Electron self-energy calculation using a general multi-pole approximation

2003 J. Phys.: Condens. Matter 15 2573

(http://iopscience.iop.org/0953-8984/15/17/312)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **15** (2003) 2573–2586 PII: S0953-8984(03)59657-6

Electron self-energy calculation using a general multi-pole approximation

J A Soininen[1], J J Rehr[1] and Eric L Shirley[2]

[1] Department of Physics, University of Washington, Seattle, WA 98195, USA
[2] Optical Technology Division, Physics Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Received 14 February 2003
Published 22 April 2003
Online at stacks.iop.org/JPhysCM/15/2573

Abstract

We present a method for calculating the inverse of the dielectric matrix in a solid using a band Lanczos algorithm. The method produces a multi-pole approximation for the inverse dielectric matrix with an arbitrary number of poles. We discuss how this approximation can be used to calculate the screened Coulomb interaction needed for electron self-energy calculations in solids.

1. Introduction

Many of the electron self-energy or quasiparticle calculations done today are based on the GW approximation (GWA) proposed by Hedin [1, 2]. It was first applied in large-scale modelling of the quasiparticle band structure of solids using a plasmon-pole model [3] for the dynamic screening. Later there were works including the full energy dependence of the screened interaction [4]. By comparing the calculated band structure to experimental results (see for example table 1 of [5]), the GWA has been found to give a clear improvement over mean-field theories such as the local-density approximation (LDA) [6]. Also, the relaxation of electrons in solids (see for example [7] and references therein) and their spectral functions have been studied, although for these properties the GWA results are sometimes less than satisfactory. The basis of the GWA and its properties have been given more complete treatment in a number of recent review articles [5, 8, 9].

For solids, the major part of a GWA calculation is the calculation of the inverse of the dielectric matrix ε^{-1} in the screened interaction, $W = \varepsilon^{-1}v$, where v is the Coulomb interaction. Numerous authors have developed plasmon-pole models for ε^{-1} that can be used for self-energy calculations in solids [3, 10–12]. These models are based on a simplified assumption that important contributions to Im W come from plasmon-like excitations. In practice, this approximation has been found to give good results for the band energies, but it has some drawbacks. For example, in the plasmon-pole approximation the quasiparticle only has a finite lifetime if its energy is higher than the plasmon energy. This means that the lifetime of a low-energy quasiparticle cannot be calculated. Also, in situations where the electron–hole

continuum gives considerable contributions to W, it is uncertain how well these simple models work. The cost of computing W can also be reduced by using space-time methods [13, 14] or by using localized basis sets [15, 16]. Another possibility is to calculate the energy integral needed for evaluating the GWA self-energy along the imaginary axis as was done, for example, in [17].

In this paper, we introduce a scheme for calculating ε^{-1} based on an iterative band Lanczos algorithm (BLA). The scheme can be used to make a general multi-pole approximation to the inverse of the dielectric matrix with the infinite-pole limit being the random-phase approximation (RPA). The approximation still retains some of the advantages of the simple plasmon-pole models when it comes to a GWA calculation, while giving one the opportunity to study the effects of electron–hole pair excitation in W on the self-energy. We start by reviewing the basic properties of ε^{-1} and expressing it in a form most useful for our current purpose. Next we discuss the BLA and how it can be used to obtain the pole structure of ε^{-1} and how this pole structure can be used in GWA calculations. Finally we use this scheme to calculate self-energies and quasiparticle properties of Si and GaAs. Although here we used this form of ε^{-1} in a GWA calculation, it can also be used for calculations of the dynamic screening of the electron–hole interaction in GWA-based Bethe–Salpeter calculations.

2. The inverse of the dielectric matrix

In linear-response theory the inverse of the dielectric matrix $\tilde{\varepsilon}^{-1}$ represents the response of the system to an external field. The diagonal of the inverse of this physical (i.e., causal) dielectric matrix ($\tilde{\varepsilon}^{-1}$) can be measured using inelastic x-ray scattering or electron energy-loss spectroscopy. The inverse of the time-ordered dielectric matrix is required (as mentioned above) to calculate quasiparticle properties in the GWA. Additionally, ε^{-1} can be used, for example, in screening the core-hole potential in core spectroscopies [18]. The relationship between the imaginary parts of these two response functions (on the real ω-axis) is $(\omega/|\omega|)\,\mathrm{Im}\,\varepsilon^{-1}(\omega) = \mathrm{Im}\,\tilde{\varepsilon}^{-1}(\omega)$, and their real parts are equal. Here we present a BLA-based method for calculating $\tilde{\varepsilon}^{-1}$, and one can obtain ε^{-1} from $\tilde{\varepsilon}^{-1}$ using the relationship given above.

As the starting point for our discussion, we use the causal RPA independent-particle polarizability which is given in the Adler–Wiser [19, 21] formulation as

$$\chi^0(\boldsymbol{r},\boldsymbol{r}',\omega) = 2\sum_{j,j'}\phi^*_j(\boldsymbol{r})\phi_{j'}(\boldsymbol{r})\phi^*_{j'}(\boldsymbol{r}')\phi_j(\boldsymbol{r}')\left[\frac{f_j(1-f_{j'})}{E_j - E_{j'} + \omega + \mathrm{i}\delta} + \frac{f_{j'}(1-f_j)}{E_{j'} - E_j - \omega - \mathrm{i}\delta}\right], \tag{1}$$

where $\phi_j(\boldsymbol{r})$, E_j, and f_j are single-particle orbitals, energies, and occupation numbers respectively, of a one-particle Hamiltonian as described below. The factor of two in front accounts for the spin degeneracy, and we have used the notation $\delta = 0^+$. Instead of using the real-space representation of χ^0 as given in equation (1), for solids it is often more convenient to use a reciprocal-space representation. Additionally, we represent the system in a particle–hole-excitation basis [20], obtaining

$$\chi^0_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q},\omega) = 2\langle 0|P_{\boldsymbol{G}}(\boldsymbol{q})\frac{1}{z^2 - H^2_{1e}}P^\dagger_{\boldsymbol{G}'}(\boldsymbol{q})|0\rangle, \tag{2}$$

where $|0\rangle$ is a mean-field ground state. Additionally, we have used z as shorthand for $\omega + \mathrm{i}\delta$. The electron–hole pair creation operator $P_{\boldsymbol{G}}(\boldsymbol{q})$ is defined as

$$P_{\boldsymbol{G}}(\boldsymbol{q}) = \sum_{\boldsymbol{j}\boldsymbol{k}}^{occ}\sum_{\boldsymbol{j}'\boldsymbol{k}+\boldsymbol{q}}^{unocc}\langle\phi_{\boldsymbol{j}\boldsymbol{k}}|\mathrm{e}^{-\mathrm{i}(\boldsymbol{q}+\boldsymbol{G})\cdot\boldsymbol{r}}|\phi_{\boldsymbol{j}'\boldsymbol{k}+\boldsymbol{q}}\rangle\Delta^{\boldsymbol{q}}_{\boldsymbol{j}\boldsymbol{j}'\boldsymbol{k}}a^\dagger_{\boldsymbol{j}\boldsymbol{k}}a_{\boldsymbol{j}'\boldsymbol{k}+\boldsymbol{q}}, \tag{3}$$

where we have $\Delta^{q}_{jj'k} = \sqrt{2(E_{j'k+q} - E_{jk})}$. The operator H_{1e} is most easily defined by the way in which it acts on a particle–hole pair ($|jj'\boldsymbol{k}\rangle_q$): $H_{1e}|jj'\boldsymbol{k}\rangle_q = (E_{j'k+q} - E_{jk})|jj'\boldsymbol{k}\rangle_q$ [20]. Time-reversal symmetry, i.e., the fact that for each state ϕ_{jk} there is state ϕ^*_{j-k} with the same eigenvalue, was used to go from equation (1) to (2). Within the RPA the response function χ fulfils a Dyson-type equation,

$$\chi_{GG'}(\boldsymbol{q},\omega) = \chi^0_{GG'}(\boldsymbol{q},\omega) + \sum_{\boldsymbol{K}} \chi^0_{G\boldsymbol{K}}(\boldsymbol{q},\omega) V_{\boldsymbol{K}}(\boldsymbol{q}) \chi_{\boldsymbol{K}G'}(\boldsymbol{q},\omega), \tag{4}$$

where $V_{\boldsymbol{K}}(\boldsymbol{q}) = 4\pi/|\boldsymbol{q}+\boldsymbol{K}|^2$ is the Coulomb potential in reciprocal space. Using equations (4) and (2) it is easy to show that χ can be cast into a form similar to that of χ_0:

$$\chi_{GG'}(\boldsymbol{q},\omega) = 2\langle 0|P_G(\boldsymbol{q})\frac{1}{z^2 - H_{RPA}}P^{\dagger}_{G'}(\boldsymbol{q})|0\rangle, \tag{5}$$

where the operator H_{RPA} is defined as $H_{RPA} = H^2_{1e} + V$ with V including the local-field effects:

$$V = 2\sum_{\boldsymbol{K}} P^{\dagger}_{\boldsymbol{K}}(\boldsymbol{q})|0\rangle V_{\boldsymbol{K}}(\boldsymbol{q})\langle 0|P_{\boldsymbol{K}}(\boldsymbol{q}). \tag{6}$$

It is important to notice that the operator $P_{\boldsymbol{K}}(\boldsymbol{q})$ given in equation (3) has the units of the square root of energy. This gives H_{RPA} the units of energy squared and χ the correct units of one over energy. The inverse of the dielectric function can be expressed in terms of the χ as

$$\tilde{\varepsilon}^{-1}_{GG'}(\boldsymbol{q},\omega) = \delta_{GG'} + V_G(\boldsymbol{q})\chi_{GG'}(\boldsymbol{q},\omega).$$

Calculating the resolvent in equation (5) is equivalent to inverting the dielectric matrix, and so we calculate the $q \to 0$ and $K \to 0$ matrix elements with an approach similar to that of [22]. In many applications, the matrix $\varepsilon^{-1}_{GG'}$ is needed for some set of reciprocal-lattice vectors $\{\boldsymbol{G}_i\}_{i=1,N}$. In this case equation (5) can be written in an electron–hole basis in matrix notation:

$$\boldsymbol{\chi}(\boldsymbol{q},\omega) = \boldsymbol{\Psi}^{\dagger}\frac{1}{z^2 - \boldsymbol{H}_{RPA}}\boldsymbol{\Psi}, \tag{7}$$

where each column of the matrix $\boldsymbol{\Psi}$ corresponds to the total electron–hole pair wavefunction created using $P^{\dagger}_{G}(\boldsymbol{q})$, and each row is associated with a single-particle excitation, specified by j, j', and $\boldsymbol{k}$.

Because we use a pseudopotential approximation for the single-particle states in our calculations, plane waves provide a natural basis for expanding the Coulomb potential and χ. However, equation (7) can also be derived using the mixed-basis approach [23] sometimes used in GWA all-electron work. This simply amounts to replacing $V_{\boldsymbol{K}}(\boldsymbol{q})$ with the expansion of the Coulomb potential in mixed-basis and plane-wave matrix elements in equation (3) with the appropriate matrix elements. Otherwise, one only has to change the indices from being reciprocal-lattice vectors $\{\boldsymbol{G}_i\}$ to mixed-basis indices in what follows.

2.1. The band Lanczos algorithm

Because the size of $\boldsymbol{H}_{RPA}$ is typically of order $10^5 \times 10^5$, its direct diagonalization would be a formidable task. However, with iterative techniques one can obtain approximate eigenvalues and eigenstates of $\boldsymbol{H}_{RPA}$. The band Lanczos [24, 25] algorithm is well suited for matrix problems such as the one given in equation (7). First, an orthonormal set of vectors $\{\boldsymbol{v}_i\}_{i=1,n}$ is created by orthogonalizing $\boldsymbol{\Psi}$ so that one has $\boldsymbol{\Psi} = \boldsymbol{V}\boldsymbol{R}$, where the ith column of $\boldsymbol{V}$ is the vector $\boldsymbol{v}_i$. We use canonical orthogonalizations [26] that can be used to optimize the basis set

(i.e., $n \leqslant N$). As in the simple Lanczos algorithm, a set of orthogonal vectors is generated by iteration:

$$T_{i+n,i}\boldsymbol{v}_{i+n} = \boldsymbol{H}\boldsymbol{v}_i - \sum_{j=i-n}^{i+n-1} T_{j,i}\boldsymbol{v}_j, \tag{8}$$

where we have $T_{j,i} = \boldsymbol{v}_j^\dagger \boldsymbol{H}\boldsymbol{v}_i$. This choice of coefficients $T_{j,i}$ guarantees (in exact arithmetic) that the new vector $\boldsymbol{v}_{i+n}$ is orthogonal to all previous vectors. Additionally, $\boldsymbol{v}_{i+n}$ is normalized. The matrix $\boldsymbol{T}$ is Hermitian, banded with a bandwidth of $2n + 1$, and we have $T_{i,j} = 0$ for $j \leqslant 0$. In our implementation we do not store all the generated vectors $\boldsymbol{v}_i$, but only the vectors that are needed for the next iteration step. The storage requirements of the method as we have implemented it are $((2n + 1) + N_G^{local})N_{p-h}$, where N_G^{local} is the number of reciprocal-lattice vectors included in local-field effects and N_{p-h} is the number of electron–hole pairs. If the number of reciprocal-lattice vectors N in $\boldsymbol{\chi}$ is equal to N_G^{local}, the orthogonality of the new Lanczos vector to all previous vectors can be used to reduce the computational cost of the iteration. This is related to the fact that, in this case, $\boldsymbol{\Psi}$ spans the same space in the particle–hole basis as the one within which V operates. Detailed analysis of this property of the BLA is given in the appendix. In this work, we choose N_G^{local} to be almost twice as large as N, and so we do not use this property in the current approach.

After some finite number (I) of iterations, the matrix $\boldsymbol{T}$ and the generated set of vectors give an approximation for the full matrix:

$$\boldsymbol{H}^{(I)} = \boldsymbol{V}^{(I)}\boldsymbol{T}\boldsymbol{V}^{(I)\dagger}.$$

Using decomposition of $\boldsymbol{T}$ in terms of eigenvalues $\{\Omega_i\}$ and eigenvectors $\{\boldsymbol{X}_i\}$ we can approximate equation (7) as

$$\boldsymbol{\chi}(\boldsymbol{q},\omega) = \boldsymbol{S}^\dagger \frac{1}{z^2 - \Omega(\boldsymbol{q})}\boldsymbol{S}.$$

We have used $\boldsymbol{S} = \boldsymbol{X}^\dagger \boldsymbol{R}$, where only the first n components of the eigenvectors $\{\boldsymbol{X}_i\}$ are included in the product. Besides the fact that the band Lanczos method makes it possible to approximate the eigenvalues and eigenvectors of $\boldsymbol{H}_{RPA}$ with a smaller band matrix $\boldsymbol{T}$, it also has other important properties. Assuming exact arithmetic and completeness of the particle–hole space, the first $2I/n - 2$ [25] frequency moments of different components of $\boldsymbol{\chi}(\boldsymbol{q},\omega)$ are correctly reproduced. Additionally, the eigenvalues Ω_i only depend on $\boldsymbol{q}$. This simplifies the use of $\boldsymbol{\chi}(\boldsymbol{q},\omega)$ in an electron self-energy calculation, as we will demonstrate in the next section.

In principle, the band Lanczos method used here can be used to reproduce all the eigenvalues and eigenvectors of $\boldsymbol{H}_{RPA}$. In practice, however, as the iteration proceeds the vectors $\{\boldsymbol{v}_i\}$ start to lose their orthogonality and multiple copies of already converged eigenvalues might be produced. The remedy for this is, for example, re-orthonormalization of the Lanczos vectors, which would be, at least in our case, extremely expensive computationally. Also, for our application the global features of the eigenspectrum (such as the moments of $\boldsymbol{\chi}(\boldsymbol{q},\omega)$) are more important than individual eigenstates. This is why we use the simple algorithm of equation (8), which is well suited for such problems [25].

3. Self-energy calculation

In the GWA [1, 2], the self-energy is given by

$$\Sigma(\boldsymbol{r},\boldsymbol{r}',\omega) = \mathrm{i}\int \frac{\mathrm{d}\omega'}{2\pi}\,\mathrm{e}^{\mathrm{i}\delta\omega'} G(\boldsymbol{r},\boldsymbol{r}',\omega+\omega')W(\boldsymbol{r},\boldsymbol{r}',\omega'), \tag{9}$$

where again we have used the notation $\delta = 0^+$. We assume that the Green function can be represented in terms of the quasiparticle band states $\phi_{nk}(\boldsymbol{r})$ and energies E_{nk},

$$G(\boldsymbol{r}, \boldsymbol{r}', \omega) = \sum_{n,k} \frac{\phi_{nk}(\boldsymbol{r})\phi^*_{nk}(\boldsymbol{r}')}{\omega - E_{nk} \pm \mathrm{i}\delta}, \tag{10}$$

where the sign of the imaginary part is negative (positive) for occupied (unoccupied) states.

In crystalline solids, the Fourier components of the screened interaction are given by

$$W_{GG'}(\boldsymbol{q}, \omega) = \varepsilon^{-1}_{GG'}(\boldsymbol{q}, \omega) V_{G'}(\boldsymbol{q}).$$

Using the general multi-pole approximation for the polarizability presented in the previous section, we can write the energy-dependent part of W as

$$\tilde{W}_{GG'}(\boldsymbol{q}, \omega) = \sum_p w_{p,q}(\boldsymbol{G}) w^*_{p,q}(\boldsymbol{G}') \left[\frac{1}{\omega - \omega_p(\boldsymbol{q}) + \mathrm{i}\delta} - \frac{1}{\omega + \omega_p(\boldsymbol{q}) - \mathrm{i}\delta} \right], \tag{11}$$

where we have $\omega_p(\boldsymbol{q}) = \sqrt{\Omega_{p,p}(\boldsymbol{q})}$ and

$$w_{p,q}(\boldsymbol{G}) = \frac{1}{\sqrt{2\omega_p(\boldsymbol{q})}} V_G(\boldsymbol{q}) S^*_{p,G}(\boldsymbol{q}).$$

The coefficients $w_{p,q}(\boldsymbol{G})$ are the Fourier components of the fluctuation potentials [8, 9] calculated within the RPA. The matrix elements for the bare-exchange part of the self-energy Σ^X, i.e. the Fock part, are given by [9]

$$\langle \phi_{mk} | \Sigma^X | \phi_{lk} \rangle = -\frac{1}{N_q \Omega} \sum_n^{occ} \sum_G \sum_q^{BZ} M^{nm}_G(\boldsymbol{k}, \boldsymbol{q}) [M^{nl}_G(\boldsymbol{k}, \boldsymbol{q})]^* V_G(\boldsymbol{q}), \tag{12}$$

where the plane-wave matrix elements

$$M^{nm}_G(\boldsymbol{k}, \boldsymbol{q}) = \langle \phi_{mk} | \mathrm{e}^{\mathrm{i}(q+G)\cdot r} | \phi_{nk-q} \rangle$$

were used. The integration over energy in the remaining energy-dependent correlation part $\Sigma^C(\omega)$ can be done analytically using equations (10) and (11). The result is [9]

$$\langle \phi_{mk} | \Sigma^C(\omega) | \phi_{lk} \rangle = -\frac{1}{N_q \Omega} \sum_n \sum_q^{BZ} \sum_p \frac{\beta^{nm}_p(\boldsymbol{k}, \boldsymbol{q}) [\beta^{nl}_p(\boldsymbol{k}, \boldsymbol{q})]^*}{\omega - E_{nk-q} \pm (\omega_p(\boldsymbol{q}) - \mathrm{i}\delta)}, \tag{13}$$

where the plus sign is for $E_{nk-q} \leqslant \mu$, and the minus sign is for $E_{nk-q} > \mu$. The coefficients are

$$\beta^{nm}_p(\boldsymbol{k}, \boldsymbol{q}) = \sum_G M^{nm}_G(\boldsymbol{k}, \boldsymbol{q}) w_{p,q}(\boldsymbol{G}). \tag{14}$$

The form of expressions (13) and (14) is the same as what one would get in variants of the plasmon-pole models [10, 12]. However, the poles are not limited to the plasmon pole but also include contributions from the electron–hole continuum excitations. This form of equation (13) is a direct result of the fact that the positions of the poles of W are independent of the reciprocal-lattice vectors (i.e., they are only functions of $\boldsymbol{q}$), and that the dependence of W on the reciprocal-lattice vectors is of separable form. The practical advantages of this are that the summation in equation (14) is over one reciprocal-lattice vector, and that it can be calculated independently of the energy argument of Σ^C.

The quasiparticles in solids satisfy the equation [5, 8, 9]

$$[-\tfrac{1}{2}\nabla^2 + V_H + V_{ext}]\psi_{nk}(\boldsymbol{r}) + \int \mathrm{d}\boldsymbol{r}'\, \Sigma(\boldsymbol{r}, \boldsymbol{r}', E_{nk})\psi_{nk}(\boldsymbol{r}') = E_{nk}\psi_{nk}(\boldsymbol{r}). \tag{15}$$

Here V_H is the Hartree potential of the electrons, V_{ext} is the external potential of the ions, and $\psi_{nk}(\boldsymbol{r})$ and E_{nk} are the quasiparticle wavefunction and energy, respectively. Instead of solving equation (15) for both the wavefunction and energy, we approximate the quasiparticle wavefunctions with single-particle LDA wavefunctions. Then the quasiparticle energies are obtained from

$$E_{nk} \approx E_{nk}^{LDA} + \langle \phi_{nk} | \Sigma(E_{nk}) - V^{xc} | \phi_{nk} \rangle, \tag{16}$$

where E_{nk}^{LDA} is the LDA band energy, and V^{xc} is the LDA exchange–correlation potential. Instead of solving equation (16) self-consistently for E_{nk}, we form an approximate band energy by linearly modifying the LDA energy according to $E_{nk}^{\alpha} = \beta E_{nk}^{LDA} + \delta$, with β and δ having different values for valence and conduction bands. Next, we write $E_{nk} = E_{nk}^{\alpha} + \Delta_{nk}$ and approximate the correction Δ_{nk} to linear order by

$$\Delta_{nk} = \frac{1}{1 - \partial \Sigma(E_{nk}^{\alpha})/\partial E_{nk}^{\alpha}} (E_{nk}^{LDA} + \langle \phi_{nk} | \Sigma(E_{nk}^{\alpha}) - V^{xc} | \phi_{nk} \rangle - E_{nk}^{\alpha}). \tag{17}$$

It is important to note that we do not use a self-consistent approach: the band energies used in the calculation of W and Σ_C are LDA energies. We use E_{nk}^{α} to keep the correction Δ_{nk} small. It is quite easy to use E_{nk}^{α} in equation (13), but often the quasiparticle correction $E_{nk} - E_{nk}^{LDA}$ exhibits behaviour that, especially for the conduction band energies, is not easily expressed as an affine function of LDA energy. We have simply chosen to have $\beta \neq 1.0$ only for valence bands and $\delta \neq 0$ for both conduction and valence bands.

4. Results and discussion

We calculate the band states needed for the Green function (equation (10)) and polarizability (equation (7)) using the LDA [6]. This was done in a plane-wave pseudopotential framework, using codes capable of calculations discussed in [27]. For Si we used 28 points in the irreducible part of the Brillouin zone (IBZ) for calculating ε^{-1}. W was calculated for 16 points in the IBZ. For GaAs, we used 56 points in the IBZ to calculate W for 22 points in the IBZ. In both cases we included reciprocal-lattice vectors in the evaluation of equation (6) with $|\boldsymbol{q} + \boldsymbol{K}|^2 < 9.3$ Ryd. For evaluating the self-energies, we used $|\boldsymbol{q} + \boldsymbol{G}|^2 < 7$ Ryd for the correlation part (equation (13)) and $|\boldsymbol{q} + \boldsymbol{G}|^2 < 16.0$ Ryd for the exchange (equation (12)). Additionally, we used a total of 100 bands when calculating both the dielectric matrix and the self-energy. We estimate that, with respect to these numerical parameters, the self-energies are converged to 0.1 eV. To facilitate comparison with other works, we do not include core-polarization effects, and treat the core–valence interaction at an LDA level [28].

Typically, the plasmon-pole models used in the GWA are obtained by taking the functional form of the dielectric function in the electron gas, and modifying it so that the f-sum rule and the static dielectric screening are correctly given by the model. In table 1 we give examples for silicon of the values of $\varepsilon^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q} \to 0, \omega = 0)$ as calculated using the approach introduced in this work. The cut-off specified translates to 137 reciprocal-lattice vectors in the calculation of local-field effects. In table 1 the value of $\varepsilon^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q} \to 0, \omega = 0)$ is also studied as a function of the effective number of poles $J = I/n$ for a few selected pairs $(\boldsymbol{G}, \boldsymbol{G}')$. Already $J = 20$ has a discrepancy less than 10^{-2} when compared to the direct inversion of the dielectric matrix done as explained in [22]. This is important because, unlike in the case of positive energy moments, for the negative moments there is no general rule for the accuracy as a function of J. Increasing J decreases the discrepancy of the values $\varepsilon^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q} \to 0, \omega = 0)$, so $J = 40$ has a maximum discrepancy of less than 10^{-3}, and $J = 160$, less than 10^{-4}. It appears that increasing the number of iterations results in increasingly more accurate values of $\varepsilon^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q} \to 0, \omega = 0)$ even with rather small values of J (keeping I/J constant).

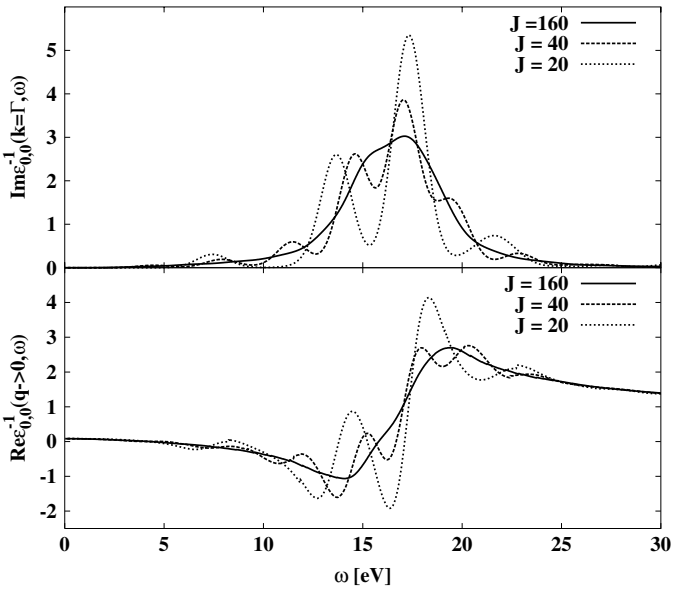

Figure 1. The energy dependence of the dielectric matrix as a function of J. The results for $J = 20$, 40, and 160 are shown. The upper panel shows the results for $\mathrm{Im}\,\varepsilon^{-1}_{\mathbf{00'}}(q \to 0, \omega)$ and the lower panel gives $\mathrm{Re}\,\varepsilon^{-1}_{\mathbf{00'}}(q \to 0, \omega)$. A Gaussian broadening of 1.0 eV half-width at half-maximum (HWHM) was used in all cases.

Table 1. Values of $\mathrm{Re}\,\varepsilon^{-1}_{GG'}(\boldsymbol{q} \to 0, \omega \to 0)$ for Si. The results are given for different values of $J = I/n$, where I is the total number of band Lanczos iterations and n is the number of initial vectors as explained in the text. The direction of $\boldsymbol{q}$ was along a cartesian axis.

G	G'	$J = 20$	$J = 40$	$J = 160$	Direct
(0, 0, 0)	(0, 0, 0)	0.0859	0.0805	0.0796	0.0796
(1, 1, 1)	(0, 0, 0)	0.0002	0.0002	0.0002	0.0002
(1, 1, 1)	(1, 1, 1)	0.600	0.5941	0.5935	0.5935
(−1, 1, 1)	(1, 1, 1)	0.0067	0.0069	0.0069	0.0069
(1, −1, 1)	(1, 1, 1)	−0.0023	−0.0021	−0.0020	−0.0020
(1, −1, 1)	(−1, 1, 1)	−0.0026	−0.0025	−0.0024	−0.0024
(1, 1, −1)	(1, −1, 1)	0.0065	0.0066	0.0065	0.0065
(1, −1, −1)	(−1, 1, 1)	0.0460	0.0481	0.0483	0.0483

Because the aim of the method introduced here is to provide an alternative way to include the full energy dependence of the dielectric matrix into a GWA calculation, we now look at its dependence on the details of the calculation. In figure 1 the results for $\varepsilon^{-1}_{\mathbf{0,0}}(\boldsymbol{q} \to 0, \omega)$ are compared for three values of J, namely $J = 20$, 40, and 160, the last of which is taken to be the converged result. From the figure it is clear that, although there are clear differences in the details of the function, the overall behaviour of $\varepsilon^{-1}_{\mathbf{0,0}}(\boldsymbol{q} \to 0, \omega)$ is reproduced for all values J shown. Also, even for $J = 20$ there are considerable contributions in $\varepsilon^{-1}_{\mathbf{0,0}}(\boldsymbol{q} \to 0, \omega)$ both above and below the plasmon energy. We also checked the result for failure of the orthogonality by using a smaller N, and it was found that this produced only small changes. It should be noted that in the case of silicon, $\varepsilon^{-1}_{\mathbf{0,0}}(\boldsymbol{q} \to 0, \omega)$ has a rather simple energy dependence, and it could be quite accurately approximated in terms of a few poles at complex frequencies.

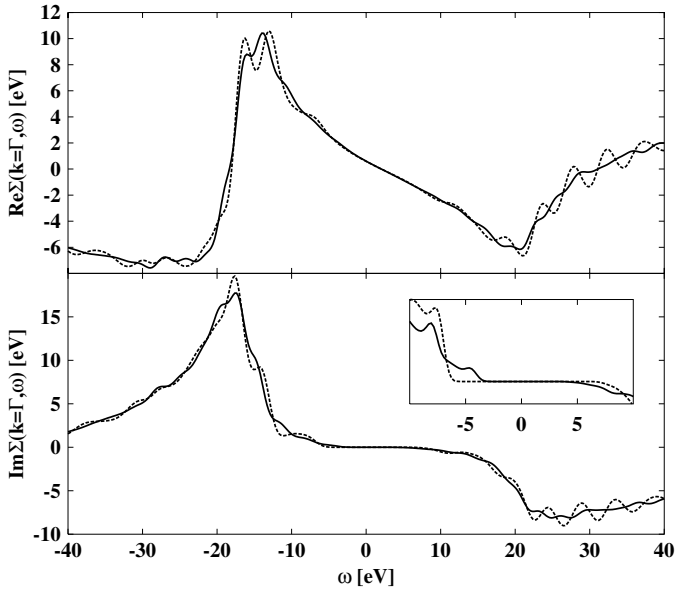

Figure 2. The energy-dependent part of the self-energy (Σ^C) of the highest occupied band Si at the Γ point for two different effective numbers of poles (J). There is a Gaussian broadening of 1.0 eV HWHM for the larger energy scale. In the inset, 0.5 eV HWHM Gaussian broadening was used to assess the convergence. Results for $J = 20$ are given by the dashed curve, and for $J = 40$, by the solid curve. The upper panel shows the real part of the self-energy, and the lower panel shows the imaginary part of the self-energy. In the inset the imaginary part of the self-energy is compared for a smaller energy window. The energy scale was shifted so that the valence band maximum energy was at 0 eV.

These poles and associated weights could be found by fitting the function $\varepsilon^{-1}_{\mathbf{0},\mathbf{0}}(\boldsymbol{q} \to 0, \omega)$ in some selected energy range. In the current approach the poles are real and they are obtained by solving an eigenvalue problem for $\boldsymbol{T}$. The same set of poles will give an approximation for the energy dependence of ε^{-1} for all $(\boldsymbol{G}, \boldsymbol{G}')$ pairs and also give the correct frequency moments. In practice this means that one would expect the same automatic scheme to work also for systems where the energy dependence of ε^{-1} is more complicated.

In figure 2, we show the energy dependence of the self-energy for the highest valence band in silicon. Both the real and imaginary parts of the self-energy are shown. In the figure the self-energy was calculated with two different effective numbers of poles in the screened interaction, namely $J = 20$ and 40. The zero of the energy is set to the value of the valence band maximum. The overall shapes of the self-energy are similar, except that the $J = 20$ case shows a large amount of oscillatory structure. Close to the LDA band energy (0 eV), the differences for the real part of the self-energy are very small. In the imaginary part there is a greater difference in this energy range. The $J = 20$ case has no contribution for small energies ($|\omega| < 5$ eV) as shown in the inset. This is directly related to the fact that W has the smallest eigenvalue of $\approx$6 eV. In figure 3, results for the self-energy for GaAs calculated using the scheme presented here are compared with the results of [29]. The results of [29] shown are for the GWA based on the exact-exchange approximation (solid thick curve) and LDA (dashed thick curve) approaches to density-functional theory. The result of the present work (given by the solid curve) follows quite closely the result of [29] based on the similar approximations, i.e., use of the LDA.

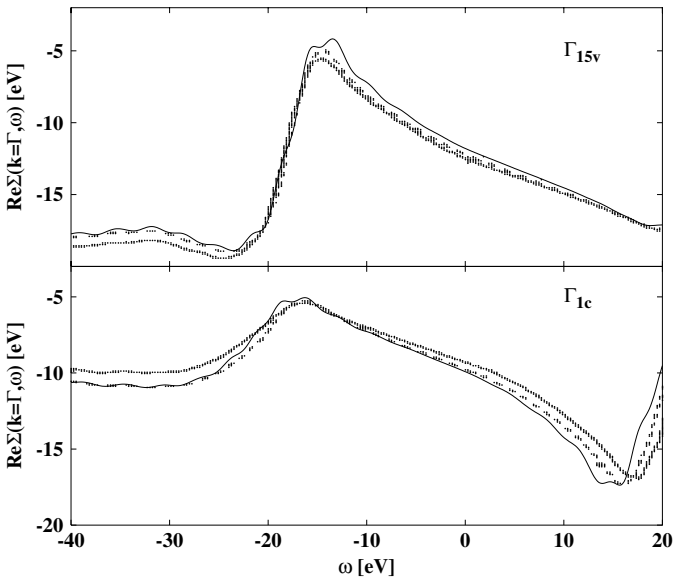

Figure 3. The self-energy (Σ) of the highest occupied band (upper panel) and lowest unoccupied (lower panel) of GaAs at the Γ point. The electronic reproduction of the results of [29] is shown by the thick solid curve and the thick dashed curve, as noted in the text. The thin solid curve shows the results from this work. A Gaussian broadening of 1.0 eV HWHM was used. Following [29] the energy scale was shifted so that the valence band maximum energy was at 0 eV.

Table 2. Calculated quasiparticle energies at points of high symmetry for Si. The energies are given in electron volts, and the valence band maximum was set to zero. The meaning of J is as explained in the text and in table 1.

Band state	LDA	$J = 20$	$J = 40$	[17]
Γ_{1v}	−12.01	−11.86	−11.60	−11.57
Γ_{15c}	2.56	3.25	3.21	3.24
Γ_{2c}	3.04	3.70	3.66	3.94
X_{1v}	−7.83	−7.40	−7.61	−7.57
X_{4v}	−2.86	−2.89	−2.86	−2.83
X_{1c}	0.66	1.34	1.32	1.35
L'_{2v}	−9.64	−9.45	−9.35	−9.35
L_{1v}	−6.99	−6.69	−6.82	−6.78
L'_{3v}	−1.21	−1.23	−1.22	−1.20
L_{1c}	1.42	2.07	2.05	2.14
L_{3c}	3.36	4.09	4.05	4.05

In table 2, the results for quasiparticle band energies of silicon are given for some high-symmetry points. The table shows the results for two different values of J ($J = 20$ and 40), as well as results from [17]. Although the $J = 20$ results usually show improvement over the LDA, the convergence is not monotonic in all cases. However, it is interesting to note that the conduction bands already appear to converge within a few hundredths of an electron volt for $J = 20$. The valence bands appear to converge more slowly. Comparing the results of this work with the results of Fleszar *et al*, there are small deviations. However, usually the differences can be traced back to differences in the LDA band structure. In figure 4, we show the quasiparticle correction $E_{nk} - E_{nk}^{LDA}$ for 16 k-points in the Brillouin zone. For the real part

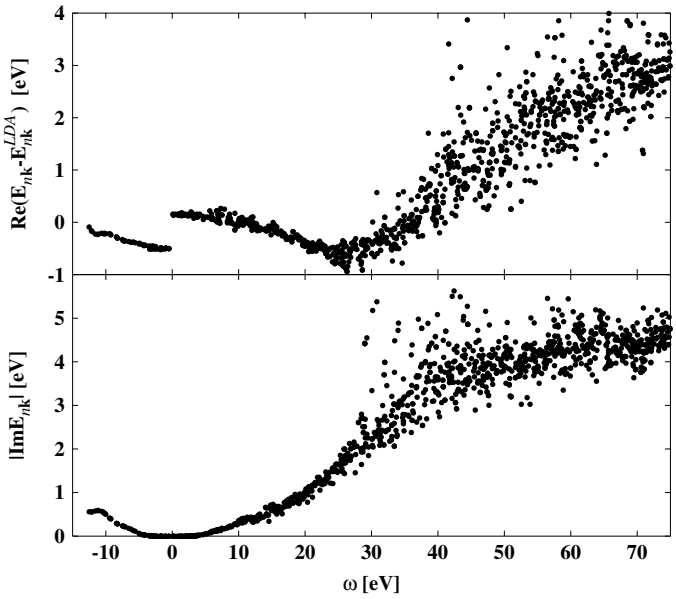

Figure 4. Quasiparticle corrections for Si as a function of the LDA band energy. The upper panel shows the real part of the quasiparticle correction, and the lower panel, the imaginary part of the quasiparticle energy. The result was calculated using a 1.0 eV HWHM Gaussian broadening.

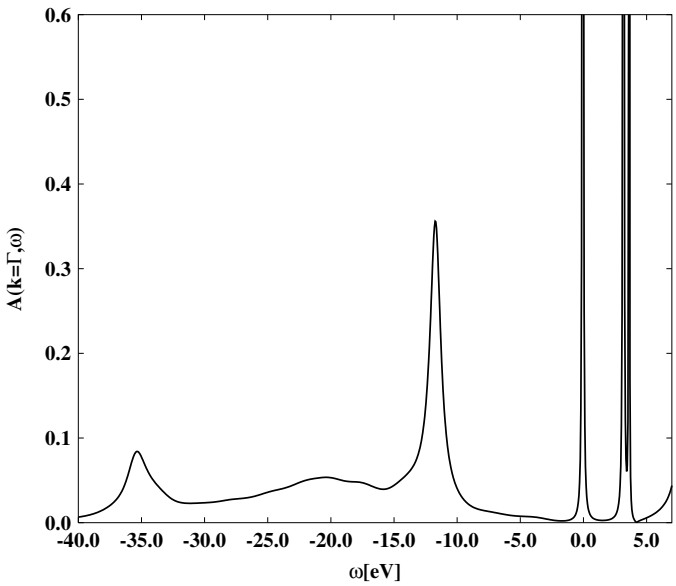

Figure 5. The spectral function for Si at the Γ point. This result was calculated using a 1.0 eV HWHM Gaussian broadening. A small imaginary part (0.005 eV) was added to the self-energy to give a finite width to the quasiparticle peak for the highest valence band states, for purposes of presentation. The energy scale was shifted so that the valence band maximum energy was at 0 eV.

the results are very similar to the ones reported in [17], although we use a linear approximation given in equation (17) and Fleszar *et al* [17] used the position of the quasiparticle peak. The

Table 3. Calculated quasiparticle energies at points of high symmetry for GaAs. The energies are given in eV, and the valence band maximum was set to zero. We have spin–orbit averaged the results of [29] (see the text).

Γ						
LDA	−12.71	0.37	3.75			
GWA (this work)	−12.33	1.00	4.36			
GWA [29]	−12.36	1.34	4.48			
X						
LDA	−10.31	−6.85	−2.64	1.42	1.65	10.12
GWA (this work)	−10.14	−6.80	−2.65	1.96	2.22	10.51
GWA [29]	−10.08	−6.65	−2.61	2.08	2.31	10.69
L						
LDA	−11.06	−6.62	−1.08	0.92	4.68	7.73
GWA (this work)	−10.89	−6.55	−1.11	1.51	5.27	8.17
GWA [29]	−10.80	−6.46	−1.11	1.73	5.35	8.31

highest valence states are moved down by approximately 0.5 eV, and the lowest unoccupied state has hardly moved from its LDA value. The imaginary part of $E_{n\boldsymbol{k}}$, which is related to the lifetime of the quasiparticle, is also shown. It is finite even for band energies close to the valence band maximum (0 eV). Figure 5 shows of the spectral function of Si at Γ calculated using the matrix definition of the spectral function,

$$A(\boldsymbol{k},\omega) = \frac{1}{\pi}\sum_n |\mathrm{Im}\,\langle n\boldsymbol{k}|\frac{1}{\omega - H_0 - \Sigma^X - \Sigma^C(\omega) + V^{xc}}|n\boldsymbol{k}\rangle|,$$

where H_0 is the LDA Hamiltonian. Again we obtain good agreement with [17] as regards the positions and shapes of the spectral features. For example, the wide satellite peak at approximately −20 eV and the relatively sharp plasmon satellite peak of the lowest valence state at approximately −35 eV are correctly reproduced.

As an additional test of the current approach, we apply it to the quasiparticle band structure of GaAs. In table 3 we compare results for the quasiparticle energies of GaAs with the results of [29], which included spin–orbit coupling. Because we do not include this effect, we have spin–orbit averaged their result. Again the agreement of the quasiparticle energies is similar to the agreement of LDA band energies. The largest difference (0.24 eV) in LDA energy between the two works was for the conduction band minimum. This is also reflected in GWA energies, which differ by 0.34 eV.

Because of several case- and implementation-specific details, an accurate comparison of the numerical cost of the method discussed in this paper with other methods is difficult. However, we will try here to give an overall picture of some of the scaling issues associated with the current approach compared to (as an example) the approach discussed in [17]. The method for obtaining the energy dependence of W in [17] is based on obtaining the dielectric matrix ε for a relatively high number of energy values (N_ω) in a predetermined energy range. ε^{-1} is then obtained by inverting the dielectric matrix for each energy. The computational cost of this process is dominated by evaluating the vector inner products required for calculating ε. The computational cost of this process behaves approximately as $N_\omega(N_{\boldsymbol{G}}^{local})^2 N_{prod}$ where N_{prod} is the cost of evaluating a vector inner product using complex vectors. The cost of inverting the matrix is smaller ($N_\omega(N_{\boldsymbol{G}}^{local})^3$). The crystal symmetry can be used to reduce the computational cost in this approach.

In the current approach and for the systems studied here, the computational cost is dominated by operating with V on a vector and subsequent re-orthogonalization at each iteration step. As explained in the appendix, the cost of using V can be reduced. However, the overall cost for the cases studied here and using the algorithm used here behaves approximately as $2J(N_G^{local})^2 N_{prod}$. There are two ways to decrease this computational cost. One can use a smaller number of poles (J) and larger Gaussian broadening and introduce only a small error into the calculation. This possibility is demonstrated by the similarity of the overall shapes of the self-energies in figure 2 for $J = 20$ and 40. Also, using a smaller number of optimized basis functions n but the same J could reduce the computational cost. This possibility was not studied in this paper but one would not expect much optimization to be achieved in this way without compromising the accuracy. The computational costs of calculating the plane-wave matrix elements and the Fock part of the self-energy in the current method and the method in [17] are identical. The cost of evaluating the correlation part of the self-energy (equation (13)) for one value of energy (like the one needed in equation (16)) using the method in [17] behaves as $(N_{i\omega}+1)N_G^2$. Here $N_{i\omega}$ is the number of imaginary energy points used in the energy integral (the second term equation (4) in [17]) and N_G is the number of reciprocal-lattice vectors used in ε^{-1}. For any extra energy points the cost would behave as N_G^2. In the current approach the cost for evaluating each energy value is approximately $(J+2)N_G$. Overall, we estimate that for a given accuracy, the current method is comparable with other ones used in the GWA in terms of computational cost.

5. Conclusions

We have presented an iterative method for calculating the eigenvalue representation of the dielectric matrix and the screened interaction. We showed that the method is capable of giving a good approximation for the energy dependence of the dielectric matrix and the fluctuation potentials. We studied the convergence of the inverse of the dielectric matrix with respect to the effective number of poles for the static case, and the effect that increasing the number of poles has on the energy dependence of ε^{-1}. It was found that in the static case results converged rather uniformly as the number of poles was increased. The overall shape of the energy dependence of ε^{-1} was well reproduced even with a relatively small number of poles. We applied this form to GWA calculations of Si and GaAs, obtaining good agreement with previous work both for the quasiparticle energies and the energy-dependent self-energy. For Si it was found that the quasiparticle energies of the conduction band states converged with a smaller number of poles than those of the valence states. Additionally, we studied the behaviour of the energy-dependent self-energy in Si as a function of the number of iterations.

Acknowledgments

J A Soininen is supported in part by the US Department of Energy through Grant number DE-FG03-97ER45623 and by the National Institute of Standards and Technology, through the DOE Computational Materials Science Network (CMSN).

Appendix

We derive here some properties of the band Lanczos method in connection of the RPA that were not used in this work but might be useful in future work. We start by writing the local-field effects in terms of the initial vectors of the Lanczos iteration as (assuming that we have

$N_G^{local} = N$)

$$V = \boldsymbol{V}\boldsymbol{U}\boldsymbol{V}^{\dagger},$$

where we have

$$\boldsymbol{U}_{ij} = \sum_{i_K} \boldsymbol{R}_{ii_K} V_{\boldsymbol{K}}(\boldsymbol{q}) \boldsymbol{R}^{*}_{ji_K}.$$

If we consider the Lanczos vectors generated by an iteration where local-field effects are excluded, we have

$$\tilde{T}_{i+n,i}\tilde{\boldsymbol{v}}_{i+n} = \boldsymbol{H}_{1e}\tilde{\boldsymbol{v}}_i - \sum_{j=i-n}^{i+n-1} \tilde{T}_{j,i}\tilde{\boldsymbol{v}}_j,$$

and where V is included, we have

$$T_{i+n,i}\boldsymbol{v}_{i+n} = \boldsymbol{H}\boldsymbol{v}_i - \sum_{j=i-n}^{i+n-1} T_{j,i}\boldsymbol{v}_j. \tag{A.1}$$

For $i \leqslant n$ the initial vectors fulfil $\boldsymbol{v}_i = \tilde{\boldsymbol{v}}_i$, and it follows that we have $T_{j,i} - \tilde{T}_{j,i} = U_{j,i}$. This follows from the choice of the coefficients $T_{j,i} = \boldsymbol{v}_j^{\dagger}\boldsymbol{H}\boldsymbol{v}_i$ and the definition of $U_{j,i}$. If we insert this into equation (A.1) we find

$$\begin{aligned}\boldsymbol{H}\boldsymbol{v}_i - \sum_{j=i-n}^{i+n-1} T_{j,i}\boldsymbol{v}_j &= \boldsymbol{H}_{1e}\boldsymbol{v}_i + \sum_{j=1}^{n} U_{j,i}\boldsymbol{v}_j - \sum_{j=1}^{n}(\tilde{T}_{j,i} + U_{j,i})\boldsymbol{v}_j + \sum_{j=n+1}^{i+n-1} \tilde{T}_{j,i}\boldsymbol{v}_j \\ &= \boldsymbol{H}_{1e}\tilde{\boldsymbol{v}}_i - \sum_{j=i-n}^{i+n-1} \tilde{T}_{j,i}\tilde{\boldsymbol{v}}_j,\end{aligned}$$

i.e., we have $\boldsymbol{v}_{i+n} = \tilde{\boldsymbol{v}}_{i+n}$. Additionally, for $i > n$ we have $V\boldsymbol{v}_i = 0$, because of the definition of V and the fact that each new vector $\boldsymbol{v}_{i+n}$ is orthogonal to each of the previous vectors. We have found this to be true to a large degree of accuracy. In a mixed-basis approach this could be used to reduce the computational time considerably. This is because, with a mixed basis, unlike with a plane-wave basis, the Coulomb potential is not diagonal, and operating on a vector with V is computationally the most expensive part of each iteration step.

In other words, both Lanczos iterations, using the full operator $\boldsymbol{H}_{RPA}$ and the diagonal one $\boldsymbol{H}_{1e}$, generated the same set of vectors $\boldsymbol{v}_i$. Additionally, the difference between $T_{j,i}$ and $\tilde{T}_{j,i}$ can be calculated without calculating the vector multiplications required in the Lanczos process using the matrix $\boldsymbol{H}_{RPA}$. This property of the band Lanczos iteration with RPA can be used to reduce the effect of finite accuracy of the floating-point operations and the computational time required especially in the case when a mixed basis is used. When plane waves are used, this reduction is only by a half, because it does not remove the need to do orthogonalization with respect to the previous $2n + 1$ vectors.

References

[1] Hedin L 1965 *Phys. Rev.* **139** 796
The name of the GW approximation arises from the fact that the electron self-energy involves the product of the electron Green function, G, and the dynamically screened Coulomb interaction, W.

[2] Hedin L and Lundqvist S 1969 *Solid State Physics* vol 23, ed F Seitz, D D Turnbull and H Ehrenreich (New York: Academic) p 1

[3] Hybertsen M S and Louie S G 1985 *Phys. Rev. Lett.* **55** 1418
Hybertsen M S and Louie S G 1986 *Phys. Rev.* B **34** 5390

[4] Godby R W, Schlüter M and Sham L J 1986 *Phys. Rev. Lett.* **56** 2415
Godby R W, Schlüter M and Sham L J 1987 *Phys. Rev.* B **37** 10159
[5] Aryasetiawan F and Gunnarsson O 1998 *Rep. Prog. Phys.* **61** 237
[6] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** 864
Kohn W and Sham L J 1965 *Phys. Rev.* **140** 1133
[7] Zhukov V P, Aryasetiawan F, Chulkov E V, de Gurtubay I G and Echenique P M 2001 *Phys. Rev.* B **64** 195122
[8] Hedin L 1999 *J. Phys.: Condens. Matter* **11** R489
[9] Aulbur W G, Jönsson L and Wilkins J W 2000 *Solid State Physics* vol 54, ed H Ehrenreich and F Spaepen (New York: Academic) p 1
[10] von der Linden W and Horsch P 1988 *Phys. Rev.* B **37** 8351
[11] Cappellini G, Del Sole R, Reining L and Bechstedt F 1993 *Phys. Rev.* B **47** 9892
[12] Engel G E and Farid B 1993 *Phys. Rev.* B **47** 15931
[13] Rojas H N, Godby R W and Needs R J 1995 *Phys. Rev. Lett.* **74** 1827
[14] Rieger M M, Steinbeck L, White I D, Rojas H N and Godby R W 1999 *Comput. Phys. Commun.* **117** 211
[15] Rohlfing M, Krüger P and Pollman J 1993 *Phys. Rev.* B **48** 11810
[16] Aryasetiawan F and Gunnarsson O 1995 *Phys. Rev. Lett.* **74** 3221
[17] Fleszar A and Hanke W 1997 *Phys. Rev.* B **56** 10228
[18] Soininen J A and Shirley E L 2001 *Phys. Rev.* B **64** 165112
[19] Adler S L 1962 *Phys. Rev.* **126** 413
[20] Here the meaning of the particle–hole basis is similar to the one used in
Benedict L X, Shirley E L and Bohn R B 1998 *Phys. Rev. Lett.* **80** 4514
[21] Wiser N 1963 *Phys. Rev.* **129** 62
[22] Hybertsen M S and Louie S G 1987 *Phys. Rev.* B **35** 5585
[23] Kotani T and van Schilfgaarde M 2002 *Solid State Commun.* **121** 461
[24] Cullum J K and Willoughby R A 1985 *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* vol 1 (Boston, MA: Birkhäuser)
[25] Meyer H-D and Pal S 1989 *J. Chem. Phys.* **91** 6195
[26] Szabo A and Ostlund N S 1982 *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (New York: Macmillan)
[27] Pickett W E 1989 *Comput. Phys. Rep.* **9** 115
[28] Shirley E L, Zhu X J and Louie S G 1992 *Phys. Rev. Lett.* **69** 2955
Shirley E L and Martin R M 1993 *Phys. Rev.* B **47** 15413
Shirley E L, Zhu X J and Louie S G 1997 *Phys. Rev.* B **56** 6648
[29] Fleszar A 2001 *Phys. Rev.* B **64** 245204