PHYSICAL REVIEW B **94**, 165109 (2016)

Cubic scaling GW: Towards fast quasiparticle calculations

Peitao Liu,[1,2,*] Merzuk Kaltak,[1] Jiří Klimeš,[3,4] and Georg Kresse[1]

[1]*University of Vienna, Faculty of Physics and Center for Computational Materials Science, Sensengasse 8/12, A-1090 Vienna, Austria*

[2]*Institute of Metal Research, Chinese Academy of Sciences, Shenyang 110016, China*

[3]*J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, CZ-18223 Prague 8, Czech Republic*

[4]*Department of Chemical Physics and Optics, Faculty of Mathematics and Physics, Charles University, Ke Karlovu 3, CZ-12116 Prague 2, Czech Republic*

(Received 8 July 2016; published 5 October 2016)

Within the framework of the full potential projector-augmented wave methodology, we present a promising low-scaling GW implementation. It allows for quasiparticle calculations with a scaling that is cubic in the system size and linear in the number of k points used to sample the Brillouin zone. This is achieved by calculating the polarizability and self-energy in the real-space and imaginary-time domains. The transformation from the imaginary time to the frequency domain is done by an efficient discrete Fourier transformation with only a few nonuniform grid points. Fast Fourier transformations are used to go from real space to reciprocal space and vice versa. The analytic continuation from the imaginary to the real frequency axis is performed by exploiting Thiele's reciprocal difference approach. Finally, the method is applied successfully to predict the quasiparticle energies and spectral functions of typical semiconductors (Si, GaAs, SiC, and ZnO), insulators (C, BN, MgO, and LiF), and metals (Cu and $SrVO_3$). The results are compared with conventional GW calculations. Good agreement is achieved, highlighting the strength of the present method.

DOI: 10.1103/PhysRevB.94.165109

I. INTRODUCTION

Kohn-Sham (KS) density functional theory (DFT) in the local density approximation (LDA) [1] has been proven to be successful in describing the ground-state properties for many weakly correlated materials. However, it fails to predict excited-state properties [2]. For instance, DFT within the LDA or generalized gradient approximation (GGA) always gives a smaller band gap than the experimental value. This is not surprising because DFT is just a ground-state theory and there is no formal justification to interpret DFT eigenvalues as quasiparticle (QP) energies. In contrast, the GW approximation of Hedin [3,4] has been widely and successfully applied to the calculations of QP energies for many kinds of systems (for reviews, see Refs. [5,6]) because it provides a good approximation for the electron's self-energy by including many-body effects in the electron-electron interaction. This is achieved by screening the bare exchange interaction with the inverse frequency-dependent dielectric function. Moreover, since the GW self-energy can be diagrammatically formulated in the same many-body framework as dynamical mean field theory (DMFT) [7,8], the GW approximation not only enables an elegant combination with DMFT, i.e., GW+DMFT [9,10], but also overcomes the fundamental double-counting problems occurring in LDA+DMFT because for GW+DMFT one actually knows which Feynman diagrams are counted twice [11].

However, conventional GW calculations are usually restricted to small systems and few k points. This is related to the fairly high computational cost, which is caused by the evaluation of the computationally demanding polarizability and self-energy at a set of real frequencies. Direct evaluation of the polarizability using the Adler and Wiser formula [12,13] involves a summation over all pairs of occupied and unoccupied states, leading to the quartic scaling in the system size and quadratic scaling in the number of k points used to sample the Brillouin zone (BZ). Furthermore, calculation of the self-energy for all occupied states in reciprocal space and real frequency is at least two times more expensive than the evaluation of the polarizability. To reduce the computational effort, the space-time method [14,15], which calculates the polarizability and self-energy in real space and imaginary time, was proposed. Nevertheless, the space-time method demands considerable storage for the Green's function and self-energy due to the huge number of real-space grid points. In addition, to obtain reasonable convergence, fairly dense imaginary-time grids are required.

To circumvent the large storage requirement of the space-time method, a promising scheme has been recently proposed by Kaltak *et al.* [16,17]. It allows to calculate the random phase approximation (RPA) correlation energy with a cubic scaling in the system size and a linear scaling in the number of k points used to sample the BZ. As in the work of Rojas *et al.* [14], this is achieved by calculating the polarizability in real space and imaginary time via contraction over the Green's functions of occupied and unoccupied states. The transformation of the polarizability from the imaginary time to the frequency domain is performed by an efficient discrete Fourier transformation with only a few nonuniform grid points [16]. Spatial fast Fourier transformations (FFT) within a supercell are utilized to go from real space to reciprocal space and vice versa [17].

Here, we extend Kaltak's scheme [16,17] to QP calculations in the GW approximation, in which the screened Coulomb interaction W is calculated within the RPA and the self-energy is efficiently evaluated via contraction over the Green's function and W in real space and imaginary time. Similar spatial FFT as discussed in Ref. [17] are employed, whenever transformations between the real and reciprocal space are required. To transform the self-energy from the imaginary time to the frequency domain, nonuniform cosine and sine

*peitao.liu@univie.ac.at

transformations are used for the even and odd parts of the self-energy, respectively. Given that DMFT is usually formulated on the imaginary frequency axis as well [7,8], our method provides a natural interface for the combination of GW with DMFT.

In this paper, we focus only on the GW QP calculations. Detailed formulations for our low-scaling GW implementation within the framework of the projector-augmented wave (PAW) are given. Considering that the Green's functions and self-energies in the present implementation are evaluated at imaginary frequencies, an analytic continuation to the real frequency axis is required to compare with experimentally measured observables, such as QP energies and spectral functions. To this end, Thiele's reciprocal difference method [18] is used. We then apply our new implementation to predict the QP energies and spectral functions for typical semiconductors (Si, GaAs, SiC, and ZnO), insulators (C, BN, MgO, and LiF), and metals (Cu and $SrVO_3$), and compare our results with the conventional GW implementation. For the sake of brevity, we just show the comparison for the single-shot GW calculations, i.e., G_0W_0, where the one-electron energies and wave functions required in G and W are fixed at the DFT level. The self-consistent low-scaling GW will be discussed in future publications. To avoid confusion with the conventional G_0W_0, we denote our low-scaling single-shot GW as G_0W_0r. It is found that the QP energies and spectral functions predicted by G_0W_0r are in good agreement with G_0W_0 but with a reduced scaling in the system size and number of k points, highlighting the power of the present method.

II. METHOD

Figure 1 shows our scheme for the low-scaling GW QP calculations. The polarizability $\chi(\mathbf{r},\mathbf{R}',i\tau)$ is calculated via the contraction (GG) of the occupied and unoccupied Green's functions within the PAW framework [17]. The contraction is performed in real space and the necessary quantities are obtained by fast Fourier transformations (FFT) within a supercell [17]. Subsequently, the screened interaction $W_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega)$

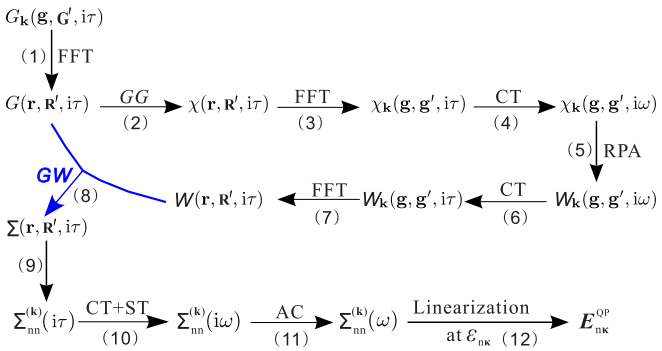

FIG. 1. Schematic work flow for the low-scaling GW QP calculations showing the necessary steps [(1)–(12)] to obtain the QP energies from the Green's function G via the polarizability χ, screened Coulomb interaction W, and self-energy Σ. FFT denote fast Fourier transformations between real and reciprocal space. CT and ST are nonuniform cosine and sine transformations between imaginary time and frequency.

is obtained within the RPA. To transform the polarizability χ and screened Coulomb interaction W from imaginary time to frequency domain and vice versa, efficient nonuniform cosine transformations (CT) [16] are used. The self-energy $\Sigma(\mathbf{r},\mathbf{R}',i\tau)$ is calculated by contracting the Green's function and W within the GW approximation. The matrix elements of the self-energy in the orbital basis are evaluated within the PAW. To transform the self-energy from the imaginary time to the frequency domain, CT and sine transformations (ST), respectively, are used for the even and odd parts of the self-energy. The self-energy along the real frequency axis is obtained by an analytic continuation (AC). Finally, the QP energies $E^{\mathrm{QP}}_{n\mathbf{k}}$ within single-shot G_0W_0r are calculated by linearizing the diagonal elements of the self-energy around the DFT one-electron eigenvalues $\epsilon_{n\mathbf{k}}$. The subsequent subsections describe these steps in detail.

A. Description of notations and definitions

In this part, we will give the description of the notations used throughout the paper, the definition of the Green's functions, as well as the spatial and temporal Fourier transformations.

1. Definitions of Green's functions

In this paper, we have defined two types of Green's functions: occupied and unoccupied Green's functions, which are evaluated for the negative and positive time, respectively:

$$\underline{G}(\mathbf{r},\mathbf{r}',i\tau) = \sum_{i}^{\mathrm{occ}} \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}')e^{-\epsilon_i\tau} \quad (\tau < 0), \tag{1}$$

$$\overline{G}(\mathbf{r},\mathbf{r}',i\tau) = -\sum_{a}^{\mathrm{unocc}} \psi_a(\mathbf{r})\psi_a^*(\mathbf{r}')e^{-\epsilon_a\tau} \quad (\tau > 0). \tag{2}$$

Here, the indices i and a label occupied and unoccupied orbitals, respectively. $\psi_i(\mathbf{r})$ [$\psi_a(\mathbf{r})$] is the one-electron orbital with the energy of ϵ_i (ϵ_a) and the Fermi energy is set to zero. This implies that all occupied (unoccupied) one-electron energies ϵ_i (ϵ_a) are negative (positive), yielding exponentially decaying Green's functions $\underline{G}$ and $\overline{G}$. With the definitions in Eqs. (1) and (2), the single-particle Green's function can be expressed as

$$G(\mathbf{r},\mathbf{r}',i\tau) = \Theta(\tau)\overline{G}(\mathbf{r},\mathbf{r}',i\tau) + \Theta(-\tau)\underline{G}(\mathbf{r},\mathbf{r}',i\tau), \tag{3}$$

where Θ is the Heaviside step function.

2. Nonuniform imaginary-time and frequency grids

The imaginary-time $\{i\tau_j\}_{j=1}^{N}$ and frequency $\{i\omega_k\}_{k=1}^{N}$ grids used in this work have been determined by minimizing the discretization error of the direct Møller-Plesset energy in the imaginary-time and frequency domains, respectively [16]. It has been found that the two grids are dual to each other. That is, given $\{i\tau_j\}_{j=1}^{N}$, the discretization error function is minimal at the grid points $\{i\omega_k\}_{k=1}^{N}$, and vice versa [16]. It has also been observed that the RPA correlation energy can be evaluated accurately with a modest number of grid points [16]. For instance, to achieve μeV accuracy per atom, 16 time and frequency points are usually sufficient [16]. In this work, we also found that with 20 grid points, we could

obtain converged QP energies with 0.01 eV accuracy for all the materials considered. With a few imaginary grid points, the memory requirements are obviously much reduced.

3. Nonuniform cosine and sine transformations

To go from the imaginary time to imaginary frequency, and vice versa, nonuniform discrete cosine and sine transformations have been exploited for the even and odd functions, respectively. Specifically, for an even function F with respect to imaginary time/frequency, such as the polarizability χ and screened Coulomb interaction W, the forward Fourier transformation is given by a CT

$$F(i\omega_k) = \sum_{j=1}^{N} \gamma_{kj} \cos(\omega_k \tau_j) F(i\tau_j). \tag{4}$$

Here, the imaginary time $\{i\tau_j\}_{j=1}^{N}$ and frequency $\{i\omega_k\}_{k=1}^{N}$ grids are precalculated. The coefficients γ_{kj} are determined in analogy to the imaginary time and frequency grids by minimizing the error function [16]

$$\eta^c(x,\gamma) = \underbrace{\frac{2x}{x^2+\omega_k^2}}_{2\int_0^\infty d\tau \cos(\omega_k\tau)e^{-x\tau}} - \sum_{j=1}^{N} \gamma_{kj} \cos(\omega_k \tau_j) e^{-x\tau_j}, \tag{5}$$

for all transition energies $\min(\epsilon_a - \epsilon_i) \leqslant x \leqslant \max(\epsilon_a - \epsilon_i)$ and each known frequency point ω_k separately.

Analogously, the inverse CT is given by [16]

$$F(i\tau_j) = \sum_{k=1}^{N} \xi_{jk} \cos(\tau_j \omega_k) F(i\omega_k), \tag{6}$$

where the matrix $\xi \cos(\tau\omega)$ is the inverse of the matrix $\gamma \cos(\omega\tau)$ in Eq. (4).

In contrast, for an odd function F with respect to imaginary time/frequency, the forward Fourier transformation is described by a ST

$$F(i\omega_k) = i \sum_{j=1}^{N} \lambda_{kj} \sin(\omega_k \tau_j) F(i\tau_j). \tag{7}$$

Again, $\{i\tau_j\}_{j=1}^{N}$ and $\{i\omega_k\}_{k=1}^{N}$ are precalculated and chosen to be identical to the cosine grid. However, the coefficients λ_{kj} are determined by minimizing the error function

$$\eta^s(x,\lambda) = \underbrace{\frac{2\omega_k}{x^2+\omega_k^2}}_{2\int_0^\infty d\tau \sin(\omega_k\tau)e^{-x\tau}} - \sum_{j=1}^{N} \lambda_{kj} \sin(\omega_k \tau_j) e^{-x\tau_j}. \tag{8}$$

To this end, similar strategies as discussed in Ref. [16] are used. The inverse ST is then obtained by

$$F(i\tau_j) = -i \sum_{k=1}^{N} \zeta_{jk} \sin(\tau_j \omega_k) F(i\omega_k), \tag{9}$$

where the matrix $\zeta \sin(\tau\omega)$ is the inverse of the matrix $\lambda \sin(\omega\tau)$ in Eq. (7). It should be noted that the matrices γ, ξ, λ, and ζ are all precalculated and stored after the imaginary time $\{i\tau_j\}_{j=1}^{N}$ and frequency $\{i\omega_k\}_{k=1}^{N}$ grids are determined.

4. Spatial fast Fourier transformation

To transform the Green's functions from reciprocal to real space, we employ fast discrete Fourier transformation within a supercell [17]. Considering the symmetry of the Green's functions, only the irreducible stripe $G(\mathbf{r},\mathbf{R}')$ needs to be calculated in two steps [17]:

$$G(\mathbf{r},\mathbf{G}') = \sum_{\mathbf{g}\in\mathcal{L}_c^*} e^{i(\mathbf{k}+\mathbf{g})\mathbf{r}} G_{\mathbf{k}}(\mathbf{g},\mathbf{g}'), \tag{10}$$

$$G(\mathbf{r},\mathbf{R}') = \sum_{\mathbf{G}'\in\mathcal{L}_s^*} G(\mathbf{r},\mathbf{G}') e^{-i\mathbf{G}'\mathbf{R}'}. \tag{11}$$

Here, position vector $\mathbf{r}$ is restricted to the unit cell (C), whereas $\mathbf{R}$ extends over the entire supercell (S). $\mathbf{g}$ and $\mathbf{G}$, respectively, represent the lattice vector of the reciprocal cell ($\mathcal{L}_c^*$) and reciprocal supercell ($\mathcal{L}_s^*$). Furthermore, $\mathbf{k}$ is a k point used to sample the Brillouin zone (BZ) and $\mathbf{G}' = \mathbf{k} + \mathbf{g}'$. The time complexity of the spatial FFT is of the order $\ln(N_b^2 N_k) N_b^2 N_k$ with N_b and N_k being the total number of considered basis vectors $\mathbf{g}$ and k points in the BZ, respectively [17].

Similarly, the inverse spatial FFT is given by [17]

$$G(\mathbf{r},\mathbf{G}') = \sum_{\mathbf{R}'\in S} G(\mathbf{r},\mathbf{R}') e^{i\mathbf{G}'\mathbf{R}'}, \tag{12}$$

$$G_{\mathbf{k}}(\mathbf{g},\mathbf{g}') = \sum_{\mathbf{r}\in C} e^{-i(\mathbf{k}+\mathbf{g})\mathbf{r}} G(\mathbf{r},\mathbf{G}'), \tag{13}$$

which has the same time complexity as the spatial FFT.

Considering that the polarizability χ has the same spatial symmetry as the Green's functions, the above-mentioned spatial and inverse spatial FFT applies to the polarizability χ as well.

B. Calculation of the polarizability $\chi(\mathbf{r},\mathbf{R}',i\tau)$ within the PAW

In this section, we discuss the steps 1 and 2 in Fig. 1 and derive a suitable expression for the polarizability in real space $\chi(\mathbf{r},\mathbf{R}',i\tau)$ within the framework of the PAW method.

It is known that the evaluation of the polarizability in reciprocal space and real frequency results in an unfavorable scaling. However, the polarizability is simply multiplicative, when evaluated in the real-space and imaginary-time domains [14,15]

$$\chi(\mathbf{r},\mathbf{R}',i\tau) = G(\mathbf{r},\mathbf{R}',i\tau) G^*(\mathbf{r},\mathbf{R}',-i\tau). \tag{14}$$

For simplicity, we restrict our considerations to positive imaginary times $\tau > 0$ in the following since the expressions for $\tau < 0$ are obtained by exchanging $\overline{G} \leftrightarrow \underline{G}$.

Inserting expression (3) for the Green's function into (14) and using the explicit representations in Eqs. (1) and (2) yields, for $\tau > 0$,

$$\chi(\mathbf{r},\mathbf{R}',i\tau) = -\sum_{a}^{\text{unocc}} \psi_a(\mathbf{r}) \psi_a^*(\mathbf{R}') e^{-\epsilon_a \tau} \sum_{i}^{\text{occ}} \psi_i(\mathbf{R}') \psi_i^*(\mathbf{r}) e^{\epsilon_i \tau}. \tag{15}$$

Fourier transforming this expression using Eqs. (12) and (13) gives the polarizability in reciprocal space for $\tau > 0$,

$$\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau) = -\sum_{i}^{\text{occ}}\sum_{a}^{\text{unocc}} \xi_{ia}(i\tau)\langle\psi_i|e^{-i(\mathbf{k}+\mathbf{g})\mathbf{r}}|\psi_a\rangle \times \langle\psi_a|e^{i(\mathbf{k}+\mathbf{g}')\mathbf{r}'}|\psi_i\rangle, \quad (16)$$

where $\xi_{ia}(\tau) = e^{-(\epsilon_a-\epsilon_i)\tau}$ describes the time dependence.

However, within the PAW method [19,20], this expression is more involved because the all-electron orbitals $\psi_{n\mathbf{k}}$ are related to the corresponding pseudo-orbitals $\tilde{\psi}_{n\mathbf{k}}$ by a linear transformation

$$|\psi_{n\mathbf{k}}\rangle = |\tilde{\psi}_{n\mathbf{k}}\rangle + \sum_{\mu}(|\phi_\mu\rangle - |\tilde{\phi}_\mu\rangle)\langle\tilde{p}_\mu|\tilde{\psi}_{n\mathbf{k}}\rangle. \quad (17)$$

The pseudo-orbitals $\tilde{\psi}_{n\mathbf{k}}$ are the variational quantity of the PAW method and are expanded in plane waves, whereas ϕ_μ and $\tilde{\phi}_\mu$ are all-electron and pseudo-partial waves, respectively, and $\tilde{p}_\mu$ are projectors, which are dual to the $\tilde{\phi}_\mu$ within the augmentation sphere. The index $\mu = (\mathbf{R}_\mu, n_\mu, l_\mu, m_\mu)$ is an abbreviation for the atomic site $\mathbf{R}_\mu$ and the energy quantum number n_μ and angular momentum numbers (l_μ, m_μ) that characterize the solution of the Schrödinger equation for a reference atom.

Inserting Eq. (17) into Eq. (15) yields four terms for $\tau > 0$:

$$\chi^{(1)}(\mathbf{r},\mathbf{R}',i\tau) = -\sum_{a}^{\text{unocc}} \tilde{\psi}_a(\mathbf{r})\tilde{\psi}_a^*(\mathbf{R}')e^{-\epsilon_a\tau} \times \sum_{i}^{\text{occ}} \tilde{\psi}_i(\mathbf{R}')\tilde{\psi}_i^*(\mathbf{r})e^{\epsilon_i\tau}, \quad (18)$$

$$\chi^{(2)}(\mathbf{r},\mathbf{R}',i\tau) = -\sum_{\mu\nu} Q_{\mu\nu}(\mathbf{r}) \sum_{a}^{\text{unocc}} \langle\tilde{p}_\nu|\tilde{\psi}_a\rangle\tilde{\psi}_a^*(\mathbf{R}')e^{-\epsilon_a\tau} \times \sum_{i}^{\text{occ}} \tilde{\psi}_i(\mathbf{R}')\langle\tilde{\psi}_i|\tilde{p}_\mu\rangle e^{\epsilon_i\tau}, \quad (19)$$

$$\chi^{(3)}(\mathbf{r},\mathbf{R}',i\tau) = -\sum_{\alpha\beta} Q_{\alpha\beta}(\mathbf{R}') \sum_{a}^{\text{unocc}} \tilde{\psi}_a(\mathbf{r})\langle\tilde{\psi}_a|\tilde{p}_\alpha\rangle e^{-\epsilon_a\tau} \times \sum_{i}^{\text{occ}} \langle\tilde{p}_\beta|\tilde{\psi}_i\rangle\tilde{\psi}_i^*(\mathbf{r})e^{\epsilon_i\tau}, \quad (20)$$

$$\chi^{(4)}(\mathbf{r},\mathbf{R}',i\tau) = -\sum_{\mu\nu\alpha\beta} Q_{\mu\nu}(\mathbf{r})Q_{\alpha\beta}(\mathbf{R}') \times \sum_{a}^{\text{unocc}} \langle\tilde{p}_\nu|\tilde{\psi}_a\rangle\langle\tilde{\psi}_a|\tilde{p}_\alpha\rangle e^{-\epsilon_a\tau} \times \sum_{i}^{\text{occ}} \langle\tilde{p}_\beta|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|\tilde{p}_\mu\rangle e^{\epsilon_i\tau}, \quad (21)$$

where the auxiliary function $Q_{\alpha\beta}(\mathbf{r})$ is defined as

$$Q_{\alpha\beta}(\mathbf{r}) = \phi_\alpha^*(\mathbf{r})\phi_\beta(\mathbf{r}) - \tilde{\phi}_\alpha^*(\mathbf{r})\tilde{\phi}_\beta(\mathbf{r}), \quad (22)$$

and describes the difference between the charge density of all-electron and pseudo-partial waves. In practice, one needs further approximations for $Q_{\alpha\beta}(\mathbf{r})$ since normally this function is oscillatory within the augmentation sphere. In the present implementation, the function is expanded in an orthogonal set of functions, and the rapid spatial oscillations are neglected beyond a certain plane-wave energy cutoff [20].

According to the definitions of the Green's functions in Eqs. (1) and (2) and four expressions in Eqs. (18)–(21), we define here four auxiliary functions for unoccupied Green's functions:

$$\overline{G}^{(1)}(\mathbf{r},\mathbf{R}',i\tau) = -\sum_{a}^{\text{unocc}} \tilde{\psi}_a(\mathbf{r})\tilde{\psi}_a^*(\mathbf{R}')e^{-\epsilon_a\tau}, \quad (23)$$

$$\overline{G}^{(2)}(\nu,\mathbf{R}',i\tau) = -\sum_{a}^{\text{unocc}} \langle\tilde{p}_\nu|\tilde{\psi}_a\rangle\tilde{\psi}_a^*(\mathbf{R}')e^{-\epsilon_a\tau}, \quad (24)$$

$$\overline{G}^{(3)}(\mathbf{r},\alpha,i\tau) = -\sum_{a}^{\text{unocc}} \tilde{\psi}_a(\mathbf{r})\langle\tilde{\psi}_a|\tilde{p}_\alpha\rangle e^{-\epsilon_a\tau}, \quad (25)$$

$$\overline{G}^{(4)}(\nu,\alpha,i\tau) = -\sum_{a}^{\text{unocc}} \langle\tilde{p}_\nu|\tilde{\psi}_a\rangle\langle\tilde{\psi}_a|\tilde{p}_\alpha\rangle e^{-\epsilon_a\tau}, \quad (26)$$

and four auxiliary functions for occupied Green's functions:

$$\underline{G}^{*(1)}(\mathbf{r},\mathbf{R}',-i\tau) = \sum_{i}^{\text{occ}} \tilde{\psi}_i(\mathbf{R}')\tilde{\psi}_i^*(\mathbf{r})e^{\epsilon_i\tau}, \quad (27)$$

$$\underline{G}^{*(2)}(\mu,\mathbf{R}',-i\tau) = \sum_{i}^{\text{occ}} \tilde{\psi}_i(\mathbf{R}')\langle\tilde{\psi}_i|\tilde{p}_\mu\rangle e^{\epsilon_i\tau}, \quad (28)$$

$$\underline{G}^{*(3)}(\mathbf{r},\beta,-i\tau) = \sum_{i}^{\text{occ}} \langle\tilde{p}_\beta|\tilde{\psi}_i\rangle\tilde{\psi}_i^*(\mathbf{r})e^{\epsilon_i\tau}, \quad (29)$$

$$\underline{G}^{*(4)}(\mu,\beta,-i\tau) = \sum_{i}^{\text{occ}} \langle\tilde{p}_\beta|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|\tilde{p}_\mu\rangle e^{\epsilon_i\tau}. \quad (30)$$

It is easy to prove that

$$G^{(2)}(\nu,\mathbf{R}',i\tau) = \sum_{\mathbf{r}\in C}\langle\tilde{p}_\nu|\mathbf{r}\rangle G^{(1)}(\mathbf{r},\mathbf{R}',i\tau), \quad (31)$$

$$G^{(3)}(\mathbf{r},\alpha,i\tau) = \sum_{\mathbf{R}'\in S} G^{(1)}(\mathbf{r},\mathbf{R}',i\tau)\langle\mathbf{R}'|\tilde{p}_\alpha\rangle, \quad (32)$$

$$G^{(4)}(\nu,\alpha,i\tau) = \sum_{\mathbf{r}\in C}\sum_{\mathbf{R}'\in S}\langle\tilde{p}_\nu|\mathbf{r}\rangle G^{(1)}(\mathbf{r},\mathbf{R}',i\tau)\langle\mathbf{R}'|\tilde{p}_\alpha\rangle, \quad (33)$$

holds for both auxiliary unoccupied and occupied Green's functions.

With the definitions in Eqs. (23)–(30), we obtain the central expression for the polarizability $\chi(\mathbf{r},\mathbf{R}',i\tau)$ at $\tau > 0$ within the PAW framework as follows:

$$\begin{aligned}\chi(\mathbf{r},\mathbf{R}',i\tau) = {}& \overline{G}^{(1)}(\mathbf{r},\mathbf{R}',i\tau)\underline{G}^{*(1)}(\mathbf{r},\mathbf{R}',-i\tau) \\ &+ \sum_{\mu\nu} Q_{\mu\nu}(\mathbf{r})\overline{G}^{(2)}(\nu,\mathbf{R}',i\tau)\underline{G}^{*(2)}(\mu,\mathbf{R}',-i\tau) \\ &+ \sum_{\alpha\beta} Q_{\alpha\beta}(\mathbf{R}')\overline{G}^{(3)}(\mathbf{r},\alpha,i\tau)\underline{G}^{*(3)}(\mathbf{r},\beta,-i\tau) \\ &+ \sum_{\mu\nu\alpha\beta} Q_{\mu\nu}(\mathbf{r})Q_{\alpha\beta}(\mathbf{R}') \\ &\times \overline{G}^{(4)}(\nu,\alpha,i\tau)\underline{G}^{*(4)}(\mu,\beta,-i\tau). \end{aligned} \quad (34)$$

Here, the atomic positions $\mathbf{R}_\mu$, $\mathbf{R}_\nu$ are restricted to the unit cell C, while $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$ take values within the supercell S. Note

that the polarizability for $\tau < 0$ is recovered from Eq. (34) by exchanging $\overline{G} \leftrightarrow \underline{G}$.

In practice, we do not store the auxiliary Green's functions in Eqs. (23)–(25) and (27)–(29) directly using the real-space grids since this would demand considerable storage due to the large number of real-space grid points. Instead, we evaluate them in the reciprocal space using a plane-wave representation first, and successively Fourier transform the functions to real space whenever required. Since the number of plane-wave coefficients is at least twice but often up to 16 times smaller than the number of real-space grid points, the storage demand is dramatically reduced. Fourier transforming Eqs. (23)–(25) to the reciprocal space yields another three auxiliary unoccupied Green's functions:

$$\overline{G}^{(1)}_{\mathbf{k}}(\mathbf{g},\mathbf{G}',i\tau) = -\sum_{a}^{\text{unocc}} \langle \mathbf{g}|\tilde{\psi}_a\rangle\langle\tilde{\psi}_a|\mathbf{G}'\rangle e^{-\epsilon_a\tau}, \tag{35}$$

$$\overline{G}^{(2)}_{\mathbf{k}}(\nu,\mathbf{G}',i\tau) = -\sum_{a}^{\text{unocc}} \langle \tilde{p}_\nu|\tilde{\psi}_a\rangle\langle\tilde{\psi}_a|\mathbf{G}'\rangle e^{-\epsilon_a\tau}, \tag{36}$$

$$\overline{G}^{(3)}_{\mathbf{k}}(\mathbf{g},\alpha,i\tau) = -\sum_{a}^{\text{unocc}} \langle \mathbf{g}|\tilde{\psi}_a\rangle\langle\tilde{\psi}_a|\tilde{p}_\alpha\rangle e^{-\epsilon_a\tau}. \tag{37}$$

Analogously, Fourier transforming Eqs. (27)–(29) to the reciprocal space yields another three auxiliary occupied Green's functions [21]:

$$\underline{G}^{*(1)}_{\mathbf{k}}(\mathbf{g},\mathbf{G}',-i\tau) = \sum_{i}^{\text{occ}} \langle \mathbf{G}'|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|\mathbf{g}\rangle e^{\epsilon_i\tau}, \tag{38}$$

$$\underline{G}^{*(2)}_{\mathbf{k}}(\mu,\mathbf{G}',-i\tau) = \sum_{i}^{\text{occ}} \langle \mathbf{G}'|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|\tilde{p}_\mu\rangle e^{\epsilon_i\tau}, \tag{39}$$

$$\underline{G}^{*(3)}_{\mathbf{k}}(\mathbf{g},\beta,-i\tau) = \sum_{i}^{\text{occ}} \langle \tilde{p}_\beta|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|\mathbf{g}\rangle e^{\epsilon_i\tau}, \tag{40}$$

where the notation

$$\langle \mathbf{g}|\tilde{\psi}\rangle = \sum_{\mathbf{r}\in C} e^{-i(\mathbf{k}+\mathbf{g})\mathbf{r}}\tilde{\psi}(\mathbf{r}), \tag{41}$$

$$\langle \tilde{\psi}|\mathbf{G}'\rangle = \sum_{\mathbf{R}'\in S} \tilde{\psi}^*(\mathbf{R}')e^{i\mathbf{G}'\mathbf{R}'} \tag{42}$$

is used. The computational complexity for evaluating both $\underline{G}^{(j)}$ and $\overline{G}^{(j)}$ is of the order: $NN_kN_b^3$ showing a roughly cubic scaling in the system size ($\approx N_b$) and linear scaling in the number of k points N_k and imaginary grid points N.

We point out that the widely used conventional GW implementation [22], where the polarizability is directly evaluated in the reciprocal space and real frequency domain, shows an unfavorable scaling that is quartic in the system size and quadratic in the number of k points. This scaling is acceptable or even beneficial for small systems, but prohibitive as the system size becomes larger. In contrast, in our new GW implementation, the computational cost in calculating the polarizability reduces to a scaling that is nearly cubic in the system size and linear in the number of k points. This definitely increases the efficiency of GW calculations for large systems. This is also true for the original space-time implementation of Godby *et al.* [14,15], but we emphasize that our implementation has the following advantages: (i) The Green's functions are stored in a plane-wave representation at a few optimized imaginary-time/frequency grid points, which dramatically reduces the memory requirement. (ii) It is implemented within the PAW method. (iii) Discrete CT and ST transformations and spatial FFT are used and the implementation is highly parallelized. (iv) Although similar strategies were used in Ref. [15], an auxiliary supercell Green's function was defined without the Bloch phase factors $e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}')}$. The present method is applicable to all electron Hamiltonians, whereas the augmentation terms cannot be straightforwardly implemented following Ref. [15].

C. Calculation of the correlated screened Coulomb interaction $\tilde{W}(\mathbf{r},\mathbf{R}',i\omega)$

Now, we describe the evaluation of the dynamical correlated screened Coulomb interaction $\tilde{W}(\mathbf{r},\mathbf{R}',i\omega)$, which corresponds to the steps 3–7 in Fig. 1. Once the polarizability $\chi(\mathbf{r},\mathbf{R}',i\tau)$ has been calculated, one has to Fourier transform it to the reciprocal space and imaginary frequency domain where the screened Coulomb interaction is much more comfortable to be calculated. The calculation of $\tilde{W}(\mathbf{r},\mathbf{R}',i\omega)$ involves five steps:

(i) $\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau)$ is determined by an inverse spatial FFT of $\chi(\mathbf{r},\mathbf{R}',i\tau)$ in two steps:

$$\chi(\mathbf{r},\mathbf{G}',i\tau) = \sum_{\mathbf{R}'\in S} \chi(\mathbf{r},\mathbf{R}',i\tau)e^{i\mathbf{G}'\mathbf{R}'}, \tag{43}$$

$$\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau) = \sum_{\mathbf{r}\in C} e^{-i(\mathbf{k}+\mathbf{g})\mathbf{r}}\chi(\mathbf{r},\mathbf{G}',i\tau). \tag{44}$$

Actually, in our implementation the polarizability $\chi(\mathbf{r},\mathbf{R}',i\tau)$ is never stored. Instead, once $\chi(\mathbf{r},\mathbf{R}',i\tau)$ is known for a specific $\mathbf{r}$ and all $\mathbf{R}'$, Eq. (43) is used to Fourier transform the second index to the reciprocal space where the reciprocal wave vectors are restricted to a cutoff sphere, and $\chi(\mathbf{r},\mathbf{G}',i\tau)$ is then stored. The second FFT in Eq. (44) cannot be performed until $\chi(\mathbf{r},\mathbf{G}',i\tau)$ for all $\mathbf{r}$ has been calculated.

(ii) $\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega)$ is computed by a CT of $\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau)$,

$$\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega_k) = \sum_{j=1}^{N} \gamma_{kj}\cos(\omega_k\tau_j)\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau_j). \tag{45}$$

(iii) The full screened Coulomb interaction $W_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega)$ is evaluated by multiplying the bare Coulomb kernel with the inverse dielectric matrix

$$W_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega) = v_{\mathbf{k}}(\mathbf{g},\mathbf{g}')\varepsilon^{-1}_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega), \tag{46}$$

where the symmetric bare Coulomb kernel $v_{\mathbf{k}}(\mathbf{g},\mathbf{g}')$ is

$$v_{\mathbf{k}}(\mathbf{g},\mathbf{g}') = \frac{4\pi e^2}{|\mathbf{k}+\mathbf{g}||\mathbf{k}+\mathbf{g}'|}. \tag{47}$$

The symmetric dielectric matrix is calculated within the RPA as

$$\varepsilon_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega) = \delta_{\mathbf{g},\mathbf{g}'} - v_{\mathbf{k}}(\mathbf{g},\mathbf{g}')\chi_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega). \tag{48}$$

To make the integral over the imaginary frequency well defined, we further define the correlated screened Coulomb interaction

$$\tilde{W}_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega) = W_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega) - v_{\mathbf{k}}(\mathbf{g},\mathbf{g}'). \tag{49}$$

(iv) $\tilde{W}_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau)$ is determined by an inverse CT of $\tilde{W}_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega)$,

$$\tilde{W}_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau_j) = \sum_{k=1}^{N} \xi_{jk} \cos(\tau_j\omega_k)\tilde{W}_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\omega_k). \quad (50)$$

(v) Finally, $\tilde{W}(\mathbf{r},\mathbf{R}',i\tau)$ is calculated by a spatial FFT in two steps:

$$\tilde{W}(\mathbf{r},\mathbf{G}',i\tau) = \sum_{\mathbf{g}\in\mathcal{L}_c^*} e^{i(\mathbf{k}+\mathbf{g})\mathbf{r}}\tilde{W}_{\mathbf{k}}(\mathbf{g},\mathbf{g}',i\tau), \quad (51)$$

$$\tilde{W}(\mathbf{r},\mathbf{R}',i\tau) = \sum_{\mathbf{G}'\in\mathcal{L}_s^*} \tilde{W}(\mathbf{r},\mathbf{G}',i\tau)e^{-i\mathbf{G}'\mathbf{R}'}. \quad (52)$$

D. Calculation of the self-energy

In this section, we give a detailed description how the matrix elements of the self-energy in the orbital basis along the imaginary frequency axis are evaluated. This corresponds to the steps 8–10 in Fig. 1.

1. Evaluation of the self-energy $\Sigma(\mathbf{r},\mathbf{R}',i\tau)$ within the GWA

Within the GW approximation, the self-energy in the reciprocal space and real frequency domain is evaluated by a convolution of the Green's function and screened Coulomb interaction and is used in the conventional GW implementation [22]. However, to obtain converged self-energies, a reasonable number of the real frequency points (∼50 or more) is required to evaluate the convolution integral, thus increasing the computational cost. In contrast, when the self-energy is evaluated in real space and time, it is simply multiplicative [14]

$$\Sigma(\mathbf{r},\mathbf{R}',i\tau) = -G(\mathbf{r},\mathbf{R}',i\tau)W(\mathbf{r},\mathbf{R}',i\tau). \quad (53)$$

In addition, only a few imaginary-time points are required due to the smooth behavior of the Green's functions and screened Coulomb interaction along the imaginary axis.

2. Evaluation of $\tilde{\underline{\Sigma}}^{(\mathbf{k})}_{nn}(i\tau)$ within the PAW

In the following, we evaluate the matrix elements of the self-energy in the orbital basis within the PAW framework. We focus only on the frequency/time-dependent correlation contribution $-\langle\psi_{n\mathbf{k}}|G\tilde{W}|\psi_{n\mathbf{k}}\rangle$ since the bare exchange part $-\langle\psi_{n\mathbf{k}}|Gv_x|\psi_{n\mathbf{k}}\rangle$ within the PAW has already been discussed elsewhere [23]. Furthermore, we define "occupied" $\tilde{\underline{\Sigma}}$ and "unoccupied" $\tilde{\overline{\Sigma}}$ correlated self-energies, i.e., the self-energies evaluated at negative and positive time, respectively, analogous to the Green's functions.

Here, we concentrate on the occupied self-energy $\tilde{\underline{\Sigma}}$ only. The evaluation of the matrix elements of the unoccupied self-energy $\tilde{\overline{\Sigma}}$ is done by replacing $\underline{G}$ with $\overline{G}$. Within the PAW, the diagonal matrix elements of the occupied self-energy ($\tau < 0$) can be calculated as

$$\begin{aligned}\tilde{\underline{\Sigma}}^{(\mathbf{k})}_{\mathrm{nn}}(i\tau) &= \langle\psi_{n\mathbf{k}}|\tilde{\underline{\Sigma}}(i\tau)|\psi_{n\mathbf{k}}\rangle = -\langle\psi_{n\mathbf{k}}|\underline{G}(i\tau)\tilde{W}(i\tau)|\psi_{n\mathbf{k}}\rangle \\ &= -\sum_{\mathbf{r}\in C}\sum_{\mathbf{R}'\in S}\langle\tilde{\psi}_{n\mathbf{k}}|\left\{|\mathbf{r}\rangle\langle\mathbf{r}| + \sum_{\mu\nu} Q_{\mu\nu}(\mathbf{r})|\tilde{p}_\mu\rangle\langle\tilde{p}_\nu|\right\}\underline{G}(i\tau)\tilde{W}(i\tau)\left\{|\mathbf{R}'\rangle\langle\mathbf{R}'| + \sum_{\alpha\beta} Q_{\alpha\beta}(\mathbf{R}')|\tilde{p}_\alpha\rangle\langle\tilde{p}_\beta|\right\}|\tilde{\psi}_{n\mathbf{k}}\rangle. \quad (54)\end{aligned}$$

Here, $|\mathbf{r}\rangle\langle\mathbf{r}| + \sum_{\mu\nu} Q_{\mu\nu}(\mathbf{r})|\tilde{p}_\mu\rangle\langle\tilde{p}_\nu|$ is the density operator within the PAW at the position $\mathbf{r}$ [19,20]. The one-center term $\sum_{\mu\nu} Q_{\mu\nu}(\mathbf{r})|\tilde{p}_\mu\rangle\langle\tilde{p}_\nu|$ arises from the additive augmentation of the PAW.

The calculation is performed in two steps, starting with the contraction $\tilde{\Sigma} = -G\tilde{W}$ in real space and imaginary time. In analogy to the four auxiliary components of the Green's function, we obtain four quantities that store the self-energy,

$$\tilde{\underline{\Sigma}}^{(1)}(\mathbf{r},\mathbf{R}',i\tau) = -\underline{G}^{(1)}(\mathbf{r},\mathbf{R}',i\tau)\tilde{W}(\mathbf{r},\mathbf{R}',i\tau), \quad (55)$$

$$\tilde{\underline{\Sigma}}^{(2)}(\mu,\mathbf{R}',i\tau) = -\sum_{\nu} D^{(2)}(\mu\nu,\mathbf{R}',i\tau)\underline{G}^{(2)}(\nu,\mathbf{R}',i\tau), \quad (56)$$

$$\tilde{\underline{\Sigma}}^{(3)}(\mathbf{r},\beta,i\tau) = -\sum_{\alpha} \underline{G}^{(3)}(\mathbf{r},\alpha,i\tau)D^{(3)}(\mathbf{r},\alpha\beta,i\tau), \quad (57)$$

$$\tilde{\underline{\Sigma}}^{(4)}(\mu,\beta,i\tau) = -\sum_{\nu\alpha} \underline{G}^{(4)}(\nu,\alpha,i\tau)D^{(4)}(\mu\nu,\alpha\beta,i\tau), \quad (58)$$

where the auxiliary quantities $D^{(2)}$, $D^{(3)}$, and $D^{(4)}$ are defined as

$$D^{(2)}(\mu\nu,\mathbf{R}',i\tau) = \sum_{\mathbf{r}\in C} Q_{\mu\nu}(\mathbf{r})\tilde{W}(\mathbf{r},\mathbf{R}',i\tau), \quad (59)$$

$$D^{(3)}(\mathbf{r},\alpha\beta,i\tau) = \sum_{\mathbf{R}'\in S} \tilde{W}(\mathbf{r},\mathbf{R}',i\tau)Q_{\alpha\beta}(\mathbf{R}'), \quad (60)$$

$$D^{(4)}(\mu\nu,\alpha\beta,i\tau) = \sum_{\mathbf{r}\in C}\sum_{\mathbf{R}'\in S} Q_{\mu\nu}(\mathbf{r})\tilde{W}(\mathbf{r},\mathbf{R}',i\tau)Q_{\alpha\beta}(\mathbf{R}'). \quad (61)$$

Again, the Green's functions and the screened interaction are stored in reciprocal space and Fourier transformed to the real space on the fly, whenever they are required. To reduce the memory requirements, the self-energy is also stored in reciprocal space. In the second step, the matrix elements of the self-energy are then obtained as

$$\begin{aligned}\tilde{\underline{\Sigma}}^{(\mathbf{k})}_{\mathrm{nn}}(i\tau) = &\sum_{\mathbf{r}\in C}\sum_{\mathbf{R}'\in S}\tilde{\psi}^*_{n\mathbf{k}}(\mathbf{r})\tilde{\underline{\Sigma}}^{(1)}(\mathbf{r},\mathbf{R}',i\tau)\tilde{\psi}_{n\mathbf{k}}(\mathbf{R}') \\ &+ \sum_{\mu}\sum_{\mathbf{R}'\in S}\langle\tilde{\psi}_{n\mathbf{k}}|\tilde{p}_\mu\rangle\tilde{\underline{\Sigma}}^{(2)}(\mu,\mathbf{R}',i\tau)\tilde{\psi}_{n\mathbf{k}}(\mathbf{R}') \\ &+ \sum_{\mathbf{r}\in C}\sum_{\beta}\tilde{\psi}^*_{n\mathbf{k}}(\mathbf{r})\tilde{\underline{\Sigma}}^{(3)}(\mathbf{r},\beta,i\tau)\langle\tilde{p}_\beta|\tilde{\psi}_{n\mathbf{k}}\rangle \\ &+ \sum_{\mu}\sum_{\beta}\langle\tilde{\psi}_{n\mathbf{k}}|\tilde{p}_\mu\rangle\tilde{\underline{\Sigma}}^{(4)}(\mu,\beta,i\tau)\langle\tilde{p}_\beta|\tilde{\psi}_{n\mathbf{k}}\rangle. \quad (62)\end{aligned}$$

One point that should be mentioned here is that in the present implementation, the core-valence exchange-correlation interaction is treated in the same way as in the conventional GW implementation, that is, the Hartree-Fock approximation is used. This is found to be more reliable than LDA since the GW self-energy approaches the bare Fock exchange operator in the short-wavelength limit [22].

3. Evaluation of $\tilde{\Sigma}^{(\mathbf{k})}_{nn}(i\omega)$ by CT+ST

After the matrix elements of the self-energy along the imaginary time have been obtained, one needs to Fourier transform them to the imaginary frequency domain to calculate the QP energies. However, the self-energy (like the Green's function) is neither an even nor an odd function in imaginary time/frequency. Hence, we split the Green's functions into even and odd parts

$$G(i\tau) = \frac{[G(i\tau) + G(-i\tau)]}{2} + \frac{[G(i\tau) - G(-i\tau)]}{2}. \quad (63)$$

Then, the self-energy along the imaginary frequency is given by the temporal Fourier transformation

$$\begin{aligned}\tilde{\Sigma}(i\omega) =& -\int_{-\infty}^{\infty} d\tau\, G(i\tau)\tilde{W}(i\tau)e^{i\omega\tau} \\ =& 2\int_0^{\infty} d\tau\, \tilde{\Sigma}^c(i\tau)\cos(\omega\tau) + 2i\int_0^{\infty} d\tau\, \tilde{\Sigma}^s(i\tau)\sin(\omega\tau),\end{aligned} \quad (64)$$

where the cosine $\tilde{\Sigma}^c$ and sine $\tilde{\Sigma}^s$ part read as

$$\tilde{\Sigma}^c(i\tau) = -\frac{1}{2}[\overline{G}(i\tau) + \underline{G}(-i\tau)]\tilde{W}(i\tau), \quad (65)$$

$$\tilde{\Sigma}^s(i\tau) = -\frac{1}{2}[\overline{G}(i\tau) - \underline{G}(-i\tau)]\tilde{W}(i\tau). \quad (66)$$

Therefore, the corresponding diagonal matrix elements are given by

$$\tilde{\Sigma}^{c(\mathbf{k})}_{\mathrm{nn}}(i\tau) = \frac{1}{2}\left[\overline{\tilde{\Sigma}}^{(\mathbf{k})}_{\mathrm{nn}}(i\tau) + \underline{\tilde{\Sigma}}^{(\mathbf{k})}_{\mathrm{nn}}(-i\tau)\right], \quad (67)$$

$$\tilde{\Sigma}^{s(\mathbf{k})}_{\mathrm{nn}}(i\tau) = \frac{1}{2}\left[\overline{\tilde{\Sigma}}^{(\mathbf{k})}_{\mathrm{nn}}(i\tau) - \underline{\tilde{\Sigma}}^{(\mathbf{k})}_{\mathrm{nn}}(-i\tau)\right]. \quad (68)$$

Finally, the diagonal matrix elements of the correlated self-energy along the imaginary frequency axis are evaluated as

$$\tilde{\Sigma}^{(\mathbf{k})}_{\mathrm{nn}}(i\omega) = \tilde{\Sigma}^{c(\mathbf{k})}_{\mathrm{nn}}(i\omega) + \tilde{\Sigma}^{s(\mathbf{k})}_{\mathrm{nn}}(i\omega), \quad (69)$$

where $\tilde{\Sigma}^{c(\mathbf{k})}_{\mathrm{nn}}(i\omega)$ and $\tilde{\Sigma}^{s(\mathbf{k})}_{\mathrm{nn}}(i\omega)$, respectively, are determined by discrete CT and ST:

$$\tilde{\Sigma}^{c(\mathbf{k})}_{\mathrm{nn}}(i\omega_k) = \sum_{j=1}^{N} \gamma_{kj}\cos(\omega_k\tau_j)\tilde{\Sigma}^{c(\mathbf{k})}_{\mathrm{nn}}(i\tau_j), \quad (70)$$

$$\tilde{\Sigma}^{s(\mathbf{k})}_{\mathrm{nn}}(i\omega_k) = i\sum_{j=1}^{N} \lambda_{kj}\sin(\omega_k\tau_j)\tilde{\Sigma}^{s(\mathbf{k})}_{\mathrm{nn}}(i\tau_j). \quad (71)$$

E. Calculation of QP energies and spectral functions

In this section, we describe the calculation of QP energies and spectral functions, which corresponds to the last two steps in Fig. 1.

1. Analytic continuation

In our present implementation, the self-energy and Green's function are calculated in the imaginary frequency domain. However, the experimental observables of interest, such as QP energies and spectral functions, are obviously measured all along the real frequency axis. This implies that an analytic continuation from the imaginary to the real frequency domain has to be performed. Given that our self-energy is exact in the sense that there are no stochastic noises [unlike the Green's functions $G(i\tau)$ obtained from quantum Monte Carlo (QMC) simulations], here we utilize the N-point Padé approximant and employ Thiele's reciprocal difference method [18]

$$P_N(z) = \frac{a_1}{1+}\,\frac{a_2(z-z_1)}{1+}\,\cdots\,\frac{a_N(z-z_{N-1})}{1+(z-z_N)g_{N+1}(z)}. \quad (72)$$

Here, the complex coefficients a_n are obtained by the following recursion relations:

$$a_n = g_n(z_n), \quad g_1(z_n) = f_n, \quad n = 1,\ldots,N \quad (73)$$

$$g_n(z) = \frac{g_{n-1}(z_{n-1}) - g_{n-1}(z)}{(z-z_{n-1})g_{n-1}(z)}, \quad n \geqslant 2. \quad (74)$$

It is straightforward to prove [18] that $P_N(z_j) = f_j$ holds for the known point-value pairs $\{z_i, f_j\}_{j=1}^{N}$ of the function $f(z)$ (the diagonal elements of the self-energy in the G_0W_0r case). It should be emphasized that Thiele's reciprocal difference method is fairly stable, whereas a naive computation of Padé coefficients usually yields numerical instabilities. Thiele's reciprocal difference method has been successfully applied to the analytic continuation of dynamic response functions [24]. In the following, we show that this method can be used for the accurate prediction of GW QP energies and spectral functions as well (see Sec. IV).

2. Evaluation of $E^{QP}_{n\mathbf{k}}$ and $A_{n\mathbf{k}}(\omega)$

After the diagonal elements of the self-energy along the real frequency axis $\Sigma^{(\mathbf{k})}_{\mathrm{nn}}(\omega)$ including contributions from the core-valence exchange correlation, bare exchange and dynamical interactions [hereafter we denote it as $\Sigma_{n\mathbf{k}}(\omega)$] have been obtained by the analytic continuation, the QP energies are evaluated as in conventional GW implementations. This means, for the single-shot GW calculations, the QP energies are calculated to first order by linearizing the self-energy around the DFT single-particle eigenvalues $\epsilon_{n\mathbf{k}}$:

$$\Sigma_{n\mathbf{k}}\left(E^{\mathrm{QP}}_{n\mathbf{k}}\right) = \Sigma_{n\mathbf{k}}(\epsilon_{n\mathbf{k}}) + \left.\frac{\partial\Sigma_{n\mathbf{k}}(\omega)}{\partial\omega}\right|_{\omega=\epsilon_{n\mathbf{k}}}\left(E^{\mathrm{QP}}_{n\mathbf{k}} - \epsilon_{n\mathbf{k}}\right). \quad (75)$$

After some simple derivations, the QP energy is calculated as [22]

$$\begin{aligned}E^{\mathrm{QP}}_{n\mathbf{k}} =& \epsilon_{n\mathbf{k}} + Z_{n\mathbf{k}}\mathrm{Re}[\langle\psi_{n\mathbf{k}}|\hat{T} + \hat{V}_{n-e} + \hat{V}_H|\psi_{n\mathbf{k}}\rangle \\ &+ \Sigma_{n\mathbf{k}}(\epsilon_{n\mathbf{k}}) - \epsilon_{n\mathbf{k}}],\end{aligned} \quad (76)$$

where $\hat{T}$ is the kinetic energy operator, $\hat{V}_{n-e}$ the nuclei potential, $\hat{V}_H$ the Hartree potential, and $Z_{n\mathbf{k}}$ the renormalization factor given by

$$Z_{n\mathbf{k}} = \left(1 - \left.\frac{\partial\,\mathrm{Re}[\Sigma_{n\mathbf{k}}(\omega)]}{\partial\omega}\right|_{\omega=\epsilon_{n\mathbf{k}}}\right)^{-1}. \quad (77)$$

In principle, one could calculate the QP energies by searching the root of equation $E^{\mathrm{QP}}_{n\mathbf{k}} = \mathrm{Re}[\langle\psi_{n\mathbf{k}}|\hat{T} + \hat{V}_{n-e} + \hat{V}_H|\psi_{n\mathbf{k}}\rangle + \Sigma_{n\mathbf{k}}(E^{\mathrm{QP}}_{n\mathbf{k}})]$ numerically. For solids, this does not make a sizable difference in the QP energies compared to the linearization. In this work, we therefore only show the calculated QP energies from the linearized version to compare with the conventional implementation where the linearization was used as well.

The spectral functions are calculated as the imaginary part of the interacting Green's function, which is calculated from the Dyson equation [25]

$$A_{n\mathbf{k}}(\omega) = \frac{1}{\pi}|\mathrm{Im}[G_{n\mathbf{k}}(\omega)]| = \frac{1}{\pi}\frac{|\mathrm{Im}[\Delta\Sigma_{n\mathbf{k}}(\omega)]|}{\{\omega - \epsilon_{n\mathbf{k}} - \mathrm{Re}[\Delta\Sigma_{n\mathbf{k}}(\omega)]\}^2 + \mathrm{Im}[\Delta\Sigma_{n\mathbf{k}}(\omega)]^2}, \tag{78}$$

where $\Delta\Sigma_{n\mathbf{k}}(\omega) = \langle\psi_{n\mathbf{k}}|\hat{T} + \hat{V}_{n-e} + \hat{V}_H|\psi_{n\mathbf{k}}\rangle + \Sigma_{n\mathbf{k}}(\omega) - \epsilon_{n\mathbf{k}}$.

III. TECHNICAL DETAILS

Our low-scaling GW scheme has been implemented in the Vienna *ab initio* simulation package (VASP) [27,28]. For all the calculations presented here, the ultrasoft (US) PAW potentials with an appendix (_GW) released with VASP.5.2 were used unless otherwise explicitly specified. These potentials are constructed by using additional projectors above the vacuum level and thus describe well the high-energy scattering properties of the atoms. The plane-wave cutoff for the orbitals was chosen to be the maximum of all elements in the considered material. The energy cutoff for the response function was chosen to be half of the plane-wave cutoff. To sample the Brillouin zone, $8 \times 8 \times 8$ k-point grids centered at the Γ point were used except for Cu where the grids were increased to $10 \times 10 \times 10$. For the tested materials, the experimental lattice constants at low temperature (if available, otherwise at room temperature) were used. The total number of bands was chosen to be 480, which is sufficient to obtain the converged QP energies for most of the materials considered, except for GaAs and ZnO where the convergence is very slow. It was suggested that thousands of orbitals are required for accurate predictions for ZnO [29], but this finite-basis-set correction is beyond the scope of this work. In fact, for the present setup the errors in some QP energies are large, with errors of, e.g., 0.5 eV for ZnO. For more accurate results, we refer to the previous publication by some of the present authors [30].

Clearly, the purpose of this work is not to basis set converge the calculations (this is of course possible with the present implementation, as it was possible in the standard framework). Instead, we restrict ourselves to validating the low-scaling GW implementation by comparing the results with the already widely used conventional GW implementation. Hence, the same setups (crystal structure, potential, k points, and so on) were used for both G_0W_0r and G_0W_0 calculations. In addition, finite-basis-set corrections for QP energies discussed in Ref. [30] are not taken into account for either G_0W_0r or G_0W_0 calculations.

The actual GW calculations involve three steps: (i) A self-consistent KS-DFT calculation was performed using the Perdew-Burke-Ernzerhof (PBE) functional [31]. (ii) The one-electron wave functions and eigenenergies of all unoccupied (virtual) orbitals spanned by the plane-wave basis set were evaluated by an exact diagonalization of the previously determined self-consistent KS Hamiltonian. (iii) The GW calculations were carried out. For all the materials considered, the number of imaginary-time/frequency points in G_0W_0r calculations was set to 20, whereas the number of real frequency points was chosen to be 200 for G_0W_0 calculations. Increasing the number of grid points further changes the QP energies by less than 0.01 eV.

IV. RESULTS

A. Results for semiconductors and insulators

Table I shows the QP energies and band gaps for the tested semiconductors and insulators predicted by G_0W_0r and G_0W_0. First, we emphasize that our G_0W_0@PBE results are consistent with previous calculations [26]. As expected, the band gaps calculated by G_0W_0@PBE are slightly underestimated compared to the experimental values. Improvements further towards experimental gaps have been achieved either by GW_0@PBE (iterating the one-electron energies only in G) [26], or by G_0W_0@HSE (using the hybrid functionals as a starting point) [32]. The best agreement with experimental values thus far has been achieved by $GW^{\mathrm{TC\text{-}TC}}$ (self-consistent GW with the vertex correction only in W) [33]. We note again that finite-basis-set corrections [30] have not been used here, which would increase the gap for ZnO by 0.3–0.4 eV, for instance. Second, one can see that the agreement between the results from G_0W_0r and G_0W_0 is remarkably good, validating our low-scaling GW implementation. Specifically, for the sp semiconductors and insulators (Si, SiC, C, BN, MgO, and LiF), the difference in QP energies and band gaps between G_0W_0r and G_0W_0 is not larger than 0.02 eV. This is even true for GaAs with localized d orbitals. Except for ZnO, the G_0W_0r seems to have the tendency to yield a slightly smaller downwards shift (below 0.02 eV) for valence and conduction bands compared to G_0W_0. However, for ZnO the difference of the calculated gaps between G_0W_0r and G_0W_0 is larger (0.04 eV) since the self-energy exhibits many poles from $d - p$ excitations at energies around -40 eV [see Fig. 2(e) below].

To further assess our low-scaling GW implementation, we plot the diagonal elements of the self-energies and spectral functions at the Γ point for some chosen bands around the Fermi level, as well as the QP shift versus the DFT eigenvalues for Si and ZnO in Fig. 2. The results obtained from the conventional G_0W_0 are also presented for comparison. Overall, the agreement between the results from G_0W_0r and G_0W_0 is very good, in particular for the region close to the Fermi level. Specifically, for Si, the self-energies and spectral functions (including the spectral background and contributions from plasmons) calculated by G_0W_0r agree nicely with the ones from G_0W_0 [see Figs. 2(a) and 2(b)]. This is achieved by employing Thiele's reciprocal difference method. Solving for the Padé coefficients directly, however, yields less satisfactory results (not shown here). For ZnO, the agreement in the self-energies and spectral functions is still good. Even the small satellites in bands Γ^{v}_{15} and Γ^{c}_{1} are reproduced [see the

TABLE I. Positions of conduction band (CB) minimum at Γ (Γ_c) and X (X_c), valence band (VB) maximum at X (X_v) with respect to the VB maximum at Γ, as well as the band gap. Spin-orbit coupling (SOC) and finite-basis-set corrections are not included. The crystal structures, lattice constants, and experimental band gaps are identical to Ref. [26] and references therein.

	Γ_c		X_c		X_v		Band gap			Crystal	Lattice
	G_0W_0r	G_0W_0	G_0W_0r	G_0W_0	G_0W_0r	G_0W_0	G_0W_0r	G_0W_0	Expt.	structure	constant (Å)
Si	3.22	3.23	1.24	1.25	−2.89	−2.89	1.15	1.16	1.17	diamond	5.430
GaAs	1.33	1.34	1.86	1.88	−2.79	−2.77	1.33	1.34	1.52	zinc-blende	5.648
SiC	7.40	7.41	2.30	2.31	−3.36	−3.35	2.30	2.31	2.40	zinc-blende	4.350
ZnO	2.10	2.06	6.73	6.66	−2.31	−2.28	2.10	2.06	3.44	zinc-blende	4.580
C	7.39	7.39	6.07	6.08	−6.66	−6.66	5.49	5.50	5.48	diamond	3.567
BN	11.14	11.14	6.16	6.17	−5.28	−5.27	6.16	6.17	6.1-6.4	zinc-blende	3.615
MgO	7.27	7.27	11.47	11.48	−1.55	−1.54	7.27	7.27	7.83	rock salt	4.213
LiF	13.68	13.68	20.20	20.20	−1.21	−1.19	13.68	13.68	14.20	rock salt	4.010

inset in Fig. 2(d)]. However, satellites far from the Fermi level have been smoothed by the analytic continuation.

In contrast, there exist larger deviations in the region far away from the Fermi level. As shown in Figs. 2(c) and 2(f), the difference in the QP shift between G_0W_0r and G_0W_0 increases as the binding energies increase above 4 eV. The reason can be easily understood. Considering band Γ_1 of Si for instance, the QP peak is not sharp. Instead, it is broadened with a width of around 5 eV, as shown in Fig. 2(b). Therefore, it is difficult to obtain the exact position of the quasiparticle. This is true for both G_0W_0r and G_0W_0. In addition, the QP peaks measured from the angle-resolved photoelectron spectroscopy (ARPES) would be as broad as in the GW approximation so that the errors are in fact negligible compared to the width of the peak.

B. Results for metals

Now, we turn to the QP calculations for metals where some extra considerations are required. For metallic systems, there exists a nonvanishing probability that an electron is excited within one and the same band. These transitions are called intraband transitions and lead to the so-called Drude term for the long-wavelength limit ($\mathbf{q} \to 0$). Following similar strategies as in Refs. [37,38], we derived the head of the intraband dielectric function in the imaginary frequency domain

$$\varepsilon^{\text{intra}}_{\alpha\beta}(i\omega) = \frac{\bar{\omega}^2_{\alpha\beta}}{\omega^2}. \tag{79}$$

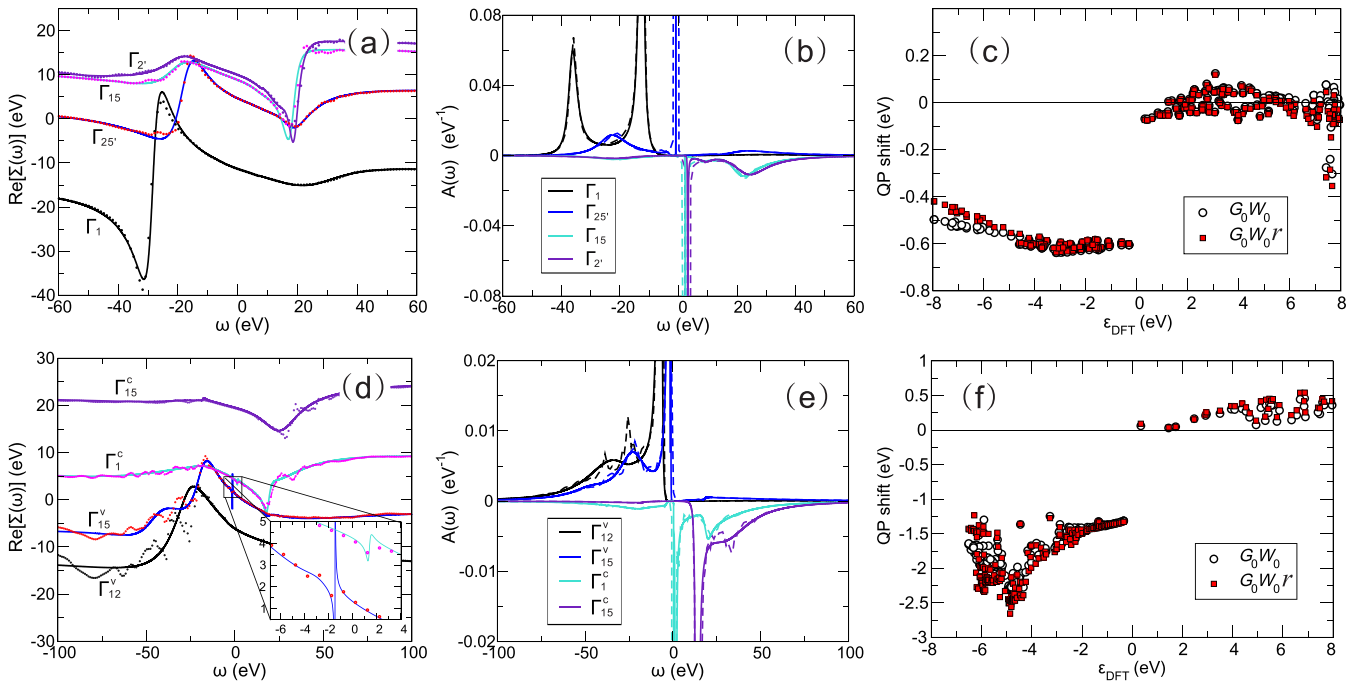

FIG. 2. The real part of the diagonal elements of the self-energy $\mathrm{Re}[\Sigma_{n\mathbf{k}}(w)]$ [(a), (d)], and the spectral functions $A_{n\mathbf{k}}(w)$ of the Green's functions at the Γ point [(b), (e)], as well as the QP shift versus the DFT eigenvalues [(c), (f)] for Si (first row) and ZnO (second row). The solid lines and dotted (broken) lines in [(a), (d), (b), and (e)], respectively, specify the results from G_0W_0r and G_0W_0. Note that the sign of the spectral functions for the unoccupied states in (b) and (e) is intentionally reversed for clarity. The inset in (d) shows the zoom-in plot for the local satellites.

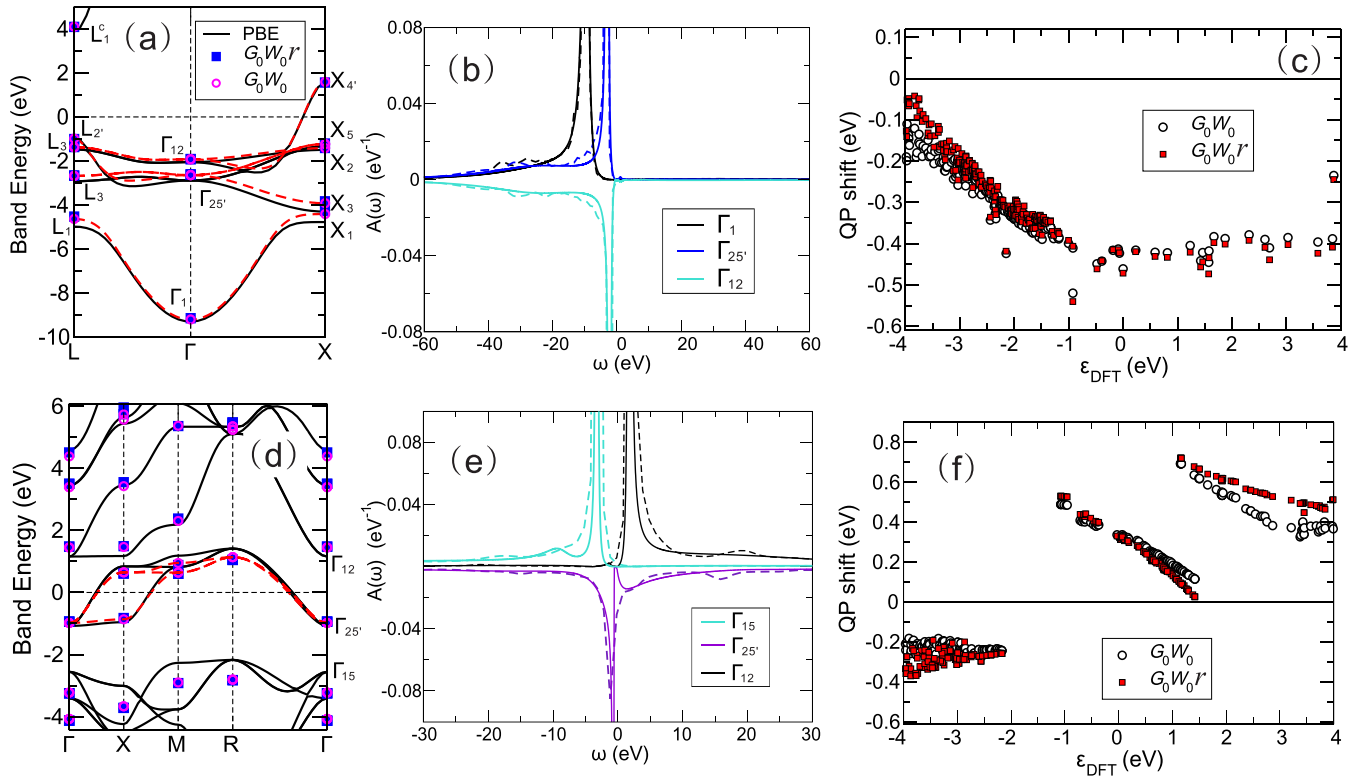

FIG. 3. Band structures [(a), (d)], spectral functions $A_{n\mathbf{k}}(w)$ of the Green's functions at the Γ point [(b), (e)], and QP shift versus the DFT eigenvalues [(c), (f)] for Cu (first row) and $SrVO_3$ (second row). Note that in (a) and (d) the red broken lines specify the Wannier interpolated band structure from G_0W_0. PBE and GW Fermi energies are aligned at zero.

Here, the tensor $\bar{\omega}_{\alpha\beta}$ is the plasma frequency and its square is defined as

$$\bar{\omega}^2_{\alpha\beta} = -\frac{4\pi e^2}{\Omega_C}\sum_{n\mathbf{k}} 2\frac{\partial f(\epsilon_{n\mathbf{k}})}{\partial \epsilon_{n\mathbf{k}}}\left(\mathbf{e}_\alpha \cdot \frac{\partial \epsilon_{n\mathbf{k}}}{\partial \mathbf{k}}\right)\left(\mathbf{e}_\beta \cdot \frac{\partial \epsilon_{n\mathbf{k}}}{\partial \mathbf{k}}\right), \quad (80)$$

where the factor of 2 is due to the spin-degenerate systems considered here, Ω_C is the volume of the unit cell, and $\mathbf{e}_\alpha$ is the unit vector along the Cartesian coordinate α. It should be noted that the intraband transitions are only nonvanishing for the head of the dielectric functions. For the wings and body, they are both zero.

As a test, we calculated the QP energies for the metals Cu and $SrVO_3$ and compare the results with the ones from the conventional G_0W_0 in Figs. 3(a) and 3(d). To guide the eye, the PBE band structures and interpolated G_0W_0 QP band structures obtained with the WANNIER90 code [39,40] are also displayed. One can see that good agreement between G_0W_0r and G_0W_0 is achieved for both Cu and $SrVO_3$, indicating that our low-scaling GW implementation is robust and applies

TABLE II. QP energies (eV) of Cu predicted by G_0W_0r and G_0W_0 using the norm-conserving (NC) GW PAW potential (Cu_sv_GW_nc). Basis-set-corrected G_0W_0 QP energies are also given for comparison. The labeling of the high-symmetry points is shown in Fig. 3(a). The results are compared to the pseudopotential plane-wave (PPW) values obtained by Marini *et al.* [34] and the full potential linear muffin-tin orbital (LMTO) calculations by Zhukov *et al.* [35]. Experimental data are taken from Ref. [36].

		PBE	G_0W_0r	G_0W_0	Corrected G_0W_0	PPW [34]	LMTO [35]	Expt. [36]
Positions of d bands	Γ_{12}	−2.05	−1.92	−1.92	−2.11	−2.81	−2.36	−2.78
	X_5	−1.33	−1.22	−1.23	−1.45	−2.04	−1.63	−2.01
	L_3	−1.47	−1.36	−1.37	−1.58	−2.24	−1.78	−2.25
	$\Gamma_{12}-\Gamma_{25'}$	0.84	0.70	0.72	0.69	0.60	0.81	0.81
Widths of d bands	X_5-X_3	2.97	2.61	2.68	2.60	2.49	2.92	2.79
	X_5-X_1	3.44	3.05	3.18	3.10	2.90	3.37	3.17
	L_3-L_3	1.44	1.30	1.31	1.26	1.26	1.43	1.37
	L_3-L_1	3.51	3.16	3.26	3.16	2.83	3.42	2.91
Positions of s/p bands	Γ_1	−9.29	−9.14	−9.20	−9.18	−9.24	−9.35	−8.60
	$L_{2'}$	−0.92	−1.00	−0.98	−1.02	−0.57	−0.92	−0.85
L gap	$L_1^c-L_{2'}$	4.80	5.09	5.08	4.98	4.76	4.78	4.95

to metals as well. However, if we take a closer look at the spectral functions, as shown in Figs. 3(b) and 3(e), we observed that although the main QP peaks are well reproduced, the plasma and some satellites are again smoothed by the analytic continuation.

Figures 3(c) and 3(f) further show the QP shift versus DFT eigenvalues for Cu and $SrVO_3$, respectively. For the noble metal Cu, in the energy region of the plot, overall, the QP shift difference between G_0W_0r and G_0W_0 is not exceeding 0.1 eV. Analogous behavior as for Si is observed for Cu. The further one moves away from the Fermi level, the larger is the QP shift difference. This is due to the large broadening of the QP peak for the corresponding bands. This is also true for the metal $SrVO_3$. The negative slope of the QP shift between -1.5 and 1.5 eV implies a shrinking of the t_{2g} bands as compared to the DFT results, which was observed in other GW studies as well [25,41]. In the region far away from the Fermi level QP differences are visible, but the maximum difference is smaller than 0.2 eV.

In Table II, we show in detail the QP energies of Cu predicted by G_0W_0r and G_0W_0 and compare our results with other theoretical calculations and experiment. The most significant error in the PBE one-electron energies is the wrong description of the absolute positions and the bandwidth of the d bands. For instance, the highest d band at X_5 is located at -1.33 eV in PBE, 0.68 eV above the experimental value of -2.01 eV [36]. The bandwidth of the d bands is widened compared to the experiment (see $X_5 - X_1$). Unfortunately, though, our G_0W_0 does not improve the results significantly. It shrinks the bandwidth of the d bands towards the experimental values, but predicts worse positions for the d bands than PBE. However, as already mentioned basis-set errors might be substantial for the 480 bands employed in the present case. To improve the results, finite-basis-set corrections were used as discussed in Ref. [30]. Indeed, with these corrections the absolute positions of the d bands are lowered by about 0.2 eV. Agreement with the full potential LMTO method [35] is then reasonable. However, our QP d-band energies are still way above those of Marini *et al.* [34]. We are pretty confident that the good agreement of these calculations with experiment is largely fortuitous: the applied pseudopotentials somehow canceled the errors introduced by the G_0W_0 approximation. Compared to the LMTO data, we note that all our QP energies are shifted upwards by 0.2 eV (except for $L_{2'}$). Of course, our QP energies are reported with respect to the G_0W_0 Fermi energy, whereas Ref. [35] does not mention how and whether the Fermi energy was determined at the G_0W_0 level. Using the DFT Fermi energy would improve agreement with Ref. [35]. For the widths of the d bands, the present results are in very good agreement with experiment, though, slightly improving upon the LMTO data, which were generally above the experimental data.

We feel that the residual errors compared to experiment are to be expected and arise from (i) the neglect of self-consistency (the DFT d orbitals of Cu are most likely too strongly hybridized with the sp states), and (ii) spurious self-interactions in the GW approximation. The latter error can be only eliminated via the inclusion of vertex corrections in the self-energy. Indeed, the importance of vertex corrections has been highlighted for predicting the ionization potentials and d-electron binding energies of solids [42], with typical corrections for the d bands of 0.7 eV.

C. Time complexity for large systems

In order to investigate the scaling with respect to the system size in our new implementation, we performed G_0W_0r calculations on different bulk Si diamond supercells with 16, 24, 36, and 54 atoms using the Γ point only. For comparison, similar calculations have been done for the conventional G_0W_0 code. Our G_0W_0r implementation displays clearly a better than cubic scaling in the system size, as shown in Table III. The reason for this good scaling is that the contraction steps such as GG and GW scale only quadratically in system size, and for the number of atoms considered here, construction of the Green's function and manipulations of the self-energy matrix, which scale profoundly cubically, are not yet dominating the total compute time. Furthermore, it needs to be mentioned that the G_0W_0r compute time includes the calculation of the full nondiagonal self-energy at all frequency points (including *all* off-diagonal elements), whereas the G_0W_0 code calculates only few diagonal elements of the self-energy for the occupied and some unoccupied states. Despite this, the G_0W_0r code substantially outperforms the older G_0W_0 code. Concerning scaling, the old G_0W_0 code shows a slightly less beneficial scaling, nevertheless, it is also closer to cubic than quartic in system size. This relates to the fact that the quartic part (construction of polarizability and self-energy in orbital basis) is done using high-efficiency BLAS level 3 calls, and hence this part becomes only dominant for very large systems, typically beyond 100 atoms.

To test the scaling with respect to the number of k points, we performed calculations on a bulk Si diamond supercell with 16 atoms using 64 cores. We note that the new code does not yet perform optimally if the number of cores exceeds the number of atoms. This and the need to use a complex

TABLE III. Timings in minutes for G_0W_0r and G_0W_0 calculations for different bulk Si diamond supercells. The calculations were done for 64 QP energies using the Γ point only and using the real valued Γ only VASP version.

		Times		Times × cores/atoms^3 × 10^3	
Atoms	Cores	G_0W_0	G_0W_0r	G_0W_0	G_0W_0r
16	16	9.18	2.50	35.86	9.78
24	20	18.94	4.14	27.40	5.99
36	48	41.68	5.65	42.88	5.82
54	64	104.53	12.07	42.49	4.91

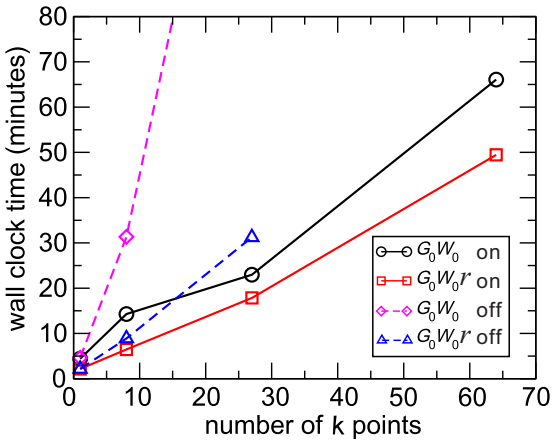

FIG. 4. Computational time for G_0W_0r and G_0W_0 calculations with symmetry switched on or off on a bulk Si diamond supercell with 16 atoms as a function of the number of k points (in the full Brillouin zone). Note that the computational time of G_0W_0 for the $3 \times 3 \times 3$ case without symmetry is about 162 minutes, which is not shown in the figure. The calculations are done for 64 QP energies using 64 cores.

code version explain why the timings for a single k point in Fig. 4 are hardly better than for 16 cores shown in Table III. As shown in Fig. 4, the computational demand increases almost perfectly linear in the number of k points for G_0W_0r. The slight deviation for the $4 \times 4 \times 4$ case arises from the need to pick a less efficient parallelization strategy for this k-point set to be able to perform the calculation using the memory available on 64 cores. In contrast, G_0W_0 shows a roughly quadratic scaling in the number of k points. The bad scaling of the old code is, however, somewhat masked by its efficient handling of symmetry. The old implementation uses small point-group operations compatible with the considered momentum transfer q, whereas the new code uses yet no symmetry when contracting GG or GW. Concomitantly, if symmetry is switched off, the new code becomes only slower by a factor 2 for the $3 \times 3 \times 3$ k points, whereas the time for the old code increases to 162 minutes (off the scale, see the blue and pink broken lines in Fig. 4). Therefore, one would expect that the G_0W_0r code outperforms the old G_0W_0 code, in particular, if large low-symmetry unit cells are used and/or if many k points are used. It is, however, also clear that the old code can be competitive or superior for small high-symmetry unit cells, even if many k points are used to sample the Brillouin zone. For instance, for a cubic diamond unit cell or for fcc Cu, the old code is usually much faster than the new GW code.

V. CONCLUSIONS

In conclusion, we present a promising low-scaling GW implementation within the PAW method, which allows for fast QP calculations with a scaling that is roughly cubic in the system size and linear in the number of k points used to sample the Brillouin zone. All implementation details have been given. We apply the method to predict the quasiparticle energies and spectral functions for typical semiconductors, insulators, and metals. Comparison of the results with the ones from conventional GW calculations shows a good agreement between the two implementations. Specifically, for semiconductors and insulators, the positions of the bands and the band gaps agree within 0.02 eV for all considered materials except for ZnO. Due to the low scaling of our new GW implementation, we believe that our method has great potential for applications, in particular for large unit cells. In addition, our GW self-energies are obtained in imaginary-time/frequency domain, which will facilitate an elegant combination of GW with DMFT, enhancing the predictive abilities of GW+DMFT for large correlated systems. Finally, we have shown that with typical compute times around 12 minutes on 64 cores for 54 silicon atoms, GW calculations are becoming a commodity. We believe this will greatly help to establish methods beyond density functional theory in the realm of materials modeling.

ACKNOWLEDGMENTS

This work was supported by the China Scholarship Council (CSC)-Austrian Science Fund (FWF) Scholarship Program and FWF within the SFB ViCoM (Grant No. F 41) and I597-N16 (research unit FOR 1346 of the Deutsche Forschungsgemeinschaft and FWF). Supercomputing time on the Vienna Scientific cluster (VSC) is gratefully acknowledged. J.K. is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 658705.

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[2] W. Aulbur, L. Jönsson, and J. Wilkins, Solid State Phys. **54**, 1 (2000).
[3] L. Hedin, Phys. Rev. **139**, A796 (1965).
[4] L. Hedin and S. Lundqvist, *Solid State Physics* (Academic, New York, 1969).
[5] F. Aryasetiawan, *Advances in Condensed Matter Science* (Gordon and Breach, New York, 2000).
[6] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[7] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[8] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006).
[9] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003).
[10] K. Held, Adv. Phys. **56**, 829 (2007).
[11] *The LDA+DMFT Approach to Strongly Correlated Materials*, Lecture Notes of the Autumn School 2011, Hands-On LDA+DMFT, edited by E. Pavarini, E. Koch, D. Vollhardt, and A. E.. Lichtenstein (Forschungszentrum, Jülich, 2011).
[12] S. Adler, Phys. Rev. **126**, 413 (1962).
[13] N. Wiser, Phys. Rev. **129**, 62 (1963).
[14] H. N. Rojas, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **74**, 1827 (1995).

[15] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. White, and W. Godby, Comput. Phys. Commun. **125**, 105 (2000).
[16] M. Kaltak, J. Klimeš, and G. Kresse, J. Chem. Theory Comput. **10**, 2498 (2014).
[17] M. Kaltak, J. Klimeš, and G. Kresse, Phys. Rev. B **90**, 054115 (2014).
[18] G. A. J. Baker, *Essentials of Padé Approximants* (Academic, New York, 1975).
[19] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[20] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[21] Note that the auxiliary Green's functions in Eqs. (39)–(41) of Ref. [17] represent Eqs. (38)–(40).
[22] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
[23] J. Paier, R. Hirschl, M. Marsman, and G. Kresse, J. Chem. Phys. **122**, 234102 (2005).
[24] K.-H. Lee and K. J. Chang, Phys. Rev. B **54**, R8285 (1996).
[25] T. Miyake, C. Martins, R. Sakuma, and F. Aryasetiawan, Phys. Rev. B **87**, 115110 (2013).
[26] M. Shishkin and G. Kresse, Phys. Rev. B **75**, 235102 (2007).
[27] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[28] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[29] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **105**, 146401 (2010).
[30] J. Klimeš, M. Kaltak, and G. Kresse, Phys. Rev. B **90**, 075125 (2014).
[31] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); **78**, 1396 (1997).
[32] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, Phys. Rev. B **76**, 115109 (2007).
[33] M. Shishkin, M. Marsman, and G. Kresse, Phys. Rev. Lett. **99**, 246403 (2007).
[34] A. Marini, G. Onida, and R. Del Sole, Phys. Rev. Lett. **88**, 016403 (2001).
[35] V. P. Zhukov, E. V. Chulkov, and P. M. Echenique, Phys. Rev. B **68**, 045102 (2003).
[36] R. Courths and S. Hüfner, Phys. Rep. **112**, 53 (1984).
[37] M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, and F. Bechstedt, Phys. Rev. B **73**, 045112 (2006).
[38] C. Ambrosch-Draxl and J. O. Sofo, Comput. Phys. Commun. **175**, 1 (2006).
[39] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).
[40] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).
[41] J. M. Tomczak, M. Casula, T. Miyake, and S. Biermann, Phys. Rev. B **90**, 165138 (2014).
[42] A. Grüneis, G. Kresse, Y. Hinuma, and F. Oba, Phys. Rev. Lett. **112**, 096401 (2014).