The *GW* method

Rep. Prog. Phys. **61** (1998) 237–312. Printed in the UK PII: S0034-4885(98)75097-3

The GW method

F Aryasetiawan† and O Gunnarsson‡

† Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-223 62 Lund, Sweden
‡ Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

Received 12 September 1997

Abstract

Calculations of ground-state and excited-state properties of materials have been one of the major goals of condensed matter physics. Ground-state properties of solids have been extensively investigated for several decades within the standard density functional theory. Excited-state properties, on the other hand, were relatively unexplored in *ab initio* calculations until a decade ago. The most suitable approach up to now for studying excited-state properties of extended systems is the Green function method. To calculate the Green function one requires the self-energy operator which is non-local and energy dependent. In this article we describe the GW approximation which has turned out to be a fruitful approximation to the self-energy. The Green function theory, numerical methods for carrying out the self-energy calculations, simplified schemes, and applications to various systems are described. Self-consistency issue and new developments beyond the GW approximation are also discussed as well as the success and shortcomings of the GW approximation.

0034-4885/98/030237+76$59.50

Contents

1. Introduction

The Hamiltonian for a many-electron system is given by (in atomic units $\hbar = m = e = 1$)

$$H = \sum_i \left[-\frac{1}{2}\nabla^2(\boldsymbol{r}_i) + V(\boldsymbol{r}_i) \right] + \frac{1}{2}\sum_{i \neq j} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1}$$

where V is a local external potential such as the field from the nuclei. Solving the above Hamiltonian has been a major problem in molecular and condensed matter physics. The Coulomb interaction in the last term makes the Hamiltonian difficult to solve. For a small system, such as an atom or a small molecule, it is possible to obtain the many-particle ground-state wavefunction by means of the configuration interaction (CI) method (Boys 1950, Dykstra 1988). In this method the wavefunction is expanded as a sum of Slater determinants (Slater 1930) whose orbitals and coefficients are determined by minimizing the total energy. A very accurate ground state wavefunction and energy can be obtained. The computational effort, however, scales exponentially with the system size so that applications to large molecules or solids are not feasible. For excited states, the computational effort becomes very large even for a small system. Fortunately, in practice we are interested in quantities which do not require the full knowledge of the wavefunctions. For example, we are interested in the total energy, excitation spectra and expectation values of single-particle operators which can be obtained from the Green function described in a later section.

Approximate theories are usually concerned with finding a good single-particle approximation for the Coulomb term. The earliest of these theories is the Hartree approximation (1928) where the non-local Coulomb term is replaced by an average local Coulomb potential (Hartree potential) from all the electrons. Although it gives reasonable results, due to a cancellation between exchange and correlation, the Hartree theory is not accurate enough in many cases. An extension of the Hartree theory which takes into account the fermionic nature of the electrons leads to the Hartree–Fock approximation (HFA) (Fock 1930) where in addition to the average local Coulomb potential there is a non-local exchange potential which reflects the Pauli exclusion principle. For systems with an energy gap in their excitation spectra, the HFA gives a qualitatively reasonable result. In fact the HFA works quite well for atoms but for insulating solids, the energy gap is in most cases too large. The reason for this can be traced back to the neglect of correlations or screening which is not too important in atoms but crucial in solids. The HFA already takes into account to a large extent correlation between electrons of the same spin since the Pauli exclusion principle (exchange) prevents them from getting close together. Two electrons of opposite spin, on the other hand, are allowed to occupy the same single-particle state at the cost of a large Coulomb energy. Correlations keep electrons away from each other, creating a screening hole around each electron which reduces the interaction with the other electrons and thereby the Coulomb energy. The energy cost of transferring an electron from one site to a neighbouring site is substantially reduced by screening. In the tight-binding limit, i.e. for localized states, the energy gap is approximately given by U and this is essentially the effective Coulomb energy of the states that form the gap. Thus, correlation or screening reduces the gap from its Hartree–Fock value. In metals, the absence of correlations in the HFA leads to qualitatively wrong results such as a zero density of states at the Fermi level due to a logarithmic singularity in the derivative of the single-particle spectra with respect to k at k_{F} (see, e.g., Ashcroft and Mermin 1976).

To simulate the effect of correlations, Slater introduced the $X\alpha$ method where the exchange potential is modelled by a local potential of the form $V^x = \alpha n^{1/3}$, derived from the electron gas and scaled by a constant α to simulate correlations. n is the local electron

density (Slater 1951a, b, 1953, 1974). This method has been quite successful in calculating ground-state properties and excitation spectra but it is semi-empirical. The $X\alpha$ theory may be regarded as a precursor of the modern density functional theory (DFT) (Hohenberg and Kohn 1964, Kohn and Sham 1965) which has become a standard method for calculating ground-state properties of molecules and solids. Recent reviews on DFT may be found in Jones and Gunnarsson (1989) and Dreizler and Gross (1990). In DFT, the ground-state energy can be shown to be a functional of the ground-state density and to satisfy the variational principle with respect to the density. The explicit form of the functional in terms of the density is not known and such an explicit functional may not exist. Using the variational property of the energy functional, one arrives at a set of single-particle equations, the Kohn–Sham (KS) equations (Kohn and Sham 1965), to be solved self-consistently:

$$[-\tfrac{1}{2}\nabla^2 + V^{\mathrm{H}} + V^{\mathrm{xc}}]\phi_i = \varepsilon_i \phi_i \tag{2}$$

$$n = \sum_i^{\mathrm{occ}} |\phi_i|^2 \tag{3}$$

where V^{H} and V^{xc} are the Hartree and exchange-correlation potentials, respectively. In practical applications, the functional containing the effects of exchange and correlations is approximated by the local density approximation (LDA) where the density in the exchange-correlation potential of the electron gas is replaced by the local density of the real system (Kohn and Sham 1965). The KS eigenvalues ε_i have no clear physical meaning except for the highest occupied which corresponds to the ionization energy (Almbladh and von Barth 1985a). Although there is no theoretical justification, they are often interpreted as single-particle excitation energies corresponding to excitation spectra of the system upon a removal or addition of an electron. In many cases, in particular in sp systems, the agreement with photoemission spectra is quite good but as will be described in the next section, there are also serious discrepancies.

A proper way of calculating single-particle excitation energies or quasiparticle energies (Landau 1956, 1957) is provided by the Green function theory (Galitskii and Migdal 1958, Galitskii 1958). It can be shown that the quasiparticle energies E_i can be obtained from the quasiparticle equation:

$$\left[-\frac{1}{2}\nabla^2(\boldsymbol{r}) + V^{\mathrm{H}}(\boldsymbol{r})\right]\Psi_i(\boldsymbol{r}) + \int \mathrm{d}^3r'\,\Sigma(\boldsymbol{r}, \boldsymbol{r}'; E_i)\Psi_i(\boldsymbol{r}') = E_i\Psi_i(\boldsymbol{r}). \tag{4}$$

The non-local and energy-dependent potential Σ, or the self-energy, contains the effects of exchange and correlations. It is, in general, complex with the imaginary part describing the damping of the quasiparticle. It can be seen that the different single-particle theories amount to approximating the self-energy operator Σ. Calculations of Σ are unfortunately very difficult even for the electron gas. We must resort to approximations and this review describes the GW approximation (GWA) (Hedin 1965a) which is the simplest working approximation beyond the HFA that takes screening into account.

1.1. Problems with the LDA

The LDA has been very successful for describing ground-state properties such as total energies and structural properties. There is no clear theoretical justification why the LDA KS eigenvalues should give excitation energies, and even the exact V^{xc} is not supposed to give the exact quasiparticle energies. Nevertheless, the LDA KS eigenvalues are often found to be in good agreement with the quasiparticle energies measured in photoemission experiments.

Despite its success, there are serious discrepancies already in sp systems and they become worse in d and f systems:

- the bandwidth in Na is 10–20% too large, 3.2 eV in LDA against 2.65 eV experimentally (see, however, section 5.1 for the experimental bandwidth);
- the bandgaps in semiconductors Si, GaAs, Ge, etc are systematically underestimated, by as much as 100% in Ge;
- the bandwidth in Ni is ∼30% too large, 4.5 eV in LDA against 3.3 eV experimentally (Hüfner *et al* 1972, Himpsel *et al* 1979);
- in f systems, the LDA density of states is in strong disagreement with experiment;
- in the Mott–Hubbard insulators of transition metal oxides the LDA bandgap is much too small (Powell and Spicer 1970, Hüfner *et al* 1984, Sawatzky and Allen 1984) and in some cases the LDA gives *qualitatively* wrong results; for example, the Mott–Hubbard insulators CoO and the undoped parent compound of the high-T_c material La_2CuO_4 are predicted to be metals (see, e.g., Pickett 1989);
- the magnetic moments in the transition metal oxides are systematically underestimated (Alperin 1962, Fender *et al* 1968, Cheetham and Hope 1983);
- in alkali-metal clusters, the ionization energies calculated within the LDA are too low compared to experiment (Ishii *et al* 1986, Saito and Cohen 1988);
- in the LDA, the image potential seen by an electron in the vacuum far from a surface decays exponentially instead of the expected $-1/4(z - z_0)$ decay where z is the coordinate normal to the surface and z_0 is the position of the image plane (Lang and Kohn 1973).

Strictly speaking, one should not blame the LDA for all of these discrepancies since many of them are related to excited-state properties which are outside the domain of DFT. However, excited-state properties of finite systems can be calculated within the time-dependent extension of DFT (Runge and Gross 1984, Gross *et al* 1996).

1.2. Theories beyond the LDA

When discussing theories beyond the LDA, a distinction should be made between theories which attempt to find better energy functionals but which lie within DFT and those theories which attempt to mimic the self-energy in order to obtain better quasiparticle energies but which are then usually outside DFT. The GWA belongs to the latter.

1.2.1. Gradient corrections. Since the LDA is based on the homogeneous electron gas, it is natural to take into account the inhomogeneity in the charge density of real systems by including gradient corrections in the energy functional. There are two main approaches. The first is a semi-empirical approach where the exchange-correlation functional is modelled by a functional containing parameters which are adjusted to give the best fit to the cohesive energies of a number of 'standard molecules'. The most successful of these models are due to Becke (1988, 1992, 1996). The other approach attempts to calculate the coefficients in the gradient expansion from first principles (Langreth and Mehl (1983), Svendsen and von Barth (1996), Springer *et al* (1996); for a recent development see Perdew *et al* (1997) and references therein). While, in general, gradient corrections give a significant improvement in the total energy (Causa and Zupan 1994, Philipsen and Baerends 1996, Dal Corso *et al* 1996) there is almost no major improvement for quasiparticle energies (Dufek *et al* 1994).

1.2.2. LDA + U. One of the problems with the LDA is the absence of orbital dependence in the exchange-correlation potential. Since the potential does not distinguish between orbitals with different m-quantum numbers, for systems containing a partially filled d or

f shell one obtains a corresponding partially-filled band with a metallic-type electronic structure unless the exchange and crystal field splitting create a gap between the up and down channel. Thus, the late transition metal oxides, which are insulators, are predicted to be metals by the (nonspin-polarized) LDA. In the LDA + U method (Anisimov *et al* 1991, 1993, Lichtenstein *et al* 1995), and its generalization (Solovyev *et al* 1994, 1996), an orbital-dependent potential U acting only on localized d or f states is introduced on top of the LDA potential. For Mott–Hubbard insulators or rare-earth metal compounds where the partially filled 3d or 4f bands are split by the Coulomb interaction, forming the upper and lower Hubbard band, the LDA + U works reasonably well (Anisimov *et al* 1997). The bandstructure, however, is unsatisfactory. Another problem with the method arises for systems with partially-filled 3d shells which are metallic, like the transition metals. In this case, the LDA + U would produce unphysical results since it would split the partially-filled band.

1.2.3. Self-interaction correction. Apart from the problem with orbital dependence in the LDA, there is another problem associated with an unphysical interaction of an electron with itself. In DFT, only the highest occupied state is free from self-interaction but, in LDA, there is in general self-interaction for all states. This self-interaction is explicitly subtracted out in the self-interaction correction SIC formalism resulting in an orbital-dependent potential (Cowan 1967, Lindgren 1971, Zunger *et al* 1980, Perdew and Zunger 1981). Self-interaction is significant for localized states and it tends to zero for extended states, since in the latter case the charge is spread over the crystal and therefore the Coulomb interaction of an electron with itself is of order $1/N$. Since self-interaction is usually positive, one would expect the LDA eigenvalues for localized states to be too high, as is indeed the case. For atoms, SIC therefore lowers the LDA eigenvalues giving better agreement with experiment. The orbital-dependent potential in SIC can describe Mott–Hubbard insulators although the bandstructure is not likely to be satisfactory. SIC predicts all the 3d monoxides to be insulators except VO which is correctly predicted to be a metal (Svane and Gunnarsson 1990, Szotek *et al* 1993, Arai and Fujiwara 1995). SIC, however, fails to give a delocalized paramagnetic solution for the doped high-T_c compounds and a similar problem is anticipated in LDA + U. For more itinerant systems, SIC does not give localized solutions, and it then reduces to the LDA. Accordingly, application to semiconductors, for example Si, does not increase the LDA gap.

1.2.4. A generalized KS scheme. A recent attempt to improve the LDA description of quasiparticle energies is to choose a non-interacting reference system having the same ground-state density as the real system, as in the conventional KS scheme, but with a non-local potential (Seidl *et al* 1996). Since the potential is non-local, the choice is not unique, different non-local potentials may generate the same ground-state density. A particular choice is a non-local screened exchange potential minus its local form. By definition, the functional from which this potential is derived, is zero at the correct density. The KS equations consist of the usual LDA exchange-correlation potential plus the chosen non-local potential. For semiconductors Si, GaAs, Ge, InP and InSb, this method improves the values of the bandgap. Applications to other systems have not been performed so far.

Another recent scheme proposed by Engel and Pickett (1996) incorporates part of the correlation energy into the kinetic energy functional which may be thought of as mass renormalization. This scheme is shown to improve the description of bandgaps in silicon and germanium but it gives negligible correction for diamond and carbon.

1.2.5. The optimized effective potential method. A natural way of improving the LDA would be to calculate the exchange energy exactly and to generate a local exchange potential by taking a functional derivative of the exchange energy functional with respect to the density (Kotani 1995, Kotani and Akai 1996, Bylander and Kleinman 1995a, b). The correlation energy can be approximated by the LDA value. The original idea of this method is due to Talman and Shadwick (1976) in their work on atoms where the Hartree–Fock total energy is minimized with orbitals restricted to be solutions to single-particle Hamiltonians with local potentials. The scheme is also known as the optimized effective potential method. Applications of this approach to several semiconductors and insulators (C, Si, Ge, MgO, CaO and MnO) yields encouraging results regarding the bandgaps, which in most cases are improved from the corresponding LDA values. However, the scheme is still within the density functional formalism and it is intended to improve the energy functional rather than the self-energy. One advantage of this scheme is the possibility of systematically improving the energy functional. With regard to GW calculations, the scheme may provide better starting points than the LDA. The scheme may be extended to include correlations by using a screened interaction potential.

1.3. Motivations for the GWA

The theories described above have drawbacks when applied to calculating quasiparticle energies. Gradient corrections attempt to improve total energies but do not address the problem of improving quasiparticle energies. Indeed, applications to transition metal oxides, where the gap is zero or grossly underestimated by the LDA, do not give any significant improvement. SIC theory only applies to localized occupied states and numerical calculations show that the resulting eigenvalues are too low. The LDA $+ U$ is designed for systems with localized states split by the Coulomb correlation, forming the upper and lower Hubbard band. Applications to transition metals, however, would lead to unreasonable results. The generalized KS scheme using screened exchange, like the other theories discussed above, has no energy dependence which can be important in some cases. Moreover, the choice of the non-local potential is rather arbitrary. Since the theory has not been applied extensively, it is difficult to judge its usefulness. The exact exchange approach should, in principle, improve total energies when correlations are also taken into account but it does not address the LDA problems with quasiparticles.

The GWA is derived systematically from many-body perturbation theory. The form of the self-energy in the GWA is the same as in the HFA but the Coulomb interaction is dynamically screened, remedying the most serious deficiency of the HFA. The self-energy in the GWA is therefore non-local and energy dependent. The GWA is physically sound because it is qualitatively correct in some limiting cases (Hedin 1995) which allows applications to a large class of materials, metals or insulators.

- In atoms, screening is small and the GWA approaches the HFA which is known to work well for atoms.
- In the electron gas, screening is very important which is taken into account in the GWA and for semiconductors it can be shown that the GWA reduces the Hartree–Fock gaps.
- For a core electron excitation, the GWA corresponds to the classical Coulomb relaxation energy of the other electrons due to the adiabatic switching on of the core hole potential, which is just what is to be expected physically.
- For a Rydberg electron in an atom, the GWA gives the classical Coulomb energy of the Rydberg electron due to the adiabatic switching on of an induced dipole in the ion core.
- For the decay rate and the energy loss per unit time of a fast electron in an electron

gas, the GWA gives the correct formula.

The GWA has been applied with success to a wide class of systems ranging from simple metals to transition metals and their compounds.

So far, the GWA has been applied mainly to calculate single-particle excitation spectra, but it is also possible to calculate the total energy (Lundqvist and Samathiyakanit 1969, Holm and von Barth 1997, Farid *et al* 1990) and the expectation value of any single-particle operator in the ground state.

1.4. A short historical survey

The earliest attempt to include correlations beyond the HFA in the form of GW theory was probably the work of Quinn and Ferrell (1958) for the electron gas. Their calculations, however, were limited to states around the Fermi energy and several approximations were made. DuBois (1959a, b) also calculated the self-energy of the electron gas within a GW-type theory but his calculations were only for small values of the electron gas parameter $r_s < 1$ or for high densities since $(4\pi/3)r_s^3 = \rho$, where ρ is the electron density. His results have, therefore, received less attention because they are not directly relevant to real metals which have $r_s \sim 2$–5 (Al $r_s \sim 2$, Cs $r_s \sim 5$). The first full calculation of the self-energy within the GWA for the electron gas was performed by Hedin (1965a). He also showed in a systematic and rigorous way how the self-energy can be expanded in powers of the dynamically screened Coulomb interaction, with the GWA as the first term in this expansion. Later on, Lundqvist (1967a, b, 1968) did extensive calculations of the self-energy of the electron gas for various densities and studied the spectral functions. Rice (1965) used a different version of what is conventionally known as the GWA, including vertex corrections (corrections beyond the GWA). His results are similar to those of Hedin. Later on, Mahan and his group (Mahan and Sernelius 1989, Frota and Mahan 1992) performed extensive self-energy calculations for the electron gas using various forms of the GW-type approximations, studying the importance of vertex corrections.

Due to computational difficulties, the GWA was not applied to real materials until the mid 1980s, starting with the work of Hybertsen and Louie (1985a, b, 1986, 1987a, b) on semiconductors with encouraging results. At about the same time, Godby *et al* (1986, 1987a, b, 1988) did the same calculations and their results are in good agreement with those of Hybertsen and Louie (1986). We should also mention an earlier calculation for diamond using the tight-binding approach by Strinati *et al* (1982), although it is superseded by later calculations. The good results for semiconductors encouraged further applications to more complicated systems, transition metals and their compounds, jellium surface, sodium clusters and to f systems. Now the GWA has become a standard method for including correlations beyond the HFA.

2. Theory

In this section we describe a brief summary of the Green function theory and derive the GWA. More details on the Green function theory may be found in standard text books on many-body theory (e.g. Nozières 1964, Fetter and Walecka 1971, Inkson 1984, Mahan 1990) and in the review article by Hedin and Lundqvist (1969).

2.1. The Green function and the self-energy

To study the electronic excitation spectrum of a solid, one performs photoemission experiments where photons with a certain energy ω are used as projectiles to knock out electrons. By measuring the kinetic energy (KE) of the photoemitted electrons along a certain direction $\boldsymbol{k}$ and using the conservation of energy and momentum, the excitation spectrum $E(\boldsymbol{k})$ of the solid can be obtained:

$$\omega = \mathrm{KE} + E(\boldsymbol{k}). \tag{5}$$

A photoemission experiment then measures the excitation spectrum of a solid with the presence of a hole (occupied density of states). An inverse photoemission experiment uses electrons as probes to measure the unoccupied density of states or the excitation spectrum with an additional electron.

In the limit of large kinetic energy ('sudden approximation') (Hedin and Lundqvist 1969) of the photoemitted or probing electron, the spectrum is directly related to the one-particle Green function which is defined as

$$\mathrm{i}G(x, x') = \langle N|T[\hat{\psi}(x)\hat{\psi}^{\dagger}(x')]|N\rangle \tag{6}$$

$$= \begin{cases} \langle N|\hat{\psi}(x)\hat{\psi}^{\dagger}(x')|N\rangle & \text{for } t > t' \text{ (electron)} \\ -\langle N|\hat{\psi}^{\dagger}(x')\hat{\psi}(x)|N\rangle & \text{for } t < t' \text{ (hole).} \end{cases} \tag{7}$$

$|N\rangle$ is the exact N-electron ground state, $\hat{\psi}(x)$ is a field operator in the Heisenberg representation which annihilates an electron at $x = (\boldsymbol{r}, t)$ and T is the time-ordering operator which arises naturally from the time development operator defined later in equation (14). The physical interpretation of the Green function is that for $t' > t$ it is the probability amplitude that a hole created at x will propagate to x' and for $t > t'$ the probability amplitude that an electron added at x' will propagate to x. Thus, the Green function describes the photoemission and inverse photoemission processes.

From the Green function we can obtain:

- the expectation value of any single-particle operator in the ground state;
- the ground-state energy;
- the one-electron excitation spectrum.

In this review, we are mainly interested in the excitation spectra. The first and second properties have not been explored for real systems.

From the Heisenberg equation of motion for the field operator

$$\mathrm{i}\frac{\partial\hat{\psi}(x)}{\partial t} = [\hat{\psi}(x), \hat{H}] \tag{8}$$

where the Hamiltonian is given by

$$\hat{H} = \int \mathrm{d}^3r\ \hat{\psi}^{\dagger}(x)h_0(x)\hat{\psi}(x) + \frac{1}{2}\int \mathrm{d}^3r\,\mathrm{d}^3r'\ \hat{\psi}^{\dagger}(\boldsymbol{r}, t)\hat{\psi}^{\dagger}(\boldsymbol{r}', t)v(\boldsymbol{r} - \boldsymbol{r}')\hat{\psi}(\boldsymbol{r}', t)\hat{\psi}(\boldsymbol{r}, t) \tag{9}$$

we obtain the equation of motion for the Green function:

$$\left[\mathrm{i}\frac{\partial}{\partial t} - h_0(x)\right]G(x, x') - \int \mathrm{d}x''\ M(x, x'')G(x'', x') = \delta(x - x') \tag{10}$$

where the mass operator (Hartree potential + self-energy) M is defined to be such that

$$\int \mathrm{d}x_1\ M(x, x_1)G(x_1, x')$$

$$= -\mathrm{i}\int \mathrm{d}^3r_1 v(\boldsymbol{r} - \boldsymbol{r}_1)\langle N|T[\hat{\psi}^{\dagger}(\boldsymbol{r}_1, t)\hat{\psi}(\boldsymbol{r}_1, t)\hat{\psi}(\boldsymbol{r}, t)\hat{\psi}^{\dagger}(\boldsymbol{r}', t')]|N\rangle. \tag{11}$$

h_0 is the kinetic energy operator plus a local external potential. The quantity on the right-hand side is a special case of a two-particle Green function:

$$G_2(1,2,3,4) = (\mathrm{i})^2\langle N|T[\hat{\psi}(1)\hat{\psi}(3)\hat{\psi}^\dagger(4)\hat{\psi}^\dagger(2)]|N\rangle \tag{12}$$

where $1 \equiv x_1 = (\boldsymbol{r}_1, t_1)$ etc.

The self-energy may be evaluated in at least two ways, either by using Wick's theorem (Wick 1950, Fetter and Walecka 1971) or by Schwinger's functional derivative method (Schwinger 1951, Martin and Schwinger 1959). We follow the latter. This is done by introducing a time-varying field $\phi(\boldsymbol{r}, t)$ which is used as a mathematical tool for evaluating the self-energy and it will be set to zero once the self-energy is obtained. Working in the interaction (Dirac) picture we have

$$|\psi_{\mathrm{D}}(\boldsymbol{r}, t)\rangle = \hat{U}(t, t_0)|\psi_{\mathrm{D}}(\boldsymbol{r}, t_0)\rangle. \tag{13}$$

The time development operator $\hat{U}$ is given by

$$\hat{U}(t, t_0) = T\exp\left[-\mathrm{i}\int_{t_0}^{t} \mathrm{d}\tau\, \hat{\phi}(\tau)\right] \tag{14}$$

$$\hat{\phi}(\tau) = \int \mathrm{d}^3 r\, \phi(\boldsymbol{r}, \tau)\hat{\psi}_{\mathrm{D}}^\dagger(\boldsymbol{r}, \tau)\hat{\psi}_{\mathrm{D}}(\boldsymbol{r}, \tau). \tag{15}$$

The relationship between operators in the Heisenberg and Dirac representations is

$$\hat{\psi}(\boldsymbol{r}, t) = \hat{U}^\dagger(t, 0)\hat{\psi}_{\mathrm{D}}(\boldsymbol{r}, t)\hat{U}(t, 0). \tag{16}$$

The field operator $\hat{\psi}_{\mathrm{D}}$ satisfies

$$\mathrm{i}\frac{\partial}{\partial t}\hat{\psi}_{\mathrm{D}} = [\hat{\psi}_{\mathrm{D}}, \hat{H}(\phi = 0)] \tag{17}$$

so it is the same as the unperturbed ($\phi = 0$) Heisenberg operator. The Green function can now be written as

$$\mathrm{i}G(1,2) = \frac{\langle N^0|T[\hat{U}(\infty, -\infty)\hat{\psi}_{\mathrm{D}}(1)\hat{\psi}_{\mathrm{D}}^\dagger(2)]|N^0\rangle}{\langle N^0|\hat{U}(\infty, -\infty)|N^0\rangle}. \tag{18}$$

By taking the functional derivative of G with respect to ϕ we get

$$\frac{\delta G(1,2)}{\delta\phi(3)} = G(1,2)G(3,3^+) - G_2(1,2,3,3^+). \tag{19}$$

Using the above result for G_2 in the definition of M in equation (11) the term GG gives the Hartree potential V^{H} and we define

$$\Sigma = M - V^{\mathrm{H}}. \tag{20}$$

The equation of motion for the Green function becomes

$$\left[\mathrm{i}\frac{\partial}{\partial t} - H_0(x)\right]G(x, x') - \int \mathrm{d}x''\, \Sigma(x, x'')G(x'', x') = \delta(x - x') \tag{21}$$

where

$$H_0 = h_0 + V^{\mathrm{H}} + \phi. \tag{22}$$

Using the identity

$$\frac{\delta}{\delta\phi}(G^{-1}G) = G^{-1}\frac{\delta G}{\delta\phi} + \frac{\delta G^{-1}}{\delta\phi}G = 0 \qquad \rightarrow \qquad \frac{\delta G}{\delta\phi} = -G\frac{\delta G^{-1}}{\delta\phi}G \tag{23}$$

and evaluating $\delta G^{-1}/\delta\phi$, where from equation (21)

$$G^{-1} = \mathrm{i}\frac{\partial}{\partial t} - H_0 - \Sigma \tag{24}$$

we get

$$\Sigma(1,2) = \mathrm{i}\int \mathrm{d}3\,\mathrm{d}4\, G(1,3^+)W(1,4)\Lambda(3,2,4). \tag{25}$$

W is the screened Coulomb potential

$$W(1,2) = \int \mathrm{d}3\,\epsilon^{-1}(1,3)v(3-2) \tag{26}$$

$$\epsilon^{-1}(1,2) = \frac{\delta V(1)}{\delta\phi(2)} \tag{27}$$

where V is the sum of the Hartree and the external potential:

$$V = V^{\mathrm{H}} + \phi. \tag{28}$$

Λ is the vertex function

$$\begin{aligned}\Lambda(1,2,3) &= -\frac{\delta G^{-1}(1,2)}{\delta V(3)} \\ &= \delta(1-2)\delta(2-3) + \frac{\delta\Sigma(1,2)}{\delta V(3)} \\ &= \delta(1-2)\delta(2-3) + \int \mathrm{d}(4567)\,\frac{\delta\Sigma(1,2)}{\delta G(4,5)}G(4,6)G(7,5)\Lambda(6,7,3).\end{aligned} \tag{29}$$

The second line is obtained from equation (24) and the last line by using the chain rule $\delta\Sigma/\delta V = (\delta\Sigma/\delta G)(\delta G/\delta V)$ and by using the identity in equation (23) and the definition of Λ.

Fourier transformation of equation (21) gives (with ϕ now set to zero)

$$[\omega - H_0(\boldsymbol{r})]G(\boldsymbol{r},\boldsymbol{r}',\omega) - \int \mathrm{d}^3r''\,\Sigma(\boldsymbol{r},\boldsymbol{r}'',\omega)G(\boldsymbol{r}'',\boldsymbol{r}',\omega) = \delta(\boldsymbol{r}-\boldsymbol{r}'). \tag{30}$$

If G_0 is the Green function corresponding to $\Sigma = 0$, then we have the Dyson equation

$$G = G_0 + G_0\Sigma G. \tag{31}$$

The first term $G_0(1,2)$ is a direct propagation from 1 to 2 without exchange-correlation interaction and Σ contains all possible exchange-correlation interactions with the system that an electron can have in its propagation from 1 to 2.

In practical applications, G_0 corresponds to $H_0 = H^{\mathrm{Hartree}} + V^{\mathrm{xc}}$ where V^{xc} is some local and energy-independent exchange-correlation potential, for example $V^{\mathrm{xc}}_{\mathrm{LDA}}$. In this case, the Dyson equation becomes

$$G = G_0 + G_0\Delta\Sigma G \tag{32}$$

where $\Delta\Sigma = \Sigma - V^{\mathrm{xc}}$.

2.2. The polarization and response function

The response function is an important quantity in the evaluation of the self-energy. It is related to the inverse dielectric function ϵ^{-1} as follows:

$$\begin{aligned}\epsilon^{-1} &= \frac{\delta V}{\delta \phi} \\ &= 1 + v\frac{\delta \rho}{\delta \phi} \\ &= 1 + v\frac{\delta \rho}{\delta V}\frac{\delta V}{\delta \phi}.\end{aligned} \tag{33}$$

The response function is defined as

$$R(1,2) = \frac{\delta \rho(1)}{\delta \phi(2)} \tag{34}$$

which gives the change in the charge density upon a change in the *external* field. We note that the above response function is a time-ordered one which is related to the physical (causal) response function R^R by (Fetter and Walecka 1971)

$$\mathrm{Re}\, R(\omega) = \mathrm{Re}\, R^R(\omega) \qquad \mathrm{Im}\, R(\omega)\mathrm{sgn}\omega = \mathrm{Im}\, R^R(\omega). \tag{35}$$

The polarization function is defined as

$$P(1,2) = \frac{\delta \rho(1)}{\delta V(2)} \tag{36}$$

which gives the change in the charge density upon a change in the *total* (external + induced) field. Noting that

$$\rho(1) = -\mathrm{i}G(1,1^+) \tag{37}$$

we can write

$$P(1,2) = -\mathrm{i}\int \mathrm{d}3\,\mathrm{d}4\, G(1,3)\Lambda(3,4,2)G(4,1^+). \tag{38}$$

In summary, we have

$$\epsilon^{-1} = 1 + vR \tag{39}$$

$$\epsilon = 1 - vP \tag{40}$$

$$R = P + PvR \tag{41}$$

$$\begin{aligned}W &= v + vPW \\ &= v + vRv.\end{aligned} \tag{42}$$

2.3. The Hedin equations

Summarizing the results in the previous sections, we arrive at the well known set of coupled integral equations (Hedin 1965a, Hedin and Lundqvist 1969). From equations (25), (31), (29) and (42) we have

$$\Sigma(1,2) = \mathrm{i}\int \mathrm{d}(34)\, G(1,3^+)W(1,4)\Lambda(3,2,4) \tag{43}$$

$$G(1,2) = G_0(1,2) + \int \mathrm{d}(34)\, G_0(1,3)\Sigma(3,4)G(4,2) \tag{44}$$

$$\Lambda(1,2,3) = \delta(1-2)\delta(2-3) + \int \mathrm{d}(4567)\,\frac{\delta\Sigma(1,2)}{\delta G(4,5)}G(4,6)G(7,5)\Lambda(6,7,3) \tag{45}$$

$$W(1,2) = v(1,2) + \int \mathrm{d}(34)\, v(1,3) P(3,4) W(4,2) \tag{46}$$

where P is given in equation (38). Like G, Λ and W satisfy Dyson-like equations. Starting from a given approximation for Σ the above set of equations can be used to generate higher-order approximations. Although the equations are exact, a straightforward expansion for the self-energy in powers of the screened interaction may yield unphysical results such as negative spectral functions (Minnhagen 1974, Schindlmayr and Godby 1997). In fact, the expansion itself is only conditionally convergent due to the long-range nature of the Coulomb potential. So far there is no systematic way of choosing which diagrams to sum. The choice is usually dictated by physical intuition.

2.4. Quasiparticles

From the classical theory of the Green functions the solution to equation (10) can be written in a spectral representation

$$G(\boldsymbol{r}, \boldsymbol{r}', \omega) = \sum_i \frac{\Psi_i(\boldsymbol{r}, \omega)\Psi_i^\dagger(\boldsymbol{r}', \omega)}{\omega - E_i(\omega)} \tag{47}$$

where Ψ_i are solutions to the quasiparticle equation:

$$H_0(\boldsymbol{r})\Psi_i(\boldsymbol{r}, \omega) + \int \mathrm{d}^3 r\, \Sigma(\boldsymbol{r}, \boldsymbol{r}', \omega)\Psi_i(\boldsymbol{r}', \omega) = E_i(\omega)\Psi_i(\boldsymbol{r}, \omega). \tag{48}$$

In a crystal, the index i may be associated with the Bloch wavevector and band index. The eigenvalues E_i are, in general, complex and the quasiparticle wavefunctions are not, in general, orthogonal because Σ is not Hermitian but both the real and imaginary part of Σ are symmetric. Suppose at some $\omega = \omega_i$ we find that $\omega_i = \mathrm{Re}\, E_i(\omega_i)$. If $\mathrm{Im}\, E_i(\omega_i)$ is small, then the imaginary part of G is expected to have a peak at this energy (quasiparticle peak) with a lifetime given by $1/\mathrm{Im}\, E_i(\omega_i)$. It may happen that $\omega - \mathrm{Re}\, E_i(\omega)$ is zero or close to zero at some other energies and if the corresponding $\mathrm{Im}\, E_i(\omega)$ are small, then we get satellites. For a non-interacting system, Σ is Hermitian and therefore E_i is real so that the quasiparticle has an infinite lifetime.

The spectral representation can also be obtained directly from the definition of G by inserting a complete set of $(N \pm 1)$-electron states in between the field operators and performing a Fourier transformation, keeping in mind that the field operators are in the Heisenberg representation, i.e. $\hat{\psi}(t) = \exp(\mathrm{i}\hat{H}t)\hat{\psi}(0)\exp(-\mathrm{i}\hat{H}t)$:

$$G(\boldsymbol{r}, \boldsymbol{r}', \omega) = \int_{-\infty}^{\mu} \mathrm{d}\omega' \frac{A(\boldsymbol{r}, \boldsymbol{r}', \omega')}{\omega - \omega' - \mathrm{i}\delta} + \int_{\mu}^{\infty} \mathrm{d}\omega' \frac{A(\boldsymbol{r}, \boldsymbol{r}', \omega')}{\omega - \omega' + \mathrm{i}\delta}. \tag{49}$$

The spectral function or density of states A is given by

$$\begin{aligned} A(\boldsymbol{r}, \boldsymbol{r}', \omega) &= -\frac{1}{\pi}\,\mathrm{Im}\, G(\boldsymbol{r}, \boldsymbol{r}', \omega)\mathrm{sgn}(\omega - \mu) \\ &= \sum_i h_i(\boldsymbol{r}) h_i^*(\boldsymbol{r}')\delta[\omega - \mu + e(N-1, i)] \qquad (50) \\ &\quad + \sum_i p_i^*(\boldsymbol{r})\, p_i(\boldsymbol{r}')\delta[\omega - \mu - e(N+1, i)] \qquad (51) \end{aligned}$$

where

$$h_i(\boldsymbol{r}) = \langle N-1, i|\hat{\psi}(\boldsymbol{r}, 0)|N\rangle \tag{52}$$

$$p_i(\boldsymbol{r}) = \langle N+1, i|\hat{\psi}^\dagger(\boldsymbol{r}, 0)|N\rangle \tag{53}$$

$|N \pm 1, i\rangle$ is the ith eigenstate of the $N \pm 1$ electrons with an excitation energy

$$e(N \pm 1, i) = E(N \pm 1, i) - E(N \pm 1) \tag{54}$$

which is positive and $E(N \pm 1)$ is the ground-state energy of the $N \pm 1$ electrons. The quantity μ is the chemical potential,

$$\begin{aligned} \mu &= E(N+1) - E(N) \\ &= E(N) - E(N-1) + \mathrm{O}(1/N). \end{aligned} \tag{55}$$

The physical meaning of the poles of G is therefore the exact excitation energies of the $N \pm 1$ electrons. Since the poles of G in equations (47) and (49) must be the same, it follows that the real parts of $E_i(\omega_i)$ are also the excitation energies of the $N \pm 1$ electrons. For a very large system, the poles are usually so close together that it is meaningless to talk about the individual poles, and in an infinitely large system the poles form a branch cut. In this case, it is more meaningful to interpret the excitation spectrum in terms of quasiparticles with energies $\mathrm{Re}E_i(\omega_i)$ and lifetimes $1/\mathrm{Im}E_i(\omega_i)$.

From equation (31), the spectral function A is schematically given by

$$\begin{aligned} A(\omega) &= \frac{1}{\pi} \sum_i |\mathrm{Im}\, G_i(\omega)| \\ &= \frac{1}{\pi} \sum_i \frac{|\mathrm{Im}\, \Sigma_i(\omega)|}{|\omega - \varepsilon_i - \mathrm{Re}\, \Delta\Sigma_i(\omega)|^2 + |\mathrm{Im}\, \Sigma_i(\omega)|^2} \end{aligned} \tag{56}$$

where G_i is the matrix element of G in an eigenstate ψ_i of the non-interacting system H_0. A is usually peaked at each energy $E_i = \varepsilon_i + \mathrm{Re}\, \Delta\Sigma_i(E_i)$ (quasiparticle peak) with a lifetime given by $1/|\mathrm{Im}\, \Sigma_i(E_i)|$ and renormalization factor (weight of the Lorentzian)

$$Z_i = \left[1 - \frac{\partial\, \mathrm{Re}\, \Delta\Sigma_i(E_i)}{\partial\omega}\right]^{-1} < 1. \tag{57}$$

At some other energies ω_p, the denominator may be small and $A(\omega_\mathrm{p})$ could also show peaks or satellite structure which can be due to plasmon excitations or other collective phenomena.

If we start with a single-particle Hamiltonian which is in some sense close to the true interacting Hamiltonian, the quasiparticles of the former Hamiltonian are just a set of δ-functions centred at the single-particle eigenvalues. If the interaction is switched on, typically the delta functions become broadened since the single-particle states can now decay to other excitations and lose some weight which might appear as collective excitations or satellite structures. The term quasiparticle is arbitrary but we usually refer to quasiparticle as an excitation originating from a single-particle state and to satellite as an excitation not contained in the approximate non-interacting system.

The quasiparticle energy can also be calculated to first order in $E_i - \varepsilon_i$ as follows:

$$\begin{aligned} E_i &= \varepsilon_i + \mathrm{Re}\, \Delta\Sigma_i(E_i) \\ &= \varepsilon_i + \mathrm{Re}\, \Delta\Sigma_i(\varepsilon_i) + (E_i - \varepsilon_i) \frac{\partial\, \mathrm{Re}\, \Delta\Sigma_i(\varepsilon_i)}{\partial\omega} \\ &= \varepsilon_i + Z_i\, \mathrm{Re}\, \Delta\Sigma_i(\varepsilon_i). \end{aligned} \tag{58}$$

2.5. The GW approximation

The GWA may be regarded as a generalization of the Hartree–Fock approximation (HFA) but with a dynamically screened Coulomb interaction. The non-local exchange potential in

the HFA is given by

$$\Sigma^x(\boldsymbol{r},\boldsymbol{r}') = -\sum_{\boldsymbol{k}n}^{\text{occ}} \psi_{\boldsymbol{k}n}(\boldsymbol{r})\psi^*_{\boldsymbol{k}n}(\boldsymbol{r}')v(\boldsymbol{r}-\boldsymbol{r}'). \tag{59}$$

In Green function theory, the exchange potential is written as

$$\Sigma^x(\boldsymbol{r},\boldsymbol{r}',t-t') = \mathrm{i}G(\boldsymbol{r},\boldsymbol{r}',t-t')v(\boldsymbol{r}-\boldsymbol{r}')\delta(t-t')$$

which when Fourier transformed yields equation (59). The GWA corresponds to replacing the bare Coulomb interaction v by a screened interaction W:

$$\Sigma(1,2) = \mathrm{i}G(1,2)W(1,2). \tag{60}$$

This is physically well motivated, especially in metals where the HFA leads to unphysical results such as a zero density of states at the Fermi level, due to the lack of screening. Formally, the GWA is obtained by neglecting the second term in the vertex function in equation (29), i.e. setting $\Lambda(1,2,3) = \delta(1-2)\delta(2-3)$. Fourier transforming equation (60) we get

$$\Sigma(\boldsymbol{r},\boldsymbol{r}',\omega) = \frac{\mathrm{i}}{2\pi}\int \mathrm{d}\omega'\, G(\boldsymbol{r},\boldsymbol{r}',\omega+\omega')W(\boldsymbol{r},\boldsymbol{r}',\omega'). \tag{61}$$

For a non-interacting G_0 the imaginary part of the correlation part of the self-energy can be evaluated explicitly:

$$\mathrm{Im}\,\Sigma^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\omega \leqslant \mu) = \pi\sum_{\boldsymbol{k}n}^{\text{occ}} \psi_{\boldsymbol{k}n}(\boldsymbol{r})\psi^*_{\boldsymbol{k}n}(\boldsymbol{r}')\,\mathrm{Im}\,W^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\varepsilon_{\boldsymbol{k}n}-\omega)\theta(\varepsilon_{\boldsymbol{k}n}-\omega) \tag{62}$$

$$\mathrm{Im}\,\Sigma^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\omega > \mu) = -\pi\sum_{\boldsymbol{k}n}^{\text{unocc}} \psi_{\boldsymbol{k}n}(\boldsymbol{r})\psi^*_{\boldsymbol{k}n}(\boldsymbol{r}')\,\mathrm{Im}\,W^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\omega-\varepsilon_{\boldsymbol{k}n})\theta(\omega-\varepsilon_{\boldsymbol{k}n}) \tag{63}$$

where

$$W^{\mathrm{c}} = W - v \tag{64}$$

is the frequency-dependent part of W. The above results are obtained by expressing G and W^{c} in their spectral representations. The spectral representation of G is given in equation (49). For W^{c} it is given by

$$W^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\omega) = \int_{-\infty}^{0}\mathrm{d}\omega'\frac{D(\boldsymbol{r},\boldsymbol{r}',\omega')}{\omega-\omega'-\mathrm{i}\delta} + \int_{0}^{\infty}\mathrm{d}\omega'\frac{D(\boldsymbol{r},\boldsymbol{r}',\omega')}{\omega-\omega'+\mathrm{i}\delta}. \tag{65}$$

D is proportional to the imaginary part of W and defined to be anti-symmetric in ω:

$$D(\boldsymbol{r},\boldsymbol{r}',\omega) = -\frac{1}{\pi}\,\mathrm{Im}\,W^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\omega)\mathrm{sgn}(\omega) \tag{66}$$

$$D(\boldsymbol{r},\boldsymbol{r}',-\omega) = -D(\boldsymbol{r},\boldsymbol{r}',\omega). \tag{67}$$

The spectral representation of the correlation part of the self-energy is

$$\Sigma^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\omega) = \int_{-\infty}^{\mu}\mathrm{d}\omega'\frac{\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega')}{\omega-\omega'-\mathrm{i}\delta} + \int_{\mu}^{\infty}\mathrm{d}\omega'\frac{\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega')}{\omega-\omega'+\mathrm{i}\delta} \tag{68}$$

where

$$\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega) = -\frac{1}{\pi}\,\mathrm{Im}\,\Sigma^{\mathrm{c}}(\boldsymbol{r},\boldsymbol{r}',\omega)\mathrm{sgn}(\omega-\mu). \tag{69}$$

The real part of Σ^{c} can be obtained by performing the principal value integration (the Kramers–Kronig relation or Hilbert transform).

A physically appealing way of expressing the self-energy is by dividing it into a screened-exchange term $\Sigma_{\rm SEX}$ and a Coulomb-hole term $\Sigma_{\rm COH}$ (COHSEX) (Hedin 1965, Hedin and Lundqvist 1969). It is straightforward to verify that the real part of the self-energy can be written as

$$\mathrm{Re}\,\Sigma_{\rm SEX}(\boldsymbol{r},\boldsymbol{r}',\omega) = -\sum_{i}^{\rm occ}\psi_i(\boldsymbol{r})\psi_i^*(\boldsymbol{r}')\mathrm{Re}\,W(\boldsymbol{r},\boldsymbol{r}',\omega-\varepsilon_i) \tag{70}$$

$$\mathrm{Re}\,\Sigma_{\rm COH}(\boldsymbol{r},\boldsymbol{r}',\omega) = \sum_{i}\psi_i(\boldsymbol{r})\psi_i^*(\boldsymbol{r}')P\int_0^\infty \mathrm{d}\omega'\,\frac{D(\boldsymbol{r},\boldsymbol{r}',\omega')}{\omega-\varepsilon_i-\omega'}. \tag{71}$$

The physical interpretation of $\Sigma_{\rm COH}$ becomes clear in the static approximation due to Hedin (1965a). If we are interested in a state with energy ω close to the Fermi level, the matrix element $\langle\psi|\mathrm{Re}\,\Sigma_{\rm COH}(\omega)|\psi\rangle$ picks up most of its weight from states with energies ε_i close to ω in energy. We may then assume that $\omega-\varepsilon_i$ is small compared to the main excitation energy of D, which is at the plasmon energy. If we set $\omega-\varepsilon_i=0$, we get

$$\mathrm{Re}\,\Sigma_{\rm COH}(\boldsymbol{r},\boldsymbol{r}') = \tfrac{1}{2}\delta(\boldsymbol{r}-\boldsymbol{r}')W^{\rm c}(\boldsymbol{r},\boldsymbol{r}',0). \tag{72}$$

This is simply the interaction energy of the quasiparticle with the induced potential due to the screening of the electrons around the quasiparticle. The factor of $1/2$ arises from the adiabatic growth of the interaction. In this static COHSEX approximation, $\Sigma_{\rm COH}$ becomes local.

The polarization function needed to evaluate W is calculated within the random phase approximation (RPA) (Pines and Bohm 1952, Bohm and Pines 1953, Lindhard 1954, Pines 1961, Gell-Mann and Brueckner 1957) which corresponds to neglecting the second term in the vertex function and using a non-interacting G_0:

$$P(\boldsymbol{r},\boldsymbol{r}',\omega) = \sum_{\rm spin}\sum_{\boldsymbol{k}n}^{\rm occ}\sum_{\boldsymbol{k}'n'}^{\rm unocc}\psi_{\boldsymbol{k}n}^*(\boldsymbol{r})\psi_{\boldsymbol{k}'n'}(\boldsymbol{r})\psi_{\boldsymbol{k}'n'}^*(\boldsymbol{r}')\psi_{\boldsymbol{k}n}(\boldsymbol{r}') \times\left\{\frac{1}{\omega-\varepsilon_{\boldsymbol{k}'n'}+\varepsilon_{\boldsymbol{k}n}+\mathrm{i}\delta}-\frac{1}{\omega+\varepsilon_{\boldsymbol{k}'n'}-\varepsilon_{\boldsymbol{k}n}-\mathrm{i}\delta}\right\}. \tag{73}$$

We have used the fact that for every $\psi_{\boldsymbol{k}n}$ there is a $\psi_{-\boldsymbol{k}n}^*$ with the same eigenvalue due to the time-reversal symmetry. The wavefunction $\psi_{\boldsymbol{k}n}$ is normalized to unity in the entire space. The physical meaning of the RPA is that the electrons respond to the total field (external + induced field) as if they were non-interacting.

3. Numerical methods

One of the main computational problems is to calculate the polarization in equation (73). Since P can be long-ranged for crystals, the conventional method of calculating it is to use Bloch basis functions. Noting that $P(\boldsymbol{r}+\boldsymbol{T},\boldsymbol{r}'+\boldsymbol{T}) = P(\boldsymbol{r},\boldsymbol{r}')$, P can be expanded in general as follows

$$P(\boldsymbol{r},\boldsymbol{r}',\omega) = \sum_{\boldsymbol{q}ij}B_{\boldsymbol{q}i}(\boldsymbol{r})P_{ij}(\boldsymbol{q},\omega)B_{\boldsymbol{q}j}^*(\boldsymbol{r}') \tag{74}$$

where $\{B_{\boldsymbol{q}i}\}$ is a set of Bloch basis functions large enough to describe P and the sum over $\boldsymbol{q}$ is restricted to the first Brillouin zone. The matrix elements of P in the basis are given by

$$P_{ij}(\boldsymbol{q},\omega) = \sum_{\mathrm{spin},\boldsymbol{k}}\sum_{n}^{\rm occ}\sum_{n'}^{\rm unocc}\langle B_{\boldsymbol{q}i}\psi_{\boldsymbol{k}n}|\psi_{\boldsymbol{k}+\boldsymbol{q}n'}\rangle\langle\psi_{\boldsymbol{k}+\boldsymbol{q}n'}|\psi_{\boldsymbol{k}n}B_{\boldsymbol{q}j}\rangle$$

$$\times \left\{ \frac{1}{\omega - \varepsilon_{\boldsymbol{k}+\boldsymbol{q}n'} + \varepsilon_{\boldsymbol{k}n} + \mathrm{i}\delta} - \frac{1}{\omega + \varepsilon_{\boldsymbol{k}+\boldsymbol{q}n'} - \varepsilon_{\boldsymbol{k}n} - \mathrm{i}\delta} \right\}.$$

Similarly, the Coulomb potential v can be expanded as in equation (74). The screened potential W can then be calculated from equation (42).

The matrix element of the imaginary part of the self-energy in a state $\psi_{\boldsymbol{q}n}$ is given by

$$\begin{aligned} \mathrm{Im}\,\Sigma^{\mathrm{c}}_{\boldsymbol{q}n}(\omega) = \pi \sum_{\boldsymbol{k}} \sum_{n' \leqslant \mu} \sum_{ij} & \langle \psi_{\boldsymbol{q}n} \psi_{\boldsymbol{k}-\boldsymbol{q},n'} | B_{\boldsymbol{k}i} \rangle \,\mathrm{Im}\, W^{\mathrm{c}}_{ij}(\boldsymbol{k}, \omega - \varepsilon_{\boldsymbol{k}-\boldsymbol{q},n'}) \\ & \times \langle B_{\boldsymbol{k}j} | \psi_{\boldsymbol{k}-\boldsymbol{q},n'} \psi_{\boldsymbol{q}n} \rangle \theta(\omega - \varepsilon_{\boldsymbol{k}-\boldsymbol{q},n'}) \qquad \text{for } \omega \leqslant \mu. \end{aligned} \tag{75}$$

If we are only interested in quasiparticle energies, it is more favourable to perform the frequency integration in the expression for the self-energy in equation (61) along the imaginary axis (Godby *et al* 1988). In this case, W must also be calculated along the imaginary axis which is advantageous since the pole structure along the real axis is avoided. For states away from the Fermi level, there is in addition a contribution to the self-energy from the poles of the Green function.

The choice of basis functions depends on the type of materials we are interested in. For sp systems, plane waves are appropriate especially when used in conjunction with pseudopotentials. For systems containing rather localized states such as the 3d and 4f systems, a large number of plane waves would be needed and therefore localized basis functions are more suitable.

3.1. Plane-wave basis

In this case $B_{\boldsymbol{k}j} \rightarrow \exp[\mathrm{i}(\boldsymbol{k}+\boldsymbol{G})\cdot r]/\sqrt{\Omega}$ where the index j is represented by the reciprocal lattice vector $\boldsymbol{G}$ and Ω is the unit cell volume. This is probably the simplest basis with the following advantages.

- Programming ease: the matrix elements can be calculated easily, particularly when the wavefunctions are also expanded in plane waves.
- The Coulomb potential is diagonal with matrix elements given by $4\pi/|\boldsymbol{k}+\boldsymbol{G}|^2$.
- There is good control over convergence.

However, the disadvantages are as follows.

- It is not feasible to do all-electron calculations. In many cases, it is essential to include core electrons. For example, the exchange of a 3d valence state with the 3s and 4p core states in the late 3d transition metals is overestimated by the LDA by as much as 1 eV which would lead to an error of the same order in the pseudopotential method.
- The size of the response matrix becomes prohibitively large for narrow-band systems due to a large number of plane waves.
- It has no direct physical interpretation.

3.2. Localized basis

For systems containing 3d or 4f electrons, plane-wave basis becomes very costly. Methods based on the linear-muffin-tin-orbital (LMTO) basis or Gaussian basis are more appropriate. In the LMTO method (Andersen 1975), the wavefunctions are expanded as follows,

$$\psi_{\boldsymbol{k}n}(\boldsymbol{r}) = \sum_{\boldsymbol{R}L} \chi_{\boldsymbol{R}L}(\boldsymbol{r}, \boldsymbol{k}) b_{\boldsymbol{k}n}(\boldsymbol{R}L) \tag{76}$$

where χ is the LMTO basis, which in the atomic sphere approximation (ASA) for $\boldsymbol{r}$ in the central cell is given by

$$\chi_{RL}(\boldsymbol{r},\boldsymbol{k}) = \phi_{RL}(\boldsymbol{r}) + \sum_{R'L'} \dot{\phi}_{R'L'}(\boldsymbol{r}) h_{R'L',RL}(\boldsymbol{k}). \tag{77}$$

$\phi_{RL}(\boldsymbol{r}) = \varphi_{Rl}(r)Y_L(\Omega)$ is a solution to the Schrödinger equation inside a sphere centred on an atom at site $\boldsymbol{R}$ for a certain energy ϵ_ν, normally chosen at the centre of the band. $\dot{\phi}_{RL}$ is the energy derivative of ϕ_{RL} at ϵ_ν. One advantage of the LMTO method is that ϕ_{RL} is independent of $\boldsymbol{k}$. When forming the polarization function in equation (73) we have products of wavefunctions for $\boldsymbol{r}$ or $\boldsymbol{r}'$ of the form

$$\psi\psi = [\phi\phi + (\phi\dot{\phi} + \dot{\phi}\phi)h + \dot{\phi}\dot{\phi}h^2]b^2. \tag{78}$$

It is then clear that the sets of products $\phi\phi$, $\phi\dot{\phi}$, $\dot{\phi}\dot{\phi}$ form a complete basis for the polarization function and the response function (Aryasetiawan and Gunnarsson 1994a) since the latter can be written as

$$R = P + PvP + PvPvP + \cdots \tag{79}$$

so that the basis for R is completely determined by that of P. Although this product basis is not complete for the Coulomb potential, it is of no consequence since v is always sandwiched between two P's. It can be easily shown that the product basis is also complete for the self-energy. Thus schematically

$$\begin{aligned}\Sigma(\boldsymbol{k}n,\omega) &= \langle\psi_{\boldsymbol{k}n}|\mathrm{i}GW|\psi_{\boldsymbol{k}n}\rangle \\ &= \langle\psi_{\boldsymbol{k}n}\psi|v|\psi\psi_{\boldsymbol{k}n}\rangle + \langle\psi_{\boldsymbol{k}n}\psi|vRv|\psi\psi_{\boldsymbol{k}n}\rangle\end{aligned} \tag{80}$$

which shows that it is sufficient to expand v in the product basis and the latter is, therefore, a complete basis for the self-energy.

The number of product functions is still large, with nine spd orbitals we have $2[9(9+1)/2]$ products of $\phi\phi$ and $\dot{\phi}\dot{\phi}$ and 9×9 products of $\phi\dot{\phi}$ giving in total 161 product functions. With spdf orbitals the number of product functions is 528. It can be reduced considerably without too much loss of accuracy by neglecting the $\dot{\phi}$ terms since they are small. Moreover, in the polarization function there are no products between conduction states. Therefore, in sp and d systems products of $\phi_d\phi_d$ and $\phi_f\phi_f$, respectively, can be neglected. After these eliminations, the remaining product functions turn out to have a significant number of linear dependences, typically 30–50%, which can be eliminated further giving $\sim$100 optimized product functions per atom for spdf orbitals. A product function is given by

$$\begin{aligned}\tilde{B}_\alpha(\boldsymbol{r}) &= \phi_{RL}(\boldsymbol{r})\phi_{RL'}(\boldsymbol{r}) \\ &= \varphi_{Rl}(r)\varphi_{Rl'}(r)Y_L(\Omega)Y_{L'}(\Omega)\end{aligned} \tag{81}$$

where $\alpha \equiv (\boldsymbol{R}, LL')$. This function is non-zero only inside an atomic sphere centred on atom $\boldsymbol{R}$. There are no products between orbitals centred on different spheres. The optimization is performed by calculating the eigenvalues of the overlap matrix $O_{\alpha\beta} = \langle\tilde{B}_\alpha|\tilde{B}_\beta\rangle$ and subsequently neglecting eigenvectors with eigenvalues less than a certain tolerance. The optimized product functions are then linear combinations

$$B_\alpha = \sum_\gamma \tilde{B}_\gamma z_{\gamma\alpha} \tag{82}$$

where z are the eigenvectors of the overlap matrix: $Oz = ez$. Due to the localized property of the basis, the calculation of the dielectric matrix scales as N^3. This approach has been used with success to calculate loss spectra (Aryasetiawan and Gunnarsson 1994b) and

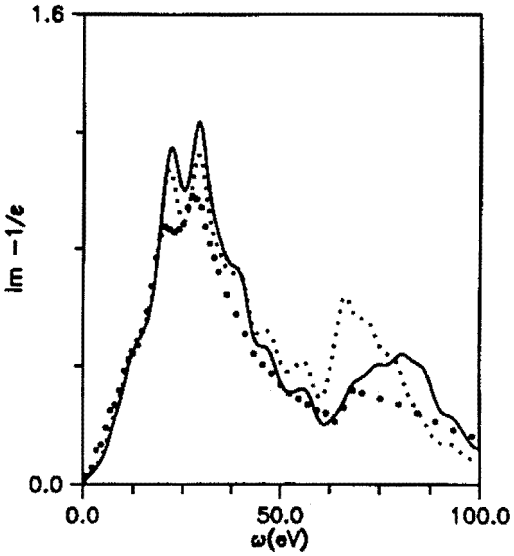

Figure 1. The energy loss spectra of Ni for $\boldsymbol{q} = (0.25, 0, 0)2\pi/a$, $a = 6.654a_0$. The large dots correspond to the experimental spectrum taken from Feldcamp *et al* (1979). The full curve and small dots are respectively the loss spectra with and without local field corrections due to the inhomogeneity in the charge density. Both spectra are calculated with 4s, 4p, 3d, 4d, 4f, and 5g orbitals, including an empty sphere at (0.5, 0.5, 0.5)a and core excitations. After Aryasetiawan and Gunnarsson (1994b).

self-energy for 3d systems (Aryasetiawan and Gunnarsson 1995). The loss spectra of Ni calculated using the product basis is shown in figure 1.

Another approach using a localized basis set is based on Gaussian functions (Rohlfing *et al* 1993). The wavefunctions and the dielectric matrix are expanded in this basis. For Si, for example, the number of Gaussian orbitals needed to perform the GW calculation is 40 to 60 whereas 350 plane waves are needed in the conventional approach. This approach has been applied with success to a number of semiconductors and insulators, and to semicore states in Si, Ge and CdS as well as to the Si surface.

3.3. The plasmon-pole approximation

One of the major computational efforts in self-energy calculations is the calculation of the screened interaction W. The physical features of W are well known; the imaginary part of W is characterized by a strong peak corresponding to a plasmon excitation at the plasmon frequency. This is particularly evident in the case of the electron gas or the alkalis such as Na and Al. The plasmon-pole approximation assumes that all the weight in Im W resides in the plasmon excitation (Lundqvist 1967a, b, Overhauser 1971, Hedin and Lundqvist 1969, Hybertsen and Louie 1986). In the case of the electron gas, this is strictly true in the limit of long wavelength $\boldsymbol{q} \to 0$. For finite $\boldsymbol{q}$, the spectrum also contains particle–hole excitations at lower energies. The particle–hole spectrum eventually merges with the plasmon excitation as $\boldsymbol{q}$ gets larger. Thus, in the simplest form, the plasmon-pole approximation is given by (Lundqvist 1967a, b, Hedin and Lundqvist 1969) $\mathrm{Im}\,\epsilon^{-1}(\boldsymbol{q}, \omega) = A_{\boldsymbol{q}}\delta(\omega - \omega_{\boldsymbol{q}})$. The two parameters $A_{\boldsymbol{q}}$ and $\omega_{\boldsymbol{q}}$ are determined from the static limit of ϵ^{-1} and the f-sum rule. In the generalized plasmon-pole approximation due to Hybertsen and Louie (1986), each matrix

component of the inverse dielectric function is written (for positive frequency)

$$\mathrm{Im}\,\epsilon^{-1}_{GG'}(\boldsymbol{q},\omega) = A_{GG'}(\boldsymbol{q})\delta(\omega - \omega_{GG'}(\boldsymbol{q})). \tag{83}$$

The corresponding real part is given by

$$\mathrm{Re}\,\epsilon^{-1}_{GG'}(\boldsymbol{q},\omega) = \delta_{GG'} + \frac{\Omega^2_{GG'}(\boldsymbol{q})}{\omega^2 - \omega^2_{GG'}(\boldsymbol{q})}. \tag{84}$$

The effective bare plasma frequency $\Omega_{GG'}$ is defined below. The unknown parameters $A_{GG'}(\boldsymbol{q})$ and $\omega_{GG'}(\boldsymbol{q})$ are determined from the static limit of ϵ^{-1}

$$\mathrm{Re}\,\epsilon^{-1}_{GG'}(\boldsymbol{q},0) = \delta_{GG'} + \frac{2}{\pi}P\int_0^\infty \mathrm{d}\omega \frac{1}{\omega}\,\mathrm{Im}\,\epsilon^{-1}_{GG'}(\boldsymbol{q},\omega) \tag{85}$$

and the f-sum rule

$$\int_0^\infty \mathrm{d}\omega\,\omega\,\mathrm{Im}\,\epsilon^{-1}_{GG'}(\boldsymbol{q},\omega) = -\frac{\pi}{2}\omega_\mathrm{p}^2\frac{(\boldsymbol{q}+\boldsymbol{G})\cdot(\boldsymbol{q}+\boldsymbol{G}')}{|\boldsymbol{q}+\boldsymbol{G}|^2}\frac{\rho(\boldsymbol{G}-\boldsymbol{G}')}{\rho(0)} \equiv -\frac{\pi}{2}\Omega^2_{GG'}. \tag{86}$$

Thus, there are no adjustable parameters. The f-sum rule is true for the exact response function since it is obtained from the double commutator $[[H,\hat{\rho}_{\boldsymbol{q}+\boldsymbol{G}}],\hat{\rho}_{\boldsymbol{q}+\boldsymbol{G}'}]$ which yields

$$\frac{1}{2}\sum_s (E_s - E_0)\{\langle 0|\hat{\rho}_{\boldsymbol{q}+\boldsymbol{G}}|s\rangle\langle s|\hat{\rho}^\dagger_{\boldsymbol{q}+\boldsymbol{G}'}|0\rangle + \mathrm{CC}\} = (\boldsymbol{q}+\boldsymbol{G})\cdot(\boldsymbol{q}+\boldsymbol{G}')\rho(\boldsymbol{G}-\boldsymbol{G}'). \tag{87}$$

The states $|s\rangle$ are the exact many-body excited-states and $\rho(\boldsymbol{G})$ is the Fourier component of the electron density. It can also be shown from the definition of the linear response dielectric function in terms of the ground-state matrix element of the commutator of the density operators that

$$\int_0^\infty \mathrm{d}\omega\,\omega\,\mathrm{Im}\,\epsilon^{-1}_{GG'}(\boldsymbol{q},\omega) = -\frac{\pi}{2}v(\boldsymbol{q}+\boldsymbol{G})\frac{1}{2}\sum_s (E_s - E_0)\{\langle 0|\hat{\rho}_{\boldsymbol{q}+\boldsymbol{G}}|s\rangle\langle s|\hat{\rho}^\dagger_{\boldsymbol{q}+\boldsymbol{G}'}|0\rangle + \mathrm{CC}\}. \tag{88}$$

The f-sum rule follows immediately from the last two equations (Hybertsen and Louie 1986) recalling that $\omega_\mathrm{p} = \sqrt{4\pi\rho(0)}$. The quality of the plasmon-pole approximation is illustrated in figure 2.

One drawback of the Hybertsen and Louie model is that the plasmon frequency may become complex, which is somewhat unphysical. A different model proposed by von der Linden and Horsch (1988) circumvents this problem by using the dielectric-bandstructure approach. Here one defines a Hermitian dielectric matrix

$$\epsilon_{GG'}(\boldsymbol{q},\omega) = \delta_{GG'} + \frac{4\pi}{|\boldsymbol{q}+\boldsymbol{G}||\boldsymbol{q}+\boldsymbol{G}'|}\alpha^0_{GG'}(\boldsymbol{q},\omega) \tag{89}$$

$$P^0_{GG'}(\boldsymbol{q},\omega) = \sum_{\mathrm{spin}}\sum_{\boldsymbol{k}}\sum_n^{\mathrm{occ}}\sum_m^{\mathrm{unocc}} \frac{\langle \boldsymbol{k},n|\mathrm{e}^{\mathrm{i}(q+G)\cdot r}|\boldsymbol{k}+\boldsymbol{q},m\rangle\langle \boldsymbol{k}+\boldsymbol{q},m|\mathrm{e}^{-\mathrm{i}(q+G')\cdot r}|\boldsymbol{k},n\rangle}{\omega - \varepsilon_{\boldsymbol{k}n} + \varepsilon_{\boldsymbol{k}+\boldsymbol{q},m}}. \tag{90}$$

The inverse static dielectric matrix is expressed in its eigenrepresentation

$$\epsilon^{-1}_{GG'}(\boldsymbol{q},0) = \delta_{GG'} + \sum_{i=1}^\infty U_{q,i}(\boldsymbol{G})[d_i^{-1}(\boldsymbol{q}) - 1]U^*_{q,i}(\boldsymbol{G}'). \tag{91}$$

The matrix U is formed by the eigenvectors of the inverse dielectric matrix. The plasmon-pole approximation is then obtained by introducing the frequency dependence in the *eigenvalues*:

$$d_i^{-1}(\boldsymbol{q},\omega) - 1 = \frac{z_i(\boldsymbol{q})}{\omega^2 - [\omega_i(\boldsymbol{q}) - \mathrm{i}\delta]^2}. \tag{92}$$

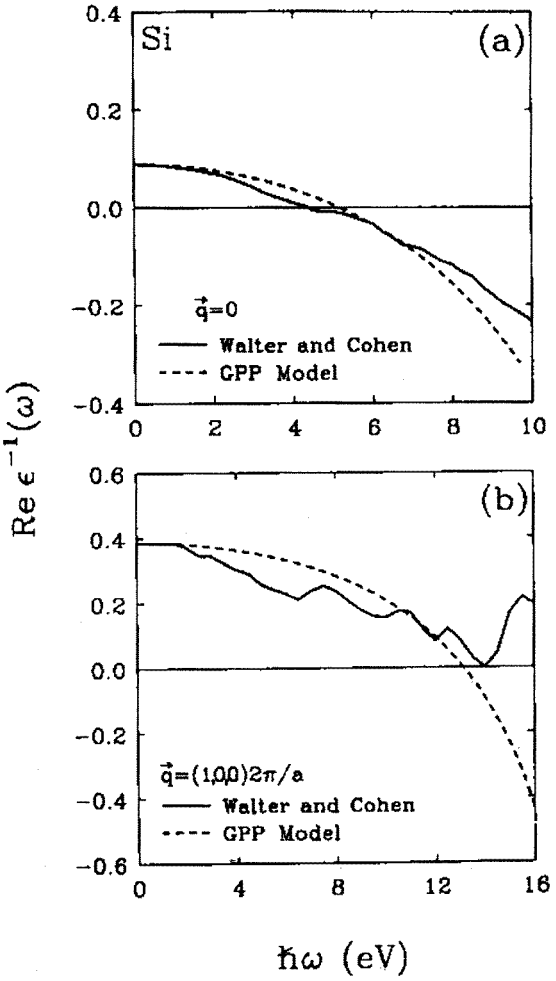

Figure 2. Comparison between the numerically calculated inverse dielectric function (Walter and Cohen 1972) and the corresponding plasmon-pole results for Si. After Hybertsen and Louie (1988a).

The eigenfunctions for $\omega \neq 0$ are approximated by the static eigenfunctions (von der Linden and Horsch 1988). As in the Hybertsen and Louie approach, the unknown quantities z_i and ω_i are determined from the static limit of ϵ^{-1} and f-sum rules. These give

$$z_i(\boldsymbol{q}) = \frac{\omega_{\mathrm{p}}^2}{\rho(0)} \sum_{GG'} U_{q,i}^*(\boldsymbol{G}) \frac{(\boldsymbol{q}+\boldsymbol{G})\cdot(\boldsymbol{q}+\boldsymbol{G}')}{|\boldsymbol{q}+\boldsymbol{G}||\boldsymbol{q}+\boldsymbol{G}'|} \rho(\boldsymbol{G}-\boldsymbol{G}') U_{q,i}(\boldsymbol{G}') \tag{93}$$

and

$$\omega_i^2(\boldsymbol{q}) = \frac{z_i(\boldsymbol{q})}{1 - d_i^{-1}(\boldsymbol{q})}. \tag{94}$$

It can be shown that the pole strength z_i can be expressed in terms of the real-space eigenpotentials

$$z_i(\boldsymbol{q}) = \frac{\omega_p^2}{\rho(0)} \int \mathrm{d}^3 r\, \rho(\boldsymbol{r}) |\nabla \Psi_{\boldsymbol{q}i}(\boldsymbol{r})|^2 \tag{95}$$

where

$$\Psi_{\boldsymbol{q}i}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} \frac{U_{\boldsymbol{q}i}(\boldsymbol{G})\, \mathrm{e}^{-\mathrm{i}(\boldsymbol{q}+\boldsymbol{G})\cdot\boldsymbol{r}}}{|\boldsymbol{q}+\boldsymbol{G}|} \tag{96}$$

so that the pole strength is positive definite. The eigenvalues d_i^{-1} of the static dielectric matrix lie between $(0, 1)$ which implies that the plasmon frequencies ω_i are all real (von der Linden and Horsch 1988). A generalization of the plasmon-pole model has been proposed by Engel *et al* (1991) and Engel and Farid (1992, 1993).

One drawback of the plasmon-pole approximation is that the imaginary part of the self-energy is zero except at the plasmon poles. As a consequence, the lifetime of the quasiparticles cannot be calculated. Another drawback is its limited applicability to other than sp systems. For more complex systems where the plasmon excitations merge with the particle–hole excitations, it is not clear anymore if the plasmon-pole approximation is appropriate.

3.4. The spacetime method

Conventional ways of performing self-energy calculations express all quantities in frequency and reciprocal space. This is natural because in solids the Bloch momentum is a good quantum number and, by working in reciprocal space, one takes advantage of the lattice translational symmetry. In the extreme case of the electron gas, the self-energy becomes diagonal in $\boldsymbol{k}$. In fact, it has been found numerically that, in the Bloch-state representation, the self-energy is almost diagonal even in systems that bear little resemblance to the electron gas (Hybertsen and Louie 1986, Aryasetiawan 1992). The reason for this has been clarified recently by Hedin (1995). The frequency representation is also natural because most experiments generally focus on energy-dependent measurements. For example, photoemission spectra are measured as functions of momentum and energy. Theoretically, however, the self-energy becomes a multidimensional convolution when expressed in momentum–energy space which is computationally very expensive as may be seen in equation (75). In the spacetime representation, on the other hand, the self-energy in the GWA takes a simple multiplicative form

$$\Sigma(\boldsymbol{r}, \boldsymbol{r}', t) = \mathrm{i} G(\boldsymbol{r}, \boldsymbol{r}', t) W(\boldsymbol{r}, \boldsymbol{r}', t). \tag{97}$$

The polarization function also has a simple form

$$P^0(\boldsymbol{r}, \boldsymbol{r}', t) = -\mathrm{i} G^0(\boldsymbol{r}, \boldsymbol{r}', t) G^0(\boldsymbol{r}', \boldsymbol{r}, -t) \tag{98}$$

and the non-interacting Green function G^0 is given by

$$G^0(\boldsymbol{r}, \boldsymbol{r}', t) = \begin{cases} \mathrm{i} \displaystyle\sum_{\boldsymbol{k}n}^{\mathrm{occ}} \psi_{\boldsymbol{k}n}(\boldsymbol{r}) \psi^*_{\boldsymbol{k}n}(\boldsymbol{r}')\, \mathrm{e}^{\mathrm{i}\varepsilon_{\boldsymbol{k}n} t} & t < 0 \\ -\mathrm{i} \displaystyle\sum_{\boldsymbol{k}n}^{\mathrm{unocc}} \psi_{\boldsymbol{k}n}(\boldsymbol{r}) \psi^*_{\boldsymbol{k}n}(\boldsymbol{r}')\, \mathrm{e}^{-\mathrm{i}\varepsilon_{\boldsymbol{k}n} t} & t > 0. \end{cases} \tag{99}$$

Evaluation of G^0 in real time is not advantageous due to the oscillatory exponential term. In imaginary time this exponential term decays rapidly and the summation can be performed easily. G^0 becomes (Rojas *et al* 1995)

$$G^0(\boldsymbol{r},\boldsymbol{r}',\mathrm{i}\tau)=\begin{cases}\mathrm{i}\displaystyle\sum_{\boldsymbol{k}n}^{\mathrm{occ}}\psi_{\boldsymbol{k}n}(\boldsymbol{r})\psi^*_{\boldsymbol{k}n}(\boldsymbol{r}')\,\mathrm{e}^{-\varepsilon_{\boldsymbol{k}n}\tau} & \tau<0\\ -\mathrm{i}\displaystyle\sum_{\boldsymbol{k}n}^{\mathrm{unocc}}\psi_{\boldsymbol{k}n}(\boldsymbol{r})\psi^*_{\boldsymbol{k}n}(\boldsymbol{r}')\,\mathrm{e}^{-\varepsilon_{\boldsymbol{k}n}\tau} & \tau>0.\end{cases}\tag{100}$$

This is obtained by analytically continuing the expression for G^0 in equation (49) from real to imaginary energy and taking the Fourier transform from imaginary energy to imaginary time. From translational and lattice symmetry, $\boldsymbol{r}$ can be restricted to lie in the irreducible wedge but $\boldsymbol{r}'$ must run over all grid points which are located inside a large sphere of radius R_{max} centred on $\boldsymbol{r}$. The method was tested for jellium and Si, and typical values for the grids are $\Delta r = 0.5$ au, $\Delta\tau = 0.3$ au, $R_{\mathrm{max}} = 18$ au and $\tau_{\mathrm{max}} = 10$ au, where au is atomic units, which give convergence in the quasiparticle energy differences to 0.05 eV and absolute values to 0.1 eV (Rojas *et al* 1995).

Although the polarization function P^0 takes a particularly simple form in the spacetime representation, calculating ϵ^{-1} in real space is still computationally prohibitive because of the large size of matrices in inverting the dielectric function. In practice, one performs a six-dimensional fast Fourier transform from $(\boldsymbol{r},\boldsymbol{r}')$ to $(\boldsymbol{k},\boldsymbol{G},\boldsymbol{G}')$ and a one-dimensional Fourier transform from imaginary time to imaginary energy and solves for ϵ^{-1} as a matrix equation in $\boldsymbol{G},\boldsymbol{G}'$ for each $\boldsymbol{k}$ and $\mathrm{i}\omega$. After forming $W_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k},\mathrm{i}\omega)$ one Fourier transforms back to real space and to the imaginary time representation. The advantages of this method are that the computational effort is much reduced compared to the conventional techniques since the double summation over $\boldsymbol{k}$ points and bands are avoided and there is no need for a plasmon-pole approximation.

For comparison with experiment, the self-energy has to be calculated at real frequencies. This is achieved by first calculating the self-energy and the matrix element of the self-energy correction directly in real space $\langle\psi_{\boldsymbol{k}n}|\Sigma(\mathrm{i}\tau)-V^{\mathrm{xc}}|\psi_{\boldsymbol{k}n}\rangle$ and then Fourier transforming the result from imaginary time to imaginary energy. The imaginary energy representation can then be analytically continued to real energies by fitting the self-energy to the multipole form

$$a_0+\sum_{i=1}^{n}\frac{a_i}{\omega-b_i}\tag{101}$$

where the parameters a_i and b_i are, in general, complex. A good fit is obtained with only $n = 2$ with a root-mean-square error of 0.2%. An example is shown in figure 3. The advantage of calculating the self-energy along the imaginary axis is that one avoids the sharp pole structures in both G and W. For quantities which require frequency integration, it is then useful to perform the integration along the imaginary axis whenever possible. However, when the detailed structure of the self-energy is required along the real axis, knowledge of it at a few points along the imaginary axis is not likely to be sufficient.

One important aspect of the spacetime method is its scaling with respect to the system size (Rojas *et al* 1995). It is found numerically that the range of W and Σ is rather material independent so that the parameters R_{max} and Δr do not change with system size. This means that the information stored as well as the computational time for the non-interacting response function should scale linearly with the system size. However, the calculation of the dielectric function involves matrix inversions which scale as N^3. It is still favourable compared with

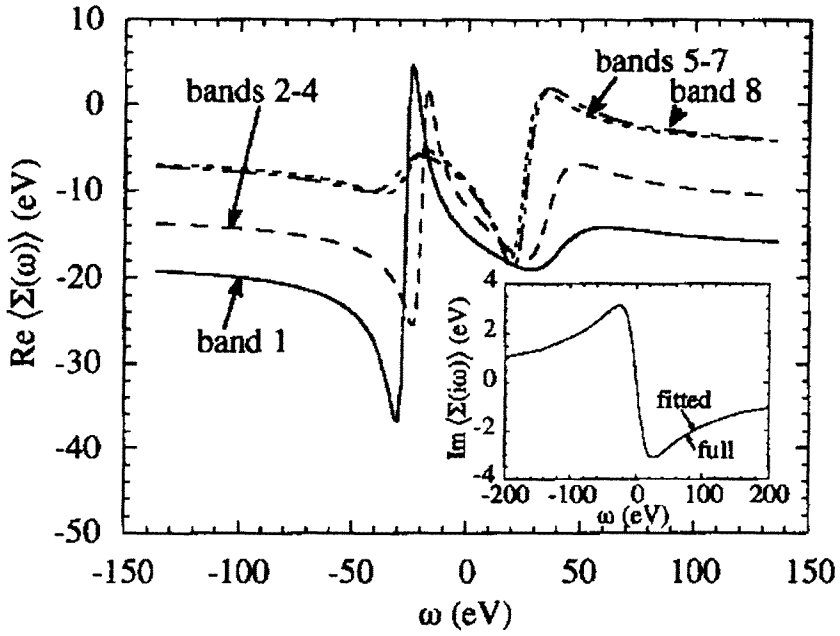

Figure 3. The real part of the matrix elements of the self-energy operator of Si continued onto the real axis for the first eight bands at $\boldsymbol{k} = 0$. Inset: The matrix elements calculated along the imaginary axis for band 4 (the valence band maximum) together with the fitted form (with two poles). After Rojas *et al* (1995).

the conventional plane-wave basis approach which scales as N^4 but comparable to the localized basis approach which also scales as N^3 (Aryasetiawan and Gunnarsson 1994b).

An interesting mixed-space approach for calculating the polarization function was recently proposed by Blase *et al* (1995). The polarization function is written as

$$P(\boldsymbol{r}, \boldsymbol{r}', \omega) = \sum_{\boldsymbol{q}} \exp[\mathrm{i}\boldsymbol{q} \cdot (\boldsymbol{r} - \boldsymbol{r}')] P_{\boldsymbol{q}}(\boldsymbol{r}, \boldsymbol{r}', \omega) \tag{102}$$

where

$$P_{\boldsymbol{q}}(\boldsymbol{r}, \boldsymbol{r}', \omega) = \sum_{\text{spin}} \sum_{\boldsymbol{k}n}^{\text{occ}} \sum_{n'}^{\text{unocc}} u^*_{\boldsymbol{k}n}(\boldsymbol{r}) u_{\boldsymbol{k}+\boldsymbol{q}n'}(\boldsymbol{r}) u^*_{\boldsymbol{k}+\boldsymbol{q}n'}(\boldsymbol{r}') u_{\boldsymbol{k}n}(\boldsymbol{r}') \times \left\{ \frac{1}{\omega - \varepsilon_{\boldsymbol{k}+\boldsymbol{q}n'} + \varepsilon_{\boldsymbol{k}n} + \mathrm{i}\delta} - \frac{1}{\omega + \varepsilon_{\boldsymbol{k}+\boldsymbol{q}n'} - \varepsilon_{\boldsymbol{k}n} - \mathrm{i}\delta} \right\}. \tag{103}$$

The function $P_{\boldsymbol{q}}(\boldsymbol{r}, \boldsymbol{r}', \omega)$ is periodic in $\boldsymbol{r}$ and $\boldsymbol{r}'$ separately and it needs to be calculated within a unit cell only which distinguishes this approach from the direct real-space method where one of the position variables is not restricted to the central cell. The former approach scales as N^3, similar to localized-basis methods. It was found that the crossover between the mixed-space and reciprocal-space methods occurs for unit cells as small as that of Si.

The real-space or the mixed-space approach is suitable for systems with large unit cells and a large variation in the electron density, or open systems.

4. Simplified GW theory

GW calculations are computationally expensive and therefore it is desirable to find simplifications which reduce the numerical effort but still maintain the accuracy of the full calculations. So far there is no simplified GW theory that is applicable to all systems. Most simplified theories are for semiconductors and insulators. Although some success has been achieved, none of these models is very reliable. While bandgaps can be reasonably well reproduced by these models, details of bandstructures are not satisfactorily described. It is a major challenge to construct a good approximation for the self-energy which includes both non-locality and energy dependence, but which is simple enough to be applicable to complex systems without significant loss of accuracy.

4.1. The static Coulomb-hole and screened-exchange (COHSEX) approximation

One of the earliest attempts to simplify GW self-energy is the static COHSEX approximation (Hedin 1965a). It is obtained formally by setting $\omega - \varepsilon_{\boldsymbol{k}n} = 0$ in equations (70) and (71) yielding

$$\Sigma_{\mathrm{SEX}}(\boldsymbol{r}, \boldsymbol{r}') = -\sum_{\boldsymbol{k}n}^{\mathrm{occ}} \phi_{\boldsymbol{k}n}(\boldsymbol{r})\phi^*_{\boldsymbol{k}n}(\boldsymbol{r}')W(\boldsymbol{r}, \boldsymbol{r}', 0) \tag{104}$$

$$\Sigma_{\mathrm{COH}}(\boldsymbol{r}, \boldsymbol{r}') = \tfrac{1}{2}\delta(\boldsymbol{r} - \boldsymbol{r}')[W(\boldsymbol{r}, \boldsymbol{r}', 0) - v(\boldsymbol{r} - \boldsymbol{r}')]. \tag{105}$$

The first term is the exchange self-energy but with a statically screened interaction. The second term is the Coulomb-hole term which is the interaction energy between the quasiparticle and the potential due the Coulomb hole around the quasiparticle, as a result of the rearrangement of the electrons (screening). Both the screened-exchange and the Coulomb-hole terms in equations (70) and (71) are energy dependent and non-local, respectively, but in the static COHSEX approximation they are energy independent and the Coulomb-hole term is local. The validity of the COHSEX approximation relies on whether $\omega - \varepsilon_{\boldsymbol{k}n}$ is small compared to the energy of the main excitation in the screened interaction which is essentially the plasmon energy. Comparison to the results of the full calculations shows that the COHSEX approximation consistently overestimates the magnitude of the self-energy, by about 20% in Si, resulting in larger bandgaps in semiconductors (Hybertsen and Louie 1986). For ω corresponding to an occupied state, most of the error resides in the Coulomb-hole term. The approximation $\omega - \varepsilon_{\boldsymbol{k}n} = 0$ is more severe for the Coulomb-hole term than for the screened-exchange term because the Coulomb-hole term involves a sum over unoccupied states as well as over occupied states whereas the screened-exchange term involves a sum over occupied states only.

Another source of error comes from the neglect of dynamical renormalization or Z-factor in equation (57). When taking the matrix element $\langle\phi_{\boldsymbol{k}n}|\Sigma|\phi_{\boldsymbol{k}n}\rangle$ and setting $\omega - \varepsilon_{\boldsymbol{k}n} = 0$, this approximately corresponds to calculating the self-energy at $\omega = \varepsilon_{\boldsymbol{k}n}$ but the true self-energy should be calculated at the quasiparticle energy, leading to the formula in equation (58). Indeed, a significant improvement can be obtained if the Z-factor is taken into account when the self-energy itself is calculated within the COHSEX approximation.

4.2. Improving the COHSEX approximation

An attempt to include dynamical renormalization into a simplified GW scheme was made by Bechstedt *et al* (1992). In this scheme, the energy-dependent part of the self-energy is expanded to linear order around the LDA eigenvalue. A proper calculation of the self-energy

derivative, or equivalently the Z-factor, requires unfortunately a full frequency-dependent response function. It is not sufficient to calculate the static response function and its energy derivative at $\omega = 0$. For the screened-exchange term, on the other hand, knowledge of the first energy derivative of the response function at $\omega = 0$ is sufficient to expand Σ_{SEX} to first order around the LDA eigenvalue.

To take into account the energy dependence of the self-energy, the screened interaction W is calculated within the plasmon-pole model. The static dielectric function is modelled by (Bechstedt *et al* 1992)

$$\epsilon(\boldsymbol{q}, \rho) = 1 + \left[(\epsilon_0 - 1)^{-1} + \alpha \left(\frac{q}{q_{\mathrm{TF}}} \right)^2 + \frac{3}{4} \left(\frac{q^2}{k_{\mathrm{F}} q_{\mathrm{TF}}} \right)^2 \right]^{-1} \tag{106}$$

where k_{F} and q_{TF} are the Fermi and Thomas–Fermi wavevectors, respectively, which depend on the average electron density ρ. This model for ϵ interpolates between the free-electron gas result $1 + (2\omega_{\mathrm{p}}/q^2)^2$ at high q, $1 + \alpha(q_{\mathrm{TF}}/q)^2$ at small q (Thomas–Fermi theory) and the $q = 0$ value ϵ_0 for the semiconductor. It allows analytical calculation of the screened Coulomb interaction. The coefficient α is obtained by fitting the model dielectric function to a full RPA calculation for $\omega = 0$. The values of α turn out to be material independent for those semiconductors considered (Si, GaAs, AlAs and ZnSe).

To take into account local field effects due to the inhomogeneity in the charge density, an LDA ansatz is used (Hybertsen and Louie 1988a)

$$W(\boldsymbol{r}, \boldsymbol{r}', \omega = 0) = \tfrac{1}{2}[W^{\mathrm{h}}(\boldsymbol{r} - \boldsymbol{r}', \rho(\boldsymbol{r})) + W^{\mathrm{h}}(\boldsymbol{r} - \boldsymbol{r}', \rho(\boldsymbol{r}'))] \tag{107}$$

where W^{h} is the screened interaction for the electron gas. The sum rule, that the total induced charge around a test charge is $-1 + 1/\epsilon_0$, is fulfilled but the induced charge density is allowed to vary according to the local density. Using this model, the static Coulomb-hole term can be calculated analytically (Bechstedt *et al* 1992)

$$V_{\mathrm{COH}}(\boldsymbol{r}) = -\left(1 - \frac{1}{\epsilon_0}\right)^{1/2} \frac{q_{\mathrm{TF}}(\boldsymbol{r})}{\sqrt{\alpha}} \left[1 + \frac{q_{\mathrm{TF}}(\boldsymbol{r})}{\alpha k_F(\boldsymbol{r})} \sqrt{\frac{3\epsilon_0}{\epsilon_0 - 1}} \right]^{-1/2}. \tag{108}$$

It is interesting to observe that the matrix element of $\Sigma^{\mathrm{dyn}} = \Sigma - \Sigma_{\mathrm{COHSEX}}$ and its energy derivative can be shown to be independent of the state if local field effects are neglected. The state dependence thus comes from the local field effects and it is found to be rather weak. We note also the dependence on $q_{\mathrm{TF}} \sim \rho^{1/3}$ as in the LDA. Σ^{dyn} gives an upward shift of $\sim$1.4 eV for all the materials mentioned above. This means that good values for the bandgaps are obtained by using $\Sigma = \Sigma_{\mathrm{COHSEX}}$ but taking into account the dynamical renormalization factor Z. The results for Si, GaAs, AlAs and ZnSe show agreement with the full GW calculations to within 0.2 eV for most of the states considered. Application has also been made recently to GaN (Palummo *et al* 1995). This scheme reduces the computational effort by two orders of magnitude. Further tests on a wide range of semiconductors and insulators are desirable to evaluate the validity of the number of approximations used in the model.

4.3. Extreme tight-binding models

One of the major problems in GW calculations is the calculation of the response function. In electron-gas-like materials such as the alkalis, it is reasonable to model the dielectric function with a single plasmon and to neglect off-diagonal elements. In semiconductors, however, local field effects, which are described by the off-diagonal elements of the dielectric matrix, are important. The local field is dominant at distances on the interatomic scale so that

high Fourier components are needed to describe it in a plane-wave basis, resulting in a large matrix. It is, therefore, advantageous to calculate the dielectric matrix and its inverse directly in real space. A real-space approach is also useful for systems with low symmetry. Ortuno and Inkson (1979) proposed an extreme tight-binding model where the valence and conduction bands were assumed to be flat so that the bandgap E_g is the only parameter entering the model. Due to its simplicity, the model allows analytic evaluation of the dielectric function to a certain degree.

The wavefunctions are expanded in localized Wannier functions

$$\psi_{\boldsymbol{k}n}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\nu \boldsymbol{T}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{T}} \phi_{n\nu}(\boldsymbol{r} - \boldsymbol{T}) c_{n\nu}(\boldsymbol{k}) \tag{109}$$

where $\phi_{n\nu}$ is a Wannier function localized in bond ν and $\boldsymbol{T}$ is a lattice translational vector. In the two-flat-band model, the coefficients $c_{n\nu}$ are equal to one. Since $\phi_{n\nu}$ is localized in each cell and in each bond, we have

$$\int \mathrm{d}^3 r\, \phi_{n\nu}(\boldsymbol{r} - \boldsymbol{T}) \phi_{n\mu}(\boldsymbol{r} - \boldsymbol{T}') \propto \delta_{\nu\mu} \delta_{\boldsymbol{T}\boldsymbol{T}'}. \tag{110}$$

Using this expansion in the expression for the polarization function and using the flat-band approximation, $\varepsilon_{\boldsymbol{k}'n'} - \varepsilon_{\boldsymbol{k}n} = E_g$, one arrives at a simple expression for the dielectric function:

$$\epsilon(\boldsymbol{r}, \boldsymbol{r}', \omega) = \delta(\boldsymbol{r} - \boldsymbol{r}') - N(\omega) \sum_{\nu \boldsymbol{T}} \int \mathrm{d}^3 r''\, v(\boldsymbol{r} - \boldsymbol{r}'') A_\nu(\boldsymbol{r}'' - \boldsymbol{T}) A_\nu^*(\boldsymbol{r}' - \boldsymbol{T}) \tag{111}$$

where

$$N(\omega) = \frac{4E_g}{\omega^2 - (E_g - \mathrm{i}\delta)^2} \tag{112}$$

$$A_\nu(\boldsymbol{r} - \boldsymbol{T}) = \phi_{\mathrm{cond},\nu}(\boldsymbol{r} - \boldsymbol{T}) \phi^*_{\mathrm{val},\nu}(\boldsymbol{r} - \boldsymbol{T}). \tag{113}$$

This has the form of a separable matrix $\epsilon = 1 - \boldsymbol{BC}$ (Hayashi and Shimizu 1969, Sinha 1969) whose inverse is given by $\epsilon^{-1} = 1 + \boldsymbol{B}(1 - \boldsymbol{CB})^{-1}\boldsymbol{C}$. Under certain approximations, the matrix ϵ can be inverted giving a screened interaction (Ortuno and Inkson 1979)

$$\begin{aligned} W(\boldsymbol{r}, \boldsymbol{r}', \omega) &= \int \mathrm{d}^3 r'' \epsilon^{-1}(\boldsymbol{r}, \boldsymbol{r}'', \omega) v(\boldsymbol{r}'' - \boldsymbol{r}') \\ &= v(\boldsymbol{r} - \boldsymbol{r}') - \frac{4E_g}{E_g^2 + \omega_p^2 - \omega^2} \sum_{\nu \boldsymbol{T}} D_\nu(\boldsymbol{r} - \boldsymbol{T}) D_\nu^*(\boldsymbol{r}' - \boldsymbol{T}) \end{aligned} \tag{114}$$

where

$$D_\nu(\boldsymbol{r} - \boldsymbol{T}) = \int \mathrm{d}^3 r'\, v(\boldsymbol{r} - \boldsymbol{r}') A_\nu(\boldsymbol{r}' - \boldsymbol{T}) \tag{115}$$

and ω_p is the plasmon energy. The approximation is good in the limit $E_g \gg \omega_p$. The screened potential is effective at $\omega = E_g$ and approaches a bare value at $\omega = \omega_p$. The quantity $D_\nu(\boldsymbol{r} - \boldsymbol{T})$ represents a dipole moment at $\boldsymbol{r}$ due to bond ν in cell $\boldsymbol{T}$. The physical interpretation of the second term in equation (114) is that an electron on a site interacts with other electrons through the Coulomb interaction, inducing dipole moments on the other sites. These dipoles in turn produce a potential which is screened by other induced dipoles by the frequency-dependent factor arising from the inversion of the dielectric function (Sterne and Inkson 1984).

Using the above screened interaction, the self-energy can be evaluated analytically. Defining a state-dependent local potential

$$\int \mathrm{d}^3 r' \, \Sigma(\boldsymbol{r}, \boldsymbol{r}', \omega = E_n)\psi_{\boldsymbol{k}n}(\boldsymbol{r}') = V_n^{\mathrm{xc}}\psi_{\boldsymbol{k}n}(\boldsymbol{r}) \tag{116}$$

gives two potentials for the valence and conduction states, respectively, (Sterne and Inkson 1984)

$$V_{\mathrm{COHSEX}}^{\mathrm{val}}(\boldsymbol{r}) = -\rho^{1/3}(\boldsymbol{r})\frac{\gamma}{2}\left[\frac{2\pi}{3}\right]^{1/3}\left[1+\frac{1}{\epsilon_0}\right] - \frac{C}{2}\left[\frac{\epsilon_0 - 1}{\epsilon_0 + \sqrt{\epsilon_0}}\right] \tag{117}$$

$$V_{\mathrm{COHSEX}}^{\mathrm{cond}}(\boldsymbol{r}) = -\rho^{1/3}(\boldsymbol{r})\frac{\gamma}{2}\left[\frac{2\pi}{3}\right]^{1/3}\left[1-\frac{1}{\epsilon_0}\right] - \frac{C}{2}\left[\frac{\epsilon_0 - 1}{\epsilon_0 - \sqrt{\epsilon_0}}\right] \tag{118}$$

where

$$C = \int \mathrm{d}^3 r \, \mathrm{d}^3 r' A_\nu^*(\boldsymbol{r} - \boldsymbol{T})v(\boldsymbol{r} - \boldsymbol{r}')A_{\nu'}(\boldsymbol{r}' - \boldsymbol{T}) \tag{119}$$

is an on-site exchange interaction and

$$\epsilon_0 \cong \frac{E_{\mathrm{g}}^2 + \omega_{\mathrm{p}}^2}{E_{\mathrm{g}}^2}. \tag{120}$$

Comparison with the LDA exchange-only potential, $V_{\mathrm{x}}^{\mathrm{LDA}} = -[3/\pi]^{1/3}\rho^{1/3}$, gives a value for $\gamma = 2[9/(2\pi^2)]^{1/3}$. It is interesting to observe that although the potentials have been derived from an extreme tight-binding picture, one arrives at a formula similar to that obtained from the electron gas (Sterne and Inkson 1984).

The method has been applied to C, Si, GaAs, Ge and ZnSe (Jenkins *et al* 1993). Good agreement with the full GW calculations are obtained for Si and GaAs but not for the other two materials. Perhaps this is not surprising in view of the flat extremal bands in Si and GaAs, making them ideal for the present approach. Moreover, the pseudopotentials used in the LDA calculations are better for the first two materials than for the last two so that the eigenfunctions used to calculate the matrix elements of the self-energy are correspondingly better. The LDA exchange only gap is 0.41 eV whereas the calculated value is 1.25 eV (experimentally 1.17 eV). The direct gap at the Γ point is also improved from 2.49 eV to 3.32 eV (experimentally 3.35 eV). In the case of Ge the improvement is not as good as in the case of Si but while the LDA predicts Ge to be a metal with a negative bandgap of -0.19 eV, the method gives a gap of 0.34 eV albeit too low compared to experiment (0.89 eV). In diamond, the method improves the gap at the Γ point from 5.46 eV in the LDA (exchange only) to 7.87 eV (experimentally 7.4 eV). Similar improvement is also obtained across the Brillouin zone. Finally, the results for ZnSe are rather poor. The bandgap is overestimated by $\sim$1 eV (3.77 eV against 2.82 eV). The problem could be related to the difficulties of treating 3d states within the pseudopotential approach resulting in relatively poor wavefunctions.

A similar tight-binding approach was also proposed by Hanke and Sham (1988) and Bechstedt and Del Sole (1988). They derived an analytical model for Σ, V^{xc} and the gap correction in insulators. They arrived at a similar $\rho^{1/3}$ formula for the exchange-correlation potential but they also included a term corresponding to the bare conduction-band exchange which is neglected in the Sterne and Inkson model. The method was applied to Si, diamond and LiCl, and the results are within a few per cent of the more elaborate calculations. A more recent work uses an orthogonalized linear combination of atomic orbitals with applications to diamond, Si, Ge, GaAs, GaP and ZnSe. The gaps are generally within 10% of the experimental values (Gu and Ching 1994).

4.4. Quasiparticle local density approximation (QPLDA)

This approximation is based on the work of Sham and Kohn (1966) who showed that for a system with a slowly-varying density the self-energy is short range in $|\boldsymbol{r}-\boldsymbol{r}'|$ although for larger energies the self-energy is expected to be long range since screening is not very effective at large energies. Numerical calculations by Hedin (1965a) show that the self-energy of the homogeneous electron gas quickly vanishes beyond $|\boldsymbol{r}-\boldsymbol{r}'| = 2r_s$ for $\omega \approx E_{\mathrm{F}}$ and it depends, therefore, on the density in the vicinity of $(\boldsymbol{r}+\boldsymbol{r}')/2$ which suggests the following approximation (Sham and Kohn 1966),

$$\Sigma(\boldsymbol{r}, \boldsymbol{r}', \omega) \approx \Sigma_{\mathrm{h}}(\boldsymbol{r}-\boldsymbol{r}', \omega - \Delta\mu(\overline{n}); \overline{n}) \tag{121}$$

where $\Sigma_{\mathrm{h}}(\boldsymbol{r}-\boldsymbol{r}', E; n)$ is the self-energy operator of the homogeneous electron gas with density n, $\overline{n} = n[(\boldsymbol{r}+\boldsymbol{r}')/2]$ and $\Delta\mu = \mu - \mu_{\mathrm{h}}(\overline{n})$ is the difference between the true chemical potential and that of the homogeneous electron gas of density $\overline{n}$. The local density approximation for the self-energy is obtained by assuming the quasiparticle wavefunction in equation (48) as a superposition of locally plane-wave-like functions (Sham and Kohn 1966)

$$\Psi(\boldsymbol{r}, \omega) \approx A(\boldsymbol{r})\, \mathrm{e}^{\mathrm{i}\boldsymbol{k}(\boldsymbol{r},\omega)\cdot\boldsymbol{r}} \tag{122}$$

which when inserted into the quasiparticle equation gives the solution $\omega(E) = E$ if k satisfies the local-density condition $k = k_{\mathrm{LD}}$ with

$$-E + \tfrac{1}{2}k_{\mathrm{LD}}^2 + V^{\mathrm{H}}(\boldsymbol{r}) + \Sigma(k_{\mathrm{LD}}, E - \Delta\mu(n); n) = 0. \tag{123}$$

In obtaining the above equation, the $\boldsymbol{r}$ dependence of A and $\boldsymbol{k}$ has been neglected. For a slowly-varying density, the local change in the chemical potential from its homogeneous value is simply given by the electrostatic potential (Thomas–Fermi)

$$V^{\mathrm{H}}(\boldsymbol{r}) = \Delta\mu[n(\boldsymbol{r})] \tag{124}$$

and by definition

$$\begin{aligned}\mu_h(n) &= \tfrac{1}{2}k_{\mathrm{F}}^2(n) + \mu^{\mathrm{xc}}(n) \\ &= \tfrac{1}{2}k_{\mathrm{F}}^2(n) + \Sigma(k_{\mathrm{F}}, \mu_{\mathrm{h}}(n), n).\end{aligned} \tag{125}$$

Using the last two expressions, the condition for the local-density wavevector becomes

$$\tfrac{1}{2}(k_{\mathrm{LD}}^2 - k_{\mathrm{F}}^2) = (E-\mu) - [\Sigma_{\mathrm{h}}(k_{\mathrm{LD}}, E - \Delta\mu(n); n) - \Sigma_{\mathrm{h}}(k_{\mathrm{F}}, \mu_{\mathrm{h}}(n); n)]. \tag{126}$$

When operating on the locally plane-wave function, the effect of the non-local operator $\Sigma_{\mathrm{h}}(\boldsymbol{r}-\boldsymbol{r}', E)$ can be reproduced exactly by a local operator $\Sigma_{\mathrm{h}}(k_{\mathrm{LD}}, E; n)\delta(\boldsymbol{r}-\boldsymbol{r}')$. The local density approximation for the self-energy operator consists of using this local operator when operating on the actual wavefunction at energy E. It may happen that for some values of E equation (126) has no solution for positive k_{LD}^2. In this case, one should analytically continue the self-energy operator into complex momentum. An early work of approximating the self-energy by a local potential using information from the local density is given by Hedin and Lundqvist (1971).

Instead of calculating the full self-energy $\Sigma_{\mathrm{h}}(k, E; n)$ it is useful to write (Wang and Pickett 1983, Pickett and Wang 1984)

$$\begin{aligned}\Sigma_{\mathrm{h}}(k, E; n) &= \mu^{\mathrm{xc}}(n) + \Sigma_{\mathrm{h}}(k, E; n) - \Sigma_{\mathrm{h}}(k_{\mathrm{F}}, \mu_{\mathrm{h}}; n) \\ &= \mu^{\mathrm{xc}}(n) + \Delta(k, E; n)\end{aligned} \tag{127}$$

since μ^{xc} has been calculated rather accurately for the electron gas. The GWA is then used to calculate the relatively small quantity Δ. Hedin and Lundqvist (1971) calculated $\Delta(k_{\mathrm{LDA}})$

and for metals, the self-energy corrections in the LDA turn out to be very small. Δ is of the order of a few 10^{-2} eV. For Ni, for example, the corrections are smaller than 0.1 eV (Watson *et al* 1976). The main reason for such small corrections is probably due to the weak energy dependence of the electron-gas self-energy. Furthermore, most of the effect of the self-energy is already accounted for by the local term μ^{xc}. The QPLDA self-energy is dominated by a constant contribution and the assumption of slowly-varying density is probably not good enough in many cases.

Above, the homogeneous electron gas was used to obtain an approximate self-energy for inhomogeneous systems. For semiconductors and insulators one can also use a homogeneous insulator model for this purpose which turns out to give a significant improvement. The presence of an energy gap in these systems brings in new physics not presence in the electron gas. The response function needed to calculate the screened interaction is obtained from a model semiconducting homogeneous electron gas. One simply introduces a gap in the otherwise metallic spectrum of the electron gas, that is to say (Penn 1962, Levine and Louie 1982)

$$\epsilon_2(k,\omega) = \begin{cases} 0 & |\omega| < \lambda E_F \\ \epsilon_2^{RPA}(k,\omega_-) & |\omega| > \lambda E_F \end{cases} \tag{128}$$

where $\omega_- = \sqrt{\omega^2 - \lambda^2 E_F^2}\,\mathrm{sgn}\,\omega$. This function satisfies the f-sum rule. The dielectric gap λE_F, which is in general larger than the minimum direct gap, is determined from the experimental value of the static, long-wavelength dielectric constant ϵ_0,

$$\lambda E_F = \frac{\omega_p}{\sqrt{\epsilon_0 - 1}} \tag{129}$$

where $\omega_p = \sqrt{4\pi n}$ is the plasmon frequency. The single-particle spectrum has a gap E_g

$$E(\boldsymbol{k}) = \tfrac{1}{2}k^2 + \tfrac{1}{2}E_g\,\mathrm{sgn}(k - k_F). \tag{130}$$

The gap E_g is taken to be the average value across the Brillouin zone of the direct gap and it is numerically different from λE_F. It can be shown that the valence bands are lowered and the conduction bands are raised by the self-energy corrections, independently of the density sampled by the wavefunctions. This means that the QPLDA will always lead to an increase in the gap over the LDA value (Wang and Pickett 1983, Pickett and Wang 1984).

The QPLDA was applied to Si and diamond with considerable success. For Si, the zone-boundary transitions $X_4 \to X_1$, $L_3' \to L_3$ and $L_3' \to L_1$ are well reproduced by the QPLDA. The zone-centre transition $\Gamma_{25'} \to \Gamma_{15}$ is underestimated by 0.25 eV, which is the worst case, but the transition $\Gamma_{25'} \to \Gamma_{15}$ is reproduced well. The calculated indirect gap of 0.93 eV still underestimates the experimental value by 0.24 eV. The QPLDA gives a valence-bandwidth of 12.5 eV, in good agreement with the experimental data 12.4 ± 0.6 eV (Grobman and Eastman 1972) and 12.6 ± 0.6 eV (Ley *et al* 1972). For diamond, the QPLDA gives an indirect gap of 5.74 eV, which is a slight overestimate compared to the experimental value of 5.47 eV. The LDA value is 4.05 eV. The $\Gamma_{25'} \to \Gamma_{15}$ gap is experimentally about 7.3 eV and the QPLDA gives 7.36 eV (compared to the LDA value of 5.51 eV). The valence bandwidth is 23.4 eV, in good agreement with the experimental data of McFeely *et al* (24.2 ± 1.0 eV) (McFeely *et al* 1974) but in disagreement with the data of Himpsel *et al* (21 eV) (Himpsel *et al* 1980). The LDA bandwidth is 20.4 eV.

4.5. LDA $+ \delta\Sigma_{COHSEX}$

In metals, the screened interaction W decreases rapidly for $|\boldsymbol{r} - \boldsymbol{r}'| > k_{TF}^{-1}$, but in insulators and semiconductors it decreases as $1/\epsilon_0|\boldsymbol{r} - \boldsymbol{r}'|$ for large $|\boldsymbol{r} - \boldsymbol{r}'|$ since the screening is not

complete. Accordingly, one writes (Gygi and Baldereschi 1989)

$$W(\boldsymbol{r}, \boldsymbol{r}', \omega) = W^{\mathrm{IEG}}(\boldsymbol{r}, \boldsymbol{r}', \omega) + \delta W(\boldsymbol{r}, \boldsymbol{r}', \omega) \tag{131}$$

where W^{IEG} is the short-range interaction potential of a metallic inhomogeneous electron gas and δW has the same long-range behaviour as W. The self-energy arising from the first term in W has been shown by Sham and Kohn (1966) to be short ranged and depends only on the density in the vicinity of $\boldsymbol{r}$ and $\boldsymbol{r}'$, and is therefore approximated by a local potential:

$$\Sigma_{\mathrm{GW}}(\boldsymbol{r}, \boldsymbol{r}', \omega) = \mu^{\mathrm{xc}}(\boldsymbol{r}, \omega)\delta(\boldsymbol{r} - \boldsymbol{r}') + \frac{\mathrm{i}}{2\pi}\int \mathrm{d}\omega'\, G(\boldsymbol{r}, \boldsymbol{r}', \omega + \omega')\delta W(\boldsymbol{r}, \boldsymbol{r}', \omega'). \tag{132}$$

Pickett and Wang (1984) found that the inclusion of energy dependence in the local exchange-correlation potential had very little effects on the eigenvalues obtained with an energy-independent LDA potential. It is then assumed that for states close to the Fermi level, μ^{xc} takes its value at the Fermi level $\mu^{\mathrm{xc}}_{\mathrm{LDA}}(\boldsymbol{r})$. Furthermore, it is assumed that δW depends only on $|\boldsymbol{r} - \boldsymbol{r}'|$, which is strictly valid only when $|\boldsymbol{r} - \boldsymbol{r}'| \to \infty$, but it has been found numerically that in non-metals the local field effects become negligible as $|\boldsymbol{r} - \boldsymbol{r}'|$ exceeds the interatomic distance (Hybertsen and Louie 1986, Godby *et al* 1988). One now makes the COHSEX approximation on $\delta\Sigma$, rather than on the full self-energy

$$\delta\Sigma(\boldsymbol{r}, \boldsymbol{r}') = -\rho(\boldsymbol{r}, \boldsymbol{r}')\delta W(\boldsymbol{r} - \boldsymbol{r}') + \delta\Sigma_{\mathrm{COH}} \tag{133}$$

where the first term is the screened-exchange contribution and

$$\delta\Sigma_{\mathrm{COH}} = \frac{\Omega}{2(2\pi)^3}\int \mathrm{d}^3q\, \delta W(\boldsymbol{q}) \tag{134}$$

is the Coulomb-hole contribution which is a constant, shifting all eigenvalues by the same amount. The Fourier transform of δW is given by

$$\delta W(q) = \frac{4\pi}{\Omega q^2}[\epsilon^{-1}_{\mathrm{SC}}(\boldsymbol{q}, \boldsymbol{q}, \omega = 0) - \epsilon^{-1}_{M}(\boldsymbol{q}, \omega = 0)] \tag{135}$$

where $\epsilon^{-1}_{\mathrm{SC}}(\boldsymbol{q}, \boldsymbol{q}, 0)$ is the diagonal part of the inverse dielectric matrix for the semiconductor calculated in the RPA and $\epsilon^{-1}_{M}(q, 0)$ is the inverse of the static Lindhard dielectric function (Lindhard 1954) of a homogeneous electron gas.

Applications to diamond, Si, Ge, GaAs, AlAs and GaP give values in general agreement to within 0.2 eV with the results of the full calculations. Larger deviations (0.3–0.4 eV) occur for the $X_{4v} \to X_{1c}$ gap in diamond and the $\Gamma_{25'v} \to \Gamma_{2'c}$ gaps in Ge. The Coulomb-hole contribution δE_{COH} is positive and cancels approximately the effect of the screened-exchange term in the valence bands. As a result, the net effect of the self-energy correction is to shift the conduction band upwards and leave the valence bands essentially unchanged. This is also in agreement with the full calculations.

5. Applications

5.1. Alkali metals

The early calculations in the GWA were performed for the electron gas, because of its simplicity (for a review see Hedin and Lundqvist (1969)). The alkali metals are therefore of particular interest, being the systems which are the closest approximation to the electron gas. For these systems the correlation effects are only moderately strong, and the GWA could therefore be expected to be relatively accurate. It therefore created a lot of interest when the first high-resolution angular-resolved photoemission experiments for the alkali

metals appeared (Jensen and Plummer 1985, Lyo and Plummer 1988, Itchkawitz *et al* 1990). Two unexpected observations were made. First the bandwidth was smaller than expected. For Na the occupied part of the band was found to be 2.65 eV broad (Lyo and Plummer 1988). This is consistent with an old angular-integrated measurement by Kowalczyk *et al* (1973). This bandwidth is substantially smaller than the width 3.2 eV predicted by nearly-free-electron (NFE) theory. Figure 4 shows the experimental peak position (crosses) as a function of the photon energy together with the prediction of NFE theory (dashed curve). The band narrowing is about a factor of two larger than the narrowing (0.27 eV) predicted by *GW* calculations for the electron gas of the appropriate density (Hedin 1965a). Although this error is not very large in absolute terms, it raised questions about the accuracy of the GWA for these systems. In addition, a set of essentially dispersionless states were observed close to the Fermi energy, which was completely unexpected according to the NFE bandstructure. This raised the interesting issue that the alkali metals may have charge density waves (Overhauser 1985).

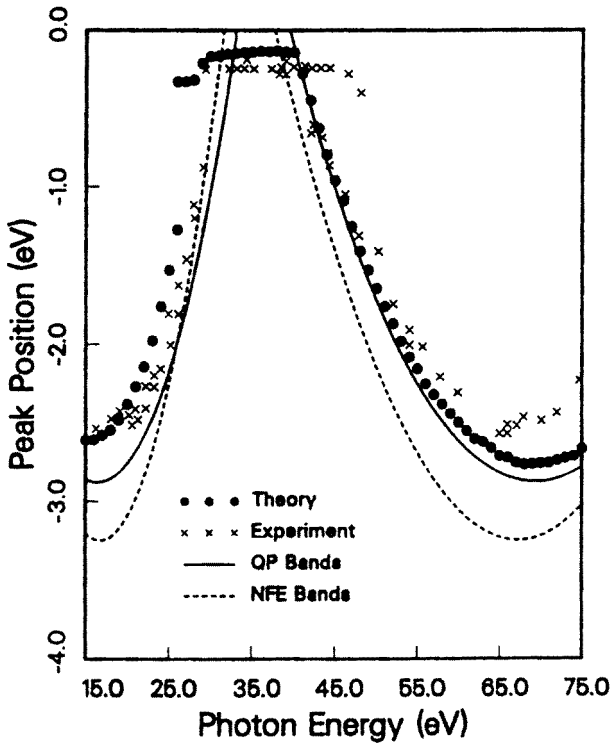

Figure 4. Peak position as a function of photon energy for photoemission at normal angle from the Na (110) surface. The crosses show experimental results (Jensen and Plummer 1985), the full circles show the photoemission calculations of Shung and Mahan (1987), the full curve the quasiparticle energies and the dashed curves the NFE theory. After Shung and Mahan (1987).

A GW calculation was performed by Northrup *et al* (1987, 1989) for Na metal and by Surh *et al* (1988) for K, using a generalized plasmon-pole approximation. These calculations gave essentially the same band narrowing (0.31 eV) for Na as a *GW* calculation for the electron gas. A much better agreement with experiment was, however, obtained by using

an improved dielectric function (Northrup *et al* 1987, 1989)

$$\epsilon^{-1} = 1 + v[1 - P(v + K^{xc})]^{-1}P \tag{136}$$

where P is the independent particle polarizability and $K^{xc} = \delta V^{xc}/\delta\rho$ with V^{xc} the (LDA) exchange-correlation potential and ρ the density. Equation (136) gives the appropriate dielectric function within the density functional formalism. Using this ϵ instead of the RPA ϵ, the bandwidth narrowing increased from 0.31 eV to 0.57 eV. Finally, a calculation was performed using a Green function with a certain self-consistency, namely with the LDA eigenvalues replaced by the quasiparticle energies. This led to a further increase of bandwidth narrowing to 0.71 eV. Lyo and Plummer (1988) also observed large effects when including the corrections to the dielectric function in equation (136).

The theoretical justification for including such corrections without simultaneously adding vertex corrections is, however, weak. We notice that the total energy of the system can be expressed in terms of the dielectric function. The quasiparticle energies can then be obtained by differentiating with respect to the occupation numbers (Rice 1965). If the dielectric function has the form of equation (136), it has been shown that there is also a vertex correction of the type (Rice 1965, Ting *et al* 1975, Mahan 1994)

$$\Gamma = \frac{1}{1 - PK^{xc}}. \tag{137}$$

Mahan and Sernelius (1989) have extensively tested various corrections to the dielectric function, including the corresponding vertex corrections when calculating the self-energy. They found that for the electron gas the vertex corrections cancel most of the effects of the corrections of the dielectric function, and the final results are rather close to the original GWA. Del Sole *et al* (1994) obtained similar conclusions for Si. The results for the electron gas suggest that there are discrepancies between appropriate self-energy calculations and the peak positions in the photoemission experiments.

The relation between the peak position in photoemission and the quasiparticle energies was studied by Mahan and coworkers in a series of papers, focusing on surface and quasiparticle lifetime effects (Shung and Mahan 1986, 1988, Shung *et al* 1987). They studied a model which includes the rapidly varying potential in the surface region as well as the lattice potential in the bulk. The effects of these potentials on the photoemission process were included, while the lattice part of the potential was neglected when calculating the states. The self-energy was calculated using the Rayleigh–Schrödinger perturbation theory, which gives results that differ slightly from the traditional GW results. For instance, the Na bandwidth is reduced by 0.37 eV due to this self-energy. Mahan and coworkers, furthermore, included the effects of the imaginary part of the self-energy of the emitted electron, by using the wavefunction $\phi^{>}$ for these electrons which decayed exponentially inside the surface. Finally, the self-energy of the initial state was included as well. Due to the exponential decay of $\phi^{>}$, the electron momentum perpendicular to the surface is not conserved, and there are non-vertical (non $k_{\perp}$-conserving) transitions.

Mahan and coworkers found that the broadening of the initial state and the instrumental resolution, as well as the interference between bulk and surface photoemission, shift the apparent peak positions towards lower binding energies relative to the quasiparticle energies by about 0.2–0.4 eV (Shung and Mahan 1988). Including these effects, as well as the quasiparticle self-energy shifts, leads to the full circles in figure 4. There is a substantial band narrowing relative to the quasiparticle energies (full curve) and the agreement with experiment is generally quite good. This suggests that the GWA gives good quasiparticle energies for Na, and that the main reason for the discrepancy between these energies

and the photoemission peak positions can be explained by considering the details of the photoemission process.

Mahan and coworkers also gave an explanation for the unexpected features close to the Fermi energy (see the essentially dispersionless structure in figure 4). These features occur for photon energies for which there are now vertical, energy-conserving transitions available. Due to the exponential decay of $\phi^{>}$, non-vertical but energy-conserving transitions are available. It has been shown that interference between surface and bulk emission puts most of the weight of these transitions close to the Fermi energy (Shung and Mahan 1986). Figure 4 illustrates that the theory can almost completely explain the experimental features, with just a few experimental points at larger photon energy unexplained. Thus, there seems to be no need to assume a charge-density wave to explain this structure.

5.2. Semiconductors and insulators: sp systems

The LDA systematically underestimates the bandgaps in semiconductors and insulators. In table 1 the calculated LDA bandgaps of some materials are compared with the experimental gaps. The discrepancies range from 30–100% and for Ge the LDA conduction and valence bands in fact overlap when relativistic corrections are included. Also, individual bands away from the Fermi level can be in error by up to 50%. It is natural to ask if the bandgap problem originates from the error in the LDA. Exchange-correlation potentials V^{xc} calculated from GW self-energies turn out to be similar to the LDA V^{xc} (Godby *et al* 1988) which indicates that even the exact V^{xc} probably does not give the correct gap, but this is still an open question. The V^{xc} may be a non-analytic function of the particle number (Almbladh and von Barth 1985b, Perdew and Levy 1983, Sham and Schlüter 1983). That is to say, V^{xc} with an extra electron, V^{xc}_{N+1}, may have an additional constant compared to V^{xc}_{N} and this constant may be large. Moreover, experience with empirical potentials shows that a local potential cannot, in general, give both the correct band structure and the ground-state electron density (Kane 1971).

In table 1 the bandgaps of some materials calculated within the GWA are shown. The agreement with experiment is very good, in most cases to within 0.1 eV. Although it is not strictly true, the self-energy correction is approximately an upward rigid shift of the conduction band relative to the valence band, the so-called scissor operator, i.e. cutting the bandstructure along the bandgap and shifting the conduction band rigidly upwards. The scissor operator is accurate to 0.1, 0.2, 0.2 and 0.4 eV in Si, GaAs, AlAs and diamond, respectively (Godby *et al* 1988). The validity of the scissor operator in Si is somewhat fortuitous, due to an almost complete cancellation between the strong energy dependence and non-locality.

The experimental values are obtained from optical measurements or photoemission and inverse photoemission experiments. The latter corresponds closer to the theoretical values whereas the former may contain excitonic binding energy, which should be subtracted off but unfortunately it is unknown. In optical experiments, the excited electron does not leave the system and may therefore form an exciton with the corresponding hole.

To discuss in more detail the features in the self-energy which are important for the quasiparticle energies, we consider Si as a prototype since it has been studied extensively. The main features are energy dependence and non-locality. We first consider non-locality within the COHSEX approximation. A measure of non-locality in the self-energy is its range, the distance $|\boldsymbol{r}-\boldsymbol{r}'|$ beyond which the self-energy is approximately zero. This range r_h is approximately given by the corresponding value for a jellium with the average density of Si ($r_s = 2$) and $r_h \sim 2r_s$. In Si more than 99% of the matrix element of the self-energy in

Table 1. Minimum bandgaps of semiconductors and insulators which have been calculated within the GWA. The energies are in eV.

	LDA	GWA	Experiment
AlAs	−1.37	2.18[a]	2.32[b]
$Al_{0.5}Ga_{0.5}As$	−1.12	2.06[c]	2.09[b]
AlN (wurtzite)	−3.9	5.8[d]	6.2[b]
AlN (zinc-blende)	−3.2	4.9[d]	
AlP	−1.52	2.59[c]	2.50[b]
AlSb	−0.99	1.64[c]	1.68[b]
CdS (zinc-blende)	−1.37[e], 0.83[f]	2.83[e], 2.45[f]	2.55[g,b]
CdS (wurtzite)	−1.36	2.79[e]	2.59[b]
CdSe (zinc-blende)	−0.76	2.01[e]	1.90[b]
CdSe (wurtzite)	−0.75	1.91[e]	1.97[b]
CdTe (zinc-blende)	−0.80	1.76[e]	1.92[b]
CdTe (wurtzite)	−0.85	1.80[e]	1.60[b]
Diamond	−3.90	5.6[h], 5.33[a], 5.67[i]	5.48[b]
GaAs	−0.67	1.58[a], 1.32[i], 1.22[c]	1.52[b], 1.63[j]
GaN (wurtzite)	−2.3	3.5[d]	3.5[b]
GaN (zinc-blende)	−2.1	3.1[d]	3.2[k], 3.3[l]
GaP	−1.82	2.55[c]	2.39[b]
GaSb	−0.10	0.62[c]	0.80[b]
Ge	< 0	0.75[h], 0.65[i]	0.744[b]
InAs	−0.39	0.40[c]	0.41[b]
$In_{0.53}Ga_{0.47}As$	−0.02	0.80[c]	0.81[b]
InP	−0.57	1.44[c]	1.42[b]
InSb	−0.51	0.18[c]	0.23[b]
LiCl	−6.0	9.1[h]	9.4[b]
Li_2O	−5.3	7.4[m]	6.6[n]
MgO	−5.0	7.7[o]	7.83[p]
Si	−0.52	1.29[h], 1.24[a], 1.25[i]	1.17[q]
SiC (β)	−1.31	2.34[i]	2.39[b]
ZnS (zinc-blende)	−2.37	3.98[e]	3.80[b]
ZnS (wurtzite)	−2.45	4.03[e]	3.92[b]
ZnSe (zinc-blende)	−1.45	2.84[e]	2.96[b]
ZnSe (wurtzite)	−1.43	2.75[e]	2.87[b]
ZnTe (zinc-blende)	−1.33	2.57[e]	2.71[b]
ZnTe (wurtzite)	−1.48	2.67[e]	

[a] Godby *et al* (1988).
[b] Hellwege and Madelung (1982).
[c] Zhu and Louie (1991).
[d] Rubio *et al* (1993).
[e] Zakharov *et al* (1994).
[f] Rohlfing *et al* (1995a).
[g] Cardona *et al* (1965).
[h] Hybertsen and Louie (1986).
[i] Rohlfing *et al* (1993).
[j] Aspnes (1976).
[k] Lei *et al* (1992a, b), Eddy *et al* (1993).
[l] Paisley *et al* (1989), Sitar *et al* (1992).
[m] Albrecht *et al* (1997).
[n] Rauch (1940).
[o] Schönberger and Aryasetiawan (1995).
[p] Whited *et al* (1973).
[q] Baldini and Bosacchi (1970).

a state ψ, $\langle\psi|\Sigma|\psi\rangle$, originates from $|\boldsymbol{r}-\boldsymbol{r}'| < r_h$. One expects non-locality to be important when r_h is comparable to or greater than the extent or wavelength of the wavefunction which is the case in Si. The matrix element in a non-local potential can be very sensitive to the nodal structure of ψ. This is contrary to the case when Σ is local, such as V^{xc}, since it is then $|\psi|^2$ that enters into the integral. In Si, non-locality has the effect of widening the gap. This can be understood as follows. The top of the valence band is bonding p whereas the bottom of the conduction band is antibonding p. Therefore non-locality has a larger effect on the conduction band than on the valence band since the antibonding state has an extra node which means a smaller wavelength than that of the bonding state. The presence of an extra node in the antibonding state reduces the matrix element $\langle\psi|\Sigma|\psi\rangle$ relative to $\langle\psi|V^{\mathrm{xc}}|\psi\rangle$ and therefore the conduction band is pushed upwards. Non-locality has a smaller effect on the valence state so the net effect is a widening of the gap (Godby *et al* 1988).

The non-locality arises from the density matrix (exchange charge)

$$\rho_{\mathrm{x}}(\boldsymbol{r},\boldsymbol{r}') = \sum_{i}^{\mathrm{occ}} \psi_i(\boldsymbol{r})\psi_i^*(\boldsymbol{r}') \tag{138}$$

and the screened potential $W(\boldsymbol{r},\boldsymbol{r}',\omega)$. In semiconductors, the screening is not complete because, due to the gap, a finite energy is required to excite a particle–hole pair. The range of the screened interaction and its nodal structure are then determined by ρ_{x}. Apart from non-locality, anisotropy also plays an important role. In a homogeneous system, ρ_{x} and W only depend on $|\boldsymbol{r}-\boldsymbol{r}'|$. In an inhomogeneous system, the screening potential $W - v$ as a function of $\boldsymbol{r}'$ may strongly depend on the location $\boldsymbol{r}$ of the test charge. In Si, for example, there is a large accumulation of charge in the bonding region as opposed to the antibonding region. It is to be expected that the screening potential of a test charge located in the bonding and antibonding regions will be very different, as shown in figure 5. This local field effect, which is entirely missing in the homogeneous case, is very important in covalent materials. Local fields are crucial in determining the strength of the screening hole but not its shape, and they contribute directly to the differing strengths of Σ at different points in the unit cell (figure 5) and therefore to the bandgap correction (Hybertsen and Louie 1986). In Si, local fields account for more than one-third of the screening potential in the region around the bond as can be seen in figure 5.

The local fields shift the centre of the screening potential, increase it in the bonding region and reduce it in the interstitial. The local field effect is one important feature in Si which distinguishes covalent-bond semiconductors from the alkalis. The screening potential in these materials is therefore considerably more complex than in the alkalis. Calculations for Si show that the local field effect in W is confined to a region near $\boldsymbol{r}$ as shown in figure 5. This is not surprising since the long-range part of the potential is contained in the diagonal element $W(\boldsymbol{q})$ corresponding to small $\boldsymbol{q}$. In the plane-wave basis, the local fields are described by the off-diagonal elements of the dielectric matrix. Nevertheless, $\Sigma(\boldsymbol{r},\boldsymbol{r}',\omega=0)$ is almost spherical as a function of $\boldsymbol{r}'$ for a fixed $\boldsymbol{r}$ and it is reasonably well reproduced by jellium Σ with the same average density as that of Si. This is illustrated in figure 6. However, the interaction of the wavefunction with the non-locality in Σ is not contained within the jellium model.

The local field effect has a large influence on Σ_{COH}. In fact, in a homogeneous system Σ_{COH} is a constant within the COHSEX approximation (equation (105)). Thus, if the local field effect is neglected, Σ_{COH} has no effects on the band dispersion. Σ_{COH} is deeper in the bonding region and shallower in the antibonding region as shown in figure 7. Since the valence state is concentrated in the bonding region and the conduction state in the

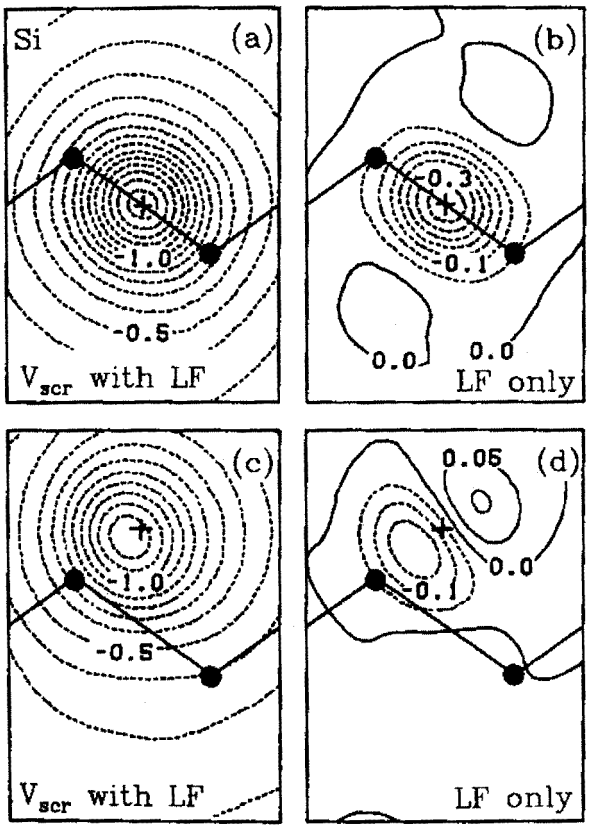

Figure 5. (a) and (c) show the screening potential in response to a single electron at $\boldsymbol{r}$ (indicated by +) in the $(1\bar{1}0)$ plane of Si in units of Ryd. (b) and (d) show the contribution from local fields only. The bond chain is indicated by a straight line. After Hybertsen and Louie (1986).

antibonding region, Σ_{COH} makes a large contribution to the gap. Σ_{COH} is, however, a local potential within the COHSEX approximation, a feature which is common to V^{xc} as opposed to Σ_{SEX} which is non-local (figure 8). This mechanism of gap opening by Σ_{COH} could, in principle, be accounted for by V^{xc}. The local field effect on Σ_{SEX}, on the other hand, is small, since the local field effect is rather localized and Σ_{SEX} is dominated by the large bare Coulomb interaction for small $|\boldsymbol{r} - \boldsymbol{r}'|$. The local field contribution to Σ_{SEX} is about 25% of that to Σ_{COH} but of opposite sign (Hybertsen and Louie 1986). Non-locality is essential in determining the correct quasiparticle energies and, in particular, the bandgap.

The COHSEX approximation is a static approximation to the self-energy. Strictly speaking, the quasiparticle energy should be obtained from the position of the peak in the spectral function. This procedure requires knowledge of the energy dependence of the self-energy, at least around the quasiparticle energy. The degree of the energy dependence is related to the renormalization factor Z (weight of the quasiparticle, equation (57)): the smaller Z, the stronger the energy dependence. The value of Z for semiconductors is 0.7 to 0.8 (Hybertsen and Louie 1986). The slope of the real part of the self-energy around the chemical potential is negative and the renormalization factor Z reduces the self-energy correction, as shown in equation (58). Thus, if the energy dependence of the self-energy is neglected, that is if the self-energy were calculated at the LDA instead of quasiparticle

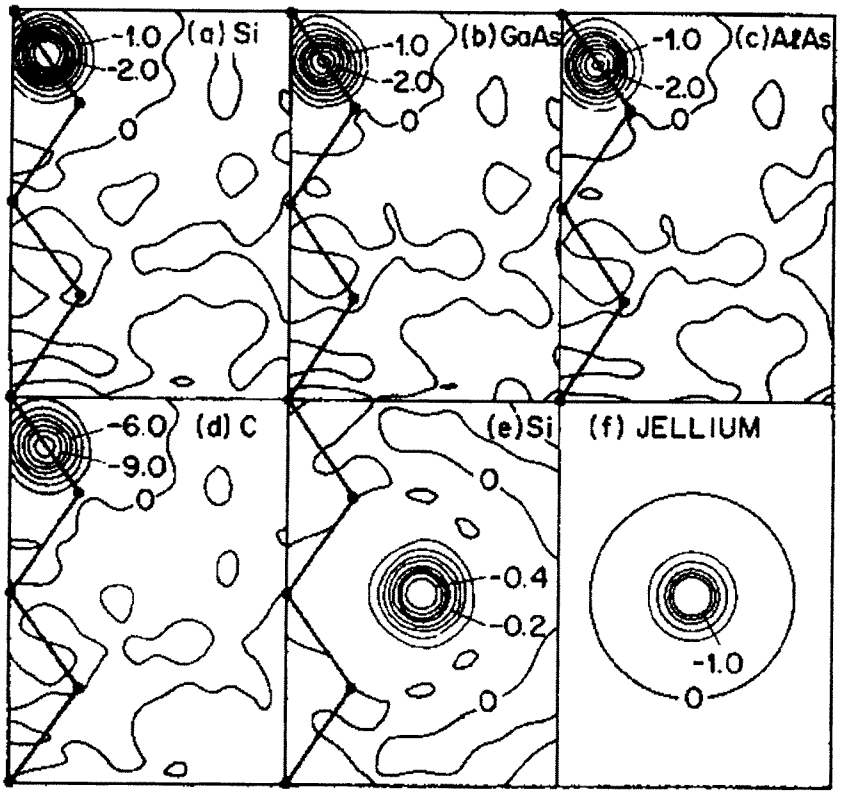

Figure 6. Contour plots of the self-energy $\Sigma(\boldsymbol{r}, \boldsymbol{r}', \omega = \text{midgap})$ in eV au^{-3} for $\boldsymbol{r}$ fixed at the bond centre and $\boldsymbol{r}'$ shown in the (110) plane for (a) Si, (b) GaAs, (c) AlAs and (d) diamond. For Si, the corresponding plot with $\boldsymbol{r}$ fixed at the tetrahedral interstitial site is also shown (e). For comparison, the self-energy operator of jellium with $r_s = 2.0$ (the average electron density of Si) is shown in (f) (from Hedin (1965a)). After Godby *et al* (1988).

energy, the bandgap correction would be overestimated. This is in agreement with the results of the COHSEX approximation which approximately corresponds to neglecting the renormalization factor Z. A similar conclusion is reached if the self-energy is approximated by its value at $\omega = 0$. In this case, the self-energy correction for the valence state would be underestimated whereas for the conduction state overestimated, leading again to an overestimated bandgap. A much better agreement is obtained if the self-energy correction $(\Sigma_{\text{COHSEX}} - V^{\text{xc}})$ is simply multiplied by Z. We note that the COHSEX approximation without local field effect generally gives a gap in better agreement with experiment, although not for Si. This means that local field effect, or in a certain sense non-locality, tends to cancel energy dependence. The importance of energy dependence is illustrated by plotting $\langle\psi_{\boldsymbol{k}n}|\Sigma(E_{\boldsymbol{k}n}) - V^{\text{xc}}|\psi_{\boldsymbol{k}n}\rangle$ and $\langle\psi_{\boldsymbol{k}n}|\Sigma(0) - V^{\text{xc}}|\psi_{\boldsymbol{k}n}\rangle$ as a function of $E_{\boldsymbol{k}n}$. The first quantity is very much like a step function (scissor operator) while the second quantity shows a strong energy dependence (figure 5(b) and 6(b) of Godby *et al* 1988). The effect of energy dependence is therefore to alter greatly the dispersion of the individual bands.

In general, the energy dependence of Σ leads to a strongly state-dependent self-energy correction $\Delta\Sigma = \Sigma - V^{\text{xc}}$ within each band as well as across the gap. The weak state

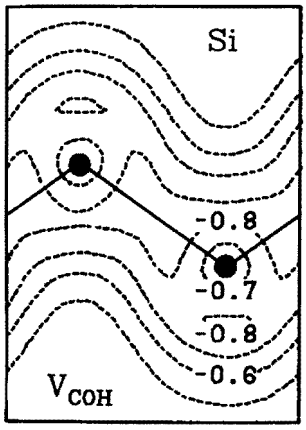

Figure 7. The local potential corresponding to the Coulomb hole part of the electron self-energy in the COHSEX approximation in the $(1\bar{1}0)$ plane of Si in units of Ryd. After Hybertsen and Louie (1986).

dependence of $\Delta\Sigma$ for Si within a band, resulting in a scissor operator, is therefore a coincidence. In diamond, for instance, the scissor operator approximation is not as good as in Si.

The highest occupied state in the exact DFT gives the exact ionization energy (Almbladh and von Barth 1985a). Assuming that the LDA V^{xc} is close to the exact one, we expect the self-energy correction to shift the conduction band but not the top of the valence band. If the RPA is used to obtain the electron gas data (von Barth and Hedin (1972) potential), it is indeed found that the top of the valence band is almost the same within the LDA and the GWA (Godby *et al* 1988). More accurate electron gas data obtained from *GW* calculations (Lundqvist and Samathiyakanit 1969) or quantum Monte Carlo simulations (Ceperley and Alder 1980), however, shift the top of the valence band upward by about 0.5 eV. Thus, vertex corrections (corrections beyond the GWA) may shift the *GW* bandstructure by 0.5 eV upwards (Godby *et al* 1988) or the LDA may not be sufficiently accurate.

Other authors have also repeated *GW* calculations for Si bandstructure with similar results (Hamada *et al* 1990, Rohlfing *et al* 1993). Calculations of the quasiparticle bandstructure and the bandgap of many semiconductors and insulators have been performed by a number of authors. Quasiparticle bandstructures of six II–VI compounds ZnS, ZnSe, ZnTe, CdS, CdSe and CdTe have been calculated by Zakharov *et al* (1994). We list in table 1 the results of these calculations. Quasiparticle energies for the F-centre defect in LiCl have also been calculated by Surh *et al* (1995) and self-energy calculations of carrier-induced bandgap narrowing in Si may be found in the work of Oschlies *et al* (1992, 1995). Berggren and Sernelius (1981) also studied the bandgap narrowing in doped Si and Ge as a function of impurity concentration. Application of the GWA to the metal–insulator transition of Si in diamond structure (Godby and Needs 1989) suggests that the metalization occurs at a much smaller volume than in the LDA which in turn indicates that the Fermi surface obtained from the Kohn–Sham DFT is not necessarily the same as that of the real system as shown by Schönhammer and Gunnarsson (1988) for model systems. Recently, application to a two-dimensional crystal found good agreement between the quasiparticle energies in the GWA and quantum Monte Carlo results (Engel *et al* 1995).

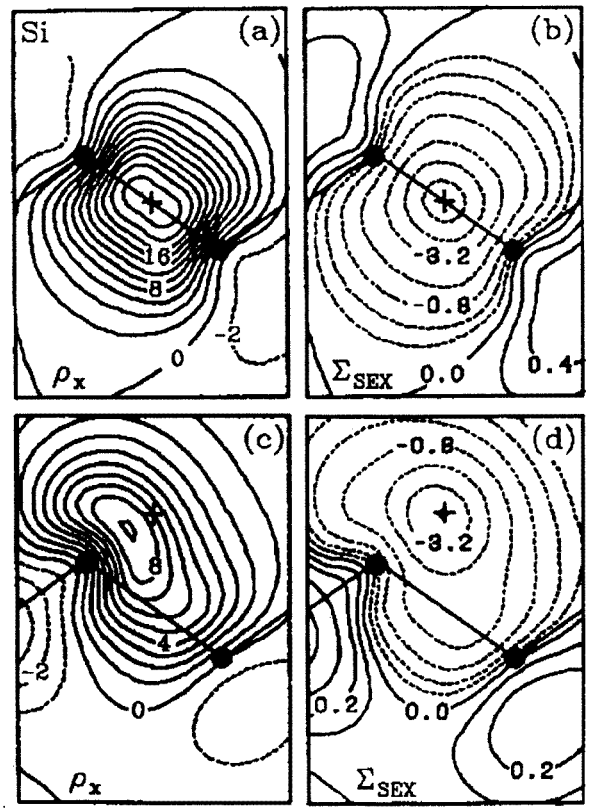

Figure 8. (a) and (c): The exchange charge in the $(1\bar{1}0)$ plane of Si with r' fixed (indicated by $+$) in units of electrons/cell. (b) and (d): The screened-exchange part of the self-energy $\times |r - r'|$ in the COHSEX approximation in units of au Ryd/cell. The contours increase in powers of two. After Hybertsen and Louie (1986).

5.2.1. Core polarization. A disadvantage of using pseudopotentials is a difficulty of including the core states in the calculation of the exchange potential as well as the polarization. Within a pseudopotential scheme, it is inevitable that core electrons are treated in an approximate manner. The core electron charge densities are frozen at their atomic values in the reference configuration used for constructing the pseudopotential. The core electrons and their potential are then eliminated so that there is no possibility for them to relax. Core relaxation gives rise to crystal field distortion and strong mixing between 3d and 3p states. These single-particle effects are small and can be included *a posteriori* by comparison with all-electron calculations. It is, however, important to take into account many-body effects arising from core relaxation since they can be large. In atoms with easily polarizable core the inclusion of core relaxation leads to an increase in the ionization energies, a contraction of the valence shell, a reduction of polarizabilities and oscillator strengths of the valence electrons.

The self-energy may be broken up into three terms (Hedin 1965b, Hedin and Lundqvist 1969):

$$\Sigma = \mathrm{i}G_c W + \mathrm{i}G_v W_v + \mathrm{i}G_v v R_c v \tag{139}$$

where $W = W_v + vR_cv$. The first term is the core–valence and core–core (screened)

exchange which is essentially the same as the bare exchange since screening is ineffective for small distances. The second term is the self-energy of the valence electron and the last term is the screened polarization potential due to the core electrons acting on the valence electrons. In pseudopotential calculations, the first and last term together are approximated by the LDA. The absolute contribution from these terms is estimated to be ∼1 eV for atomic Na (Hedin and Lundqvist 1969) and solid Al (Arbman and von Barth 1975) but the difference $\Sigma - V^{\mathrm{xc}}$, which is a more relevant quantity, is much smaller. For s–p semiconductors, this difference can be significantly larger. As a result, the calculated direct gaps and also the orderings and splittings of the conduction bands in Ge and GaAs, materials of technological interest, are in disagreement with experiment. It is important to get the orderings of the conduction band right since they affect the transport properties and lifetime of thermally excited carriers. Atomic calculations in Ge estimate the error for the 4s and 4p states to be 0.3 and 0.04 eV, respectively (Hybertsen and Louie 1986). If this error is taken into account, the result for the bandgap becomes even better. In transition metals it is crucial to include the core electrons in the calculations of the bare exchange since the error can be as large as ∼1 eV.

While the core-valence exchange can be treated straightforwardly in the Hartree–Fock theory, core-valence correlation is more complicated. A good approximation for taking into account core polarization is provided by the core polarization potential (CPP) method used in quantum chemistry (Müller *et al* 1984). The physical idea behind this method is that core polarization functions are characterized by sharp high-frequency excitations and are rather insensitive to the valence environment. This allows evaluation of the core polarization function for isolated atoms and neglect of frequency dependence. Consider a valence electron in the presence of a core. The electron polarizes the core resulting in polarization $\boldsymbol{p} = \alpha_c \nabla(1/r)$, where $\boldsymbol{r}$ is the position of the electron with respect to the core. From the classical theory of electrostatics, the electric field arising from this polarization at $\boldsymbol{r}'$ is given by

$$\boldsymbol{E}(\boldsymbol{r}') = \frac{3(\boldsymbol{p}\cdot\boldsymbol{r}')\boldsymbol{r}' - r'^2\boldsymbol{p}}{r'^5}. \tag{140}$$

The potential experienced by another electron at $\boldsymbol{r}'$ due to core polarization is then

$$V_{\mathrm{e\text{–}e}}(\boldsymbol{r},\boldsymbol{r}') = -\alpha_c \frac{\boldsymbol{r}\cdot\boldsymbol{r}'}{(rr')^3} \tag{141}$$

assuming that both r and r' are large. For a set of cores and many valence electrons, the core polarization potential is

$$V_{\mathrm{CPP}} = -\frac{1}{2}\sum_c \alpha_c \boldsymbol{f}_c \cdot \boldsymbol{f}_c \tag{142}$$

where $\boldsymbol{f}_c$ is the electric field acting on core c due to the valence charges at i and all other cores,

$$\boldsymbol{f}_c = \sum_i \frac{\boldsymbol{r}_{ci}}{r_{ci}^3} C(r_{ci}) - \sum_{c'\neq c} \frac{\boldsymbol{R}_{cc'}}{R_{cc'}^3} Z_{c'} \tag{143}$$

where C is a cut-off function, $Z_{c'}$ is the net charge of core c' and $\boldsymbol{R}_{cc'} = \boldsymbol{R}_{c'} - \boldsymbol{R}_c$. Inserting $\boldsymbol{f}_c$ into equation (142) yields (Müller *et al* 1984)

$$V_{\mathrm{CPP}} = -\frac{1}{2}\sum_c \alpha_c \left\{ \sum_i \frac{1}{r_{ci}^4} C^2(r_{ci}) + \sum_{i\neq j} \frac{\boldsymbol{r}_{ci}\cdot\boldsymbol{r}_{cj}}{r_{ci}^3 r_{cj}^3} C(r_{ci})C(r_{cj}) \right.$$
$$\left. -2\sum_i \sum_{c'\neq c} \frac{\boldsymbol{r}_{ci}\cdot\boldsymbol{R}_{cc'}}{r_{ci}^3 R_{cc'}^3} Z_{c'} C(r_{ci}) + \sum_{c',c''\neq c} \frac{\boldsymbol{R}_{cc'}\cdot\boldsymbol{R}_{cc''}}{R_{cc'}^3 R_{cc''}^3} Z_{c'} Z_{c''} \right\}. \tag{144}$$

The first term is the CPP in atoms with a single valence electron. Without the cut-off function, it would diverge at small r. The cut-off function C is semi-empirical with a parameter related to the size of the core (Biermann 1943, Biermann and Lübeck 1948, Bates 1947). The first term is a one-particle operator which is added to the pseudopotential. The second term is the two-electron interaction which was qualitatively discussed above. The third term is an indirect interaction of a valence electron with another core. This term is repulsive and cancels the attractive potentials from the first and last term. The latter is a core–core interaction and, together with the third term, essential to make sure of a vanishing long-range interaction with a neutral atom.

The effective interaction among the *valence* electrons is given by (Hedin and Lundqvist 1969)

$$W_C = v + v\sum_c R_c v + v\sum_c R_c v \sum_{c'\neq c} R_{c'} v + \cdots \tag{145}$$

R_c is the full or self-consistent response function of core c given by

$$v R_c v = V_{\mathrm{e-e}}. \tag{146}$$

Thus, the screened interaction becomes

$$\begin{aligned} W &= \epsilon^{-1} v \\ &= [1 - W_C P^0]^{-1} W_C \end{aligned} \tag{147}$$

where P^0 is the valence RPA polarization function. Using this CPP formalism, various transition energies for Si, Ge, AlAs and GaAs were calculated by Shirley *et al* (1992). The results for the fundamental bandgaps of Si, Ge, AlAs and GaAs shown in table 2 are in systematically better agreement with experiment compared with previous calculations where core relaxation effects were taken into account within the LDA only. The correction can be as large as 0.4 eV for the fundamental gap in GaAs. Notable also (not shown in the table) is the correct L–Γ–X ordering of conduction-band states in Ge obtained in the CPP approach. X–L and X_{6c}–X_{7c} splittings in GaAs are also improved (Shirley *et al* 1992).

Table 2. Fundamental bandgaps of Si, Ge, AlAs and GaAs calculated within the LDA and the GWA including core polarization within the CPP formalism (Shirley *et al* 1992) compared with experiment (Madelung 1984). Energies are in eV.

	LDA	CPP	Experiment
Si			
$\Gamma_{8v} \rightarrow 0.85X_{5c}$	1.29	1.16	1.17
Ge			
$\Gamma_{8v} \rightarrow \Gamma_{7c}$	0.53	0.85	0.89
$\Gamma_{8v} \rightarrow L_{6c}$	0.58	0.73	0.744
AlAs			
$\Gamma_{8v} \rightarrow X_{6c}$	2.09	2.01	2.24
GaAs			
$\Gamma_{8v} \rightarrow X_{6c}$	1.02	1.42	1.52

The CPP formalism is relatively easy to implement in many-body valence calculations without increasing computational cost significantly. Apart from its use within the pseudopotential approach, it can also be used in all-electron calculations with a frozen core.

5.3. Transition metals

The success of the GWA in semiconductors has encouraged applications to more complicated systems of transition metals and their compounds. From a numerical point of view, transition metals require a new approach for calculating the response function and the self-energy. The conventional plane-wave basis is not suitable in this case because the localized nature of the 3d states results in a prohibitively large number of plane waves. From a theoretical point of view, it has been argued that the localized nature of the 3d states makes the atomic approach more suitable for explaining the characteristic properties of transition metals. The presence of a Fermi surface as shown by de Haas–van Alphen measurements, on the other hand, strongly suggests that the itinerant character of the 3d electrons should be taken into account. Moreover, the bandwidths in transition metals are not too small and the ratio between the Hubbard U and the bandwidth is of order one. This is crucial when we consider screening of a photoemission hole where 3d electrons from neighbouring cells can take part in the screening, whereas such a possibility is absent in the atomic case. It then seems that an RPA type of approach such as the GWA is meaningful for these systems. GW calculations for transition metals have not been extensively performed. We concentrate, therefore, on two materials, Ni (Aryasetiawan 1992, Aryasetiawan and von Barth 1992) and NiO (Aryasetiawan and Gunnarsson 1995), for which full GW calculations have been done in some detail and on MnO for which a model GW calculation has been performed (Massidda *et al* 1995a).

5.3.1. Nickel. Among the transition elements, Ni is the most anomalous case in many respects. The LDA bandstructure deviates significantly from angle-resolved photoemission data. The occupied 3d bandwidth is 30% smaller than that of the LDA (3.3 eV against 4.5 eV) (Hüfner *et al* 1972, Himpsel *et al* 1979) and the experimental exchange splitting is half the LDA value (0.25–0.30 eV against 0.6 eV) (Eberhardt and Plummer 1980). Furthermore, there is the famous 6 eV satellite below the Fermi level (Hüfner *et al* 1972, Hüfner and Wertheim 1973, Kemeny and Shevchik 1975) which is entirely missing in the LDA or in any single-particle theory. These discrepancies are related to excited-state properties. An indication that single-particle theories would have difficulties in describing quasiparticle properties in Ni is the fact that the two lowest atomic configurations $3d^9 4s$ and $3d^8 4s^2$ are almost degenerate, differing by only 0.025 eV (Moore 1958). Another indication of many-body effects in Ni is the unusually large quasiparticle widths (Eberhardt and Plummer 1980)—up to 2 eV at the bottom of the 3d band—which implies strong interaction between the quasiparticles and the rest of the system resulting in short lifetimes. The photoemission process introduces an additional 3d hole to an already existing one, causing on-site many-body correlations not amenable to a single-particle treatment. This is in contrast to Cu where there is only one 3d hole after photoemission and where the LDA bandstructure is good, apart from a somewhat too high position (0.5 eV) of the 3d band relative to the 4s band (see, e.g., Jones and Gunnarsson 1989). Ground-state properties such as the equilibrium lattice constant, bulk modulus and magnetic moment are well reproduced by the LDA with the exception of the cohesive energy, where the LDA value is about 1 eV too large (Moruzzi *et al* 1978).

The GW results for Ni may be summarized as follows (Aryasetiawan 1992, Aryasetiawan and von Barth 1992). The LDA bandstructure is much improved, in particular the 3d bandwidth is narrowed by almost 1 eV, as shown in figure 9. The quasiparticle lifetimes are also given rather well by the GWA but the exchange splittings remain essentially unchanged from their LDA values and the 6 eV satellite is not reproduced.

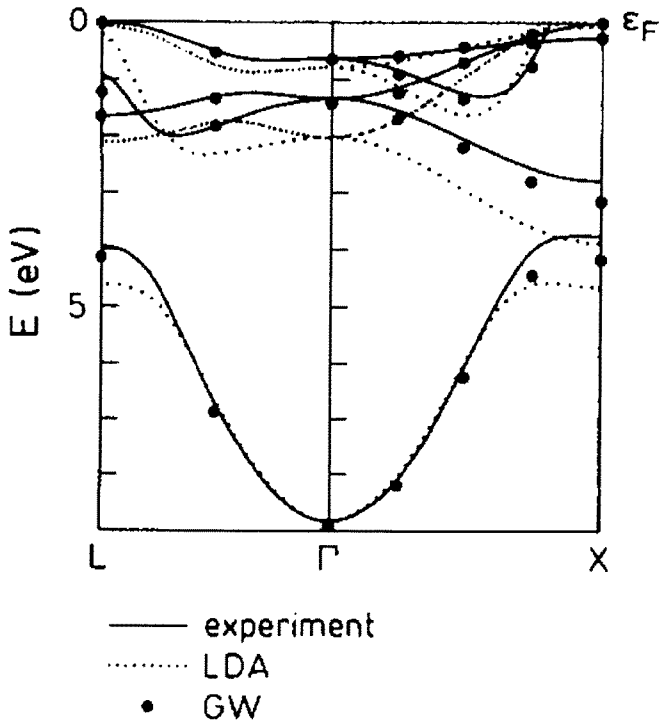

Figure 9. The bandstructure of Ni along $\Gamma = (0, 0, 0)$ and $X = (1, 0, 0)$ and along Γ and $L = (0.5, 0.5, 0.5)$. The full curves are the experiment and the dotted curves are the LDA (from Mårtensson and Nilsson 1984). The full circles are the quasiparticle energies in the GWA. After Aryasetiawan (1992) and Aryasetiawan and von Barth (1992).

The self-energies show a number of interesting features. As an illustration, the self-energy as a function of frequency for the Γ'_{25} state is shown in figure 10, later. The imaginary part of the self-energy is significantly more complicated compared with those of the alkalis or semiconductors. The latter are typically characterized by a large peak associated with a plasmon excitation but they otherwise show no other distinct structures. The frequency structure of the imaginary part of the self-energy is determined essentially by the imaginary part of the screened interaction W, as may be seen in equation (62). In alkalis or s–p semiconductors Im W is dominated by a plasmon peak which is then mirrored in Im Σ^c. Similarly in Ni, there is a strong similarity between Im W (figure 1) and Im Σ^c. In transition metals, there is no well defined plasmon excitation, rather it merges with the single-particle excitations forming a broad spectrum. There is a two-peak structure at about 20 and 30 eV which is probably due to plasmon excitation. An estimate based on the electron gas formula gives a plasmon energy of 30.8 eV when the 3d electrons are included in the density. This coincides rather well with the second large peak in the two-peak structure. A smaller structure at around 5–6 eV originates from transitions from the occupied valence band to states just above the Fermi level which constitute a large density of states. Interesting to observe is the behaviour of Im Σ^c at large frequencies. The hole and particle parts show similar behaviour and they therefore tend to cancel each other when one performs a Hilbert transform to obtain the real part of the self-energy. This justifies the use of energy cut-off in the calculation of the response function. It also agrees with our physical intuition that the

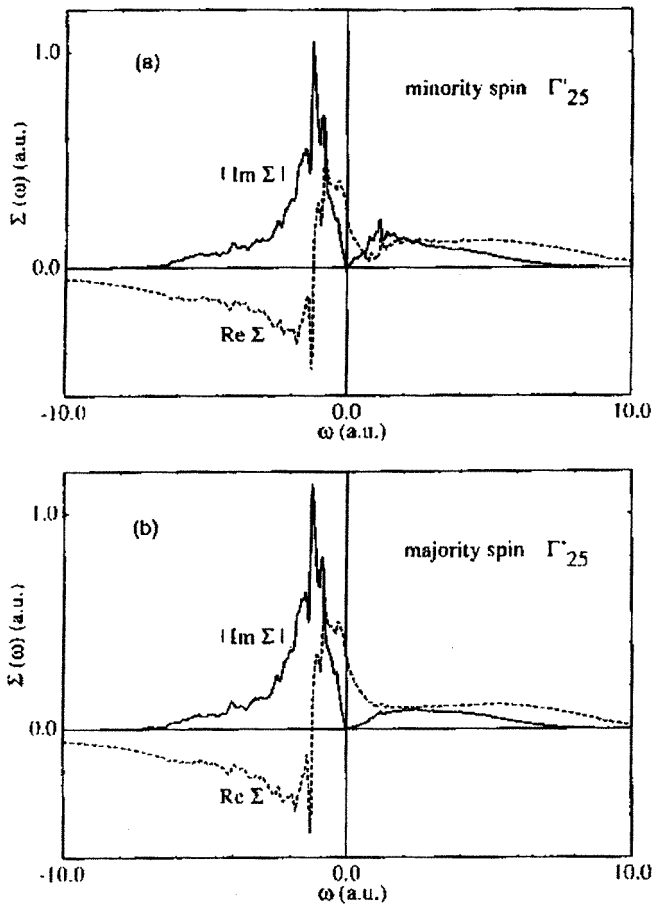

Figure 10. The real and imaginary parts of the correlation part of the self-energy for the minority and majority spin state Γ'_{25} (the lowest 3d state at the Γ point). The units are in au (Hartree = 27.2 eV). After Aryasetiawan (1992).

main contribution to the self-energy should come from energies up to the plasmon energy. For states lying a few eV below the Fermi level, the hole part (negative energy) of Im Σ^c has larger weight than the particle part (positive energy). This simply reflects the fact that the hole (occupied) states have larger overlap and correlation with other occupied states resulting in a larger correlated part of the self-energy for the hole part. As we go towards the Fermi level, the hole and particle parts become of almost equal weight. This is the reason why the self-energy correction for states at the bottom of the band is larger than for those at the top of the band, resulting in band narrowing. Im Σ^c around the Fermi level shows a quadratic Fermi liquid behaviour but it becomes linear rather quickly.

The real part of the self-energy is obtained by Hilbert transforming the imaginary part. A notable feature is a large derivative at around the Fermi level which implies large renormalization of the quasiparticle weights. A typical value for the renormalization factor is 0.5 for the 3d states (Aryasetiawan 1992). This is significantly smaller than in the electron gas (0.7) (Hedin and Lundqvist 1969) or semiconductors (0.8) (Hybertsen and Louie 1986) which reflects a larger loss of single-particle character of the quasiparticles.

The s states, on the other hand, have a renormalization factor ~ 0.7, comparable to those in the alkalis and semiconductors. It is in agreement with our physical picture that the 3d states are more correlated than the 4s–4p states.

In comparison with semiconductors, the self-energy corrections in Ni is considerably more complicated. While in semiconductors the self-energy correction is approximately a scissor operator which increases the bandgap by shifting the conduction band upwards and leaving the valence band unchanged, the self-energy correction in Ni shows a rather strong state and energy dependence. The self-energy correction can be positive or negative depending on the state and its magnitude varies throughout the Brillouin zone. For example, at the X-point the bottom of the 3d band experiences a self-energy correction of 0.8 eV while the top of the band is almost unchanged. The correction to the state Γ'_{25} is positive whereas at the L'_2 state it is negative. The state dependence of the self-energy correction is demonstrated clearly in figure 9 by the lowest valence band which is of 4s character at the bottom and a mixture between s and d at the top. The self-energy correction is approximately zero at the bottom of the band and positive at the band edges. The free-electron-like s states are well described by the LDA but the description of the more localized d states is less satisfactory.

As can be seen in figure 9, the 3d bandwidth is reduced by almost 1 eV. The exchange splittings on the other hand remain essentially unchanged from their LDA values. The discrepancy between the LDA exchange splitting and the experiment is rather small, 0.3 eV, which is slightly larger than the numerical accuracy (0.1–0.2 eV) but the results seem to indicate inadequacy in the GWA itself. GW calculations on transition metal atoms also show that strong correlations among 3d electrons of opposite spin are not well accounted for by the GWA (Shirley and Martin 1993). The quasiparticle widths or the inverse lifetimes are in reasonable agreement with available experimental data (Aryasetiawan 1992). The large width at the bottom of the 3d band indicates a strong interaction between the quasiparticles and the rest of the system.

As mentioned before, the 6 eV satellite is not reproduced in the GWA. That the 6 eV satellite is missing in the GWA can be seen directly in the imaginary part of the self-energy. For the existence of a satellite, there should be a strong peak at around 5–6 eV reflecting the presence of a stable excitation. However, as can be seen in figure 10, such a peak is missing.

3p-resonance photoemission measurements at 67 eV photon energy corresponding to the binding energy of the 3p core exhibit an asymmetric (Fano) resonant enhancement of the satellite and the main 3d line shows a strong antiresonance (Guillot *et al* 1977). This is explained as an Auger process where a 3p core electron is excited to fill the empty 3d states followed by a super Coster–Kronig decay

$$3p^6 3d^9 4s + \hbar\omega \rightarrow 3p^5 3d^{10} 4s \rightarrow 3p^6 3d^8 4s + e. \tag{148}$$

Although there is some indication that the 6 eV structure might arise from single-particle states (Kanski *et al* 1980) the 3p resonance photoemission provides strong evidence of the many-body character of the 6 eV satellite. The standard explanation for the presence of the satellite (Penn 1979, Liebsch 1979, 1981) is that a 3d hole created in a photoemission experiment introduces a strong perturbation due to its localized nature, causing another 3d electron to be excited to an empty state just above the Fermi level. In atomic picture, the state with two 3d holes correspond to the configuration $3d^7 4s^2$ which is separated from the main line configuration $3d^8 4s$ by more than 6 eV but which is reduced considerably by metallic screening. The two holes scatter each other repeatedly forming a 'bound state' at 6 eV. In a simple picture, the photon energy is used to emit a d electron and to excite another

into an empty d state so that the emitted electron appears to have a lower binding energy (satellite). The excited electrons mainly come from the bottom of the d band since they have the largest mixing with the s–p states and, therefore, according to the dipole selection rule, a large transition to the empty d states. This then has been argued as the source of band narrowing. The GW calculations, however, show that the largest contribution to band narrowing comes mainly from screening rather than two-hole interactions.

The reduction in the exchange splittings is related to the satellite. For simplicity we consider a two-level model with fully occupied majority channel and one occupied minority channel. A photoemission hole in the majority channel can induce another hole in the minority channel but not *vice versa* since the majority channel has no empty states. Thus, the effects of two-hole interactions are larger for the majority than for the minority channel resulting in a reduction in the exchange splitting. Calculations based on the Hubbard model within the T-matrix approach (Kanamori 1963) confirm this picture (Liebsch 1979, 1981, Penn 1979). These model calculations also assign the reduction in the bandwidth as originating from the two-hole interactions, which is not in complete agreement with the GW calculations. These results may be reconciled as follows. In the T-matrix calculations, it was found that there was no value of U that gave the correct satellite position and the bandwidth. To obtain the correct bandwidth, the value of U was such that the satellite energy became too large (Liebsch 1981). If we assume that the T-matrix theory gives the correct physical description for the satellite, the appropriate value of U would not give a large reduction in the bandwidth, but this is taken care of by the GWA. Thus one concludes that self-energy calculations which include the diagrams of the GWA and T-matrix theory may give both the correct bandwidth and the satellite structure. Diagrammatic comparison between the GWA and the T-matrix theory reveals that hole–hole interactions described by the T-matrix are not included in the GWA except to second order only. Direct comparison between GW calculations on real systems and Hubbard model calculations is, however, difficult if not impossible. This is because the Hubbard U cannot be easily related to the screened interaction in the GW calculations.

An extension of the T-matrix theory including Faddeev's three-body interaction (Faddeev 1963) was made by Igarashi (1983, 1985) and by Calandra and Manghi (1992). The theory has been applied to Ni (Igarashi *et al* 1994, Manghi *et al* 1997) and NiO (Manghi *et al* 1994).

5.3.2. Nickel oxide. NiO is a prototype of the Mott–Hubbard insulators. It was pointed out by Mott in the late 1940s that a system with an on-site Coulomb energy larger than the single-particle bandwidth tends to become an insulator and that single-particle theory is bound to give a wrong prediction for the state of the system (Mott 1949). Indeed, the LDA predicts NiO to be a metal when the calculation is performed in a paramagnetic state (Mattheiss 1972a, b). Slater (1974) suggested that a gap could be opened by an interplay between antiferromagnetism and crystal-field splittings. A detailed work along this direction can be found in the paper by Terakura *et al* (1984). The LDA does produce a gap in an antiferromagnetic state but of only 0.2 eV, in contrast to the experimental gap of 4.0 eV (Powell and Spicer 1970, Hüfner *et al* 1984, Sawatzky and Allen 1984). As expected, the free-electron like O p band is well described by the LDA but the magnetic moment is too small (1.0 μ_B) compared to experiment (1.7–1.9 μ_B) (Alperin 1962, Fender *et al* 1968, Cheetham and Hope 1983). Clearly there is something seriously wrong with the LDA. A more convincing evidence is provided by CoO, where the number of electrons in the paramagnetic structure is odd, making it impossible for any single-particle theory

to predict CoO as an insulator without doubling the unit cell. Experimentally, CoO is an antiferromagnetic insulator so that one might argue like Slater that single-particle theory could still give the correct prediction if the calculation is performed in an antiferromagnetic structure. One realizes, however, that the difference in magnetic energy distinguishing the paramagnetic and antiferromagnetic states is only a fraction of an eV, which is much smaller than the bandgap. This means that the result of a theoretical calculation for the bandgap should not depend on whether the calculation is performed in a paramagnetic or an antiferromagnetic structure.

The basic physics of the Mott–Hubbard insulators was explained by Mott several decades ago (Mott 1949). From the tight-binding limit, switching on hopping matrix elements causes the formation of a band of width w centred around the atomic eigenvalue. The possibility of occupying states with lower energy favours electron hopping but it costs a Coulomb energy U for an electron to hop from one site to the neighbouring site. If U is larger than w, the gain in kinetic energy is overwhelmed by the loss in Coulomb energy and the system prefers to be an insulator with a gap approximately given by U, splitting the lower and upper Hubbard bands. While the Mott picture is essentially correct, there are a number of experimental data which cannot be explained. The value of U, for instance, is estimated to be 8–10 eV which is much larger than the experimental gap. More recent studies initiated by Fujimori *et al* (1984) and Sawatzky and Allen (1984) based on the cluster approach and Anderson impurity model show that the gap in NiO is a charge-transfer gap. If an electron is removed from a Ni site the number of holes increases leading to a state with high energy due to an increase in the Coulomb interaction among the holes. The hole created on the Ni site may be filled by the transfer of an electron from an O site. Although it costs some energy transfer this leads to a state with a small binding energy. The states at the top of the valence band therefore have a large O p character. The lowest conduction state is of d character as in the Mott picture and the gap is therefore formed between the valence O p and conduction Ni d states. Much of the Ni d weight goes into a satellite located below the O p, opposite to the Mott and the Slater pictures. This model is able to explain experimental data which would otherwise be difficult to explain by the Mott and Slater pictures. The most convincing evidence supporting the charge-transfer picture is the 2p resonant photoemission experiment in CuO (Tjeng *et al* 1991) which has a similar electronic structure as that of NiO. In this experiment, a 2p core electron is excited and the remaining hole is subsequently filled by a valence electron. Since the dipole transition matrix element is largest between p and d states, resonance in the valence energy region can be identified as the position of d states which turns out to be below the O p band, rather than above as in the Mott and Slater pictures.

The LDA antiferromagnetic bandstructure is shown in figure 11. The highest valence state is formed by the majority e_g and minority t_{2g} Ni states and the lowest conduction band is formed by the minority e_g state. The GW calculation for NiO gives a gap of $\sim$5.5 eV (Aryasetiawan and Gunnarsson 1995). Starting from the LDA antiferromagnetic bandstructure with a gap of 0.2 eV, a GW calculation shifts the Ni e_g conduction upwards, increasing the gap to $\sim$1 eV. This upward shift of the e_g conduction band leads to a substantial change of the character of the wavefunctions, reducing the amount of minority e_g character in the occupied states. To include a limited self-consistency, we introduce a new non-interacting Hamiltonian H^0 with a non-local potential $\Delta|e_g\rangle\langle e_g|$, where Δ is chosen so that the bandgap obtained from H^0 agrees with the gap obtained from the previous GW calculation. This H^0 is then used to generate a new G^0 and a new self-energy in the GWA. This procedure is iterated to self-consistency. The non-local potential modifies the eigenvalues as well as the wavefunctions used to construct the zeroth-order Green function

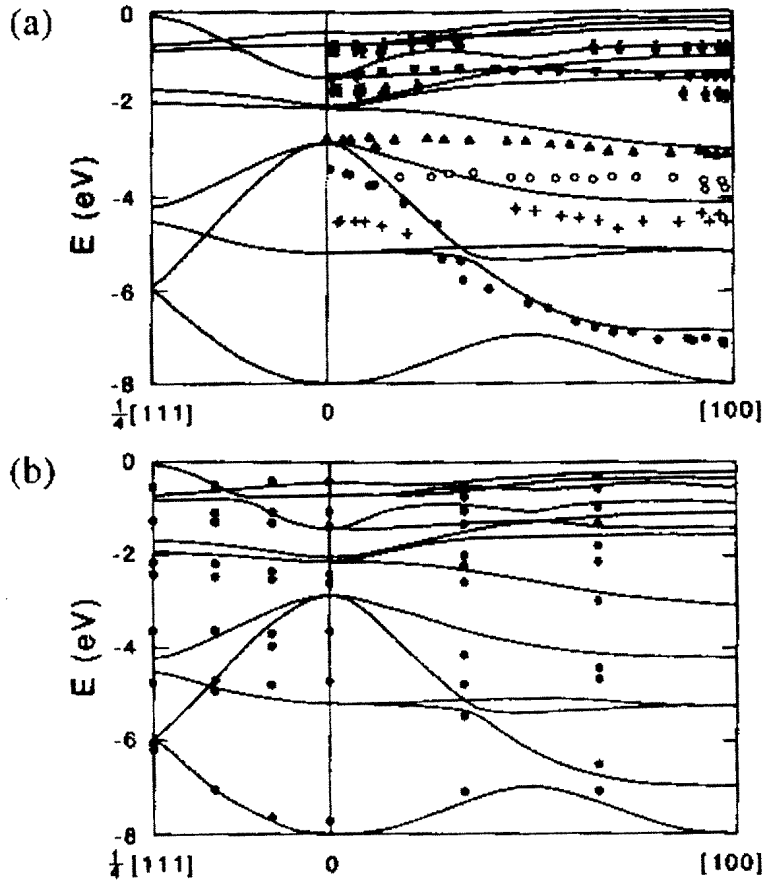

Figure 11. (a) Comparison between the LDA (full curve) and the experimental bandstructure of antiferromagnetic II NiO. From Shen *et al* (1990, 1991a, 1991b). (b) Comparison between the LDA bandstructure (full curve) and the quasiparticle bandstructure in the GWA (full circles) for NiO. After Aryasetiawan and Gunnarsson (1995).

G^0. The raising of the unoccupied majority e_g band by the self-energy correction reduces the hybridization with the O p band and has the effects of raising the bottom of the O p band and pushing down the top of the O p band at the Γ-point resulting in better agreement with photoemission data. In addition, the width of the unoccupied e_g band is reduced. The reduction in hybridization also reduces the magnitude of the exchange interaction of the e_g band with the occupied states which has the consequence of widening the gap. Thus, it is important that the wavefunctions are also modified in the self-consistent procedure. The final position of the unoccupied e_g band is just below the Ni 4s. As a check, the calculation has also been performed in the ferromagnetic state. A gap of $\sim$5.2 eV was obtained, close to the antiferromagnetic value. In contrast to the Slater model, the gap does not depend on the antiferromagnetic ordering and the results correctly predict that NiO remains an insulator above the Néel temperature. The *GW* calculation clearly improves the LDA gap markedly and it is in reasonable agreement with the experimental value of 4.0 eV. An estimate of the magnetic moment yields a value of 1.6 μ_B (Aryasetiawan and Gunnarsson 1995) in good agreement with the experimental value of 1.7–1.9 μ_B.

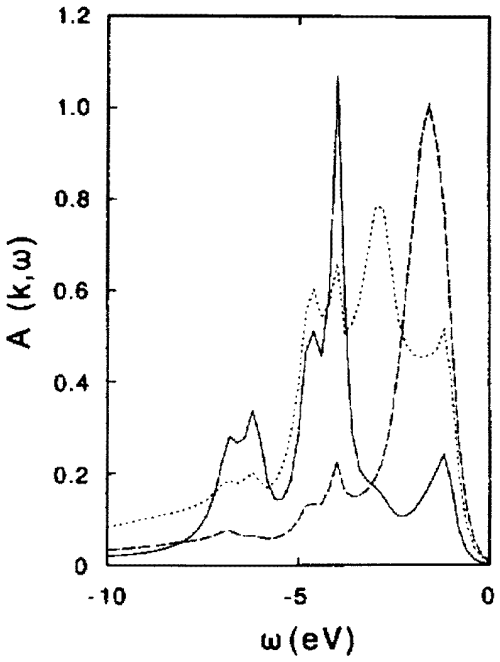

Figure 12. The spectral function of antiferromagnetic NiO at $\boldsymbol{k} = \frac{2}{3}(1, 0, 0)2\pi/a$ within the GWA projected on the O p orbitals (full curve) and the Ni d orbitals (the dotted curve corresponds to the majority spin and the dashed curve the minority spin). After Aryasetiawan and Gunnarsson (1995).

To study the character of the states at the top of the valence band, the spectral function has been calculated (figure 12). Projection of the spectrum into the Ni 3d and O p orbitals shows that there is an increase of the O p character at the top of the valence band but the main character is primarily Ni 3d. The satellite at -8 eV (Shen *et al* 1990, Shen *et al* 1991a, b) is not reproduced at the final self-consistent spectrum but a more detailed study of the satellite structure reveals that it is rather sensitive to the starting Hamiltonian (Aryasetiawan and Karlsson 1996). Starting from the LDA Hamiltonian in fact gives a satellite at about -10 eV but this satellite diminishes in intensity as the gap opens up. The origin of this behaviour can be traced back to the presence of a plasmon-like peak at low energy which is related to the incorrect LDA bandstructure. As the gap opens up, this peak structure becomes broadened and consequently the satellite structure diminishes. As in the case of Ni, it appears that the satellite structure is due to short-ranged correlations which are not properly taken into account by the RPA. The T-matrix approach could be appropriate and might remove some Ni d weight from the top of the valence band to the satellite region, but it is not clear how this could increase the charge transfer from the O p orbital.

5.3.3. Manganese oxide. A calculation on MnO based on a simplified GW scheme described in section 4 has been performed by Massidda *et al* (1995a). The electronic structure of MnO is the simplest among the transition metal oxides and in some respects similar to NiO. As in the case of NiO, the LDA gives a too small bandgap of 1.0 eV compared with the experimental value of 3.8–4.2 eV. The magnetic moment is also somewhat too small (LDA 4.3 μ_B, experiment 4.6–4.8 μ_B) although the relative discrepancy is not as large as in NiO. The larger LDA bandgap is due to the fact that in MnO the majority

spin is fully occupied and the minority spin is empty resulting in a large magnetic moment so that the exchange splitting is also large and dominates the ligand-field splitting and band broadening due to intersublattice coupling. In NiO the magnetic moment is smaller and the exchange splitting is comparable to the ligand-field splitting and band broadening.

The semi-empirical model GW calculation gives a gap of 4.2 eV which compares well with the experimental value of 3.8–4.2 eV. The LDA magnetic moment is also improved to 4.52 μ_B. There is an increase of O p character and a decrease of Ni 3d character at the top of the valence band, which are percentagewise large but small in absolute terms, so that the main character is still primarily Ni 3d (figure 13). The results are qualitatively similar to the full GW calculations on NiO described above. Calculation on $CaCuO_2$ using the model GW scheme has also been performed recently (Massidda *et al* 1997).

5.3.4. 3d semicore states. It is well known that the LDA eigenvalues for localized states are usually too high compared to experiment. The discrepancies can be several eV. For example, the Zn semicore 3d states in ZnSe are too high by 2.5 eV, the Ga semicore 3d states in GaAs by 4 eV and the Ge 3d semicore states by as much as 5 eV. In free atoms these deviations are even larger and in, for instance, a free Zn atom the 3d eigenvalue is about 6.5 eV too high. This raises interesting questions about whether or not GWA can describe these shallow core levels and why the LDA eigenvalues are substantially worse for free atoms than for solids, although the semicore states are almost completely localized.

For *deep* core levels, an important contribution to the GWA self-energy comes from the polarization (relaxation) of the more weakly bound electrons (Hedin and Lundqvist 1969, Lundqvist 1969). This relaxation is a classical electrostatic effect, which can be described in ΔSCF calculations, performing ground-state calculations with and without the core hole. Explicit ΔSCF calculations for free atoms were performed by Hedin and Johansson (1969), who obtained quite accurate results. This suggests that GWA may be rather accurate for such deep core levels, since it includes similar physics as the ΔSCF calculations.

It is not *a priori* clear that GWA can describe semicore states. Unlike the deep core states, the *shallow* core states are screened by states which are almost degenerate with the hole, which may introduce new effects. Calculations for semicore states in ZnSe, GaAs and Ge (Aryasetiawan and Gunnarsson 1996) showed, however, that GWA improves the LDA eigenvalues significantly also for these core states, leaving a discrepancy of only 0.5–1.0 eV from experimental values (too small a binding energy).

The errors in the LDA eigenvalues are to a large extent due the unphysical interaction of an electron with itself. This interaction is only incompletely cancelled by the exchange correlation potential (Gunnarsson *et al* 1974). This leads to an ionization energy which is too small. In addition, relaxation effects are neglected, which tends to overestimate the ionization energy and therefore partly cancels the error from the self-interaction. As discussed above, essential relaxation effects are included in ΔSCF calculations. In addition, the LDA total energies are rather accurate and the self-interaction errors are rather small. This is due to an exact sum rule satisfied by the LDA (Gunnarsson and Lundqvist 1976). While this sum rule is very important for total energy calculations, it does not necessarily imply a good exchange-correlation potential or accurate eigenvalues. It is therefore crucial that ΔSCF calculations involve total energy differences rather than eigenvalues.

To understand the different accuracy of the LDA eigenvalues for a solid and a free atom, Aryasetiawan and Gunnarsson (1996) performed transition-state calculations in the Slater transition state approach (Slater 1974). These calculations showed that in the solid the creation of a core hole led to a substantial charge transfer to the site where the hole

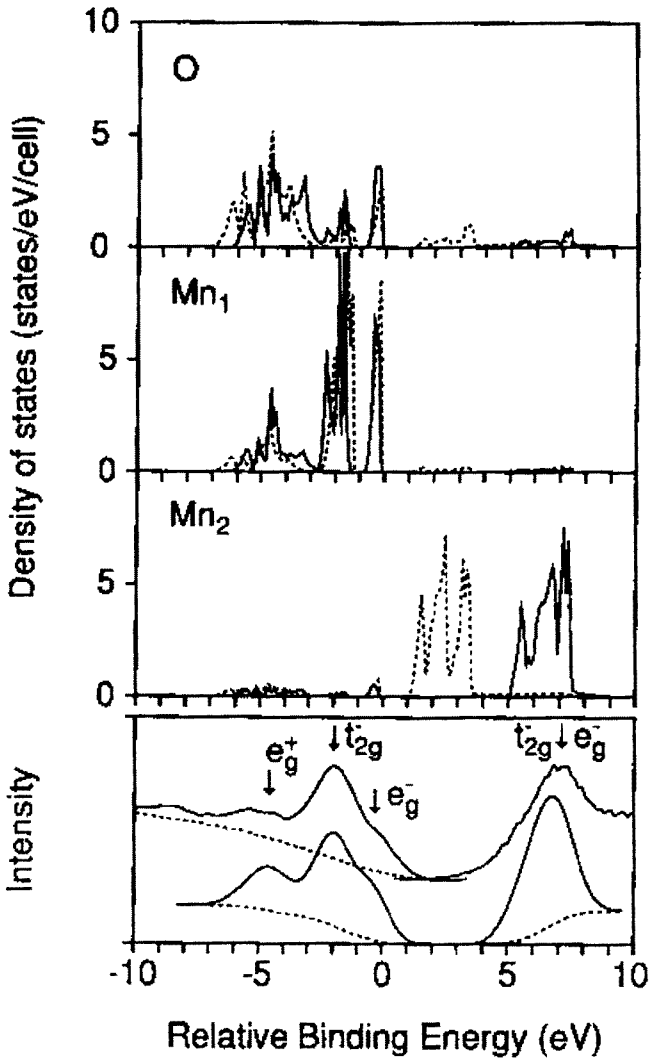

Figure 13. Partial densities of spin-up states of antiferromagnetic MnO. Top panel: The total O 2p projection into the O spheres. Second and third panels: d projection into the two inequivalent Mn spheres. Full (dashed) curves correspond to the model *GW* (spin polarized LDA) calculations. Bottom panel, upper curves: Inverse photoemission spectrum (from van Elp *et al* 1991) and the difference between the on- and off-resonance photoemission spectra (from Lad and Henrich 1988), representing the Mn contribution. Lower curves: *GW* d-projected density of states into both Mn spheres. Dotted curves illustrate the integrated-intensity background. After Massidda *et al* (1995a).

was created (Lang and Williams 1977, Zunger and Lindefelt 1983). Due to this charge transfer the relaxation energy is larger than for a free atom. The cancellation of the error in the eigenvalues due to the self-interaction and the error due to the neglect of relaxation effects therefore becomes more complete in the solid than in a free atom. This explains why the errors in the eigenvalues are much smaller in the solid (Aryasetiawan and Gunnarsson 1996).

The ΔSCF gave ionization energies for the 3d semicore states which were about 1 eV too large. The discrepancy between ΔSCF and GWA may be due to an error in the LDA, which tends to overestimate the exchange-correlation energy between a 3d electron and the 3s–3p core (Gunnarsson and Jones 1980) which is about 1 eV in the series K–Cu (Harris and Jones 1978). It could also be due to the difference in the RPA screening and the static LDA screening. Transition-state calculations for a number of Zn compounds have also been performed by Zhang *et al* (1995). We observe, however, that the Slater transition-state approach requires that the semicore states are sufficiently localized to form a bound state in the transition-state calculations. The bound state has no dispersion, in contrast to the *GW* results which show the full bandstructure. Thus, if a bound state is not formed in the transition-state calculation, the result would be identical to that of the LDA. The problem can be illustrated for Cu metal where the 3d band is 0.5 eV too high. The transition state would give a single number rather than a band assuming that a bound state is formed in the first place. A *GW* calculation, on the other hand, lowers the position of the band by 0.2–0.3 eV while maintaining the LDA bandstructure (Aryasetiawan and Gunnarsson 1997).

A model *GW* calculation for ZnO by Massidda *et al* (1995b) also improved the LDA result from −5.4 eV to −6.4 eV, but a significant error remains when compared to experimental results −8.6 eV and −7.5 eV.

5.4. Surfaces

5.4.1. Jellium surface. *GW* calculations for a model jellium–vacuum interface have been performed in some detail. The problem of interest here is what the surface barrier looks like. According to classical electrostatics, the image potential seen by an electron in the vacuum far from the surface behaves like

$$V_{\text{image}} = -\frac{1}{4(z - z_0)} \tag{149}$$

where z is the coordinate normal to the surface and z_0 is the position of the effective image plane. In the LDA, the image potential is known to decay exponentially (Lang and Kohn 1973). This unphysical behaviour is due to the fact that in the LDA the exchange-correlation hole is determined by the local density and it does not feel the surface directly. The exchange-correlation hole obeys a sum rule that it must integrate to one but since the density is small outside the surface it means that the hole becomes very extended. In fact, about half the hole resides far inside the surface. Due to the very extended structure of the hole, the resulting V^{xc} decays exponentially outside the surface. The situation is similar to the atomic case where the LDA V^{xc} decays exponentially instead of decaying as $1/r$ as in the exact V^{xc} (Almbladh and von Barth 1985a). Similarly, the exact DFT exchange-correlation potential should be capable of reproducing the correct behaviour of the image potential which is important in many applications. For example, the LDA potential cannot produce the Rydberg series of image states. Binding energies and the lifetime of surface states bound by the image potential depend crucially on the $1/z$ dependence, as do tunnelling currents in the scanning-tunnelling microscope and positions of alkali ions adsorbed on metal substrates. Interpretation of inverse photoemission data relies on the existence of the image tail of the surface barrier.

By calculating the self-energy for the surface, V^{xc} can be derived simply by taking the trace of the following Dyson equation (Sham and Schlüter 1983, 1985, van Leeuwen 1996)

$$G = G^{\text{DF}} + G^{\text{DF}}(\Sigma - V^{\text{xc}})G \tag{150}$$

and noting that the exact density is equal to the exact DF density (i.e. the trace of G is equal to the trace of G^{DF}). This yields

$$\int \mathrm{d}^3 r_1\, V^{\mathrm{xc}}(\boldsymbol{r}_1) \int \mathrm{d}\omega\, G^{\mathrm{DF}}(\boldsymbol{r}, \boldsymbol{r}_1, \omega) G(\boldsymbol{r}_1, \boldsymbol{r}, \omega)$$
$$= \int \mathrm{d}^3 r_1 \int \mathrm{d}^3 r_2 \int \mathrm{d}\omega\, G^{\mathrm{DF}}(\boldsymbol{r}, \boldsymbol{r}_1, \omega) \Sigma(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega) G(\boldsymbol{r}_2, \boldsymbol{r}, \omega). \tag{151}$$

The exchange-correlation potential obtained from the above equation using the GW self-energy exhibits the correct asymptotic behaviour as can be seen in figure 14. It can also be seen from the figure that it is the correlation potential originating from the Coulomb correlation that determines the asymptotic behaviour of the image potential. The exchange potential, on the other hand, is numerically found to decay as $v^x \sim -a/z^2$ and to contribute significantly to the determination of z_0. The position of the effective image plane z_0 deduced from V^{xc} is therefore different from the one for a classical test charge. Thus, for $r_s = 2.07$ the value of z_0 deduced from the image tail of V^{xc} is $z_0 = 0.72 \pm 0.1$ au measured from the jellium edge, while the value obtained from a linear response calculation is $z_0 = 1.49$ au (Eguiluz *et al* 1992).

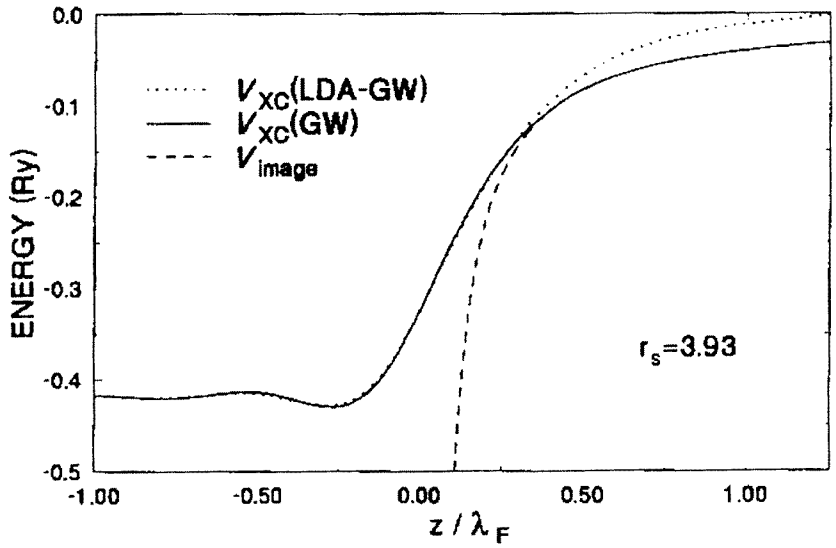

Figure 14. $V^{\mathrm{xc}}(z)$ at a jellium surface for $r_s = 3.93$ ($\lambda_{\mathrm{F}} = 12.9$ au). The full curve is the solution to equation (151) using the GWA for Σ, and the dotted curve is the corresponding LDA potential. The dashed curve is the image potential $V_{\mathrm{image}}(z) = -1/4(z - z_0)$. After Eguiluz *et al* (1992).

The Kohn–Sham eigenvalues calculated from the V^{xc} deduced from Σ_{GW} turn out to be very close to the quasiparticle energies obtained from the Dyson equation (Deisz *et al* 1993). The difference $E_{\mathrm{QP}} - E_{\mathrm{KS}} = 0.02$ eV for $q_{\parallel} = 0$, which is much smaller than the binding energy of the state which is about 0.5 eV. Furthermore, the Kohn–Sham and quasiparticle eigenfunctions are practically identical (the overlap is greater than 0.999). This does not, however, imply that the physics of the self-energy at the surface can be completely described by a local potential. The imaginary part of the self-energy associated with damping is intrinsically non-local and energy dependent. It cannot be mimicked by a local complex potential. As the electron moves away from the surface, the maximum

of Im Σ remains at the surface, reflecting a high degree of non-locality. The range of this non-locality is comparable with the decay length of the electron density in the vacuum (Deisz *et al* 1993). It has also been found that Im Σ deviates from the quadratic behaviour $\simeq (E - E_F)^2$ when the electron is in the vacuum outside the surface and the departures grow as the electron–surface separation increases. The effect is attributed to the suppression of one-electron decay channels near the Fermi level (Deisz and Eguiluz 1997).

5.4.2. Si(1 0 0) surface. There have also been studies for more detailed microscopic models of surfaces. In particular, the Si(100) surface has been extensively studied experimentally and theoretically, because of the technological importance of Si. At room temperature there is a 2×1 reconstruction which at low temperatures goes over in a $c(4\times2)$ reconstruction (see, e.g., Johansson *et al* 1990 for references to experiment). The surface atoms are believed to form buckled dimers, where one atom moves out of the surface and the other into the surface. The room temperature photoemission spectrum has been measured by Johansson *et al* (1990).

The electronic structure has been calculated in the GWA by Northrup (1993) ($c(4\times2)$), by Kress *et al* (1994) (2×1) and by Rohlfing *et al* (1995b) (2×1). In figure 15 we show the results of Rohlfing *et al* (1995b). These calculations were performed using a Gaussian basis set and a super cell containing eight Si layers. The figure shows the LDA bulk bands (dashed) and *GW* surface states. The *GW* surface band agrees rather well with an experimental band, while another experimental band has no correspondence in the theoretical calculation. It is interesting that the calculation of Northrup, using the low-temperature structure, produced two bands in close agreement with experiment. Actually, the experimental samples may have some domains with this $c(4\times2)$ structure (Johansson *et al* 1990). The *GW* calculation shifts the conduction band by about $+0.50$ to $+0.65$ eV relative to the valence band (Rohlfing *et al* 1995b). According to optical experiments, the bottom of the valence band at Γ is 1.1 eV above the top of the valence band in good agreement with the theoretical result 0.95 eV. For the indirect bandgap there are experimental estimates in the range 0.44–0.9 eV compared with the theoretical result 0.7 eV (Rohlfing *et al* 1995b).

In this context we also mention that there has been work within the GWA for interfaces, for example, for calculating Schottky barriers. This work will not be discussed further here, but we refer the reader to Charlesworth *et al* (1993) and Godby and Sham (1994) for an example of such work and further references. For other work on semiconductor surfaces we refer to Hybertsen and Louie (1988b) and Bechstedt and Del Sole (1990). Rohlfing *et al* (1996) performed calculations on a clean, H and S terminated Ge(001) surface and Zhang *et al* (1988) have calculated the valence band off-set of AlAs–GaAs(001).

5.5. Clusters

Electronic excitations and optical spectra in clusters have been studied mainly within the configuration interaction approach (Bonacic-Koutecky *et al* 1990a, b, 1992). While it gives accurate results, its applications are limited to systems containing a small number of electrons, typically less than 10. GWA provides an alternative for calculating excitation properties in clusters with relatively large numbers of electrons which cannot be handled by the configuration interaction method.

GW calculations have been performed for a jellium-sphere model for alkali metals (Saito *et al* 1989) and more recently for the real cluster Na_4 (Onida *et al* 1995). In alkali-metal clusters, it is known that the ionization energies calculated within the LDA are too

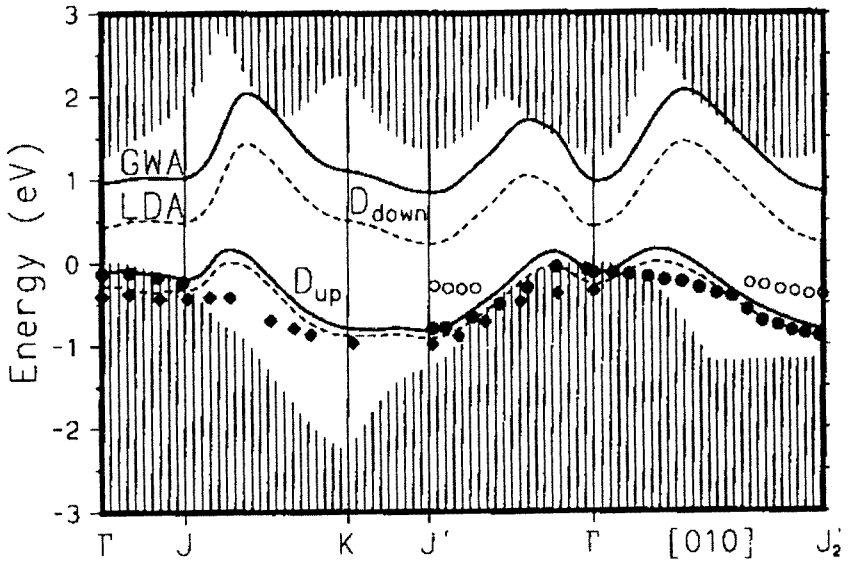

Figure 15. Calculated dangling-bond bands. Full curves, GWA energies; dashed curves, LDA energies. The experimental results are shown by diamonds (Uhrberg *et al* 1981) and circles (full and open) (Johansson *et al* 1990). (After Rohlfing *et al* 1995b).

low compared to experiment (Ishii *et al* 1986, Saito and Cohen 1988) and the discrepancy becomes worse the smaller the cluster. The size dependence of the ionization energy in the LDA is too weak. This discrepancy is attributed to self-interaction which is not taken into account properly in LDA eigenvalues. That the discrepancy gets worse for smaller clusters is intuitively clear, since the larger the clusters the more they resemble an electron gas on which the LDA is based. On the other hand, the LDA gives the correct sequence of eigenvalues for the valences states in the jellium-sphere model which is 1s, 1p, 1d, 2s, 1f, 2p, etc (Saito *et al* 1989).

In the jellium-sphere model, the ionic charges are smeared to form a sphere of a uniform positive background. The *GW* calculation for this model corresponding to Na_{20} lowers the occupied LDA eigenvalues and increases the unoccupied ones, thus increasing the energy gap similar to the results for bulk semiconductors and insulators. The ionization energy as a function of cluster size is also in better agreement with experiment although the absolute values are somewhat too large (Saito *et al* 1989). This could be due to the absence of core polarization in the jellium-sphere model which gives a positive contribution to the self-energy. The GW results are actually rather close to the HF results, the reason being that screening due to long-range correlations is much less important in a small finite system than in a solid (Saito *et al* 1989).

The *GW* calculation for real Na_4 yields similar results (Onida *et al* 1995). The calculated absorption spectra are shown in figure 16. The empty states, including the lowest unoccupied molecular orbital (LUMO), are raised by between 0.75 and 0.90 eV while the highest occupied molecular orbital (HOMO) and the lowest occupied molecular orbital state are lowered by 1.55 and 1.40 eV, respectively, giving a HOMO–LUMO gap of 3.0 eV compared with the LDA value of 0.55 eV. Again the *GW* result is close to the HF value of 3.4 eV for the reason discussed above. Indeed, the calculation of the static dielectric function for this cluster shows a metallic behaviour for small distances but it drops to less than unity starting

at about 6 au. For comparison, the value of the dielectric function for metallic sodium at 5 au is about 50 whereas in the cluster it is 1.2. This phenomenon of antiscreening is typical of a small system. If a positive test charge is introduced at the centre of the cluster, say, electrons will surround the test charge and the surface of the cluster will therefore be positively charged. Thus, as the distance from the centre increases, the screening quickly vanishes and becomes negative (Onida *et al* 1995). A similar behaviour is also observed in the C_{60} molecules (Gunnarsson 1992).

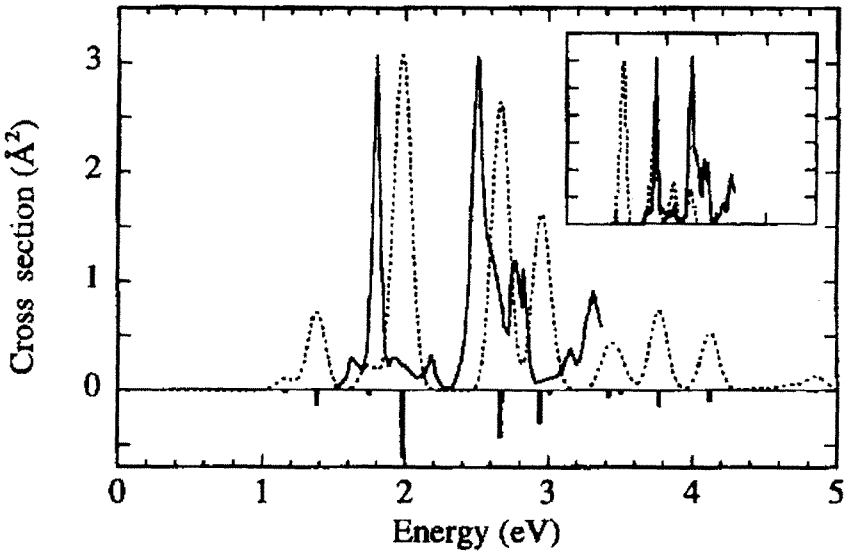

Figure 16. Calculated absorption spectra of Na_4 including self-energy and excitonic effects (dotted curve) in arbitrary units and using a Gaussian broadening of 0.06 eV. The full curve is the experimental photodepletion cross sections from Wang *et al* (1990). The vertical bars show the unbroadened spectrum and the inset shows the LDA results. After Onida *et al* (1995).

5.6. Fullerenes

Solid C_{60} (fullerite) is a molecular solid, where the orbitals of a free C_{60} molecule essentially keep their character and the hopping between the molecules only leads to a small broadening of the discrete molecular states. The bandstructure of C_{60} is shown in figure 17. In undoped C_{60} the H_u band is occupied and the T_{1u} band is empty. The LDA (figure 17(a)) bandgap is about 1 eV (Erwin and Pickett 1991, Troullier and Martins 1992, Satpathy *et al* 1992), which is substantially smaller than the experimental value, 2.3 eV (Lof *et al* 1992), obtained from photoemission and inverse photoemission.

Shirley and Louie (1993) have performed a GW calculation for solid C_{60} on a face-centred cubic (fcc) lattice with all molecules having the same orientation ($Fm\bar{3}$ structure). This differs somewhat from the experimental $T = 0$ structure $Pa3$, where the molecules take four different orientations. This difference should not be important for the present discussion. Shirley and Louie (1993) used the Levine and Louie (1982) model for the static dielectric function together with a plasmon-pole approximation. Their results are shown in figure 17. The bandgap is increased to 2.15 eV in good agreement with the experimental results. The bandwidths were also increased by about 30%. There are no reliable experimental results

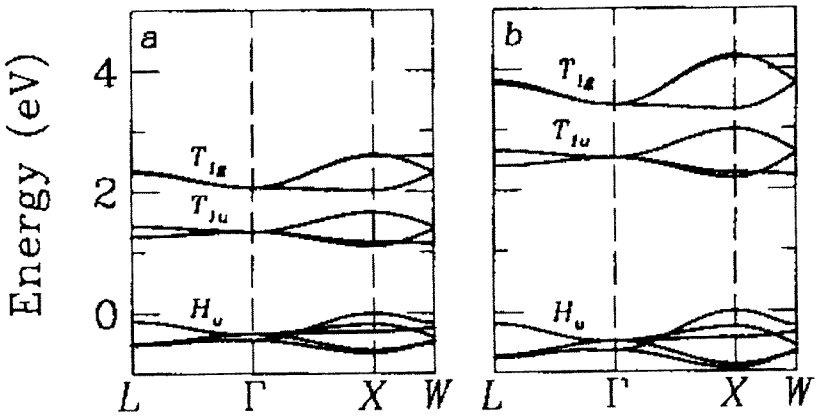

Figure 17. LDA (a) and GW (b) bandstructures for solid C_{60} in the $Fm\bar{3}$ structure. After Shirley and Louie (1993).

for the dispersional bandwidths. It was concluded that undoped solid C_{60} is a standard band insulator with dispersive bands.

The calculation by Shirley and Louie was performed for the undoped C_{60} solid. In A_3C_{60} (A = K, Rb) the T_{1u} band is half-filled. The Hubbard $U \sim 1.5$ eV (Lof *et al* 1992), describing the interaction between two electrons on the same molecule, is large compared with the width ($W \sim 0.6$ eV) of the partly-filled T_{1u} band. One may, therefore, expect strong correlation effects for these systems. Although it is not clear if the GW approximation can describe such strong correlation effects, it is interesting to perform such a calculation.

To screen the Hubbard U, Gunnarsson (1997) considered a model dielectric function

$$\epsilon(\boldsymbol{q}, \omega) = \epsilon_0 - \frac{\omega_{0\mathrm{p}}^2}{m^*\omega^2} \tag{152}$$

which describes the coupling to a plasmon at $\omega = \omega_{0\mathrm{p}}/\sqrt{m^*\epsilon_0}$. This plasmon is due to the oscillations of the three electrons donated by the alkali atoms to the T_{1u} band. The model (152) is therefore complimentary to the calculation of Shirley and Louie (1993), since this T_{1u} plasmon does not exist for the undoped system, while the physics considered by Shirley and Louie is neglected in equation (152). This model of the dielectric function was combined with a tight-binding (TB) model (Satpathy *et al* 1992) for the bandstructure, which reproduces the LDA T_{1u} band well.

First, an HF calculation was performed for this model. It was found that in the HFA the width of the T_{1u} band is increased by about 75%. Next a GW calculation was performed. Including the coupling to the T_{1u} plasmons was found to reduce the T_{1u} width to a value 35% smaller than the original TB width. In fact, for a model of this type one can show under rather general assumptions that the bandwidth is always reduced if there are no other bands above or below the band considered (Gunnarsson 1997). The density of the T_{1u} electrons is very small and corresponds to the electron gas parameter $r_s \sim 7$. Thus the density is substantially smaller than for the free-electron-like metals. It is then not too surprising that the quasiparticle weight $Z \sim 0.4$–0.5 is also smaller than for these metals. This means that much of the spectral weight appears in satellites. If the narrowing of the T_{1u} band due to the coupling to the T_{1u} plasmon is combined with the broadening found by Shirley and Louie due to other couplings, the net result is a small change of the bandwidth.

6. Self-consistency

The set of Hedin's equations (43)–(46), in the original formulation of the self-energy expansion in powers of the screened interaction W, constitutes a self-consistent cycle (Hedin 1965a, Hedin and Lundqvist 1969). Within the GWA, starting from a (usually) non-interacting Green function G_0, one calculates the polarization function $P_0 = -\mathrm{i}G_0G_0$ and the corresponding screened interaction W_0. The self-energy is then obtained from $\Sigma_0 = \mathrm{i}G_0W_0$. In most GW calculations that have been performed so far, Σ_0 is taken to be the final self-energy. The interacting Green function G obtained from the Dyson equation $G = G_0 + G_0\Sigma_0G$ is, however, not necessarily the same as G_0. To achieve self-consistency, the Green function obtained from the Dyson equation should be used to form a new polarization function $P = -\mathrm{i}GG$, a new screened interaction W and a new self-energy Σ which in turns yields a new Green function through the Dyson equation. This process is continued until G obtained from the Dyson equation is the same as G used to calculate the self-energy. Self-consistency is evidently an important issue since it guarantees that the final results are independent of the starting Green function. Moreover, according to Baym–Kadanoff theory (1961), a self-consistent GW scheme ensures conservation of particle number and energy when the system is subjected to an external perturbation. Conservation of particle number means that the continuity equation

$$-\partial_t n(\boldsymbol{r}, t) = \nabla \cdot \boldsymbol{j}(\boldsymbol{r}, t) \tag{153}$$

is satisfied. Conservation of energy means that the energy change when an external potential is applied to the system is equal to the work done by the system against the external potential when calculated using the self-consistent G (Baym and Kadanoff 1961, Baym 1962). Moreover, self-consistency ensures that

$$N = \frac{1}{\pi}\mathrm{tr}\int_{-\infty}^{\mu} \mathrm{d}\omega\, \mathrm{Im}\, G(\omega) \tag{154}$$

gives the correct total number of particles, when n and $\boldsymbol{j}$ are obtained from the self-consistent Green function. The first self-consistent GW calculation was probably by de Groot *et al* (1995) for a model quasi-one-dimensional semiconducting wire. The relevance of this model to real solids is, however, unclear.

Five aspects may be distinguished in relation to self-consistency (von Barth and Holm 1996):

(1) modification of quasiparticle wavefunctions;
(2) shift of quasiparticle energies;
(3) modification of quasiparticle weights (Z factors);
(4) modification of quasiparticle lifetimes;
(5) modification of the screening properties of the system.

These aspects have recently been studied in detail by von Barth and Holm (1996) and by Shirley (1996) for the electron gas. Naturally, the first aspect cannot be addressed for the electron gas since the quasiparticle wavefunctions remain plane waves. The results of these studies are:

- the bandwidth is increased from its non-self-consistent value as can be seen in figure 18, worsening the agreement with experiment;
- the weight of the quasiparticles is increased, reducing the weight in the plasmon satellite;
- the quasiparticles are narrowed, increasing their lifetime;
- the plasmon satellite is broadened and shifted towards the Fermi level (figure 19). In fact, it almost disappears at full self-consistency.

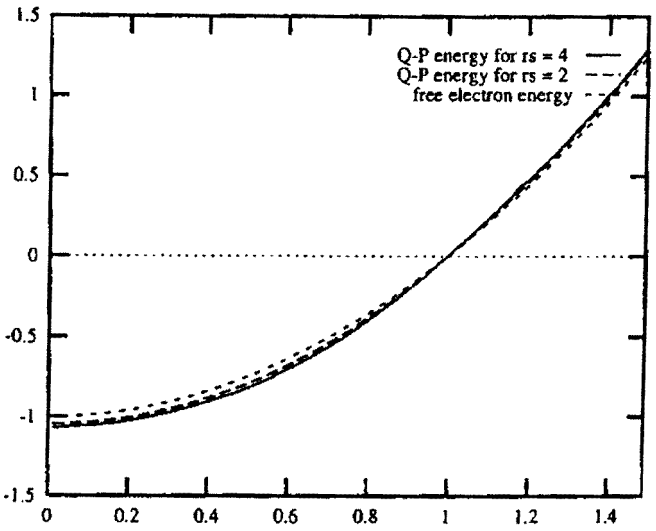

Figure 18. The quasiparticle dispersion for the electron gas for $r_s = 2$ and 4 from partial self-consistent GW_0 calculations (W_0 is held fixed at the RPA whereas G is allowed to vary for self-consistency) compared with the free-electron dispersion. The largest change in the bandwidth occurs for $r_s = 4$. After von Barth and Holm (1996).

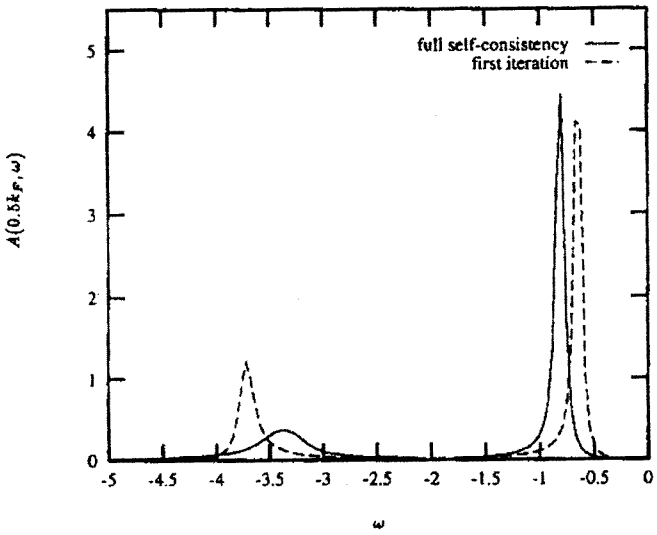

Figure 19. The spectral function from a partial self-consistent GW_0 calculation (see previous figure) compared to that of the first iteration for $k = 0.5k_F$ and $r_s = 4$. The self-consistent quasiparticle energy is lowered (band broadening) compared to the non-self-consistent one whereas the satellite position is somewhat improved. After von Barth and Holm (1996).

As discussed later, the main effects of self-consistency are caused by allowing the quasiparticle weight to vary (von Barth and Holm 1996). These results are true for the case when the screened interaction W is fixed at the RPA level (calculated using the non-interacting G_0) and only the Green function is allowed to vary for self-consistency and

also for the case when both G and W are allowed to vary (full self-consistency case). The increase in the bandwidth is disturbing and can be understood as follows. Let us consider the first case with fixed $W = W_0$ for simplicity. First we note that the GW result for the bandwidth after one iteration is close to the free-electron one. This means that there is almost a complete cancellation between exchange and correlation. After one iteration the quasiparticle weight is reduced to typically 0.7 and the rest of the weight goes to the plasmon satellite. Correspondingly, when the new G is used to calculate a new $\mathrm{Im}\,\Sigma^{\mathrm{c}}$, its weight is transferred away from low energy to high energy, due to the sum rule (von Barth and Holm 1996)

$$\int_{-\infty}^{\infty} \mathrm{d}\omega |\mathrm{Im}\,\Sigma^{\mathrm{c}}(\boldsymbol{k}, \omega)| = \sum_{\boldsymbol{q}} \int_{0}^{\infty} \mathrm{d}\omega |\mathrm{Im}\, W_0(\boldsymbol{q}, \omega)| \tag{155}$$

which shows that the left-hand side is a constant depending only on the pre-chosen W_0 but independent of $\boldsymbol{k}$ and self-consistency. For a state at the Fermi level, this has little effect since $\mathrm{Im}\,\Sigma^{\mathrm{c}}$ has almost equal weights for the hole ($\omega \leqslant \mu$) and the particle part ($\omega > \mu$) which cancel each other when calculating $\mathrm{Re}\,\Sigma^{\mathrm{c}}$, as can be seen in equation (68). However, for the state at the bottom of the valence band, $\mathrm{Im}\,\Sigma^{\mathrm{c}}$ has most of its weight in the hole part so that the shifting of the weight in $\mathrm{Im}\,\Sigma^{\mathrm{c}}$ to higher energy causes $\mathrm{Re}\,\Sigma^{\mathrm{c}}$ to be less positive than its non-self-consistent value. A similar effect is found for the exchange part, which becomes less negative, but because the bare Coulomb interaction has no frequency dependence, the renormalization factor has a smaller effect on Σ^{x} so that the reduction in Σ^{x} is less than the reduction in Σ^{c}. The net effect is then an increase in the bandwidth. The shifting of the weight in $\mathrm{Im}\,\Sigma^{\mathrm{c}}$ to higher energy has the immediate consequences of increasing the lifetime and the renormalization weight Z (through a decrease in $|\partial\,\mathrm{Re}\,\Sigma^{\mathrm{c}}/\partial\omega|$) of the quasiparticles and of broadening the plasmon satellite, compared to the results of one iteration (von Barth and Holm 1996).

When W is allowed to vary (full self-consistency) the results become even worse: the bandwidth becomes even wider and the plasmon satellite becomes broad and featureless, in contradiction to experiment. The quasiparticle weight is increased further. These results can be explained by the disappearance of a well defined plasmon excitation in W. The quantity $P = -\mathrm{i}GG$ no longer has the physical meaning of a response function, rather it is an auxiliary quantity needed to construct W. Indeed, it does not satisfy the usual f-sum rule. The equations $\mathrm{Re}\,\epsilon(\boldsymbol{q}, \omega_{\mathrm{p}}) = \mathrm{Im}\,\epsilon(\boldsymbol{q}, \omega_{\mathrm{p}}) = 0$ determining the plasmon energy are not satisfied any more since the Green function now always has weight around $\omega = \omega_{\mathrm{p}}$. This has the effect of transferring even more weight in $\mathrm{Im}\,\Sigma^{\mathrm{c}}$ to higher energy with the consequences discussed in the previous paragraph. Shirley (1996) included the second-order self-energy diagram (vertex correction) and found that it cancelled the effects of self-consistency to some degree.

A very interesting outcome of the full self-consistent calculation (Holm and von Barth 1997) is that the total energy calculated with the Galitskii–Migdal formula (1958) turns out to be strikingly close to the total energy calculated from a much more elaborate quantum Monte Carlo technique (Ceperley and Alder 1980). For $r_s = 2$ and 4 quantum Monte Carlo analysis gives 0.004 and -0.155 Ryd, respectively, while the self-consistent GW calculation gives 0.005 and -0.156 Ryd, respectively. This unexpected result may be related to the fact that the self-consistent GW scheme is energy conserving. The result is also partly explained by consideration of the Luttinger–Ward energy functional (1960) which is variational with respect to G and has a minimum equal to the value obtained from the Galitskii–Migdal formula. What is not clear is why the first-order energy diagram (giving the GW self-energy upon taking a functional derivative with respect to G) appears to represent a very

good energy functional. Furthermore, the chemical potential calculated from $\mu = \partial E/\partial N$ is in agreement with the value obtained from $\mu = k_F^2/2 + \Sigma(k_F, \mu)$ and the particle density $n = 2\sum_{\boldsymbol{k}} \int_{-\infty}^{\mu} d\omega\, A(\boldsymbol{k}, \omega)$ yields $n = k_F^3/(3\pi^2)$, i.e. particle number is conserved, as proven by Baym (1962). It has also been proven that with a fixed $W = W_0$, particle number is also conserved (Holm and von Barth 1997).

The conclusion is that it is not a good idea to perform fully self-consistent GW calculations for quasiparticle energies. If full self-consistency is not introduced, an important question is how to choose H_0 determining G_0. Farid (1996) argued that H_0 should be chosen so that its ground-state density is the same as that resulting from the GW calculation. It could be more favourable to perform partially self-consistent calculations by, for instance, fixing W at the RPA level and modifying the quasiparticle energies in G_0, or by choosing a single-particle Hamiltonian such that the resulting quasiparticle energies obtained from a GW calculation are the same as those of the single-particle Hamiltonian. In any case, efforts should be directed towards finding vertex corrections (beyond GW).

7. Vertex corrections

By vertex corrections we mean corrections to the self-energy beyond the GW–RPA approximation and corrections to the response function beyond the RPA. Vertex corrections to the RPA response function naturally involve interaction between screening electrons which are not taken into account in the RPA. This interaction must include the effect of exchange and correlation, which makes the interaction depend upon the relative spin states of the electrons. Vertex corrections to the self-energy, on the other hand, involve interaction of the hole with its surroundings which is not taken into account in the GWA. A consistent treatment should include vertex corrections in both the response function and the self-energy (Ward 1950). This preserves conservation laws and is a natural outcome of the perturbation expansion of the self-energy in the functional derivative technique of Baym and Kadanoff (1961) and Baym (1962). For example, calculations of the optical spectra of Si indicate a cancellation between vertex corrections in the response function and the renormalization of the quasiparticle energies (Bechstedt *et al* 1997, Del Sole and Girlanda 1996). Mahan and Sernelius (1989) found a similar effect in the calculations of bandwidths of the electron gas. Conserving approximations including vertex corrections with exchange effects only have been considered by Hong and Mahan (1994) for the electron gas.

In the GWA, the screened interaction is calculated within the RPA which takes into account primarily long-range correlation which gives rise to collective excitations (plasmons). The photoemission hole is coupled to one plasmon only in the GWA. Vertex corrections may be loosely divided into the short-range and long-range parts. Short-range vertex corrections improve the description of the quasiparticles and low-energy satellites whereas long-range vertex corrections improve description of high-energy satellites (plasmons).

The RPA pair distribution function, which is the probability of finding another electron at a certain distance from a given electron, is negative for small distances, which is unphysical (Pines 1961). Exchange and correlations should therefore increase the probability of finding another electron at small distances. The same applies to holes. For a test charge, the screening becomes more effective when the effects of exchange and correlations are taken into account. But for an electron, the RPA screening is *reduced* when the effects of exchange and correlations between the screened electron and the surrounding electrons are taken into account. The physical interpretation of vertex corrections is as follows: an electron pushes away other electrons in its vicinity creating a screening hole around the

electron. Taking into account exchange and correlations of the screening holes increases the screening since the probability of the holes getting closer together is increased leading to more screening holes. If we now take into account the effects of exchange and correlations between the screened electron and the other electrons, then screening will be reduced because the probability of finding electrons at small distances to the screened electron is increased, leading to stronger effective interactions between the electrons. The net effect is that the RPA screening is reduced. Thus, vertex corrections for electrons will in general reduce screening. Vertex corrections in the response function can also take the form of interaction between electrons and holes in electron–hole pairs created by the perturbation in the system due to the presence of a photoemission hole. This interaction may actually create an additional bound state (exciton) much lower in energy, of the order of a few eV, than the plasmon energy.

Vertex corrections to the self-energy are particularly important for systems with localized states, such as those containing 3d and 4f orbitals. This is because the electron–hole pairs created in the screening process can be rather localized and therefore interact strongly with the localized photoemission hole. This interaction can significantly modify the quasihole energy and its weight, as well as creating new collective excitations appearing as low-energy satellites. This type of vertex correction is short range in nature. Long-range vertex corrections modify the structure of the plasmon satellites and may create multiple plasmons as observed in the alkalis. In the GWA, the photoemission hole only couples to one plasmon, resulting in a one-plasmon satellite. The hole, however, can in general interact with several plasmons during its propagation, producing multiple plasmon satellites.

7.1. Direct evaluation of the second-order self-energy

Formally, the GWA is the first-order term in the expansion of the self-energy in the screened interaction W. It is then thought that the simplest vertex correction is the second-order self-energy. This procedure, however, warrants some precautions. First, the physical meaning of the second-order term is not clear. Second, this second-order term when evaluated with a frequency-dependent interaction can give a self-energy with wrong analytic properties which result in a negative density of states, as shown in the electron-gas case (Minnhagen 1974).

This type of vertex correction was calculated for the bandgap of Si both with a bare and a screened interaction (Daling and van Haeringen 1989, Daling *et al* 1991, Bobbert and van Haeringen 1994). The second-order self-energy with the bare Coulomb interaction for the Γ_{15} and Γ'_{25} states was found to be factors of 18 and 37 smaller than the first-order self-energy. This second-order self-energy gives a correction of $\sim$1% to the Hartree–Fock direct gap. A calculation using a screened interaction yields a correction to the gap of the order of $\sim$4%. Since this calculation was performed for the gap states only, it is not clear if the density of states became negative for some energies. Both calculations indicate in any case that the second-order term is small. This is encouraging since it suggests that higher-order terms are probably small too since the GWA already gives results in agreement with experiment. An interesting result is that the second-order vertex correction does not shift the absolute position of the LDA valence-band maximum which, in the approximation used, is too low by 0.5 eV. It is speculated that higher-order vertex corrections could account for the required shift (Bobbert and van Haeringen 1994).

7.2. Vertex corrections based on the LDA exchange-correlation potential

The set of equations (43)–(46) derived by Hedin provides a systematic way of developing a perturbation series for the self-energy in powers of the screened interaction W. In the original derivation, the zeroth-order Green function is taken to be the Hartree one. However, the bandstructure and wavefunctions in the Hartree theory are less accurate compared with those of the LDA, making the Hartree Green function a poor starting point. In practical calculations, one uses a Green function constructed from the LDA bandstructure. The response function is also calculated using the LDA Green function.

Consistent vertex corrections have been derived by Rice (1965) in relation to the dielectric function as discussed in the alkali section. A similar approach was also made by Ting *et al* (1975) and Mahan (1994). Alternatively, consistent vertex corrections can also be derived from the set of equations (43)–(46) where one regards $V^{\rm xc}_{\rm LDA}$ as a self-energy correction to the Hartree approximation, albeit an *ad hoc* one (not based on a systematic diagrammatic expansion). Based on this starting point, the vertex function Λ can be easily evaluated yielding (Hybertsen and Louie 1986, Del Sole *et al* 1994)

$$\Lambda(1,2,3) = \delta(1,2)\delta(1,3) - \mathrm{i}\int \mathrm{d}(5,6,7)K^{\rm xc}(1,5)G(5,6)G(7,5)\Lambda(6,7,3) \tag{156}$$

where

$$K^{\rm xc}(1,5) = \frac{\delta V^{\rm xc}(1)}{\delta\rho(5)} \tag{157}$$

remembering that $\rho(1) = -\mathrm{i}G(1,1^+)$. The new self-energy with vertex corrections has the form of the GWA but with a new screened potential

$$\tilde{W} = v[1 - P^0(v + K^{\rm xc})]^{-1} \tag{158}$$

corresponding to a dielectric function

$$\tilde{\epsilon} = 1 - P^0(v + K^{\rm xc}). \tag{159}$$

The same result was already discussed in section 5.1 where the self-energy was expressed as a corrected dielectric function ϵ (equation (136)) and a vertex correction Γ (equation (137)). The product $\epsilon^{-1}\Gamma$ is equal to $\tilde{\epsilon}^{-1}$ with $\tilde{\epsilon}$ given above. This dielectric function may also be derived straightforwardly from the time-dependent LDA and may therefore be interpreted as the dielectric function that screens the external potential felt by an electron, as opposed to a test charge (Hybertsen and Louie 1986). This distinguishes itself from the RPA (time-dependent Hartree) dielectric function in that the induced charge does not only generate the Hartree potential but also an exchange-correlation potential. It is clear that one could start with a different local zeroth-order self-energy other than $V^{\rm xc}_{\rm LDA}$ and arrive at a similar formula. Note that in the formula for the self-energy in equation (43), the vertex function Λ enters both W, through P, and Σ. A problem with starting with a local zeroth-order self-energy is that the self-energy with the vertex corrections becomes asymmetric in $\boldsymbol{r}$ and $\boldsymbol{r}'$.

Application of this scheme to the electron gas gives small changes compared with the original GWA (Mahan and Sernelius 1989). Similarly, for Si it yields practically the same gap (0.70 eV) and valence bandwidth (11.4 eV) as those of the standard GW calculation (Del Sole *et al* 1994). However, the absolute position of the top of the valence band is shifted 10 meV upwards by the GW + vertex corrections and 400 meV downwards by the standard GW calculation. This could improve the calculations of band off-sets at interfaces.

A calculation with the vertex corrections in the response function R only but not in Σ, that is with

$$W = v[1 + v[1 - P^0(v + K^{\mathrm{xc}})]^{-1} P^0] \tag{160}$$

$$R = [1 - P^0(v + K^{\mathrm{xc}})]^{-1} P^0 \tag{161}$$

gives a smaller gap (0.57 eV) and a smaller bandwidth (10.9 eV) (Del Sole *et al* 1994). The latter result is similar to the case of the alkalis (Northrup *et al* 1987, Surh *et al* 1988). The top of the valence band is shifted downward by 0.4 eV as in the standard GWA. The result for the bandwidth should be taken with caution because the plasmon-pole approximation was used in the calculations. The worse results obtained by including vertex corrections in R only, is consistent with the violation of conservation laws.

7.3. The cumulant expansion

A diagrammatic approach for including vertex corrections is provided by the cumulant expansion method. One of the first applications of the cumulant expansion method was in studying the x-ray spectra of core electrons in metals (Nozières and de Dominicis 1969, Langreth 1970). Later on the method was extended to valence states by Bergersen *et al* (1973) and Hedin (1980). The core-electron problem is modelled by a Hamiltonian consisting of a core electron interacting with a set of plasmons:

$$H = \varepsilon c^\dagger c + \sum_{\boldsymbol{q}} \omega_{\boldsymbol{q}} b_{\boldsymbol{q}}^\dagger b_{\boldsymbol{q}} + \sum_{\boldsymbol{q}} c c^\dagger g_{\boldsymbol{q}} (b_{\boldsymbol{q}} + b_{\boldsymbol{q}}^\dagger) \tag{162}$$

where c is the annihilation operator for the core electron with energy ε, $b_{\boldsymbol{q}}^\dagger$ is the creation operator for a plasmon of wavevector $\boldsymbol{q}$ and energy $\omega_{\boldsymbol{q}}$, and the last term is the coupling of the core electron to the plasmon field. The Hamiltonian can be solved exactly and it can be shown that the cumulant expansion also gives the exact solution (Langreth 1970):

$$A_\pm(\omega) = \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-a} a^n}{n!} \delta(\omega - \varepsilon - \Delta\varepsilon \mp n\omega_{\mathrm{p}}) \tag{163}$$

where $+$ refers to absorption spectrum and $-$ to emission spectrum. $a = \sum_{\boldsymbol{q}} g_{\boldsymbol{q}}^2/\omega_{\mathrm{p}}^2$ and $\Delta\varepsilon = a\omega_{\mathrm{p}}$ is the shift in core energy due to the interaction with the plasmon field. It is assumed that the plasmon excitations have no dispersion although this assumption is not necessary. The spectra consist, therefore, of the main quasiparticle peak at $\omega = \varepsilon + \Delta\varepsilon$ and a series of plasmon excitations at multiples of the plasmon energy below the quasiparticle peak which is in accordance with experiment. This is in contrast to the GW spectra which only has one plasmon excitation located at too high an energy, typically $1.5\omega_{\mathrm{p}}$ below the quasiparticle peak (Hedin *et al* 1970). More recently the cumulant expansion method was applied to a model Hamiltonian with electron–boson interaction and the cumulant was calculated to higher order (Gunnarsson *et al* 1994).

In the cumulant expansion approach, the Green function for the hole ($t < 0$) is written as (Langreth 1970, Bergersen *et al* 1973, Hedin 1980)

$$\begin{aligned} G(k, t < 0) &= \mathrm{i}\theta(-t)\langle N|\hat{c}_k^\dagger(0)\hat{c}_k(t)|N\rangle \\ &= \mathrm{i}\theta(-t)\,\mathrm{e}^{-\mathrm{i}\varepsilon_k t + C^{\mathrm{h}}(k,t)} \end{aligned} \tag{164}$$

and the hole spectral function is

$$A(k, \omega \leqslant \mu) = \frac{1}{\pi} \operatorname{Im} G(k, \omega \leqslant \mu)$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\langle N|\hat{c}_k^{\dagger}(0)\hat{c}_k(t)|N\rangle$$
$$= \frac{1}{\pi}\,\mathrm{Im}\,\mathrm{i}\int_{-\infty}^{0} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{e}^{-\mathrm{i}\varepsilon_k t + C^{\mathrm{h}}(k,t)} \tag{165}$$

where k denotes all possible quantum labels and $C^{\mathrm{h}}(k,t)$ is defined to be the cumulant. Expanding the exponential in powers of the cumulant we get

$$G(k,t) = G_0(k,t)[1 + C^{\mathrm{h}}(k,t) + \tfrac{1}{2}[C^{\mathrm{h}}(k,t)]^2 + \cdots] \tag{166}$$

where $G_0(k,t) = \mathrm{i}\exp(-\mathrm{i}\varepsilon_k t)$. In terms of the self-energy, the Green function for the hole can be expanded as

$$G = G_0 + G_0\Sigma G_0 + G_0\Sigma G_0\Sigma G_0 + \cdots. \tag{167}$$

To lowest order in the screened interaction W, the cumulant is obtained by equating

$$G_0 C^{\mathrm{h}} = G_0\Sigma G_0 \tag{168}$$

where $\Sigma = \Sigma_{GW} = \mathrm{i}G_0 W$. If G_0 corresponds to, for example G_{LDA}, then $\Sigma = \Sigma_{GW} - V^{\mathrm{xc}}$. The first-order cumulant is therefore (Hedin 1980, Almbladh and Hedin 1983, Aryasetiawan *et al* 1996)

$$C^{\mathrm{h}}(k,t) = \mathrm{i}\int_t^{\infty}\mathrm{d}t'\int_{t'}^{\infty}\mathrm{d}\tau\,\mathrm{e}^{\mathrm{i}\varepsilon_k\tau}\,\Sigma(k,\tau). \tag{169}$$

The cumulant may be conveniently divided into a quasiparticle part and a satellite part: $C^{\mathrm{h}} = C^{\mathrm{h}}_{\mathrm{QP}} + C^{\mathrm{h}}_{\mathrm{S}}$ where

$$C^{\mathrm{h}}_{\mathrm{QP}}(k,t) = (\mathrm{i}\alpha_k + \gamma_k) + (-\mathrm{i}\Delta\varepsilon_k + \eta_k)t \tag{170}$$

$$C^{\mathrm{h}}_{\mathrm{S}}(k,t) = \int_{-\infty}^{\mu}\mathrm{d}\omega\frac{\mathrm{e}^{\mathrm{i}(\varepsilon_k-\omega-\mathrm{i}\delta)t}}{(\varepsilon_k-\omega-\mathrm{i}\delta)^2}\Gamma(k,\omega) \tag{171}$$

with

$$\mathrm{i}\alpha_k + \gamma_k = \left.\frac{\partial\Sigma(k,\omega)}{\partial\omega}\right|_{\omega=\varepsilon_k} \tag{172}$$

$$\Delta\varepsilon_k = P\int_{-\infty}^{\infty}\mathrm{d}\omega\frac{\Gamma(k,\omega)}{\varepsilon_k-\omega} \qquad \eta_k = \pi\Gamma(k,\varepsilon_k) = |\mathrm{Im}\,\Sigma(k,\varepsilon_k)|. \tag{173}$$

η_k is the inverse lifetime of the quasiparticle and $\Gamma(k,\omega)$ is the spectral function of the self-energy which is proportional to $\mathrm{Im}\,\Sigma(k,\omega)$. A similar derivation can be carried out for the particle Green function

$$G(k,t>0) = -\mathrm{i}\theta(t)\,\mathrm{e}^{-\mathrm{i}\varepsilon_k t + C^{\mathrm{p}}(k,t)}. \tag{174}$$

The result is

$$C^{\mathrm{p}}(k,t) = -\mathrm{i}\Delta\varepsilon_k t - \eta_k t + \left.\frac{\partial\Sigma(k,\omega)}{\partial\omega}\right|_{\omega=\varepsilon_k} + \int_{\mu}^{\infty}\mathrm{d}\omega\frac{\mathrm{e}^{\mathrm{i}(\varepsilon_k-\omega+\mathrm{i}\delta)t}}{(\varepsilon_k-\omega+\mathrm{i}\delta)^2}\Gamma(k,\omega). \tag{175}$$

It is physically appealing to extract the quasiparticle part from the Green function:

$$G^{\mathrm{h}}_{\mathrm{QP}}(k,t) = \mathrm{i}\theta(-t)\,\mathrm{e}^{\mathrm{i}\alpha_k+\gamma_k}\,\mathrm{e}^{(-\mathrm{i}E_k+\eta_k)t} \qquad E_k = \varepsilon_k + \Delta\varepsilon_k. \tag{176}$$

The spectral function for this quasiparticle can be calculated analytically (Almbladh and Hedin 1983):

$$A_{\mathrm{QP}}(k,\omega<\mu) = \frac{\mathrm{e}^{-\gamma_k}}{\pi}\frac{\eta_k\cos\alpha_k - (\omega-E_k)\sin\alpha_k}{(\omega-E_k)^2+\eta_k^2}. \tag{177}$$

From equation (165) we have

$$\langle N|\hat{c}_k^\dagger(0)\hat{c}_k(t)|N\rangle = \mathrm{e}^{-\mathrm{i}\varepsilon_k t + C^{\mathrm{h}}(k,t)} \qquad \text{for } t < 0.$$

By analytically continuing to $t > 0$ the spectral function in equation (165) can be rewritten as

$$A(k, w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{e}^{-\mathrm{i}\varepsilon_k t + C^{\mathrm{h}}(k,t)} \tag{178}$$

where for $t > 0$ we have $C^{\mathrm{h}}(k, t) = C^{\mathrm{h}*}(k, -t)$. The total spectra can be written as a sum of A_{QP} and a convolution between the quasiparticle and the satellite part:

$$\begin{aligned} A(k, \omega) &= A_{\mathrm{QP}}(k, \omega) + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{e}^{C^{\mathrm{h}}_{\mathrm{QP}}(k,0)}\, \mathrm{e}^{(-\mathrm{i}E_k + \eta_k)t}[\mathrm{e}^{C^{\mathrm{h}}_{\mathrm{S}}(k,t)} - 1] \\ &= A_{\mathrm{QP}}(k, \omega) + A_{\mathrm{QP}}(k, \omega) * A_{\mathrm{S}}(k, \omega) \end{aligned} \tag{179}$$

where

$$\begin{aligned} A_{\mathrm{S}}(k, \omega) &= \frac{1}{2\pi}\int \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\{\mathrm{e}^{C^{\mathrm{h}}_{\mathrm{S}}(k,t)} - 1\} \\ &= \frac{1}{2\pi}\int \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}\left\{C^{\mathrm{h}}_{\mathrm{S}}(k, t) + \frac{1}{2!}[C^{\mathrm{h}}_{\mathrm{S}}(k, t)]^2 + \cdots\right\}. \end{aligned} \tag{180}$$

The second term $A_{\mathrm{QP}} * A_{\mathrm{S}}$ is responsible for the satellite structure. The Fourier transform of $C^{\mathrm{h}}_{\mathrm{S}}$ can be done analytically (Aryasetiawan *et al* 1996)

$$C^{\mathrm{h}}_{\mathrm{S}}(k, \omega < 0) = \frac{\Gamma(k, \varepsilon_k + \omega) - \Gamma(k, \varepsilon_k) - \omega\Gamma'(k, \varepsilon_k)}{\omega^2}. \tag{181}$$

As follows from equations (177) and (179), the quasiparticle energy in the cumulant expansion is essentially determined by E_k, which is the quasiparticle energy in the GWA.

By comparing the diagrammatic expansions in the GWA and the cumulant expansion we can gain some idea about the vertex corrections. In figure 20 the Green function diagrams are shown to second order in the screened interaction, which should be sufficient for our purpose. The cumulant expansion diagrams are obtained by considering the three possible time orderings of the integration time variables t' in $C^{\mathrm{h}^2}(k, t)$ with $C^{\mathrm{h}}(k, t)$ given by equation (169). The cumulant expansion contains second-order diagrams which are not included in the GWA. It is these additional diagrams that give rise to the second plasmon satellite and they are quite distinct from the second-order diagram common to both approximations. The interpretation of the latter diagram is that a hole emits a plasmon, which is reabsorbed at a later time, and the hole returns to its original state before plasmon emission. This process is repeated once at a later time. Thus there is only one plasmon coupled to the hole at one time. In contrast, the other two diagrams, not contained in the GWA, describe an additional plasmon emission before the first one is reabsorbed, giving two plasmons coupled to the hole simultaneously. Similar consideration can be extended to the higher-order diagrams. If self-consistency is taken into account then the second second-order diagram is also included in the GWA.

The cumulant expansion contains only boson-type diagrams describing emission and reabsorbtion of plasmons but it does not contain diagrams corresponding to the interaction between a hole and particle–hole pairs. For this reason, the cumulant expansion primarily corrects the satellite description whereas the quasiparticle energies are to a large extent determined by the GWA as mentioned previously. The interaction between hole–hole and particle–hole is described by the ladder diagrams which in a Hubbard model study were found to improve the low-energy satellite (Verdozzi *et al* 1995).

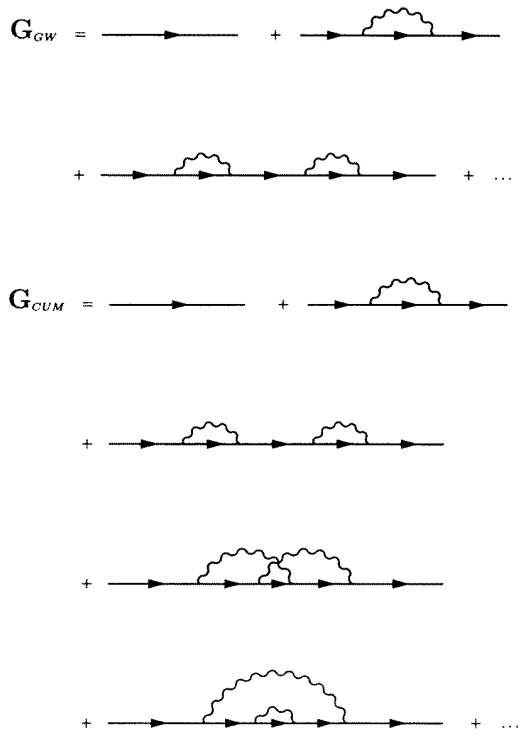

Figure 20. A diagrammatic expansion for the Green function to second order in the GWA and the cumulant expansion. The full lines represent non-interacting Green functions G_0, and the wiggly lines represent the screened interaction W.

The cumulant expansion was applied recently to calculate the photoemission spectra in Na and Al (Aryasetiawan *et al* 1996) (figure 21). The experimental spectra consist of a quasiparticle peak with a set of plasmon satellites separated from the quasiparticle by multiples of the plasmon energy. The spectra in the GWA shows only one plasmon satellite located at a too high energy, approximately $1.5\omega_p$ below the quasiparticle which is similar to the core electron case. The cumulant expansion method remedies this problem and yields spectra in good agreement with experiment regarding the position of the satellites. The relative intensities of the satellites with respect to that of the quasiparticle are still in discrepancy. This is likely to be due to extrinsic effects corresponding to the interaction of the photoemitted electron with the bulk and the surface on its way out of the solid resulting in energy loss. These are not taken into account in the sudden approximation.

When applied to valence electrons with band dispersions the cumulant expansion does not yield the exact result anymore as in the core electron case. Surprisingly, the numerical results show that the cumulant expansion works well even in Al with a bandwidth of $\sim$11 eV. Considering its simplicity, it is a promising approach for describing plasmon satellites.

8. Summary and conclusions

The GWA has been applied by now to a large number of systems, ranging from atoms, simple metals, semiconductors, transition metals, clusters, and surfaces and interfaces. In practically all of these systems, the GWA improves the quasiparticle energies relative to the

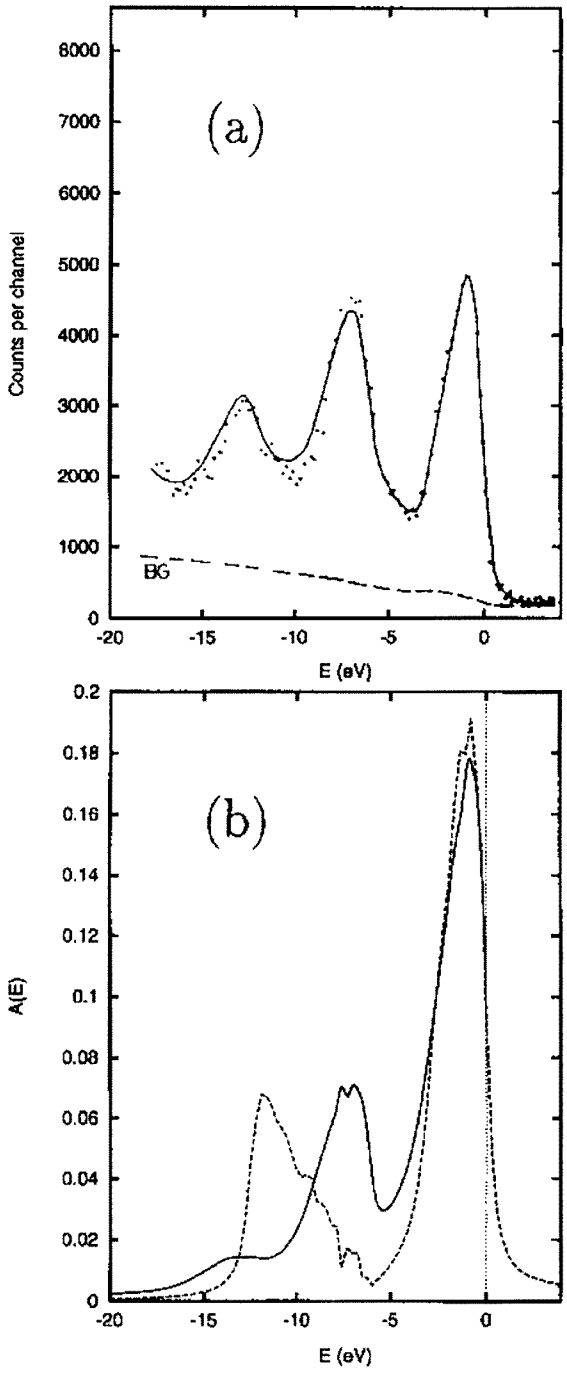

Figure 21. (a) The experimental spectral function of Na (dots). The full curve is a synthetic spectrum obtained by convoluting the density of states from a bandstructure calculation and the experimental core level spectrum. BG is the estimated background contribution. After Steiner *et al* (1979). (b) The calculated total spectral function of Na for the occupied states. The full and dashed curves correspond to the cumulant expansion and the GWA, respectively. After Aryasetiawan *et al* (1996).

LDA eigenvalues. The reason for the success of the GWA may be understood qualitatively by the fact that it is correct in some limiting cases as described in the introduction. The GWA includes an important physical ingredient in extended systems, namely screening or polarization of the medium which is absent in the HFA. It is well known that the neglect of screening leads to unphysical results in metals, such as a zero density of states at the Fermi level, and in semiconductors and insulators to too large bandgaps. Even in atoms, the inclusion of polarization effects leads to a significant improvement in the HF eigenvalues. Since screening is a common feature in all electronic systems, it is perhaps not surprising that the GWA works in a wide variety of materials. It is also known empirically that second-order perturbation theory often takes into account most of the physical effects, in particular the shifting of quasiparticle energies.

Despite the success of the GWA, it has naturally some shortcomings. One of these is related to satellite structure in the photoemission spectra. Since the GWA describes a coupling of the electrons to one plasmon excitation, represented by W, multiple plasmon excitations observed in the alkalis are clearly beyond the scope of the GWA. This problem is remedied by the cumulant expansion theory described in section 7.3. Apart from plasmon-related satellites, there is also satellite structure originating from short-range interactions. This type of satellite appears in strongly correlated materials containing 3d or 4f orbitals. The GWA is based on the RPA screening which takes into account the dominant part of long-range (small momentum) screening. Short-range (large momentum) intrasite interactions of multiple holes usually present in strongly correlated systems are therefore not well described by the GWA. Here a T-matrix-type approach may be appropriate as has been shown by model calculations. Apart from problems with the satellite, there are also in some cases discrepancies in the quasiparticle energies. For example, the bandwidth in Na is $\approx$10% too large within the GWA compared with experiment, but these discrepancies are relatively small.

Another shortcoming is the absence of spin dependence in the screened interaction W since the screening is purely Coulombic. The spin dependence enters only through the Green function. One would therefore expect some problems when applying the GWA to magnetic systems where spin–spin correlations are important as indicated by GW calculations on transition metal atoms. This area of research has not been explored extensively and it would require inclusion of vertex corrections to take into account spin–spin correlations. A T-matrix approach may be a first step in this direction.

Most GW calculations performed so far are not self-consistent, i.e. the Green function used to calculate the self-energy is not equal to the Green function obtained from the Dyson equation with the very same self-energy. Only very recently were such self-consistent calculations performed for the electron gas. The results turned out to be worse than the straight GW calculations (one iteration) and clearly show a cancellation between the effects of self-consistency and vertex corrections. In some way it is a blessing, since fully self-consistent calculations are numerically difficult to perform for real systems. One interesting aspect of the self-consistent calculations is that the total energies are in very good agreement with the quantum Monte Carlo (QMC) results. A recent finding shows that when the total energies are calculated within the Luttinger–Ward functional and its extension (Almbladh *et al* 1997) using approximate G and W the results are almost equally good, which circumvents the need for self-consistent calculations (Hindgren 1997). This could be due to the variational property of the functional and the fact that the self-consistent GWA is conserving in the sense of Baym and Kadanoff. Calculating total energies within the GWA could be an alternative to the QMC method, particularly for extended systems.

Due to computational difficulties, applications of the GWA to large and complex systems are still not feasible. Simplified GW schemes, which are computationally efficient and yet

maintain the accuracy of the full calculations, are therefore very desirable. Many schemes have been proposed but most of them are designed for semiconductors. While they give reliable bandgaps, details of the bandstructure are not fully accounted for. Reliable schemes must probably take into account non-locality as well as the energy dependence of the self-energy. With efficient schemes, many interesting problems can then be tackled. These include chemisorption at surfaces, 3d impurities in semiconductors, interfaces, band off-sets in heterojunctions and exotic materials such as fullerenes in their diverse forms.

Acknowledgments

It is a pleasure to thank Professor Lars Hedin and Dr Ulf von Barth for their many valuable comments and careful reading of the manuscript.

References

Albrecht S, Onida G and Reining L 1997 *Phys. Rev.* B **55** 10 278–81
Almbladh C-O and von Barth U 1985a *Phys. Rev.* B **31** 3231
——1985b *Density Functional Methods in Physics (NATO Advanced Study Institute, Series B)* vol 123, ed R M Dreizler and J da Providencia (New York: Plenum)
Almbladh C-O, von Barth U and van Leeuwen R 1997 Private communication
Almbladh C-O and Hedin L 1983 *Handbook on Synchrotron Radiation* vol 1, ed E E Koch (Amsterdam: North-Holland) p 686
Alperin H A 1962 *J. Phys. Soc. Japan Suppl.* B **17** 12
Andersen O K 1975 *Phys. Rev.* B **12** 3060
Anisimov V I, Aryasetiawan F and Lichtenstein A I 1997 *J. Phys.: Condens. Matter* **9** 767–808
Anisimov V I, Solovyev I V, Korotin M A, Czyzyk M T and Sawatzky G A 1993 *Phys. Rev.* B **48** 16 929
Anisimov V I, Zaanen J and Andersen O K 1991 *Phys. Rev.* B **44** 943
Arai M and Fujiwara T 1995 *Phys. Rev.* B **51** 1477–89
Arbman G and von Barth U 1975 *J. Phys. F: Met. Phys.* **5** 1155
Aryasetiawan F 1992 *Phys. Rev.* B **46** 13 051–64
Aryasetiawan F and von Barth U 1992 *Phys. Scr.* T **45** 270–1
Aryasetiawan F and Gunnarsson O 1994a *Phys. Rev.* B **49** 16 214–22
——1994b *Phys. Rev.* B **49** 7219
——1995 *Phys. Rev. Lett.* **74** 3221–4
——1996 *Phys. Rev.* B **54** 17 564–8
——1997 unpublished
Aryasetiawan F, Hedin L and Karlsson K 1996 *Phys. Rev. Lett.* **77** 2268–71
Aryasetiawan F and Karlsson K 1996 *Phys. Rev.* B **54** 5353–7
Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Saunders College)
Aspnes D E 1976 *Phys. Rev.* B **14** 5331
Baldereschi A and Tosatti E 1979 *Solid State Commun.* **20** 131
Baldini G and Bosacchi B 1970 *Phys. Status Solidi* **38** 325
Bates D R 1947 *Proc. R. Soc. London* A **188** 350
Baym G 1962 *Phys. Rev.* **127** 1391
Baym G and Kadanoff L P 1961 *Phys. Rev.* **124** 287
Bechstedt F and Del Sole R 1988 *Phys. Rev.* B **38** 7710–16
——1990 *Solid State Commun.* **74** 41–4
Bechstedt F, Del Sole R, Cappellini G and Reining L 1992 *Solid State Commun.* **84** 765–70
Bechstedt F, Tenelsen K, Adolph B and Del Sole R 1997 *Phys. Rev. Lett.* **78** 1528–31
Becke A D 1988 *Phys. Rev.* A **38** 3098
——1992 *J. Chem. Phys.* **96** 2155
——1996 *J. Chem. Phys.* **104** 1040-6
Bergersen B, Kus F W and Blomberg C 1973 *Can. J. Phys.* **51** 102–10
Berggren K-F and Sernelius B E 1981 *Phys. Rev.* B **24** 1971–86
Biermann L 1943 *Z. Astrophys.* **22** 157
Biermann L and Lübeck K 1948 *Z. Astrophys.* **25** 325

Blase X, Rubio A, Louie S G and Cohen M L 1995 *Phys. Rev.* B **52** R2225–8
Bobbert P A and van Haeringen W 1994 *Phys. Rev.* B **49** 10 326–31
Bohm D and Pines D 1953 *Phys. Rev.* **92** 609
Bonacic-Koutecky V, Fantucci P and Koutecky J 1990a *J. Chem. Phys.* **96** 3802
——1990b *Chem. Phys. Lett.* **166** 32
Bonacic-Koutecky V, Pittner J, Scheuch C, Guest M F and Koutecky J 1992 *J. Chem. Phys.* **96** 7938
Boys S F 1950 *Proc. R. Soc. London* A **200** 542
Bylander D M and Kleinman L 1995a *Phys. Rev. Lett.* **74** 3660
——1995b *Phys. Rev.* B **52** 14 566
Calandra C and Manghi F 1992 *Phys. Rev.* B **45** 5819
Cardona M, Weinstein M and Wolff G A 1965 *Phys. Rev.* **140** A633
Causa M and Zupan A 1994 *Chem. Phys. Lett.* **220** 145–53
Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566
Charlesworth J P A, Godby R W and Needs R J 1993 *Phys. Rev. Lett.* **70** 1685–8
Cheetham A K and Hope D A O 1983 *Phys. Rev.* B **27** 6964
Cowan R D 1967 *Phys. Rev.* **163** 54
Dal Corso A, Pasquarello A, Baldereschi A and Car R 1996 *Phys. Rev.* B **53** 1180–5
Daling R and van Haeringen W 1989 *Phys. Rev.* B **40** 11 659–65
Daling R, Unger P, Fulde P and van Haeringen W 1991 *Phys. Rev.* B **43** 1851–4
de Groot H J, Bobbert P A and van Haeringen W 1995 *Phys. Rev.* B **52** 11 000
Deisz J J and Eguiluz A 1997 *Phys. Rev.* B **55** 9195–9
Deisz J J, Eguiluz A and Hanke W 1993 *Phys. Rev. Lett.* **71** 2793–6
Del Sole R and Girlanda R 1996 *Phys. Rev.* B **54** 14 376–80
Del Sole R, Reining L and Godby R W 1994 *Phys. Rev.* B **49** 8024–8
Dreizler R M and Gross E K U 1990 *Density Functional Theory* (New York: Springer)
DuBois D F 1959a *Ann. Phys.* **7** 174
——1959b *Ann. Phys.* **8** 24
Dufek P, Blaha P, Sliwko V and Schwarz K 1994 *Phys. Rev.* B **49** 10 170–5
Dykstra C E 1988 *Ab Initio Calculation of the Structure and Properties of Molecules* (Amsterdam: Elsevier)
Eberhardt W and Plummer E W 1980 *Phys. Rev.* B **21** 3245
Eddy C R, Moustakas T D and Scanlon J 1993 *J. Appl. Phys.* **73** 448
Eguiluz A G, Heinrichsmeier M, Fleszar A and Hanke W 1992 *Phys. Rev. Lett.* **68** 1359–62
Engel G E and Farid B 1992 *Phys. Rev.* B **46** 15 812–27
——1993 *Phys. Rev.* B **47** 15 931–4
Engel G E, Farid B, Nex C M M and March N H 1991 *Phys. Rev.* B **44** 13 356–73
Engel G E, Kwon Y and Martin R M 1995 *Phys. Rev.* B **51** 13 538–46
Engel G E and Pickett W E 1996 *Phys. Rev.* B **54** 8420–9
Erwin S C and Pickett W E 1991 *Science* **254** 842
Faddeev L D 1963 *Sov. Phys.–JETP* **12** 275
Farid B 1996 *Phil. Mag.* B **76** 145–92
Farid B, Godby R W and Needs R J 1990 *20th Int. Conf. on the Physics of Semiconductors* vol 3, ed E M Anastassakis and J D Joannopoulos (Singapore: World Scientific) pp 1759–62
Feldcamp L A, Stearns M B and Shinozaki S S 1979 *Phys. Rev.* B **20** 1310
Fender B E F, Jacobson A J and Wedgwood F A 1968 *J. Chem. Phys.* **48** 990
Fetter A L and Walecka J D 1971 *Quantum Theory of Many-Particle Systems* (New York: McGraw-Hill)
Fock V 1930 *Z. Phys.* **61** 126
Frota H O and Mahan G D 1992 *Phys. Rev.* B **45** 6243
Fujimori A, Minami F and Sugano S 1984 *Phys. Rev.* B **29** 5225
Galitskii V M 1958 *Sov. Phys.–JETP* **7** 104
Galitskii V M and Migdal A B 1958 *Sov. Phys.–JETP* **7** 96
Gell-Mann M and Brueckner K 1957 *Phys. Rev.* **106** 364
Godby R W and Needs R J 1989 *Phys. Rev. Lett.* **62** 1169–72
Godby R W, Schlüter M and Sham L J 1986 *Phys. Rev. Lett.* **56** 2415–18
——1987a *Phys. Rev.* B **36** 6497–500
——1987b *Phys. Rev.* B **35** 4170–1
——1988 *Phys. Rev.* B **37** 10 159–75
Godby R W and Sham L J 1994 *Phys. Rev.* B **49** 1849–57
Grobman W D and Eastman D E 1972 *Phys. Rev. Lett.* **29** 1508

Gross E K U, Dobson J F and Petersilka M 1996 *Density Functional Theory* ed R F Nalewajski (New York: Springer)
Gu Z-q and Ching W Y 1994 *Phys. Rev.* B **49** 10 958–67
Guillot C, Ballu Y, Paigné J, Lecante J, Jain K P, Thiry P, Pinchaux R, Pétroff Y and Falicov L M 1977 *Phys. Rev. Lett.* **39** 1632
Gunnarsson O 1997 *J. Phys.: Condens. Matter* **9** 5635
Gunnarsson O and Jones R O 1980 *J. Chem. Phys.* **72** 5357
Gunnarsson O and Lundqvist B I 1976 *Phys. Rev.* B **13** 4274
Gunnarsson O, Lundqvist B I and Wilkins J W 1974 *Phys. Rev.* B **10** 1319
Gunnarsson O, Meden V and Schönhammer K 1994 *Phys. Rev.* B **50** 10 462
Gunnarsson O, Rainer D and Zwicknagl G 1992 *Int. J. Mod. Phys.* B **6** 3993
Gygi F and Baldereschi A 1989 *Phys. Rev. Lett.* **62** 2160
Hamada N, Hwang M and Freeman A J 1990 *Phys. Rev.* B **41** 3620–6
Hanke W and Sham L J 1975 *Phys. Rev.* B **12** 4501–11
——1988 *Phys. Rev.* B **38** 13 361–70
Harris J and Jones R O 1978 *J. Chem. Phys.* **68** 3316
Hartree D R 1928 *Proc. Camb. Phil. Soc.* **24** 89, 111, 426
Hayashi E and Shimizu M 1969 *J. Phys. Japan* **26** 1396
Hedin L 1965a *Phys. Rev.* A **139** 796
Hedin L 1965b *Arkiv för Fysik* **30** 231–58
——1980 *Phys. Scr.* **21** 477–80
——1995 *Int. J. Quantum Chem.* **56** 445–52
Hedin L and Johansson A 1969 *J. Phys. B: At. Mol. Phys.* **2** 1336
Hedin L and Lundqvist B I 1971 *J. Phys. C: Solid State Phys.* **4** 2064–83
Hedin L, Lundqvist B I and Lundqvist S 1970 *J. Res. Natl Bur. Stand. Sect.* A **74** 417
Hedin L and Lundqvist S 1969 *Solid State Physics* vol 23, ed H Ehrenreich, F Seitz and D Turnbull (New York: Academic)
Hellwege K-H and Madelung O (eds) 1982 *Numerical Data and Functional Relationships in Science and Technology (Landolt-Börnstein, New Series, Group 3)* vols 17a and 22a (Berlin: Springer)
Himpsel F J, Knapp J A and Eastman D E 1979 *Phys. Rev.* B **19** 2919
Himpsel F J, van der Veen J F and Eastman D E 1980 *Phys. Rev.* B **22** 1967
Hindgren M 1997 *PhD Thesis* University of Lund
Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
Holm and von Barth 1997 Private communication
Hong S and Mahan G D 1994 *Phys. Rev.* B **50** 8182–8
Hüfner S, Osterwalder J, Riester T and Hullinger F 1984 *Solid State Commun.* **52** 793
Hüfner S and Wertheim G K 1973 *Phys. Rev.* B **8** 4857
Hüfner S, Wertheim G K, Smith N V and Traum M M 1972 *Solid State Commun.* **11** 323
Hybertsen M S and Louie S G 1985a *Phys. Rev. Lett.* **55** 1418–21
——1985b *Phys. Rev.* B **32** 7005–8
——1986 *Phys. Rev.* B **34** 5390–413
——1987a *Phys. Rev.* B **35** 5585–601, 5602–10
——1987b *Comments Condens. Matter Phys.* **13** 223
——1988a *Phys. Rev.* B **37** 2733–6
——1988b *Phys. Rev.* B **38** 4033–44
Igarashi J 1983 *J. Phys. Soc. Japan* **52** 2827
——1985 *J. Phys. Soc. Japan* **54** 260
Igarashi J, Unger P, Hirai K and Fulde P 1994 *Phys. Rev.* B **49** 16 181
Inkson J C 1984 *Many-body Theory of Solids* (New York: Plenum)
Ishii Y, Ohnishi S and Sugano S 1986 *Phys. Rev.* B **33** 5271
Itchkawitz B S, Lyo I-W and Plummer E W 1990 *Phys. Rev.* B **41** 8075
Jenkins S J, Srivastava G P and Inkson J C 1993 *Phys. Rev.* B **48** 4388–97
Jensen E and Plummer E W 1985 *Phys. Rev. Lett.* **55** 1912
Johansson L S O, Uhrberg R I G, Mårtensson P and Hansson G V 1990 *Phys. Rev.* B **42** 1305
Jones R O and Gunnarsson O 1989 *Rev. Mod. Phys.* **61** 689–746
Kanamori J 1963 *Prog. Theor. Phys.* **30** 275
Kane E O 1971 *Phys. Rev.* B **4** 1910–7
Kanski J, Nilsson P O and Larsson C G 1980 *Solid State Commun.* **35** 397

Kemeny P C and Shevchik 1975 *Solid State Commun.* **17** 255
Kohn W and Sham L J 1965 *Phys. Rev.* A **140** 1133
Kotani T 1995 *Phys. Rev. Lett.* **74** 2989
Kotani T and Akai H 1996 *Phys. Rev.* B **54** 16 502
Kowalczyk S P, Ley L, McFeely F R, Pollak R A and Shirley D A 1973 *Phys. Rev.* B **8** 3583
Kress C, Fiedler M and Bechstedt F 1994 *Proc. 4th Int. Conf. on the Formation of Semiconductor Interfaces* ed B Lengeler *et al* (Singapore: World Scientific) p 19
Lad R J and Henrich V E 1988 *Phys. Rev.* B **38** 10 860
Landau L D 1956 *Zh. Eksp. Teor. Fiz.* **30** 1058
——1957 *Zh. Eksp. Teor. Fiz.* **32** 59
Lang N D and Kohn W 1973 *Phys. Rev.* B **7** 3541–50
Lang N D and Williams A R 1977 *Phys. Rev.* B **16** 2408
Langreth D C 1970 *Phys. Rev.* B **1** 471
Langreth D C and Mehl M J 1983 *Phys. Rev.* B **28** 1809
Lei T, Fanciulli M, Molnar R J, Moustakas T D, Graham R J and Scanlon J 1992b *Appl. Phys. Lett* **59** 944
Lei T, Moustakas T D, Graham R J, He Y and Berkowitz S J 1992a *J. Appl. Phys.* **71** 4933
Levine Z H and Louie S G 1982 *Phys. Rev.* B **25** 6310
Ley L, Kowalczyk S P, Pollak R A and Shirley D A 1972 *Phys. Rev. Lett.* **29** 1088
Lichtenstein A I, Zaanen J and Anisimov V I 1995 *Phys. Rev.* B **52** R5467
Liebsch A 1979 *Phys. Rev. Lett.* **43** 1431–4
——1981 *Phys. Rev.* B **23** 5203–12
Lindgren I 1971 *Int. J. Quantum Chem.* **5** 411
Lindhard J 1954 *Kgl. Danske Videnskap. Selskap, Mat.-fys. Medd.* **28** 8
Lof R W, van Veenendaal M A, Koopmans B, Jonkman H T and Sawatzky G A 1992 *Phys. Rev. Lett.* **68** 3924
Lundqvist B I 1967a *Phys. Kondens. Mater.* **6** 193
——1967b *Phys. Kondens. Mater.* **6** 206
——1968 *Phys. Kondens. Mater.* **7** 117
——1969 *Phys. Kondens. Mater.* **9** 236
Lundqvist B I and Samathiyakanit V 1969 *Phys. Kondens. Mater.* **9** 231
Luttinger J M and Ward J C 1960 *Phys. Rev.* **118** 1417–27
Lyo I-W and Plummer E W 1988 *Phys. Rev. Lett.* **60** 1558
Madelung O 1984 *Crystal and Solid State Physics (Landolt-Börnstein, New Series, Group 3)* vol 17 (Berlin: Springer)
Mahan G D 1990 *Many-particle Physics* (New York: Plenum)
——1994 *Comments Cond. Mater. Phys.* **16** 333
Mahan G D and Sernelius B E 1989 *Phys. Rev. Lett.* **62** 2718
Manghi F, Bellini V and Arcangeli C 1997 *Phys. Rev.* B **56** 7149–61
Manghi F, Calandra C and Ossicini S 1994 *Phys. Rev. Lett.* **73** 3129–32
Martin P C and Schwinger J 1959 *Phys. Rev.* **115** 1342–73
Mårtensson H and Nilsson P O 1984 *Phys. Rev.* B **30** 3047
Massidda S, Continenza A, Posternak M and Baldereschi A 1995a *Phys. Rev. Lett.* **74** 2323–6
——1997 *Phys. Rev.* B **55** 13 494–502
Massidda S, Resta R, Posternak M and Baldereschi A 1995b *Phys. Rev.* B **52** R16 977
Mattheiss L F 1972a *Phys. Rev.* B **5** 290
——1972b *Phys. Rev.* B **5** 306
McFeely F R, Kowalczyk S P, Ley L, Cavell R G, Pollak R A and Shirley D A 1974 *Phys. Rev.* B **9** 5268
Minnhagen P 1974 *J. Phys. C: Solid State Phys.* **7** 3013
Moore C E 1949, 1952, 1958 *Atomic Energy Levels, Natl. Bur. Stand. Ref. Data Ser., Natl. Bur. Stand. (US) Circ. No.* 467 (Washington DC: US GPO)
Moruzzi V L, Janak J F and Williams A R 1978 *Calculated Electronic Properties of Metals* (New York: Pergamon)
Mott N F 1949 *Phys. Soc. London* A **62** 416
Müller W, Flesch J and Meyer W 1984 *J. Chem. Phys.* **80** 3297–310
Northrup J E 1993 *Phys. Rev.* B **47** 10 032–5
Northrup J E, Hybertsen M S and Louie S G 1987 *Phys. Rev. Lett.* **59** 819
——1989 *Phys. Rev.* B **39** 8198
Nozières P 1964 *Theory of Interacting Fermi Systems* (New York: Benjamin)
Nozières P and de Dominicis C J 1969 *Phys. Rev.* **178** 1097
Onida G, Reining L, Godby R W, Del Sole R and Andreoni W 1995 *Phys. Rev. Lett.* **75** 818–21

Ortuno M and Inkson J C 1979 *J. Phys. C: Solid State Phys.* **12** 1065–71
Oschlies A, Godby R W and Needs R J 1992 *Phys. Rev.* B **45** 13 741–4
——1995 *Phys. Rev.* B **51** 1527–35
Overhauser A W 1971 *Phys. Rev.* B **3** 1888
——1985 *Phys. Rev. Lett.* **55** 1916
Paisley M J, Sitar Z, Posthill J B and Davis R F 1989 *J. Vac. Sci. Technol.* A **7** 701
Palummo M, Del Sole R, Reining L, Bechstedt F and Cappellini G 1995 *Solid State Commun.* **95** 393–8
Penn D R 1962 *Phys. Rev.* **128** 2093
——1979 *Phys. Rev. Lett.* **42** 921–4
Perdew J P, Burke K and Ernzerhof M 1997 *Phys. Rev. Lett.* **77** 3865–8
Perdew J P and Levy M 1983 *Phys. Rev. Lett.* **51** 1884
Perdew J P and Zunger A 1981 *Phys. Rev.* B **23** 5048
Philipsen P H T and Baerends E J 1996 *Phys. Rev.* B **54** 5326–33
Pickett W E 1989 *Rev. Mod. Phys.* **62** 433
Pickett W E and Wang C S 1984 *Phys. Rev.* B **30** 4719–33
Pines D 1961 *Elementary Excitations in Solids* (New York: Benjamin)
Pines D and Bohm D 1952 *Phys. Rev.* **85** 338
Powell R J and Spicer W E 1970 *Phys. Rev.* B **2** 2182
Quinn J J and Ferrell R A 1958 *Phys. Rev.* **112** 812
Rauch W 1940 *Z. Phys.* **116** 652
Rice T M 1965 *Ann. Phys.* **31** 100
Rohlfing M, Krüger P and Pollmann J 1993 *Phys. Rev.* B **48** 17 791–805
——1995a *Phys. Rev. Lett.* **19** 3489–92
——1995b *Phys. Rev.* B **52** 1905–17
——1996 *Phys. Rev.* B **54** 13 759–66
Rojas H N, Godby R W and Needs R J 1995 *Phys. Rev. Lett.* **74** 1827–30
Rubio A, Corkill J L, Cohen M L, Shirley E L and Louie S G 1993 *Phys. Rev.* B **48** 11 810–16
Runge E and Gross E K U 1984 *Phys. Rev. Lett.* **52** 997
Saito S and Cohen M L 1988 *Phys. Rev.* B **38** 1123
Saito S, Zhang S B, Louie S G and Cohen M L 1989 *Phys. Rev.* B **40** 3643–6
Satpathy S, Antropov V P, Andersen O K, Jepsen O, Gunnarsson O and Lichtenstein A I 1992 *Phys. Rev.* B **46** 1773
Sawatzky G A and Allen J W 1984 *Phys. Rev. Lett.* **53** 2339
Schindlmayr A and Godby R W 1997 *Preprint*
Schönberger U and Aryasetiawan F 1995 *Phys. Rev.* B **52** 8788–93
Schönhammer K and Gunnarsson O 1988 *Phys. Rev.* B **37** 3128
Schwinger J 1951 *Proc. Natl Acad. Sci., USA* **37** 452
Seidl A, Görling A, Vogl P, Majewski J A and Levy M 1996 *Phys. Rev.* B **53** 3764–74
Sham L J and Kohn W 1966 *Phys. Rev.* **145** 561–7
Sham L J and Schlüter M 1983 *Phys. Rev. Lett.* **51** 1888
——1985 *Phys. Rev.* B **32** 3883
Shen Z X, Shih C K, Jepsen O, Spicer W E, Lindau I and Allen J W 1990 *Phys. Rev. Lett.* **64** 2442
Shen Z X *et al* 1991a *Phys. Rev.* B **44** 3604
——1991b *Solid State Commun.* **79** 623
Shirley E L 1996 *Phys. Rev.* B **54** 7758–64
Shirley E L and Louie S G 1993 *Phys. Rev. Lett.* **71** 133
Shirley E L and Martin R M 1993 *Phys. Rev.* B **47** 15 404–27
Shirley E L, Zhu X and Louie S G 1992 *Phys. Rev. Lett.* **69** 2955–8
Shung K W-K and Mahan G D 1986 *Phys. Rev. Lett.* **57** 1076–79
——1988 *Phys. Rev.* B **38** 3856
Shung K W-K, Sernelius B E and Mahan G D 1987 *Phys. Rev.* B **36** 4499
Sinha S K 1969 *Phys. Rev.* B **177** 1256
Sitar Z *et al* 1992 *J. Matter. Sci. Lett.* **11** 261
Slater J C 1930 *Phys. Rev.* **35** 210
——1951a *Phys. Rev.* **81** 385
——1951b *Phys. Rev.* **82** 538
——1953 *Phys. Rev.* **91** 528
——1974 *The Self-consistent Field for Molecules and Solids (Quantum Theory of Molecules and Solids IV)*

(New York: McGraw-Hill)
Solovyev I V, Dederichs P H and Anisimov V I 1994 *Phys. Rev.* B **50** 16 861–71
Solovyev I, Hamada N and Terakura K 1996 *Phys. Rev.* B **53** 7158
Springer M, Svendsen P S and von Barth U 1996 *Phys. Rev.* B **54** 17 392–401
Steiner P, Höchst H and Hüfner S 1979 *Photoemission in Solids II (Topics in Applied Physics 27)* ed L Ley and M Cardona (Heidelberg: Springer)
Sterne P A and Inkson J C 1984 *J. Phys. C: Solid State Phys.* **17** 1497–510
Strinati G, Mattausch H J and Hanke W 1982 *Phys. Rev.* B **25** 2867–88
Surh M P, Chacham H and Louie S G 1995 *Phys. Rev.* B **51** 7464–70
Surh M P, Northrup J E, and Louie S G 1988 *Phys. Rev.* B **38** 5976–80
Svane A and Gunnarsson O 1990 *Phys. Rev. Lett.* **65** 1148–51
Svendsen P S and von Barth U 1996 *Phys. Rev.* B **54** 17 402–13
Szotek Z, Temmerman W M and Winter H 1993 *Phys. Rev.* B **47** 4029
Talman J D and Shadwick W F 1976 *Phys. Rev.* A **14** 36
Terakura K, Oguchi T, Williams A R and Kübler J 1984 *Phys. Rev.* B **30** 4734–47
Ting C S, Lee T K and Quinn J J 1975 *Phys. Rev. Lett.* **34** 870
Tjeng L H, Chen C T, Ghijsen J, Rudolf P and Sette F 1991 *Phys. Rev. Lett.* **67** 501
Troullier N and Martins J L 1992 *Phys. Rev.* B **46** 1754
Uhrberg R I G, Hansson G V, Nicholls J M and Flodström S A 1981 *Phys. Rev.* B **24** 4684
van Elp J *et al* 1991 *Phys. Rev.* B **44** 1530
van Leeuwen R 1996 *Phys. Rev. Lett.* **76** 3610–13
Verdozzi C, Godby R W and Holloway S 1995 *Phys. Rev. Lett.* **74** 2327–30
von Barth U and Hedin L 1972 *J. Phys. C: Solid State Phys.* **5** 1629–42
von Barth U and Holm B 1996 *Phys. Rev.* B **54** 8411–19
von der Linden W and Horsch P 1988 *Phys. Rev.* B **37** 8351
Walter J P and Cohen M L 1972 *Phys. Rev.* B **5** 3101
Wang C S and Pickett W E 1983 *Phys. Rev. Lett.* **51** 597–600
Wang C, Pollack S, Cameron D and Kappes M M 1990 *J. Chem. Phys.* **93** 3787
Ward J C 1950 *Phys. Rev.* **78** 182
Watson R E, Herbst J F, Hodges L, Lundqvist B I and Wilkins J W 1976 *Phys. Rev.* B **13** 1463
Whited R C, Flaten C J and Walker W C 1973 *Solid State Commun.* **13** 1903
Wick G C 1950 *Phys. Rev.* **80** 268
Zakharov O, Rubio A, Blase X, Cohen M L and Louie S G 1994 *Phys. Rev.* B **50** 10 780–7
Zhang S B, Tomànek D, Louie S G, Cohen M L and Hybertsen M S 1988 *Solid State Commun.* **66** 585–8
Zhang S B, Wei S-H and Zunger A 1995 *Phys. Rev.* B **52** 13 975
Zhu X and Louie S G 1991 *Phys. Rev.* B **43** 14 142–56
Zunger A and Lindefelt U 1983 *Solid State Commun.* **45** 343
Zunger A, Perdew J P and Oliver G L 1980 *Solid State Commun.* **34** 933