

1 A First-Principles Scheme for Calculating the Electronic Structure of Strongly Correlated Materials: GW+DMFT

Ferdi Aryasetiawan, Silke Biermann, and Antoine Georges

1.1 Introduction

The last few decades have witnessed substantial progress in the field of electronic structure of materials. Using density functional theory (DFT) [1,2] within the local density approximation (LDA) or generalized gradient approximation (GGA) [3] it is quite routine to calculate the electronic structure of relatively complicated materials containing tens of atoms per unit cell. The success of LDA, however, is also accompanied by a number of serious problems. It was noticed very early on that, when applied to calculate the band structures of s–p semiconductors and insulators, the band gaps are systematically underestimated by some tens of percents. Apart from the too small gaps, the band dispersions are very reasonable. This remarkable property of the LDA is still waiting for an explanation since formally there is no theoretical justification for identifying the one-particle Kohn–Sham eigenvalues as quasiparticle energies observed in photoemission experiments. Applications to alkali metals also indicate some problems, albeit less serious. When the band dispersions are compared with photoemission data, they are found to be too wide by 10–30%. Some many-body calculations of the electron gas [4, 5], however, suggest that the band widths are actually widened compared with the free-electron values and that the LDA performs better than is commonly believed. If this turns out to be true, photoemission data would presumably need a complete revision. In any case, the LDA errors in s–p metals are probably less significant than the band gap errors in semiconductors and insulators.

A much more serious problem of the LDA arises when it is applied to calculate the electronic structures of so-called “strongly correlated systems”. We have to be more precise with what we mean by correlations. Even in the electron gas, correlation as conventionally defined is rather large. It is as large as exchange so that the two almost cancel each other leaving the free-electron band essentially unchanged. Thus, it is more appropriate in our case to define correlation as anything beyond the LDA rather than anything beyond the Fock exchange since the former is usually our starting point in electronic structure calculations of solids.

Strongly correlated systems are characterized by partially occupied localized orbitals such as found in transition metal oxides or 4f metals. Here the problem is often more of a *qualitative* rather than a *quantitative* nature. It is often found that the LDA predicts a transition metal oxide to be a metal whereas experimentally it is an antiferromagnetic insulator. To cite some examples, LaMnO_3 , famous for its colossal magnetoresistance, and La_2CuO_4 , a well-known parent compound of high-temperature superconductors, are antiferromagnetic insulators but

predicted to be metals by the LDA [6]. In cases where the LDA does predict the correct structure, it is legitimate to ask if the one-particle spectrum is also reproduced correctly. According to the currently accepted interpretation, transition metal oxides may be classified as charge-transfer insulators [7, 8], which are characterized by the presence of occupied and unoccupied 3d bands with the oxygen 2p band in between. The gap is then formed by the oxygen 2p and unoccupied 3d bands, unlike the gap in LDA, which is formed by the 3d states (Mott–Hubbard gap). A more appropriate interpretation is to say that the highest valence state is a charge-transfer state: During photoemission a hole is created in the transition metal site but due to the strong 3d Coulomb repulsion it is energetically more favorable for the hole to hop to the oxygen site despite the cost in energy transfer. A number of experimental data, notably 2p core photoemission resonance, suggest that the charge-transfer picture is more appropriate to describe the electronic structure of transition metal oxides. And of course in the case of 4f metals, the LDA, being a one-particle theory, is totally incapable of yielding the incoherent part of the spectral function or satellite structures.

The above difficulties encountered by the LDA have prompted a number of attempts at improving the LDA. Notable among these is the GW approximation (GWA), developed systematically by Hedin in the early sixties [9]. He showed that the self-energy can be formally expanded in powers of the screened interaction W , the lowest term being iGW , where G is the Green function. Due to computational difficulties, for a long time the applications of the GWA were restricted to the electron gas but with the rapid progress in computer power, applications to realistic materials eventually became possible about two decades ago. Numerous applications to semiconductors and insulators reveal that in most cases the GWA [10, 11] removes a large fraction of the LDA band-gap error. Applications to alkali metals show band narrowing from the LDA values and account for more than half of the LDA error (although controversy about this issue still remains [12]).

The success of the GWA in sp materials has prompted further applications to more strongly correlated systems. For this type of materials the GWA has been found to be less successful. For example, GW calculation on nickel [13] does reproduce the photoemission quasiparticle band structure rather well, as compared with the LDA one where the 3d band width is too large by about 1 eV, but the too large LDA exchange splitting of 0.6 eV (experimentally 0.3 eV) remains essentially unchanged. Moreover, the famous 6 eV satellite is not reproduced. Application to NiO [14], a prototype of transition metal oxides, also reveals some shortcomings. One problem is related to the starting Green's function, usually constructed from the LDA Kohn–Sham orbitals and energies. In the LDA the band gap is very small, about 0.2 eV compared with the 4 eV experimental band gap. A commonly used procedure of performing a one-iteration GW calculation yields about 1 eV gap, much too small. This problem is solved by performing a partial self-consistency, where knowledge of the self-energy from the previous iteration is used to construct a better starting one-particle Hamiltonian [14]. This procedure improves the band gap considerably to a self-consistent value of 5.5 eV and at the same time increases the LDA magnetic moment from $0.9 \mu_B$ to about $1.6 \mu_B$ much closer to the experimental value of $1.8 \mu_B$. However, the GWA maintains the Mott–Hubbard gap, i.e., the gap is formed by the 3d states as in the LDA, instead of the charge-transfer gap. In other words, the top of the valence band is dominated by the Ni 3d. A more recent calculation using a more refined procedure of partial self-consistency has also confirmed these results [15]. The

problem with the GWA appears to arise from inadequate account of short-range correlations, probably not properly treated in the random-phase approximation (RPA).

Attempts at improving the LDA to treat strongly correlated systems were initiated by the LDA+U method [16–19], which introduces, on top of the LDA Hamiltonian, a Hubbard U term and a double-counting correction term, usually applied to partially filled 3d or 4f shells. The LDA+U method is essentially a Hartree–Fock approximation to the LDA+U Hamiltonian. In the LDA, the Kohn–Sham potential does not explicitly distinguish between occupied and unoccupied orbitals so that they experience the same potential. In, for example, transition metal oxides, where the 3d orbitals are partially occupied, this leads to metallicity or to underestimation of the band gap. The LDA+U cures this problem by approximately pushing down the occupied orbitals by $U/2$ and pushing up the unoccupied orbitals by $U/2$, creating a lower and upper Hubbard band, thus opening up a gap of the order of the Hubbard U . The LDA+U method has been successfully applied to late transition metal oxides, rare earth compounds such as CeSb, as well as to problems involving metal–insulator transition and charge-orbital ordering.

More recently, the idea of the LDA+U was extended further by treating the Hubbard U term in a more sophisticated fashion utilizing the dynamical mean-field theory (DMFT) [20]. The DMFT is remarkably well suited for treating systems with strong on-site correlations because the on-site electronic Coulomb interactions are summed to all orders. This is achieved by using a mapping onto a self-consistent quantum impurity problem, thereby including the effects of the surrounding in a mean-field approximation. The strength of the DMFT is its ability to properly describe Mott phenomenon or the formation of local moments, which is the key to understanding many physical properties in strongly correlated materials. The combination of LDA and DMFT takes advantage of the first-principles nature of LDA while at the same time incorporates local correlation effects not properly treated within the LDA. The LDA+DMFT method [19, 21, 22] has now been successfully applied to a number of systems.

In both the LDA+U and LDA+DMFT methods, two fundamental problems remain unaddressed. First, the Hubbard U is usually treated as a parameter, and second, the Hubbard U term contains interaction already included in the LDA but it is not clear how to take into account this double-counting term in a precise way. Thus, a truly first-principles theory for treating strongly correlated systems is still lacking. In this chapter, we describe a dynamical mean-field approach for calculating the electronic structure of strongly correlated materials from first-principles [23, 24]. The DMFT is combined with the GW method, which enables one to treat strong interaction effects [25]. One of the main features of the new scheme is that the Hubbard U is calculated from first principles through a self-consistency requirement on the on-site screened Coulomb interaction, analogous to the self-consistency in the local Green’s function in the DMFT. Since the GWA has an explicit diagrammatic representation, the on-site contribution of the GW self-energy can be readily identified and the scheme then allows for a precise double-counting correction.

In the next two sections, we will give a summary of the GWA and DMFT, describing their main features. In the fourth section we lay out the GW+DMFT scheme, followed by a simplified application of the scheme to the excitation spectrum of nickel. Finally we discuss some future challenges and directions.

1.2 The GW Approximation

1.2.1 Theory

It can be shown that the self-energy may be expressed as [9]

$$\Sigma(1, 2) = -i \int d3 d4 v(1, 4)G(1, 3) \frac{\delta G^{-1}(3, 2)}{\delta \phi(4)} \quad (1.1)$$

where v is the bare Coulomb interaction, G is the Green function and ϕ is an external time-dependent probing field. We have used the short-hand notation $1 = (x_1 t_1)$. From the equation of motion of the Green function

$$G^{-1} = i \frac{\partial}{\partial t} - H_0 - \Sigma \quad (1.2)$$

$$H_0 = h_0 + \phi + V_H \quad (1.3)$$

h_0 is the kinetic energy and V_H is the Hartree potential. We then obtain

$$\begin{aligned} \frac{\delta G^{-1}(3, 2)}{\delta \phi(4)} &= -\delta(3-2) \left[\delta(3-4) + \frac{\delta V_H(3)}{\delta \phi(4)} \right] - \frac{\delta \Sigma(3, 2)}{\delta \phi(4)} \\ &= -\delta(3-2) \epsilon^{-1}(3, 4) - \frac{\delta \Sigma(3, 2)}{\delta \phi(4)} \end{aligned} \quad (1.4)$$

where ϵ^{-1} is the inverse dielectric matrix. The GWA is obtained by neglecting the vertex correction $\delta \Sigma / \delta \phi$, which is the last term in Eq. (1.4). This is just the random-phase approximation (RPA) for ϵ^{-1} . This leads to

$$\Sigma(1, 2) = iG(1, 2)W(1, 2) \quad (1.5)$$

where we have defined the screened Coulomb interaction W by

$$W(1, 2) = \int d3 v(1, 3) \epsilon^{-1}(3, 2) \quad (1.6)$$

The RPA dielectric function is given by

$$\epsilon = 1 - vP \quad (1.7)$$

where

$$\begin{aligned} P(\mathbf{r}, \mathbf{r}'; \omega) &= -2i \int \frac{d\omega'}{2\pi} G(\mathbf{r}, \mathbf{r}'; \omega + \omega') G(\mathbf{r}', \mathbf{r}; \omega') \\ &= 2 \sum_i^{\text{occ}} \sum_j^{\text{unocc}} \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}') \psi_j^*(\mathbf{r}) \psi_j(\mathbf{r}') \\ &\quad \times \left\{ \frac{1}{\omega - \epsilon_j + \epsilon_i + i\delta} - \frac{1}{\omega + \epsilon_j - \epsilon_i - i\delta} \right\} \end{aligned} \quad (1.8)$$

with the Green function constructed from a one-particle band structure $\{\psi_i, \varepsilon_i\}$. The factor of 2 arises from the sum over spin variables. In frequency space, the self-energy in the GWA takes the form

$$\Sigma(r, r'; \omega) = \frac{i}{2\pi} \int d\omega' e^{i\eta\omega'} G(\mathbf{r}, \mathbf{r}'; \omega + \omega') W(\mathbf{r}, \mathbf{r}'; \omega') \quad (1.9)$$

We have so far described the zero temperature formalism. For finite temperature we have

$$P(\mathbf{r}, \mathbf{r}'; i\nu_n) = \frac{2}{\beta} \sum_{\omega_k} G(\mathbf{r}, \mathbf{r}'; i\nu_n + i\omega_k) G(\mathbf{r}', \mathbf{r}; i\omega_k) \quad (1.10)$$

$$\Sigma(r, r'; i\omega_n) = -\frac{1}{\beta} \sum_{\nu_k} G(r, r'; i\omega_n + i\nu_k) W(r, r'; i\nu_k) \quad (1.11)$$

In the Green function language, the Fock exchange operator in the Hartree–Fock approximation (HFA) can be written as iGv . We may therefore regard the GWA as a generalization of the HFA, where the bare Coulomb interaction v is replaced by a screened interaction W . We may also think of the GWA as a mapping to a polaron problem where the electrons are coupled to some bosonic excitations (e.g., plasmons) and the parameters in this model are obtained from first-principles calculations.

The replacement of v by W is an important step in solids where screening effects are generally rather large relative to exchange, especially in metals. For example, in the electron gas, within the GWA, exchange and correlation are approximately equal in magnitude, to a large extent canceling each other, modifying the free-electron dispersion slightly. But also in molecules, accurate calculations of the excitation spectrum cannot neglect the effects of correlations or screening. The GWA is physically sound because it is qualitatively correct in some limiting cases [26].

1.2.2 The GW Approximation in Practice

The quality of the GWA may be seen in Figure 1.1, where a plot of band gaps of a number of well known semiconductors and insulators is displayed. It is clear from the plot that the LDA systematically underestimates the band gaps and that the GWA substantially improves the LDA band gaps. It has been found that for some materials, like MgO and InN, significant error still remains within the GWA. The reason for the discrepancy has not been understood well. One possible explanation is that the result of the one-iteration GW calculation may depend on the starting one-particle band structure. For example, in the case of InN, the starting LDA band structure has no gap. This may produce a metal-like (over)screened interaction W which fails to open up a gap or yields too small a gap in the GW calculation. Similar behavior is also found in the more extreme case of NiO, where a one-iteration GW calculation only yields a gap of about 1 eV starting from an LDA gap of 0.2 eV (the experimental gap is 4 eV) [10, 14].

The problems with the GWA arise when it is applied to strongly correlated systems. Application to ferromagnetic nickel [13] illustrates some of the difficulties with the GWA. Starting from the LDA band structure, a one-iteration GW calculation does improve significantly the LDA band structure. In particular it reduces the too large 3d band width bringing it into

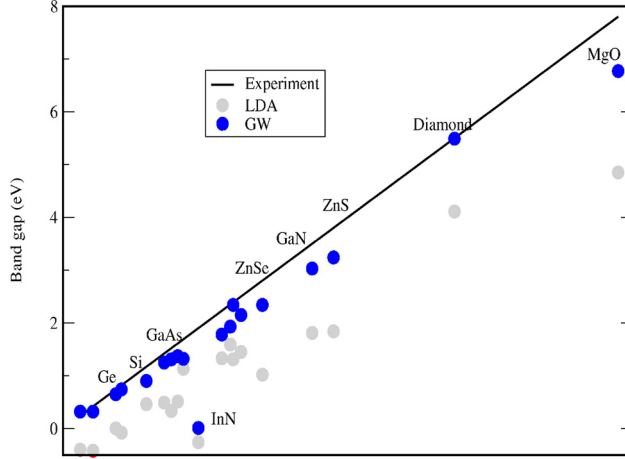


Figure 1.1: Band gaps of some selected semiconductors and insulators calculated within the GWA compared with the LDA and experimental values. The GW data are taken from [27].

much better agreement with photoemission data. However, the too large LDA exchange splitting (0.6 eV compared with the experimental value of 0.3 eV) remains essentially unchanged. Moreover, the famous 6 eV satellite, which is of course missing in the LDA, is not reproduced. These problems point to deficiencies in the GWA in describing short-range correlations since we expect that both exchange splitting and satellite structure are influenced by on-site interactions. In the case of exchange splitting, long-range screening also plays a role in reducing the HF value and the problem with the exchange splitting indicates a lack of spin-dependent interaction in the GWA: In the GWA the spin dependence only enters in G not in W .

Application to NiO, the prototype of Mott–Hubbard transition metal oxides, reveals another difficulty with the one-iteration GWA. As already mentioned previously, when the starting band structure is far from the experimental quasiparticle band structure, a one-iteration GW calculation may not be sufficient. This problem may be circumvented by performing a partial self-consistent calculation in which the self-energy from the previous iteration at a given energy, such as the Fermi energy of the center of the band of interest, is used to construct a new set of one-particle orbitals. This procedure is continued to self-consistency such that the starting one-particle band structure gives zero self-energy correction [10, 14, 15]. A more serious problem, however, is describing the charge-transfer character of the top of the valence band. The GWA essentially still maintains the Mott–Hubbard band gap as in the LDA, i.e., the top of the valence band is mainly of 3d character rather than the charge-transfer character dominated by the 2p oxygen hole. As in nickel, the problem with the satellite arises again. Depending on the starting band structure, a satellite may be reproduced albeit at a too high energy. Thus there is a strong need to improve the short-range correlations in the GWA which may be achieved by using a suitable approach based on the dynamical mean-field theory described in the next section.

1.3 Dynamical Mean Field Theory

Dynamical mean field theory (DMFT) [20] was originally developed within the context of *models* for correlated fermions on a lattice, where it has proven very successful for determining the phase diagrams or for calculations of excited states properties. It is a non-perturbative method and as such appropriate for systems with any strength of the interaction. In recent years, combinations of DMFT with band structure theory, in particular density functional theory with the local density approximation (LDA) have emerged. The idea is to correct for shortcomings of DFT-LDA due to strong Coulomb interactions and localization (or partial localization) phenomena that cause effects very different from a homogeneous itinerant behavior. Such signatures of correlations are well-known in transition metal oxides or f-electron systems but are also present in several elemental transition metals.

The application of DMFT to *real solids* relies on a *representability conjecture* assuming that local quantities, for example the local Green's function or self-energy of a solid, can be calculated from a local impurity model, that is one can find a *dynamical mean field* \mathcal{G}_0 and a Hubbard parameter U , such that the Green's function calculated from the effective action

$$\begin{aligned}
S = & \int_0^\beta d\tau \sum_{m\sigma} c_{m\sigma}^\dagger(\tau) \mathcal{G}_{0mm'\sigma}^{-1}(\tau - \tau') c_{m'\sigma}(\tau') \\
& + \frac{1}{2} \int_0^\beta d\tau \sum_{mm'\sigma} U_{mm'} n_{m\sigma}(\tau) n_{m'-\sigma}(\tau) \\
& + \frac{1}{2} \int_0^\beta d\tau \sum_{m \neq m'\sigma} (U_{mm'} - J_{mm'}) n_{m\sigma}(\tau) n_{m'\sigma}(\tau)
\end{aligned} \tag{1.12}$$

coincides with the local Green's function of the solid. For a model of correlated fermions on a lattice with infinite coordination number this conjecture can rigorously be proven: it is a consequence of the absence of any non-local contributions to the self-energy of the system. For a real solid the situation is somewhat more complicated, not only due to the finite coordination number but also to the difficulty in defining the notion of locality. This notion is crucial at both stages, for the construction of the impurity model, where *long-range* Coulomb interactions are mimicked by *local* Hubbard parameters¹, and for the resolution of the model within DMFT, which approximates the full self-energy of the model by a local quantity. Applications of DMFT to electronic structure calculations (e.g. the LDA+DMFT method) are therefore always defined within a specific basis set using localized basis functions. Within an LMTO implementation for example locality can naturally be defined as referring to the same muffin tin sphere. This amounts to defining matrix elements $G_{LR,L'R'}(i\omega)$ of the full Green's function

$$G(\mathbf{r}, \mathbf{r}', i\omega) = \sum_{LL'RR'} \chi_{LR}^*(\mathbf{r}) G_{LR,L'R'}(i\omega) \chi_{L'R'}(\mathbf{r}')$$

and assuming that its local, that is “on-sphere” part equals the Green's function of the local impurity model (1.12). Here \mathbf{R}, \mathbf{R}' denote the coordinates of the centers of the muffin tin

¹ For a discussion of the appropriateness of local Hubbard parameters see [28].

spheres, while r, r' can take any values. The index $L = (n, l, m)$ regroups all radial and angular quantum numbers. The dynamical mean field \mathcal{G}_0 in Eq. (1.12) has to be determined in such a way that the Green's function $G_{\text{impurity}L, L', \sigma}$ of the impurity model Eq. (1.12) coincides with $G_{LR, L'R'}(i\omega)$ if the impurity model self-energy is used as an estimate for the true self-energy of the solid. This self-consistency condition reads

$$G_{\text{impurity}}(i\omega_n) = \sum_k (i\omega_n + \mu - H_o(k) - \Sigma(i\omega_n))^{-1}$$

where Σ, H_0 and G are matrices in orbital and spin space, and $i\omega + \mu$ is a matrix proportional to the unit matrix in that space.

Together with Eq. (1.12) this defines the DMFT equations that have to be solved self-consistently. Note that the main approximation of DMFT is hidden in the self-consistency condition where the local self-energy has been promoted to the full lattice self-energy.

The representability assumption can actually be extended to other quantities of a solid than its local Green's function and self-energy. In “extended DMFT” [29–32] e.g. a two particle correlation function is calculated and can then be used in order to represent the local screened Coulomb interaction W of the solid. This is the starting point of the “GW+DMFT” scheme described in Section 1.4.

1.3.1 DMFT in Practice

Combinations of DFT-LDA and DMFT, so-called “LDA+DMFT” techniques [21], have so far been applied to transition metals (Fe, Ni, Mn) and their oxides (e.g. La/YTiO₃, V₂O₃, Sr/CaVO₃) as well as elemental f-electron materials (Pu, Ce) and their compounds. In the most general formulation, one starts from a many-body Hamiltonian of the form

$$\begin{aligned} H = & \sum_{\{im\sigma\}} (H_{im, i'm'}^{\text{LDA}} - H^{\text{dc}}) a_{im\sigma}^{\dagger} a_{i'm'\sigma} \\ & + \frac{1}{2} \sum_{imm'\sigma} U_{mm'}^i n_{im\sigma} n_{im'\sigma} \\ & + \frac{1}{2} \sum_{im \neq m'\sigma} (U_{mm'}^i - J_{mm'}^i) n_{im\sigma} n_{im'\sigma}, \end{aligned} \quad (1.13)$$

where H^{LDA} is the effective Kohn–Sham Hamiltonian derived from a self-consistent DFT-LDA calculation. This one-particle Hamiltonian is then corrected by Hubbard terms for direct and exchange interactions for the “correlated” orbitals, e.g. d or f orbitals. In order to avoid double counting of the Coulomb interactions for these orbitals, a correction term H^{dc} is subtracted from the original LDA Hamiltonian. The resulting Hamiltonian (1.13) is then treated within dynamical mean field theory by assuming that the many-body self-energy associated with the Hubbard interaction terms can be calculated from a multi-band impurity model.

This general scheme can be simplified in specific cases, e.g. in systems with a separation of the correlated bands from the “uncorrelated” ones, an effective model of the correlated bands can be constructed; symmetries of the crystal structure can be used to reduce the number of components of the self-energy etc.

Despite the huge progress made in the understanding of the electronic structure of correlated materials thanks to such LDA+DMFT schemes, certain conceptual problems remain open: These are related to the choice of the Hubbard interaction parameters and to the double counting corrections. An a priori choice of which orbitals are treated as correlated and which orbitals are left uncorrelated has to be made, and the values of U and J have to be fixed. Attempts at calculating these parameters from constrained LDA techniques are appealing in the sense that one can avoid introducing external parameters to the theory, but suffer from the conceptual drawback in that screening is taken into account in a static manner only [28]. Finally, the double counting terms are necessarily ill defined due to the impossibility of singling out in the LDA treatment contributions to the interactions stemming from specific orbitals. These drawbacks of LDA+DMFT provide a strong motivation to attempt the construction of an electronic structure method for correlated materials beyond combinations of LDA and DMFT.

1.4 GW+DMFT

The idea behind combining the GWA and the DMFT is to take advantage of the strong features of the two theories. The GWA, being based on RPA screening, is capable of taking into account long-range correlations but does not describe properly short-range correlations. On the other hand, the strength of the DMFT is its ability to describe on-site correlations by means of a mapping of the many-body problem onto an impurity problem where the on-site interactions are summed to all orders. While the GWA is a fully first-principles theory, the DMFT has been traditionally used in conjunction with a Hubbard model. Clearly, it is desirable to combine the two theories into a consistent theory where the parameters in the Hubbard model are determined self-consistently from first-principles via the GWA.

Recently, first steps have been undertaken towards a combination of the GWA and DMFT, both in a model context [24] and within the framework of realistic electronic structure calculations [23]. The basic physical idea of GW+DMFT is to separate the lattice into an on-site part and the rest. The on-site self-energy is taken to be the impurity self-energy calculated by the DMFT and the off-site self-energy is calculated by the GWA. Viewed from the GWA, we replace the on-site GW self-energy by that of DMFT, correcting the GW treatment of on-site correlations. Viewed from the DMFT, we add off-site contributions to the self-energy approximated within the GWA, giving a momentum dependent self-energy.

The impurity problem contains a Hubbard interaction U that is usually treated as a parameter. In order to calculate the U , we introduce a self-consistency condition that the U screened by the effective bath in the impurity model be equal to the local projection of the global screened interaction W . This condition is complementary to the condition imposed in the DMFT that the impurity Green's function be equal to the local Green's function.

The above physical ideas can be formulated in a precise way using the free-energy functional of Luttinger and Ward (LW). A generalization of the original LW functional takes the

form [33] (see also Ref. [34])

$$\begin{aligned} \Gamma(G, W) = & \text{Tr} \ln G - \text{Tr}[(G_{\text{H}}^{-1} - G^{-1})G] - \frac{1}{2} \text{Tr} \ln W \\ & + \frac{1}{2} \text{Tr}[(v^{-1} - W^{-1})W] + \Psi[G, W] \end{aligned} \quad (1.14)$$

$G_{\text{H}}^{-1} = i\omega_n + \mu + \nabla^2/2 - V_{\text{H}}$ corresponds to the Hartree Green's function with V_{H} being the Hartree potential. The functional $\Psi[G, W]$ is a generalization of the original LW $\Phi[G]$ functional, whose derivative with respect to G gives the self-energy. A more general derivation of (1.14) using a Hubbard-Stratonovich transformation and a Legendre transformation with respect to both G and W may be found in a later work [34]. It is straightforward to verify that at equilibrium the stationarity of Γ yields

$$\frac{\delta \Gamma}{\delta G} = 0 \rightarrow G^{-1} = G_{\text{H}}^{-1} - \Sigma, \quad \Sigma = \frac{\delta \Psi}{\delta G} \quad (1.15)$$

$$\frac{\delta \Gamma}{\delta W} = 0 \rightarrow W^{-1} = v^{-1} - P, \quad P = -2 \frac{\delta \Psi}{\delta W} \quad (1.16)$$

The functional Ψ is divided into on-site and off-site components:

$$\Psi = \Psi_{GW}^{\text{off-site}}[G^{RR'}, W^{RR'}] + \Psi_{\text{imp}}^{\text{on-site}}[G^{RR}, W^{RR}] \quad (1.17)$$

where R denotes a lattice site. In the GWA the functional Ψ is given by

$$\Psi_{GW}[G, W] = \frac{1}{2} G W G \quad (1.18)$$

The impurity part Ψ_{imp} is generated from a local quantum impurity problem defined on a single atomic site with an effective action

$$\begin{aligned} S = & \int d\tau d\tau' \left[- \sum c_L^{\dagger}(\tau) \mathcal{G}_{LL'}^{-1}(\tau - \tau') c_{L'}(\tau') \right. \\ & \left. + \frac{1}{2} \sum : c_{L_1}^{\dagger}(\tau) c_{L_2}(\tau) : \mathcal{U}_{L_1 L_2 L_3 L_4}(\tau - \tau') : c_{L_3}^{\dagger}(\tau') c_{L_4}(\tau') : \right] \end{aligned} \quad (1.19)$$

The double dots denote normal ordering and L refers to an orbital of angular momentum L on a given sphere where the impurity problem is defined. These orbitals are usually partially filled localized 3d or 4f orbitals.

The GW+DMFT set of equations can now be readily derived from Eqs. (1.14), (1.17), and (1.18) by taking functional derivatives of Ψ with respect to G and W as in Eqs. (1.15) and (1.16):

$$\Sigma = \Sigma_{GW}^{RR'}(1 - \delta_{RR'}) + \Sigma_{\text{imp}}^{RR} \delta_{RR'} \quad (1.20)$$

$$P = P_{GW}^{RR'}(1 - \delta_{RR'}) + P_{\text{imp}}^{RR} \delta_{RR'} \quad (1.21)$$

In practice the self-energy is expanded in some basis set $\{\phi_L\}$ localized in a site. The polarization function on the other hand is expanded in a set of two-particle basis functions $\{\phi_L \phi_{L'}\}$

(product basis) since the polarization corresponds to a two-particle propagator. For example, when using the linear muffin-tin orbital (LMTO) band-structure method, the product basis consists of products of LMTOs. These product functions are generally linearly dependent and a new set of optimized product basis (OPB) [10] is constructed by forming linear combinations of product functions, eliminating the linear dependences. We denote the OPB set by $B_\alpha = \sum_{LL'} \phi_L \phi_{L'} c_{LL'}^\alpha$. To summarize, one-particle quantities like G and Σ are expanded in $\{\phi_L\}$ whereas two-particle quantities such as P and W are expanded in the OPB set $\{B_\alpha\}$. It is important to note that the number of $\{B_\alpha\}$ is generally smaller than the number of $\{\phi_L \phi_{L'}\}$ so that quantities expressed in $\{B_\alpha\}$ can be expressed in $\{\phi_L \phi_{L'}\}$, but not vice versa. In momentum space, Eqs. (1.20) and (1.21) read

$$\Sigma^{LL'}(\mathbf{k}, i\omega_n) = \Sigma_{GW}^{LL'}(\mathbf{k}, i\omega_n) - \sum_{\mathbf{k}} \Sigma_{GW}^{LL'}(\mathbf{k}, i\omega_n) + \Sigma_{\text{imp}}^{LL'}(i\omega_n) \quad (1.22)$$

$$P^{\alpha\beta}(\mathbf{k}, i\omega_n) = P_{GW}^{\alpha\beta}(\mathbf{k}, i\omega_n) - \sum_{\mathbf{k}} P_{GW}^{\alpha\beta}(\mathbf{k}, i\omega_n) + P_{\text{imp}}^{\alpha\beta}(i\omega_n) \quad (1.23)$$

The second terms in the above two equations remove the on-site contributions of the GW self-energy and polarization, which are already included in Σ_{imp} and P_{imp} .

We are now in a position to outline the self-consistency loop which determines \mathcal{G} and \mathcal{U} as well as the full G and W self-consistently.

- The impurity problem (1.19) is solved, for a given choice of Weiss field $\mathcal{G}_{LL'}$ and Hubbard interaction $\mathcal{U}_{\alpha\beta}$: the ‘‘impurity’’ Green’s function

$$G_{\text{imp}}^{LL'} \equiv -\langle T_\tau c_L(\tau) c_{L'}^\dagger(\tau') \rangle_S \quad (1.24)$$

is calculated, together with the impurity self-energy

$$\Sigma_{\text{imp}} \equiv \delta\Psi_{\text{imp}}/\delta G_{\text{imp}} = \mathcal{G}^{-1} - G_{\text{imp}}^{-1}. \quad (1.25)$$

The two-particle correlation function

$$\chi_{L_1 L_2 L_3 L_4} = \langle : c_{L_1}^\dagger(\tau) c_{L_2}(\tau) :: c_{L_3}^\dagger(\tau') c_{L_4}(\tau') : \rangle_S \quad (1.26)$$

must also be evaluated.

- The impurity effective interaction is constructed as follows:

$$W_{\text{imp}}^{\alpha\beta} = \mathcal{U}_{\alpha\beta} - \sum_{L_1 \dots L_4} \sum_{\gamma\delta} \mathcal{U}_{\alpha\gamma} O_{L_1 L_2}^\gamma \chi_{L_1 L_2 L_3 L_4} [O_{L_3 L_4}^\delta]^* \mathcal{U}_{\delta\beta} \quad (1.27)$$

Here all quantities are evaluated at the same frequency² and $O_{L_1 L_2}^\gamma$ is the overlap matrix $\langle B_\gamma | \phi_L \phi_{L'} \rangle$. The polarization operator of the impurity problem is then obtained as:

$$P_{\text{imp}} \equiv -2\delta\Psi_{\text{imp}}/\delta W_{\text{imp}} = \mathcal{U}^{-1} - W_{\text{imp}}^{-1}, \quad (1.28)$$

where the matrix inversions are performed in the OPB $\{B_\alpha\}$.

² Note that $\chi_{L_1 \dots L_4}$ does *not* denote the matrix element $\langle L_1 L_2 | \chi | L_3 L_4 \rangle$, but is rather defined by $\chi(r, r') = \sum_{L_1 \dots L_4} \phi_{L_1}^*(r) \phi_{L_2}^*(r) \chi_{L_1 \dots L_4} \phi_{L_3}(r') \phi_{L_4}(r')$.

- From Eqs. (1.22) and (1.23) the full \mathbf{k} -dependent Green's function $G(\mathbf{k}, i\omega_n)$ and effective interaction $W(\mathbf{q}, i\nu_n)$ can be constructed. The self-consistency condition is obtained, as in the usual DMFT context, by requiring that the on-site components of these quantities coincide with G_{imp} and W_{imp} . In practice, this is done by computing the on-site quantities

$$G_{\text{loc}}(i\omega_n) = \sum_{\mathbf{k}} [G_{\text{H}}^{-1}(\mathbf{k}, i\omega_n) - \Sigma(\mathbf{k}, i\omega_n)]^{-1} \quad (1.29)$$

$$W_{\text{loc}}(i\nu_n) = \sum_{\mathbf{q}} [V_{\mathbf{q}}^{-1} - P(\mathbf{q}, i\nu_n)]^{-1} \quad (1.30)$$

and using them to update the Weiss dynamical mean field \mathcal{G} and the impurity model interaction \mathcal{U} according to:

$$\mathcal{G}^{-1} = G_{\text{loc}}^{-1} + \Sigma_{\text{imp}} \quad (1.31)$$

$$\mathcal{U}^{-1} = W_{\text{loc}}^{-1} + P_{\text{imp}} \quad (1.32)$$

This cycle is iterated until self-consistency for \mathcal{G} and \mathcal{U} is obtained (as well as on G , W , Σ^{xc} and P). When self-consistency is reached, $G_{\text{imp}} = G_{\text{loc}}$ and $W_{\text{imp}} = W_{\text{loc}}$. This implies that at self-consistency, the second term in Eq. (1.22) can be rewritten as (in imaginary-time)

$$\sum_{\mathbf{k}} \Sigma_{GW}^{LL'}(\mathbf{k}, \tau) = - \sum_{L_1 L_1'} W_{\text{imp}}^{LL_1 L' L_1'}(\tau) G_{\text{imp}}^{L_1 L_1}(\tau) \quad (1.33)$$

This shows that the local or on-site contribution of the GW self-energy is precisely subtracted out, thus avoiding double counting. Eventually, self-consistency over the local electronic density can also be implemented, (in a similar way as in LDA+DMFT [35, 36]) by recalculating $\rho(\mathbf{r})$ from the Green's function at the end of the convergence cycle above, and constructing an updated Hartree potential. This new density is used as an input of a new GW calculation, and convergence over this external loop must be reached. While implementing self-consistency within the GWA is known to yield unsatisfactory spectra [37], we expect a more favorable situation in the proposed GW+DMFT scheme since part of the interaction effects are treated to all orders.

1.4.1 Simplified Implementation of GW+DMFT and Application to Ferromagnetic Nickel

The full implementation of the proposed approach in a fully dynamical and self-consistent manner is at the present stage computationally very demanding and we regard it as a major challenge for future research. Here, we apply a simplified scheme of the approach [23] to the electronic structure of nickel in order to demonstrate its feasibility and potential. The main simplifications made are:

1. The DMFT local treatment is applied only to the d-orbitals, and we replace the dynamical impurity problem by its static limit, solving the impurity model (1.19) for a frequency-independent $\mathcal{U} = \mathcal{U}(\omega = 0)$.

2. We perform a one-iteration GW calculation in the form [10]: $\Sigma_{GW} = G_{LDA} \cdot W[G_{LDA}]$, from which the off-site part of the self-energy is obtained.

The local Green's function is taken to be

$$G_{loc}^{\sigma}(i\omega_n) = \sum_{\mathbf{k}} \left[G_H^{-1}(\mathbf{k}, i\omega_n) - \Sigma_{GW}^{off-site} - \left(\Sigma_{imp,\sigma} - \frac{1}{2} \text{Tr}_{\sigma} \Sigma_{imp,\sigma}(0) + V_{xc}^{on-site} \right) \right]^{-1} \quad (1.34)$$

Thus, the off-site part is obtained from the GW self-energy whereas the on-site part is derived from the impurity self-energy with a double-counting correction of the form proposed in [38].

We have performed finite temperature GW and LDA+DMFT calculations (within the LMTO-ASA [39] with 29 irreducible \mathbf{k} -points) for ferromagnetic nickel (lattice constant 6.654 a.u.), using 4s4p3d4f states, at the Matsubara frequencies $i\omega_n$ corresponding to $T = 630$ K, just below the Curie temperature. The GW self-energy is calculated from a paramagnetic Green's function, leaving the spin-dependence to the impurity self-energy. The resulting self-energies are inserted into Eq. (1.34), which is then used to calculate a new Weiss field according to Eq. (1.31). The Green's function $G_{loc}^{\sigma}(\tau)$ is recalculated from the impurity effective action by QMC and analytically continued using the maximum entropy algorithm. The resulting spectral function is plotted in Figure (1.2). Comparison with the LDA+DMFT results in [38] shows that the good description of the satellite structure, exchange splitting and band narrowing is indeed retained within the (simplified) GW+DMFT scheme. We have also calculated the quasiparticle band structure, from the poles of Eq. (1.34), after linearization of $\Sigma(\mathbf{k}, i\omega_n)$ around the Fermi level³. Figure (1.3) shows a comparison of GW+DMFT with the LDA and experimental band structure. It is seen that GW+DMFT correctly yields the bandwidth reduction compared to the (too large) LDA value and renormalizes the bands in a (\mathbf{k} -dependent) manner.

Because of the static approximation 3), we could not implement self-consistency on W_{loc} (Eq. (1.30)). We chose the value of $\mathcal{U}(\omega = 0)$ ($\simeq 3.2$ eV) by calculating the correlation function χ and ensuring that Eq. (1.27) is fulfilled at $\omega = 0$, given the GW value for $W_{loc}(\omega = 0)$ ($\simeq 2.2$ eV for nickel [40]).

1.5 Conclusions

The proposed GW+DMFT scheme avoids the conceptual problems inherent to ‘‘LDA+DMFT’’ methods, such as double counting corrections and the use of Hubbard parameters assigned to correlated orbitals. The notion of a self-consistency condition on the on-site Green's function in the DMFT is extended to the screened interaction. Analogous to the usual condition that the impurity Green's function be equal to the on-site Green's function, we demand that the on-site screened Hubbard \mathcal{U} of the impurity be equal to the on-site projection of the global screened interaction W . In this fashion, the Hubbard \mathcal{U} is determined from first-principles. Since the

³ Note however that this linearization is no longer meaningful at energies far away from the Fermi level. We therefore use the unrenormalized value for the quasi-particle residue for the s-band ($Z_s = 1$).

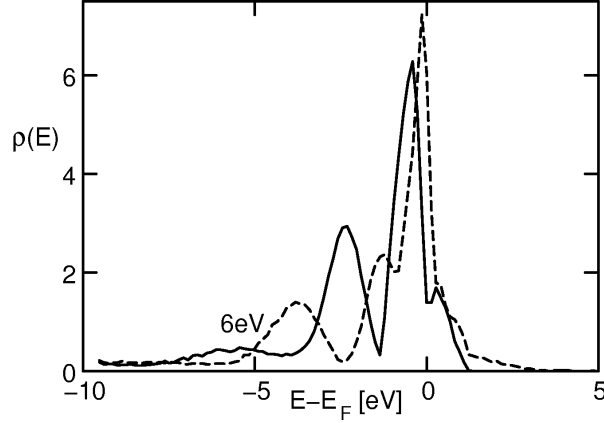


Figure 1.2: Partial density of states of 3d orbitals of nickel (solid/dashed lines give the majority/minority spin contribution) as obtained from the combination of GW and DMFT [23]. For comparison with LDA and LDA+DMFT results, see Ref. [38]; for experimental spectra, see Ref. [42].

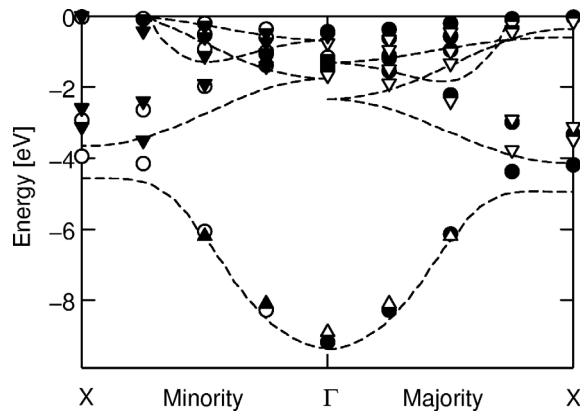


Figure 1.3: band structure of nickel (majority and minority spins) from GW+DMFT scheme [23] (circles) in comparison with the LDA band structure (dashed lines) and experiments [41] (triangles down) and [42] (triangles up).

GWA has a well-defined diagrammatic interpretation, it is also possible to precisely take into account the double-counting correction.

A number of issues are of immediate importance. Solving impurity models with frequency-dependent interaction parameters [24, 43, 44] is one of the most urgent tasks as well as studies of various possible self-consistency schemes. To study these aspects, we are now carrying out GW+DMFT calculations on the electron gas. Applications to real materials are both theoretically and computationally very challenging. Extension to the two-particle Green's function is another field for future research [45].

Acknowledgments

This work was supported in part by NAREGI Nanoscience Project, Ministry of Education, Culture, Sports, Science and Technology, Japan and by a grant of supercomputing time at IDRIS Orsay, France (project number 031393).

References

- [1] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).
- [2] W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
- [3] See, e.g., J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996) and references therein.
- [4] H. Yasuhara, S. Yoshinaga and M. Higuchi, Phys. Rev. Lett. **83**, 3250 (1999).
- [5] R. Maezono, M. D. Towler, Y. Lee, and R. J. Needs, Phys. Rev. B **68**, 165103 (2003).
- [6] See, e.g., W. E. Pickett, Rev. Mod. Phys. **62**, 433 (1989).
- [7] A. Fujimori, F. Minami, and S. Sugano, Phys. Rev. B **29**, 5225 (1984).
- [8] G. A. Sawatzky and J. W. Allen, Phys. Rev. Lett. **53**, 2339 (1984).
- [9] L. Hedin, Phys. Rev. A **139**, 796 (1965); L. Hedin and S. Lundqvist, *Solid State Physics* vol. 23, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1969).
- [10] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
- [11] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (2000).
- [12] W. Ku, A. G. Eguiluz, and E. W. Plummer, Phys. Rev. Lett. **85**, 2410 (2000); H. Yasuhara, S. Yoshinaga, and M. Higuchi, Phys. Rev. Lett. **85**, 2411 (2000).
- [13] F. Aryasetiawan, Phys. Rev. B **46**, 13051 (1992).
- [14] F. Aryasetiawan and O. Gunnarsson, Phys. Rev. Lett. **74**, 3221 (1995).
- [15] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, unpublished results.
- [16] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
- [17] V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyzyk, and G. A. Sawatzky, Phys. Rev. B **48**, 16929 (1993).
- [18] A. I. Lichtenstein, J. Zaanen, and V. I. Anisimov, Phys. Rev. B **52**, R5467 (1995).
- [19] For reviews, see V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys.: Condens. Matter **9**, 767 (1997).
- [20] For reviews, see A. Georges, G. Kotliar, W. Krauth, and M. J. Rosenberg, Rev. Mod. Phys. **68**, 13 (1996); T. Pruschke, M. Jarrel, and J. K. Freericks, Adv. Phys. **44**, 187 (1995).
- [21] For reviews, see *Strong Coulomb Correlations in Electronic Structure Calculations*, edited by V. I. Anisimov, Advances in Condensed Material Science (Gordon and Breach, New York, 2001).
- [22] A. I. Lichtenstein and M. I. Katsnelson, Phys. Rev. B **57**, 6884 (1998).
- [23] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003).
- [24] P. Sun and G. Kotliar, Phys. Rev. B **66**, 085120 (2002).

- [25] For related ideas, see G. Kotliar and S. Savrasov, in *New Theoretical Approaches to Strongly Correlated Systems*, edited by A. M. Tsvelik (Kluwer Dordrecht, 2001) (and the updated version: cond-mat/0208241).
- [26] L. Hedin, *Int. J. Quantum Chem.* **54**, 445 (1995).
- [27] T. Kotani and M. van Schilfgaarde, *Solid State Commun.* **121**, 461 (2002).
- [28] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A. I. Lichtenstein, cond-mat/0401620
- [29] Q. Si and J. L. Smith, *Phys. Rev. Lett.* **77**, 3391 (1996).
- [30] A. M. Sengupta and A. Georges, *Phys. Rev. B* **52**, 10295 (1995).
- [31] G. Kotliar and H. Kajueter, unpublished results.
- [32] H. Kajueter, Ph.D. Thesis, Rutgers University, 1996.
- [33] C.-O. Almbladh, U. von Barth and R. van Leeuwen, *Int. J. Mod. Phys. B* **13**, 535 (1999).
- [34] R. Chitra and G. Kotliar, *Phys. Rev. B* **63**, 115110 (2001).
- [35] S. Savrasov and G. Kotliar, cond-mat/0106308.
- [36] S. Savrasov, G. Kotliar and E. Abrahams, *Nature (London)* **410**, 793 (2000).
- [37] B. Holm and U. von Barth, *Phys. Rev. B* **57**, 2108 (1998).
- [38] A. I. Lichtenstein, M. I. Katsnelson and G. Kotliar, *Phys. Rev. Lett.* **87**, 067205 (2001).
- [39] O. K. Andersen, *Phys. Rev. B* **12**, 3060 (1975); O. K. Andersen, T. Saha-Dasgupta, S. Erzhov, *Bull. Mater. Sci.* **26**, 19 (2003).
- [40] M. Springer and F. Aryasetiawan, *Phys. Rev. B* **57**, 4364 (1998).
- [41] J. Bünemann, F. Gebhard, T. Ohm, R. Umstatter, S. Weiser, W. Weber, R. Claessen, D. Ehm, A. Harasawa, A. Kakizaki, A. Kimura, G. Nicolay, S. Shin, V. N. Strocov, *Europhys. Lett.* **61**, 667 (2003).
- [42] H. Mårtensson and P. O. Nilsson, *Phys. Rev. B* **30**, 3047 (1984).
- [43] Y. Motome and G. Kotliar, *Phys. Rev. B* **62**, 12800 (2000).
- [44] J. K. Freericks, M. Jarrell and D. J. Scalapino, *Phys. Rev. B* **48**, 6302 (1993).
- [45] G. Onida, L. Reining and A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002).

2 A Many-body Approach to the Electronic and Optical Properties of Copper and Silver

Andrea Marini

2.1 Introduction

Quasiparticles, plasmons and excitons are the fundamental quantities used to interpret the electronic and optical properties of solids. More important than their isolated description is the comprehension of their mutual interaction. However, due to the high complexity and large computational requirements of many-body calculations the experimental band-structures, optical absorption and electron-energy-loss spectra are often compared with the results of (simpler) calculations performed within density functional theory (DFT). The consequences of this approach must however be considered with great care. DFT is based on the idea that the ground state spatial density of a system of interacting particles can be exactly described by a non-interacting gas of fictitious independent particles, moving under the action of an effective potential. Thus DFT, as a ground state theory, makes the comparison with experiments inadequate, in particular for the description of excited states properties. An alternative approach is time-dependent DFT (TDDFT) where all neutral excitations are, in principle, exactly described [1]. However only very recently an efficient approximation for the exchange-correlation kernel of TDDFT has been proposed [2], mainly focused on semiconductors and insulators. The shortcomings of DFT are particularly evident in noble metals where the electronic and optical properties are only qualitatively described: (a) compared with experiment, the usual local density (LDA) approximation to the exchange-correlation potential yields an overestimated d-band width and a too small binding energy for the d-bands top; (b) the experimental electron–energy loss spectrum of silver is dominated by a sharp plasmon peak at 3.83 eV, underestimated in position and almost completely damped when calculated within the DFT-LDA random-phase approximation (RPA); (c) the experimental optical spectra intensity of copper and silver are overall overestimated by $\sim 30\%$ when calculated within RPA.

In the following sections I will show how these deficiencies of DFT can be successfully corrected using many-body perturbation theory (MBPT). In Section 2.2 I will present the quasiparticle band structure of copper. The effect of the quasiparticle corrections on the optical absorption and EELS of silver will be described in Section 2.3 while in Section 2.4 I will review the most recent results concerning excitonic effects on the optical spectra of metals.

2.2 Quasiparticle Electronic Structure of Copper

The electronic properties of solids are routinely calculated within DFT in the LDA, by expanding the Kohn–Sham orbitals (KS) in plane waves, see e.g. Ref. [3]. This is made possible by the use of modern norm–conserving pseudopotentials, which allow one to obtain highly accurate valence and conduction band energies without explicitly including the core electrons in the calculation. “Freezing” the core electrons is crucial when a plane-wave basis is used: the number of basis functions needed to describe the 1s of the Si atom is, in fact, 1000 times larger than in the case of the valence shell. The study of noble metals like copper using first-principles methods based on plane-waves and ab-initio pseudopotentials (PPs) presents some peculiar complication with respect to the case of simple metals or semiconductors. In fact, in addition to metalicity, which implies the use of an accurate sampling of the Brillouin zone in order to describe properly the Fermi surface, one must also take into account the contribution of d-electrons to the bonding and to the valence bandstructure. This means that, within the PP scheme, d states cannot be frozen into the core part, but must be explicitly included in the valence, yielding a large total number of valence electrons (11 for bulk copper) [4]. Using soft Martins–Troullier [5] PPs it is possible to work at full convergence with a reasonable kinetic energy cutoff (60 Ry if the 3s and 3p atomic states are frozen into the core, 160 Ry when they are explicitly included) [4]. In Figure 2.1 the DFT band-structure (dashed line) of bulk copper is compared with the experimental data (circles). In contrast to the case of semiconductors, the disagreement between theory and experiment is far from being limited to a rigid shift of the Kohn–Sham occupied eigenvalues with respect to the empty ones. In particular, the comparison clearly shows substantial differences with respect to the experiment for both the d-band width (3.70 instead of 3.17 eV) and position (more than 0.5 eV up-shifted in the DFT), in agreement with previous results [6, 7]. The reason for these important deviations of the DFT band-structure is the approximate inclusion of exchange and correlation effects in the LDA single-particle Kohn–Sham (KS) potential. Moreover, in DFT, KS eigenvalues cannot be identified with electron addition or removal energies, since there is no equivalent of Koopman’s theorem. MBPT represents an exact method to correct the DFT single particle levels, as the band energies can be obtained in a rigorous way, i.e. as the poles of the one-particle Green’s function $G(\mathbf{r}, \mathbf{r}'; \omega)$ [8]. The latter are determined by an equation of the form:

$$\left[-\frac{\hbar^2}{2m} \Delta_{\mathbf{r}} + V_{\text{external}}(\mathbf{r}) + V_{\text{Hartree}}(\mathbf{r}) \right] \psi_{n\mathbf{k}}(\mathbf{r}, \omega) + \int d\mathbf{r}' \Sigma(\mathbf{r}, \mathbf{r}'; \omega) \psi_{n\mathbf{k}}(\mathbf{r}', \omega) = E_{n\mathbf{k}}(\omega) \psi_{n\mathbf{k}}(\mathbf{r}, \omega), \quad (2.1)$$

containing the non-local, non-hermitian and frequency dependent self-energy operator Σ . The poles of G are the QP energies $\epsilon_{n\mathbf{k}}^{\text{QP}}$, the solutions of $\epsilon_{n\mathbf{k}}^{\text{QP}} = E_{n\mathbf{k}}(\epsilon_{n\mathbf{k}}^{\text{QP}})$. The corresponding quasiparticle wavefunctions are $\phi_{n\mathbf{k}}(\mathbf{r}) = \psi_{n\mathbf{k}}(\mathbf{r}, \epsilon_{n\mathbf{k}}^{\text{QP}})$. The off-diagonal matrix elements of the self-energy operator $\langle n\mathbf{k} | \Sigma(\mathbf{r}_1, \mathbf{r}_2, \omega) | n'\mathbf{k}' \rangle$ are usually much smaller than the diagonal elements. Thus the quasiparticle and the KS wavefunctions can be assumed identical and

Eq. (2.1) can be rewritten as a scalar equation for $\epsilon_{n\mathbf{k}}^{\text{QP}}$,

$$\epsilon_{n\mathbf{k}}^{\text{QP}} = \epsilon_{n\mathbf{k}}^{\text{DFT}} + M_{n\mathbf{k}} \left(\epsilon_{n\mathbf{k}}^{\text{QP}} \right) + \Sigma_x^{n\mathbf{k}} - V_{xc}^{n\mathbf{k}}. \quad (2.2)$$

Following Ref. [9], we have separated the static, bare-exchange part $\Sigma_x(\mathbf{r}, \mathbf{r}')$ from $M(\mathbf{r}, \mathbf{r}'; \omega)$, the energy-dependent correlation contribution, or mass-operator. $\Sigma_x^{n\mathbf{k}}$ is given by:

$$\Sigma_x^{n\mathbf{k}} = - \sum_{n_1} \int_{\text{BZ}} \frac{d\mathbf{q}}{(2\pi)^3} f_{n_1(\mathbf{k}-\mathbf{q})} \iint d\mathbf{r} d\mathbf{r}' \phi_{n\mathbf{k}}^*(\mathbf{r}) \phi_{n_1(\mathbf{k}-\mathbf{q})}(\mathbf{r}) v(\mathbf{r}, \mathbf{r}') \phi_{n_1(\mathbf{k}-\mathbf{q})}^*(\mathbf{r}') \phi_{n\mathbf{k}}(\mathbf{r}'), \quad (2.3)$$

where $v(\mathbf{r}, \mathbf{r}')$ is the bare Coulomb interaction and $0 \leq f_n(\mathbf{k}) \leq 1$ represents the occupation number. M is usually evaluated according to the so-called GW approximation, derived by Hedin in 1965 [8, 10], which is based on an expansion in terms of the dynamically screened Coulomb interaction $W(\mathbf{r}, \mathbf{r}'; \omega)$:

$$\begin{aligned} M_{n\mathbf{k}}(\omega) &= \langle n\mathbf{k} | M(\mathbf{r}_1, \mathbf{r}_2, \omega) | n\mathbf{k} \rangle \\ &= - \sum_{n_1} \int_{\text{BZ}} \frac{d\mathbf{q}}{(2\pi)^3} \int_{-\infty}^{\infty} d\omega' \left[\frac{\Gamma_{nn_1}^v(\mathbf{k}, \mathbf{q}, \omega')}{\omega - \omega' - \epsilon_{n_1(\mathbf{k}-\mathbf{q})}^{\text{DFT}} + i\delta} \right. \\ &\quad \left. + \frac{\Gamma_{nn_1}^c(\mathbf{k}, \mathbf{q}, \omega')}{\omega - \omega' - \epsilon_{n_1(\mathbf{k}-\mathbf{q})}^{\text{DFT}} - i\delta} \right]. \end{aligned} \quad (2.4)$$

Γ^c and Γ^v are the conduction, valence contributions to the self-energy spectral function:

$$\Gamma_{nn_1}^c(\mathbf{k}, \mathbf{q}, \omega) = 2(1 - f_{n_1(\mathbf{k}-\mathbf{q})}) \iint d\mathbf{r} d\mathbf{r}' \phi_{n\mathbf{k}}^*(\mathbf{r}) \phi_{n_1(\mathbf{k}-\mathbf{q})}(\mathbf{r}) W^\delta(\mathbf{r}, \mathbf{r}'; \omega) \theta(-\omega) \phi_{n_1(\mathbf{k}-\mathbf{q})}^*(\mathbf{r}') \phi_{n\mathbf{k}}(\mathbf{r}'), \quad (2.5)$$

$$\Gamma_{nn_1}^v(\mathbf{k}, \mathbf{q}, \omega) = 2f_{n_1(\mathbf{k}-\mathbf{q})} \iint d\mathbf{r} d\mathbf{r}' \phi_{n\mathbf{k}}^*(\mathbf{r}) \phi_{n_1(\mathbf{k}-\mathbf{q})}(\mathbf{r}) W^\delta(\mathbf{r}, \mathbf{r}'; \omega) \theta(\omega) \phi_{n_1(\mathbf{k}-\mathbf{q})}^*(\mathbf{r}') \phi_{n\mathbf{k}}(\mathbf{r}'). \quad (2.6)$$

Here W^δ is the delta-like part of the Lehman representation of the screened Coulomb interaction function. W is expressed in terms of the microscopical inverse dielectric function, $W(\mathbf{r}, \mathbf{r}'; \omega) = v(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}'' v(\mathbf{r}, \mathbf{r}'') \epsilon^{-1}(\mathbf{r}'', \mathbf{r}'; \omega)$.

Most GW calculations on semiconductor systems use a plasmon-pole approximation (PPA) for $W(\omega)$ [11], based on the observation that the Fourier components of the inverse dielectric function are generally peaked functions of ω , and can be approximated by a single pole. Since the evaluation of M involves an integration over the energy, the fine details of the ω -dependence are not critical, and the PPA turns out to work reasonably well for most applications. However, in the case of copper, the use of a PPA becomes more critical. The presence

of flat d-bands 2 eV below the Fermi level implies strong transitions in the inverse dielectric function spread over a large energy range. These transitions are not at all well described as a single-pole function, leading to instabilities when determining the plasmon-pole parameters. Instead, the screened electron-hole interaction must be explicitly computed over a grid of about 200 frequencies from zero to ~ 130 eV, and the energy integral performed numerically.

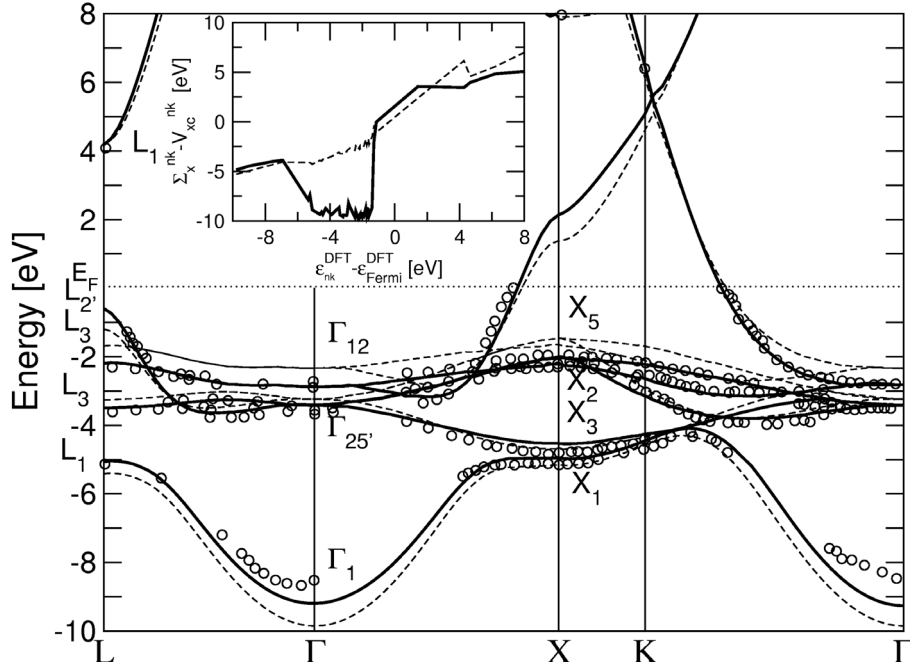


Figure 2.1: Full line: quasiparticle GW results for the bulk copper bandstructure [12], compared with the DFT-LDA results (dashed line), and with the experimental data reported in Ref. [7] (circles). Inset: values of $\Sigma_x^{nk} - V_{xc}^{nk}$, plotted as a function of the non-interacting energies $\epsilon_{nk}^{\text{DFT}}$. The dashed line represents the results obtained without the contributions from the 3s and 3p core states.

The set of Eqs. (2.1)–(2.6) constitutes a well-defined and successful scheme to calculate quasiparticle band structure in many different materials [8]. Normally (e.g., in GW calculations for semiconductors), the calculation of G and W to correct the DFT valence band structure can be performed by including only valence states, and fully neglecting the core states which have been frozen in the pseudopotential approach. Among transition metals, full quasiparticle calculations have been carried out only for Ni [13]. In the case of Ni, GW yields a good description of photoemission data, except for the 6 eV satellite, which is due to strong short-range correlations within the partially filled d-shell. For copper, we would expect band-theory to work better than for transition metals, since d-shells are completely filled. Instead, as showed in Ref. [12], when Σ is computed neglecting the 3s and 3p atomic core states (which in the solid create two flat bands, at about 112 and 70 eV, respectively, below

the Fermi level), the resulting QP corrections on the d-bands are clearly unphysical: GW corrections move the highest occupied d-bands above the DFT-LDA Fermi level. On the other hand, the situation for s/p states (e.g. for the state $L_{2'}$) is much more reasonable, with correlation and exchange parts of the self-energy which largely cancel each other (as in the case of semiconductors), and negative QP corrections of the order of eV. The solution of this puzzling situation is provided by the role of the above-mentioned 3s and 3p states, which, despite being well separated *in energy* from the 3d ones, have a large *spatial* overlap with the latter. As a consequence, non-negligible contributions to the self-energy are expected from the *exchange* contributions between 3d and 3s/3p states as clearly shown in the inset of Figure 2.1, where the difference $\Sigma_x^{nk} - V_{xc-LDA}^{nk}$ is strongly affected by the presence of core levels in the n_1 summation of Eq. (2.4). The role of core levels in the calculation of the bare exchange contributions (whose importance has already been addressed for transition metals by Aryasetiawan and Gunnarsson [8], but estimated to be of the order of 1 eV) is hence crucial in the case of copper. Moreover their effect is unexpected on the basis of DFT-LDA calculations, where the band structure does not change appreciably even if the 3s/3p orbitals are fully included in the valence [4].

In Figure 2.1 the full GW theoretical band structure [12] is compared with experimental data: the agreement is remarkably good and the fact that the GW corrections cannot be reproduced by any rigid shift of the LDA bands clearly appears. Hence, the GW method, originally devised to describe the *long-range* charge oscillations [10], is hence shown to yield a good description also of copper, a system characterized by localized orbitals and short-range correlation effects.

2.3 The Plasmon Resonance of Silver

In the previous section we have seen how to correct the DFT single particle levels using MBPT obtaining an excellent agreement with the experimental results. However in a fully interacting electronic system quasiparticles coexist with collective excitations, i.e. plasmons. Plasmons occur at energies for which the real part of the dielectric function vanishes with a corresponding small imaginary part; they can be observed experimentally as sharp peaks in electron energy loss spectra (EELS). A well established technique to calculate EELS uses the independent particles or random phase approximation (RPA) for the polarization function, obtained in terms of ab-initio single-particle energy bands. In the case of silicon, Olevano and Reining [14] showed that using the quasiparticle energies without going beyond the RPA, including excitonic effects (as we will discuss in the next section), the shape of the plasmon peak worsens with respect to the experiment. Inclusion of self-energy corrections and excitonic effects yields a spectrum very similar to the DFT-LDA one and to the experiment. Ku and Eguiluz [15] obtained a correct positive dispersion of the plasmon width in K using the single particle approximation, with no many-body corrections beyond DFT-LDA. In these cases many-body effects (quasiparticle corrections and/or excitonic effects) are not required to describe correctly the experimental data. These results agree with the general feeling that excitonic effects partially cancel self-energy corrections (we will discuss in detail this cancellation in the next section). A similar result has been found for copper [4, 16], where the RPA

response function calculated without many-body corrections yields good agreement with the experimental EEL and optical spectra.

In this framework the case of silver is rather surprising: the experimental EELS (circles in Figure 2.2) is dominated by a sharp plasmon peak at 3.83 eV [17], whose position and width are badly reproduced in DFT-LDA RPA [18] (dashed line in Figure 2.2). In particular, a width of about 0.5 eV is obtained within this approach, to be contrasted with a much narrower experimental width (~ 100 meV). A similar discrepancy occurs in the reflectance spectrum (see Figure 2.3), where a very narrow dip at 3.92 eV is hardly reproduced by DFT-LDA calculations.

Following the same GW scheme used for copper in the previous section the quasiparticle band structure of silver [19] at high symmetry points is compared with DFT-LDA results in Table 2.1. While the deeper energy levels remain mostly unchanged, a downward shift of about 1.3 eV of the top d bands leads to a decrease of the bandwidth, and hence to an excellent agreement with experiment.

Table 2.1: Theoretical band widths and band energies for silver [19], at high-symmetry points. GW energies are relative to the QP Fermi Level. The striking agreement with the experimental results shows that the silver band-structure is very well described at the GW level. The values in the last column are taken from Ref. [20] where spin-orbit splittings have been removed by making degeneracy-weighted averages.

		DFT-LDA	GW	Experiment
Positions of d-bands	Γ_{12}	-3.57	-4.81	-4.95
	X_5	-2.49	-3.72	-3.97
	$L_3(2)$	-2.71	-3.94	-4.15
Widths of d-bands	$\Gamma_{12} - \Gamma_{25'}$	1.09	0.94	1.11
	$X_5 - X_3$	3.74	3.39	3.35
	$X_5 - X_1$	3.89	3.51	3.40
	$L_3(2) - L_3(1)$	1.98	1.85	1.99
	$L_3 - L_1$	3.64	3.17	2.94
	$X_5 - X_2$	0.27	0.29	0.38

To calculate the EEL spectra, the most simple expression for the dielectric function is obtained within RPA, where the electron and holes excited by the external perturbation are assumed to move independently. The EELS is given by the imaginary part of the inverse dielectric function $\epsilon^{-1}(\omega)$:

$$\epsilon^{-1}(\omega) = \left[\epsilon_{\text{ib}}(\omega) - \frac{\omega_{\text{D}}^2}{\omega(\omega + i\eta)} \right]^{-1}, \quad (2.7)$$

where $\epsilon_{\text{ib}}(\omega)$ is the interband contribution and $\omega_{\text{D}} = 9.48$ eV is the Drude plasma frequency, both calculated ab-initio following the procedure described in Ref. [4]. The interband RPA

dielectric function is given by

$$\epsilon_{\text{ib}}(\omega) = 1 - 4\pi \lim_{\mathbf{q} \rightarrow \mathbf{0}} \int_{\text{BZ}} \frac{d^3 \mathbf{k}}{(2\pi)^3} \sum_{n \neq n'} \frac{|\langle n' \mathbf{k} - \mathbf{q} | e^{-i\mathbf{q} \cdot \mathbf{r}} | n \mathbf{k} \rangle|^2}{|\mathbf{q}|^2} \frac{f_{n', \mathbf{k} - \mathbf{q}} - f_{n, \mathbf{k}}}{\omega + E_{n, \mathbf{k}} - E_{n', \mathbf{k} - \mathbf{q}} + i\eta}, \quad (2.8)$$

where $\langle n' \mathbf{k} - \mathbf{q} | e^{-i\mathbf{q} \cdot \mathbf{r}} | n \mathbf{k} \rangle = \int d\mathbf{r} e^{-i\mathbf{q} \cdot \mathbf{r}} \phi_{n', \mathbf{k} - \mathbf{q}}^*(\mathbf{r}) \phi_{n, \mathbf{k}}(\mathbf{r})$. The $\mathbf{q} \rightarrow \mathbf{0}$ limit of Eq. (2.8) has been done including the effects of the pseudopotential non-locality, as described in Ref. [4]. Using as single-particle energies $E_{n\mathbf{k}}$ the KS levels we obtain the DFT-LDA RPA dielectric function, showed in the inset of Figure 2.2 (dashed line). Because of the underestimation of the d-bands top position (see Table 2.1) the interband onset is too low compared to the experiment (circles) [21]. Therefore the corresponding plasmon peak is strongly damped and its energy underestimated (dashed line in the main frame of Figure 2.2). When the quasiparticle single-particle energies are used in Eq. (2.8) instead of the DFT-LDA ones, the plasmon peak, underestimated in intensity and position in DFT-LDA, is shifted toward higher energies and strongly enhanced by *GW* corrections, in striking agreement with experiment. This resonance can be interpreted as a collective (Drude-like) motion of electrons in the partially filled band. However, its energy ω_p does not coincide with the bare Drude frequency ω_D , the difference arising from the screening of the electron–electron interaction by virtual interband transitions. The plasmon resonance, although blue shifted with respect to DFT-LDA, remains *below* the main interband threshold, but overlaps the weak low-energy tail of interband transitions acquiring a small, yet finite, width.

The polarization of the “medium” where the plasmon oscillates (the d electrons) is hence important to determine its energy and width. This polarization is absent in the homogeneous electron gas because there are no localized d orbitals and no interband transitions; it is weak in semiconductors (like Si), because interband transitions occur at energies far from that of the plasma resonance. The same polarization effect is present, but destructive in copper due to the lower onset of interband transitions. EELS peaks occur *above* this onset and are therefore strongly broadened. Consequently, the delicate interplay of plasmon-frequency renormalization with the shift of the interband-transition onset, both due to QP corrections, may yield (in silver) or may not yield (in copper) a sharp plasmon resonance.

Another important quantity is the reflectance, $R(\omega) = (|N(\omega) - 1| / |N(\omega) + 1|)^2$, where N is the complex refraction index defined by $[N(\omega)]^2 = \epsilon(\omega)$. In Figure 2.3 we compare the *GW* $R(\omega)$ with the DFT-LDA one, and with experimental results [22]. The latter shows a very narrow dip at 3.92 eV, close to the plasmon frequency, arising from the zero-reflectance point ω_0 , defined as $\epsilon(\omega_0) = 1$. Again, the width and depth of this reflectance dip are related to the imaginary part of $\epsilon(\omega)$. *GW* corrections make ω_0 occur below the main onset of interband transitions, and hence produce a very narrow and deep reflectance minimum. Here the agreement between *GW* results and experiments for the intensity and width of the dip at 3.92 eV is even more striking than in the EELS.

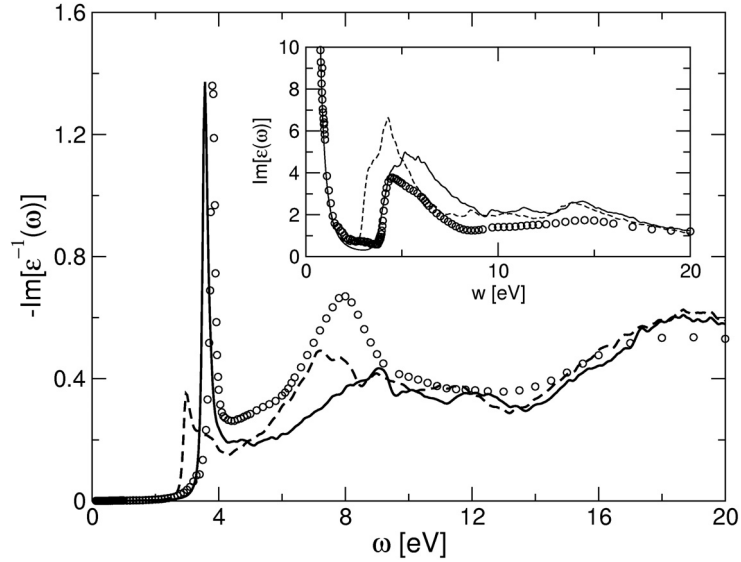


Figure 2.2: Electron energy loss spectrum (EELS) of silver from Ref. [19]. Solid line: GW . Dashed line: DFT-LDA. Circles: experiment [22]. The non trivial quasiparticle GW corrections improve considerably the DFT-LDA plasmon peak, yielding a striking agreement with the experiment. Inset: optical absorption of silver within RPA. Solid line: GW energy levels are used in Eq. (2.8). Dashed line: DFT-LDA energy levels are used. Circles: experiment [22].

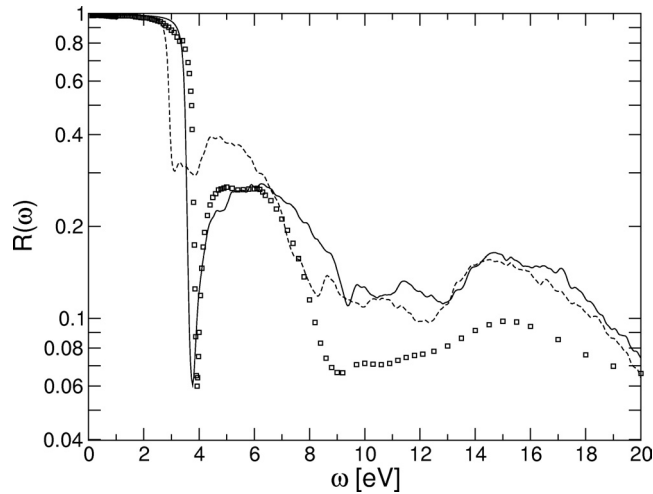


Figure 2.3: Reflectivity spectrum of silver from Ref. [19]. Solid line: GW . Dashed line: DFT-LDA. Boxes: experiment [22]. The experimental sharp dip at 3.92 eV is correctly reproduced by GW , with a substantial improvement on the DFT-LDA spectrum.

2.4 Dynamical Excitonic Effects in Metals

In the inset of Figure 2.2 we observe a strong overestimation of the absorption spectrum intensity with respect to the experiment (similar to that found in copper [4]). However this residual discrepancy cannot be traced back to the standard phenomenology observed in semiconductors and insulators where the observed light-absorption spectra largely deviate from independent-particle, RPA calculations [1]. These deviations are corrected by including the electron–hole interaction [23]. The strength of these modifications increases as the inverse of the dielectric constant of the system. In insulators, where the electron–hole interaction is only weakly screened, sharp peaks with energy below the optical gap (bound excitons) can be observed in the experimental spectra [24]. If this description is extrapolated to the metallic case the natural conclusion is that the electron–hole interaction has a negligible effect on the optical spectra of metals as the static electron–hole interaction is completely screened by the long–range part of the dielectric function. This simple argument has been considered definitive to assert that there are no excitonic effects in metals. However the standard approach to account for the electron–hole interaction in the optical spectra involves approximations whose validity is assumed a priori, and not sustained by theoretical motivations. These approximations are the key to understanding the apparently inexplicable overestimation observed in the RPA optical spectra of silver and copper.

To introduce the theoretical framework commonly used to calculate optical properties beyond the RPA we have to rewrite the dielectric function as $\epsilon(\omega) \equiv 1 - 8\pi\mathbf{A}^\dagger \mathbf{P}(\omega) \mathbf{A}$, where $\mathbf{P}(\omega)$ is the matrix representation of the polarization function in the non-interacting electron–hole basis and \mathbf{A} is a vector embodying the corresponding optical oscillators $\langle n' \mathbf{k} - \mathbf{q} | e^{-i\mathbf{q} \cdot \mathbf{r}} | n \mathbf{k} \rangle$ already introduced in Eq. (2.8). The polarization function is obtained by solving the Bethe–Salpeter equation (BSE), an integral equation for the four point electron–hole Green’s function $\mathbf{L}(t_1, t_2; t_3, t_4)$ [1, 23, 25]. As we are interested in the polarization function $\mathbf{P}(t) \equiv -i\mathbf{L}(t, 0; t, 0)$, the BSE can be rewritten as:

$$\begin{aligned} \mathbf{P}(t) = & \mathbf{P}^{(0)}(t) - \int dt_1 \mathbf{P}^{(0)}(t - t_1) \mathbf{V} \mathbf{P}(t_1) \\ & + \iint dt_1 dt_2 \mathbf{L}^{(0)}(t, t_2; t, t_1) \widetilde{\mathbf{W}}(t_1 - t_2) \mathbf{L}(t_1, 0; t_2, 0). \end{aligned} \quad (2.9)$$

Equation (2.8) is obtained when no electron–hole effects are included in \mathbf{P} , i.e. $\mathbf{P} \approx \mathbf{P}^{(0)} = -i\mathbf{L}^{(0)}$. Like \mathbf{P} , also \mathbf{V} , \mathbf{L} and $\widetilde{\mathbf{W}}$ are matrix representations in the non-interacting electron–hole basis. Using generalized indexes $\mathbf{K} := (c v \mathbf{k})$; c, v, \mathbf{k} being conduction, valence band and k-point indexes, we can express those quantities as:

$$\widetilde{W}_{\mathbf{K}\mathbf{K}'}(t_1 - t_2) = i \langle c \mathbf{k}, v \mathbf{k} | W(\mathbf{r}, \mathbf{r}'; t_1 - t_2) - v(\mathbf{r}, \mathbf{r}') | c' \mathbf{k}', v' \mathbf{k}' \rangle, \quad (2.10)$$

$$V_{\mathbf{K}\mathbf{K}'}(t_1 - t_2) = i \iint d\mathbf{r} d\mathbf{r}' \phi_{c\mathbf{k}}^*(\mathbf{r}) \phi_{v\mathbf{k}}(\mathbf{r}) v(\mathbf{r}, \mathbf{r}') \phi_{v'\mathbf{k}'}^*(\mathbf{r}') \phi_{c'\mathbf{k}'}(\mathbf{r}'). \quad (2.10')$$

$L^{(0)}$ is the non-interacting electron–hole Green’s function

$$L_{\mathbf{K},\mathbf{K}'}^{(0)}(t_1, t_2; t_3, t_4) = \delta_{v,v'} \delta_{c,c'} \delta_{\mathbf{k},\mathbf{k}'} Z_{c\mathbf{k}} Z_{v\mathbf{k}} \theta(t_1 - t_4) e^{-iE_{c\mathbf{k}}(t_1 - t_4)} \theta(t_3 - t_2) e^{iE_{v\mathbf{k}}(t_3 - t_2)}. \quad (2.11)$$

$Z_{n\mathbf{k}}$ (smaller than 1) are the QP energies and renormalization factors, respectively. The latter represent the weights of the QP peak in the many-body single-particle spectral function. The more $Z_{n\mathbf{k}}$ differs from 1, the more the high energy structures in the spectral function (like plasmonic replicas) become important. Those high-energy peaks are not visible in the optical energy range but, nevertheless, subtract intensity from the QP peaks. Very little is known about the role played by the Z factors in optical spectra calculations [26]. In Section 2.3 we had to assume $Z = 1$, because if the Z factors are included in an RPA calculation [27], or even in the BSE (see below), the intensity of the resulting spectra is strongly underestimated, both in metals and in semiconductors. Thus the Z factors are commonly set to 1 *by hand* in the solution of the BSE or in the calculation of the independent-QP spectra. This is the first approximation needed to reproduce the experimental results, and it is important to note that it lacks sound theoretical justification.

But there is another important approximation needed to solve Eq. (2.9), related to the time dependent term $\widetilde{W}(t_1 - t_2)$. Indeed, the BSE with a time-dependent interaction is considered hardly solvable (if not “practically unsolvable” [26]) and *for computational convenience* the electron–hole interaction is assumed to be instantaneous; this is equivalent to approximating W with its static value, $W(\mathbf{r}_1, \mathbf{r}_2, \omega = 0)$. This approximation is verified a posteriori through comparison with experiment and physically corresponds to the assumption that the electron–hole scattering time is much longer than the characteristic screening time of the system (roughly speaking, the inverse of the plasma frequency). Indeed, the static approximation is expected to work well for transition energies much smaller than the plasma frequency [23]. However the most striking examples of systems that do not fulfill this condition are silver and copper. From Figure 2.2 it is evident that the plasmon of silver, which dominates the EELS, is located just above the interband gap (~ 3.9 eV). Similarly, the EELS of copper shows strong, broad peaks in the optical range [4]. When $\widetilde{W}(t) \approx \widetilde{W}(\omega = 0) \delta(t)$ [1, 23]. Equation (2.9) can be formally solved by means of a Fourier transform:

$$P(\omega) = P^{(0)}(\omega) - P^{(0)}(\omega) \left(V + \widetilde{W} \right) P(\omega). \quad (2.12)$$

This is the static BSE (SBSE) commonly applied neglecting the renormalization factors in Eq. (2.9), i.e. taking $Z_{n\mathbf{k}} = 1$. It yields optical spectra in good agreement with experiments in semiconductors and insulators [23]. When applied to copper and silver however, the SBSE result (dotted lines in Figure 2.5) is indistinguishable from the independent-QP calculation, without improving the agreement with experiment.

In Ref. [28] a solution of Eq. (2.9) is proposed without the two major approximations employed in the SBSE, i.e. keeping the $Z_{n\mathbf{k}}$ factors smaller than 1 and W frequency dependent. To this end $L^{(0)} \widetilde{W} L$ is expanded in powers of \widetilde{W} . The first order term of this expansion,

$\mathbf{P}^{(1)}(t)$, is given by:

$$P_{\mathbf{K}_1\mathbf{K}_2}^{(1)}(t) = \iint dt_1 dt_2 \theta(t_1 - t_2) \left[L_{\mathbf{K}_1}^{(0)}(t, t_2; t, t_1) \widetilde{W}_{\mathbf{K}_1\mathbf{K}_2}(t_1 - t_2) L_{\mathbf{K}_2}^{(0)}(t_1, 0; t_2, 0) + L_{\mathbf{K}_1}^{(0)}(t, t_1; t, t_2) \widetilde{W}_{\mathbf{K}_1\mathbf{K}_2}(t_2 - t_1) L_{\mathbf{K}_2}^{(0)}(t_2, 0; t_1, 0) \right]. \quad (2.13)$$

From Eq. (2.10) it is straightforward to see that

$$L_{\mathbf{K}_1}^{(0)}(t, t_2; t, t_1) = i \left[P_{\mathbf{K}_1}^{(0)}(t - t_1) e^{iE_{v_1\mathbf{k}_1}(t_1 - t_2)} \theta(t_1 - t_2) + P_{\mathbf{K}_1}^{(0)}(t - t_2) e^{-iE_{c_1\mathbf{k}_1}(t_2 - t_1)} \theta(t_2 - t_1) \right], \quad (2.14)$$

$$L_{\mathbf{K}_2}^{(0)}(t_1, 0; t_2, 0) = i \left[P_{\mathbf{K}_2}^{(0)}(t_2) e^{-iE_{c_2\mathbf{k}_2}(t_1 - t_2)} \theta(t_1 - t_2) + P_{\mathbf{K}_2}^{(0)}(t_1) e^{iE_{v_2\mathbf{k}_2}(t_2 - t_1)} \theta(t_2 - t_1) \right], \quad (2.14')$$

that inserted in Eq. (2.13), casts $\mathbf{P}^{(1)}(t)$ as a time convolution of three terms (as shown diagrammatically in Figure 2.4).

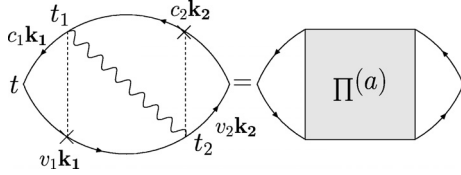


Figure 2.4: Diagrammatic representation of the first order contribution to the polarization function $\mathbf{P}(t)$ according to the BSE. Crosses indicate the time points where the incoming and outgoing non-interacting Green's functions are “cut” according to Eqs. (2.14). The right-hand side represents the first order polarization diagram of the dynamical Bethe–Salpeter as discussed in the text. It can be used to sum all orders of BSE with non overlapping interaction lines.

As a consequence, in the frequency domain $\mathbf{P}^{(1)}(\omega)$ has the form:

$$\mathbf{P}^{(1)}(\omega) = -\mathbf{P}^{(0)}(\omega) \left[\mathbf{\Pi}^{(a)}(\omega) + \mathbf{\Pi}^{(b)}(\omega) \right] \mathbf{P}^{(0)}(\omega), \quad (2.15)$$

with

$$\mathbf{\Pi}_{\mathbf{K}_1\mathbf{K}_2}^{(a)}(\omega) = \widetilde{W}_{\mathbf{K}_1\mathbf{K}_2}^{(+)}(\omega + E_{v_1\mathbf{k}_1} - E_{c_2\mathbf{k}_2}), \quad (2.16)$$

and

$$\mathbf{\Pi}_{\mathbf{K}_1\mathbf{K}_2}^{(b)}(\omega) = \widetilde{W}_{\mathbf{K}_1\mathbf{K}_2}^{(+)}(\omega + E_{v_2\mathbf{k}_2} - E_{c_1\mathbf{k}_1}); \quad (2.16')$$

$\widetilde{W}^{(+)}(\omega)$ being the Laplace transform of $\widetilde{W}(t)$. The two terms denoted by (a) and (b) correspond to the two possible time orderings of the interaction ends ($t_1 > t_2$ for term (a), shown in Figure 2.4; $t_2 > t_1$ for term (b), not shown). Equation (2.15) can be thought of as the first order expansion of $P(\omega)$ in the frequency-dependent interaction $\mathbf{II}(\omega) = \mathbf{II}^{(a)}(\omega) + \mathbf{II}^{(b)}(\omega)$, which replaces \widetilde{W} of the SBSE. Thus a partial summation of the BSE can be performed writing:

$$\mathbf{P}(\omega) = \mathbf{P}^{(0)}(\omega) - \mathbf{P}^{(0)}(\omega) [\mathbf{V} + \mathbf{II}(\omega)] \mathbf{P}(\omega). \quad (2.17)$$

This is the Dynamical Bethe–Salpeter equation (DBSE) [28]. The diagrams summed up in Eq. (2.17) are those containing the ladder series of repeated electron–hole interactions with *non overlapping* (in time) interaction lines. The poles of $\mathbf{P}(\omega)$, Ω_λ , will be given by the solution of the equation $[\mathbf{P}^{(0)}(\Omega_\lambda)]^{-1} + \mathbf{V} + \mathbf{II}(\Omega_\lambda) = 0$. In contrast to the kernel of the SBSE, $\mathbf{II}(\Omega_\lambda)$ is not hermitian and, consequently, Ω_λ is in general complex. Its imaginary part gives the inverse excitonic lifetime. Thus the interacting electron–hole states are actually dressed excitons, or *quasiexcitons*. This agrees with what has been already found in the core exciton limit [29] and emphasizes the analogy between the DBSE and the Dyson equation. Consequently, as in the single-particle problem, we expect to find similar renormalization effects on the quasiexcitonic Green’s function. To develop further this aspect we expand linearly the smooth function $\widetilde{W}^{(+)}(\omega)$ around the non-interacting electron–hole energies, obtaining $\mathbf{II}_{\mathbf{K}_1\mathbf{K}_2}(\omega) \approx \mathbf{II}_{\mathbf{K}_1\mathbf{K}_2}^{(\text{st})} + \Theta_{\mathbf{K}_1\mathbf{K}_2}(\omega - E_{c_2\mathbf{k}_2} + E_{v_2\mathbf{k}_2})$. $\mathbf{II}_{\mathbf{K}_1\mathbf{K}_2}^{(\text{st})} = \mathbf{II}_{\mathbf{K}_1\mathbf{K}_2}(\omega)|_{\omega=E_{c_2\mathbf{k}_2}-E_{v_2\mathbf{k}_2}}$ is the static limit of the dynamical Bethe–Salpeter kernel which turns out to be quite similar to the kernel of the SBSE. $\Theta_{\mathbf{K}_1\mathbf{K}_2} = \partial\mathbf{II}_{\mathbf{K}_1\mathbf{K}_2}(\omega)/\partial\omega|_{\omega=E_{c_2\mathbf{k}_2}-E_{v_2\mathbf{k}_2}}$ are the excitonic dynamical-renormalization factors. Thus Eq. (2.17) can be strongly simplified in the case of copper and silver where the effect of $\mathbf{II}^{(\text{st})} + \mathbf{V}$ is very small. The corresponding polarization function $\mathbf{P}(\omega)$ is approximately given by:

$$P_{\mathbf{K}_1\mathbf{K}_2}(\omega) \approx \frac{[(Z^{\text{eh}})^{-1} + \Theta]_{\mathbf{K}_1\mathbf{K}_2}^{-1}}{\omega - E_{c_2\mathbf{k}_2} + E_{v_2\mathbf{k}_2} + i0^+}, \quad (2.18)$$

with $Z_{\mathbf{K}_1\mathbf{K}_2}^{\text{eh}} = Z_{c_1\mathbf{k}_1} Z_{v_1\mathbf{k}_1} \delta_{\mathbf{K}_1\mathbf{K}_2}$.

The connection between dynamical excitonic and self-energy effects is now clear. $Z_{n\mathbf{k}}^{-1} = 1 - \beta_{n\mathbf{k}}$, where the negative factor $\beta_{n\mathbf{k}}$, the frequency derivative of the self-energy, is the weight lost by the QP because of the coupling with the excitations of $W(\omega)$. The excitonic factors Θ , instead, are due to the modification of such coupling as a consequence of the electron–hole interaction. Those two effects tend to cancel each other but *the cancellation is, in general, not complete*, as exemplified in Figure 2.5 for copper and silver. The SBSE calculation (dotted line), with $Z_{n\mathbf{k}} = 1$ and $\Theta = 0$, overestimates the experimental intensity (circles), while the inclusion of the $Z_{n\mathbf{k}}$ factors only (dashed line) underestimates it. In the DBSE (full line) the dynamical Θ factors partially compensate for the Z^{eh} factors yielding a spectral intensity in good agreement with experiment.

Similarly the optical spectra of semiconductors can be obtained with the DBSE in excellent agreement with experiment, as shown in the case of silicon in Ref. [28]. In contrast to the

metallic case, however, the DBSE kernel of silicon must contain second-order contributions in order to reproduce correctly the experimental optical spectrum. The main effect of the first order kernel $\mathbf{II}(\omega)$ is indeed to balance the reduction of optical strengths due to self-energy renormalization factors, as suggested by Bechstedt et al. [26]. However, the renormalized QP weights also imply a reduction of the statically screened electron–hole of almost $\sim 30\%$, which is the reason for the wrong relative intensities of the two peaks in the SBSE result. This shortcoming is fixed by the second-order diagrams.

Thus in both the metallic and semiconducting cases the DBSE correctly describes the measured optical spectra without the a priori approximations commonly used in solving the BSE. This result can be interpreted by thinking of the electron–hole pair as a neutral excitation, thus, less efficient than the electron and the hole alone in exciting virtual plasmons, which is the main process leading to QP renormalization. Only when dynamical effects are coherently included both in the self energy and in the electron–hole interaction does this (physically expected) result emerges from the bundle of many-body equations. *This confirms the SBSE results but not the separate approximations involved therein.*

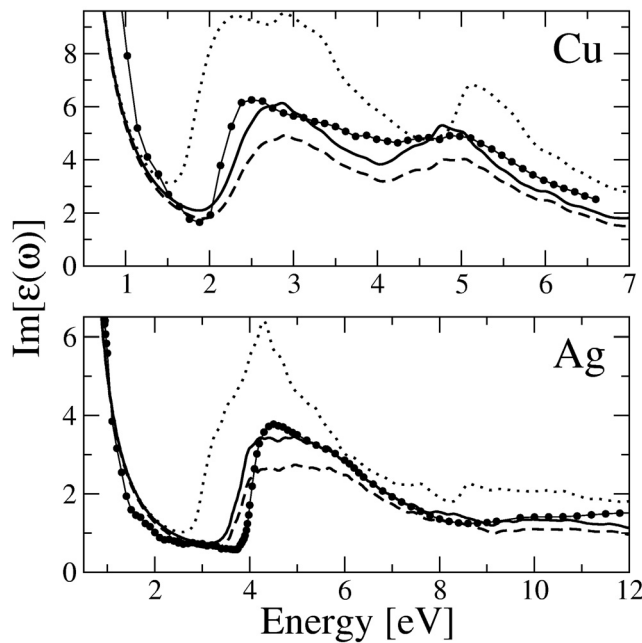


Figure 2.5: Absorption spectrum of bulk copper and silver from Ref. [28]. Dotted line: SBSE without renormalization factors. Dashed line: SBSE including the QP renormalization factors. Full line: result of the dynamical Bethe–Salpeter equation including dynamical QP and excitonic effects. Circles: experimental spectra (Ref. [22]).

2.5 Conclusions

Several conclusions can be drawn from the presented results. Starting from DFT we have reviewed all the steps needed to correct the observed disagreements between the experiment and the calculated band-structure (Section 2.2), electron energy loss spectrum (Section 2.3) and optical spectra (Section 2.4). We have observed in all cases a key role played by the localized d-orbitals. They contribute to the self-energy with strong exchange effects with the 3s/3p core states, balancing the large correlation part of the mass-operator. Thus core levels cannot be neglected in the quasiparticle calculations of noble metals. Then, in contrast to semiconductors and simple metals, the plasma frequency of silver is renormalized by the interband transitions between d and s/p bands when QP single particle energies are used to calculate the RPA polarization function. This clearly shows that the low-energy peaks in the dielectric function due to the d states affect also the excitations occurring at the Fermi surface where the main contribution to the metallic screening arises. Finally, even if the s/p metallic bands completely screen the electron-hole interaction at *zero* energy (static contribution), the localized d-orbitals induce a dynamical term in the Bethe-Salpeter kernel that partially cancel the dynamical self-energy effects (renormalization factors). In conclusion single-particle, two-particle excitations, as well as plasmons in copper and silver, can be successfully described within many-body perturbation theory. The role of the localized, d-orbitals must be correctly included in the calculations and, more importantly, in the theory. These unique features of noble and, in general, d metals make them a stringent test for modern ab-initio Many-Body theories.

References

- [1] For a review, see G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* **74** (2002) 601.
- [2] A. Marini, A. Rubio, and R. Del Sole, *Phys. Rev. Lett.* **91** (2003) 256402; F. Sottile, V. Olevano and L. Reining, *Phys. Rev. Lett.* **91** (2003) 056402.
- [3] O. Pulci et al., *Phys. Rev. B* **55** (1997) 6685.
- [4] A. Marini, G. Onida and R. Del Sole, *Phys. Rev. B* **64** (2001) 195125.
- [5] N. Troullier and J.L. Martins, *Phys. Rev. B* **43** (1991) 1993.
- [6] See, for example, V.N. Strocov et al., *Phys. Rev. Lett.* **81** (1998) 4943; *Phys. Rev. B* **63** (2001) 205108.
- [7] R. Courths and S. Hüfner, *Phys. Rep.* **112** (1984) 53.
- [8] F. Aryasetiawan and O. Gunnarsson, *Rep. Prog. Phys.* **61** (1998) 237-312.
- [9] R. W. Godby, M. Schlüter and L. J. Sham, *Phys. Rev. B* **37** (1988) 10159.
- [10] L. Hedin, *Phys. Rev. A* **139** (1965) A796.
- [11] M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* **34** (1986) 5390; R.W. Godby and R.J. Needs, *Phys. Rev. Lett.* **62** (1989) 1169.
- [12] A. Marini, R. Del Sole, G. Onida, *Phys. Rev. Lett.* **88** (2002) 016403.
- [13] F. Aryasetiawan, *Phys. Rev. B* **46** (1992) 13051.
- [14] V. Olevano and L. Reining, *Phys. Rev. Lett.* **86** (2001) 5962.
- [15] W. Ku and A. G. Eguiluz, *Phys. Rev. Lett.* **82** (1999) 2350.

- [16] I. Campillo, A. Rubio and J.M. Pitarke, Phys. Rev. B **59** (1999) 12188.
- [17] H. Herenreich and H. R. Phillipp, Phys. Rev. **128** (1962) 1622.
- [18] M. A. Cazalilla et al., Phys. Rev. B **61** (2000) 8033.
- [19] A. Marini, R. Del Sole, G. Onida, Phys. Rev. B **66** (2002) 115101.
- [20] G. Fuster et al., Phys. Rev. B **42** (1990) 7322.
- [21] V. P. Zhukov, F. Aryasetiawan, E. V. Chulkov, I. G. de Gurtubay, and P. M. Echenique, Phys. Rev. B **64** (2001) 195122.
- [22] E.D. Palik, *Handbook of Optical Constants of Solids*, Academic Press, New York, 1985, p. 835.
- [23] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, Phys. Rev. Lett. **80** (1998) 4510; L.X. Benedict, E.L. Shirley, and R.B. Bohn, Phys. Rev. Lett. **80** (1998) 4514; M. Rohlfing and S.G. Louie, Phys. Rev. Lett. **81** (1998) 2312.
- [24] See+, for instance, D. M. Roessel and W. C. Walker, J. Opt. Soc. Am. **57** (1967) 835 for the case of LiF.
- [25] G. Strinati, Rivista del nuovo cimento, **11** (1988) 1.
- [26] F. Bechstedt, K. Tenelsen, B. Adolph, and R. Del Sole, Phys. Rev. Lett. **78** (1997) 1528.
- [27] R. Del Sole and R. Girlanda, Phys. Rev. B **54** (1996) 14376.
- [28] A. Marini and R. Del Sole Phys. Rev. Lett. **91** (2003) 176402.
- [29] G. Strinati, Phys. Rev. Lett. **49** (1982) 1519; Phys. Rev. B **29** (1984) 5718.