STERNHEIMERGW: a program for calculating GW quasiparticle band structures and spectral functions without unoccupied states

Martin Schlipf*, Henry Lambert1, Nourdine Zibouche, Feliciano Giustino*

Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom

Abstract

The STERNHEIMERGW software uses time-dependent density-functional perturbation theory to evaluate GW quasiparticle band structures and spectral functions for solids. Both the Green’s function $G$ and the screened Coulomb interaction $W$ are obtained by solving linear Sternheimer equations, thus overcoming the need for a summation over unoccupied states. The code targets the calculation of accurate spectral properties by convoluting $G$ and $W$ using a full frequency integration. The linear response approach allows users to evaluate the spectral function at arbitrary electron wavevectors, which is particularly useful for indirect band gap semiconductors and for simulations of angle-resolved photoelectron spectra. The software is parallelized efficiently, integrates with version 6.3 of Quantum Espresso, and is continuously monitored for stability using a test farm.

Keywords: first-principles calculations, many-body perturbation theory, solid state physics, linear response

Program Summary

Program Title: STERNHEIMERGW
Licensing provisions: GNU General Public License v3.0
Programming language: Fortran 2003
Nature of problem: The lack of the exchange-correlation discontinuity in density-functional theory (DFT) leads to a systematic underestimation of the band gap between conduction and valence states. Many-body perturbation theory in the GW approximation provides an effective solution to this problem, as well as other limitations faced by DFT in the description of electronic excitations. However, the GW method comes with its own set of limitations: (i) The perturbation of the system is typically expressed in terms of the unoccupied states, and achieving numerical convergence with respect to these states is often cumbersome. Since the underlying DFT codes rely only on the occupied states, their default behavior is often ill-suited to provide a sufficient amount of empty states. Combined with the large number of parameters that need to be converged, this limits the use of GW codes by non-expert users and automatic scripts. (ii) Currently, GW codes require that a homogeneous $k$-point mesh is used in the calculation. Hence, features close to the band edges are often only accessible via prohibitively expensive dense $k$-point meshes or interpolation techniques. (iii) To evaluate the frequency convolution of the Green’s function $G$ and the screened Coulomb interaction $W$, many current GW calculations rely on approximations such as the plasmon-pole approximation, the analytic continuation, or the contour deformation. The relative merits and accuracy of the various approximations are not fully understood.

Solution method: In STERNHEIMERGW, we address (i) by replacing the summation over unoccupied states with a linear response equation. The solution depends on the occupied states only, and it employs linear response algorithms already provided by the Quantum ESPRESSO suite to compute phonons and related properties. As an additional benefit, transforming the problem in a linear response equation removes the restriction to homogeneous $k$-point meshes (ii), so that the GW self-energy for any arbitrary point can be evaluated. Finally (iii), we provide the possibility to perform a full frequency convolution along the real frequency axis. This feature can serve as a benchmark for approximate integrations using models or analytic continuation.
1. Introduction

Kohn-Sham density function theory (DFT) is a powerful and extremely popular formalism for studying the total energy of an interacting electron system in its ground state. When used in the study of electronic excitations, such as the calculation of electron band structures and wavefunctions, DFT exhibits some well known deficiencies, for example the lack of the exchange-correlation discontinuity in the exchange-correlation potential, which leads to an underestimation of the band gap in insulators, and the excessive delocalization of d and f electrons. Applying many-body perturbation theory corrections in the GW approximation allows one to correct some of these deficiencies using the non-local and frequency-dependent electron self-energy \( \Sigma(\omega) \). Early numerical implementations of this method demonstrated an improved band gap for diamond and similar improvement were subsequently shown for other typical semiconductors. Since then, GW has been used to accurately describe the band gaps for a variety of systems ranging from solids to interfaces and molecules. It yields accurate band offsets, defect energies, and improved effective masses. Recent developments focused on the self-consistency of the GW method for molecules and solids, as well as total energy calculations.

More recent developments of the GW method relate to photoelectron spectroscopy. Progress in high-energy resolution angle-resolved photoelectron spectroscopy (ARPES) had led to a renewed interest in the spectral function of materials, in particular the signatures of electron correlations and bosonic excitations. These features were originally examined for the homogeneous electron gas, but have recently been investigated both experimentally and theoretically for charge carriers in semiconductors coupling to plasmons. While the GW approximation does produce satellite features in the spectral function, their strength is overestimated and their energy is blue-shifted. This mismatch is intrinsic to the GW approximation and cannot be overcome by self-consistency. Nevertheless, the GW spectral function can provide a starting point for the cumulant expansion, in which the satellite and quasiparticle spectral function are convoluted and the position and weight of the satellite features are corrected.

At present, there are two major limitations to obtain accurate spectral functions with the GW approximation. First, the frequency dependence of the screened Coulomb interaction is commonly approximated by a Dirac delta function in the plasmon-pole model. To increase the accuracy, more recent approaches employ an analytic continuation or contour deformation. A detailed assessment of the accuracy of these techniques is still lacking and may prove crucial for re-
lubricant calculations of quasiparticle lifetimes. Secondly, most GW implementations rely on a summation over the empty Kohn-Sham states. This requires a very accurate description of the unoccupied subspace and therefore imposes additional constraints on the construction of pseudopotentials. Furthermore, converging the eigenvalues occasionally demands the inclusion of a considerable number of empty states, and the convergence with respect to these states is interdependent with other convergence parameters. To address these issues, several techniques have been proposed which improve the rate of convergence of the summations. However, some of these methods introduce additional convergence parameters. Alternatively, one can obtain the dielectric matrix in density-functional perturbation theory by solving linear Sternheimer equations. This approach has been subsequently optimized by iterative diagonalization to determine the most significant eigenvalue or employing an optimal representation of the polarization. One can extend this approach, so that both the Green’s function and the screened Coulomb interaction can be obtained by solving linear response equations. This approach is referred to as the Sternheimer GW method and has been successfully applied to large systems with a single \( k \) point.

2. Description of the code

2.1. Kohn-Sham reference system

To evaluate the many-body perturbation correction, we start from an unperturbed reference system within Kohn-Sham density functional theory. The important ingredients of this reference system for our calculation are the exchange-correlation potential \( V^x \) and the Kohn-Sham Hamiltonian \( \hat{H} \) along with its eigenvalues \( \epsilon_{n,k} \) and wavefunctions \( \phi_{n,k}(r) \). Here \( n \) is a band index, \( r \) is the position vector, and \( k \) a Bloch vector in the Brillouin zone of the crystal. We employ a planewaves basis set, and express the wavefunction \( \phi_{n,k}(r) \) in terms of the expansion coefficients \( u_{n,k,G} \):

\[
\phi_{n,k}(r) = \frac{1}{\sqrt{\Omega}} \sum_{G} u_{n,k,G} \exp[i(k + G) \cdot r],
\]

where \( G \) denotes reciprocal lattice vectors and \( \Omega \) is the volume of the unit cell. In a planewaves basis only the kinetic energy part of the Hamiltonian and the non-local part of the pseudopotentials are represented in reciprocal space; the local Kohn-Sham potential is applied in real space via a fast Fourier transform.

2.2. Green’s function

From the Kohn-Sham Hamiltonian, we obtain the electronic Green’s function, which is one of the ingredients to evaluate the GW self-energy. The Green’s function is given by the linear equation:

\[
\mathcal{V}_{k,G} \mathbf{L}_k^G(\omega) \mathbf{G}_G^G(\omega) = -\delta_{GG}. \tag{2}
\]

Here, \( \omega \) is a complex frequency, chosen to satisfy the time ordering, \( \mathbf{L}_k^G \) is the linear operator for the Green’s function:

\[
\mathcal{L}_k^G(\omega) = \hat{H}_k - \hbar\omega, \tag{3}
\]

and \( \delta_{GG} \) is the Kronecker \( \delta \). Most GW codes expand Eq. (2) in the basis of the eigenstates of the Hamiltonian and solve it by summing over many unoccupied states. In contrast, in STERNHEIMERGW we utilize an iterative Krylov subspace solver to explicitly find the solution to the linear equation (2). We employ the fact that shifted linear problems described by the linear operator \( \mathcal{L}_k^G \) differing only in the frequency \( \omega \) span the same Krylov subspace. This property allows us to solve the linear problem for all frequencies at the computational complexity of a single frequency.
2.3. **Dielectric matrix**

The dielectric matrix of a system can be evaluated from the electron density response to a planewave perturbation:

\[ \epsilon_{qG}^{GG}(\omega) = \delta_{GG'} - \Delta n_{qG}^{GG'}(\omega), \]  

(4)

where \( \omega \) is the frequency. The calculation of the density response is performed by starting from the linear variation of the wavefunctions in real space, and then carrying out a Fourier transform into reciprocal space:

\[ \Delta n_{qG}^{GG'}(\omega) = 2 \frac{1}{N_k} \sum_{nk,z} u_{nk,z}^* \Delta u_{nk+q,GG'}(\pm \omega), \]  

(5)

where the factor 2 accounts for the spin degeneracy and we are assuming a uniform grid of \( N_k \) \( k \)-vectors in the Brillouin zone. The change of the wavefunction coefficients is obtained from the Sternheimer equation, which can be written as:

\[ V_{nkG} L^c_{k+q} \left( \frac{\delta_{nk+q}}{\hbar} + \omega \right) \Delta u_{nk+q,GG'}(\omega) = \delta_{nkGG'} . \]  

(6)

Since the linear operator of the Coulomb response,

\[ L^c_{k}(\omega) = \hat{H}_k - \hbar \omega + \delta^{occ}, \]  

(7)

is very similar to the one for the Green’s function \( \frac{1}{\hbar} \), the same Krylov subspace solvers can be employed. The only difference with respect to \( \frac{1}{\hbar} \) is that now we have an additional projector on the occupied states \( \delta^{occ} \), this term improves the condition number of the linear operator. The right-hand side of (7) is given by a Fourier transform of:

\[ \delta_{nkGG'} = (1 - \delta^{occ}) \epsilon^{G'} \hat{u}_{nk,G'}, \]  

(8)

and describes a driving field corresponding to a periodic perturbation.

2.4. **Correlation self-energy**

For the correlation self-energy, we combine the inverse of the dielectric matrix with the truncated Coulomb potential (see Sec. 2.6):

\[ \tilde{W}^{GG'}_{q,G}(\omega) = v_{q+G} \left( \epsilon_{qG}^{-1}(\omega) - \delta_{GG'} \right), \]  

(9)

where \( \tilde{W} \) is the correlation part of the screened Coulomb interaction. To overcome numerical issues associated with the inversion of the dielectric matrix for \( q \approx 0 \), we evaluate that element for a small but finite \( q \) and set the off-diagonal elements to zero if \( q + G = 0 \) or \( q + G' \neq 0 \). Subsequently, we perform a Fourier transform of \( \tilde{W} \) and \( \tilde{W}^{GG'}_{q,G} \) to real space, because the self-energy can be expressed as a product there:

\[ \Sigma_{kG}^{\tilde{W}}(\omega) = \sum_{q} \frac{w_q w_q'}{2\pi} G_{k-qG}(\omega') \tilde{W}_{q,G}(\omega' - \omega). \]  

(10)

\( w_q \) and \( w_q' \) are weights for \( q \) and \( \omega \) integration respectively. The factor 10 depends on whether the frequency integration is performed along the real axis (\( \alpha = i \)) or along the imaginary axis (\( \alpha = -1 \)). We transform the resulting self-energy back to reciprocal space to evaluate the expectation values with the Kohn-Sham wavefunctions.

2.5. **Exchange self-energy**

To evaluate the exchange self-energy, we evaluate the Fock potential by summing over all occupied wavefunctions:

\[ \Sigma^{\tilde{F}}_{kGG'} = - \sum_{nqG} \frac{w_{nk-qG}^* u_{nk-GG'}^* \hat{v}_{nk-GG'} \hat{v}_{q,G'}}{\Omega}, \]  

(11)

where the weight \( w_{nk-qG} \) contains the weight of the \( q \) integration. The Coulomb potential is truncated as described in the following section.

2.6. **Truncation**

Owing to the long-range nature of the Coulomb potential, the Fourier transform of this quantity diverges for small arguments \( q \rightarrow 0 \), leading to instabilities in the convergence with respect to the grid in the Brillouin zone. Furthermore, when describing systems of reduced dimensionality, such as slabs or nanowires, it is advantageous to truncate the Coulomb potential along the non-periodic direction in order to avoid spurious interactions between periodic images. In order to address these issues, it is common practice to truncate the Coulomb potential at a certain distance. This approach yields a Fourier transform with a finite value in the limit \( q \rightarrow 0 \). In STERNHEIMERGW, we implement three different truncation schemes: First, one can truncate the Coulomb potential at a distance \( R_{cut} \) (spherical truncation) resulting in a potential:

\[ v_q = \frac{e^2}{4\pi\epsilon_0 \epsilon_q} \sqrt{1 - \cos(qR_{cut})}. \]  

(12)

Secondly, one can truncate the Coulomb interaction at a height \( z_{cut} \) (slab truncation), which yields the following potential:

\[ v_q = \frac{e^2}{4\pi\epsilon_0 \epsilon_q} \sqrt{1 - \exp(-q^2z_{cut}^2)} \cos(qz_{cut}) \]  

(13)
Lastly, one can truncate the Coulomb interaction in the Wigner-Seitz supercell. In this latter case, since an analytic expression for the potential is not known, we tabulate the results of the truncation and look up the values for the relevant vectors as necessary.

2.7. Analytic continuation

Quantities such as the dielectric matrix and the self-energy exhibit significant structure along the real axis, for example due to electron-hole excitations. In order to improve numerical stability and accuracy, it can be advantageous to evaluate quantities along the imaginary axis and perform an analytic continuation to the real axis. In STERNHEIMERGW, there are two levels where such an analytic continuation may be employed. On the one hand, one can evaluate the dielectric function on the imaginary axis, perform an analytic continuation to the real axis, and subsequently convolute it with the Green’s function according to Eq. (10). On the other hand, one can resolve the convolution along the imaginary axis, thereby obtaining a self-energy at imaginary frequency values. The self-energy at real frequencies is then obtained via analytic continuation. To perform the analytic continuation, one can use a Padé expansion or the adaptive Antoulas-Anderson (AAA) algorithm. In general, determining the Padé approximant using a continued fraction expansion suffers from numerical instabilities, since a very high precision (number of significant digits) is required. In contrast, the AAA algorithm relies on a singular-value decomposition, and it refines iteratively the data points used to construct the approximant, until the largest deviation falls below a specified threshold.

2.8. Spectral function

The spectral function is related to the imaginary part of the retarded Green’s function:

\[ A_{k,GG}^{\omega} = -\frac{1}{\pi} \text{Im}(G_{k,GG}^{\omega}), \]  

and represents a \(k\)-resolved many-body density of states. Since the Green’s function is the resolvent of the many-body Hamiltonian, this leads to:

\[ A_{k,GG}^{\omega} = -\frac{1}{\pi} \text{Im}\left(\hbar \omega - H_k - \Sigma_k^{xc}(\omega) + V_k^{xc} \right)_{GG}^{-1}. \]

Approximating the GW wavefunctions by their Kohn-Sham counterpart \(\phi_{nk}(r)\), the diagonal part of the spectral function in the Kohn-Sham basis can be expressed as:

\[ A_{nk}(\omega) = \frac{-\text{Im}\Sigma_{nk}^{xc}(\omega)}{\pi \left|\hbar \omega - \epsilon_{nk} - \Sigma_{nk}^{xc}(\omega) + V_{nk}^{xc} \right|}. \]

The peaks of the spectral function correspond to the quasiparticle eigenvalues. Linearizing the equation in the vicinity of the eigenvalue \(\epsilon_{nk}\) results in the following quasiparticle eigenvalue:

\[ \epsilon_{nk}^{QP} = \epsilon_{nk} + Z_{nk} \left(\text{Re} \Sigma_{nk}^{xc}(\epsilon_{nk}/\hbar) - V_{nk}^{xc} \right), \]

where \(Z_{nk} = 1 - \hbar^{-1} \text{Re}(\partial \Sigma_{nk}^{xc}/\partial \omega)\).
2.9. Software design

Architecture. In Fig. [2] we illustrate the multitier architecture of STERNHEIMERGW. The code builds on functionality provided by the Quantum Espresso software package and provides additional functionality handling low-level operations in the data tier. On top of this layer, we have an algo tier handling the interface with Quantum Espresso. The algo tier also provides integration grids, linear solvers, Coulomb truncation, and analytic continuation functionalities. In the phys tier, we evaluate the GW self-energy and matrix elements, employing the functionality of the lower lying tiers. The highest tier provides the user interface and the infrastructure for testing the software.

User input. The typical approach for specifying the user interface of a code requires that changes to input variables are manually translated into an update of the documentation. As this scheme poses the risk that the documentation becomes outdated, we employ an inverse scheme in STERNHEIMERGW. The core file controlling the user input is a human- and machine-readable YAML file documenting all input variables (see example in Fig. [3]). Starting from the documentation, we run a script to generate the Fortran file that processes the input file (Fig. [3]). As a consequence, introducing a new input variable requires an update of the documentation. This strategy allows us to make sure that the code and the documentation are always in sync. Furthermore, we process the YAML file to generate the online documentation of STERNHEIMERGW (Fig. [4]) on the website http://sternheimergw.org.

Continuous integration and testing. To ensure the long-term stability of the code and its alignment with the Quantum Espresso suite, we employ two techniques. First, we examine whether the code compiles and reproduces a set of reference results using a buildbot test farm. These integration tests involve 12 different calculations to check most of the functionality of STERNHEIMERGW. Secondly, we perform unit testing of some individual packages of the code, by benchmarking.

Table 1: Combined code coverage of integration and unit tests, i.e., the fraction of F90 lines and subroutines of the code that are executed at least once by the test set. The higher tiers are covered exclusively by the integration tests, for the lower tiers unit tests play a large role.

|                     | main | phys | algo | data |
|---------------------|------|------|------|------|
| **F90 lines**       |      |      |      |      |
| integration tests   | 100.0% | 92.1% | 70.3% | 31.4% |
| unit tests          | 0.0%  | 0.0%  | 41.3% | 74.3% |
| combined            | 100.0% | 92.1% | 86.7% | 92.3% |
| **subroutines**     |      |      |      |      |
| integration tests   | 100.0% | 97.3% | 64.6% | 30.0% |
| unit tests          | 0.0%  | 0.0%  | 53.3% | 80.0% |
| combined            | 100.0% | 97.3% | 87.7% | 93.3% |

Figure 3: Automatic generation of the user-input logic in SternheimerGW: a) The gw_input.yml file defines which input variables are used in STERNHEIMERGW along with their type, default value, and a description. b) A script processes the YAML file to generate a Fortran file containing a type with all input variables and the logic to read them. c) The YAML file is also processed to generate the description of the input variables on the website [http://sternheimergw.org].
the parallelization, the algebra, the analytic continuation, the linear solver, and the input/output (io) package independently from the rest of the code. For the unit testing, at variance with the tests of the complete code, we compare our implementation with analytic results. The unit tests are implemented in the pFUnit framework. Overall, all these tests cover at least 80% of the code in each tier (see Table 1).

2.10. Parallelization

GW calculations are computationally much more demanding than Kohn-Sham DFT calculations. To allow for a reasonable time to solution, we employ two different parallelization levels in STERNHEIMERGW. As we inherit the distribution logic from Quantum Espresso, we refer to these levels as pool and image parallelization. In Fig. 4 we compare the performance of these parallelization levels for a bulk system of MgO for different parts of the code. With the pool parallelization, we distribute different \( k \) or \( q \) points to different processors. This method is very efficient for both \( G \) and \( W \), if the number of these points is a multiple of the number of CPUs. The image parallelization distributes \( G \) vectors used in the linear response and Fourier transform to different CPU. For the calculation of \( W \), this parallelization is very efficient because the solution of the linear-response equation for each \( G \) vector is independent of the other \( G \) vectors. When evaluating \( \Sigma \), the speedup of this parallelization saturates when the number of grid points along the \( z \) axis reduces to one per CPU. The benefit of the image parallelization is that it reduces the memory footprint of the code.

3. Examples

3.1. Benchmark results

To assess the accuracy of STERNHEIMERGW, we verify the implementation by comparing to reference calculations from the literature. We use seven reference systems in the diamond (C, Si, and Ge), the zincblende (BN, CdS, and GaAs), or rocksalt structure (MgO). All lattice constants are taken from Ref. 114, except for that of Ge which is from Ref. 120.

The DFT ground state calculation was run with Quantum ESPRESSO v6.3 using a \( 12 \times 12 \times 12 \) \( k \)-point mesh and the PBE exchange-correlation potential. The pseudopotentials are verified according to the \( \Delta \)-test; as a compromise between speed and accuracy, we employ the ONCV pseudopotentials from the SG15 library for compounds not including \( d \) electrons, and the stringent pseudopotentials v0.4 from pseudodojo.org otherwise.

In Table 2 we present the comparison between the reference calculations and the results obtained with...
Table 3: Convergence parameters used in the calculation of the GW corrections reported in Table 2. The pseudopotentials (PP) are from the SG15 library or the stringent ones from pseudo-dojo.org. The energy cutoffs are employed for the DFT self-consistent field cycle (scf), the exchange (x) or the correlation (c) self energy; the integration grids for the self-energy and the density response are homogeneous grids with $N_q$ and $N_k$ points; the coarse frequency mesh for $W$ are obtained from a $(N_\omega - 1)$-node Gauss-Laguerre quadrature along the imaginary axis including the origin, and extending up to $\hbar \tilde{\omega}$; and the dense mesh for $W$ as well as the coarse one for $\Sigma$ are automatically constructed by the code. The row with the dagger contains the loose values used for the convergence test and the row with the asterisk is used for the convolution along the real frequency axis.

| material | PP     | $E_{\text{cut}}^{\text{scf}}$ (Ry) | $E_{\text{cut}}^{\text{x}}$ (Ry) | $E_{\text{cut}}^{\text{c}}$ (Ry) | $N_q^{1/3}$ | $N_k^{1/3}$ | $h\tilde{\omega}$ (eV) | $h\tilde{\omega}$ (eV) | threshold |
|----------|--------|----------------------------------|----------------------------------|----------------------------------|-------------|-------------|-----------------|-----------------|-----------|
| BN       | SG15   | 60 50 40 8 31 104 80 320 20 150 | 10^{-4} 10^{-5} |
| C        | SG15   | 70 50 24 8 31 104 50 280 30 180 | 10^{-4} 10^{-5} |
| CdS      | Dojo   | 60 55 25 8 31 104 150 150 30 180 | 10^{-4} 10^{-5} |
| GaAs     | Dojo   | 90 70 40 8 31 104 130 130 20 150 | 10^{-4} 10^{-5} |
| Ge       | Dojo   | 90 70 25 8 31 104 130 130 20 150 | 10^{-4} 10^{-5} |
| MgO      | SG15   | 70 65 30 6 31 104 50 280 30 180 | 10^{-4} 10^{-5} |
| Si       | SG15   | 20 19 15 8 31 104 50 280 20 150 | 10^{-4} 10^{-5} |
| BN$^\dagger$ | SG15 | 60 30 10 1 4 21 67 50 50 2 100 | 10^{-4} 10^{-5} |
| Si$^\star$ | SG15 | 16 15 12 8 201 768 800 100 351 35 | 10^{-10} 10^{-5} |

Figure 5: Convergence of the band gap of BN in the zincblende structure with respect to various convergence parameters in STERNHEIMER GW. The quantity $\Delta E_{\text{gap}}$ indicates the difference between the band gap with a given set of input parameters and the fully converged value $E_{\text{gap}} = 6.27$ eV. The blue line is obtained by varying one parameter at a time, while keeping the remaining parameters to loose values. The red line corresponds to calculations with the remaining parameters being set to stringent values. See section 3.2 for a detailed discussion of all parameters and Table 3 for the loose and stringent values used in the calculation.
SternheimerGW v0.15. The SternheimerGW calculations were converged using the algorithm described in Sec. 3.2 resulting in the parameters listed in Table 3. Overall, we find that our results agree well with the literature, within an error bar of 0.1 eV usually quoted in the literature for GW calculations.

3.2. Converging a GW calculation

Predictive GW calculations require the convergence of several numerical parameters. An efficient strategy is required to obtain these values at a reasonable computational cost. In order to tackle this multi-dimensional optimization, a possible strategy is to keep all convergence parameters at a loose setting, and only focus on the convergence behavior of a single parameter. To determine reasonable values for the loose setting it is a good strategy to focus on parts of the calculation that can be evaluated separately from the rest. First, one considers the exchange contribution to the GW self-energy, which is the most challenging quantity to converge in SternheimerGW, but at the same time its calculation is significantly less demanding than the correlation part. The analysis of the exchange self-energy yields estimates for the wavefunction cutoff \( E_{\text{scf}} \) (Fig. 5a), the exchange cutoff \( E_{\text{x}} \) (Fig. 5b), and the number \( N_q \) of points in the \( q \)-point grid used to evaluate the Brillouin-zone integrals (Fig. 5c). The wavefunction and the exchange cutoff are somewhat intertwined, and occasionally the former needs to be increased to allow for convergence of the latter. Secondly, we investigate the convergence of the head of the dielectric function (\( q = 0 \) point) with respect to the number of \( k \) points \( N_k \) (Fig. 5d) and the coarse frequency mesh \( N_{\text{coarse}} \) used for the analytic continuation (Fig. 5e).

Guided by the initial studies of the exchange self-energy and the dielectric matrix, we construct the loose setting for the convergence study. In general, we will choose the smallest possible values in the smooth region of convergence. For the correlation, we need to set three further numerical parameters: For the dense frequency mesh \( N_{\text{dense}} \) (Fig. 5f), we choose a mesh twice as dense as the coarse mesh; for the correlation cutoff \( E_{\text{c}} \) (Fig. 5g), we select typically a third of the value used for the exchange self-energy; and we use just two frequencies \( N_{\text{coarse}} = 2 \) (Fig. 5h) to perform the analytic continuation of the correlation self-energy from the imaginary to the real frequency axis.

With the constructed loose mesh, we then perform the convergence study individually increasing every single one of the parameters until the quasiparticle energies of interest are converged. The advantage of this strategy is that the computational cost of the complete conver-

![Figure 6: Spectral function of silicon along high-symmetry directions in the Brillouin zone. Results are obtained with full-frequency integration along the imaginary axis. For comparison the DFT bandstructure is shown as blue lines. Numerical parameters used for the calculations are listed in Table 3.](image-url)
Figure 7: Comparison between spectral function $A(\omega)$ of silicon at a) the Γ point, b) the X point, and c) the L point when $\Sigma^c$ is obtained by convolution along real (red) or imaginary (blue) frequency axis. The small shift in the valence band top is a consequence of the finite broadening $\eta = 0.15$ eV employed in the calculation of the Green’s function. Numerical parameters used for the calculations are listed in Table 3.

Figure 8: Impact of numerical technique on spectral function of silicon at Γ obtained by convolution along the real frequency axis. a) Increasing the smearing from $\eta = 0.08$ eV (blue) to 0.15 eV (red) or 0.3 eV (yellow) results in a decrease of the noise but also a shift of the valence band maximum to lower energies (see inset). b) Using the plasmon pole (PP) model (purple) instead of the AAA analytic continuation (blue) results in vanishing linewidths. We note that the vertical axes are on a logarithmic scale. Numerical parameters used for the calculations are listed in Table 3.

satellites are not accurately described in $GW$. An accurate description of these features requires the inclusion of higher order effects via the cumulant expansion.\cite{14,54,61–67}

3.4. Real frequency integration

As outlined in Sec. 2.4, $SternheimerGW$ provides the option to perform the frequency convolution along the real or the imaginary frequency axis. In Fig. 7, we demonstrate that these two implementations give nearly identical results for the spectral function at high-symmetry points of silicon. The main identifiable differences is a small shift of the valence band top to lower energies and an increased noise when using integration along the real axis. Both of these differences originate from the small parameter $\eta = 0.15$ eV that is used to shift the frequency slightly off the real axis. This shift is necessary in order to avoid the singularities on the real axis, so as to obtain numerically stable results. On the one hand, if $\eta$ is decreased to 0.08 eV the resulting spectral function becomes very noisy (see Fig. 8).
Figure 9: Inverse of the dielectric matrix of silicon at the Γ point obtained by solving the Sternheimer equation (6) along the real axis (blue) compared to a) plasmon pole model (purple), b) Padé expansion (yellow), and c) AAA algorithm (red). For the analytic continuation with either Padé or AAA, 201 frequencies along the imaginary axis were evaluated.

other hand, if \( \eta \) is increased to 0.3 eV, the energy of the quasiparticle becomes inaccurate (see inset of Fig. 8a).

In general, the integration along the imaginary axis with subsequent analytic continuation is one to two orders of magnitude cheaper because the Green’s function and the screened Coulomb interaction have less structure there. As the results of both methods are very similar, we chose the integration along the imaginary axis as the default in STERNHEIMER GW, and we provide the integration along the real axis as an optional feature for validation.

In Fig. 9b, we show that obtaining accurate quasiparticle lifetimes requires going beyond the plasmon-pole approximation, because in this approximation there is only one sharp pole at the plasmon frequency (cf. Fig 9a). In contrast, analytic continuation techniques can produce multiple poles and thereby capture the broadening of the quasiparticle spectra. In Fig. 9c, we show that the Padé expansion can reproduce the shape of the dielectric function more accurately than the plasmon-pole approximation, since it naturally incorporates multiple poles. However, this Padé algorithm is numerically not stable when using a large number of frequencies, therefore it cannot be used for the convolution along the real frequency axis. To address this issue, we employ the AAA analytic continuation technique, whereby only significant frequencies are selected via a singular value decomposition. Subsequently, poles with weak residues are removed so that only the most important contributions remain. In Fig. 9b, we show that the AAA method reduces the number of poles compared to the Padé expansion while still reproducing the overall shape and magnitude of the dielectric matrix.

4. Conclusion

We presented the STERNHEIMER GW code, which calculates many-body GW self-energies, quasiparticle band structures, and spectral functions by solving linear-response Sternheimer equations. The linear response scheme allows us to calculate the GW self-energy at arbitrary \( k \) points. This is particularly useful for computing band structures without resorting to interpolation techniques, and for comparison to ARPES experiments. STERNHEIMER GW goes beyond the plasmon-pole approximation by performing full-frequency integration along the real or the imaginary axis.

STERNHEIMER GW is aligned with the version 6.3 of Quantum Espresso. The dedicated website [http://www.sternheimer.org](http://www.sternheimer.org) provides users with installation guidelines, tutorials, and a comprehensive documentation of all input variables. The code is efficiently parallelized and has been tested extensively on high-performance computing architectures. Updated versions of the code are distributed in the GitHub repository [https://github.com/QEF/SternheimerGW](https://github.com/QEF/SternheimerGW).

Acknowledgments

The development of STERNHEIMER GW has received funding from the Leverhulme Trust (Grant RL-2012-001), the Graphene Flagship (Horizon 2020 Grant No. 785219 - GrapheneCore2), and the UK Engineering and Physical Sciences Research Council (Grant No. EP/M020517/1). Furthermore, the authors acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility [http://dx.doi.org/10.5281/zenodo.22558], the ARCHER UK National Supercomputing Service under
the T-Dops project, the Cambridge Service for Data Driven Discovery (CSD3) funded by EPSRC Tier-2 (Grant EP/P020259/1), the DECI resource Cartesius based in the Netherlands at SURFsara and Abel based in Oslo with support from the PRACE AISBL, and PRACE for awarding us access to MareNostrum at BSC-CNS, Spain.

References

[1] B. Gumhalter, V. Kovač, F. Caruso, H. Lambert, F. Giustino, On the combined use of GW approximation and cumulative expansion in the calculations of quasiparticle spectra: The paradigm of Si valence bands, Phys. Rev. B 94 (2016) 035103. doi:10.1103PhysRevB.94.035103

[2] C. Heil, M. Schlüpf, F. Giustino, Quasiparticle GW band structures and Fermi surfaces of bulk and monolayer NbSe₂, Phys. Rev. B 98 (2018) 075120. doi:10.1103PhysRevB.98.075120

[3] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864–B871. doi:10.1103PhysRev.136.B864

[4] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138. doi:10.1103PhysRev.140.A1133

[5] R. O. Jones, Density functional theory: Its origins, rise to prominence, and future, Rev. Mod. Phys. 87 (2015) 897–923. doi:10.1103RevModPhys.87.897

[6] J. P. Perdew, M. Levy, Physical content of the exact Kohn-Sham orbital energies: Band gaps and derivative discontinuities, Phys. Rev. Lett. 51 (1983) 1884–1887. doi:10.1103PhysRevLett.51.1884

[7] L. J. Sham, M. Schlüter, Density-functional theory of the energy gap, Phys. Rev. Lett. 51 (1983) 1888–1891. doi:10.1103PhysRevLett.51.1888

[8] L. Hedin, New method for calculating the one-particle Green’s function with application to the electron-gas problem, Phys. Rev. Lett. 39 (1977) 991–994. doi:10.1103PhysRevLett.39.991

[9] L. Hedin, S. O. Lundqvist, Effects of Electron-Electron and Electron-Phonon Interactions on the One-Electron States of Solids, Vol. 23 of Solid State Physics, Academic Press, 1969, pp. 1–181. doi:10.1006S0081-1947(08)60615-3

[10] G. Strinati, H. J. Mattausch, W. Hanke, Dynamical correlation effects on the quasiparticle Bloch states of a covalent crystal, Phys. Rev. Lett. 45 (1980) 290–294. doi:10.1103PhysRevLett.45.290

[11] G. Strinati, H. J. Mattausch, W. Hanke, Dynamical aspects of correlation corrections in a covalent crystal, Phys. Rev. B 25 (1982) 2867–2888. doi:10.1103PhysRevB.25.2867

[12] M. S. Hybertsen, S. G. Louie, Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies, Phys. Rev. B 34 (1986) 5390–5413. doi:10.1103Phys RevB.34.5390

[13] R. W. Godby, M. Schlüter, L. J. Sham, Accurate exchange-correlation potential for silicon and its discontinuity on addition of an electron, Phys. Rev. Lett. 56 (1986) 2415–2418. doi:10.1103PhysRevLett.56.2415

[14] F. Aryasetiawan, O. Gunnarsson, The GW method, Rep. Prog. Phys. 61 (1998) 237. doi:10.10880034488561/3.002

[15] W. G. Aulbur, M. Städele, A. Göring, Exact-exchange-based quasiparticle calculations, Phys. Rev. B 62 (2000) 7121–7132. doi:10.1103PhysRevB.62.7121

[16] G. Onida, L. Reining, A. Rubio, Electronic excitations: density-functional versus many-body Green’s-function approaches, Rev. Mod. Phys. 74 (2002) 601–659. doi:10.1103RevModPhys.74.601

[17] M. Giamoussis, M. Stankovski, R. Shaltatf, M. Grüning, F. Bruneval, P. Rinke, G.-M. Rignanese, Electronic properties of interfaces and defects from many-body perturbation theory: Recent developments and applications, Phys. Status Solidi B 248 (2011) 275–289. doi:10.1002ps sb.201046004

[18] R. Shaltatf, G.-M. Rignanese, X. Gonze, F. Giustino, A. Pasquarello, Band offsets at the Si/SiO₂ interface from many-body perturbation theory, Phys. Rev. Lett. 100 (2008) 186401. doi:10.1103PhysRevLett.100.186401

[19] P. Rinke, A. Janotti, M. Scheffler, C. G. Van de Walle, Defect formation energies without the band-gap problem: Combining density-functional theory and the GW approach for the silicon self-interstitial, Phys. Rev. Lett. 102 (2009) 026402. doi:10.1103PhysRevLett.102.026402

[20] M. R. Filip, C. Verdi, F. Giustino, GW band structures and carrier effective masses of CH₃NH₂PbI₃ and theoretical perovskites of the type APbI₃: A = NH₃, PH₃, A1H₃, and SbH₃, J. Phys. Chem. C 119 (2015) 25209–25219. doi:10.1021jacs.5b07898

[21] C. Rostgaard, K. W. Jacobsen, K. S. Thygesen, Fully self-consistent GW calculations for molecules, Phys. Rev. B 81 (2010) 085103. doi:10.1103PhysRevB.81.085103

[22] F. Caruso, P. Rinke, X. Ren, M. Scheffler, A. Rubio, Unified description of ground and excited states of finite systems: The self-consistent GW approach, Phys. Rev. B 86 (2012) 081102. doi:10.1103PhysRevB.86.081102

[23] F. Caruso, P. Rinke, X. Ren, A. Rubio, M. Scheffer, Self-consistent GW: All-electron implementation with localized basis functions, Phys. Rev. B 88 (2013) 075105. doi:10.1103PhysRevB.88.075105

[24] M. van Schilfgaarde, T. Kotani, S. Faleev, Quasiparticle self-consistent GW theory, Phys. Rev. Lett. 96 (2006) 226402. doi:10.1103PhysRevLett.96.226402

[25] M. Gatti, F. Bruneval, V. Olevano, L. Reining, Understanding correlations in vanadium dioxide from first principles, Phys. Rev. Lett. 99 (2007) 266402. doi:10.1103PhysRevLett.99.266402

[26] S. Ismail-Beigi, Correlation energy functional within the GW-RPA: Exact forms, approximate forms, and challenges, Phys. Rev. B 81 (2010) 195126. doi:10.1103PhysRevB.81.195126

[27] B. I. Lundqvist, Single-particle spectrum of the degenerate electron gas, Phys. konden. Mater. 6 (1967) 206–217. doi:10.1007BF02422717

[28] B. I. Lundqvist, Single-particle spectrum of the degenerate electron gas, konden. Mater. 7 (1968) 117–123. doi:10.1007BF02422899

[29] D. C. Langreth, Singularities in the X-ray spectra of metals, Phys. Rev. B 1 (1970) 471–477. doi:10.1103PhysRevB.1.471

[30] Y. Baer, G. Busch, X-ray photoemission from aluminium, Phys. Rev. Lett. 30 (1973) 280–282. doi:10.1103PhysRevLett.30.280

[31] A. Bostwick, F. Speck, T. Seyller, K. Horn, M. Polini, R. Asgari, A. H. MacDonald, E. Rotenberg, Observation of plasmarons in quasi-free-standing doped graphene, Science 328 (2010) 999–1002. doi:10.1126science.1186489

[32] M. Guzzo, G. L. Fanti, S. Toffoli, F. Romaniello, M. Gatti, J. J. Kas, J. J. Rehr, M. G. Silly, E. Scrocco, L. Reining, Valence electron photoemission spectrum of semiconductors: Ab initio description of multiple satellites, Phys. Rev. Lett. 107 (2011) 12
L. Hedin, On correlation effects in electron spectroscopies and the GW approximation, J. Phys.: Condens. Matter 11 (42) (1999) R489. doi:10.1088/0953-8984/11/42/201

B. Gunshater, Ultrafast dynamics and decoherence of quasiparticles in surface bands: Development of the formalism, Phys. Rev. B 72 (2005) 165406. doi:10.1103/PhysRevB.72.165406

B. Gunshater, Stages of hot electron dynamics in multie excitations processes at surfaces: General properties and benchmarks examples, Prog. Surf. Sci. 87 (2012) 163 – 188. doi:https://doi.org/10.1016/j.progsurf.2012.05.004

J. S. Zhou, J. J. Kas, L. Sponza, I. Reshetnyak, M. Guzzo, G. E. Engel, B. Farid, Generalized plasmon-pole model and dynamical effects in electron spectroscopy, J. Chem. Phys. 143 (2015) 184109. doi:10.1063/1.4934965

F. Caruso, C. Verdi, F. Giustino, Many-Body Calculations of Plasmon and Phonon Satellites in Angle-Resolved Photoelectron Spectra Using the Cumulant Expansion Approach, Springer International Publishing, Cham, 2018, pp. 1–25. doi:10.1007/978-3-319-42913-7_2

R. W. Godby, R. J. Needs, Metal-insulator transition in Kohn-Sham theory and quasiparticle theory, Phys. Rev. Lett. 62 (1989) 1169–1172. doi:10.1103/PhysRevLett.62.1169

G. E. Engel, B. Farid, Generalized plasmon-pole model and plasmon band structures of crystals, Phys. Rev. B 47 (1993) 15931–15934. doi:10.1103/PhysRevB.47.15931

R. Daling, W. van Haeringen, B. Farid, Plasmon dispersion in silicon obtained by analytic continuation of the random-phase-approximation dielectric matrix, Phys. Rev. B 44 (1991) 2952–2960. doi:10.1103/PhysRevB.44.2952

H. N. Rojas, R. W. Godby, R. J. Needs, Space-time method for ab initio calculations of self-energies and dielectric response functions of solids, Phys. Rev. Lett. 74 (1995) 1827–1830. doi:10.1103/PhysRevLett.74.1827

Y.-G. Jin, K. J. Chang, Dynamic response function and energy-loss spectrum for li using an N-point Padé approximant, Phys. Rev. B 59 (1999) 14841–14844. doi:10.1103/PhysRevB.59.14841

R. W. Godby, M. Schlüter, L. J. Sham, Self-energy operators and exchange-correlation potentials in semiconductors, Phys. Rev. B 37 (1988) 10159–10175. doi:10.1103/PhysRevB.37.10159

T. Kotani, M. van Schilfgaarde, All-electron GW approximation with the mixed basis expansion based on the full-potential LMTO method, Solid State Commun. 121 (9) (2002) 461 – 465. doi:https://doi.org/10.1016/S0038-1098(02)00028-5

S. Lebègue, B. Arnaud, M. Alouani, P. E. Bloechl, Implementation of an all-electron GW approximation based on the projector augmented wave method without plasmon pole approximation: Application to Si, S, C, AI, As, InAs, NaF, and KH, Phys. Rev. B 67 (2003) 155208. doi:10.1103/PhysRevB.67.155208

C. D. Spataru, M. A. Cazalilla, A. Rubio, L. X. Benedict, P. M. Echenique, S. G. Louie, Anomalous quasiparticle lifetime in graphite: Band structure effects, Phys. Rev. Lett. 87 (2001) 246405. doi:10.1103/PhysRevLett.87.246405

C.-J. Park, F. Giustino, C. D. Spataru, M. L. Cohen, S. G. Louie, Angle-resolved photoemission spectra of graphene from first-principles calculations, Nano Lett. 9 (2009) 4234–4239. doi:10.1021/nl092448y

B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, S. G. Louie, Quasiparticle band gap of ZnO: High accuracy from the conventional GW approach, Phys. Rev. Lett. 105 (2010) 146401. doi:10.1103/PhysRevLett.105.146401

C. Friedrich, M. C. Müller, S. Blügel, Band convergence and linearization error correction of all-electron GW calculations: The extreme case of zinc oxide, Phys. Rev. B 83 (2011) 081101. doi:10.1103/PhysRevB.83.081101

M. Stankovski, G. Antonius, D. Waroquiers, A. Miglio, H. Dixit, K. Sankaran, M. Giantomassi, X. Gonze, M. Côté, G.-M. Rignanese, G^W band gap of ZnO: Effects of plasmon-pole models, Phys. Rev. B 84 (2011) 241201. doi:10.1103/PhysRevB.84.241201

M. J. van Setten, M. Giantomassi, X. Gonze, G.-M. Rignanese, G. Haueter, Automation methodologies and large-scale validation for GW: Towards high-throughput GW calculations, Phys. Rev. B 96 (2017) 155207. doi:10.1103/PhysRevB.96.155207

F. Brunoval, X. Gonze, Accurate GW self-energies in a plane-wave basis using only a few empty states: Towards large systems, Phys. Rev. B 78 (2008) 085125. doi:10.1103/PhysRevB.78.085125

W. Kang, M. S. Hybertsen, Enhanced static approximation to the electron self-energy operator for efficient calculation of quasiparticle energies, Phys. Rev. B 82 (2010) 195108. doi:10.1103/PhysRevB.82.195108

J. A. Berger, L. Reining, F. Sottile, Ab initio calculations of electronic excitations: Collapsing spectral sums, Phys. Rev. B 82 (2010) 041103. doi:10.1103/PhysRevB.82.041103

J. A. Berger, L. Reining, F. Sottile, Efficient GW calculations for SnO2, ZnO, and rubrene: The effective-energy technique, Phys. Rev. B 85 (2012) 085126. doi:10.1103/PhysRevB.85.085126

J. Deslippe, G. Samsonidze, M. Jain, M. L. Cohen, S. G. Louie, Coulomb-hole summations and energies for GW calculations with limited number of empty orbitals: A modified static remainder approach, Phys. Rev. B 87 (2013) 165124. doi:10.1103/PhysRevB.87.165124

W. Gao, W. Xia, X. Gao, P. Zhang, Speeding up GW calculations to meet the challenge of large scale quasiparticle predictions, Sci. Rep. 6 (2016) 36849. doi:10.1038/srep36849

S. Baroni, P. Giannozzi, A. Testa, Green’s-function approach to linear response in solids, Phys. Rev. Lett. 58 (1987) 1861–1864. doi:10.1103/PhysRevLett.58.1861

S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory, Rev. Mod. Phys. 73 (2001) 515–562. doi:10.1103/RevModPhys.73.515

L. Reining, G. Oúida, R. W. Godby, Elimination of unoccupied-state summations in ab initio self-energy calculations for large supercells, Phys. Rev. B 56 (1997) R4301–R4304. doi:10.1103/PhysRevB.56.84301

H. F. Wilson, F. Gygi, G. Galli, Efficient iterative method for calculations of dielectric matrices, Phys. Rev. B 78 (2008) 113303. doi:10.1103/PhysRevB.78.113303

H. F. Wilson, D. Lu, F. Gygi, G. Galli, Iterative calculations of dielectric eigenvalue spectra, Phys. Rev. B 79 (2009) 245106. doi:10.1103/PhysRevB.79.245106

T. A. Pham, H.-V. Nguyen, D. Rocca, G. Galli, GW calculations using the spectral decomposition of the dielectric matrix: Verification, validation, and comparison of methods, Phys. Rev. B 87 (2013) 155148. doi:10.1103/PhysRevB.87.155148

P. Umari, G. Stentut, S. Baroni, Optimal representation of the polarization propagator for large-scale GW calculations, Phys. Rev. B 79 (2009) 201104. doi:10.1103/PhysRevB.79.201104

P. Umari, G. Stentut, S. Baroni, GW quasiparticle spectra from occupied states only, Phys. Rev. B 81 (2010) 115104. doi:
[123] D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, Phys. Rev. B 88 (2013) 085117. doi:10.1103/PhysRevB.88.085117