Computer Physics Communications 219 (2017) 407–414

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Linearized self-consistent quasiparticle GW method: Application to semiconductors and simple metals☆

A.L. Kutepov [a,*], V.S. Oudovenko [a], G. Kotliar [a,b]

[a] Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08856, United States
[b] Brookhaven National Laboratory, NY, United States

ARTICLE INFO

Article history:
Received 27 June 2016
Received in revised form 28 April 2017
Accepted 11 June 2017
Available online 23 June 2017

ABSTRACT

We present a code implementing the linearized quasiparticle self-consistent GW method (LQSGW) in the LAPW basis. Our approach is based on the linearization of the self-energy around zero frequency which differs it from the existing implementations of the QSGW method. The linearization allows us to use Matsubara frequencies instead of working on the real axis. This results in efficiency gains by switching to the imaginary time representation in the same way as in the space time method. The all electron LAPW basis set eliminates the need for pseudopotentials. We discuss the advantages of our approach, such as its N^3 scaling with the system size N, as well as its shortcomings.

We apply our approach to study the electronic properties of selected semiconductors, insulators, and simple metals and show that our code produces the results very close to the previously published QSGW data. Our implementation is a good platform for further many body diagrammatic resummations such as the vertex-corrected GW approach and the GW+DMFT method.

Program summary

Program Title: LqsgwFlapw

Program Files doi: http://dx.doi.org/10.17632/cpchkfty4w.1

Licensing provisions: GNU General Public License

Programming language: Fortran 90

External routines/libraries: BLAS, LAPACK, MPI (optional)

Nature of problem: Direct implementation of the GW method scales as N^4 with the system size, which quickly becomes prohibitively time consuming even in the modern computers.

Solution method: We implemented the GW approach using a method that switches between real space and momentum space representations. Some operations are faster in real space, whereas others are more computationally efficient in the reciprocal space. This makes our approach scale as N^3.

Restrictions: The limiting factor is usually the memory available in a computer. Using 10 GB/core of memory allows us to study the systems up to 15 atoms per unit cell.

Published by Elsevier B.V.

1. Introduction

The GW method was proposed by Hedin [1] and was first applied to real materials by Hybertsen and Louie [2] and by Godby et al. [3]. One can find many successful implementations of this method using plane wave basis sets in the literature such as BerkeleyGW [4], ABINIT [5], and West [6]. There is also code implementing the GW in an all electron basis set such as exciting [7] and SPEX [8].

Numerous computational developments exist in this area (see for example [9–11] and references therein). Our present study builds on work by Rieger et al. on the space–time method [12] and from the work by Ku and Eguiluz on the application of Matsubara time in GW calculations [13].

In this paper we focus on the implementation and testing of the self-consistent quasi-particle GW (QSGW) method which is a promising tool for studying the electronic structure of the moderately correlated solids, atoms, and molecules [14]. While the QSGW usually overestimates the widths of the spectral features in

☆ This paper and its associated computer program are available via the Computer Physics Communication homepage on ScienceDirect (http://www.sciencedirect.com/science/journal/00104655).

* Corresponding author.
E-mail address: kutepov@physics.rutgers.edu (A.L. Kutepov).

http://dx.doi.org/10.1016/j.cpc.2017.06.012
0010-4655/Published by Elsevier B.V.

materials (such as band gaps, band widths, and exchange splitting), it is generally more accurate than the local density approximation (LDA). The QSGW also has clear advantages in comparison with another popular approximation — the so called "one shot" GW method (one GW iteration after the self-consistent LDA calculation), as it does not depend on the starting point.

All previous implementations of the QSGW method [15–17] are real frequency implementations. We have found however, that similar results can be obtained with an approach based on the imaginary frequency representation (Matsubara). We exploit the fact that one can easily transform the functions from the imaginary frequency to the imaginary time (and back) to improve the computational efficiency. We use all-electron approximation (Linear Augmented Plane Wave method, LAPW) as a basis for our approach, thus avoiding the need in pseudopotentials. We use abbreviation LQSGW for our (linearized) version of QSGW method.

Our implementation of the GW method was outlined in our previous work [18], where we presented the general scheme of the approach with detailed description of the self-consistent GW method (scGW) in its fully relativistic form and the application of the scheme to Am and Pu. Also, we provided there the total energy evaluation which was used earlier in Ref. [19]. In this work we present the non-relativistic version of scGW/LQSGW method with special emphasis on the scaling of the most time consuming steps of the algorithm. Particularly, we stress on the overall scaling N^3 with the system size which is similar to the scaling of LDA. In the end we present a few numerical examples obtained using quasiparticle approximation. The code, implementing the scGW approach was used (as an initial step) to solve the Hedin's equations [1] with a non-trivial vertex function for a number of materials [20,21]. The implementation of the LQSGW approach was used recently in Ref. [22] in the combination of the LQSGW and the one-shot DMFT (Dynamical Mean Field Theory).

2. The basics of the GW method

Below we outline the basic formulas of the method introduced earlier in Ref. [18]. In a self-consistent calculation, one has to perform a certain number of iterations until the self-consistency is achieved. In the GW method, the input for every iteration is the Green function G. For a given G, we need to perform a few intermediate steps, such as calculation of the function of the polarizability

$$P(12;\tau) = -G(12;\tau)G(21;\beta-\tau), \tag{1}$$

the screened interaction

$$W(12;\nu) = V(12) + \int d(34)V(13)P(34;\nu)W(42;\nu), \tag{2}$$

the self-energy

$$\Sigma(12;\tau) = -G(12;\tau)W(21;\tau), \tag{3}$$

and the new Green's function

$$G(12;\omega) = G_0(12;\omega) + \int d(34)G_0(13;\omega)\Sigma(34;\omega)G(42;\omega). \tag{4}$$

In the above equations, the numbers (e.g. 12, 34) are used to denote combined space+spin variables, τ is the Matsubara's time, ν and ω are the Matsubara's frequencies (bosonic and fermionic correspondingly), and β is the inverse temperature. In the LQSGW approach, the last step (Eq. (4)) is replaced with a special construction of the quasiparticle Green's function, which was introduced in [18]

$$G[G_0;\Sigma] \to G. \tag{5}$$

The details of this step in our implementation of the LQSGW approach are given in Section 5.

3. Representation of the band states in the LAPW method

In the LAPW method [23], one represents the band states in the interstitial region as a linear combination of the plane waves

$$\Psi_\lambda^{\alpha\mathbf{k}}(\mathbf{r}) = \frac{1}{\Omega_0}\sum_{\mathbf{G}} A_{\mathbf{G}}^{\alpha\mathbf{k}\lambda} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}, \tag{6}$$

where α is the spin index, $\mathbf{k}$ is the point in the Brillouin zone, $\mathbf{G}$ labels the plane waves, and Ω_0 is the volume of the unit cell. Index λ stands for the band states, which in this work are the eigen states of an effective Hartree–Fock Hamiltonian constructed with the quasi-particle Greens function [18]. Inside the muffin-tin sphere (MT) at atom $\mathbf{t}$, it is convenient to represent the band states as linear combinations of the orbitals belonging to that MT sphere

$$\Psi_\lambda^{\alpha\mathbf{k}}(\mathbf{r})|_{\mathbf{t}} = \sum_L Z_{\mathbf{t}L}^{\alpha\mathbf{k}\lambda}\varphi_L^{\alpha\mathbf{t}}(\mathbf{r}), \tag{7}$$

where index L combines angular momentum indexes l, m and any additional indexes to distinguish the orbitals inside the sphere (for example, it distinguishes between the solutions of radial equations and their energy derivatives). Representations (6) and (7) will be used throughout the paper.

4. Product basis conventions

We define the product basis functions $M_K^{\mathbf{q}}(\mathbf{r})$ as the plane waves in the interstitial region and as the optimized basis functions inside the MT spheres. We use a joint index K to label the product basis functions in all of the MT spheres and in the interstitial region. When K runs over the functions inside the MT sphere $\mathbf{t}$, the product basis functions can be written as the following

$$M_K^{\mathbf{q}}(\mathbf{r}+\mathbf{R}) = \begin{cases} 0 & \mathbf{r}\notin\mathbf{t} \\ e^{i\mathbf{q}\mathbf{R}}M_K^{\mathbf{t}}(\mathbf{r}) & \mathbf{r}\in\mathbf{t}. \end{cases} \tag{8}$$

When K runs over the functions in the interstitial region, we associate the index with the plane waves $\mathbf{G}_K$:

$$M_K^{\mathbf{q}}(\mathbf{r}+\mathbf{R}) = \begin{cases} e^{i\mathbf{q}\mathbf{R}}e^{i(\mathbf{q}+\mathbf{G}_K)\mathbf{r}} & \mathbf{r}\in Int \\ 0 & \mathbf{r}\notin Int. \end{cases} \tag{9}$$

The word "optimized" means that we build the space of all products of the orbitals $\varphi_L^{\alpha\mathbf{t}}$ (solutions of the radial Kohn–Sham equations, as well their energy derivatives) in each MT sphere and construct linearly independent ortho-normal basis in this space, following the Ref. [8].

The product basis defined above is not ortho-normal in the interstitial region. We also define the dual product basis

$$\widetilde{M}_K^{\mathbf{q}}(\mathbf{r}) = \begin{cases} \sum_{K'\in Int}\langle M_{K'}^{\mathbf{q}}|M_K^{\mathbf{q}}\rangle^{-1}M_{K'}^{\mathbf{q}} & \mathbf{r}\in Int \\ M_{\mathbf{t}K}^{\mathbf{q}}(\mathbf{r}) & \mathbf{r}\in\mathbf{t}, \end{cases} \tag{10}$$

which is ortho-normal to the basis (8): $\langle\widetilde{M}_{K'}^{\mathbf{q}}|M_K^{\mathbf{q}}\rangle = \delta_{KK'}$.

We expand the polarizability in dual basis

$$P(\mathbf{r};\mathbf{r}';\tau) = \frac{1}{N_{\mathbf{k}}}\sum_{\mathbf{q}}\sum_{KK'}\widetilde{M}_K^{\mathbf{q}}(\mathbf{r})P_{KK'}^{\mathbf{q}}(\tau)\widetilde{M}_{K'}^{*\mathbf{q}}(\mathbf{r}'), \tag{11}$$

with $N_{\mathbf{k}}$ being the number of points in the Brillouin zone.

To express the interaction (in the product basis), we calculate the integral

$$W_{KK'}^{\mathbf{q}}(\tau) = \int\int d\mathbf{r}d\mathbf{r}'\widetilde{M}_K^{*\mathbf{q}}(\mathbf{r})W(\mathbf{r};\mathbf{r}';\tau)\widetilde{M}_{K'}^{\mathbf{q}}(\mathbf{r}'). \tag{12}$$

Because of the orthogonality of the original and dual product basis sets it is convenient to think of the interaction as an expansion in the original product basis set:

$$W(\mathbf{r};\mathbf{r}';\tau) = \frac{1}{N_{\mathbf{k}}}\sum_{\mathbf{q}}\sum_{KK'}M_K^{\mathbf{q}}(\mathbf{r})W_{KK'}^{\mathbf{q}}(\tau)M_{K'}^{*\mathbf{q}}(\mathbf{r}'). \tag{13}$$

5. Quasiparticle approximation

Our method is exclusively based on imaginary axis data, contrasting with previously proposed methods (see for example Kotani et al. [15]) .

We proceed as follows. In the Dyson equation for the Green function

$$G^{-1}_{\lambda\lambda'}(\mathbf{k};\omega) = (i\omega + \mu - \varepsilon^{\mathbf{k}}_{\lambda})\delta_{\lambda\lambda'} - \Sigma^{c}_{\lambda\lambda'}(\mathbf{k};\omega), \tag{14}$$

with the band indices (λ, λ') corresponding to the effective exchange Hamiltonian [18], we approximate the frequency dependence of the self-energy by a linear function

$$\Sigma^{c}_{\lambda\lambda'}(\mathbf{k};\omega) = \Sigma^{c}_{\lambda\lambda'}(\mathbf{k};\omega=0) + \frac{\partial\Sigma^{c}_{\lambda\lambda'}(\mathbf{k};\omega)}{\partial(i\omega)}|_{\omega=0}(i\omega). \tag{15}$$

With this approximation, the Dyson equation can be simplified

$$G^{-1}_{\lambda\lambda'}(\mathbf{k};\omega) = Z^{-1}_{\lambda\lambda'}(\mathbf{k})(i\omega) + (\mu - \varepsilon^{\mathbf{k}}_{\lambda})\delta_{\lambda\lambda'} - \Sigma^{c}_{\lambda\lambda'}(\mathbf{k};0), \tag{16}$$

where we have introduced a renormalization factor Z matrix (not to be confused with the expansion coefficients in Eq. (7)):

$$Z^{-1}_{\lambda\lambda'}(\mathbf{k}) = \delta_{\lambda\lambda'} - \frac{\partial\Sigma^{c}_{\lambda\lambda'}(\mathbf{k};\omega)}{\partial(i\omega)}|_{\omega=0}. \tag{17}$$

Representing Z-factor as a symmetrical product

$$Z^{-1}_{\lambda\lambda'}(\mathbf{k}) = \sum_{\lambda''} Z^{-1/2}_{\lambda\lambda''}(\mathbf{k})Z^{-1/2}_{\lambda''\lambda'}(\mathbf{k}), \tag{18}$$

we reduce the Dyson equation to the following form

$$\sum_{\lambda''\lambda'''} Z^{1/2}_{\lambda\lambda''}(\mathbf{k})G^{-1}_{\lambda''\lambda'''}(\mathbf{k};\omega)Z^{1/2}_{\lambda'''\lambda'}(\mathbf{k}) = i\omega\delta_{\lambda\lambda'} + \sum_{\lambda''\lambda'''} Z^{1/2}_{\lambda\lambda''}(\mathbf{k})[(\mu - \varepsilon^{\mathbf{k}}_{\lambda''})\delta_{\lambda''\lambda'''} - \Sigma^{c}_{\lambda''\lambda'''}(\mathbf{k};0)]Z^{1/2}_{\lambda'''\lambda'}(\mathbf{k}). \tag{19}$$

The second term on the right hand side of this equation is a Hermitian matrix. It is diagonalized in subroutine BANDS_QP

$$\begin{aligned}\mu\delta_{\lambda\lambda'} - H^{\mathbf{k}}_{\lambda\lambda'} &= \sum_{\lambda''\lambda'''} Z^{1/2}_{\lambda\lambda''}(\mathbf{k})[(\mu - \varepsilon^{\mathbf{k}}_{\lambda''})\delta_{\lambda''\lambda'''} - \Sigma^{c}_{\lambda''\lambda'''}(\mathbf{k};0)]Z^{1/2}_{\lambda'''\lambda'}(\mathbf{k}) \\ &= \sum_{i} Q^{\mathbf{k}}_{\lambda i}E^{\mathbf{k}}_{i}Q^{\dagger\mathbf{k}}_{i\lambda'},\end{aligned} \tag{20}$$

where $E^{\mathbf{k}}_{i}$ are the effective eigenvalues. After the diagonalization, we can rewrite (19) as follows

$$\sum_{\lambda''\lambda'''} Z^{1/2}_{\lambda\lambda''}(\mathbf{k})G^{-1}_{\lambda''\lambda'''}(\mathbf{k};\omega)Z^{1/2}_{\lambda'''\lambda'}(\mathbf{k}) = \sum_{i} Q^{\mathbf{k}}_{\lambda i}\left[i\omega + \mu - E^{\mathbf{k}}_{i}\right]Q^{\dagger\mathbf{k}}_{i\lambda'}, \tag{21}$$

or, for the Green function

$$G^{\mathbf{k}}_{\lambda\lambda'}(\omega) = \sum_{i} \frac{(Z^{1/2}Q)^{\mathbf{k}}_{\lambda i}(Q^{\dagger}Z^{1/2})^{\mathbf{k}}_{i\lambda'}}{i\omega + \mu - E^{\mathbf{k}}_{i}}. \tag{22}$$

This expression differs from the full GW Greens function by a linear approximation for the frequency dependent self-energy.

At this point, we construct the quasiparticle Greens function (step (5) in Section 2) by setting $Z^{\mathbf{k}}_{\lambda\lambda'} = \delta_{\lambda\lambda'}$ in the above equation to obtain

$$G^{\mathbf{k}}_{\lambda\lambda'}(\omega) = \sum_{i} \frac{Q^{\mathbf{k}}_{\lambda i}Q^{\dagger\mathbf{k}}_{i\lambda'}}{i\omega + \mu - E^{\mathbf{k}}_{i}}. \tag{23}$$

6. Polarizability calculation and scaling

In accordance with the MT geometry, there are essentially three different contributions to the polarizability corresponding to (i) when both space arguments of P belong to MT spheres (Mt-Mt); (ii) one of them belongs to a MT sphere and another belongs to the interstitial region (Mt-Int); (iii) both arguments belong to the interstitial region (Int-Int). Below we consider three cases separately, describing how the Green function is transformed from the band representation to the real space. We also show the details of the polarizability calculations and the way it is transformed from the real to the reciprocal space.

Our parallelization strategy here is to use the two-dimensional grid of the MPI-processes. The first MPI-dimension in the polarizability calculations is associated with the variable τ in such a way that each process performs the calculation only on its own set of the τ indexes. This is very efficient because all formulas in this section are totally independent for different τ's. The second dimension of the MPI grid is used whenever it is appropriate as described briefly below. Namely, every process associated with the second dimension of the MPI grid carries out the calculations on its own set of $\mathbf{k}$ points, or on its own set of the triplets $\mathbf{Rtt}'$ (with $\mathbf{t}$ and $\mathbf{t}'$ being the coordinates of the atoms, belonging to the unit cells separated by a translation vector $\mathbf{R}$).

In the following sections we will present the scalings associated with the principle steps of the algorithm. For convenience, we summarize main notations here: N_{at} is the number of atoms in the unit cell; N_{orb} is the number of the basis functions (per atom) in the LAPW + LO representation (we provide typical numbers below). The number of bands is equal to $N_{at}N_{orb}$, so we will not use a separate variable for the number of bands. Furthermore, the number of plane waves in the interstitial region used to represent the fermionic functions approximately equals the number of bands, so we do not introduce a new variable for it as well. Bosonic functions have a major impact on the calculation time, so it is practical to take into account their numbers more carefully. N^{Mt}_{pb} is the number of the product basis orbitals inside MT sphere (per atom); N^{Int}_{pb} is the number of plane waves associated with product basis in the interstitial region (per atom); N_r is the number of points in the regular real space mesh in the unit cell (per atom); N_k is the number of points in the whole Brillouin zone; N_τ is the number of points in τ-mesh. The number of points in the fermionic and bosonic frequency meshes is about the same as the number of τ-points, so we use the latter in all cases.

6.1. Mt-Mt part of the polarizability

When both space arguments belong to MT-spheres, real space representation means that we represent G as an expansion in local orbitals inside the spheres at $\mathbf{t}$ and $\mathbf{t}'$ in the unit cells separated by a translation vector $\mathbf{R}$

$$G^{\alpha}(\mathbf{r};\mathbf{r}';\tau)|_{\mathbf{t}+\mathbf{R};\mathbf{t}'} = \sum_{LL'} \varphi^{\alpha\mathbf{t}}_{L}(\mathbf{r})G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{t}'L'}(\tau)\varphi^{\alpha\mathbf{t}'}_{L'}(\mathbf{r}'), \tag{24}$$

where the coefficients $G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{t}'L'}(\tau)$ are found using (7):

$$G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{t}'L'}(\tau) = \frac{1}{N_{\mathbf{k}}}\sum_{\mathbf{k}} e^{i\mathbf{kR}}\sum_{\lambda\lambda'} Z^{\alpha\mathbf{k}\lambda}_{\mathbf{t}L}G^{\alpha\mathbf{k}}_{\lambda\lambda'}(\tau)Z^{*\alpha\mathbf{k}\lambda'}_{\mathbf{t}'L'}. \tag{25}$$

The first step in Mt-Mt case is to transform Green's function from the band representation to the representation (24) using (25). The scaling associated with the evaluation of (25) is $\left[(N_{at}N_{orb})^3N_k + (N_{at}N_{orb})^2N_k\ln N_k\right]N_\tau$. The first term corresponds to the sum over (λ, λ') indexes which scales as $(N_{at}N_{orb})^3$ for every $\mathbf{k}$-point and τ. The second term is related to the fast Fourier transform (FFT) from $\mathbf{k}$-space to the $\mathbf{R}$-space which scales as $N_k\ln N_k$ for

each matrix element and τ. We use the second dimension of the MPI-grid to calculate matrix products (sums over band indexes) distributing different **k**'s over the MPI processes. Then we switch the MPI parallelization to perform FFT for different indexes ($\mathbf{t}'L'$). In the above code, the Green function transformation is performed in G_RS_FROM_KS_MM subroutine.

The expression for the polarizability then follows from (1) and (24)

$$P^{\mathbf{R}}_{\mathbf{t}K;\mathbf{t}'K'}(\tau) = -\sum_{\alpha}\sum_{LL''}\langle M^{\mathbf{t}}_K|\varphi^{\alpha\mathbf{t}}_L\varphi^{\alpha\mathbf{t}}_{L''}\rangle \times \sum_{L'} G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{t}'L'}(\tau)\sum_{L'''} G^{\alpha;\mathbf{R}}_{\mathbf{t}L'';\mathbf{t}'L'''}(\beta-\tau) \times \langle\varphi^{\alpha\mathbf{t}'}_{L'}\varphi^{\alpha\mathbf{t}'}_{L'''}|M^{\mathbf{t}'}_{K'}\rangle. \quad (26)$$

The scaling of (26) is $\left[2N^3_{orb}N^{Mt}_{pb} + (N_{orb}N^{Mt}_{pb})^2\right]N^2_{at}N_kN_\tau$. To evaluate (26) we use second MPI dimension to parallelize the triplets ($\mathbf{R}$; $\mathbf{t}$; $\mathbf{t}'$). In the code, Eq. (26) is implemented in P_MM_R subroutine.

The transformation to the reciprocal space consists of a single FFT operation

$$P^{\mathbf{q}}_{\mathbf{t}K;\mathbf{t}'K'}(\tau) = \sum_{\mathbf{R}} e^{-i\mathbf{qR}}P^{\mathbf{R}}_{\mathbf{t}K;\mathbf{t}'K'}(\tau), \quad (27)$$

which is implemented in the subroutine P_MM_Q_FROM_R. The scaling of (27) is $(N^{Mt}_{pb}N_{at})^2N_k\ln N_kN_\tau$.

As one can see, some steps in the polarizability calculation scale as N^3_{at} whereas others scale as N^2_{at}. However, for the systems with a manageable number of atoms (less than 20), the steps which scale as N^2_{at} (Eq. (26)) have a larger prefactor. Namely, comparing (25) with (26) one can estimate that the number of atoms should be equal to $N^{Mt}_{PB}(2+N^{Mt}_{PB}/N_{orb})$ in order to make (25) as time consuming as (26) is. Realistically, N^{Mt}_{PB} is of order of a few hundred, so the turning point for the Mt-Mt polarizability (when it starts to scale as N^3_{at}) would be around $N_{at} \approx 1000$.

For large systems, N_k is small, so that $lnN_k \ll N_{orb}$. Thus, the steps which have N_klnN_k in their prefactors are not important from the point of view of the time consumption.

One point about the scaling should be clarified here. The above relations assume that the number of points in the Brillouin zone (N_k) does not change when N_{at} changes. This assumption is justified for the very large unit cells when one may use just one k-point. However, for the systems of moderate size (up to 10 atoms/unit cell), the N_k is roughly proportional to $1/N_{at}$ (the volume of the Brillouin zone is inversely proportional to the volume of the unit cell). This fact should be kept in mind when one compares the timings (see for example the Table 3).

6.2. Mt-Int part of the polarizability

In this case, the second space argument in Eq. (28) runs over the regular **r**-mesh in the whole unit cell, whereas for the first space argument, we use an expansion in local orbitals:

$$G^{\alpha}(\mathbf{r};\mathbf{r}';\tau)|^{\mathbf{r}\in\mathbf{t}+\mathbf{R}}_{\mathbf{r}'\in Int} = \sum_{L}\varphi^{\alpha\mathbf{t}}_L(\mathbf{r})G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{r}'}(\tau). \quad (28)$$

The corresponding coefficients $G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{r}'}(\tau)$ are obtained in two steps (subroutine G_RS_FROM_KS_MI):

$$G^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{G}'}(\tau) = \frac{1}{\sqrt{\Omega_0}}\sum_{\lambda\lambda'} Z^{\alpha\mathbf{k}\lambda}_{\mathbf{t}L}G^{\alpha\mathbf{k}}_{\lambda\lambda'}(\tau)A^{*\alpha\mathbf{k}\lambda'}_{\mathbf{G}'}, \quad (29)$$

and

$$G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{r}'}(\tau) = \frac{1}{N_{\mathbf{k}}}\sum_{\mathbf{k}} e^{i\mathbf{kR}}\sum_{\mathbf{G}'} e^{-i(\mathbf{k}+\mathbf{G}')\mathbf{r}'}G^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{G}'}(\tau). \quad (30)$$

The scaling of (29) is $N^3_{at}N^3_{orb}N_kN_\tau$. The scaling associated with the evaluation of (30) is $N^2_{at}N_{orb}N_rN_kN_\tau\,[\ln(N_{at}N_r)+\ln N_k]$. The second dimension of the MPI grid is used in (29) and (30) to perform calculations for different **k**'s independently.

The expression for the polarizability follows from (28) and (1)

$$P^{\mathbf{R}}_{\mathbf{t}K;\mathbf{r}'}(\tau) = -\sum_{\alpha}\sum_{LL'}\langle M^{\mathbf{t}}_K|\varphi^{\alpha\mathbf{t}}_L\varphi^{\mathbf{t}}_{L'}\rangle G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{r}'}(\tau)G^{\alpha\mathbf{R}}_{\mathbf{t}L';\mathbf{r}'}(\beta-\tau). \quad (31)$$

The scaling of (31) is $(N_{at}N_{orb})^2N^{Mt}_{pb}N_rN_kN_\tau$. The second dimension of the MPI grid of the processes is used in (31) to perform the calculations for different **R**'s independently (subroutine P_IM_R).

The reciprocal space representation in the product basis is obtained after two FFTs:

$$\widetilde{P}^{\mathbf{q}}_{\mathbf{t}K;\mathbf{G}'}(\tau) = \frac{1}{N_{\mathbf{r}}}\sum_{\mathbf{r}'} e^{i(\mathbf{q}+\mathbf{G}')\mathbf{r}'}\sum_{\mathbf{R}} e^{-i\mathbf{qR}}P^{\mathbf{R}}_{\mathbf{t}K;\mathbf{r}'}(\tau). \quad (32)$$

The scaling of (32) is $N^2_{at}N_kN_\tau N^{Mt}_{pb}N_r\,[\ln N_k+\ln(N_{at}N_r)]$. Representation in the dual basis is obtained using an additional step

$$P^{\mathbf{q}}_{\mathbf{t}K;K'}(\tau) = \sum_{\mathbf{G}'}\widetilde{P}^{\mathbf{q}}_{\mathbf{t}K;\mathbf{G}'}(\tau)\langle e^{i(\mathbf{q}+\mathbf{G}')\mathbf{r}'}|M^{\mathbf{q}}_{K'}\rangle_{Int}, \quad (33)$$

where $\langle e^{i(\mathbf{q}+\mathbf{G}')\mathbf{r}'}|M^{\mathbf{q}}_{K'}\rangle_{Int}$ represents the integral of the product of two plane waves over the interstitial region and it is done analytically.

The scaling of (33) is $N^3_{at}N_kN_\tau N^{Mt}_{pb}N^{2Int}_{pb}$. The second dimension of the MPI grid is used in (32) and (33) to perform independent calculations for different **q**'s (subroutine P_IM_Q_FROM_R).

Similar to the Mt-Mt case, the most time consuming step (for practically achievable system sizes) scales as N^2_{at} (Eq. (31)). Again, the reason is that it has a prefactor which scales roughly at least as N^4_{orb} (both N_r and N^{Mt}_{PB} are a few times larger than N_{orb}). The steps which scale as N^3_{at} or $N^2_{at}lnN_{at}$ (Eqs. (29), (30), (32) and (33)) have lesser prefactors.

6.3. Int-Int part of the polarizability

In this case, both space arguments run over the regular mesh in the unit cell. The real space representation for G is obtained in two steps (subroutines G_K_G_R1 and G_RR_R_R1_STAR):

$$G^{\alpha\mathbf{k}}_{\mathbf{G};\mathbf{G}'}(\tau) = \frac{1}{\Omega_0}\sum_{\lambda\lambda'} A^{\alpha\mathbf{k}\lambda}_{\mathbf{G}}G^{\alpha\mathbf{k}}_{\lambda\lambda'}(\tau)A^{*\alpha\mathbf{k}\lambda'}_{\mathbf{G}'}, \quad (34)$$

and

$$G^{\alpha\mathbf{R}}_{\mathbf{r};\mathbf{r}'}(\tau) = \frac{1}{N_{\mathbf{k}}}\sum_{\mathbf{k}} e^{i\mathbf{kR}}\sum_{\mathbf{G};\mathbf{G}'} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}G^{\alpha\mathbf{k}}_{\mathbf{G};\mathbf{G}'}(\tau)e^{-i(\mathbf{k}+\mathbf{G}')\mathbf{r}'}. \quad (35)$$

The scalings of (34) and (35) are $(N_{at}N_{orb})^3N_kN_\tau$ and $(N_{at}N_r)^2N_kN_\tau\,[2\ln(N_{at}N_r)+\ln N_k]$ respectively. The second dimension of the MPI grid is used in (34) and (35) to perform calculations for different **k**'s independently.

The formula for the polarizability is very simple in this case

$$P^{\mathbf{R}}_{\mathbf{r}\mathbf{r}'}(\tau) = -\sum_{\alpha} G^{\alpha\mathbf{R}}_{\mathbf{r}\mathbf{r}'}(\tau)G^{\alpha\mathbf{R}}_{\mathbf{r}\mathbf{r}'}(\beta-\tau). \quad (36)$$

The scaling of (36) is $(N_{at}N_r)^2N_kN_\tau$. In (36) we use the MPI processes associated with index $\mathbf{r}'$ and τ.

The reciprocal space representation in the original product basis is obtained using three FFTs:

$$\widetilde{P}^{\mathbf{q}}_{\mathbf{G}\mathbf{G}'}(\tau) = \frac{1}{N_{\mathbf{r}}}\sum_{\mathbf{r}} e^{i(\mathbf{q}+\mathbf{G})\mathbf{r}}\frac{1}{N_{\mathbf{r}}}\sum_{\mathbf{r}'} e^{-i(\mathbf{q}+\mathbf{G}')\mathbf{r}'} \times \sum_{\mathbf{R}} e^{-i\mathbf{qR}}P^{\mathbf{R}}_{\mathbf{r};\mathbf{r}'}(\tau). \quad (37)$$

The scaling of Eq. (37) is $N_{at}^2 N_r N_k N_\tau \left[(N_r + N_{pb}^{Int}) \ln(N_{at} N_r) + N_r \ln N_k\right]$. Representation in the dual basis follows as additional matrix multiplications

$$P^{\mathbf{q}}_{K;K'}(\tau) = \sum_{\mathbf{GG'}} \langle e^{i(\mathbf{q}+\mathbf{G})\mathbf{r}} | M^{\mathbf{q}}_K \rangle^*_{Int} \widetilde{P}^{\mathbf{q}}_{\mathbf{GG'}}(\tau) \langle e^{i(\mathbf{q}+\mathbf{G'})\mathbf{r'}} | M^{\mathbf{q}}_{K'} \rangle_{Int}. \tag{38}$$

The scaling of (38) is $(N_{at} N_{pb}^{Int})^3 N_k N_\tau$. The second dimension of the MPI grid is used in (37) and (38) to perform the calculations for different $\mathbf{q}$'s independently. Formulas (36)–(38) are implemented in the subroutine `P_II_SOLID`.

Looking at the scaling relations, one can conclude that in the Int-Int case the most time consuming step (Eq. (38)) scales as N_{at}^3. However, the corresponding prefactor is still not big enough ($N_{PB}^{Int} \approx 5 - 10 N_{orb}$), and, the overall scaling of the polarizability evaluation is N_{at}^2 for the systems of practically achievable size.

7. Screened interaction

Eq. (2), written in the reciprocal space, reads as the following

$$W^{\mathbf{q}}_{KK'}(\nu) = V^{\mathbf{q}}_{KK'} + \sum_{K''K'''} V^{\mathbf{q}}_{KK''} P^{\mathbf{q}}_{K''K'''}(\nu) W^{\mathbf{q}}_{K'''K'}(\nu). \tag{39}$$

The scaling of (39) is $(N_{at}[N_{pb}^{Mt} + N_{pb}^{Int}])^3 N_k N_\tau$. We associate the two-dimensional grid of the MPI-processes with variables $\mathbf{q}$ and ν. Formula (39) is implemented in the subroutine `WS_K_NU_SOLID_O`. As one can see, this step scales as N_{at}^3. The prefactor is big enough to make the time needed to accomplish this operation comparable to the time needed for the polarizability evaluation. For example, comparing (39) with (26) which scales as N_{at}^2, one can conclude that the turning point (when (39) takes roughly the same time as (26) does) is $N_{at} = \frac{N_{PB}^{Mt}(2N_{orb}^3 + N_{PB}^{Mt} N_{orb}^2)}{N_{PB}^{Mt} + N_{PB}^{Int}}$, which is approximately equal to 7 atoms/unit cell in practice (assuming $N_{orb} = 100$ and $N_{PB}^{Mt} = N_{PB}^{Int} = 500$).

8. Dynamic self-energy

Due to the division of the screened interaction into the bare Coulomb interaction V and the dynamic part $\widetilde{W}$ ($W = V + \widetilde{W}$) the self-energy is also divided into static and dynamic. Here we consider the evaluation of the dynamic part only. The static part is evaluated similarly with obvious simplifications in the formulas.

Similar to the polarizability, there are essentially three different contributions to the self-energy corresponding to (i) when both space arguments of belong to MT spheres (Mt-Mt); (ii) one of them belongs to a MT sphere and another belongs to the Interstitial (Mt-Int); (iii) both arguments belong to the interstitial region (Int-Int). Below we consider the three cases separately, describing how the screened interaction is transformed from the reciprocal space to the real space, how we calculate the dynamic self-energy, and how we transform it from the real space back to the reciprocal space and band representation.

Our parallelization strategy here is similar to the strategy we use in the polarizability calculations. Also, the details of the scaling relations are very similar to the scaling in the polarizability evaluation, so we do not discuss their details here.

8.1. Mt-Mt part of the self-energy

When both space arguments belong to MT spheres, the real space representation means that we represent $\widetilde{W}$ as an expansion in the product basis functions inside the spheres at $\mathbf{t}$ and $\mathbf{t'}$ in the unit cells separated by the translation vector $\mathbf{R}$ (compare with the formula (13))

$$\widetilde{W}^{\mathbf{R}}_{\mathbf{t}K;\mathbf{t'}K'}(\tau) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{q}} e^{i\mathbf{qR}} \widetilde{W}^{\mathbf{q}}_{\mathbf{t}K;\mathbf{t'}K'}(\tau). \tag{40}$$

We use the second dimension of the MPI grid to calculate the matrix products (sums over the band indexes) spreading different $\mathbf{q}$'s over the MPI processes. Formula (40) is implemented in the subroutine `W_MM`. The scaling of (40) is $(N_{at} N_{pb}^{Mt})^2 N_\tau N_k \ln N_k$.

The expression for the self-energy follows from (24) and (3) (subroutine `SIGC_MM_R`)

$$\begin{aligned}\Sigma^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{t'}L'}(\tau) = &- \sum_{L''L'''} \sum_{KK'} \langle \varphi^{\alpha\mathbf{t}}_L | \varphi^{\alpha\mathbf{t}}_{L''} M^{\mathbf{t}}_K \rangle \\ &\times G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{t'}L'}(\tau) \widetilde{W}^{\mathbf{R}}_{\mathbf{t}K;\mathbf{t'}K'}(\beta - \tau) \\ &\times \langle \varphi^{\alpha\mathbf{t'}}_{L'} | \varphi^{\alpha\mathbf{t'}}_{L'''} M^{\mathbf{t'}}_{K'} \rangle.\end{aligned} \tag{41}$$

The scaling of (41) is $\left[2N_{orb}^3 N_{pb}^{Mt} + (N_{orb} N_{pb}^{Mt})^2\right] N_{at}^2 N_k N_\tau$. To evaluate (41) we use the second MPI dimension to parallelize the triplets ($\mathbf{Rtt'}$).

The transform to the reciprocal space consists in one FFT transform (subroutine `SIGC_MM_K_FROM_R`).

$$\Sigma^{\mathbf{k}}_{\mathbf{t}L;\mathbf{t'}L'}(\tau) = \sum_{\mathbf{R}} e^{-i\mathbf{kR}} \Sigma^{\mathbf{R}}_{\mathbf{t}L;\mathbf{t'}L'}(\tau), \tag{42}$$

which scales as $(N_{at} N_{orb})^2 N_\tau N_k \ln N_k$.

8.2. Mt-Int part of the self-energy

In this case the transform of W to the real space involves two FFT's (subroutine `V_IM_R_FROM_K`):

$$\widetilde{W}^{\mathbf{R}}_{\mathbf{t}K;\mathbf{r}}(\tau) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{q}} e^{i\mathbf{qR}} \sum_{\mathbf{G}} e^{-i(\mathbf{q}+\mathbf{G})\mathbf{r}} \widetilde{W}^{\mathbf{q}}_{\mathbf{t}K;\mathbf{G}}(\tau). \tag{43}$$

The scaling of (43) is $N_{at}^2 \ln(N_{at} N_r) N_{pb}^{Mt} N_r N_k N_\tau$. The second dimension of the MPI grid is used in (43) to perform the calculations for different $\mathbf{q}$'s independently.

The expression for the self-energy follows from (28) and (3)

$$\Sigma^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{r'}}(\tau) = - \sum_{L'K} \langle \varphi^{\alpha\mathbf{t}}_L | \varphi^{\mathbf{t}}_{L'} M^{\mathbf{t}}_K \rangle G^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{r'}}(\tau) \widetilde{W}^{\mathbf{R}}_{\mathbf{t}K;\mathbf{r'}}(\beta - \tau), \tag{44}$$

which scales as $N_{at}^2 N_{orb}^2 N_{pb}^{Mt} N_r N_k N_\tau$.

The second dimension of the MPI grid is used in (44) to perform calculations for different $\mathbf{R}$'s independently. Formula (44) is implemented in the subroutine `SIGC_IM_R`.

The transform to the band states representation is achieved in a few steps. They are implemented in the subroutine `SIGC_IM_K_FROM_R`.

First we apply FFT

$$\Sigma^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{r'}}(\tau) = \sum_{\mathbf{R}} e^{-i\mathbf{kR}} \Sigma^{\alpha\mathbf{R}}_{\mathbf{t}L;\mathbf{r'}}(\tau), \tag{45}$$

with scaling $N_{at}^2 N_{orb} N_r N_\tau N_k \ln N_k$.

At this point the function is represented by its values at the homogeneous $\mathbf{r'}$-mesh in the whole unit cell. In order to perform integration over the interstitial region we again apply FFT to transform it into an equivalent linear combination of plane waves

$$\Sigma^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{r'}}(\tau) = \sum_{\mathbf{G'}} \tilde{\Sigma}^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{G'}} e^{-i(\mathbf{k}+\mathbf{G'})\mathbf{r'}}, \tag{46}$$

with the coefficients

$$\tilde{\Sigma}^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{G'}}(\tau) = \frac{1}{N_{\mathbf{r}}} \sum_{\mathbf{r'}} e^{i(\mathbf{k}+\mathbf{G'})\mathbf{r'}} \Sigma^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{r'}}(\tau). \tag{47}$$

The scaling of (47) is $N_{at}^2 N_{orb} N_r N_k N_\tau \ln(N_{at} N_r)$. The form (46) allows us to integrate over the interstitial region analytically, and in doing so we obtain

$$\Sigma^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{G'}}(\tau) = \frac{1}{\sqrt{\Omega_0}} \sum_{\mathbf{G''}} \tilde{\Sigma}^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{G''}}(\tau) S^{\mathbf{k}}_{\mathbf{G''G'}}, \tag{48}$$

which scales as $N_{at}^3 N_{orb}^2 N_{pb}^{Int} N_k N_\tau$. $S^{\mathbf{k}}_{\mathbf{G}''\mathbf{G}'}$ in (48) is the integral of the product of two plane waves ($e^{-i(\mathbf{k}+\mathbf{G}'')\mathbf{r}}$ and $e^{i(\mathbf{k}+\mathbf{G}''')\mathbf{r}}$) taken over the interstitial region.

Finally, the contribution to the band state representation follows

$$\Sigma^{\alpha\mathbf{k}}_{\lambda\lambda'}(\tau)|^{Mt}_{Int} = \sum_{\mathbf{t}L}\sum_{\mathbf{G}'} Z^{*\alpha\mathbf{k}\lambda}_{\mathbf{t}L} \Sigma^{\alpha\mathbf{k}}_{\mathbf{t}L;\mathbf{G}'}(\tau) A^{\alpha\mathbf{k}\lambda'}_{\mathbf{G}'} + H.C. \tag{49}$$

The scaling of (49) is $(N_a N_{orb})^3 N_k N_\tau$. The second dimension of the MPI grid is used in (45)–(49) to perform the calculations for different **k**'s independently.

8.3. Int-Int part of the self-energy

In this case, both space arguments run over the regular mesh in the unit cell. A real space representation for W is obtained with three FFT's (subroutines `W_Q_G_R1` and `W_RR_R_R1_STAR`):

$$\widetilde{W}^{\mathbf{R}}_{\mathbf{r};\mathbf{r}'}(\tau) = \frac{1}{N_{\mathbf{k}}}\sum_{\mathbf{q}} e^{i\mathbf{q}\mathbf{R}} \sum_{\mathbf{G}} e^{-i(\mathbf{q}+\mathbf{G})\mathbf{r}} \sum_{\mathbf{G}'} e^{i(\mathbf{q}+\mathbf{G}')\mathbf{r}'} \times \widetilde{W}^{\mathbf{q}}_{\mathbf{G}\mathbf{G}'}(\tau), \tag{50}$$

with scaling $N_{at}^2 N_r N_k N_\tau \times \left[N_{pb}^{Int} \ln(N_{at} N_r) + N_r \ln(N_{at} N_r) + N_r \ln N_k\right]$.

The second dimension of the MPI grid is used in (50) to perform the calculations for different **k**'s independently.

The formula for the self-energy is very simple in this case

$$\Sigma^{\alpha\mathbf{R}}_{\mathbf{r}\mathbf{r}'} = -G^{\alpha\mathbf{R}}_{\mathbf{r}\mathbf{r}'}(\tau)\widetilde{W}^{\mathbf{R}}_{\mathbf{r}\mathbf{r}'}(\beta-\tau). \tag{51}$$

The scaling of (51) is $(N_{at} N_r)^2 N_\tau N_k$. In (51), the second dimension of the MPI grid is associated with the index $\mathbf{r}'$.

Then we apply FFT

$$\Sigma^{\alpha\mathbf{k}}_{\mathbf{r}\mathbf{r}'}(\tau) = \sum_{\mathbf{R}} e^{-i\mathbf{k}\mathbf{R}} \Sigma^{\alpha\mathbf{R}}_{\mathbf{r}\mathbf{r}'}(\tau). \tag{52}$$

The scaling of (52) is $(N_{at} N_r)^2 N_\tau N_k \ln N_k$. Similar to the Mt-Int case, we use the FFT to transform it into an equivalent linear combination of the plane waves

$$\Sigma^{\alpha\mathbf{k}}_{\mathbf{r}\mathbf{r}'}(\tau) = \sum_{\mathbf{G}}\sum_{\mathbf{G}'} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} \tilde{\Sigma}^{\alpha\mathbf{k}}_{\mathbf{G};\mathbf{G}'}(\tau) e^{-i(\mathbf{k}+\mathbf{G}')\mathbf{r}'}, \tag{53}$$

with the coefficients

$$\tilde{\Sigma}^{\alpha\mathbf{k}}_{\mathbf{G};\mathbf{G}'}(\tau) = \frac{1}{N_{\mathbf{r}}^2}\sum_{\mathbf{r}\mathbf{r}'} e^{-i(\mathbf{k}+\mathbf{G})\mathbf{r}} e^{i(\mathbf{k}+\mathbf{G}')\mathbf{r}'} \Sigma^{\alpha\mathbf{k}}_{\mathbf{r}\mathbf{r}'}(\tau). \tag{54}$$

The scaling of (54) is $N_{at}^2 N_r \ln(N_{at} N_r) N_k N_\tau \left[N_r + N_{pb}^{Int}\right]$. The form (53) allows us to integrate over the interstitial region analytically and, as a result, we obtain

$$\Sigma^{\alpha\mathbf{k}}_{\mathbf{G};\mathbf{G}'}(\tau) = \frac{1}{\Omega_0}\sum_{\mathbf{G}''\mathbf{G}'''} S^{\mathbf{k}}_{\mathbf{G}\mathbf{G}''} \tilde{\Sigma}^{\alpha\mathbf{k}}_{\mathbf{G}'';\mathbf{G}'''}(\tau) S^{\mathbf{k}}_{\mathbf{G}'''\mathbf{G}'}, \tag{55}$$

with scaling $N_{at}^3 N_{pb}^{Int} N_{orb} N_k N_\tau \left[N_{pb}^{Int} + N_{orb}\right]$.

Finally, the contribution to the band state representation from the interstitial region is given by

$$\Sigma^{\alpha\mathbf{k}}_{\lambda\lambda'}(\tau)|^{Int}_{Int} = \sum_{\mathbf{G}\mathbf{G}'} A^{*\alpha\mathbf{k}\lambda}_{\mathbf{G}} \Sigma^{\alpha\mathbf{k}}_{\mathbf{G};\mathbf{G}'}(\tau) A^{\alpha\mathbf{k}\lambda'}_{\mathbf{G}'}. \tag{56}$$

The scaling of (56) is $(N_{at} N_{orb})^3 N_k N_\tau$. The second dimension of the MPI grid is used in (52)–(56) to perform the calculations for different **k**'s independently. Formulas (51)–(56) have been implemented in the subroutine `SIGC_II_SOLID`.

Table 1
Band gaps (eV) of selected semiconductors and insulators in comparison with the experimental data. We also include the error (%) relative to the experiment. For the materials with known effects of the electron–phonon and/or the spin–orbit interaction (Si, SiC, C, GaAs, MgO, ZnS, ZnSe, LiF, BN, AlP), the corresponding effects were excluded from the raw experimental data to facilitate the comparison with the calculations. Experimental data have been cited from Refs. [17,24,15,25–27].

	[15,24]	[16]	[17]	Present work	Exp.
Si	1.23	1.41	1.47	1.41(14.8%)	1.22
SiC	2.14	2.88	2.90	3.08(22.7%)	2.51
C	6.52	6.18	6.40	6.71(14.1%)	5.88
GaAs	1.93	1.85	1.75	2.08(23.1%)	1.69
ZnO	3.87	3.8	4.61	4.47(24.2%)	3.60
MgO		9.16	9.29	9.42(18.0%)	7.98
ZnS	4.04	4.15		4.19(6.3%)	3.94
ZnSe	3.08			3.17(5.7%)	3.00
LiF		15.9		16.63(14.8%)	14.48
NaCl				9.81(15.4%)	8.5
BN		7.14	7.51	7.06(7.0%)	6.6
AlP		2.90	3.10	2.80(13.4%)	2.47
NiO	4.8		4.97	4.47(3.9%)	4.3
Cu_2O	2.36		2.65	2.42(10.0%)	2.20
TiO_2	3.78		4.22	3.80(22.6%)	3.1
$Sr\,TiO_3$	4.19			4.01(21.5%)	3.3
CeO_2	∼5			5.83(70.1%)	3–3.5

Table 2
Band widths (eV) of alkali metals. Experimental data have been taken from Refs. [28] and [29]. For the present results we also include the error (%) relative to the experiment.

	[24]	Present work	Exp.
Na	3.0	3.16(19.2%)	2.65
K		2.07(29.4%)	1.60

9. Results

In this section we demonstrate performance of our version of the linearized QSGW program and compare our results and results obtained with other non-linearized implementations of the QSGW method with experimental data. Let us also notice that other implementations are done exploiting different basis set. In Table 1 we provide results for calculated band gaps of selected semiconductors and insulators obtained with our code and compare them with previously published by other groups results and experimental data. As one can see, our results and the non-linearized SQGW results are very close and all systematically overestimate the band gaps. The overestimation is generally in the 10–24% range for all studied materials, excluding antiferromagnetically ordered NiO (error is only 3.9%) and the f-electron compound CeO_2 where the error is large (70%). Let us mention that other QSGW methods also produce large error for this material.

Table 2 presents the band widths of the alkali metals Na and K. Formally, alkali metals belong to the s-materials, but as one can see from the table, the error in the calculated band width (20–30%) is a bit larger than the error in the calculated band gaps for sp-semiconductors/insulators. Let us point out that the error increases when the density of valence electrons is reduced (when going from Sodium to Potassium). This fact was expected because the electron gas of lower density corresponds to a more correlated situation.

As an example of the code scaling with the number of MPI processes, we provide Fig. 1 where we plot computation time for TiO2 versus number of MPI processes used. This material has 6 atoms in the unit cell and the corresponding calculations are rather expensive in terms of computer time. As one can estimate from the graph, the overall efficiency of the parallelization (time reduction/increase in the number of the processes) when the number of MPI processes is increased from 64 to 384 is 0.7 (=4.2/6). Taking into account the complexity of the code (one has to switch many times from one parallelization strategy to another during

Table 3
Main parameters of the calculations and timings. Time is measured in seconds during one iteration. N_{atom} is the number of atoms in the unit cell. N_{PB} is the size of product basis. N_k is the number of **k**-points in the Brillouin zone. In all cases 24 processes were used along the τ-dimension of the MPI grid and 4 processes were used along the k-dimension of the MPI grid.

	N_{atom}	N_{PB}	$N_{LAPW+LO}$	N_k	P	W	Σ	G
Si	2	1316	288	8^3	260	76	494	16
SiC	2	1237	202	8^3	317	109	555	12
C	2	1202	186	8^3	193	64	337	2
GaAs	2	1178	212	6^3	148	38	258	7
NiO	4	1906	388	$4 \times 4 \times 3$	255	124	399	34
Cu_2O	6	4232	926	4^3	337	895	560	228
TiO_2	6	3268	668	$2 \times 2 \times 4$	249	329	299	84
$Sr\,TiO_3$	5	3163	562	4^3	263	416	422	87
CeO_2	3	1633	292	6^3	172	137	331	20
K	1	741	78	16^3	393	154	586	4
Ni	1	724	126	12^3	368	81	467	10

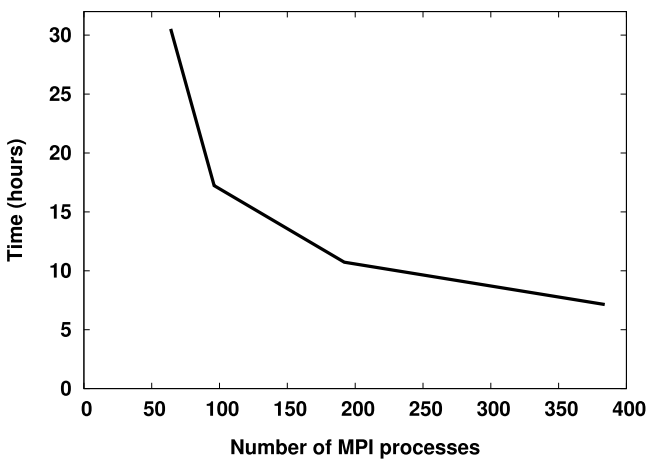

Fig. 1. Full calculation time as a function of the number of MPI processes used (TiO_2). The calculation consisted of 36 LDA iterations followed by 30 LQSGW iterations.

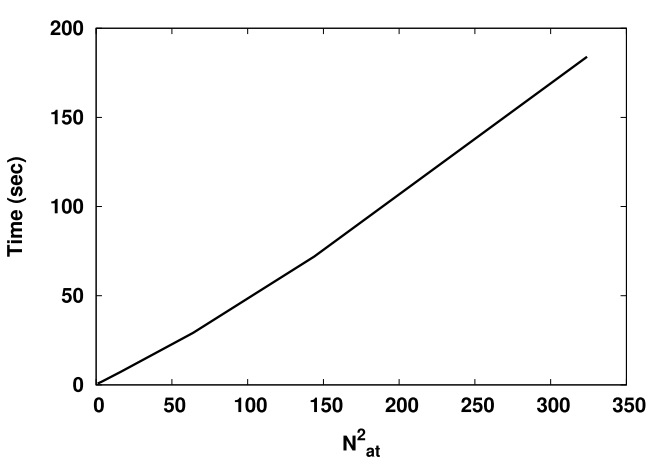

Fig. 2. Calculation time for supercells of Na atoms. The time for one LQSGW iteration is presented as a function of square of the number of atoms in the supercell.

the calculation) and the fact that we did not use the fastest interprocess communication (infiniband) allow us to think about the obtained efficiency as a satisfactory one. One of the possible ways to improve the efficiency (besides of using infiniband) is to use the SCALAPACK routines to assist with the diagonalization/inversion of large matrices, particularly when evaluating the screened interaction W (Eq. (39)).

In Fig. 2 we show how the calculation time scales with the number of atoms in unit cell. To make the plot we used artificial supercells of Na atoms. In this figure, the results obtained with N_{at} equal to 1, 2, 4, 8, 12, and 18 are presented as a function of N_{at}^2. As one can see, in materials containing only a few atoms per unit cell the calculation time scales almost quadratically (we see almost linear dependence in the plot) with only slight admixture of N_{at}^3 dependence which begins to get noticeable when $N_{at} > 8$. The Fig. 2 supports our estimations made above in Sections 6–8.

Table 3 shows the detailed timings in the evaluation of the main quantities (P, W, Σ, and G) during one iteration. As one can see for the materials studied, the calculation of the self-energy is the most time consuming step. However, the increase in the size of the product basis (which is proportional to the number of atoms in the unit cell) will eventually make the evaluation of W the most time consuming step. This is clear from the scaling considerations, while many parts of the algorithm scale as N_{at}^3, the evaluation of W has the biggest prefactor.

Conclusions

We presented an implementation of the LQSGW method in the LAPW basis set which scales for large systems as N_{at}^3 with the number of atoms. The scaling is N_{at}^2 for small number of atoms. Further improvements of the algorithm for large systems would require a removal of the computational bottleneck which is the matrix inversion in Eq. (39). In its current form, this code can serve as a starting point for further diagrammatic many body studies on the Matsubara axis in an all electron basis. Recently, self-consistent vertex-corrected GW calculations were performed for a number of semiconductors/insulators and simple metals in Refs. [20,21]. Also, the non-self-consistent implementation of the GW+DMFT method was presented in the Ref. [22]. Our code was also used in Ref. [30] for the calculation of spin susceptibilities in 3d metals.

The computer code realizing our implementation of the LQSGW approach can be obtained from https://github.com/andreykutepov65/LqsgwFlapw repository.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Science and Basic Energy Sciences as a part of the Computational Materials Science Program. We thank Sangkook Choi for many fruitful discussions.

References

[1] L. Hedin, Phys. Rev. 139 (1965) A796.
[2] M.S. Hybertsen, S.G. Louie, Phys. Rev. B 34 (1986) 5390.
[3] R.W. Godby, M. Schlüter, L.J. Sham, Phys. Rev. B 37 (1988) 10159.
[4] J. Deslippe, G. Samsonidze, D.A. Strubbe, M. Jain, M.L. Cohen, S.G. Louie, Comput. Phys. Comm. 183 (2012) 1269.

[5] X. Gonze, G.-M. Rignanese, M. Verstraete, J.-M. Beuken, Y. Pouillon, R. Caracas, F. Jollet, M. Torrent, G. Zerah, M. Mikami, Ph. Ghosez, M. Veithen, J.-Y. Raty, V. Olevano, F. Bruneval, L. Reining, R.W. Godby, G. Onida, D.R. Hamann, D.C. Allan, Z. Kristallogr. 220 (2005) 558.
[6] M. Govoni, G. Galli, J. Chem. Theor. Comp. 11 (2015) 2680.
[7] A. Gulans, S. Kontur, C. Meisenbichler, D. Nabok, P. Pavone, S. Rigamonti, S. Sagmeister, U. Werner, C. Draxl, J. Phys.: Condens. Matter 26 (2014) 363202.
[8] C. Friedrich, S. Blügel, A. Schindlmayr, Phys. Rev. B 81 (2010) 125102.
[9] H. Jiang, P. Blaha, Phys. Rev. B 93 (2016) 115203.
[10] I.-H. Chu, J.P. Trinastic, Y.-P. Wang, A.G. Eguiluz, A. Kozhevnikov, T.C. Schulthess, H.P. Cheng, Phys. Rev. B 93 (2016) 125210.
[11] M. Govoni, G. Galli, Large scale GW calculations, arXiv.cond.mat.:1501.03141.
[12] M.M. Rieger, L. Steinbeck, I.D. White, H.N. Rojas, R.W. Godby, Comput. Phys. Comm. 117 (1999) 211.
[13] W. Ku, A.G. Eguiluz, Phys. Rev. Lett. 89 (2002) 126401.
[14] F. Bruneval, M. Gatti, Top Curr. Chem. 347 (2014) 99.
[15] T. Kotani, M. van Schilfgaarde, S.V. Faleev, Phys. Rev. B 76 (2007) 165106.
[16] M. Shishkin, M. Marsman, G. Kresse, Phys. Rev. Lett. 99 (2007) 246403.
[17] W. Chen, A. Pasquarello, Phys. Rev. B 92 (2015) 041115.
[18] A. Kutepov, K. Haule, S.Y. Savrasov, G. Kotliar, Phys. Rev. B 85 (2012) 155129.
[19] A. Kutepov, S.Y. Savrasov, G. Kotliar, Phys. Rev. B 80 (2009) 041103.
[20] A.L. Kutepov, Phys. Rev. B 94 (2016) 155101.
[21] A.L. Kutepov, Self-consistent solution of Hedin's equations: semiconductors/insulators, arXiv.cond.mat.:1610.08395.
[22] S. Choi, A. Kutepov, K. Haule, M. van Schilfgaarde, G. Kotliar, NPJ Quantum Mater. 1 (2016) 16001.
[23] O.K. Andersen, Phys. Rev. B 12 (1975) 3060.
[24] M. van Schilfgaarde, T. Kotani, S. Faleev, Phys. Rev. Lett. 96 (2006) 226402.
[25] F.J. Himpsel, J.A. Knapp, J.A. VanVechten, D.E. Eastman, Phys. Rev. B 20 (1979) 624.
[26] P. Lautenschlager, M. Garriga, S. Logothetidis, M. Cardona, Phys. Rev. B 35 (1987) 9174.
[27] A. Mang, K. Reimann, St. Rübenacke, M. Steube, Phys. Rev. B 53 (1996) 16283.
[28] I.-W. Lyo, E.W. Plummer, Phys. Rev. Lett. 60 (1988) 1558.
[29] E.L. Shirley, Phys. Rev. B 54 (1996) 7758.
[30] A.L. Wysocki, A. Kutepov, V.P. Antropov, Phys. Rev. B 94 (2016) 140405(R).