GW calculations using the Lanczos method and density-functional perturbation theory

Feliciano Giustino, UCB July 2008

The electron self-energy due to the electron-electron interaction in the GW approximation reads [Eq. (10) of Ref. [1]]:

$$\Sigma(\mathbf{r}, \mathbf{r}'; \omega) = \frac{i}{2\pi} \int_{-\infty}^{+\infty} d\omega' e^{-i\delta\omega'} \, G(\mathbf{r}, \mathbf{r}', \omega - \omega') \, W(\mathbf{r}, \mathbf{r}', \omega'), \tag{1}$$

where δ is a positive infinitesimal. The calculation of the screened Coulomb interaction requires a sum over empty states through the static dielectric matrix $\epsilon_{\mathbf{GG}'}(\mathbf{q}, \omega = 0)$, cf. Eqs. (23), (25) of Ref. [1]. The calculation of the noninteracting Green's function requires a sum over empty states, Eq. (15) of Ref. [1]. This latter sum only appears in the Coulomb hole term in Eq. (34b) Ref. [1]. The purpose of the present notes is to discuss the feasibility of GW calculations without resorting to the expansion over empty states. The main idea is that the dielectric response can in principle be calculated using a self-consistent Sternheimer equation, as it is commonly done in linear-response calculations of phonons in the spirit of Ref. [2].

Self-consistent linear-response calculation of the screened Coulomb interaction

In this section we describe the main idea by writing down a few equations. We consider the static screened Coulomb interaction ($\omega = 0$), and we temporarily ignore the translational invariance of the crystal lattice ($\mathbf{q} = 0$) [Eq. (20) of Ref. [1]]:

$$W(\mathbf{r}, \mathbf{r}'; \omega = 0) = \int d\mathbf{r}' \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; \omega = 0) v(\mathbf{r}'', \mathbf{r}'). \tag{2}$$

We note that the previous expression assumes a linear response for the dielectric screening. We now rewrite this expression by introducing a Dirac delta:

$$W(\mathbf{r}, \mathbf{r}'; \omega = 0) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; \omega = 0) \underbrace{\int d\mathbf{r}''' v(\mathbf{r}'', \mathbf{r}''') \underbrace{\delta(\mathbf{r}''', \mathbf{r}')}_{\text{pert. charge}}}_{\text{bare perturbation}} . \tag{3}$$

$$\underbrace{\phantom{W(\mathbf{r}, \mathbf{r}'; \omega = 0) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; \omega = 0) \int d\mathbf{r}''' v(\mathbf{r}'', \mathbf{r}''') \delta(\mathbf{r}''', \mathbf{r}')}}_{\text{screened perturbation}}$$

In order to stress the analogy with the linear-response methods for phonons, we rewrite the bare perturbation and the screened perturbation as parametric functions of the real-space coordinate $\mathbf{r}'$ as follows:

$$\Delta V_{\mathrm{b}}^{(\mathbf{r}')}(\mathbf{r}) = \int d\mathbf{r}''' v(\mathbf{r}, \mathbf{r}''') \delta(\mathbf{r}''', \mathbf{r}'), \tag{4}$$

$$\Delta V_{\mathrm{s}}^{(\mathbf{r}')}(\mathbf{r}) = W(\mathbf{r}, \mathbf{r}'; \omega = 0). \tag{5}$$

At this point we can rewrite Eq. (3):

$$\Delta V_{\mathrm{s}}^{(\mathbf{r}')}(\mathbf{r}) = \int d\mathbf{r}'' \epsilon_0^{-1}(\mathbf{r}, \mathbf{r}'') \Delta V_{\mathrm{b}}^{(\mathbf{r}')}(\mathbf{r}''), \tag{6}$$

where ϵ_0^{-1} indicates the static dielectric matrix.

In order to calculate *precisely* the same expression for the screened Coulomb interaction as given in Ref. [1], we derive a Dyson-like equation for W using Eqs. (20)-(26) of that work.

From Eq. (20) of Ref. [1] (omitting the frequency for simplicity):

$$W(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}'') v(\mathbf{r}'', \mathbf{r}'). \tag{7}$$

The explicit expression for $\epsilon(\mathbf{r}, \mathbf{r}')$ (not its inverse!) in real space can be taken from Eq. (16) of Ref. [15]:

$$\epsilon(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r}, \mathbf{r}') - \int d\mathbf{r}'' v(\mathbf{r}, \mathbf{r}'') \chi_0(\mathbf{r}'', \mathbf{r}'), \tag{8}$$

which corresponds to Eq. (23) of Ref. [1]. Now we take the susceptibility χ_0 from Eq. (166) of Ref. [2]:

$$\chi_0(\mathbf{r}, \mathbf{r}') = \sum_{n,m} \frac{f_n - f_m}{\epsilon_n - \epsilon_m} \psi_n^\star(\mathbf{r}) \psi_m(\mathbf{r}) \psi_m^\star(\mathbf{r}') \psi_n(\mathbf{r}'). \tag{9}$$

It is precisely the last equation which will allow us to rewrite the screened Coulomb interaction in terms of a self-consistent linear-response problem. For this we need to get rid of ϵ^{-1} in Eq. (7). We first multiply both sides of this equation by ϵ, integrate, and take into account that $\epsilon \epsilon^{-1} = \delta$:

$$v(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}'' \epsilon(\mathbf{r}, \mathbf{r}'') W(\mathbf{r}'', \mathbf{r}'). \tag{10}$$

Now we use Eq. (8) in the last equation and rearrange:

$$W(\mathbf{r}, \mathbf{r}') = v(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}''' v(\mathbf{r}, \mathbf{r}''') \int d\mathbf{r}'' \chi_0(\mathbf{r}''', \mathbf{r}'') W(\mathbf{r}'', \mathbf{r}'). \tag{11}$$

This equation corresponds to Eq. (5a) of Ref. [1], i.**e**. it is one of Hedin's equations. In the following χ_0 will be taken at the RPA level, i.**e**. we will make use of Eq. (7) of Ref. [1] instead of Eq. (5b).

At this point we replace Eq. (9) in the last equation:

$$W(\mathbf{r}, \mathbf{r}') = v(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}''' v(\mathbf{r}, \mathbf{r}''') \int d\mathbf{r}'' \sum_{n,m} \frac{f_n - f_m}{\epsilon_n - \epsilon_m} \psi_n^\star(\mathbf{r}''') \psi_m(\mathbf{r}''') \psi_m^\star(\mathbf{r}'') \psi_n(\mathbf{r}'') W(\mathbf{r}'', \mathbf{r}').$$
$$\tag{12}$$

After some manipulations we obtain:

$$\Delta V_{\mathrm{s}}^{(\mathbf{r}')}(\mathbf{r}) = v(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}''' v(\mathbf{r}, \mathbf{r}''') \sum_n \psi_n^\star(\mathbf{r}''') \sum_m \frac{f_n - f_m}{\epsilon_n - \epsilon_m} \langle m | \Delta V_{\mathrm{s}}^{(\mathbf{r}')} | n \rangle \psi_m(\mathbf{r}'''), \tag{13}$$

where $\langle m|\Delta V_{\mathrm{s}}^{(\mathbf{r}')}|n\rangle = \int d\mathbf{r}'' \psi_m^\star(\mathbf{r}'') W(\mathbf{r}'', \mathbf{r}') \psi_n(\mathbf{r}'')$ and we have used the notation of Eq. (4) for the screened Coulomb interaction. Now we note that the sum connects only occupied to empty states and every (m, n) pair appears twice. Therefore we can restrict n to the occupied manifold, let m span all states, and include a factor 2 (this is not the spin factor!):

$$\Delta V_{\mathrm{s}}^{(\mathbf{r}')}(\mathbf{r}) = v(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}''' v(\mathbf{r}, \mathbf{r}''') \, 2 \underbrace{\sum_{n \in \mathrm{occ}} \psi_n^\star(\mathbf{r}''') \underbrace{\sum_m \frac{f_n - f_m}{\epsilon_n - \epsilon_m} \langle m|\Delta V_{\mathrm{s}}^{(\mathbf{r}')}|n\rangle \psi_m(\mathbf{r}''')}_{\Delta\psi_n(\mathbf{r}''')}}_{\Delta n(\mathbf{r}''')[\Delta V_{\mathrm{s}}^{(\mathbf{r}')}]}. \quad (14)$$

The main step at this stage is to realize that $\Delta\psi_n(\mathbf{r})$ defined by

$$\Delta\psi_n(\mathbf{r}) = \sum_m \frac{f_n - f_m}{\epsilon_n - \epsilon_m} \langle m|\Delta V_{\mathrm{s}}^{(\mathbf{r}')}|n\rangle \psi_m(\mathbf{r}), \quad (15)$$

is the change of the wavefunction $\psi_n(\mathbf{r})$ induced by the *screened* perturbation $\Delta V_{\mathrm{s}}^{(\mathbf{r}')}$. After some manipulations, it can be shown that $\Delta\psi_n(\mathbf{r})$ is the solution of the following linear system (*self-consistent Sternheimer equation*):

$$(H_{\mathrm{SCF}} - \epsilon_n)|\Delta\psi_n\rangle = -\Delta V_{\mathrm{s}}|\psi_n\rangle. \quad (16)$$

As a consequence, the quantity $\Delta n(\mathbf{r})$ defined by

$$\Delta n(\mathbf{r}) = 2 \sum_{n \in \mathrm{occ}} \psi_n^\star(\mathbf{r})\Delta\psi_n(\mathbf{r}) \quad (17)$$

is the variation of the electronic charge density induced by the screened perturbation ΔV_{s}. Using again the notation of Eq. (4) for the Coulomb interaction, we can now rewrite Eq. (14) as follows:

$$\Delta V_{\mathrm{s}}(\mathbf{r}) = \underbrace{\Delta V_{\mathrm{b}}(\mathbf{r})}_{\text{bare perturbation}} + \underbrace{\int d\mathbf{r}' \frac{\Delta n(\mathbf{r}')[\Delta V_{\mathrm{s}}]}{|\mathbf{r} - \mathbf{r}'|}}_{\text{Hartree screening}}. \quad (18)$$

This is similar to Eq. (27) of Ref. [2] used to calculate lattice-dynamical properties. In that case the bare perturbation is the change of the ionic potential corresponding to some displacement pattern. In addition, here we do not include the exchange and correlation contribution as it is done in the case of phonons. The neglect of the XC contribution corresponds to the RPA approximation.

The procedure to obtain the screened Coulomb interaction can be summarized as follows:

1. start with a delta function charge distribution centered in $\mathbf{r}_0$,

2. calculate the corresponding Hartree potential - this is going to be the bare perturbation.

3. calculate self-consistently the screened perturbation using Eq. (25) of Ref. [2], *without* the XC term.

3

4 This is going to be $W(\mathbf{r}, \mathbf{r}_0)$.

5 Repeat the procedure for all $\mathbf{r}_0$.

NOTE 1: we certainly need to use the Kohn-Sham Hamiltonian, but when we make the update of the bare perturbation in the SCF procedure we do not include the XC term. This is the only way to obtain the RPA response.

NOTE 2: The bare perturbation discussed so far is a Delta function. This choice is motivated by some localization requirements for the Haydock method used in the calculation of the Green's function. At this stage I am not certain that this is really necessary. Certainly the screened Coulomb interaction can be calculated by setting plane-wave like bare perturbations. The use of localized perturbations could be convenient to accelerate the self-consistent solution of the linear system, since the first iteration could be performed with the screened perturbation corresponding to a neighboring Delta function... This is something I need to check.

0.0.1 Vertex correction

If we include the XC term in the SCF response, do we end up with the approximate vertex correction of Ref. [1] ? I think the answer is yes, we do. Let try to look at the derivation quickly. We begin from the Sternheimer equation:

$$(H_{\mathrm{SCF}} - \epsilon_n)|\Delta\psi_n\rangle = -\Delta V_{\mathrm{SCF}}|\psi_n\rangle. \tag{19}$$

This is precisely the equation solved for the calculation of the dynamical matrix. Note the ΔV_{SCF} on the right-end side - we need a self-consistent loop for this. The variation of the SCF potential is given by the bare potential plus the Hartree potential plus the XC potential, as in Eq. (27) of Ref. [2]:

$$\Delta V_{\mathrm{SCF}} = \Delta V_{\mathrm{b}} + v\Delta n[\Delta V_{\mathrm{SCF}}] + K_{\mathrm{xc}}\Delta n[\Delta V_{\mathrm{SCF}}]. \tag{20}$$

Here I used the notation $K_{\mathrm{xc}} = \delta V_{\mathrm{xc}}/\delta n$ borrowed from Ref. [1]. Using Eqs. (49) and (56) we can write the induced charge density as follows:

$$\Delta n[\Delta V_{\mathrm{SCF}}] = \chi_0 \Delta V_{\mathrm{SCF}}. \tag{21}$$

By combining the previous two equations we obtain:

$$\Delta n[\Delta V_{\mathrm{SCF}}] = [1 - \chi_0(v + K_{\mathrm{xc}})]^{-1}\chi_0\Delta V_{\mathrm{b}}. \tag{22}$$

Let now construct the total electrostatic potential seen by a test charge (i.e. the bare potential plus the Hartree screening):

$$\Delta V_{\mathrm{s}} = \Delta V_{\mathrm{b}} + v\Delta n[\Delta V_{\mathrm{SCF}}]. \tag{23}$$

By combining the two previous equations we obtain:

$$\Delta V_{\mathrm{s}} = \{1 + v[1 - \chi_0(v + K_{\mathrm{xc}})]^{-1}\chi_0\}\Delta V_{\mathrm{b}}. \tag{24}$$

Using $\Delta V_{\rm b} = v$ and $\Delta V_{\rm s} = W$ we have

$$W = v\{1 + v[1 - \chi_0(v + K_{\rm xc})]^{-1}\chi_0\}, \tag{25}$$

which is the screened Coulomb interaction in the "$GW + K_{\rm xc}$" approximation introduced in Ref. [1] and investigated further in Ref. [16] (see pag. 8026, above Table I).

To summarize, the procedure to obtain the approximate vertex correction in W is the following:

1 start with a delta function charge distribution centered in $\mathbf{r}_0$,

2 calculate the corresponding Hartree potential - this is going to be the bare perturbation.

3 calculate self-consistently the screened perturbation using Eq. (25) of Ref. [2], *including* the XC term.

4 This is going to be $W(\mathbf{r}, \mathbf{r}_0)$.

5 Repeat the procedure for all $\mathbf{r}_0$.

In conclusion, the difference between the RPA and the RPA+XC calculation consists in the way we update the variation of the self-consistent potential: only through the Hartree term in the former case, and including the XC term in the latter. In both cases, the final W is constructed only using the Hartree term (electrostatic potential for a test charge). Both approaches can be tested quite easily since the entire infrastructure is exactly the same as the one used in the calculation of phonons using density-functional perturbation theory (the XC term can be included or neglected by commenting one line of code...).

Plasmon-pole approximation

The plasmon-pole model of Ref. [1] is a way to take into account the frequency dependence of the screened Coulomb interaction $W(\mathbf{r}, \mathbf{r}', \omega)$ based on the calculated static interaction $W(\mathbf{r}, \mathbf{r}', \omega = 0)$, a sum rule relating the dielectric function to the charge density, and the Kramers-Krönig relations.

Technically, the plasmon-pole model consists in assuming the following $\omega-$dependence for the inverse dielectric matrix [Eq. (26) of Ref. [1]]:

$$\mathrm{Im}\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}, \omega) = A_{\mathbf{G}\mathbf{G}'}(\mathbf{q})\big\{\delta[\omega - \tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})] - \delta[\omega + \tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})]\big\}. \tag{26}$$

For every planewave component, the free parameters are the peak frequency and strength. These parameters are determined by using the dielectric function at $\omega = 0$ and a generalized $f-$sum rule. The real part of $\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}, \omega)$ is then obtained by causality. This model is based on the observation that "realistic calculations of the response function show that $\mathrm{Im}\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}, \omega)$ is generally a peaked function in ω. [...] For cases where there is not a single well-defined peak, the amplitude tends to be small and fluctuates in sign" (Ref. [1]).

Willing to rewrite a plasmon-pole model for the screened Coulomb interaction, we realize that the various planewave components become mixed in $W(\mathbf{r}, \mathbf{r}', \omega)$, therefore the "single-peak" approximation no longer applies. Furthermore, Steven made the observation that the plasmon is a long-wavelenght excitation, therefore a "local" plasmon-pole approximation on $(\mathbf{r}, \mathbf{r}')$ it is not physically motivated and will not work.

This is to say that the only way to implement the plasmon-pole model seems to use exactly the same prescription as in Ref. [1]:

1. Calculate $W(\mathbf{r}, \mathbf{r}', \mathbf{q}, \omega = 0)$,

2. Fourier-transform to $W(\mathbf{G}, \mathbf{G}', \mathbf{q}, \omega = 0)$,

3. Obtain $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega = 0)$ through $W(\mathbf{G}, \mathbf{G}', \mathbf{q}, \omega = 0)/v(\mathbf{q} + \mathbf{G}')$,

4. Apply the plasmon-pole model using Eqs. (26)-(32) of Ref. [1],

5. Obtain $W(\mathbf{G}, \mathbf{G}', \mathbf{q}, \omega)/v(\mathbf{q} + \mathbf{G}')$ through $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega)v(\mathbf{q} + \mathbf{G}')$,

6. Fourier-transform to $W(\mathbf{r}, \mathbf{r}', \mathbf{q}, \omega)$.

In principle the Sternheimer calculation of the screened Coulomb interaction can be conducted directly in Fourier space (the localized basis is only needed for the recursive calculation of the Green's function), therefore we could in principle replace steps [1.] and [2.] by

1.-2. Calculate $W(\mathbf{G}, \mathbf{G}', \mathbf{q}, \omega = 0)$.

An important alternative to this expensive procedure is to avoid the plasmon-pole approximation and evaluate the frequency-dependence directly by modifying the Sternheimer equation as discussed at the end of the corresponding paragraph (cf. above). The frequency-dependent RPA suscepibility is [Eq. (14) of Ref. [15]]

$$\chi_0(\mathbf{r}, \mathbf{r}'; \omega) = \sum_{n,m} \left[\frac{f_n - f_m}{\epsilon_n - \epsilon_m + \omega + i\delta} + \frac{f_n - f_m}{\epsilon_n - \epsilon_m - \omega + i\delta} \right] \psi_n^\star(\mathbf{r})\psi_m(\mathbf{r})\psi_m^\star(\mathbf{r}')\psi_n(\mathbf{r}'). \qquad (27)$$

This can be obtained by simply replacing ϵ_n by $\epsilon_n \pm \omega + i\delta$. It should be still possible to perform the calculation using the self-consistent Sternheimer equation. Note that this is only a mathematical tool for obtaining the frequency-dependent RPA response, and does not require any considerations on time-dependent density-functional theorems or else. Indeed, by performing the perturbation expansion of the modified Eq. (16), I seem to get the frequency-dependent Adler-Wiser dielectric function. We note that the frequency grid needed for a direct evaluation of $W(\omega)$ does not need to be uniform: the plasmon-pole aims at describing correctly the first few eV's of the excitation spectrum. Accordingly, one may use a grid which is fine below the plasmon energy and coarse above that (a bit like the logarithmic grids used in the construction of pseudopotentials).

TO UNDERSTAND: The imaginary component $i\delta$ makes the operator in the left-hand side of the Sternheimer equation $(H_{\mathrm{KS}} - \epsilon_n \mp \omega - i\delta)$ non-Hermitian. Does the iterative solution of the self-consistent Sternheimer equation work in this case?

Recursive calculation of the Green's function

We wish to calculate the Green's function $G(\mathbf{r}, \mathbf{r}'; \omega)$ without resorting to the sum over empty states Eq. (15) of Ref. [1]. In Ref. [4] the authors argue that the calculation of the diagonal elements of the Green's function in a localized basis can be performed with a linear scaling using the recursion method by Haydock [6, 7]. The matrix elements are calculated by generating orthogonal states which "propagate" an initial state. If the initial state is localized, the recursion method generates new states which probe the neighborhood of that state, and then farther regions at higher orders.

The determination of the non-diagonal elements of the Green's function in the psinc basis can be done using either 4 chains [6] or 2 chains [8]. In both cases one has the propagation of delocalized objects (size $|\mathbf{r} - \mathbf{r}'|$) therefore it is not clear to me whether the algorithm should converge as fast as for the diagonal case. Maybe a test calculation with a model Hamiltonian would be appropariate. Note that the simplest model, corresponding to free electrons (Sec. 6 of Ref. [6]), is not appropriate in this case since the chain does not terminate [Eqs. (6.11) and (6.12) of the same reference].

Let consider the representation of the Green's function in the p-sinc basis [9] (the motivation for using this basis and the definition will be given below. For now it suffices to say that it is an orthonormal basis and it is localized - a periodic Dirac delta function). We temporarily consider the Γ-only sampling for clarity:

$$G_{mn}(\omega) = \langle D_m(\mathbf{r})|G(\mathbf{r}, \mathbf{r}'; \omega)|D_n(\mathbf{r}')\rangle = \langle D_m|G(\omega)|D_n\rangle, \tag{28}$$

where $D_n(\mathbf{r}) = D(\mathbf{r} - \mathbf{r}_n)$, $D(\mathbf{r})$ is localized around the origin, and $\mathbf{r}_n$ is a real-space grid point (the index n runs over a three-dimensional grid). In Sec. II(17) of Ref. [6] it is shown that such (possibly) non-diagonal matrix element can be rewritten in terms of diagonal matrix elements as follows:

$$\langle D_m|G(\omega)|D_n\rangle = \left[\langle u|G(\omega)|u\rangle - \langle v|G(\omega)|v\rangle\right] - i\left[\langle w|G(\omega)|w\rangle - \langle z|G(\omega)|z\rangle\right], \tag{29}$$

with the definitions:

$$u = \frac{1}{2}(D_m + D_n), \tag{30}$$

$$v = \frac{1}{2}(D_m - D_n), \tag{31}$$

$$w = \frac{1}{2}(D_m + iD_n), \tag{32}$$

$$z = \frac{1}{2}(D_m - iD_n). \tag{33}$$

With respect to Ref. [6] I included a factor $1/2$ in the definitions in order to ensure the normalization of the new functions (since the psinc functions are all normalized to unity). As a consequence, we miss a factor $1/4$ in the right-hand side of Eq. (29). In the following we focus on the first matrix element $G_{u,u}(\omega) = \langle u|G(\omega)|u\rangle$, the others are calculated similarly.

The calculation of each diagonal matrix element requires a Lanczos chain, therefore each nondiagonal matrix element requires 4 chains. In Ref. [8] an alternative procedure has been proposed which only involves 2 chains per element. We will be exploring that possibility only after the preliminary practical implementations.

The diagonal matrix element $G_{u,u}(\omega)$ can be calculated from the continued fraction:

$$G_{u,u}(\omega) = G^{(0)}_{u,u}(\omega), \tag{34}$$

$$G^{(n)}_{u,u}(\omega) = [\omega - a_n - b^2_{n+1}G^{(n+1)}_{u,u}(\omega)]^{-1}, \tag{35}$$

once the sets of coefficients $\{a_n\}$ and $\{b_n\}$ have been determined. Importantly, these coefficients do not depend on the frequency ω, therefore they must be computed once for all. The coefficients are determined from a recursive calculation. Ideally, the chain is terminated when the coefficients no longer change in the recursion procedure. Practically, one has to use "terminators" for the $G^{(N)}_{u,u}(\omega)$ element (e.g. zero) if the chain is truncated at the N-th iteration.

The chain is generated by propagating the state $|u^{(0)}\rangle = |u\rangle$ as follows [Eq. (5.11) of Ref. [6]]:

$$b_n|u^{(n)}\rangle = (H_{\text{KS}} - a_{n-1})|u^{(n-1)}\rangle - b_{n-1}|u^{(n-2)}\rangle. \tag{36}$$

In this expression, $|u^{(n-1)}\rangle$ and $|u^{(n-2)}\rangle$ are known from previous iterations (we start with $|u^{(-1)}\rangle = 0$), while the coefficients on the right-hand side are given by:

$$a_{n-1} = \langle u^{(n-1)}|H_{\text{KS}}|u^{(n-1)}\rangle \tag{37}$$
$$b_{n-1} = \langle u^{(n-1)}|H_{\text{KS}}|u^{(n-2)}\rangle. \tag{38}$$

The coefficient b_n on the left-hand side is determined by requiring $|u^{(n)}\rangle$ to be normalized:

$$|b_n|^2 = \left[(H_{\text{KS}} - a_{n-1})|u^{(n-1)}\rangle - b_{n-1}|u^{(n-2)}\rangle\right]^\dagger\left[(H_{\text{KS}} - a_{n-1})|u^{(n-1)}\rangle - b_{n-1}|u^{(n-2)}\rangle\right]. \tag{39}$$

In Ref. [6] the authors claim that b^2_n is real and positive and that it is convenient to choose the positive root for b_n. In Ref. [8] the authors claim that it is not guaranteed to have real and positive values of b^2_n and that this does not constitute a problem. Some careful analysis should be done here using some test cases.

Periodic-Sinc basis

For the computation of the Green's function by the recursion method, it is convenient to use a localized basis, since the chain probes first the surroundings of the initial state (faster convergence). In order to choose the basis, we require the following: (i) the new basis should preserve the "direct" product form of the self-energy in terms of G and W Eq. (1). (ii) the new basis should preserve the accuracy of the plane-waves calculation (real-space grids would lead to inaccuracies in the treatment of the kinetic energy). (iii) the new basis should be compatible with existing plane-waves implementations (for ease of implementation).

After some lterature survey, it appears that the periodic-Sinc (psinc) basis introduced in Ref. [9] satisfies all of the above criteria. The psinc function is a periodic Dirac delta function represented using a finite number of G-vectors:

$$D_n(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}_n)}. \tag{40}$$

By expanding a real-space function $f(\mathbf{r})$ in terms of psinc functions we have:

$$f(\mathbf{r}) = \sum_n f_n D_n(\mathbf{r}), \tag{41}$$

where the coefficients f_n in the psinc basis correspond simply to the values of the function $f(\mathbf{r})$ on the grid points where the psinc's are centered:

$$f_n = f(\mathbf{r}_n). \tag{42}$$

This indicates that in order to switch from the G-space to the psinc space and vice-versa, we simply need to perform an ordinary Fourier transform (already implemented as FFT in plane-waves codes).

PROBLEM: the one-to-one correspondece between the psinc basis and the reciprocal space only applies when the box of G-vectors is considered. In practical calculations, out of the G-vectors box we retain only the sphere of vectors smaller than the plane-wave cutoff. Therefore the one-to-one correspondence is lost. On the other hand, willing to keep all of the G-vectors in the box, we may create a strange anisotropy in the system since some directions will have more G-vectors than others...

Factorization and crystal periodicity

Now we need to make sure that, within the psinc representation, the self-energy can still be written as a direct $(\mathbf{r}, \mathbf{r}') \times (\mathbf{r}, \mathbf{r}')$ product. The crystal periodicity will lead to $\mathbf{q}-$dependent quantities.

Let first state Bloch's theorem. The self-energy $\Sigma(\mathbf{r}, \mathbf{r}')$ must be invariant under the crystal translations (I am omitting the frequency dependence for ease of notation). We expand the self-energy in terms of Bloch states (complete basis):

$$\Sigma(\mathbf{r}, \mathbf{r}') = \Sigma_{mn} \Sigma_{\mathbf{k},\mathbf{k}'} \Sigma_{mn}(\mathbf{k}, \mathbf{k}') \psi^{\star}_{m\mathbf{k}'}(\mathbf{r}') \psi_{n\mathbf{k}}(\mathbf{r}). \tag{43}$$

The condition of translationa invariace reads:

$$\Sigma(\mathbf{r} + \mathbf{T}, \mathbf{r}' + \mathbf{T}) = \Sigma(\mathbf{r}, \mathbf{r}'). \tag{44}$$

By combining the previous two equations we obtain:

$$\Sigma_{mn}(\mathbf{k}, \mathbf{k}') = \Sigma_{mn}(\mathbf{k}, \mathbf{k}') e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{T}}, \ \forall \mathbf{T}, \tag{45}$$

which implies:

$$\Sigma_{mn}(\mathbf{k}, \mathbf{k}') = \Sigma_{mn}(\mathbf{k})\delta(\mathbf{k} - \mathbf{k}'). \tag{46}$$

Introducing the Bloch-periodic part of the wavefunctions we can rewrite the self-energy as follows:

$$\Sigma(\mathbf{r}, \mathbf{r}') = \Sigma_{\mathbf{k}} \underbrace{\Sigma_{mn} \Sigma_{mn}^{\mathbf{k}} u_{m\mathbf{k}}^{\star}(\mathbf{r}') u_{n\mathbf{k}}(\mathbf{r})}_{\text{cell-periodic part}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}. \tag{47}$$

At this point the cell-periodic part can be expanded in the psinc basis introduced earlier:

$$\Sigma(\mathbf{r}, \mathbf{r}'; \omega) = \Sigma_{\mathbf{k}} \Sigma_{mn} \sigma_{mn}(\mathbf{k}; \omega) D_m^{\star}(\mathbf{r}') D_n(\mathbf{r}) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}, \tag{48}$$

where $\sigma_{mn}(\mathbf{k}; \omega)$ indicates the components of the self-energy in the psinc-basis. The Green's function and the screened Coulomb interaction can be written in a similar way:

$$G(\mathbf{r}, \mathbf{r}'; \omega) = \Sigma_{\mathbf{k}} \Sigma_{mn} g_{mn}(\mathbf{k}; \omega) D_m^{\star}(\mathbf{r}') D_n(\mathbf{r}) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}, \tag{49}$$

$$W(\mathbf{r}, \mathbf{r}'; \omega) = \Sigma_{\mathbf{k}} \Sigma_{mn} w_{mn}(\mathbf{k}; \omega) D_m^{\star}(\mathbf{r}') D_n(\mathbf{r}) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}. \tag{50}$$

At this stage we wish to rewrite Eq. (1) in the psinc basis. The idea is that the basis function are orthonormal, so we do the usual scalar product on both sides and see what happens. We need the following two generalized orthonormality properties:

$$\int d\mathbf{r} D_m^{\star}(\mathbf{r}) D_n(\mathbf{r}) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} = \delta(\mathbf{k}, \mathbf{k}')\delta_{mn}, \tag{51}$$

$$\int d\mathbf{r} D_m^{\star}(\mathbf{r}) D_n(\mathbf{r}) D_p(\mathbf{r}) e^{i(\mathbf{k}-\mathbf{k}'-\mathbf{k}'')\cdot\mathbf{r}} = \delta(\mathbf{k}, \mathbf{k}' + \mathbf{k}'')\delta_{mn}\delta_{mp}, \tag{52}$$

which can be obtained starting from the definition of the psinc functions through a direct calculation. Using these two properties we find:

$$\sigma_{mn}(\mathbf{k}, \omega) = \frac{i}{2\pi} \int_{-\infty}^{+\infty} d\omega' \int d\mathbf{q} \, e^{-i\delta\omega'} g_{mn}(\mathbf{k} - \mathbf{q}, \omega - \omega') \, w_{mn}(\mathbf{q}, \omega'). \tag{53}$$

The last equation indicates that the direct product form of the self-energy in terms of $(\mathbf{r}, \mathbf{r}')$ is maintained when transforming to the psinc basis (this is somewhat obvious since this is a basis of Dirac deltas on the real-space grid points).

q-dependent Haydock

The recursive calculation of the Green's function has been discussed for the $\mathbf{q} = 0$ case, i.e. for $G(\mathbf{r}, \mathbf{r}', \omega)$. In the general case, we need to consider that the Green's function is [Eq. (15) of Ref. [1]]:

$$G(\mathbf{r}, \mathbf{r}', \omega) = \sum_{n\mathbf{k}} \frac{\psi_{n\mathbf{k}}^{\star}(\mathbf{r}')\psi_{n\mathbf{k}}(\mathbf{r})}{\omega - \epsilon_{n\mathbf{k}}}, \tag{54}$$

where I omitted the imaginary infinitesimal for clarity (we will not need it in the following). This Green's function is the resolvent of the Kohn-Sham Hamiltonian H_{KS}. We can also write:

$$G(\mathbf{r}, \mathbf{r}'; \omega) = \sum_{\mathbf{k}} G(\mathbf{r}, \mathbf{r}'; \mathbf{k}, \omega) e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')}, \tag{55}$$

with the definition:

$$G(\mathbf{r}, \mathbf{r}'; \mathbf{q}, \omega) = \sum_{n} \frac{u_{n\mathbf{k}}^{\star}(\mathbf{r}') u_{n\mathbf{k}}(\mathbf{r})}{\omega - \epsilon_{n\mathbf{k}}}. \tag{56}$$

It can easily be verifyied that (as expected) $G(\mathbf{r}, \mathbf{r}'; \mathbf{q}, \omega)$ is the resolvent of the $\mathbf{k}-$dependent Hamiltonian:

$$H_{\mathbf{k}} = H_{KS} + \frac{k^2}{2} + \mathbf{k} \cdot \frac{1}{i} \nabla. \tag{57}$$

Therefore, in order to calculate $G(\mathbf{r}, \mathbf{r}'; \mathbf{q}, \omega)$ we need to apply the Haydock recursion to the new Hamiltonian $H_{\mathbf{k}}$.

OLD QUESTIONS: (i) Does the recursion method work for this modified Hamiltonian? (ii) The propagation of a periodic delta-function is compatible with the recursion method or we need a truly localized function?

ANSWER: It seems that the method converges for both the diagonal and the off-diagonal matrix elements. The requirement of localization does not seem to be necessary (although I am not certain at this stage). The difference between Haydock's calculations and the present ones is that we are solving for the Bloch-periodic wavefunctions, therefore we are dealing with a *discrete* spectrum. This is very different from the situation originally considered by Haydock.

Test calculation

It is important to assess the feasibility of the entire procedure using a test calculation. In particular the calculation of the off-diagonal elements of the Green's function using Haydock's recursion has never been attempted to my knowledge.

A good test-case carrying (almost) the full complexity of a real density-functional calculation is the one of a tetrahedral semiconductor within the *empirical pseudopotential model* as given in Ref. [12] for silicon and other semiconductors, and in Ref. [14] for diamond. Probably the diamond case is better than silicon since the gaps are larger. In a second step we may try to fit the form factors to a LDA calculation.

Decomposition of the Coulomb interaction into monochromatic perturbation

In the previous sections we have seen how to separate the problem of calculating the self-energy, the Green's function and the screened Coulomb interaction by using periodic real-space variables and $\mathbf{k}$-vectors. In the case of the screened Coulomb interaction we have:

$$W(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{q}} w_{\mathbf{q}}(\mathbf{r}, \mathbf{r}') \, e^{i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')}, \tag{58}$$

with $w_{\mathbf{q}}(\mathbf{r}, \mathbf{r}')$ cell-periodic. In the same way, the bare Coulomb interaction can be written as

$$v(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{q}} v_{\mathbf{q}}(\mathbf{r}, \mathbf{r}') \, e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')}. \tag{59}$$

Let use $\mathbf{r}' = \mathbf{r}_0$ as a parametric coordinate. The bare perturbation needed to start the SCF linear response calculation is (coefficients not checked)

$$\Delta v_{\mathbf{q}}(\mathbf{G}; \mathbf{r}_0) = v_{\mathbf{q}}(\mathbf{G}, \mathbf{r}_0) = \frac{4\pi e^2}{\Omega} \frac{e^{-i\mathbf{G}\cdot\mathbf{r}_0}}{|\mathbf{q}+\mathbf{G}|^2}. \tag{60}$$

Now, in order to calculate the charge-density response to this perturbation, we need to perform a *sampling* over the Brillouin zone. The SCF equation to be solved is [Eq. (33) of Ref. [2]]:

$$(H_{\mathbf{k}+\mathbf{q}} + \alpha \sum_{m\in\text{occ}} |u_{m,\mathbf{k}+\mathbf{q}}\rangle\langle u_{m,\mathbf{k}+\mathbf{q}}| - \epsilon_{n\mathbf{k}})\Delta u_{n,\mathbf{k}+\mathbf{q}} = -\Big[1 - \sum_{m\in\text{occ}} |u_{m,\mathbf{k}+\mathbf{q}}\rangle\langle u_{m,\mathbf{k}+\mathbf{q}}|\Big]\Delta v_{\text{SCF},\mathbf{q}} u_{n\mathbf{k}}, \tag{61}$$

with n referring to occupied states. The use of the projector on the valence manifold (left) is to lift the singularity of the operator $(H - \epsilon_n)^{-1}$ at $\omega = \epsilon_n$. The induced charge density is [Eq. (35) of Ref. [2]]:

$$\Delta n_{\mathbf{q}}(\mathbf{r}) = \sum_{n\in\text{occ}} u_{n\mathbf{k}}^{\star}(\mathbf{r})\Delta u_{n,\mathbf{k}+\mathbf{q}}(\mathbf{r}). \tag{62}$$

In practice, for every $\mathbf{q}$ we need to sample the whole BZ in $\mathbf{k}$. Since we will need a uniform grid of $\mathbf{q}$ for performing the convolution with the Green's function, it may be convenient to choose uniform and identical $\mathbf{q}$ and $\mathbf{k}$ grids from the very beginning.

Scaling

TO BE DONE

References

[1] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[2] S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).

[3] R. M. Pick, M. H. Cohen, and R. M. Martin, Phys. Rev. B **1**, 910 (1970).

[4] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1992).

[5] R. Haydock, V. Heine, and M. J. Kelly, J. Phys. C **5**, 2845 (1972).

[6] R. Haydock, in *Solid State Physics*, eds. H. Ehrenreich, F. Seitz and D. Turnbull (Academic Press, New York, 1980) Vol. 35, p. 215.

[7] M. J. Kelly, in *Solid State Physics*, eds. H. Ehrenreich, F. Seitz and D. Turnbull (Academic Press, New York, 1980) Vol. 35, p. 295.

[8] L. E. Ballentine and M Kolář, J. Phys. C **19**, 981 (1986).

[9] A. A. Mostofi, C. Skylaris, P. D. Haynes, and M. C. Payne, Comp. Phys. Comm. **147**, 788 (2002).

[10] P. Giannozzi, G. Grosso, S. Moroni, and G. Pastori Parravicini, Appl. Numer. Math. **4**, 273 (1988).

[11] U. Brüstel and K. Unger, Phys. Stat. Sol. B **123**, 229 (1984).

[12] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).

[13] A. Cordelli, G. Grosso, and G. Pastori Parravicini, Phys. Rev. B **38**, 2154 (1988).

[14] W. Saslow, T. K. Bergstresser, and M. L. Cohen, Phys. Rev. Lett. **16**, 354 (1966).

[15] M. L. Tiago and J. R. Chelikowsky, Phys. Rev. B **73**, 205334 (2006).

[16] R. Del Sole, L. Reining, and R. W. Godby, Phys. Rev. B **49**, 8024 (1994).