A low resources space time approach to the GW approximation

Dietrich Foerster, Saber Gueddida

Version of Record: https://www.sciencedirect.com/science/article/pii/S0927025620305693
Manuscript_21aa0bc5478054ab4e486440b8ec3cbf

A low resources space time approach to the GW approximation

Dietrich Foerster[a], Saber Gueddida[b,*]

[a] *Univ. Bordeaux, LOMA, CNRS UMR7598, F-33400 Talence, France.*
[b] *Univ. Lorraine, LPCT, CNRS UMR7019, F-54506 Vandoeuvre-Les-Nancy, France.*

Abstract

The GW approximation for electronic excitations is considered computationally hard because its CPU time scales as $O(N^4)$ with the number N of atoms in the unit cell. The space time approach to this approximation scales, in principle, as $O(N^3)$, but for crystals it is nearly impossible to exploit this feature because it requires too much computer memory. Here we remove this memory bottleneck by restricting the screened Coulomb interaction in the GW self-energy to its Nyquist content. This reduces both the needed memory and the CPU time by several orders of magnitude, while keeping satisfactory agreement with experimentally measured gaps. Our method allows exploiting the $O(N^3)$ scaling feature of the GW space time approach with modest computational resources and it will facilitate the computation of the fundamental gaps of organic semiconductors.

Keywords: Low scaling algorithm, GW approximation, Large unit cell semiconductors.

1. Introduction and Motivation

Kohn-Sham's density functional theory (DFT) gives an adequate description of the ground state of weakly correlated electronic systems, but it provides only a rough estimate of their excited states. The latter can be treated by many body techniques, of which Hedin's GW approximation [1] is the first necessary step because it provides quasi electrons and quasi holes with residual interactions. In the next step, one may then study excitonic features by using the Bethe-Salpeter method [2, 3, 4]. In non magnetic semiconductors, the GW approximation gives fairly accurate results [5, 6] for the transport gap that is closely related to the open circuit voltage of photovoltaic cells [7].

The GW approximation and its success in describing the fundamental gaps of weakly correlated semiconductors has been reviewed in great detail [8, 9, 10, 11]. While the method is successful where it can be applied, it is also known to be computationally difficult for semiconductors with a large number of atoms in their unit cell. This is because in most implementations of the GW approximation the needed CPU scales rather steeply, as $O(N^4)$,

*I am corresponding author
Email addresses: dietrich.foerster@u-bordeaux.fr (Dietrich Foerster), saber.gueddida@univ-lorraine.fr (Saber Gueddida)

with the number N of atoms in the unit cell which may lead to computations taking months on a supercomputer [12].

The situation is more favorable in the "space time approach" to the GW approximation [13, 14] where CPU time scales as $O(N^3)$ and there is also a promising stochastic implementation of GW [15, 16], where the needed CPU time scales only linearly as $O(N)$. Unfortunately, this latter method has given no results, so far, for periodic systems that are our focus here.

There are also methods where one parametrizes the response in an effective way [17, 18, 19], but in these computational approaches CPU time scales again as $O(N^4)$. The GW method works fairly well for molecules [20, 21, 22, 23] we focus here on periodic bulk systems, where the situation is much less satisfactory.

Because of its favorable $O(N^3)$ scaling, the space time method would be the natural choice for semiconductors with hundreds of atoms in their unit cell, except that it suffers from a "memory bottleneck" because entire correlation functions must be written into the computer memory at intermediate stages of the computation. A few years ago, Kaltak et *al.* [24] and Liu et *al.* [25] made important progress by lowering the number of frequencies in the imaginary time correlation functions. This reduced the needed memory enough for the resulting code to run on special purpose computers.

The aim of the present paper is to show that the screened Coulomb interaction W that enters the self-energy $\Sigma = iGW$ of the GW approximation may be expanded in waves that satisfy Nyquist's condition in each of their three components, see Eq. (9) further below. This reduces the needed memory by a factor of $2^6 = 64$ and it also reduces the CPU time of the construction of W by a factor of $8^3 = 512$. These reductions of both memory and CPU time by several orders of magnitude move the space time approach to GW from special purpose machines to more widely available laptops and work stations. Our method should also facilitate the study of organic semiconductors that have many atoms in their unit cell.

The remaining sections of this paper are organized as follows. In section 2 we introduce a "Nyquist approximation" to the GW equations. In section 3 we describe our algorithm. In section 4 we test our method on a few semiconductors. Our conclusions are given in section 5.

2. Introducing Nyquist's condition into the GW approximation

A density functional theory (DFT) computation [26, 27] of a periodic electronic system provides, as output, dispersing bands $E(\mathbf{p})$ and Kohn-Sham eigenstates $\psi^E_{\mathbf{p}}(\mathbf{r})$, where $\mathbf{p}$ ranges over a Brillouin zone of Bloch momenta. From the bands $E(\mathbf{p})$ and the wave functions $\psi^E_{\mathbf{p}}(\mathbf{r})$ one may define a propagator of noninteracting electrons

$$G_{\mathbf{p}}(\mathbf{r},\mathbf{r}',t) = -i\theta(t)\sum_{E(\mathbf{p})>0}\psi^E_{\mathbf{p}}(\mathbf{r})\psi^{*E}_{\mathbf{p}}(\mathbf{r}')e^{-iE(\mathbf{p})t} + i\theta(-t)\sum_{E(\mathbf{p})<0}\psi^E_{\mathbf{p}}(\mathbf{r})\psi^{*E}_{\mathbf{p}}(\mathbf{r}')e^{-iE(\mathbf{p})t} \quad (1)$$

where $\mathbf{r},\mathbf{r}'$ are points in a piece of periodic crystal and we use atomic units where $\hbar = 1$. The purpose of the GW approximation is to find the propagator of electrons that interact with each other via the screened Coulomb interaction by using Dyson's equation,

$G_{\text{interacting}} = (G^{-1}_{\text{non-interacting}} - \Sigma)^{-1}$. In the GW approach, one uses the noninteracting propagator $G_{\mathbf{p}}(\mathbf{r}, \mathbf{r}', t)$ to define a noninteracting density response function χ according to

$$\chi_{\mathbf{q}}(\mathbf{r}, \mathbf{r}', t) = -2i \sum_{\mathbf{p}_1 - \mathbf{p}_2 = \mathbf{q}} G_{\mathbf{p}_1}(\mathbf{r}, \mathbf{r}', t) G_{\mathbf{p}_2}(\mathbf{r}', \mathbf{r}, -t) \tag{2}$$

where the factor of 2 is from summing over spins in this non relativistic theory. From the density reponse χ one computes the RPA (Random Phase Approximation) inverse dielectric function $\frac{1}{\varepsilon_{\mathbf{q}}}$ and the correlated part of the screened Coulomb interaction. Because the Coulomb interaction is diagonal in Fourier space, one switches from $\chi_{\mathbf{q}}(\mathbf{r}, \mathbf{r}', t)$ to $\chi_{\mathbf{q}}(\mathbf{g}, \mathbf{g}', \omega)$ where $\mathbf{g}, \mathbf{g}'$ are periodic waves in the unit cell. It is easier to find the inverse of a kernel in the frequency domain, so one also switches to (imaginary) frequencies:

$$\begin{aligned} \left(\frac{1}{\varepsilon_{\mathbf{q}}}\right)(\mathbf{g}, \mathbf{g}', \omega) &= \frac{1}{\delta_{\mathbf{g}\mathbf{g}'} - V^{1/2}_{\mathbf{q}}(\mathbf{g}) \chi(\mathbf{g}, \mathbf{g}', \omega) V^{1/2}_{\mathbf{q}}(\mathbf{g}')} \\ W^c_{\mathbf{q}}(\mathbf{g}, \mathbf{g}', \omega) &= V^{1/2}_{\mathbf{q}}(\mathbf{g}) \left[\left(\frac{1}{\varepsilon_{\mathbf{q}}}\right)(\mathbf{g}, \mathbf{g}', \omega) - \delta_{\mathbf{g}\mathbf{g}'} \right] V^{1/2}_{\mathbf{q}}(\mathbf{g}') \end{aligned} \tag{3}$$

with $V_{\mathbf{q}}(\mathbf{g}) = \frac{4\pi}{(\mathbf{q}+\mathbf{g})^2}$ and where the fraction in Eq. (3) denotes the inverse of a matrix. From the DFT electron propagator $G_{\mathbf{p}}(\mathbf{r}, \mathbf{r}', t)$ of noninteracting electrons and from the screened Coulomb interaction $W^c_{\mathbf{q}}$ transformed back into space time, one finally computes the GW self-energy $\Sigma = iGW$ and its average in the unperturbed states $|\mathbf{p}E>$

$$\Sigma^{GW}_{\mathbf{p}\, c}(\mathbf{r}, \mathbf{r}', t) = i \sum_{\mathbf{p}_1 + \mathbf{p}_2 \equiv \mathbf{p}} G_{\mathbf{p}_1}(\mathbf{r}, \mathbf{r}', t) W^c_{\mathbf{p}_2}(\mathbf{r}, \mathbf{r}', t) \tag{4}$$

$$\delta E^{GW}_c(\mathbf{p}, E) = Z_{\mathbf{p}E} < \mathbf{p}E | \Sigma^{GW}_{\mathbf{p}\, c}(\omega = E_{\mathbf{p}}) | \mathbf{p}E >$$

$$Z_{\mathbf{p}E} = \frac{1}{1 - Re < \mathbf{p}E | \frac{\partial \Sigma^{GW}_{\mathbf{p}}}{\partial \omega}(\omega = E_{\mathbf{p}}) | \mathbf{p}E >}$$

see [8, 9, 10] for detailed derivations of these results. An exchange energy E^{GW}_x and an exchange correlation energy $E^{DFT}_{x,c}$ must also be taken into account

$$E^{GW}_x(\mathbf{p}, E) = < \mathbf{p}E | \Sigma^{GW}_{\mathbf{p},x} | \mathbf{p}E > \tag{5}$$

$$E^{DFT}_{x,c}(\mathbf{p}, E) = < \mathbf{p}E | V_{x,c} | \mathbf{p}E >$$

$$\Sigma^{GW}_{\mathbf{p},x}(\mathbf{r}, \mathbf{r}', t - t') = i \sum_{\mathbf{p}_1 + \mathbf{q} = \mathbf{p}} G_{\mathbf{p}}(\mathbf{r}, \mathbf{r}', 0^-) V_{\mathbf{q}}(\mathbf{r}, \mathbf{r}') \delta(t - t')$$

and, finally, the change $\delta E^{GW}(\mathbf{p})$ of the dispersion $E(\mathbf{p})$, according to the GW approximation, is given by

$$\delta E^{GW}(\mathbf{p}, E) = E^{GW}_x(\mathbf{p}, E) + E^{GW}_c(\mathbf{p}, E) - E^{DFT}_{x,c}(\mathbf{p}, E) \tag{6}$$

where $E_{x,c}^{DFT}$ is subtracted to avoid including exchange and correlations twice. $E_x^{GW}(\mathbf{p}, E)$, the largest term in the last equation, represents the Hartree-Fock exchange energy and it is partially compensated by the (negative) correlation energy $E_c^{GW}(\mathbf{p}, E)$. The construction of the screened Coulomb interaction W that enters the self-energy $\Sigma = iGW$ is the computational bottleneck of the GW approximation referred to in the introduction.

2.1. The Nyquist condition in plane wave DFT codes

To reproduce music up to a frequency of 20000 Hz, one famously samples the signal from the microphone 40000 times per second. From the perspective of the discrete Fourier transform, a periodic signal $F(n)$, with N_{ff} equidistant data points in the time domain, should only contain $\frac{N_{ff}}{2}$ modes in the frequency domain, or

$$F(n) = \sum_{k=-\frac{N_{ff}}{4}}^{\frac{N_{ff}}{4}} f(k) e^{(-\frac{2\pi i}{N_{ff}} kn)}, \tag{7}$$

with $n = -\frac{N_{ff}}{2}..\frac{N_{ff}}{2}$. In DFT computations in a plane wave basis, the Kohn-Sham wave functions $\psi_{\mathbf{p}}^E(\mathbf{r})$ are spatially periodic in an extended crystal of $N_{\mathbf{pp}}$ unit cells glued together in each direction, with the boundaries of this crystal identified. The results of a DFT computation are dispersing energy bands $E(\mathbf{p})$ and associated Kohn-Sham wave functions $\psi_{\mathbf{p}}^E(\mathbf{r}) = e^{i\mathbf{p}\cdot\mathbf{r}}\phi_{\mathbf{p}}^E(\mathbf{r})$, with $e^{i\mathbf{p}\cdot\mathbf{r}}$ a Bloch phase that is periodic in the extended crystal and where the functions $\phi_{\mathbf{p}}^E(\mathbf{r})$ are expanded in terms of waves $e^{i\mathbf{g}\cdot\mathbf{r}}$ that are periodic over the unit cell,

$$\phi_{\mathbf{p}}^E(\mathbf{r}) = \frac{1}{\sqrt{V_{uc}}} \sum_{\mathbf{g}} \varphi_{\mathbf{p}}^E(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}}, \tag{8}$$

$$\int_{\text{uc}} \phi_{\mathbf{p}}^{*E}(\mathbf{r}) \phi_{\mathbf{p}'}^{E'}(\mathbf{r}) d\mathbf{r} = \delta_{\mathbf{p},\mathbf{p}'}\delta_{EE'}.$$

Here V_{uc} is the volume of the unit cell and the wave numbers $\mathbf{g}$ that occur in the DFT code are limited, in each component, by Nyquist's condition

$$-\frac{N_{gg}}{4} \leq g_i \leq \frac{N_{gg}}{4}, i = 1..3 \tag{9}$$

In the plane wave DFT code ABINIT [28] that we use, one specifies a cutoff energy e_{cut} and ABINIT will choose a number of allowed modes N_{gg} in each direction of the unit cell and find, using an algorithm that remains unpublished, the spectral data $E(\mathbf{p})$, $\varphi_{\mathbf{p}}^E(\mathbf{g})$ for bands that oscillate increasingly in space. From its definition, the noninteracting electron propagator in Eq. (1) satisfies the Nyquist condition in both $\mathbf{g}$ and $\mathbf{g}'$ and, therefore, it contains $\frac{1}{64}N_{gg}^3$ rather than N_{gg}^3 modes, and requires, correspondingly, a factor $\frac{1}{64}$ less memory to store.

2.2. A Nyquist approximation to the exchange energy

The Coulomb interaction $V(\mathbf{r},\mathbf{r}')$ satisfies $\triangle V(\mathbf{r},\mathbf{r}') = -4\pi\delta(\mathbf{r}-\mathbf{r}')$ where $\delta(\mathbf{r}-\mathbf{r}')$ is the identity for a given set of plane waves $e^{i(\mathbf{q}+\mathbf{G})\mathbf{r}}$ and with $\mathbf{q}$ the Bloch momentum

$$V(\mathbf{r},\mathbf{r}') = \sum_{\mathbf{q}} V_{\mathbf{q}}(\mathbf{r},\mathbf{r}'),\ V_{\mathbf{q}}(\mathbf{r},\mathbf{r}') = \frac{4\pi}{\Omega}\sum_{\mathbf{g}} V_{\mathbf{q},\mathbf{g}} e^{i(\mathbf{q}+\mathbf{g})(\mathbf{r}-\mathbf{r}')} \tag{10}$$

$$V_{\mathbf{q},\mathbf{g}} = \frac{4\pi}{(\mathbf{q}+\mathbf{g})^2} \tag{11}$$

$V(\mathbf{r},\mathbf{r}')$ in Eq. (10) contains *all* the wave numbers $\mathbf{g}$ that are compatible with the unit cell, including wave numbers beyond the Nyquist condition. Let us now consider the error that imposing Nyquist's condition introduces in the GW exchange energy of Eq. (5). Using Eq. (10) for the Coulomb interaction, the exchange interaction can be brought into the following form

$$E_x^{GW}(\mathbf{p}_0,E_0) = <\mathbf{p}_0E_0|\Sigma_{\mathbf{p},x}^{GW}|\mathbf{p}_0E_0> \tag{12}$$

$$= -\frac{1}{\Omega}\sum_{\mathbf{q},\mathbf{g}}\frac{4\pi}{(\mathbf{q}+\mathbf{g})^2}\sum_{\mathbf{p}_1=\mathbf{p}_0-\mathbf{q}}\sum_{E(\mathbf{p}_1)\leq 0}|\int_\Omega <\mathbf{p}_0E_0|e^{i(\mathbf{g}+\mathbf{q})\mathbf{r}}|\mathbf{p}_1E_1> d\mathbf{r}|^2$$

where Ω stands for the volume of the extended crystal generated by gluing $N_{\mathbf{pp}}^3$ unit cells together. Imposing Nyquist's condition (see Eq. (9)) on the wave number $\mathbf{g}$ in the Coulomb interaction induces a relative error of

$$\delta_x = \frac{E_x^{GW}(\mathbf{p}_0,E_0)_{Nyquist-Coulomb} - E_x^{GW}(\mathbf{p}_0,E_0)_{full-Coulomb}}{E_x^{GW}(\mathbf{p}_0,E_0)_{full-Coulomb}} \tag{13}$$

We considered this error for silicon (Si) with $N_{gg} = \{8,12,16\}$ and found that δ_x is, respectively, of order $\{10^{-3},10^{-4},10^{-5}\}$ and, therefore, sufficiently small to be ignored. To understand this result, let us consider the Fourier transform of the overlaps of the KS waves that occur in Eq. (12). They are of the form

$$F(\mathbf{g})_{E_1\mathbf{p}_1E_2\mathbf{p}_2} = \int_\Omega <E_1\mathbf{p}_1|e^{i(\mathbf{g}+\mathbf{q})\mathbf{r}}|E_2\mathbf{p}_2> d\mathbf{r}$$

and we may decompose $F(\mathbf{g})_{E_1\mathbf{p}_1E_2\mathbf{p}_2}$ into Nyquist and non-Nyquist parts. To quantify the non-Nyquist part in the amplitude $F(\mathbf{g})_{E_1\mathbf{p}_1E_2\mathbf{p}_2}$, we compute the fraction

$$error(E_1\mathbf{p}_1E_2\mathbf{p}_2) = \frac{\sum_{\text{non-Nyquist}\mathbf{g}}|F(\mathbf{g})_{E_1\mathbf{p}_1E_2\mathbf{p}_2}|^2}{\sum_{\text{all}\mathbf{g}}|F(\mathbf{g})_{E_1\mathbf{p}_1E_2\mathbf{p}_2}|^2}$$

For a 10×10 block of bands, we find, for a given pair of $\mathbf{p}_1,\mathbf{p}_2$ and for $N_{gg} = \{8,12,16\}$, averages of $\{5.5{*}10^{-3},1.4{*}10^{-3},5.6{*}10^{-5}\}$, so clearly the non-Nyquist part of such products is small enough to lead to negligible changes in the exchange energy of Eq. (12). The exchange energy in Eq. (12) is a sum over occupied states with energies limited by the width of the valence bands and therefore it is plausible that the non-Nyquist part of the products $<\mathbf{p}_0E_0|e^{i\mathbf{g}\cdot\mathbf{r}}|\mathbf{p}_1E_1>$ is small enough to be neglected.

2.3. A Nyquist approximation to the solution of the GW equations

Consider now Adler's [29] expression for the density response at imaginary frequency $i\omega$

$$\widetilde{\chi}_{\mathbf{q}}(\mathbf{g},\mathbf{g}',i\omega) = -\frac{1}{\Omega}\sum_{\substack{\mathbf{p}_1-\mathbf{p}_2\equiv\mathbf{q}\\ E_1,E_2}}(n_{E_1}-n_{E_2})\frac{< E_2\mathbf{p}_2|e^{-i(\mathbf{g}+\mathbf{q})\mathbf{r}}|E_1\mathbf{p}_1 >< E_1\mathbf{p}_1|e^{i(\mathbf{g}'+\mathbf{q})\mathbf{r}'}|E_2\mathbf{p}_2 >}{i\omega-(E_1(\mathbf{p}_1)-E_2(\mathbf{p}_2))} \tag{14}$$

This correlator depends, again, on the overlaps $F(\mathbf{g})_{E_1\mathbf{p}_1E_2\mathbf{p}_2}$, but unlike in the exchange energy, here one of the energies $E_{1,2}$ ranges over all the empty states and the situation is therefore more complicated. For a given set of bands E_1, E_2, we expect the non-Nyquist contribution to drop out with increasing N_{gg}, but unfortunately we have no estimate of the non-Nyquist contribution at fixed N_{gg}. The inverse dielectric function $\frac{1}{\varepsilon_{\mathbf{q}}}$ depends on a modified response $\chi^1_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega)$ according to

$$\begin{aligned}\chi^1_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega) &= V^{1/2}_{\mathbf{q}}(\mathbf{g})\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega)V^{1/2}_{\mathbf{q}}(\mathbf{g}') \text{ with } V_{\mathbf{q}}(\mathbf{g}) = \frac{4\pi(1-\delta_{\mathbf{q},\mathbf{0}}\delta_{\mathbf{g},\mathbf{0}})}{(\mathbf{q}+\mathbf{g})^2} \qquad (15)\\ \left(\frac{1}{\varepsilon_{\mathbf{q}}}\right)_c(\mathbf{g},\mathbf{g}',\omega) &= \frac{1}{\delta_{\mathbf{g}\mathbf{g}'}-\chi^1_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega)}-\delta_{\mathbf{g}\mathbf{g}'}\end{aligned}$$

Finally the screened Coulomb interaction $W^c_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega)$ enters in the correlated part of the self-energy

$$W^c_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega) = V^{1/2}_{\mathbf{q}}(\mathbf{g})\left(\frac{1}{\varepsilon_{\mathbf{q}}}\right)_c(\mathbf{g},\mathbf{g}',\omega)V^{1/2}_{\mathbf{q}}(\mathbf{g}) \tag{16}$$

The matrix $\delta_{\mathbf{g}\mathbf{g}'}-\chi^1(\mathbf{g},\mathbf{g}',\omega)$ to be inverted naturally decomposes into four blocks with respect to the Nyquist properties of $\mathbf{g},\mathbf{g}'$ and we have found that the matrix norm of the Nyquist-Nyquist block dominates. The increasing dominance, as N_{gg} grows, of the norm of the Nyquist-Nyquist block (at fixed sub space dimension) of $\chi^1_{\mathbf{q}}, 1/\varepsilon$ and $W^c_{\mathbf{q}}$ suggests to retain only the Nyquist-Nyquist block in the computation. Because the error of this approximation should disappear as the number of modes grows, we view our approximation as a "convergence accelerator" of the solution of the GW equations. Lacking a mathematical proof that the non-Nyquist terms disappear with increasing $N_{\mathbf{gg}}$, we simply impose this condition and check whether it works.

3. The algorithm we used

Apart from using Nyquist's sampling condition Eq. (9) in the construction of the screened Coulomb interaction and apart from doing Fourier transformations in imaginary time differently, our algorithm remains essentially that of Godby and collaborators [13, 14] who do most of the computation for imaginary times and frequencies to avoid oscillations and to provide computational stability. The propagator and the response in imaginary time read

as

$$
\begin{aligned}
G^{+}_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau) &= i \sum_{E(\mathbf{p})<0} \psi^{E}_{\mathbf{p}}(\mathbf{r})\psi^{*E}_{\mathbf{p}}(\mathbf{r}')e^{E(\mathbf{p})\tau} \\
G^{-}_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau) &= -i \sum_{E(\mathbf{p})>0} \psi^{E}_{\mathbf{p}}(\mathbf{r})\psi^{*E}_{\mathbf{p}}(\mathbf{r}')e^{-E(\mathbf{p})\tau} \\
\chi_{\mathbf{q}}(\mathbf{r},\mathbf{r}',\tau) &= -2i \sum_{\mathbf{p}_1-\mathbf{p}_2=\mathbf{q}} G^{+}_{\mathbf{p}_1}(\mathbf{r},\mathbf{r}',\tau)G^{-}_{\mathbf{p}_2}(\mathbf{r}',\mathbf{r},\tau)
\end{aligned}
\tag{17}
$$

Following again [13, 14], Fourier transformations between imaginary time and imaginary frequency are defined according to

$$
f(\omega) = -i\int_{-\infty}^{\infty} F(\tau)e^{-i\omega\tau}d\tau,\; F(\tau) = \frac{i}{2\pi}\int_{-\infty}^{\infty} f(\omega)e^{i\omega\tau}d\omega \tag{18}
$$

The guiding principle of our own algorithm, and where we differ from [13, 14], is to express correlation functions in the smaller space of waves $\mathbf{g},\mathbf{g}'$ that satisfy Nyquist's condition. Our starting point for the (periodic) electron propagator in the domain of waves $\mathbf{g},\mathbf{g}'$ is, therefore,

$$
\begin{aligned}
G^{+}_{\mathbf{p}}(\mathbf{g},\mathbf{g}',\tau) &= i \sum_{E(\mathbf{p})<0} \sum_{\mathbf{g},\mathbf{g}'} \varphi^{E}_{\mathbf{p}}(\mathbf{g})\varphi^{*E}_{\mathbf{p}}(\mathbf{g}')\; e^{E(\mathbf{p})\tau} \\
G^{-}_{\mathbf{p}}(\mathbf{g},\mathbf{g}',\tau) &= -i \sum_{E(\mathbf{p})>0} \sum_{\mathbf{g},\mathbf{g}'} \varphi^{E}_{\mathbf{p}}(\mathbf{g})\varphi^{*E}_{\mathbf{p}}(\mathbf{g}')e^{-E(\mathbf{p})\tau} \\
\sum_{\mathbf{g}} \varphi^{E}_{\mathbf{p}}(\mathbf{g})\varphi^{*E'}_{\mathbf{p}}(\mathbf{g}) &= \delta_{EE'}
\end{aligned}
\tag{19}
$$

Because coordinates $\mathbf{r},\mathbf{r}'$ in real space are computationally expensive, they should occur only when needed to compute products of correlation functions. This then leads, rather naturally, to the following computational steps:

1. Import Kohn-Sham data $\{E(\mathbf{p}),\varphi^{E}_{\mathbf{p}}(\mathbf{g})\}$ and the crystal geometry for a semiconductor from a DFT code such as ABINIT [28].
2. Find the uncorrelated (denoted hereafter "unc") or Hartree-Fock (denoted hereafter

"HF") gap $\Delta_{\rm HF}$ from the exchange and exchange-correlation energies

$$\begin{aligned}
\delta E^{GW}_{\rm unc}(\mathbf{p},E) &= E^{GW}_{x}(\mathbf{p},E) - E^{DFT}_{xc}(\mathbf{p},E) \qquad (20)\\
E^{GW}_{x}(\mathbf{p},E) &= <\mathbf{p}E|\Sigma^{GW}_{\mathbf{p},x}|\mathbf{p}E>\\
E^{DFT}_{xc}(\mathbf{p},E) &= <\mathbf{p}E|V^{DFT}_{xc}|\mathbf{p}E>\\
E_{\rm HF}(\mathbf{p}) &= E(\mathbf{p}) + \delta E^{GW}_{\rm unc}(\mathbf{p},E)\\
\Delta_{\rm HF} &= \min_{\mathbf{p}}\left\{E^{\rm lumo}_{\rm HF}(\mathbf{p})\right\} - \max_{\mathbf{p}}\left\{E^{\rm homo}_{\rm HF}(\mathbf{p})\right\}
\end{aligned}$$

3. Prepare nonuniform grids τ in imaginary time and in imaginary frequencies ω for storing the electron propagator and other correlation functions and find matrices $M(\omega,\tau)$ and $M_{\pm\pm}(\omega,\tau)$ that represent the Fourier transformations for the response and the self-energy on these grids (further details will be given below).
4. Construct the electron propagator in wave form on the τ grid according to Eq. (19) and transform the electronic propagator $G^{\pm}_{\mathbf{p}}$ into its "dual" representation $G^{\pm}_{\mathbf{n}}$ that is needed for fast convolutions in Bloch momenta

$$G^{\pm}_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau) = \sum_{\mathbf{p}} e^{2\pi i \mathbf{n}\cdot\mathbf{p}/N_{\mathbf{pp}}} G^{\pm}_{\mathbf{p}}(\mathbf{g},\mathbf{g}',\tau)$$

5. Use $G^{\pm}_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau)$ to construct the noninteracting response $\chi_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau)$

$$G^{\pm}_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau) \ \underrightarrow{\text{two fold Fourier transformation}} \ G^{\pm}_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) \qquad (21)$$

$$\chi_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) = iG^{+}_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau)\overline{G^{-}_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau)}$$

$$\chi_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) \ \underrightarrow{\text{two fold Fourier transformation}} \ \chi_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau)$$

and return from $\chi_{\mathbf{n}}$ to $\chi_{\mathbf{q}}$

$$\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\tau) = \frac{1}{N^3_{\mathbf{pp}}}\sum_{\mathbf{n}} e^{-2\pi i \mathbf{n}\cdot\mathbf{q}/N_{\mathbf{pp}}}\chi_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau) \qquad (22)$$

6. Find the Fourier image $\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega) = -i\int_{-\infty}^{\infty}\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\tau)e^{i\omega\tau}d\tau$ on the nonuniform ω grid according to

$$\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega) = \sum_{\tau\geq 0} M(\omega,\tau)\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\tau) \qquad (23)$$

and construct the correlated part of the RPA dielectric function and of the screened Coulomb interaction by inverting in the Nyquist subspace

$$\frac{1}{\varepsilon_{\mathbf{q}}}(\mathbf{g},\mathbf{g}',\omega)_{\rm c} = \frac{1}{\delta_{\mathbf{gg}'} - V^{1/2}_{\mathbf{q},\mathbf{g}}\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega)V^{1/2}_{\mathbf{q},\mathbf{g}'}} - \delta_{\mathbf{gg}'} \qquad (24)$$

7. Fourier transform $\frac{1}{\varepsilon_{\mathbf{q}}}(\mathbf{g},\mathbf{g}',\omega)_{\rm c} \rightarrow \frac{1}{\varepsilon_{\mathbf{q}}}(\mathbf{g},\mathbf{g}',\tau)_{\rm c}$ via

$$\begin{aligned} \frac{1}{\varepsilon_{\mathbf{q}}}(\mathbf{g},\mathbf{g}',\tau)_{\rm c} &= \frac{i}{2\pi}\int_{-\infty}^{\infty}\frac{1}{\varepsilon_{\mathbf{q}}}(\mathbf{g},\mathbf{g}',\omega)_{\rm c}e^{i\omega\tau}d\omega \\ &\simeq \sum_{\omega\geq 0} M^{-1}_{\rm pseudo}(\tau,\omega)\frac{1}{\varepsilon_{\mathbf{q}}}(\mathbf{g},\mathbf{g}',\omega)_{\rm c} \end{aligned} \tag{25}$$

where $M^{-1}_{\rm pseudo}(\tau,\omega)$ is a suitably defined pseudo inverse of the (forward) matrix $M(\omega,\tau)$ in Eq. (23). Use $\frac{1}{\varepsilon_{\mathbf{q}}}$ to construct the screened Coulomb interaction W^c first for Bloch momenta $\mathbf{q}$ and then in the space $\mathbf{n}$ that is "conjugate" to $\mathbf{q}$:

$$\begin{aligned} W^c_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\tau) &= V^{1/2}_{\mathbf{q}}(\mathbf{g})\frac{1}{\varepsilon_{\mathbf{q}}}(\mathbf{g},\mathbf{g}',\tau)_{\rm c}V^{1/2}_{\mathbf{q}}(\mathbf{g}') \\ W^c_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau) &= \sum_{\mathbf{q}} e^{2\pi i\mathbf{n}\cdot\mathbf{q}/N_{\mathbf{pp}}}W^c_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\tau) \end{aligned} \tag{26}$$

8. Construct the self-energy correction for a set of bands near the Fermi energy

$$\begin{pmatrix} G_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau) \\ W_{\mathbf{n}}(\mathbf{g},\mathbf{g}',\tau) \end{pmatrix} \xrightarrow{\text{two fold Fourier transformation}} \begin{pmatrix} G_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) \\ W_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) \end{pmatrix} \tag{27}$$

$$\Sigma_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) = iG_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau)W_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau)$$

return to Bloch space :

$$\Sigma_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau) = \frac{1}{N^3_{\mathbf{pp}}}\sum_{\mathbf{n}} e^{-2\pi inq/N_{\mathbf{pp}}}\Sigma_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau)$$

find the self energies for some bands :

$$\Sigma_{\mathbf{p}E}(\tau) = \int_{\Omega}\psi^*_{E\mathbf{p}}(\mathbf{r})\Sigma_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau)\psi_{E\mathbf{p}}(\mathbf{r}')d\mathbf{r}d\mathbf{r}'$$

9. Construct $\Sigma_{\mathbf{p}E}(\omega) = -i\int_{-\infty}^{\infty}\Sigma_{\mathbf{p}E}(\tau)e^{i\omega\tau}d\tau$ in a notation with $\omega,\tau\geq 0$ using

$$\begin{pmatrix} \Sigma_+(\omega) \\ \Sigma_-(\omega) \end{pmatrix} = \sum_{\tau\in grid,\ \tau\geq 0}\begin{pmatrix} M_{++}(\omega,\tau) & M_{+-}(\omega,\tau) \\ M_{-+}(\omega,\tau) & M_{--}(\omega,\tau) \end{pmatrix}\begin{pmatrix} \Sigma_+(\tau) \\ \Sigma_-(\tau) \end{pmatrix} \tag{28}$$

with matrices $M_{\pm\pm}(\omega,\tau)$ that were precomputed at the beginning.

10. Do a Pade analytic continuation to real frequencies and compute the correlated part of the GW self energies according to Eq. (4)

$$\Sigma_{\mathbf{p}E}(\omega) \rightarrow \Sigma_{\mathbf{p}E}(E(\mathbf{p})),\ Z_{\mathbf{p}E} = \left(1 - Re\frac{\partial\Sigma_{\mathbf{p}}}{\partial\omega}\right)^{-1}_{\omega=E(\mathbf{p})} \tag{29}$$

Find the GW gap according to

$$\Delta_{\rm GW} = \Delta_{\rm HF} + \left[Z_{\mathbf{p}E}\Sigma^c_{\mathbf{p}E}(\mathbf{p},E)\right]_{\rm lumo} - \left[Z_{\mathbf{p}E}\Sigma^c_{\mathbf{p}E}(\mathbf{p},E)\right]_{\rm homo} \tag{30}$$

Below we provide details on the above computational steps.

3.1. Construction of grids in imaginary time and frequency

In imaginary time, the electron propagator $G_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau)$ is a sum of rapidly decaying exponentials in τ with little structure. Adler's exact expression of the response function $\chi_{\mathbf{q}}(\mathbf{r},\mathbf{r}',t)$ in Eq. (14) is again a sum of decaying exponentials, and it follows from their spectral representations that the screened Coulomb interaction $W_{\mathbf{q}}(\mathbf{r},\mathbf{r}',t)$ and the self-energy $\Sigma^{GW}_{\mathbf{p}}(\mathbf{r},\mathbf{r}',t)$ are also of this form. Therefore we construct appropriate nonuniform finite grids, both in imaginary time and in imaginary frequencies, in a construction that may be similar to that of Kaltak et al. [24]. To find a nonuniform grid in τ over the interval $(0,\tau_{\rm max})$ we impose that the grid points be uniformly spaced near $\tau = 0$ and that they should expand exponentially for growing τ. We used the grid points

$$\tau_n = A\sinh\frac{n\cdot d\tau}{A},\ n = 0..M \tag{31}$$

with the parameter A fixed by requiring τ_M to coincide with $\tau_{\rm max}$, where M of the order of 20 gives sufficient accuracy. Near $n = 0$ the grid remains uniform with $\tau_n = n\cdot d\tau + O(n^3)$. Except for a scale factor, we used the same grid in imaginary frequency as in imaginary time

$$\omega_n = \frac{\omega_{\rm max}}{\tau_{\rm max}}\tau_n,\ n = 0..M \tag{32}$$

3.2. Choice of the extension $(0,\tau_{\rm max})$, $(0,\omega_{\rm max})$ of the grids

We have found that $\omega_{\rm max} = 200..250 * \Delta_{DFT}$ where Δ_{DFT} represents the gap according to DFT is sufficient to represent a simplified response

$$\chi_0(\tau) = e^{-2\Delta|\tau|},\ \chi_0(\omega) = -i\int_{-\infty}^{\infty}\chi_0(\tau)e^{-i\omega\tau}d\tau = -\frac{4i\Delta}{(2\Delta)^2+\omega^2} \tag{33}$$

There is a uniform grid spacing $d\omega = \frac{4\omega_{\rm max}}{N_{ff}}$ associated with the interval $(0,\omega_{\rm max})$ where $N_{ff} = 512$ is a remnant of an earlier fft type method we developed before. To find the interval $(0,\tau_{\rm max})$ we used $\tau_{\rm max} = \frac{N_{ff}d\tau}{4}$ and the fft type relation $d\omega * d\tau = \frac{2\pi}{N_{ff}}$.

3.3. Fourier transformations between time and frequency for the density response

Again, our technique may be similar to that of Kaltak et al. [24], but we were unable to understand their work. The density response in the frequency domain is given by

$$\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega) = -i\int_{-\infty}^{\infty}\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',t)e^{-i\omega\tau}d\tau \tag{34}$$

We now consider exponential functions $F(t) = e^{-\gamma|t|}$ on the grid $\{\tau_n,\ n = 0..M\}$ and their Fourier images $f(\omega)$

$$f(\omega) = -i\int_{-\infty}^{\infty} F(t)e^{-i\omega\tau}d\tau = -\frac{2i\gamma}{\gamma^2+\omega^2} \tag{35}$$

On the τ grid, the space of functions is of dimension $M+1$ and we span this space by functions $X(\tau,k) = e^{-\varepsilon_k|\tau|}$, where we used the frequencies from the frequency grid as exponents, but shifted by $\varepsilon_k = \omega_k + 2\Delta_{DFT}$, because the minimal exponent must be twice the gap:

$$X(\tau,k) = e^{-\varepsilon_k|\tau|} \quad \Rightarrow Y(\omega,k) = -\frac{2i\varepsilon_k}{{\varepsilon_k}^2+\omega^2} \tag{36}$$

So a set of M functions on the τ grid is transformed into a set of M functions $Y(\omega,k) = -\frac{2i\varepsilon_k}{{\varepsilon_k}^2+\omega^2}$ on the ω grid. Clearly there must exist a matrix, say $M(\omega,\tau)$, that describes the map between these two sets of functions

$$Y(\omega,k) = \sum_{\tau\in grid,\ \tau\geq 0} M(\omega,\tau)X(\tau,k) \tag{37}$$

Because both the response and the functions $X(\tau,k)$ are even in τ, we restrict ourselves to the $\tau \geq 0$ part of the grid in the summation. The matrix $M(\omega,\tau)$ is *formally* given by $M = YX^{-1}$ but we expect the inverse of the matrix X to be singular because of linear dependences among the functions in the τ domain. Searching for $M(\omega,\tau)$ by minimizing a suitably defined quadratic error, we are lead to the result

$$M(\omega,\tau) = \sum_{k,\tau'} Y(\omega,k) * X^{+}(k,\tau') * (XX^{+})^{-1}_{\tau',\tau_{\text{pseudo}}} \tag{38}$$

where we have replaced the inverse of the hermitian matrix $A = XX^{+}$ by its "pseudo inverse" defined as follows

$$\begin{aligned} (A)^{-1}_{\text{pseudo}} &= \sum_{\lambda>\text{threshold}} \frac{Z_\lambda Z^{*}_\lambda}{\lambda} \\ AZ_\lambda &= \lambda Z_\lambda \end{aligned} \tag{39}$$

The condition $\lambda >$ threshold means that the matrix $A(\tau',\tau)$ is inverted only in the space spanned by eigenvectors corresponding to eigenvalues above a threshold adjusted to minimize the error. We also need the inverse Fourier transform $F(t) = \frac{i}{2\pi}\int_{-\infty}^{\infty} f(\omega)e^{-i\omega\tau}d\omega$ for the (inverse) dielectric function in imaginary time. This Fourier inverse should be represented by $M^{-1}(\tau,\omega)$, but again we must replace the straight forward inverse by its pseudo inverse with a suitable threshold. We tested the matrices $M(\omega,\tau)$, $M^{-1}_{pseudo}(\tau,\omega)$ by comparing with the exact result given by Adler and found, in the case of silicon, relative forward and backward errors of $\sim 10^{-5}..10^{-4}$ that are small enough for our purpose.

3.4. Fourier transformations for the self-energy

According to Eq. (4) the self-energy $\Sigma(\tau)$ is discontinuous at $\tau = 0$ and therefore we parametrize both $\Sigma(\tau)$ and its Fourier image $\Sigma(\omega) = -i\int_{-\infty}^{\infty}\Sigma(\tau)e^{-i\omega\tau}d\tau$ in terms of, respectively, two functions $\Sigma_{\pm}(\tau)$, $\Sigma_{\pm}(\omega)$ at $\tau \geq 0$, $\omega \geq 0$ according to

$$\begin{aligned}\Sigma(\tau) &= \theta(\tau)\Sigma_{+}(\tau) + \theta(-\tau)\Sigma_{-}(-\tau)\\ \\ \Sigma(\omega) &= \theta(\omega)\Sigma_{+}(\omega) + \theta(-\omega)\Sigma_{-}(-\omega)\end{aligned}$$

The functions $\Sigma_{\pm}(\omega)$ on the frequency grid are related to the functions $\Sigma_{\pm}(\tau)$ on the τ grid as

$$\left(\begin{array}{c}\Sigma_{+}(\omega)\\ \Sigma_{-}(\omega)\end{array}\right) = \sum_{\tau\in grid,\ \tau\geq 0}\left(\begin{array}{cc}M_{++}(\omega,\tau) & M_{+-}(\omega,\tau)\\ M_{-+}(\omega,\tau) & M_{--}(\omega,\tau)\end{array}\right)\left(\begin{array}{c}\Sigma_{+}(\tau)\\ \Sigma_{-}(\tau)\end{array}\right)$$

For convenience, we place the Fermi energy at the homo level, so the functions $\Sigma_{\pm}(\tau)$ decay asymptotically at least as fast as $e^{-c\tau}$, with $c = 2\Delta_{DFT}$, where Δ_{DFT} denotes the DFT gap. We therefore generate the space of functions $\Sigma_{\pm}(\tau)$ as $\theta(\tau)e^{-(c+\omega_k)t}$ where we use the same frequency grid $\{\omega_k, k = 0..M\}$ as before. The construction is very similar to the one given for the response functions. We estimated the precision of these Fourier matrices by comparing with exact results for the self-energy for an artificial plasmon oscillator and found, in the case of silicon, an error of $\sim 10^{-7}$ for a frequency $\omega_{plasmon} = 10\Delta_{DFT}$

3.5. Fourier transformations of correlators between points in space and waves

These are Fourier transformations in $d = 3 + 3$ dimensions

$$\begin{aligned}\chi_{\mathbf{q}}(\mathbf{r},\mathbf{r}',\omega) &= \frac{1}{\Omega}\sum_{\mathbf{g},\mathbf{g}'}\chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega)e^{i\mathbf{q}(\mathbf{r}-\mathbf{r}')+i(\mathbf{g}\mathbf{r}-\mathbf{g}'\mathbf{r}')} \qquad (40)\\ \\ \chi_{\mathbf{q}}(\mathbf{g},\mathbf{g}',\omega) &= \frac{1}{\Omega}\int_{tore}\chi_{\mathbf{q}}(\mathbf{r},\mathbf{r}',\omega)e^{-i\mathbf{q}(\mathbf{r}-\mathbf{r}')-i(\mathbf{g}\mathbf{r}-\mathbf{g}'\mathbf{r}')}d\mathbf{r}d\mathbf{r}'\end{aligned}$$

and they must be adequately distributed over processors in a future parallel version of the code.

3.6. Fourier transformations between Bloch momenta and cell indices

The equations for the density response (Eqs. (2)) and the self-energy (Eq. (4)) contain convolutions that are done by switching from Bloch momenta $\overrightarrow{\mathbf{p}}$ to cell indices $\overrightarrow{\mathbf{n}}$. Consider Eq. (4) for the self-energy. By defining "dual" quantities that are labelled by cell indices $\overrightarrow{\mathbf{n}}$

$$\Sigma_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) = \sum_{\mathbf{p}} e^{2\pi i\mathbf{n}\cdot\mathbf{p}/N_{\mathbf{pp}}}\Sigma_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau),\ G_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) = \sum_{\mathbf{p}} e^{2\pi i\mathbf{n}\cdot\mathbf{p}/N_{\mathbf{pp}}}G_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau) \qquad (41)$$

we find

$$\begin{aligned}\Sigma_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau) &= iG_{\mathbf{n}}(\mathbf{r},\mathbf{r}',t)W_{\mathbf{n}}(\mathbf{r},\mathbf{r}',t) \qquad (42)\\ \\ \Sigma_{\mathbf{p}}(\mathbf{r},\mathbf{r}',\tau) &= \frac{1}{N_{\mathbf{pp}}^3}\sum_{\mathbf{n}} e^{-2\pi i\mathbf{n}\cdot\mathbf{p}/N_{\mathbf{pp}}}\Sigma_{\mathbf{n}}(\mathbf{r},\mathbf{r}',\tau)\end{aligned}$$

This amounts to a fast convolution in Bloch momenta in Eq. (4).

3.7. Pade analytic continuation

In step 9, we obtain $\Sigma_{\mathbf{p}E}$ on a nonuniform grid on the imaginary frequency axis. This is represented in Pade fashion via

$$\Sigma_{\mathbf{p}E}(i\omega) = \sum \frac{c^{(k)}_{\mathbf{p}E}}{i\omega - z_k} = \frac{P_N(i\omega)}{Q_{N+1}(i\omega)} \tag{43}$$

where P_N and Q_{N+1} are polynomials of, respectively, degree N and $N+1$. The Thiele type continued fraction representation for $\Sigma_{\mathbf{p}E}(i\omega)$ is analytically continued by replacing $i\omega \to \omega$ in $\Sigma_{\mathbf{p}E}$ [30].

4. Testing the method on some semiconductors

To test our Nyquist approximation, we use Eq. (30) to compute the gaps of a few simple semiconductors such as silicon, diamond, a cubic phase of boron nitrate and of crystals of acetylene. In principle, a GW computation should include an infinite number of bands up to the ionization threshold and this is obviously unfeasible. To solve this difficulty in the space time approach, Godby and collaborators replaced DFT electron propagators near the ionization threshold by propagators of free electrons [14]. A large number of increasingly accurate extrapolation schemes have been developed since within various $O(N^4)$ algorithms [31, 32]. Here we simply follow [33] and extrapolate the gap from data for two or three data points for large numbers of bands n to $n = \infty$ by using the representations

$$\Delta_{GW}(n) = a_0 + \frac{a_1}{n} \quad \text{or} \quad \Delta_{GW}(n) = a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} \tag{44}$$

4.1. Some technical details

Computations for small systems such as silicon were done on an old laptop with 8 Gbyte of memory while for larger systems we used a single core of a somewhat obsolete workstation. To save memory, we used a periodic crystal with only $N_{pp} = 3$ unit cells glued together. That such small systems are large enough to represent infinite crystals was discovered by [13, 14] and rationalized later in more detail [34]. Like Kaltak et *al.* [24] and Liu et *al.* [25] we use time and frequency grids with 20 grid points to represent correlators, although 10 grid points were also used for exploratory computations. The correlators that we must deal with satisfy hermiticity and time reversal relations that read, respectively, in the case of the electron propagator

$$\begin{aligned} G^*_{\mathbf{n}}(\mathbf{g}', \mathbf{g}, \tau) &= -G_{-\mathbf{n}}(\mathbf{g}, \mathbf{g}', \tau) \\ G_{\mathbf{n}}(\mathbf{g}, \mathbf{g}', \tau) &= -G^*_{\mathbf{n}}(-\mathbf{g}, -\mathbf{g}', \tau) \end{aligned} \tag{45}$$

The above symmetry relations are useful because they reduce the needed memory by a factor of 4.

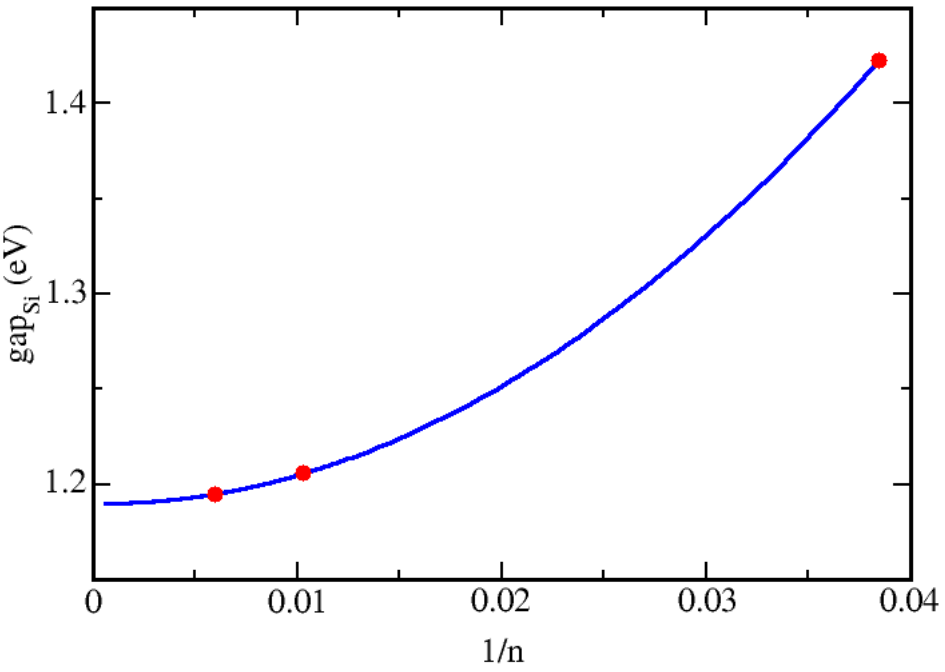

Figure 1: Computed gap (in eV) of silicon (1.422 eV, 1.21 eV, and 1.20 eV corresponding to 26, 97, and 168 bands with red dots representing the raw data) versus the inverse of the number of bands.

4.2. Computational results

We begin with silicon, see figure 1 where we find $\Delta_{GW} = 1.18$ eV and 1.19 eV by extrapolationg, respectively, two and three data points. The close agreement with the experimental gap of 1.17 eV for silicon [35] may be a coincidence. In the perturbative approach, we use the GW correlation energy only at $\omega = E_{\rm homo}, E_{\rm lumo}$, but it is instructive to examine the global behavior of the GW correlation energy $\Sigma_{\mathbf{p}E}(\omega)$ Eq. (29) before and after analytic continuation, see figure 2. The figure 2 (right panel) shows half parabolas in the imaginary part of the self-energy below and above the fermi energy (at zero) because we switch from the lower to the upper imaginary axis at this point. Therefore the imaginary part of the self-energy is seen to vanish quadratically at the Fermi energy, as it should. Our results for the self-energy at imaginary and real frequencies resemble those of Godby and coworkers [13, 14].

To judge the convergence of our results with respect to the number of gridpoints, we recomputed the gap of silicon for $N_{gg} = 16$ with gridpoints $M = \{10, 15, 20, 25\}$ and obtained the gaps of $\{1.17, 1.18, 1.23, 1.20\}$. Although we consider twenty gridpoints sufficient for our computations, this suggests the choice of the gridpoints should be variationally optimized.

For the diamond phase of carbon, see figure 3, we obtain a gap of 5.7 eV by interpolating

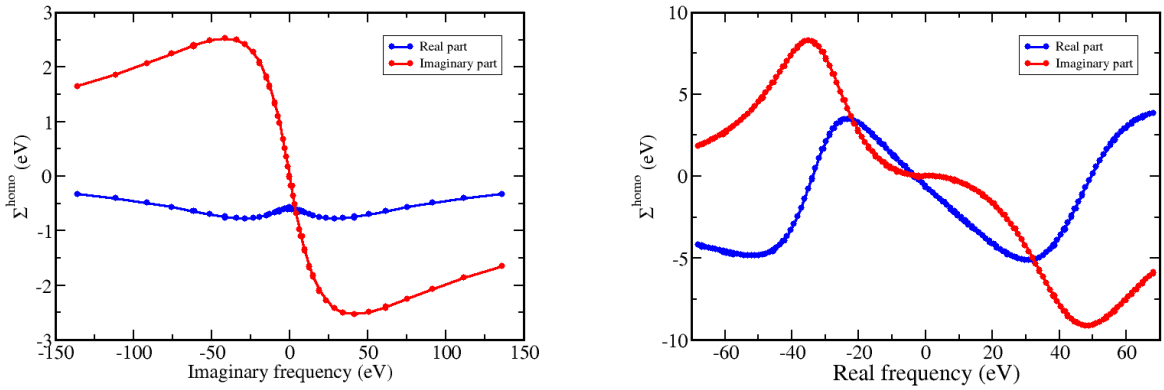

Figure 2: The self-energy of the homo level vs real (right panel) and imaginary (left panel) frequency.

linearly in $1/n$, which is reasonably close to the experimental gap of 5.8 eV [35].

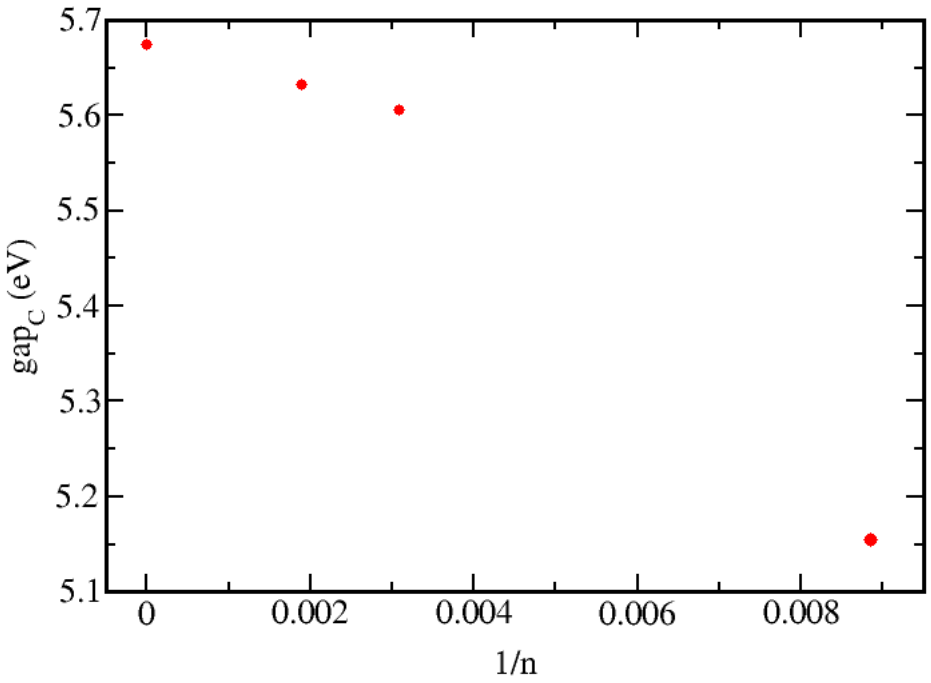

Figure 3: Computed gap (in eV) of diamond (5.15, 5.60, 5.63 eV corresponding to 113, 323, 528 bands (red dots)) versus the inverse of the number of bands.

For Boron Nitride, we consider only the cubic or Zinkblende phase. Our data extrapolate

to a gap of 6.5 eV while the experimental gap is 6.4 eV [35]

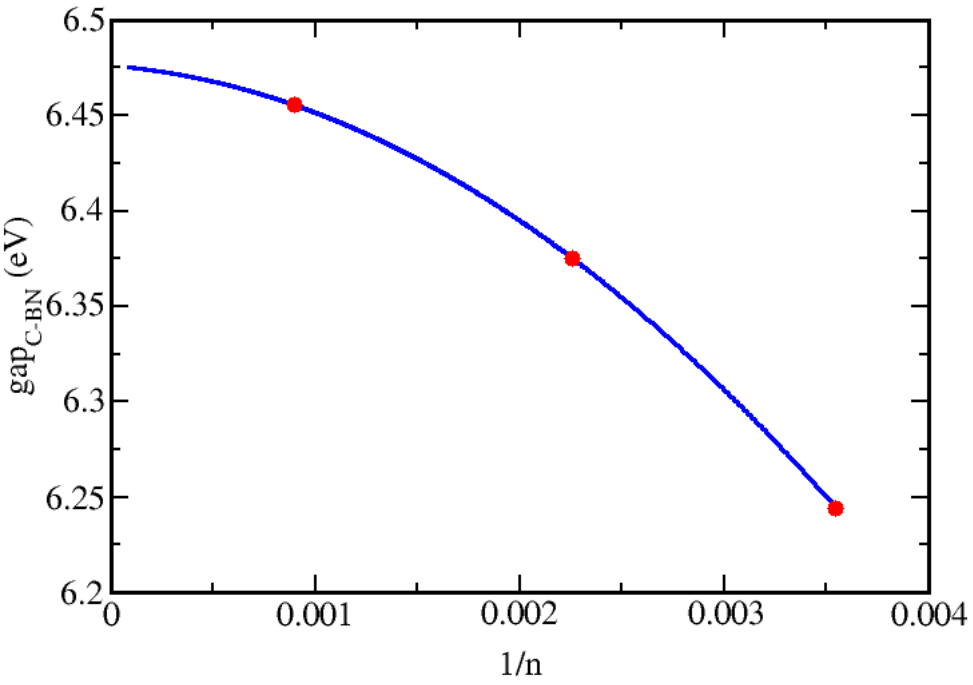

Figure 4: Computed gap (in eV) of cubic Boron nitride (6.24 eV, 6.38 eV, and 6.46 eV corresponding to 282, 443 and 1117 bands, red dots) versus the inverse of the number of bands.

Since our method was motivated by the need to predict transport gaps in organic semiconductors, we also considered crystals of acetylene [36, 37] although we are lacking experimental data to compare with. By extrapolating linearly, we found a transport gap of 6.4 eV.

4.3. Growth of CPU time with system size

It is encouraging that sequential computations for silicon at $N_{gg} = 8$ take only about 10 seconds on a PC. Although our work station needs about a day for computing the gap for crystals of acetylene at $N_{gg} = 28$ when using a single core, this should reduce to roughly an hour with computations done in parallel by all of its cores.

Let $N = (\frac{N_{gg}}{2} + 1)^3$ be the dimension of the correlation matrices that depend on $\mathbf{g}, \mathbf{g}'$. The operations in the algorithm that scale as N^3 are the inversion for finding $\frac{1}{\varepsilon}$ in Eq. (15) and the construction of the electron propagator in Eq. (19). The remaining operations, such as forming the self-energy and doing Fourier transforms of correlation functions from waves $\mathbf{g}, \mathbf{g}'$ to space points $\mathbf{r}, \mathbf{r}'$ scale as N^2, apart from logarithmic corrections.

To find how the needed CPU time grows with the number of atoms, one may construct artificial unit cells that contain multiple copies of the true unit cell of a crystal and observe

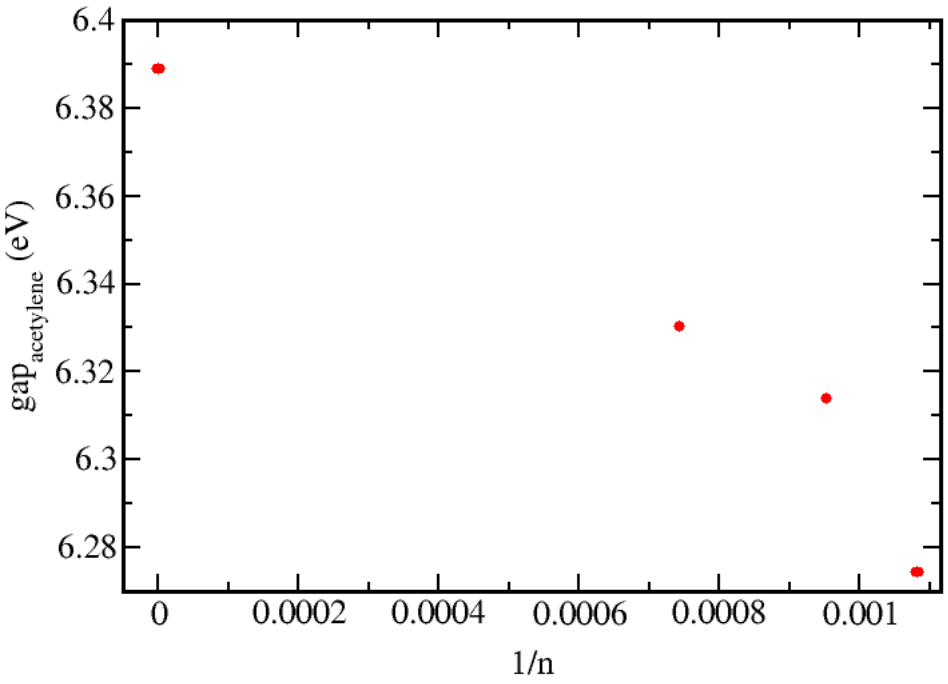

Figure 5: Computed gap (in eV) of acetylene crystals (6.27, 6.31, 6.33 eV corresponding to 925, 1050, 1345 bands (red dots)) versus the inverse of the number of bands.

how the time for computations grows with the number of atoms in it. Here we use a technically simpler method where we keep the unit cell, but vary the number of waves N_{gg} and find an exponent x

$$CPU \text{ time } \sim N_{gg}^x$$

In a second step we observe that the number of modes needed in a plane wave DFT computation for a material with n atoms in the unit cell is about $N_{gg} \sim 10\ n^{1/3}$ and we *could* have put approximately $n \sim \frac{N_{gg}^3}{1000}$ atoms in this unit cell. We conclude that CPU time grows with the number of atoms as

$$CPU \text{ time} \sim n^{x/3}$$

To find the exponent x of growth of CPU time with N_{gg}, we compute the gap of cubic Boron nitride for $N_{gg} = \{12, 16, 20, 24, 28\}$. Figure 6 gives the logarithm of the CPU time as a function of the logarithm of N_{gg} and from the slope we find $\frac{x}{3} = 2.35$. This shows that for the systems we considered, the computational cost of the operations that scale as N^3 is still small, although for very large systems their cost will dominate.

We found that the $O(N^4)$ GW code of abinit has a runtime comparable to our code for a small system with $N_{gg} = 8$. As its run time scales as N^4 or, equivalently as N_{gg}^{12}, this leads,

at $N_{gg} = 28$, to an estimated run time larger by a factor of $(\frac{28}{8})^{4.95} \sim 500$ than that of our code, turning a day into a year, when computing on a single core.

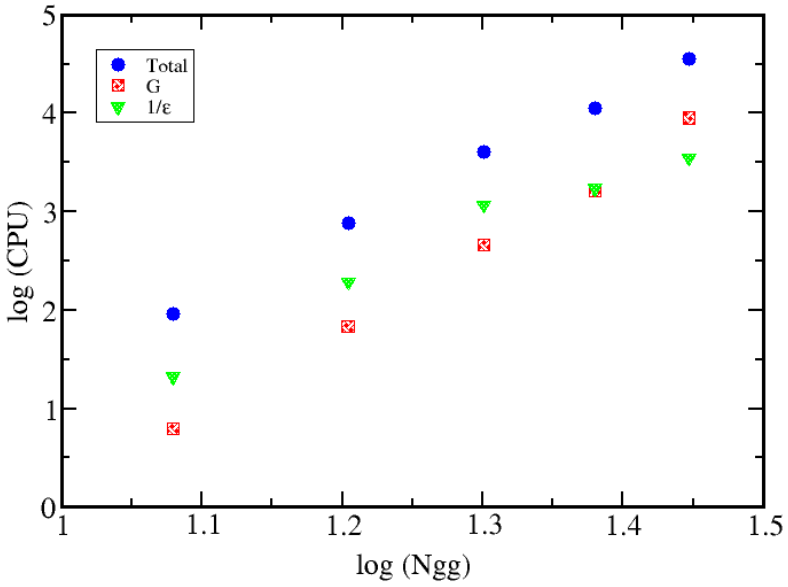

Figure 6: Plot of log(CPU_{Total}) vs log(Ngg) to identify the scaling exponent x in the relation CPU$_{Total}$ $\sim$ Ngg^x. The figure also gives the CPU times for setting up the propagator and for finding the inverse dielectric function that should both scale as Ngg^9.

5. Conclusions

We considered the $O(N^3)$ space time approach to GW by Godby and collaborators and the memory bottleneck that still remained after significant progress made by Kaltak et *al.* [24] and Liu et *al.* [25]. We found that restricting the density response and the screened Coulomb interaction to their Nyquist subspace reduces memory and CPU time by several orders of magnitudes, while keeping satisfactory agreement with the experimental data. Our "Nyquist approximation" allows using the space time approach to GW on widely available laptops and workstations and special purpose computers are no longer needed.

Here we presented a "proof of principle" of our method. A parallel implementation of our algorithm is planned to allow predictions for organic semiconductors and other challenging materials where the $O(N^3)$ scaling of the space time approach is crucial.

The important reduction in CPU time and memory that we found might also facilitate computations in extensions of the GW approximation that include spin [38] or external electrodes [39]. Here we considered GW as applied to data imported from DFT computations that use pseudo potentials and we did not consider the more precise DFT method developed by Bloechl and his collaborators [40, 41].

Acknowledgments

The present work started as the project "ORGAVOLT" of the ANR "MODELES NUMERIQUES" and which funded one of us (S. G.). We acknowledge useful advice and insight provided by Mark E. Casida of the University of Grenoble, one of our partners in this project. The synthesis and characterization of organic semiconductors by the groups of Georges Hadzioannou, Lionel Hirsch and Guillaume Wantz at the University of Bordeaux motivated this work.

Discussions with Koussay Lazaar from the University of Monastir helped to bring into focus some of the ideas of our project. Sébastien Lebegue of Nancy University sent us an early version of ABINIT that helped us write an interface for importing the results from this software package and he also commented on the manuscript. We are indebted to Georg Kresse from the University of Vienna for comments on the need for special hardware to run the Vienna $O(N^3)$ GW code and to Charles Patterson from the University of Dublin for suggesting to consider the transport gap in crystals of acetylene.

The University of Bordeaux contributed to the funding of a dedicated work station that Alain Bluck and Eric Basset of LOMA installed and helped us to run and we are also indebted to Jean-Paul Caltagirone of TREFLE for advice on computing. A free licence of "Parallel Studio" was kindly provided by INTEL, and the speed and reliability of this software, along with inspiring training sessions organized by INTEL on its use, contributed to this project in an essential way.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] L. Hedin, New method for calculating the one-particle Green's function with application to the electron-gas problem, Phys. Rev. 139 (1965) A796–A823. doi:10.1103/PhysRev.139.A796.
URL `https://link.aps.org/doi/10.1103/PhysRev.139.A796`

[2] C. Faber, P. Boulanger, C. Attaccalite, I. Duchemin, X. Blase, Excited states properties of organic molecules: from density functional theory to the GW and bethe 2013;salpeter Green's function formalisms, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 372 (2011) (2014) 20130271. doi:10.1098/rsta.2013.0271.
URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2013.0271`

[3] T. Rangel, A. Rinn, S. Sharifzadeh, F. H. da Jornada, A. Pick, S. G. Louie, G. Witte, L. Kronik, J. B. Neaton, S. Chatterjee, Low-lying excited states in crystalline perylene, Proceedings of the National Academy of Sciences 115 (2) (2018) 284–289. arXiv:https://www.pnas.org/content/115/2/284.full.pdf, doi:10.1073/pnas.1711126115.
URL `https://www.pnas.org/content/115/2/284`

[4] P. Puschnig, C. Meisenbichler, C. Draxl, Excited state properties of organic semiconductors: Breakdown of the Tamm-Dancoff approximation (2013). arXiv:1306.3790.

[5] H. Jiang, P. Blaha, *GW* with linearized augmented plane waves extended by high-energy local orbitals, Phys. Rev. B 93 (2016) 115203. doi:10.1103/PhysRevB.93.115203.
URL `https://link.aps.org/doi/10.1103/PhysRevB.93.115203`

[6] M. Grumet, P. Liu, M. Kaltak, J. c. v. Klimeš, G. Kresse, Beyond the quasiparticle approximation: Fully self-consistent *GW* calculations, Phys. Rev. B 98 (2018) 155143. doi:10.1103/PhysRevB.98.155143.
URL https://link.aps.org/doi/10.1103/PhysRevB.98.155143
[7] P. Wurfel, Physics of Solar Cells: From Basic Principles to Advanced Concepts, 2nd Edition, 2009.
[8] F. Aryasetiawan, O. Gunnarsson, The GW method, Rep. Prog. Phys. 61 (3) (1998) 237–312. doi:10.1088/0034-4885/61/3/002.
[9] F. Bruneval, Exchange and correlation in the electronic structure of solids, from silicon to cuprous oxide: GW approximation and beyond, Ph.D. thesis, Ecole Polytechnique.
[10] C. Friedrich, A. Schindlmayr, Many-body perturbation theory: The GW approximation, Condensed Matter Physics, Matter and Materials 32.
[11] D. Golze, M. Dvorak, P. Rinke, The GW compendium: A practical guide to theoretical photoemission spectroscopy, Front. Chem. 7 (2019) 377. doi:10.3389/fchem.2019.00377.
URL https://www.frontiersin.org/article/10.3389/fchem.2019.00377
[12] S. Yanagisawa, Y. Morikawa, A. Schindlmayr, HOMO band dispersion of crystalline rubrene: Effects of self-energy corrections within the *GW* approximation, Phys. Rev. B 88 (2013) 115438. doi:10.1103/PhysRevB.88.115438.
URL https://link.aps.org/doi/10.1103/PhysRevB.88.115438
[13] M. M. Rieger, L. Steinbeck, I. White, H. Rojas, R. Godby, The GW space-time method for the self-energy of large systems, Comput. Phys. Commun. 117 (3) (1999) 211 – 228. doi:https://doi.org/10.1016/S0010-4655(98)00174-X.
URL http://www.sciencedirect.com/science/article/pii/S001046559800174X
[14] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. White, R. Godby, Enhancements to the GW space-time method, Comput. Phys. Commun. 125 (1) (2000) 105 – 118. doi:https://doi.org/10.1016/S0010-4655(99)00466-X.
URL http://www.sciencedirect.com/science/article/pii/S001046559900466X
[15] D. Neuhauser, Y. Gao, C. Arntsen, C. Karshenas, E. Rabani, R. Baer, Breaking the theoretical scaling limit for predicting quasiparticle energies: The stochastic *GW* approach, Phys. Rev. Lett. 113 (2014) 076402. doi:10.1103/PhysRevLett.113.076402.
URL https://link.aps.org/doi/10.1103/PhysRevLett.113.076402
[16] V. Vlcek, E. Rabani, D. Neuhauser, R. Baer, Stochastic GW calculations for molecules, J. Chem. Theory Comput. 13 (10) (2017) 4997–5003. doi:10.1021/acs.jctc.7b00770.
URL https://doi.org/10.1021/acs.jctc.7b00770
[17] 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.
URL https://link.aps.org/doi/10.1103/PhysRevB.87.155148
[18] M. Govoni, G. Galli, Large scale GW calculations, J. Chem. Theory Comput. 11 (6) (2015) 2680–2696. doi:10.1021/ct500958p.
URL https://doi.org/10.1021/ct500958p
[19] P. Umari, G. Stenuit, S. Baroni, GW quasiparticle spectra from occupied states only, Phys. Rev. B 81 (2010) 115104. doi:10.1103/PhysRevB.81.115104.
URL https://link.aps.org/doi/10.1103/PhysRevB.81.115104
[20] D. Foerster, P. Koval, D. Sanchez-Portal, An O(N3) implementation of hedin's GW approximation for molecules, J. Chem. Phys. 135 (7) (2011) 074105. doi:10.1063/1.3624731.
URL https://doi.org/10.1063/1.3624731
[21] J. Wilhelm, D. Golze, L. Talirz, J. Hutter, C. A. Pignedoli, Toward GW calculations on thousands of atoms, J. Phys. Chem. Lett. 9 (2) (2018) 306–312. doi:10.1021/acs.jpclett.7b02740.
URL https://doi.org/10.1021/acs.jpclett.7b02740
[22] F. Bruneval, T. Rangel, S. M. Hamed, M. Shao, C. Yang, J. B. Neaton, molgw 1: Many-body perturbation theory software for atoms, molecules, and clusters, Comput. Phys. Commun. 208 (2016) 149 – 161. doi:https://doi.org/10.1016/j.cpc.2016.06.019.

URL http://www.sciencedirect.com/science/article/pii/S0010465516301990
[23] M. J. van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, P. Rinke, GW100: Benchmarking G0W0 for molecular systems, J. Chem. Theo. Comput. 11 (12) (2015) 5665–5687. doi:10.1021/acs.jctc.5b00453.
URL https://doi.org/10.1021/acs.jctc.5b00453
[24] M. Kaltak, J. c. v. Klimeš, G. Kresse, Cubic scaling algorithm for the random phase approximation: Self-interstitials and vacancies in Si, Phys. Rev. B 90 (2014) 054115. doi:10.1103/PhysRevB.90.054115.
URL https://link.aps.org/doi/10.1103/PhysRevB.90.054115
[25] P. Liu, M. Kaltak, J. c. v. Klimeš, G. Kresse, Cubic scaling GW: Towards fast quasiparticle calculations, Phys. Rev. B 94 (2016) 165109. doi:10.1103/PhysRevB.94.165109.
URL https://link.aps.org/doi/10.1103/PhysRevB.94.165109
[26] R. Martin, Electronic structure: basic theory and practical methods, 2008.
[27] R. Martin, L. Reining, D. Ceperley, Interacting electrons, Theory and Computational Approaches, 2016.
[28] X. Gonze, B. Amadon, G. Antonius, F. Arnardi, L. Baguet, J.-M. Beuken, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, N. Brouwer, F. Bruneval, G. Brunin, T. Cavignac, J.-B. Charraud, W. Chen, M. Coté, S. Cottenier, J. Denier, G. Geneste, P. Ghosez, M. Giantomassi, Y. Gillet, O. Gingras, D. R. Hamann, G. Hautier, X. He, N. Helbig, N. Holzwarth, Y. Jia, F. Jollet, W. Lafargue-Dit-Hauret, K. Lejaeghere, M. A. Marques, A. Martin, C. Martins, H. P. Miranda, F. Naccarato, K. Persson, G. Petretto, V. Planes, Y. Pouillon, S. Prokhorenko, F. Ricci, G.-M. Rignanese, A. H. Romero, M. M. Schmitt, M. Torrent, M. J. van Setten, B. van Troeye, M. J. Verstraete, G. Zérah, J. W. Zwanziger, The abinitproject: Impact, environment and recent developments, Comput. Phys. Commun. 248 (2020) 107042. doi:https://doi.org/10.1016/j.cpc.2019.107042.
URL http://www.sciencedirect.com/science/article/pii/S0010465519303741
[29] S. L. Adler, Quantum theory of the dielectric constant in real solids, Phys. Rev. 126 (1962) 413–420. doi:10.1103/PhysRev.126.413.
URL https://link.aps.org/doi/10.1103/PhysRev.126.413
[30] H. J. Vidberg, J. W. Serene, Solving the eliashberg equations by means of n-point padé approximants, J. Low. Temp. Phys. 29 (1977) 179–192. doi:10.1007/BF00655090.
[31] F. Bruneval, 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.
URL https://link.aps.org/doi/10.1103/PhysRevB.78.085125
[32] G. Weiwei, X. Weiyi, G. Xiang, Z. Peihong, Speeding up GW calculations to meet the challenge of large scale quasiparticle predictions, Sci. Rep. 6 (2016) 36849. doi:10.1038/srep36849.
[33] P. Liu, M. Kaltak, J. c. v. Klimeš, G. Kresse, Cubic scaling GW: Towards fast quasiparticle calculations, Phys. Rev. B 94 (2016) 165109. doi:10.1103/PhysRevB.94.165109.
URL https://link.aps.org/doi/10.1103/PhysRevB.94.165109
[34] A. Schindlmayr, Decay properties of the one-particle Green function in real space and imaginary time, Phys. Rev. B 62 (2000) 12573–12576. doi:10.1103/PhysRevB.62.12573.
URL https://link.aps.org/doi/10.1103/PhysRevB.62.12573
[35] G. S. G. Michael S Shur, Handbook Series on Semiconductor Parameters - Volume 1: Si, Ge, C (Diamond), Gaas, Gap, Gasb, Inas, Inp, Insb (1999).
URL http://www.ioffe.ru/SVA/NSM/Semicond/
[36] R. K. McMullan, Å. Kvick, P. Popelier, Structures of cubic and orthorhombic phases of acetylene by single-crystal neutron diffraction, Acta Cryst. Section B 48 (5) (1992) 726–731. doi:10.1107/S0108768192004774.
URL https://doi.org/10.1107/S0108768192004774
[37] C. H. Patterson, Excited states of molecular and crystalline acetylene: application of TDHF and BSE via density fitting methods, Mol. Phys. 0 (0) (2020) 1–9. doi:10.1080/00268976.2020.1792568.
URL https://doi.org/10.1080/00268976.2020.1792568

[38] F. Aryasetiawan, S. Biermann, Generalized hedin's equations for quantum many-body systems with spin-dependent interactions, Phys. Rev. Lett. 100 (2008) 116402. doi:10.1103/PhysRevLett.100.116402.
URL https://link.aps.org/doi/10.1103/PhysRevLett.100.116402

[39] K. S. Thygesen, A. Rubio, Nonequilibrium GW approach to quantum transport in nano-scale contacts, J. Chem. Phys. 126 (9) (2007) 091101. doi:10.1063/1.2565690.
URL https://doi.org/10.1063/1.2565690

[40] 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, SiC, AlAs, InAs, NaH, and KH, Phys. Rev. B 67 (2003) 155208. doi:10.1103/PhysRevB.67.155208.
URL https://link.aps.org/doi/10.1103/PhysRevB.67.155208

[41] M. Shishkin, G. Kresse, Implementation and performance of the frequency-dependent *GW* method within the paw framework, Phys. Rev. B 74 (2006) 035101. doi:10.1103/PhysRevB.74.035101.
URL https://link.aps.org/doi/10.1103/PhysRevB.74.035101