

10 Linear Response and More: the Bethe-Salpeter Equation

Lucia Reining

Laboratoire des Solides Irradiés and

European Theoretical Spectroscopy Facility

École Polytechnique, CNRS, CEA,

Université Paris-Saclay, 91128 Palaiseau, France

Contents

1	Introduction	2
2	Green functions and Dyson equations	2
3	Linear response	5
4	Self-energies and generalized response: the Bethe-Salpeter equation	9
5	The Bethe-Salpeter equation from the GW approximation	14
6	A two-body Schrödinger equation	17
7	Excitons and correlation	20

1 Introduction

In this lecture we will present the Bethe-Salpeter equation (BSE) from the point of view of its use in condensed matter physics and chemistry. In this context it is most often applied to the calculation of optical spectra. For this reason we will work in the framework of linear response theory, although, as we will see at the end, there are other uses of the BSE.

The BSE solves a many-body problem, expressed in terms of electrons, holes, and their interaction. It is convenient to formulate such a problem in terms of Green functions. Therefore, we will start by briefly recalling the necessary tools.

2 Green functions and Dyson equations

Green functions are often encountered in scattering theory. Suppose a system is described by \hat{H}_0 , and \hat{V} indicates an extra potential that acts as a center for scattering. Assuming that the energies ω form a continuum, one only has to determine the wavefunctions of the scattered states. The Lippmann-Schwinger equation [1] gives the relation between an unperturbed state $|\phi^0\rangle$ and an eigenstate of the full Hamiltonian $|\phi\rangle$ at the same energy ω :

$$|\phi\rangle = (1 - G_0 \hat{V})^{-1} |\phi^0\rangle \quad \text{with} \quad G_0 \equiv (\omega + i\eta - \hat{H}_0)_{|\eta \rightarrow 0^+}^{-1}. \quad (1)$$

The Green function G_0 depends only on the unperturbed system. Moreover, it contains a boundary condition: one imposes that the scattering contribution $|\phi\rangle - |\phi^0\rangle$ contains only outgoing contributions. This boundary condition, which guarantees that the solution is causal, is fulfilled thanks to the positive infinitesimal η . Eq. (1) is equivalent to $|\phi\rangle = GG_0^{-1}|\phi^0\rangle$ with the definition of the full Green function $G \equiv (1 - G_0 \hat{V})^{-1} G_0$, which fulfills the *Dyson equation* $G = G_0 + G_0 \hat{V} G$.

Like the Hamiltonian, the Green functions are non-local in space. In general they can also be non-local in a spin coordinate. Moreover, their dependence on the frequency ω (see Eq. (1)) corresponds to a dependence on a time difference (whereas G depends on two times when the Hamiltonian is time-dependent). In the following we denote a space-spin-time argument with $1 \rightarrow (x_1, t_1) \rightarrow (\mathbf{r}_1, \sigma_1, t_1)$, and we use the convention that arguments with a bar are integrated over: $f(\bar{1})g(\bar{1}) \rightarrow \int d1 f(1)g(1)$. Then the Dyson equation can be written as

$$G(1, 2) = G_0(1, 2) + G_0(1, \bar{3}) \hat{V}(\bar{3}) G(\bar{3}, 2). \quad (2)$$

The Dyson equation is a general way to move from the Green function of a simpler system to the Green function of a system in presence of an extra potential, which may depend on space, spin and time. In the following we do not display spin unless necessary, supposing that we are interested in spin-unpolarized systems.¹ Moreover for simplicity we assume the temperature to

¹In the most general case the Green function depends on two spin arguments. When the interaction is spin-independent, the Green function is spin-diagonal, and in the absence of spin polarization, the two spin components are equal. We mostly suppose to be in that case, and do not display spin for simplicity. Details on the spin-dependent BSE can be found in [2].

be zero, and the system to be in its ground state. To simplify notation we also omit the hat on operators whenever it does not create confusion.

In the many-body electron system, there is scattering because of the Coulomb interaction between electrons, so the same considerations as above hold. The starting G_0 is an independent-particle Green function, describing electrons that are *not* scattered by other electrons.² Its retarded version G_0^R reads

$$G_0^R(\mathbf{r}, t, \mathbf{r}', t') = -i\Theta(t - t') \sum_s \psi_s(\mathbf{r}) \psi_s^*(\mathbf{r}') e^{-i\varepsilon_s(t-t')}, \quad (3)$$

where ε_s and ψ_s are the eigenvalues and eigenfunctions of the single-electron Hamiltonian h_0 . The ε_s are the poles of G_0 in frequency space, where $G_0^R(\mathbf{r}, \mathbf{r}'; \omega) = \sum_s \frac{\psi_s(\mathbf{r}) \psi_s^*(\mathbf{r}')}{\omega - \varepsilon_s + i\eta}$. Often it is more convenient to work with time-ordered Green functions instead of retarded ones; we will also adopt this framework here. The time-ordered Green function G_0 of an independent-particle system reads

$$G_0(\mathbf{r}, t, \mathbf{r}', t') = -i \left[\Theta(t - t') \Theta(\varepsilon_s - \mu) - \Theta(t' - t) \Theta(\mu - \varepsilon_s) \right] \sum_s \psi_s(\mathbf{r}) \psi_s^*(\mathbf{r}') e^{-i\varepsilon_s(t-t')} \quad (4)$$

so that electrons (states above the chemical potential μ) and holes (states below μ) contribute with opposite sign.

The independent-particle Green function yields some important observables of the independent-particle system. In particular, the density is $n_0(\mathbf{r}) = \sum_s^{\text{occ}} |\psi_s(\mathbf{r})|^2 = -iG_0(\mathbf{r}, t, \mathbf{r}, t^+)$, where t^+ stands for $t + \eta$. The diagonal of the *spectral function* $A_{ss}(\omega) = \frac{1}{\pi} \text{Im} G_{0,ss}(\omega)$, which is the imaginary part of the Green function in frequency space, is $A_{ss}^0(\omega) = \delta(\omega - \varepsilon_s)$: it exhibits the spectrum of electron addition and removal energies.

Suppose now that we add an extra static potential v_a to the system. The new Green function G can be obtained from the Dyson equation (2) where \hat{V} is replaced by v_a . It will have the same form as Eq. (4), but the eigenvalues and eigenfunctions that appear in Eq. (4) are those of $h_0 + v_a$. They can also be used to evaluate the density and spectral function as above. Importantly for our purpose, the independent-particle expressions are also valid when v_a is not some *external* potential, but a system-*internal* mean-field potential such as the Hartree (v_H) or Kohn-Sham (v_{xc}) ones. They even hold when one introduces a spatially non-local mean field, such as the Hartree-Fock potential Σ_{Hx} , for which the Dyson equation reads

$$G(1, 2) = G_0(1, 2) + G_0(1, \bar{3}) \Sigma_{Hx}(\bar{3}, \bar{4}) G(\bar{4}, 2). \quad (5)$$

Note that like the Hartree or Kohn-Sham potentials, the Hartree-Fock potential is instantaneous, i.e., local in time, which means that $\Sigma_{Hx}(3, 4)$ is proportional to $\delta(t_3^+ - t_4)$.

When v_a depends explicitly on time, i.e., the system is out of equilibrium, G is no longer of the simple form of Eq. (3). Although we will implicitly apply such a potential to our system later, we do not need to consider the resulting Green functions explicitly here, since we will limit ourselves to linear response.

²An independent-particle system can be the non-interacting one in some external potential, or it can have also a part of the interaction included through a static mean field.

However, below we will replace the instantaneous Hartree-Fock mean field by a more general self-energy $\Sigma(3, 4)$, which is *not* instantaneous, i.e., it depends on two time arguments (t_3, t_4) , or one time difference $(t_3 - t_4)$ (one frequency) in equilibrium. As a consequence, G has no longer the same simple structure as G_0 . Still, it can be calculated from the Dyson equation, if the self-energy is known.

There are two ways to understand why the full self-energy is not instantaneous, or equivalently, why the full Green function has a form different from Eq. (4). They are linked to the definition and meaning of G . The one-body Green function is defined such that it describes electron addition and removal from the full many-body system. In the Hartree and Hartree-Fock single-particle Schrödinger equation the eigenvalues are directly the addition and removal energies; this is Koopmans' theorem. It implies that electrons are added or removed without influencing the already present system electrons. However, in reality the system should react to the addition of a charge, which leads to *screening*. This reaction is not instantaneous, but it needs time to build up a screening cloud, and charge oscillations can be excited.³ This explains why the total effective potential, which includes the self-energy, is not instantaneous but depends on a time difference.

The fundamental difference between a static mean-field and a fully interacting system can also be understood by looking directly at the Green function. The generalization of Eq. (4) to the fully interacting Green function at zero temperature is [3]

$$G(\mathbf{r}, t, \mathbf{r}', t') = -i \langle N | T [\hat{\psi}(\mathbf{r}, t) \hat{\psi}^\dagger(\mathbf{r}', t')] | N \rangle, \quad (6)$$

where $|N\rangle$ is the N -particle many-body ground state, $\hat{\psi}$ are field operators in the Heisenberg picture, and T is the time-ordering operator defined as

$$T[A(t_1)B(t_2)] \equiv \begin{cases} A(t_1)B(t_2), & t_1 > t_2 \\ B(t_2)A(t_1), & t_1 < t_2. \end{cases} \quad (7)$$

Eq. (6) shows that the Green function G is the probability amplitude to find an electron in (\mathbf{r}, t) if it has been inserted in (\mathbf{r}', t') (and vice versa for a hole). This definition reduces to Eq. (4) in absence of interaction, where the ground state $|N\rangle$ is a Slater determinant built with the single particle orbitals $\psi_s(\mathbf{r})$. As one can see from the definition (6) of the Green function, in analogy to the non-interacting case,

$$n(\mathbf{r}) = -iG(\mathbf{r}, t, \mathbf{r}t^+). \quad (8)$$

The spectral function becomes

$$A_{ss}(\omega) = -\frac{1}{\pi} \text{Im} G_{ss}(\omega) = \sum_{\lambda} |f_{s\lambda}|^2 \delta(\omega - \varepsilon_{\lambda}), \quad (9)$$

³Note that especially in finite systems the Hartree or Hartree-Fock approximations are also used in a Δ self-consistent field (Δ -SCF) approach, where charges are explicitly added to the system, which is allowed to relax self-consistently. In this case, one implicitly includes screening, although in an adiabatic approximation that does not lead to excitations.

where ε_λ are electron addition and removal energies, and the $f_{s\lambda}$ are Dyson amplitudes projected on single particle orbitals. These projections are in general not sharp. Therefore, contrary to the non-interacting case the interacting spectral function is not just a single δ -peak, and it shows a continuous distribution of weight in an extended system, although a pronounced “quasi-particle” peak may still dominate the spectrum. Such a spectral function can only be produced if G is solution of a Dyson equation with self-energy that is not instantaneous, and which has therefore a Fourier transform that depends on frequency.

3 Linear response

In the following we will concentrate mainly on spectroscopic measurements, such as optical absorption, electron energy loss experiments, or inelastic x-ray scattering. The experimental results can often be understood in terms of linear response theory. In particular, they are related to the frequency-dependent dielectric function $\epsilon(\omega)$, or equivalently, to the linear density-density response function $\chi(\omega)$. The linear response is the first-order change of the density δn in a system due to an external perturbation v_{ext} , given as $\delta n = \chi v_{\text{ext}}$. The response function and the inverse dielectric function are related by

$$\epsilon^{-1}(\mathbf{r}_1, \mathbf{r}_2; \omega) = \delta(\mathbf{r}_1 - \mathbf{r}_2) + \int d\mathbf{r}_3 v_c(|\mathbf{r}_1 - \mathbf{r}_3|) \chi(\mathbf{r}_3, \mathbf{r}_2; \omega), \quad (10)$$

where v_c is the bare Coulomb interaction. More specifically, in a periodic system, this reads

$$\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}; \omega) = \delta_{\mathbf{G}\mathbf{G}'} + v_c(\mathbf{q} + \mathbf{G}) \chi_{\mathbf{G}\mathbf{G}'}(\mathbf{q}; \omega), \quad (11)$$

where \mathbf{q} is a vector in the first Brillouin zone and \mathbf{G} is a reciprocal lattice vector. Spectra are obtained from ϵ or χ . The most important quantities are:

- **The loss function** $-\text{Im} \epsilon_{\mathbf{G}\mathbf{G}}^{-1}(\mathbf{q}; \omega) = -v_c(\mathbf{q} + \mathbf{G}) \text{Im} \chi_{\mathbf{G}\mathbf{G}}(\mathbf{q}; \omega)$. This quantity can be measured in an electron microscope by performing a momentum-resolved electron energy loss experiment with a selected momentum transfer $\mathbf{Q} = \mathbf{q} + \mathbf{G}$.
- **The dynamic structure factor** $S(\mathbf{Q}, \omega) = -\frac{1}{\pi} \text{Im} \chi_{\mathbf{G}, \mathbf{G}}(\mathbf{q}, \omega)$. One can measure S as a function of energy and momentum transfer $\mathbf{Q} = \mathbf{q} + \mathbf{G}$ at a synchrotron by performing inelastic x-ray scattering (IXS).
- **The optical absorption spectrum** $\text{Im} \epsilon_M(\mathbf{q}, \omega)$. Under certain conditions, it is given by the macroscopic dielectric function $\epsilon_M(\mathbf{q}, \omega) = 1/[\epsilon(\mathbf{q}, \omega)]_{\mathbf{G}=\mathbf{G}'=0}^{-1}$ in the limit of long wavelength, $\mathbf{q} \rightarrow 0$. In rather homogeneous systems, where the off-diagonal elements of the matrix $\epsilon_{\mathbf{G}, \mathbf{G}'}$ are small, $\epsilon_M(\mathbf{q}, \omega) \approx \epsilon_{\mathbf{G}=\mathbf{G}'=0}(\mathbf{q}, \omega)$.

Note that ϵ^{-1} yields the screened Coulomb interaction, $W = \epsilon^{-1} v_c$, which is the effective interaction between classical charges in a medium. It is a key quantity in Hedin’s equations and in the GW approximation to the self-energy. Both will be briefly recalled in Sec. 4.

We will now see how to calculate linear response using Green functions. Let us concentrate on the density-density response function $\chi = \delta n / \delta v_{\text{ext}}$. With Eq. (8), this can be written as

$$\chi(1, 2) = iG(1, \bar{3}) \frac{\delta G^{-1}(\bar{3}, \bar{4})}{\delta v_{\text{ext}}(2)} G(\bar{4}, 1^+), \quad (12)$$

where we have used $\delta G / \delta v = -G [\delta G^{-1} / \delta v] G$. Eq. (12) is a convenient starting point for the various approximations that we will consider in the following.

Let us first look at the case of non-interacting particles. We can apply a potential v_{ext} to the system and evaluate the linear response. In that case we have

$$G^{-1}[v_{\text{ext}}] = G_0^{-1} - v_{\text{ext}}, \quad (13)$$

where we have indicated explicitly that now G is a functional of v_{ext} . Using Eq. (12) we obtain the independent-particle response function

$$\chi_0(1, 2) = -iG_0(1, 2)G_0(2, 1^+). \quad (14)$$

Because of the opposite time-ordering of the two Green functions, one of the two G_0 contributes only with a sum over occupied states v in Eq. (4), and the other one with a sum over empty states c . Altogether, using Eq. (4) for G_0 , one finds the usual expression for the linear response function χ_0 of a non-interacting system, consisting of a sum over all possible transitions from occupied to empty states (the *resonant* part) and vice versa (the *anti-resonant* part):

$$\chi_0(\mathbf{r}, \mathbf{r}', \omega) = \sum_{vc} \left[\frac{\psi_v^*(\mathbf{r}) \psi_c(\mathbf{r}) \psi_c^*(\mathbf{r}') \psi_v(\mathbf{r}')}{\omega - (\varepsilon_c - \varepsilon_v) - i\eta} - \frac{\psi_v(\mathbf{r}) \psi_c^*(\mathbf{r}) \psi_c(\mathbf{r}') \psi_v^*(\mathbf{r}')}{\omega + (\varepsilon_c - \varepsilon_v) + i\eta} \right]. \quad (15)$$

This *independent-particle approximation* is frequently used to describe absorption spectra. Instead, it is not at all appropriate to directly access loss spectra. The difference between absorption and loss is illustrated in Fig. 1: The left panel shows the absorption spectrum of bulk silicon, the right panel shows the imaginary part of the full density-density response function χ , as measured in electron energy loss spectroscopy at vanishing momentum transfer. The absorption spectrum rises steeply above 3 eV, which corresponds to the direct band gap. Instead, the loss spectrum has its main feature at much higher energy, around 17 eV. In order to understand how this difference comes about, let us move on and put some interaction in our system.

Suppose that we do this first on a mean-field level, by adding a potential v_{mf} . Since this potential is supposed to stem from the interaction between all electrons, in the spirit of DFT it is reasonable to assume that it is a functional of the density, $v_{\text{mf}} = v_{\text{mf}}[n]$. Now we have

$$G^{-1} = G_0^{-1} - v_{\text{ext}} - v_{\text{mf}}[n], \quad (16)$$

which leads to

$$\chi(1, 2) = \chi_0(1, 2) - iG(1, \bar{3}) \frac{\delta v_{\text{mf}}[n](\bar{3})}{\delta v_{\text{ext}}(2)} G(\bar{3}, 1^+), \quad (17)$$

where $\chi_0 = -iG G_{|v_{\text{ext}} \rightarrow 0}$ is now built with a pair of equilibrium mean-field Green functions.

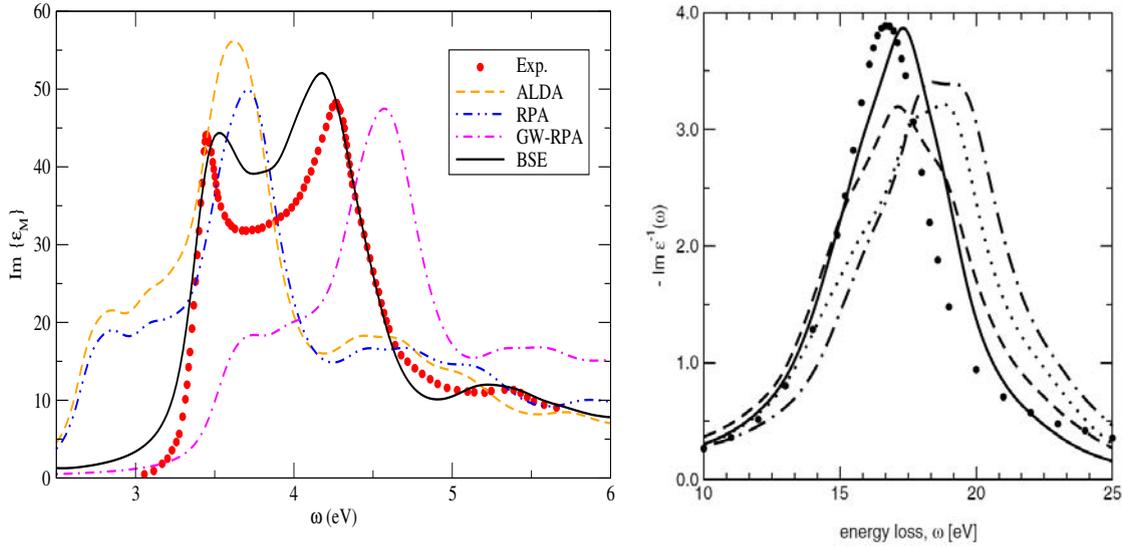


Fig. 1: Electronic spectra of bulk silicon. Left panel: absorption spectrum $\text{Im } \epsilon_M(\omega)$ (figure by Francesco Sottile; experiment from [4]). Right panel: loss spectrum for vanishing momentum transfer, from [5]. The important curves are experiment (dots), RPA (dashed), and BSE (continuous).

Since v_{mf} is a density functional, and the density itself depends on the external potential v_{ext} , we can use a chain rule in the functional derivative, which leads to a Dyson equation for χ

$$\begin{aligned} \chi(1, 2) &= \chi_0(1, 2) - iG(1, \bar{3}) \frac{\delta v_{\text{mf}}[n](\bar{3})}{\delta n(\bar{5})} \frac{\delta n(\bar{5})}{\delta v_{\text{ext}}(2)} G(\bar{3}, 1^+) \\ &= \chi_0(1, 2) + \chi_0(1, \bar{3}) \frac{\delta v_{\text{mf}}[n](\bar{3})}{\delta n(\bar{5})} \chi(\bar{5}, 2). \end{aligned} \quad (18)$$

One important example for such a mean-field response is the time-dependent Hartree approximation, where $v_{\text{mf}}[n](\mathbf{r}, t) = v_H[n](\mathbf{r}, t) = \int d\mathbf{r}' v_c(\mathbf{r} - \mathbf{r}') n(\mathbf{r}', t)$. In that case the functional derivative, the *kernel* of the Dyson equation, equals simply v_c , and χ_0 is built with Hartree Green functions. More generally, an approximation of the form $\chi = \chi_0 + \chi_0 v_c \chi$ for any $\chi_0 = iGG$, where G is some mean-field Green function, is called *Random Phase Approximation* (RPA). The RPA has been proposed for the homogeneous electron gas by Pines and Bohm [6–8], and it is today used in many contexts and for many materials.

The natural way to go beyond the time-dependent Hartree approximation is Time-Dependent Density-Functional Theory (TDDFT). We can do this by adding a Kohn-Sham exchange-correlation potential v_{xc} such that $v_{\text{mf}} = v_H + v_{\text{ext}} + v_{\text{xc}}$. This potential should be a functional of the density in the whole space, and at all past times.⁴ Following the same path that has led to Eq. (18) this yields

$$\chi(1, 2) = \chi_0(1, 2) + \chi_0(1, \bar{3}) [\delta(t_{\bar{3}} - t_{\bar{5}}) v_c(\mathbf{r}_{\bar{3}} - \mathbf{r}_{\bar{5}}) + f_{\text{xc}}(\bar{3}, \bar{5})] \chi(\bar{5}, 2), \quad (19)$$

⁴This is a requirement of causality. TDDFT is usually formulated in a causal framework. We use mostly a time-ordered formulation in these lecture notes, because this facilitates many-body perturbation theory which is the main topic here. However, it is easy to move from one to the other, as one can see for example by comparing Eqs. (3) and (4). One only has to be careful to be consistent.

where we have defined the exchange-correlation kernel $f_{xc}(3, 5) = \delta v_{xc}(3)/\delta n(5)$. Note that, contrary to the Hartree part, it is not instantaneous.

If the exact v_{xc} were known, χ from Eq. (19) would be the exact density-density response function, because the potential would yield the exact time-dependent density. However, this is not the case. Most often very simple approximations are used, such as the adiabatic local density approximation (ALDA), where $v_{xc}(\mathbf{r}, t)$ depends only on the density at point \mathbf{r} and time t and $f_{xc}(3, 5)$ is therefore proportional to $\delta(t_3 - t_5) \delta(\mathbf{r}_3 - \mathbf{r}_5)$. In extended systems, the ALDA often yields results close to the RPA ones, when the same χ_0 is used. This is illustrated in the left panel of Fig. 1. The curves labeled RPA and ALDA have been obtained using LDA Green functions for G . The ALDA shows only minor modifications with respect to the RPA. Both are not very good: the Kohn-Sham gap in χ_0 underestimates the experimental gap by about 50%. This is not recovered by the RPA or ALDA kernels, so the onset of absorption is underestimated with respect to experiment. Also the lineshape differs from the measured one, since there is not enough oscillator-strength on the low-energy side, although one can still recognize a correspondence between calculated and measured spectra. The loss spectrum in the right panel, instead, is reasonably well described by the RPA.

In order to understand the difference between absorption and loss spectra, it is enough to look at the RPA, and to neglect off-diagonal elements of the matrices in reciprocal space. Then, as outlined in the beginning of this section, absorption is given approximately by

$$\text{Im } \epsilon_{\mathbf{G}=\mathbf{G}'=0}(\mathbf{q}, \omega) = -v_c(\mathbf{q}) \text{Im } \chi_{0,\mathbf{G}=\mathbf{G}'=0}(\mathbf{q}, \omega) \quad \text{for } \mathbf{q} \rightarrow 0,$$

whereas the loss function at vanishing momentum transfer is

$$-\text{Im} [1/(1 - v_c(\mathbf{q})\chi_{0,\mathbf{G}=\mathbf{G}'=0}(\mathbf{q}, \omega))] \quad \text{for } \mathbf{q} \rightarrow 0.$$

To first order in the Coulomb interaction the two expressions are equal. However, the Coulomb interaction is strong, and the difference is very obvious for bulk silicon in Fig. 1. As anticipated, comparison of the two panels shows that the loss function in the right panel has its main structures at much higher energies than the absorption spectrum, which is shown in the left panel. In the RPA the difference between the two only stems from the long-range Coulomb interaction v_c . This interaction causes a correlated motion of all particles as response to an external perturbation. These are long-range charge oscillations, called plasmons. They give rise to the strong peak in the loss function in Fig. 1. With $v_c(q) = 4\pi/q^2$, the Coulomb kernel is particularly important for small momentum transfer q in extended systems. It is the dominant effect in loss spectra.

At larger momentum transfer $4\pi/q^2$ is smaller, and the two kinds of spectra are more similar. The energy of the plasmon changes as a function of momentum transfer. With this plasmon dispersion, the sharp peak moves into the continuum of electron-hole transitions and decays into a broad structure. Plasmons are a broad topic, and more can be found for example in [9]. Here the important lesson to take away is that the RPA contains the physics of plasmons, because it includes the long-range variation of the Hartree potential.

The neglect of off-diagonal elements is an approximation. As Eq. (11) shows, because of scattering at the periodic crystal potential, $\epsilon_{\mathbf{G}\mathbf{G}'}$ is not diagonal, so its inversion (or, equivalently, the solution of the Dyson equation) mixes different momentum transfers. Therefore, even when one is interested in the response to a macroscopic, long-range, perturbation the system can respond with charge fluctuations on a microscopic scale. These are included in a calculation through the $\mathbf{G} \neq 0$ components of $v_c(\mathbf{q} + \mathbf{G})$. The effects of the microscopic components of v_c are called *crystal local-field effects* (LFE) [10–12]. Microscopic components of the induced Hartree potential gain in importance when the system is inhomogeneous, and when one probes shorter distances, with increased momentum transfer. In absorption spectra, only the microscopic components contribute, since the macroscopic component of v_c is eliminated by the double inversion $\epsilon_M = 1/\epsilon^{-1}$. The effect of the microscopic components (the LFE) is usually moderate in extended systems, with variations of spectra of the order of 10% or less for simple semiconductors, but they can have more important effects in systems with localized electrons, or in layered systems, which are more inhomogeneous.

The fact that the RPA contains plasmons and crystal local-field effects explains its generally good performance in describing loss spectra, the dynamic structure factor, and therefore also the screened Coulomb interaction W . The limitations of the RPA, instead, can be best detected when one looks at optical absorption in the first few eV spectral range. We have seen the example of silicon in Fig. 1 above. It is typical for simple semiconductors.

The situation is even worse for insulators, where strongly bound *excitons* can occur. Roughly speaking, excitons are due to an effective interaction between the excited electron and the hole left behind. Since these are a positive and a negative charge, similar to a hydrogen atom, the interaction can lead to bound states that are clearly detected in experimental absorption spectra. However, neither in the RPA nor in the ALDA bound excitons can be described. To date, a few exchange-correlation kernels exist that can produce bound excitons, but their reliability and/or computational efficiency are not yet satisfactory. Moreover, none of them can overcome the problem that the Kohn-Sham band gap is usually smaller than the threshold of optical absorption. This means that the kernel should shift the spectrum to higher energies, which turns out to be a very difficult task. These problems suggest to move away from density functional theory, towards Green function functional theory, using self-energies and many-body perturbation theory.

4 Self-energies and generalized response: the Bethe-Salpeter equation

In order to understand how moving to the framework of Green functions and self-energies can cure the problems of approximate density functionals, we can look at the simplest approximation to the self-energy, the Fock exchange operator Σ_x .

Let us start with the band gap. In Hartree-Fock (HF) we have Koopmans' theorem, which states that HF eigenvalues equal electron removal and addition energies, expressed as Hartree-Fock

total energy differences. Since relaxation is not included in this framework, the HF band gap is in general much too large compared to experiment. For example, in silicon the direct HF gap is almost 9 eV, whereas the experimental value is about 3 eV (see Fig. 2).

Let us now try to calculate the HF response function using, again, Eq. (12). From Eq. (5) the equivalent to Eq. (16) is

$$G^{-1} = G_0^{-1} - v_{\text{ext}} - v_H - \Sigma_x. \quad (20)$$

Strictly speaking, like v_H also the exchange self-energy Σ_x is a functional of the density. However, its explicit form is not known, maybe not existent or at best non-analytic. Instead, we know $\Sigma_x(1, 2)$ as a functional of the one-body spin-resolved density *matrix* $\rho(1, 2) = -iG(x_1, t_1; x_2, t_1^+)$; it is $\Sigma_x(1, 2) = -\delta(t_1^+ - t_2) v_c(\mathbf{r}_1 - \mathbf{r}_2) \rho(1, 2)$. Let us now try to follow the lines of Eq. (18) while using ρ instead of the density n in the chain rule for Σ_x . This yields

$$\chi(1, 2) = \chi_0(1, 2) + \chi_0(1, \bar{3})\delta(t_{\bar{3}} - t_{\bar{5}})v_c(\mathbf{r}_{\bar{3}} - \mathbf{r}_{\bar{5}})\chi(\bar{5}, 2) + \chi_0^{\text{nl}}(1; \bar{4}, \bar{3})\frac{\delta\Sigma_x[\rho](\bar{3}, \bar{4})}{\delta\rho(\bar{6}, \bar{5})}\chi_{\text{nl}}(\bar{6}, 2, \bar{5}), \quad (21)$$

where we have defined the three-point response functions

$$\chi_0^{\text{nl}}(1; 4, 3) \equiv -iG(1, 3)G(4, 1^+) \quad \text{and} \quad \chi_{\text{nl}}(6, 2, 5) \equiv \frac{\delta\rho(6, 5)}{\delta v_{\text{ext}}(2)}. \quad (22)$$

The additional non-locality in χ_0^{nl} stems from the non-locality of Σ_x and is not problematic. Instead, the functional derivative with respect to the non-local density matrix creates a problem: Eq. (21) is not a closed equation for the desired $\chi(1, 2)$. In order to obtain a closed equation, we have to generalize the equation to make it fully three-point, by looking from the very start at $\delta\rho(1, 1')/\delta v_{\text{ext}}(2)$. With $\delta\Sigma_x[\rho](3, 4)/\delta\rho(6, 5) = -\delta(t_3^+ - t_4)\delta(3, 6)\delta(4, 5)v_c(\mathbf{r}_3 - \mathbf{r}_4)$ and carrying out the same steps as before, this leads to

$$i\chi_{\text{nl}}(1, 2, 1') = G(1, 2)G(2, 1') + G(1, \bar{3})G(\bar{4}, 1')i\Xi_{H_x}(3, 5, 4, 6)\chi_{\text{nl}}(\bar{6}, 2, \bar{5}), \quad (23)$$

where the kernel of this Dyson equation reads

$$i\Xi_{H_x}(3, 5, 4, 6) \equiv i\frac{\delta\Sigma_{H_x}(3, 4)}{\delta G(6, 5)} = \delta(3, 4)\delta(5, 6)v_c(3, 5) - \delta(3, 6)\delta(4, 5)v_c(3, 4), \quad (24)$$

and $v_c(1, 2)$ includes the δ -function in time. This is now a closed equation for the response of the spin-resolved density *matrix*. In order to obtain the desired density response, one has *first to solve* for the density matrix response, and *then* use the fact that by definition $\chi(1, 2) = \chi_{\text{nl}}(1, 2, 1)$.⁵ Note that G are now HF Green functions at $v_{\text{ext}} = 0$. Indeed, Eq. (23) is the linear response in the time-dependent Hartree Fock (TDHF) approximation.

Fig. 2 shows the result for bulk silicon, taken from [13]. The dots are the experimental spectrum, the same as in Fig. 1. At the far right of the figure, the dot-dashed curve represents $\text{Im}\chi_0$ built with HF Green functions. Since the HF gap is almost 9 eV, the result is far off experiment.

⁵For the response of the total density one also has to sum over spin.

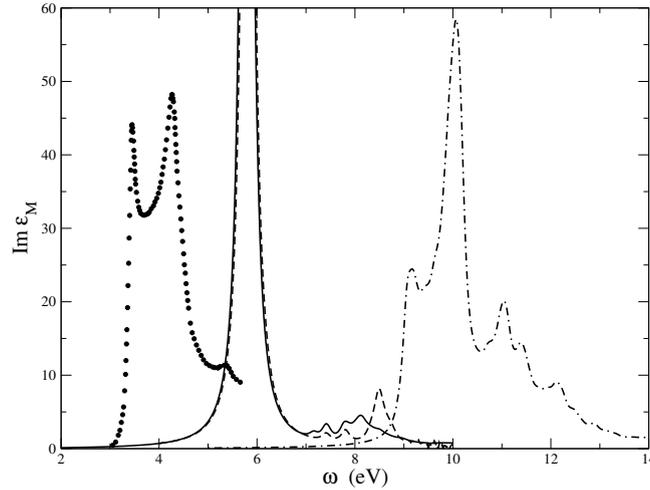


Fig. 2: Absorption spectrum of bulk silicon in the TDHF approximation, from [13]. Dots: experiment. Dot-dashed curve: independent-particle spectrum using Hartree-Fock ingredients. Continuous curve: TDHF result.

Instead, the kernel Ξ_{Hx} used in Eq. (23) has a dramatic effect: the spectrum becomes sharper and moves to lower energies, recovering more than half of the discrepancy to experiment. This very important effect is due to the second part of Ξ_{Hx} , the derivative of Σ_x . The effect of the first part is a shift of spectral weight to higher energies, but it is very small since, as explained earlier, only the microscopic part of the variation of the Hartree potential contributes to the final absorption spectrum.

The TDHF case contains almost everything we need to understand the Bethe-Salpeter equation:

- A Dyson equation for the density-matrix response has to be solved in order to obtain the density response.
- The starting χ_0 has a gap that can be interpreted as a difference between electron addition and removal energies. This is called the quasi-particle gap, and it is the gap that would be measured for example in direct and inverse photoemission.
- The variation of the Hartree potential is the same as in the RPA. Its effect on absorption spectra is moderate (whereas it is responsible for plasmons that are seen in loss spectra).
- The variation of the Fock exchange moves the spectrum to lower energies. Its effect is strong. We find now spectral weight within the quasi-particle gap: this means that we have a bound exciton. In other words, the variation of the exchange is responsible for the electron-hole attraction.

All this might seem to be meaningless, since what judges a theory at the end is agreement with experiment – and TDHF visibly does less well than the simple RPA based on an LDA G shown in Fig. 1! However, the problem of HF is clear: it is the absence of screening (or more generally formulated, of correlation), which makes band gaps too large, and interactions too strong. The

introduction of screening, even in relatively simple ways, brings remarkable improvement. For example, hybrid functionals include screening in an effective way by adding a fraction of Fock exchange to local Kohn-Sham potentials. Moving from time-dependent HF to time-dependent hybrid-functional calculations leads to very decent absorption spectra in semiconductors, as has been shown for example in [14].

A more systematic way to introduce screening is to get better self-energies Σ_{xc} instead of Σ_x from many-body perturbation theory. We will do this in the next section; here we will conclude by generalizing Eq. (23) to a form that is usually called the *Bethe-Salpeter equation* (BSE).

The BSE describes a generalized response, where a non-local (in space, spin and time) “potential” is applied to the system, and the variation of the Green function, instead of its equal-time limit (the density matrix) is determined. Moreover, since now we are heading for a more general self-energy, we can no longer suppose that it is known as functional of the density matrix; instead, we will have approximations that are explicit functionals of the Green function. With this in mind, all steps can be carried out in close analogy to the derivation of the TDHF equations.

With the definition

$$L(1, 2, 1', 2') \equiv \frac{\delta G(1, 1')}{\delta v_{\text{ext}}(2', 2)} \quad (25)$$

we find the Bethe-Salpeter equation [15]

$$L(1, 2, 1', 2') = L_0(1, 2, 1', 2') + L_0(1, \bar{3}', 1', \bar{3}) \Xi(\bar{3}, \bar{2}, \bar{3}', \bar{2}') L(\bar{2}', 2, \bar{2}, 2'), \quad (26)$$

with

$$\Xi(\bar{3}, \bar{2}, \bar{3}', \bar{2}') \equiv -i\delta(\bar{3}, \bar{3}') \delta(\bar{2}', \bar{2}) v_c(\bar{3}, \bar{2}) + \frac{\delta \Sigma_{xc}(\bar{3}, \bar{3}')}{\delta G(\bar{2}', \bar{2})}. \quad (27)$$

The uncorrelated $L_0(1, 2, 1', 2') = G(1, 2') G(2, 1')$ contains the Green function G , solution of the Dyson equation

$$G(1, 2) = G_0(1, 2) + G_0(1, \bar{3}) \Sigma(\bar{3}, \bar{4}) G(\bar{4}, 2). \quad (28)$$

This is analogous to Eq. (5), with Σ_{Hx} replaced by the full self-energy $\Sigma = v_H + \Sigma_{xc}$.

From the definition of L in Eq. (25) it follows that

$$\chi(1, 2) = -iL(1, 2; 1^+, 2^+); \quad (29)$$

as in the case of TDHF, χ can only be obtained after solving the full BSE for L .

The four-point function L is a two-particle correlation function. One can calculate it formally from Eq. (25), starting from a one-body Green function G in the presence of an external potential. The derivation is delicate since the potential can be non-local in time [16], but the result is qualitatively intuitive: the applied potential contributes $v_{\text{ext}} \psi^\dagger \psi$ to the time evolution in the Heisenberg picture. The derivative of $G = -i\langle N | T [\hat{\psi} \hat{\psi}^\dagger] | N \rangle$ with respect to v_{ext} leads therefore to an expression with four field operators. It is closely linked to the two-particle Green function

$$G_2(1, 2, 1', 2') = (-i)^2 \langle T [\hat{\psi}(1) \hat{\psi}(2) \hat{\psi}^\dagger(2') \hat{\psi}^\dagger(1')] \rangle, \quad (30)$$

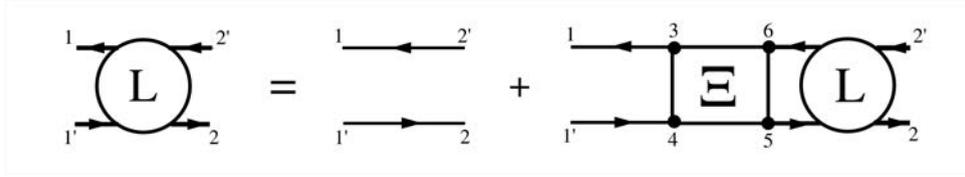


Fig. 3: The Bethe Salpeter equation for the propagation of an electron-hole pair.

via

$$L(1, 2, 1', 2') = -G_2(1, 2, 1', 2') + G(1, 1') G(2, 2'). \quad (31)$$

The two-particle correlation function, and therefore the BSE, contain much more than just the density-density response function, which is the small part given by Eq. (29). In particular, the order of times is crucial to select different pieces of physics, since it determines the order of the four field operators. In the density-density response this order alternates ψ and ψ^\dagger , which means that electron-hole pairs are described. Alternatively, the order can be for example $\psi^\dagger\psi^\dagger\psi\psi$: this describes the propagation of two holes. Also various combinations of spin-resolved response can be described. We will not delve into these subjects, but it should be clear that, given the general Bethe-Salpeter equation, these problems can be treated in strict analogy to the electron-hole case. The BSE for the propagation of an electron-hole pair is expressed graphically by the diagrams in Fig. 3.

The time structure of the BSE is a particular complication. The quantities that appear in the BSE depend on four time arguments, which corresponds in equilibrium to three time differences, or three frequencies in the Fourier transform. With the definition of the time differences [17]

$$\tau_2 = t_1 - t_{1'}, \quad \tau_3 = t_2 - t_{2'}, \quad \tau_1 = \frac{1}{2} [(t_1 + t_{1'}) - (t_2 + t_{2'})] \quad (32)$$

and of the Fourier transform

$$C(t_1, t_2, t_{1'}, t_{2'}) = \frac{1}{(2\pi)^3} \int d\bar{\omega}_1 d\bar{\omega}_2 d\bar{\omega}_3 C(\bar{\omega}_1, \bar{\omega}_2, \bar{\omega}_3) e^{-i\bar{\omega}_1\tau_1} e^{-i\bar{\omega}_2\tau_2} e^{-i\bar{\omega}_3\tau_3}, \quad (33)$$

the frequency structure of the BSE (26) is [18]

$$L(\omega_1, \omega_2, \omega_3) = L_0(\omega_1, \omega_2, \omega_3) + \frac{L_0(\omega_1, \omega_2, \bar{\omega}_4)}{(2\pi)^2} \Xi(\omega_1, \bar{\omega}_4, \bar{\omega}_5) L(\omega_1, \bar{\omega}_5, \omega_3). \quad (34)$$

The definitions are not unique, and the only requirement is to be consistent. With the present choice, τ_2 and τ_3 are differences in the time where the electron and the hole are considered, and τ_1 is the average time of propagation. In frequency space the density-density response from Eq. (29) reads

$$\chi(\omega) = \frac{1}{(2\pi)} \int d\omega_2 L(\omega, \omega_2) = \frac{1}{(2\pi)^2} \int d\omega_2 d\omega_3 L(\omega, \omega_2, \omega_3), \quad (35)$$

where we have used the same symbol L for the integrated function that depends only on two frequencies.

Since ω_3 appears as a dummy index in Eq. (34), it can be integrated before the equation is solved. The new equation reads

$$L(\omega_1, \omega_2) = L_0(\omega_1, \omega_2) + \int d\omega_4 d\omega_5 \frac{L_0(\omega_1, \omega_2, \omega_4)}{(2\pi)^2} \Xi(\omega_1, \omega_4, \omega_5) L(\omega_1, \omega_5), \quad (36)$$

with

$$L_0(\omega_1, \omega_2) = -iG(\omega_2 + \frac{\omega_1}{2}) G(\omega_2 - \frac{\omega_1}{2}). \quad (37)$$

If one performs also the integration over ω_2 in Eq. (36), one runs into a problem similar to the density response in HF, namely, the equation is no longer of closed form. Therefore, one has to solve Eq. (36) and only subsequently perform the integration.

5 The Bethe-Salpeter equation from the GW approximation

At this stage, we have everything in hand to calculate the response function starting from an arbitrarily complicated self-energy. The task is hence to find a good approximation for Σ , beyond Σ_{Hx} . We have already anticipated that the most important missing ingredient for our purpose is screening. The task of the present section is to put this hand-waving argument on a more rigorous basis. A widely used approach is diagrammatic expansions in the framework of many-body perturbation theory. Here we take another (though strictly analogous) way, which is closer to the spirit of this lecture about linear response. There is no space for a detailed derivation; more can be found in the book [19], which we closely follow here.

From the definition (6) of the Green function one can derive its equation of motion

$$G(1, 1') = G_0(1, 1') + G_0(1, \bar{2}) v_H(\bar{2}) G(\bar{2}, 1') + iG_0(1, \bar{2}) v_c(\bar{2}, \bar{3}) L(\bar{2}, \bar{3}^+, 1', \bar{3}^{++}). \quad (38)$$

It expresses the fact that the propagation of a particle in a system of many electrons equals the propagation of a single particle, modified by the classical electrostatic (Hartree) potential of all electrons, the Fock exchange that is contained in the last term (and that can be obtained with the approximation $L \approx L_0$), and correlation effects such as the reaction of the other electrons, which is expressed by the fact that L is related to a variation of G via Eq. (25).

We can now transform Eq. (38) into a Dyson equation by using Eq. (25) and the trick $\delta G/\delta v = -G[\delta G^{-1}/\delta v]G$. This defines a self-energy $\Sigma = v_H - iv_c G[\delta G^{-1}/\delta v_{\text{ext}}]$. Moreover, we can introduce screening by using the chain rule $\delta/\delta v_{\text{ext}} = [\delta/\delta v_{\text{cl}}][\delta v_{\text{cl}}/\delta v_{\text{ext}}]$, where $v_{\text{cl}} \equiv v_{\text{ext}} + v_H$. With this choice, $\delta v_{\text{cl}}/\delta v_{\text{ext}} = \epsilon^{-1}$, which makes the screened Coulomb interaction $W = \epsilon^{-1} v_c$ appear. Altogether, these manipulations lead to a set of equations known as *Hedin's equations*:

$$\Sigma_{\text{xc}}(1, 2) = iG(1, \bar{4}) W(1^+, \bar{3}) \tilde{\Gamma}(\bar{4}, 2; \bar{3}) \quad (39)$$

$$W(1, 2) = v_c(1, 2) + v_c(1, \bar{3}) P(\bar{3}, \bar{4}) W(\bar{4}, 2) \quad (40)$$

$$P(1, 2) = -iG(1, \bar{3}) G(\bar{4}, 1) \tilde{\Gamma}(\bar{3}, \bar{4}; 2) \quad (41)$$

$$\tilde{\Gamma}(1, 2; 3) = \delta(1, 2) \delta(1, 3) + \frac{\delta \Sigma_{\text{xc}}(1, 2)}{\delta G(\bar{4}, \bar{5})} G(\bar{4}, \bar{6}) G(\bar{7}, \bar{5}) \tilde{\Gamma}(\bar{6}, \bar{7}; 3) \quad (42)$$

$$G(1, 2) = G_0(1, 2) + G_0(1, \bar{3}) \Sigma(\bar{3}, \bar{4}) G(\bar{4}, 2). \quad (43)$$

These equations contain

- the *irreducible polarizability* P , Eq. (41). It is the response of the density to the total classical perturbation v_{cl} . It is linked to χ via $\chi = P + P v_c \chi$. When the *vertex function* $\tilde{\Gamma}$ is set to 1, P describes non-interacting electron-hole pairs: this is the RPA. Otherwise, $\tilde{\Gamma}$ contains the information that the two particles interact. The equation for $\tilde{\Gamma}$ can be transformed into a BSE for the irreducible part \tilde{L} (the generalization to four points of P) of L by integrating with two Green functions, or into the BSE (26) for L by including the Hartree potential.
- The *screened interaction* W is different from the bare v_c when the system is polarizable, i.e., P is not zero. It is dynamical, which means, frequency-dependent, because of the frequency-dependence of P .
- The *vertex function* $\tilde{\Gamma}$ also appears in the self-energy: it corrects for the fact that W is the screened interaction between classical charges, whereas the system particles are fermions. Moreover, P is the polarizability of all electrons, including the one that should be screened. This self-screening error is removed by the $\tilde{\Gamma}$ in the self-energy.
- Variations of the self-energy, in turn, determine the vertex function. Similarly to the bare Coulomb interaction, which is a first derivative of the Hartree potential, $\delta\Sigma_{\text{xc}}/\delta G$ plays the role of an effective exchange-correlation interaction.

The GW approximation consists in setting $\tilde{\Gamma}$ to 1 everywhere. This means that W is calculated in the RPA, and $\Sigma = iGW$. This approximation is exactly what we have been heading for: it is a sort of dynamically screened Hartree-Fock. Contrary to Hartree-Fock, the GW self-energy is not instantaneous, because W depends on a time difference.

The Dyson equation for G in Hedin's equations is the only equation that exhibits the external potential via the non-interacting G_0 . The other equations are universal, and can be used to create expressions for the quantity of interest with increasing accuracy. Typically one starts with a guess, like $\Sigma_{\text{xc}} = 0$, and from this calculates $\tilde{\Gamma}$, P , W , and then again Σ_{xc} as functional of G . This yields expressions with terms of higher orders in W . The formulae become increasingly complex, and one cannot go too far. However, at least one update of W has become a standard ingredient in the toolbox of condensed matter calculations: in a first step $\tilde{\Gamma} = 1$, which yields the RPA for P and W and the GW approximation for Σ_{xc} . Then $\tilde{\Gamma}$ is recalculated. This can be transformed into a BSE (26) for L , with $\Xi_{\text{xc}} = \delta\Sigma_{\text{xc}}^{GW}/\delta G$. The resulting kernel reads

$$\Xi_{\text{xc}}^{GWA}(1, 2, 3, 4) = i\delta(1, 4) \delta(2, 3) W(1, 2) + iG(1, 3) \frac{\delta W(1, 3)}{\delta G(4, 2)}. \quad (44)$$

The first term is very similar to Ξ_{x} of time-dependent Hartree-Fock, but now it is screened. The diagrams for the contribution to $\Xi_{H_{\text{xc}}}$ that contains only the variation of the Hartree potential and this first term are shown in Fig. 4.

The second term is of higher order in W [20], and it goes beyond the linear response of the electrons to an added charge: the screened interaction itself changes when the system is perturbed.

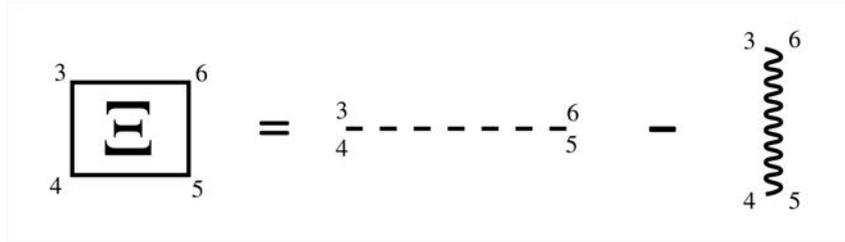


Fig. 4: Kernel of the GW-BSE, as explained in the text. The dashed line is the electron-hole exchange interaction that stems from the variation of the Hartree potential. The wiggly line represents the direct electron-hole attraction. It is depicted with a minus sign in order to stress the fact that it is attractive.

Usually this term is neglected, both because it is of higher order, and, very pragmatically, because it would be much more cumbersome to calculate. It has not been investigated very much, but there has been no strong evidence to date that it would be crucial to include it. Therefore, one usually sets

$$\Xi_{xc}^{GW}(\omega_1, \omega_2, \omega_3) \approx iW(\omega_2 - \omega_3). \quad (45)$$

Here we can see a major difference to TDHF: since the GW self-energy is not instantaneous, the BSE cannot be solved frequency by frequency. From Eq. (36), the coupling of frequencies reads

$$L(\omega_1, \omega_2) = L_0(\omega_1, \omega_2) + \frac{1}{2\pi} L_0(\omega_1, \omega_2) \int d\bar{\omega}_3 [v_c - W(\omega_2 - \bar{\omega}_3)] L(\omega_1, \bar{\omega}_3). \quad (46)$$

This looks very annoying. However, it should be noted that also L_0 has acquired a non-trivial frequency dependence: the Green functions that yield $L_0 = G G$ are now GW ones, which means, they are no longer of an independent-particle type but have a complicated spectral function, where weight is transferred from the quasi-particle to satellites. If the remaining quasi-particle weight is $Z < 1$ (let us say, 0.7), a transition between quasi-particle peaks would show up in the spectrum of L_0 with weight Z^2 [21] (which would yield a weight reduction by a factor of 1/2). This, however, is not what is observed. The reason is that these *dynamical effects* in L_0 cancel to a large extent with the dynamical effects in Ξ_{xc}^{GW} [22]. Therefore, in realistic calculations most often the Green functions G in L_0 are replaced by quasi-particle ones with weight $Z = 1$, and the frequency dependence of W in Ξ_{xc}^{GW} is neglected. Besides the fact that the quasi-particle G 's are usually derived as an approximation from the fully frequency dependent GW self-energy, one might call the resulting method “linear-response time-dependent screened Hartree-Fock”.

Now one can integrate ω_2 in Eq. (46), and the GW-BSE becomes

$$\begin{aligned} L(x_1, x_2, x_{1'}, x_{2'}; \omega) &= L_0(x_1, x_2, x_{1'}, x_{2'}; \omega) \\ &\quad - iL_0(x_1, \bar{x}_3, x_{1'}, \bar{x}_3; \omega) v_c(\bar{x}_3, \bar{x}_4) L(\bar{x}_4, x_2, \bar{x}_4, x_{2'}; \omega) \\ &\quad + iL_0(x_1, \bar{x}_4, x_{1'}, \bar{x}_3; \omega) W(\bar{x}_3, \bar{x}_4) L(\bar{x}_3, x_2, \bar{x}_4, x_{2'}; \omega). \end{aligned} \quad (47)$$

This equation still exhibits all spin arguments. The spin structure of Ξ_{Hxc}^{GW} is

$$\Xi_{\sigma_3\sigma_2\sigma_3'\sigma_2'}^{GW} = -i\delta_{\sigma_3\sigma_3'}\delta_{\sigma_2'\sigma_2}v_c + i\delta_{\sigma_3\sigma_2'}\delta_{\sigma_3'\sigma_2}W. \quad (48)$$

In spin space, the equation can be decoupled into singlet and triplet contributions [2]. In a spin-unpolarized material and without a spin-dependent interaction the result is

$$\begin{aligned} L^{singlet} &= L_0 + iL_0[W - 2v_c]L^{singlet} \\ L^{triplet} &= L_0 + iL_0WL^{triplet}. \end{aligned} \quad (49)$$

As we can see, the variation of the Hartree potential enters with a factor of two in the singlet, whereas it is absent in the triplet. This difference causes the so-called singlet-triplet splitting. In the following we concentrate on singlets, which can be optically allowed.

6 A two-body Schrödinger equation

There are various ways to solve the BSE (47), for example, by iterative inversion. For small (less than 100 electrons) systems sometimes the equation is transformed into the form of an effective two-particle Hamiltonian that is then diagonalized. This is interesting, because it suggests a simple physical interpretation. To obtain this form, we first write the equation in the basis of the orthonormal orbitals $\psi_n(\mathbf{r})$ that diagonalize G_0 , and therefore L_0 , which reads

$$L_{0n_1n_3}^{n_4n_2}(z) = 2i \frac{(f_{n_1} - f_{n_2})\delta_{n_1n_4}\delta_{n_2n_3}}{z - (\varepsilon_{n_1} - \varepsilon_{n_2})}, \quad (50)$$

where z is a complex frequency containing the appropriate infinitesimal imaginary part, f are occupation numbers, and spin has been summed. We also define interaction matrix elements as

$$v_{n_1n_3}^{n_4n_2} = \int d\mathbf{r}_1 d\mathbf{r}_2 \psi_{n_1}^*(\mathbf{r}_1)\psi_{n_4}(\mathbf{r}_1)v(\mathbf{r}_1, \mathbf{r}_2)\psi_{n_3}(\mathbf{r}_2)\psi_{n_2}^*(\mathbf{r}_2), \quad (51)$$

for both v_c and W . This transforms the BSE into

$$L_{n_1n_3}^{n_4n_2}(z) = [L_0^{-1} + \frac{i}{2}\Xi]^{-1}_{n_1n_3}{}^{n_4n_2} = 2i[H^{2p} - \mathbb{I}z]^{-1}_{n_1n_3}{}^{n_4n_2}(f_{n_2} - f_{n_4}) \quad (52)$$

with \mathbb{I} the identity matrix. Here we have defined the *effective two-particle Hamiltonian* H^{2p}

$$H_{n_1n_3}^{2p}{}^{n_4n_2} \equiv (\varepsilon_{n_2} - \varepsilon_{n_1})\delta_{n_1n_4}\delta_{n_2n_3} + (f_{n_1} - f_{n_3})\Xi_{n_1n_3}^{n_4n_2}, \quad (53)$$

where

$$\Xi_{n_1n_3}^{n_4n_2} \equiv 2v_c{}_{n_1n_3}{}^{n_3n_2} - W_{n_1n_3}^{n_4n_2}. \quad (54)$$

Optical transitions happen between occupied and empty states. Therefore the only combination of indices that is needed in a non-metal at $T = 0$ is couples of occupied and empty states, which means, we only need terms of the form

$$\Xi_{v_c}^{v'c'} = 2v_c{}_{vv'}{}^{cc'} - W_{v_c}^{v'c'}. \quad (55)$$

As we can see, the variation of the Hartree potential gives rise to a dipole-dipole interaction, called *electron-hole exchange*. The variation of the GW self-energy, instead, is called *direct electron-hole interaction*, since it contains the interaction between the charge densities of electrons and holes.

The full two-particle Hamiltonian is only pseudo-hermitian [23], because of the coupling between resonant ($v \rightarrow c$) and anti-resonant ($c \rightarrow v$) transitions. In the Tamm–Dancoff approximation (TDA) [24, 25, 3], this coupling is neglected. This is often a very good approximation for optical spectra of bulk materials [26]. It is more critical for finite systems [23]. One also has to be careful when calculating loss spectra [5], because they are influenced by the long-range part of v_c , which gives rise to strong coupling. For simplicity, in the following we give expressions in the TDA; the appropriate formula for the pseudo-hermitian full case can be found for example in [27, 19].

To perform BSE calculations in practice, one first determines W and the quasi-particle band structure, typically from a GW calculation. With this, the two-particle Hamiltonian H^{2p} is built using the expressions above. The next step is its diagonalization,

$$\sum_{n_3 n_4} H^{2p}_{n_1 n_2} A_\lambda^{n_3 n_4} = E_\lambda A_\lambda^{n_1 n_2}. \quad (56)$$

In the TDA, the retarded L is then built from

$$L_{n_1 n_2}^{n_3 n_4}(\omega) = 2i \sum_\lambda \frac{A_\lambda^{n_1 n_2} A_\lambda^{*n_3 n_4}}{\omega - E_\lambda + i\eta} (f_{n_4} - f_{n_3}). \quad (57)$$

Each couple (nn') corresponds to a pair (vc) of an occupied and an empty state. In the absence of electron-hole interaction, each eigenstate $A_\lambda^{*n_1 n_2}$ would correspond to a given electron-hole pair, $A_\lambda^{vc} = \delta_{v v_\lambda} \delta_{c c_\lambda}$, and the transition energy would be $\varepsilon_c - \varepsilon_v$. Instead, when the electron-hole interaction is switched on, one can no longer associate a transition λ with one independent-quasiparticle transition (vc) : transitions are mixed by the interaction. Note that this already occurs when only the variation v_c of the Hartree potential is considered, i.e., in the RPA: the self-consistent response of the electron system, even on a classical electrostatic level, has non-trivial effects. Of course, the interaction also affects the transition energies, which are now E_λ instead of $\varepsilon_c - \varepsilon_v$.

In a solid and for vanishing momentum transfer $\mathbf{q} \rightarrow 0$, the resonant part of the independent-particle retarded response function reads

$$\chi_{00}^0(\mathbf{q} \rightarrow 0, \omega) = 2 \sum_{v\mathbf{k}} \frac{|\tilde{\rho}_{v\mathbf{k}c}|^2}{\omega - (\varepsilon_{c\mathbf{k}} - \varepsilon_{v\mathbf{k}}) + i\eta}, \quad (58)$$

where the $\tilde{\rho}$ are dipole transition matrix elements between quasi-particle states. Instead, with all the above approximations the result obtained from the BSE is

$$\chi_{00}(\mathbf{q}, \omega) = 2 \sum_\lambda \frac{|\sum_{v\mathbf{k}} A_\lambda^{v\mathbf{k}c} \tilde{\rho}_{v\mathbf{k}c}|^2}{\omega - E_\lambda + i\eta}. \quad (59)$$

Again we can see the mixing of transitions induced by the eigenstates A_λ of the BSE two-particle Hamiltonian, and the modification of transition energies. Note that this expression yields directly the optical spectra, if the long-range ($G = 0$) contribution to v_c is omitted in the BSE.

Let us now illustrate and analyze the BSE with two examples. We will concentrate on the effect of the direct electron-hole interaction W , since the contribution v_c has been discussed earlier in the framework of the RPA.

The first example is the optical absorption of bulk silicon. The modest performance of the RPA and the ALDA has been shown in the left panel of Fig. 1, and the crucial need for screening was illustrated by the TDHF results in Fig. 2. The pink dot-dashed curve in the left panel of Fig. 1 is GW-RPA, which means, the GW-BSE is solved by neglecting W in the kernel. Since now the starting band structure is the GW one, the spectrum is at higher energies than the RPA or ALDA ones. However, it is now too much displaced to higher energies. On the other hand, since screening is included, the overshooting is not as drastic as in HF in Fig. 2. The GW-BSE results including W in the kernel are given by the continuous black curve: it shows very good agreement with experiment, both concerning position and spectral shape.

One might wonder why there is any effect of W at all in bulk silicon, since its dielectric constant is about $\epsilon_M \approx 12$, so screening is very strong, and W should be small. Indeed, a closer analysis shows that the transition *energies* are almost unchanged with respect to the GW-RPA ones. It should be noted that in infinite systems transition energies do not change to first order in W , because the first order is given by the diagonal of the matrix, which tends to zero for an infinitely dense k-point sampling, so for small W no effect on energies should be expected. The coefficients A_λ , instead, can change already to first order, since matrix elements are summed in the first order correction to eigenstates. Moreover, in the first-order perturbation correction to states, matrix elements appear in the numerator, and differences between zero-order transition energies in the denominator. The bandstructure of silicon shows almost parallel bands in large portions of the Brillouin zone, and has therefore many independent-particle transitions at similar energies. This makes the denominator small and creates a strong effect of the electron-hole interaction, even though the numerator is small because W is so strongly screened. The shift to lower energies of spectral weight in the optical spectrum of silicon is hence a pure interference effect.

On the opposite side, we find large gap insulators or low-dimensional systems with weak screening. In this case even the transition energies can be strongly affected. In particular, W is an attractive interaction, because here W is the interaction between an electron and a hole. This leads to new transition energies within the quasi-particle gap: these are the energies of bound excitons. The difference between the transition energy and the quasi-particle gap is called *exciton binding energy*. It can be as large as several eV.

Indeed, when one approximates the band structure by two parabolic bands, calculates the matrix element of W using plane waves for the orbitals, and replaces sums over k-points in the Brillouin zone by integrals over the whole k-space, the GW-BSE takes the form of a Schrödinger equation for the hydrogen atom, with a modified electron and proton mass and a screened

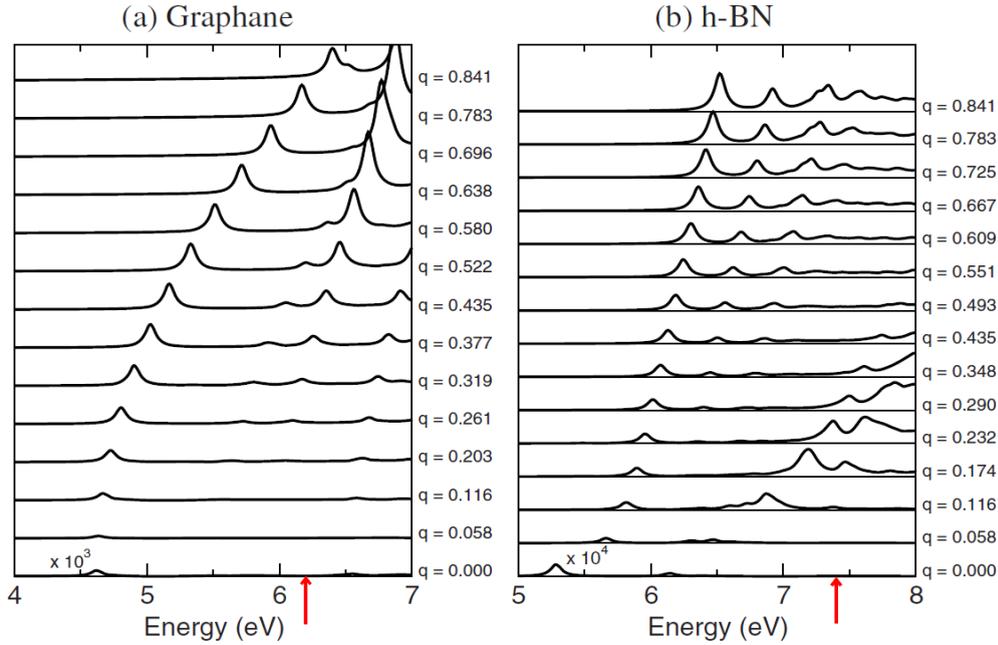


Fig. 5: Excitonic spectra $q^2 \text{Im } \epsilon_M(\mathbf{q}, \omega)$ in graphane (left) and a hexagonal BN sheet (right), for various momentum transfers \mathbf{q} along the Γ -M direction, indicated on the right side in \AA^{-1} . The quasi-particle band gap is shown by the red arrow. From [28].

Coulomb interaction. This is the Wannier model for excitons, which predicts Rydberg series of bound electron-hole states for three-dimensional solids. It works surprisingly well for not too strongly bound excitons, and can yield reasonable results even for binding energies in the eV range. For even stronger bound excitons, where the onsite interaction dominates, the Frenkel model is more appropriate; it can also be derived from the GW-BSE (see [19]).

As an example for bound excitons, Fig. 5 shows $q^2 \text{Im } \epsilon_M(\mathbf{q}, \omega)$ for graphane (a hydrogenated graphene sheet, left panel) and one layer of hexagonal boron nitride (right panel). In both cases, important structures are found within the quasi-particle band gap, which is indicated by the red arrow. Binding energies are larger than one eV in both cases. The different nature of the excitons in these two materials can be inferred from their dispersion: the bound exciton in graphane changes its position in a parabolic way, whereas in h-BN the dispersion is rather linear, after a first, more rapid, rise. This is discussed in [28].

7 Excitons and correlation

In the context of these lecture notes, it is interesting to comment about various aspects of correlation concerning excitons. Here we would like to concentrate on two points:

- Excitons are strongly correlated electron-hole pairs.
- The BSE contains cancellation effects between self-energy corrections and electron-hole interaction. These can be particularly important in correlated materials with localized electrons.

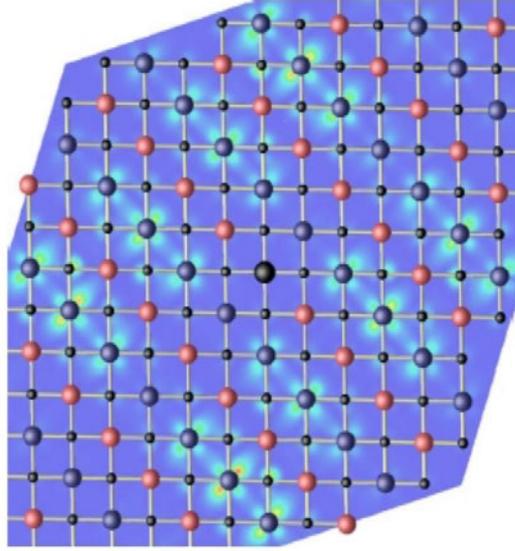


Fig. 6: Excitonic effects in MnO, from [29]. Distribution of spin-up density in the (001) plane for an excitation at 5.2 eV. The position of the hole is indicated by the large black ball in the center. Light red balls represent Mn atoms with occupied 3d spin-up orbitals and dark blue balls are Mn atoms with occupied 3d spin-down orbitals. Oxygen atoms are indicated by small black balls.

To illustrate the first point, it is enough to look at the exciton wavefunction. It is a superposition of products of the wavefunctions of the single electrons and holes, determined by the coefficients A_λ

$$\Psi_\lambda(\mathbf{r}_e, \mathbf{r}_h) = \sum_{v\mathbf{k}c} A_\lambda^{v\mathbf{k}c} \psi_{v\mathbf{k}}^*(\mathbf{r}_e) \psi_{c\mathbf{k}}(\mathbf{r}_h). \quad (60)$$

This is a correlated two-particle wavefunction: in order to know the probability distribution of the electron, one has to fix the position of the hole, and vice versa.

Let us look at such a wavefunction, for the example of the magnetic material MnO [29]. Its antiferromagnetic ordering consists of alternating planes of occupied spin-up and spin-down Mn 3d orbitals. In order to visualize the charge density of an excited electron, one has to choose the excitation energy, and the position of the hole. In Fig. 6, which is taken from [29], the excitation energy corresponds to a peak in the absorption spectrum at 5.2 eV. For the hole position one usually chooses an occupied orbital; here the hole is fixed on an Mn atom where the spin-up orbital is occupied. Because of the dipole transition rules, a spin-up electron is excited from an occupied to an empty spin-up orbital. This determines the distribution that is observed in Fig. 6.

Surprisingly, the picture in Fig. 6 breaks translational invariance, since the exciton extends over several unit cells. How is this possible? The reason is that we had to fix the position of the hole, because of the electron-hole correlation. However, the probability to find the hole in a given unit cell is periodic, so overall translational invariance is respected.

Let us now come to the second point anticipated above, cancellations. To get an idea, think of a single electron. Even if you excite it, there should be no interaction effects – there is no other electron to interact with. However, the BSE starts with L_0 , in other words, with a difference of electron addition and removal. In the addition process the system gets two electrons: now there *is* interaction! This spurious effect has to be removed by the electron-hole interaction. In a solid with delocalized electrons some cancellation is seen (look for example at silicon in the left panel of Fig. 1), but it is far from complete; otherwise, we would never need the BSE. However, when electrons are localized, and especially at low density, in a certain sense one comes closer to the regime of single electrons, and cancellations are more important. Therefore in these materials sometimes the RPA evaluated with Kohn-Sham wavefunctions gives surprisingly decent excitation spectra, for example in transition metal oxides like V_2O_3 [30].

There are many more aspects of optical or loss spectra in the BSE that one might want to address; more can be found for example in [19]. However, one should not forget that the BSE yields in principle the full two-particle correlation function, and more information can be gained from it. Applications such as the calculation of correlated two-hole states [31, 32], or total energies [33], promise an increasingly broad horizon for people interested in the Bethe-Salpeter equation.

References

- [1] B. A. Lippmann and J. Schwinger, *Phys. Rev.* **79**, 469 (1950)
- [2] F. Bechstedt: *Many-Body Approach to Electronic Excitations, Concepts and Applications* Springer Series in Solid-State Sciences, Vol. 181 (Springer, Heidelberg, 2015)
- [3] A. L. Fetter and J. D. Walecka: *Quantum Theory of Many-particle Systems* (McGraw-Hill, New York, 1971)
- [4] P. Lautenschlager, M. Garriga, L. Vina, and M. Cardona, *Phys. Rev. B* **36**, 4821 (1987)
- [5] V. Olevano and L. Reining, *Phys. Rev. Lett.* **86**, 5962 (2001)
- [6] D. Bohm and D. Pines, *Phys. Rev.* **82**, 625 (1951)
- [7] D. Pines and D. Bohm, *Phys. Rev.* **85**, 338 (1952)
- [8] D. Bohm and D. Pines, *Phys. Rev.* **92**, 609 (1953)
- [9] G. Giuliani and G. Vignale: *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, U.K., 2005)
- [10] V. Ambegaokar and W. Kohn, *Phys. Rev.* **117**, 423 (1960)
- [11] S.L. Adler, *Phys. Rev.* **126**, 413 (1962)
- [12] N. Wiser, *Phys. Rev.* **129**, 62 (1963)
- [13] F. Bruneval, F. Sottile, V. Olevano, and L. Reining, *J. Chem. Phys.* **124** (2006)
- [14] J. Paier, M. Marsman, and G. Kresse, *Phys. Rev. B* **78**, 121201 (2008)
- [15] E.E. Salpeter and H.A. Bethe, *Phys. Rev.* **84**, 1232 (1951)
- [16] R. van Leeuwen, N.E. Dahlen, and A. Stan, *Phys. Rev. B* **74**, 195105 (2006)
- [17] G. Strinati, *Rivista del Nuovo Cimento* **11**, 1 (1988)
- [18] P. Romaniello, D. Sangalli, J.A. Berger, F. Sottile, L.G. Molinari, L. Reining, and G. Onida, *J. Chem. Phys.* **130** (2009)
- [19] R.M. Martin, L. Reining, and D.M. Ceperley: *Interacting Electrons: Theory and Computational Approaches* (Cambridge University Press, 2016)
- [20] A. Schindlmayr and R.W. Godby, *Phys. Rev. Lett.* **80**, 1702 (1998)
- [21] R. Del Sole and R. Girlanda, *Phys. Rev. B* **54**, 14376 (1996)
- [22] F. Bechstedt, K. Tenelsen, B. Adolph, and R. Del Sole, *Phys. Rev. Lett.* **78**, 1528 (1997)

-
- [23] M. Gruning, A. Marini, and X. Gonze, *Nano Lett.* **9**, 2820 (2009)
- [24] I. Tamm, *J. Phys. (USSR)* **9**, 449 (1945)
- [25] S.M. Dancoff, *Phys. Rev.* **78**, 382 (1950)
- [26] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, *Phys. Rev. Lett.* **80**, 4510 (1998)
- [27] G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002)
- [28] P. Cudazzo, L. Sponza, C. Giorgetti, L. Reining, F. Sottile, and M. Gatti, *Phys. Rev. Lett.* **116**, 066803 (2016)
- [29] C. Rödl, F. Fuchs, J. Furthmüller, and F. Bechstedt, *Phys. Rev. B* **77**, 184408 (2008)
- [30] F. Iori, F. Rodolakis, M. Gatti, L. Reining, M. Upton, Y. Shvyd'ko, J.-P. Rueff, and M. Marsi, *Phys. Rev. B* **86**, 205132 (2012)
- [31] Y. Noguchi, S. Ishii, K. Ohno, I. Solovyev, and T. Sasaki, *Phys. Rev. B* **77**, 035132 (2008)
- [32] Y. Noguchi, S. Ishii, and K. Ohno, *J. Chem. Phys.* **125** (2006)
- [33] E. Maggio and G. Kresse, *Phys. Rev. B* **93**, 235113 (2016)