalues $\varepsilon_v - \varepsilon_{v'} + \alpha - \omega$ and $\varepsilon_c - \varepsilon_{v'} - \omega$.

Let consider first the simplest case where $\omega = 0$ and $\alpha > 0$. In this case we find by inspection that the smallest eigenvalue is $\min(E_{\mathrm{g}}, |\alpha - W_{\mathrm{occ}}|)$, E_{g} being the fundamental energy gap and W_{occ} the valence bandwidth. It is common practice to set $\alpha = 2W_{\mathrm{occ}}$ to avoid null eigenvalues.[28] With this choice the smallest eigenvalue becomes E_{g}. On the other hand, the largest eigenvalue can be approximated by the cutoff energy of the wave-function basis set E_{cut}. In this case the condition number reads $\kappa = E_{\mathrm{cut}}/E_{\mathrm{g}}$. As an example, if we are using a plane-waves basis with a kinetic-energy cutoff of 30 Ry, we have an electron energy gap of 1 eV, and the our desired accuracy is $\varepsilon = 10^{-10}$, then according to Eq. (B1) the minimum number of iterations required to solve the linear system would be $N_{\min} = 240$. Empirical tests show that this estimate is quite accurate for the system considered in the present work. In order to improve the convergence rate it is useful to employ preconditioning techniques. We here adopt the Teter-Payne-Allan preconditioner[54] in order to "compress" the eigenvalue spectrum and thereby reduce the condition number. An ideal preconditioner would make the linear operator perfectly well conditioned ($\kappa = 1$). In this case the optimal number of iterations (for a relative accuracy $\varepsilon = 10^{-10}$) would be as small as $N_{\mathrm{min,pc}} = 12$. We have found empirically that by using the Teter-Payne-Allan preconditioner the number of iterations required to achieve convergence was in all cases in the range $N_{\mathrm{TPA}} = 15$–40.

We now consider the case of $\omega < 0$. Simple algebra shows that in this case $\kappa = E_{\mathrm{cut}}/(E_{\mathrm{g}} + w)$ when $\alpha = 2W_{\mathrm{occ}}$. Hence in this case the larger the frequency ω, the better conditioned the linear system. We checked this result by explicit calculations.

The worst case in terms of condition number is found when $\omega > 0$. In fact, as soon as the frequency exceeds the optical excitation threshold $\omega > E_{\mathrm{g}}$, the linear operator acquires null eigenvalues corresponding to the resonance condition $\omega = \varepsilon_c - \varepsilon_{v'}$. In this latter case the condition number $\kappa(\omega)$ exhibits significant structure, reflecting the joint density of states of the system. Even after preconditioning the system, the number of iterations required to achieve convergence can be as high as $N_{\min} = 500$, thus rendering this avenue unpractical. The calculation of the screened Coulomb interaction for frequencies slightly off the real axis $\omega + i\eta$, with $\omega > 0$ and small η, leads to only a small improvement of the convergence rate. The difficulty of solving iteratively the linear system Eq. (4) for large positive frequencies is accompanied by the additional difficulty of adequately sam-

pling the Brillouin zone to describe the singularities at $\omega = \varepsilon_c - \varepsilon_{v'}$.

Altogether these considerations suggest that an iterative solution of the linear system along the real axis is not convenient from the computational viewpoint. For this reason we decided to evaluate the screened Coulomb interaction along the imaginary axis and then to analytically continue the functions to the real axis using *Padé approximants*.[18,52,53] The motivation behind our choice becomes obvious after considering a simple plasmon-pole model of the screened Coulomb interaction,[2]

$$W(\omega) = v + \frac{W_0 - v}{2}\left[\frac{\omega_p}{\omega + \omega_p} - \frac{\omega_p}{\omega - \omega_p}\right], \tag{B2}$$

where ω_p is the pole frequency and W_0 the static screened Coulomb interaction. Analytical continuation of this function to the imaginary axis yields

$$W(\omega = i\beta) = v + \frac{W_0 - v}{1 + (\beta/\omega_p)^2}. \tag{B3}$$

Equation (B3) indicates that the screened Coulomb interaction along the imaginary axis contains the same amount of information as the one on the real axis (ω_p and W_0), and at the same time does not exhibit any singularities. In this case the condition number reads (assuming no preconditioning and $\alpha=0$ for simplicity) $\kappa=[(E_g^2+\beta^2)/(E_{cut}^2+\beta^2)]^{1/2}$, and tends to unity for large imaginary frequencies. As a result, the worst case scenario for the solution of the linear system corresponds to the static case $\omega=0$.

In summary, by solving iteratively the linear system along the imaginary axis we circumvent the difficulties associated with the ill conditioning of the linear system in Eq. (4) occurring at real frequencies and the necessity of dense Brillouin-zone sampling. The details of the analytic continuation are discussed in Appendix C.

2. Green's function

A similar analysis can be carried out for the calculation of the Green's function using the method introduced in Sec. II B. It is straightforward to establish that in this case the condition number of the system is given by $\kappa=E_{cut}/\delta$. As the infinitesimal δ is typically taken to be 0.1 eV, we are effectively dealing with a situation analogous to a small band-gap semiconductor. The TPA preconditioner can be adopted to reduce the condition number to $\kappa=E_{kin}^{VBM}/\delta$, where E_{kin}^{VBM} is the expectation value of the kinetic energy of the highest occupied state and is independent of the basis set cutoff. Numerical tests confirm that this is indeed a sensible and effective strategy.

APPENDIX C: ANALYTIC CONTINUATION USING PADÉ APPROXIMANTS

In order to perform the analytic continuation of the screened Coulomb interaction from the imaginary axis to the real axis, we employ diagonal Padé approximants.[18,52,53] The Padé approximant of order N is the optimal rational approximation to a target function $f(\omega)$ known in N distinct points ω_n, $n=1,\ldots,N$. When N is an odd integer the diagonal Padé approximant reads

$$P_N(\omega) = \frac{p_0 + p_1\omega + \cdots + p_{(N-1)/2}\omega^{(N-1)/2}}{1 + q_1\omega + \cdots + q_{(N-1)/2}\omega^{(N-1)/2}}, \tag{C1}$$

and its coefficients $p_0, p_1, \ldots, p_{(N-1)/2}$, $q_1, \ldots, q_{(N-1)/2}$ are determined by matching the approximant to the target function in N points $P_N(\omega_n)=f(\omega_n)$, $n=1,\ldots,N$. Both the coefficients and the Padé approximant can be calculated very efficiently using a simple recursive algorithm.[53] Some experimentation indicates that approximants of order $N\geq 5$ are necessary to reproduce a plasmon-pole spectral shape including a finite linewidth. This observation can be rationalized by considering that a plasmon-pole spectral function is completely defined by the values of the function at $\omega=0$, the location, strength, and width of the pole, and the asymptotic value at $\omega=+i\infty$. Some of this information is redundant and can be obtained by using sum rules.[2] The parity of the screened Coulomb interaction can also be exploited to minimize the number of input frequencies. The advantage of the Padé approximant is that a more refined description of the frequency-dependent screened Coulomb interaction can simply be achieved by calculating additional points along the imaginary axis.

We also investigated the possibility of analytically continuing the screened Coulomb interaction by using a multipole expansion as suggested in Ref. 17. We tried one-, two-, and three-pole expansions by determining the coefficients using the simplex method of Nelder and Mead.[67] The single-pole approximation appears robust but the quality of the real-axis continuation is poorer than what we obtained by using Padé approximants. Multipole approximations were found to be unreliable because of their high sensitivity to the initial guesses for the coefficients. Our experience therefore is that the multipole expansion is not an optimal choice for an automated procedure where the analytic continuation has to be performed for every $\mathbf{G}$, $\mathbf{G}'$, and $\mathbf{q}$ of the screened Coulomb interaction without manual intervention.

APPENDIX D: SIMULTANEOUS CALCULATION OF THE SUSCEPTIBILITY AT MULTIPLE FREQUENCIES

The linear systems, Eqs. (11) and (18), can be solved efficiently by using the "multishift" cBiCG method of Ref. 29. Multishift methods exploit the knowledge gained during the iterative solution of the *seed* system $Ax=b$ to determine the solutions of the *shifted* system $Ax+\omega x=b$ with only a small computational overhead. The rationale behind such method is that the seed system and the shifted system share the same Krilov subspaces $\{b, Ab, A^2b, \ldots\}$, therefore the residuals of the seed and of the shifted systems can be taken to be collinear.[29]

This multishift technique allows us to determine the entire frequency dependence of the dielectric matrix by performing one single static calculation for each set of parameters $[\mathbf{q},\mathbf{G}]$ in Eq. (34). For the seed system the algorithm is still given

by Eqs. (A2)–(A8). For the shifted system we replace the calculation of the residuals $r_{n,\omega}$ and of the coefficients $\alpha_{n,\omega}$, $\beta_{n,\omega}$ corresponding to the frequency ω by the following relations,

$$r_{n,\omega} = \frac{r_n}{\pi_{n,\omega}}, \quad \alpha_{n,\omega} = \frac{\pi_{n,\omega}}{\pi_{n+1,\omega}}\alpha_n, \quad \beta_{n,\omega} = \left(\frac{\pi_{n,\omega}}{\pi_{n+1,\omega}}\right)^2 \beta_n, \tag{D1}$$

where the scaling factor $\pi_{n+1,\omega}$ is calculated through the recurrence relation

$$\pi_{n+1,\omega} = (1+\omega\alpha_n)\pi_{n,\omega} + \frac{\alpha_n\beta_{n-1}}{\alpha_{n-1}}(\pi_{n,\omega} - \pi_{n-1,\omega}). \tag{D2}$$

In order to obtain collinear residuals r_n and $r_{n,\omega}$ we need to initialize the algorithm using $x_0=0$.

The use of Eqs. (D1) and (D2) allows us to skip the time-consuming operations involving the Hamiltonian in Eqs. (A2) and (A6). This method is extremely convenient for determining the frequency-dependent susceptibility for many frequencies at the cost of one single calculation.

We point out that this method still carries some drawbacks. One limitation is that this method cannot be applied to the self-consistent system of Eq. (4) because the know term on the right-hand side depends on the frequency ω itself. Therefore the use of the shifted cBiCG method is only possible for *non-self-consistent calculations of the dielectric matrix* and requires explicit matrix inversions to determine the screened Coulomb interaction. This approach can be regarded as an improved version of the technique proposed in Ref. 24.

Another limitation is that the shifted cBiCG method does not allow for the use of preconditioners. In fact the preconditioned seed system $M^{-1}Ax=M^{-1}b$ and the preconditioned shifted system $M^{-1}Ax+\omega M^{-1}x=M^{-1}b$ do not share the same Krilov subspaces, hence the residuals cannot be taken to be collinear.[68] The practical consequence is that for systems with large basis set energy cutoffs and small band gaps, the number of iterations required to achieve convergence could be impractically large.

*feliciano.giustino@materials.ox.ac.uk

[1] L. Hedin, Phys. Rev. **139**, A796 (1965).
[2] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[3] G. Onida, L. Reining, R. W. Godby, R. Del Sole, and W. Andreoni, Phys. Rev. Lett. **75**, 818 (1995).
[4] M. Rohlfing and S. G. Louie, Phys. Rev. Lett. **80**, 3320 (1998).
[5] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, Phys. Rev. Lett. **80**, 4510 (1998).
[6] M. Rohlfing and S. G. Louie, Phys. Rev. Lett. **81**, 2312 (1998).
[7] M. Rohlfing and S. G. Louie, Phys. Rev. B **62**, 4927 (2000).
[8] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[9] C. D. Spataru, S. Ismail-Beigi, L. X. Benedict, and S. G. Louie, Phys. Rev. Lett. **92**, 077402 (2004).
[10] K. S. Thygesen and A. Rubio, Phys. Rev. B **77**, 115333 (2008).
[11] C. D. Spataru, M. A. Cazalilla, A. Rubio, L. X. Benedict, P. M. Echenique, and S. G. Louie, Phys. Rev. Lett. **87**, 246405 (2001).
[12] C.-H. Park, F. Giustino, C. D. Spataru, M. L. Cohen, and S. G. Louie, Nano Lett. **9**, 4324 (2009).
[13] M. Gatti, F. Bruneval, V. Olevano, and L. Reining, Phys. Rev. Lett. **99**, 266402 (2007).
[14] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).
[15] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
[16] A. Schindlmayr and R. W. Godby, Phys. Rev. Lett. **80**, 1702 (1998).
[17] H. N. Rojas, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **74**, 1827 (1995).
[18] S. Lebegue, B. Arnaud, M. Alouani, and P. E. Bloechl, Phys. Rev. B **67**, 155208 (2003).
[19] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, in *Solid State Physics*, edited by H. Ehrenreich and F. Spaepen (Academic, San Diego, 2000), Vol. 54, p. 1.
[20] L. Hedin and S. Lundqvist, in *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), Vol. 23, p. 1.
[21] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[22] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).
[23] F. Bruneval and X. Gonze, Phys. Rev. B **78**, 085125 (2008).
[24] L. Reining, G. Onida, and R. W. Godby, Phys. Rev. B **56**, R4301 (1997).
[25] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **79**, 201104(R) (2009).
[26] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **81**, 115104 (2010).
[27] H. F. Wilson, D. Lu, F. Gygi, and G. Galli, Phys. Rev. B **79**, 245106 (2009).
[28] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozi, Rev. Mod. Phys. **73**, 515 (2001).
[29] A. Frommer, Computing **70**, 87 (2003).
[30] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).
[31] A. Fleszar and R. Resta, Phys. Rev. B **31**, 5305 (1985).
[32] K. Kunc and E. Tosatti, Phys. Rev. B **29**, 7045 (1984).
[33] S. Baroni, P. Giannozzi, and A. Testa, Phys. Rev. Lett. **58**, 1861 (1987).
[34] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
[35] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[36] S. Kümmel and L. Kronik, Rev. Mod. Phys. **80**, 3 (2008).
[37] S. de Gironcoli, Phys. Rev. B **51**, 6773 (1995).
[38] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch,

J. Phys.: Condens. Matter **21**, 395502 (2009).
[39] R. Del Sole, L. Reining, and R. W. Godby, Phys. Rev. B **49**, 8024 (1994).
[40] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **35**, 5585 (1987).
[41] The factor of 2 for the spin degeneracy is not included in Eq. (13) as only the states with the same spin couple when forming the matrix elements of the GW self-energy (Ref. 2).
[42] Had we truncated the integral along the real axis as proposed for the Coulomb term, the Σ^{ex} term would have acquired a spurious frequency dependence associated with the semicircle closing the complex integration path.
[43] R. M. Pick, M. H. Cohen, and R. M. Martin, Phys. Rev. B **1**, 910 (1970).
[44] A. Baldereschi and E. Tosatti, Phys. Rev. B **17**, 4710 (1978).
[45] S. Baroni and R. Resta, Phys. Rev. B **33**, 7017 (1986).
[46] J. Spencer and A. Alavi, Phys. Rev. B **77**, 193110 (2008).
[47] D. A. H. Jacobs, IMA J. Numer. Anal. **6**, 447 (1986).
[48] S. G. Johnson and M. Frigo, IEEE Trans. Signal Process. **55**, 111 (2007).
[49] A. Natan, A. Benjamini, D. Naveh, L. Kronik, M. L. Tiago, S. P. Beckman, and J. R. Chelikowsky, Phys. Rev. B **78**, 075109 (2008).
[50] J. M. Soler, E. Artacho, J. D. Gale, A. Garcìa, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[51] J. P. Walter and M. L. Cohen, Phys. Rev. B **2**, 1821 (1970).
[52] K.-H. Lee and K. J. Chang, Phys. Rev. B **54**, R8285 (1996).
[53] H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).
[54] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12255 (1989).
[55] D. D. Johnson, Phys. Rev. B **38**, 12807 (1988).
[56] J. A. Appelbaum and D. R. Hamman, Phys. Rev. B **8**, 1777 (1973).
[57] J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B **10**, 5095 (1974).
[58] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[59] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[60] *Numerical Data and Functional Relationships in Science and Technology*, Landolt-Börnstein, New Series, Group III, Vol. 17, Pt. A, edited by K. H. Hellwege and O. Madelung (Springer, Berlin, 1982); and *Numerical Data and Functional Relationships in Science and Technology*, Landolt-Börnstein, New Series, Group III, Vol. 22, Pt. A, edited by K. H. Hellwege and O. Madelung (Springer, Berlin, 1982).
[61] J. E. Ortega and F. J. Himpsel, Phys. Rev. B **47**, 2130 (1993).
[62] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[63] D. Vanderbilt, Phys. Rev. B **32**, 8412 (1985).
[64] X. Gonze, Phys. Rev. B **55**, 10337 (1997).
[65] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[66] G. H. Golub and C. F. Van Loan, *Matrix Computations* (John Hopkins University Press, Baltimore, 1983).
[67] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
[68] V. Simoncini and D. B. Szyld, Numer. Linear Algebra Appl. **14**, 1 (2006).