Excitons in time-dependent density-functional theory

Carsten A. Ullrich and Zeng-hui Yang

Abstract This article gives an overview of the description of the optical and dielectric properties of bulk insulators and semiconductors in time-dependent density-functional theory (TDDFT), with an emphasis on excitons. We review the linear-response formalism for periodic solids, discuss excitonic exchange-correlation kernels, calculate exciton binding energies for various materials, and compare the treatment of excitons with TDDFT and with the Bethe-Salpeter equation.

Keywords Time-dependent density-functional theory; Excitons; Dielectric function; Exchange-correlation kernel; Bethe-Salpeter equation

1 Introduction

Time-dependent density-functional theory (TDDFT) is a universal approach to the dynamical many-body problem. A detailed, up-to-date coverage of TDDFT can be found in two books [1, 2]. An easy and concise introduction is given in a recent review article by Ullrich and Yang [3].

At present, the majority of applications of TDDFT take place in the field of computational (bio)chemistry, to obtain excitation energies and excited-state properties of molecules. However, applications in condensed-matter physics and materials science are emerging at a rapid rate. In this article we will give an introduction and overview of TDDFT for extended periodic systems, focusing on the optical properties of semiconducting and insulating systems. In particular, we will address the question how TDDFT can be used to calculate excitonic binding energies and optical spectra with excitonic features. The present state of the art approach in this field is given by Green’s function based many-body techniques, most notably, the
Carsten A. Ullrich and Zeng-hui Yang

combination of GW [4, 5] and the Bethe-Salpeter equation (BSE) [6]–[13]. We will compare and contrast this approach with TDDFT and discuss their performance and the various pros and cons for bulk semiconductors and insulators.

TDDFT for periodic solids was reviewed a few years ago by Onida et al. and Botti et al. [14, 15]. Since then, many new developments have occurred, and in this article we will attempt to cover the more recent progress in this field, including our own recent work [16]–[19]. We shall use atomic units ($\hbar = m = e = 4\pi\epsilon_0 = 1$) unless otherwise indicated.

2 What is an exciton?

The optical properties of materials are determined by the way in which the electrons and the ions respond to light. In this article we shall focus exclusively on the electronic response and ignore the lattice dynamics or any effects related to the coupling of electronic and lattice excitations (such as polarons; for details see, e.g., [20]).

The response of a system of $N$ electrons is often characterized as having either “single-particle” or “collective” character. What do we mean by this? If a system consists of noninteracting particles, the response is always purely single-particle, or can be viewed as the sum of many individual, uncorrelated single-particle excitations. In the presence of interactions, this simple picture is no longer valid, because any change of the state of one electron will have an immediate influence on all other electrons in the system; the response is collective and involves, in principle, all electrons. The question is how dominant these effects are.

There are many situations where the interactions give rise to rather straightforward behavior. Imagine a thought experiment where we can turn off the interactions between the electrons in a molecule, and measure the resulting excitation spectrum. If we now gradually turn the interactions back on, all the while keeping an eye on the excitation spectrum, we find that the peaks in the spectrum will shift, but we can keep track of each of them and hence can in principle interpret them as single-particle excitations of an effective system such as Kohn-Sham.

On the other hand, interacting systems have certain excitations that do not have a counterpart in any corresponding noninteracting system. A most drastic example are plasmons in a metal, where all electrons respond collectively and with a fixed phase relationship. A plasmon requires dynamical electron-electron interactions at least at the level of the random-phase approximation (RPA).

An exciton is another example of a collective excitation, which occurs in nonmetallic systems. The ideal exciton can be described [21] as an electrically neutral quantum of electronic excitation energy travelling in the periodic structure of a crystal. It can be viewed as a bound electron-hole pair and can hence be associated with the transportation of energy, but not of net charge. Excitons are a crucial stage in the photovoltaic process, where free carriers are generated after separation of the electron-hole pairs.

Excitons come in different flavors [22]:
• Frenkel excitons [23, 24] are excitations that are localized at the atomic sites of wide-gap insulators such as solid rare gases (neon, argon) or certain ionic solids (e.g., LiF).

• Davydov excitons [25] refer to excitons in molecular crystals with ring units, such as benzene and anthracene. Since the excitations remain localized on the individual molecules, Davydov excitons can be viewed as a subclass of Frenkel excitons.

• Mott-Wannier excitons [26, 27] are the excitons that typically occur in semiconductors such as GaAs, CdSe, or Cu$_2$O. They tend to be delocalized over several atomic unit cells.

The concept of excitons was originally introduced in bulk crystals, but they also exist in many lower-dimensional systems such as surfaces, quantum wells, quantum wires, nanotubes, polymers, nanocrystals, and quantum dots [28]–[33]. In this article we shall limit ourselves to three-dimensional periodic crystals.

Excitons are usually described as bound electron-hole pairs, i.e., as an effective two-particle system. Within the effective-mass approximation [34, 35], where conduction band electrons have effective mass $m_e$ and valence band holes have effective mass $m_h$, we can define a reduced effective mass $m^r = m_e m_h / (m_e + m_h)$. Next, we separate center-of-mass and relative degrees of freedom. The former describes how the exciton travels through the crystal, and the latter determines the exciton binding energy according to the following hydrogen-like Schrödinger equation:

$$\left\{ -\frac{\hbar^2 \nabla^2}{2m^r} - \frac{e^2}{4\pi\varepsilon_0\varepsilon r} \right\} \psi_j (\mathbf{r}) = E_j \psi_j (\mathbf{r}).$$

(1)

Here, $\varepsilon_0$ is the vacuum permittivity, $e$ is the free electron charge, and $\varepsilon$ is the dielectric constant of the material (we shall talk about the dielectric constant in much detail in the following section). Equation (1) is also known as the Wannier equation. It yields a Rydberg series of bound states as well as a continuum of unbound states [36]. The lowest (1s) excitonic state determines the exciton binding energy $E_0^{\text{ex}}$ and the exciton bohr radius $a_0^{\text{ex}}$. In GaAs, a material in which the Wannier equation works particularly well, one obtains $E_0^{\text{ex}} = 4.6$ meV and $a_0^{\text{ex}} = 118$ Å, which clearly shows that Wannier excitons are weakly bound and extend over many lattice constants [which, a posteriori, justifies the simplified treatment via Eq. (1)].

The Wannier picture of excitons as bound electron-hole pairs, described by Eq. (1), is generally not quantitatively accurate, and breaks down completely if the exciton radius becomes comparable to a lattice constant. In this article, we shall present an \textit{ab initio} approach, based on TDDFT and/or other many-body techniques, which is universally valid and in principle exact. It will turn out that this approach reveals an alternative point of view, in which excitons are described as collective excitations of the many-electron system. This picture is schematically illustrated in Fig. 1.

The left panel of Fig. 1 shows a single-particle transition in a simple model of an insulator, going vertically from the filled valence band to the empty conduction band. The energy associated with this transition is just the difference between the levels in the two bands. By contrast, an exciton arises from a superposition of many
Optical transitions in a two-band insulator, where the lower (valence) band is filled and the upper (conduction) band is empty. The vertical direction is energy, the horizontal direction is wavevector. (a) A single-particle transition, in which only one single-particle state gets excited and all other states do not participate. The associated excitation energy is the difference of the initial and final single-particle states. (b) Excitonic transition, which is a collective excitation in which many states participate. The thickness of the arrows indicates that transitions close to the band gap are dominant. The associated excitation energy can be lower than the band gap, since the collective nature of the response is energetically favorable.

single-particle transitions, as illustrated in the right panel of Fig. 1. Not all transitions contribute with an equal weight, as indicated by the different thickness of the arrows, but all of them have a fixed phase relationship; hence, the exciton is a collective excitation. The energy of the exciton (i.e., the energy of this collective excitation) is lower than the lowest single-particle transition. This happens because the collective behavior induced by the dynamical many-body effects is energetically favorable compared to any single-particle transition. We will see later on how the two viewpoints of the nature of an exciton can be reconciled with each other [18].

3 A tale of three gaps

The defining characteristic of insulators and semiconductors is that they have an electronic band gap (we only consider materials at zero temperature), which dominates their optical and transport properties. Before we deal with the optical response of solids, it is crucial to have a clear understanding and a good description of the gap. However, it turns out that there are in fact three different kinds of gap (for the nonmagnetic materials we are interested in), and it is important to distinguish them carefully [37].

The fundamental band gap $E_g$ of an $N$-electron system is defined as follows:

$$E_g(N) = I(N) - A(N),$$

where $I(N)$ and $A(N)$ are the ionization potential and the electron affinity of the system. These two quantities can be obtained in a straightforward manner from
ground-state DFT: the ionization potential is formally exactly given by the highest occupied Kohn-Sham eigenvalue of the $N$-electron system, $\epsilon_N(N)$, and the electron affinity is the corresponding quantity of the $N+1$-electron system. Hence, we obtain

$$E_g(N) = \epsilon_{N+1}(N+1) - \epsilon_N(N).$$

(3)

It is important to note that the right-hand side of Eq. (3) contains the highest occupied Kohn–Sham eigenvalues of two different systems, namely with $N$ and with $N+1$ electrons. In a macroscopic solid with $10^{23}$ electrons, it would be impossible (or, at least, highly impractical) to calculate the band gap according to this definition.

The band gap in the noninteracting Kohn–Sham system, also known as the Kohn–Sham gap, is defined as

$$E_{g,s}(N) = \epsilon_{N+1}(N) - \epsilon_N(N).$$

(4)

In contrast with the interacting gap $E_g$, the Kohn–Sham gap $E_{g,s}$ is simply the difference between the highest occupied and lowest unoccupied single-particle levels in the same $N$-particle system. This quantity is what is usually taken as the band gap in standard DFT band-structure calculations. We can relate the two gaps by

$$E_g = E_{g,s} + \Delta_{xc},$$

(5)

which defines $\Delta_{xc}$ as a many-body correction to the Kohn–Sham gap. By making use of the previous relations, we find $\Delta_{xc} = \epsilon_{N+1}(N+1) - \epsilon_{N+1}(N)$. It turns out that the many-body gap correction $\Delta_{xc}$ can be related to a very fundamental property of density functionals, known as derivative discontinuities [38]–[42].

The so-called band-gap problem of DFT reflects the fact that in practice $E_{g,s}$ is often a poor approximation to $E_g$, typically underestimating the exact band gap by as much as 50%. The reason for this is twofold: commonly used approximate xc functionals (such as LDA and GGA) tend to underestimate the exact Kohn–Sham gap $E_{g,s}$, and they do not yield any discontinuity correction $\Delta_{xc}$. An extreme example for the second failure are Mott insulators, which are typically predicted to be metallic by DFT. This is no accident: in Mott insulators, the exact Kohn–Sham system is metallic (i.e., $E_{g,s} = 0$) so that $E_g = \Delta_{xc}$. Clearly, standard xc functionals (where $\Delta_{xc}$ vanishes) are unfit to describe Mott insulators.

It is important to distinguish between the fundamental band gap and the optical gap [43]. The band gap describes the energy that an electron must have so that, when it is added to an $N$-electron system, the result is an $N+1$ electron system in its ground state. The total charge of the system changes by $-1$ in this process. By contrast, the optical gap describes the lowest neutral excitation of an $N$-electron system: here, the number of electrons remains unchanged. The two gaps are schematically illustrated in Fig. 2 together with the Kohn–Sham gap.

The band gap of insulators can be accurately obtained from the so-called quasiparticle energies, which are defined as the single-particle energies of a noninteracting system whose one-particle Green’s function is the same as that of the real interacting system (notice that this effective noninteracting system is very different
Fig. 2 Schematic illustration of the different types of gaps in DFT and TDDFT. The Kohn-Sham gap is defined as the difference of the highest occupied and lowest unoccupied Kohn-Sham eigenvalues of the $N$-electron system, see Eq. (4). The fundamental band gap [or quasiparticle (QP) gap] is the Kohn-Sham gap plus the derivative discontinuity, see Eq. (5). The optical gap is the band gap minus the lowest exciton binding energy $E_{0}^{ex}$. The Kohn-Sham gap can be viewed as an approximation for the optical gap.

from the Kohn-Sham system, which is defined as that noninteracting system which reproduces the exact density). In practice, quasiparticle calculations are often done using the GW method [4, 5, 14]. GW calculations are more demanding than DFT, but they produce band structures of solids that agree very well with experiment.

In recent years, generalized Kohn-Sham schemes [37, 44, 45] have become quite popular for calculating band gaps of solids. Generalized Kohn-Sham theory means, in essence, using exchange-correlation (xc) functionals which contain an admixture of Hartree-Fock (HF) nonlocal exchange. These functionals (for instance, B3LYP [46] or PBE0 [47]) have been crucial for the enormous success of DFT in theoretical chemistry. Hybrid xc functionals for solids are not unproblematic: HF exchange is difficult to implement in periodic systems, and B3LYP fails for metals [48]. But for insulators, hybrid functionals generally produce excellent results, comparable to what can be achieved by more sophisticated many-body calculations [49]–[59]. Another promising approach are meta-GGA functionals such as the Tran-Blaha exchange potential [60]–[64], which gives good band structures and band gaps and is much less costly than hybrid functionals. However, it has the drawback of not being derivable from an energy functional.

It is to be emphasized that current implementations of hybrid xc functionals do not actually calculate the Kohn-Sham gap (this would require a self-consistent calculation with a local potential); instead, hybrid functionals yield an approximation to the quasiparticle gap. Local xc potentials can, in principle, be constructed from hybrid functionals using the OEP (optimized effective potential) method, which was successfully done for the case of exact exchange (see [65] and references therein).

While the band gap can be measured using techniques in which electrons are added or removed from the system (such as photoemission spectroscopy), the opti-
Excitons in TDDFT

cal gap refers to the lowest neutral excitation. The difference between quasiparticle band gap and optical gap is the lowest exciton binding energy, $E_{\text{ex}}^0$. In the previous section we have seen that excitons can be viewed as bound electron-hole pairs, whose bound states form a Rydberg series, analogous to the hydrogen atom. The band gap is given by the asymptotic limit of the excitonic Rydberg series \[66\] (at least for direct-gap insulators and semiconductors).

4 Linear response and optical properties in periodic solids

4.1 The microscopic and macroscopic dielectric functions

The interactions of electromagnetic fields and matter are governed by Maxwell’s equations,

$$\nabla \cdot D = n_f,$$

$$\nabla \times E = -\frac{\partial B}{\partial t},$$

$$\nabla \cdot B = 0,$$

$$\nabla \times H = j_f + \frac{\partial D}{\partial t},$$

where all fields ($D$, $E$, $B$, and $H$) and all source terms (the density of free charges $n_f$ and the free current density $j_f$) are functions of position $r$ and time $t$. We consider situations where all time dependence is periodic, and we Fourier transform from time $t$ to frequency $\omega$. Of main interest to us (since we will be concerned with nonmagnetic materials) is the relationship between the electric displacement $D$ and the total electric field $E$:

$$D(r, \omega) = \int d^3r' \epsilon(r, r', \omega) E(r', \omega),$$

where $\epsilon(r, r', \omega)$ is the nonlocal, frequency-dependent dielectric tensor. In lattice-periodic systems, translational symmetry implies $\epsilon(r, r', \omega) = \epsilon(r+R, r'+R, \omega)$, where $R$ is a lattice vector. We can then Fourier analyze $\epsilon(r, r', \omega)$ and obtain

$$\epsilon(r, r', \omega) = \frac{1}{V} \sum_{k \in \text{BZ}, G, G'} e^{-i(k+G) \cdot r} e^{i(k+G') \cdot r'} \epsilon(k + G, k + G', \omega),$$

where $V$ is the crystal volume, $k$ is a wave vector in the first Brillouin zone (BZ), and $G$ and $G'$ are reciprocal lattice vectors. In the following we shall use the notation

$$\epsilon_{GG'}(k, \omega) = \epsilon(k + G, k + G', \omega).$$

Using these definitions, we can recast Eq. (10) into
\[ D_G(k, \omega) = \sum_{G'} \varepsilon_{GG'}(k, \omega) E_{G'}(k, \omega). \]  

(13)

For comparison with experiment, one is usually interested in macroscopic quantities, i.e., quantities which are defined as averages over the unit cell of the crystal. For instance, the macroscopic limit of Eq. (13) is defined as

\[ D_{\text{mac}}(\omega) = \varepsilon_{\text{mac}}(\omega) E_{\text{mac}}(\omega). \]  

(14)

An important observation from Eq. (13) is that the microscopic \( \varepsilon_{GG'}(k, \omega) \) is in general nondiagonal in \( G \) and \( G' \), for inhomogeneous systems. Therefore, even a uniform external field will induce nonuniform microscopic fluctuations in the solid; these are called local-field effects. As a consequence, the macroscopic \( \varepsilon_{\text{mac}}(\omega) \) cannot be calculated directly; instead, one must take a detour via microscopic linear-response theory. Otherwise, local-field effects would not be properly included.

In the following, we shall restrict the discussion to crystals with cubic symmetry, since this leads to the considerable simplification that the dielectric tensor becomes isotropic in the long-wavelength limit and one can carry out a decomposition into longitudinal and transverse components similar to that for the homogeneous case. It can then be shown that the macroscopic dielectric constant is given by \([67]–[69]\]

\[ \varepsilon_{\text{mac}}(\omega) = \lim_{k \to 0} \left[ \varepsilon_{GG'}(k, \omega)^{-1} \right]^{-1}, \]  

(15)

where \( \varepsilon_{GG'}(k, \omega) \), the longitudinal component of the dielectric tensor for the cubic system, is often called the dielectric matrix. Calculating the macroscopic dielectric tensor without imposing cubic symmetry is technically more involved [70].

In the optical spectroscopy of materials, a central quantity is the complex refractive index \( \tilde{n} \), defined as [20]

\[ \varepsilon_{\text{mac}}(\omega) = \tilde{n}^2. \]  

(16)

The real and imaginary parts of \( \tilde{n} \) determine two key optical properties of materials: the refractive index \( n \) and the extinction coefficient \( \kappa \), where

\[ \Re \varepsilon_{\text{mac}} = n^2 + \kappa^2, \]  

(17)

\[ \Im \varepsilon_{\text{mac}} = 2n\kappa. \]  

(18)

The extinction coefficient \( \kappa \) is proportional to the optical absorption coefficient; therefore, optical absorption spectra are essentially determined by \( \Im \varepsilon_{\text{mac}}(\omega) \).

4.2 Linear-response theory and TDDFT

We now make a connection between the dielectric function and the linear-response formalism. The linear density response \( n_1(r, \omega) \) caused by a frequency-dependent
Excitons in TDDFT

scalar perturbation $v_1(r, \omega)$ is given by

$$n_1(r, \omega) = \int d^3r' \chi(r, r', \omega)v_1(r', \omega),$$

where $\chi(r, r', \omega)$ is the density-density response function of the interacting many-body system. In analogy with Eq. (10), the scalar dielectric function can be introduced as follows [71]:

$$v_1(r, \omega) = \int d^3r' \epsilon(r, r', \omega) \left[ v_1(r', \omega) + \int d^3r'' n_1(r'', \omega) \right].$$

Combining Eqs. (19) and (20), we obtain the inverse dielectric function as

$$\epsilon^{-1}(r, r', \omega) = \delta(r - r') + \int d^3r'' \chi(r'', r', \omega) \frac{n_1(r'', \omega)}{|r'' - r'|},$$

and for a periodic system we have

$$\epsilon_{GG'}^{-1}(\mathbf{k}, \omega) = \delta_{GG'} + v_G(\mathbf{k})\chi_{GG'}(\mathbf{k}, \omega),$$

where the Fourier transform of the 3D Coulomb potential is given by

$$v_G(\mathbf{k}) = \frac{4\pi}{|\mathbf{k} + \mathbf{G}|^2}.$$

Thus, the inverse dielectric function follows directly from the response function.

In linear-response TDDFT [72], the interacting response function $\chi$ can be expressed in terms of the response function of the Kohn–Sham system $\chi_s$ and the Hartree and xc kernels:

$$\chi(r, r', \omega) = \chi_s(r, r', \omega)$$

$$+ \int d^3x \int d^3x' \chi_s(r, x, \omega) \left\{ \frac{1}{|x - x'|} + f_{xc}(x, x', \omega) \right\} \chi(x', r', \omega).$$

Here, $\chi_s$ is the response function of the noninteracting Kohn-Sham system, given by

$$\chi_s(r, r', \omega) = \sum_{j,k=1}^{\infty} (f_k - f_j) \frac{\varphi_j(r)\varphi_k^*(r)\varphi_j^*(r')\varphi_k(r')}{\omega - \omega_{jk} + i\eta},$$

where $f_j$ and $f_k$ are occupation numbers referring to the configuration of the Kohn-Sham ground state (1 for occupied and 0 for empty Kohn-Sham orbitals), $\varphi_j(r)$ are the Kohn-Sham orbitals, and the $\omega_{jk}$ are defined as the differences of the Kohn-Sham eigenvalues,

$$\omega_{jk} = \varepsilon_j - \varepsilon_k.$$

The key quantity in linear-response TDDFT is the xc kernel, defined as the functional derivative of the time-dependent xc potential with respect to the time-depen-
dent density, evaluated at the ground-state density:

\[ f_{xc}(r, r', t, t') = \left. \frac{\delta v_{xc}[n](r, t)}{\delta n(r', t')} \right|_{n_0(r)}. \]  

(27)

The frequency-dependent xc kernel, \( f_{xc}(r, r', \omega) \), is the Fourier transform of this with respect to \((t - t')\).

In lattice-periodic systems, Eq. (24) can be cast into the following form:

\[ \chi_{GG'}(k, \omega) = \chi_{G0}(k, \omega) + \sum_{G_1, G_2} \chi_{G_1 G_2}(k, \omega) \left\{ v_{G_1}(k) \delta_{G_1 G_2} + f_{xc} G_1 G_2(k, \omega) \right\} \chi_{G_2 G'}(k, \omega), \]  

(28)

where the Kohn–Sham response function (25) is transformed into

\[ \chi_{G0}(k, \omega) = \frac{1}{V} \sum_\kappa \sum_{j=1}^m \frac{f_{j+k'} - f_{j+k}}{\omega + \epsilon_{j+k'} - \epsilon_{j+k} + i\eta} \times \int d^3r \varphi_{jk}^*(r)e^{-i(k+G') \cdot r} \int d^3r' \varphi_{jk}^*(r')\varphi_{jk}^*(r'), \]  

(29)

featuring the Kohn–Sham band structure \( \epsilon_{jk} \) and Bloch functions \( \varphi_{jk}(r) \). The so-called head \((G = G' = 0)\) of the xc kernel \( f_{xc} G0(k, \omega) \) gives the largest contribution to the change from \( \chi \) to \( \chi' \); the contributions from bigger \( G \)'s decay rapidly. Thus the sums in Eq. (28) can usually be restricted to a small number of reciprocal lattice vectors, which reduces the computational effort significantly.

Let us now come back to the macroscopic dielectric constant. It can be shown \([1, 14, 15]\) that \( \epsilon_{mac}(\omega) \) takes on the following form:

\[ \epsilon_{mac}(\omega) = 1 - \lim_{k \to 0} v_0(k) \bar{\chi}_{G0}(k, \omega). \]  

(30)

Here, \( \bar{\chi}_{G0}(k, \omega) \) differs from the full response function \( \chi_{G0}(k, \omega) \), as defined in Eq. (28), in the following way: instead of using the full Coulomb interaction \( v_G(k) \) [see Eq. (23)], it uses the modified Coulomb interaction

\[ \bar{v}_G(k) = \begin{cases} 
0 & \text{for } G = 0, \\
\frac{4\pi}{|k + G|^2} & \text{for } G \neq 0,
\end{cases} \]  

(31)

in which the long-range part \( v_0(k) = 4\pi/k^2 \) has been left out. This seemingly small modification turns out to be quite important.
5 TDDFT for excitons in solids

5.1 Why are excitons a difficult problem?

Up until a few years ago, common wisdom held that TDDFT may be good for molecular excitations, but it fails for excitons in solids. We now know better, as we shall see in this Section. But let us first discuss why excitons are such a hard problem for TDDFT.

Figure 3 shows the imaginary part of the macroscopic dielectric function of bulk silicon [73], comparing experimental data with calculations using the RPA (where $f_{xc} = 0$) and the adiabatic local-density approximation (ALDA), which is the simplest and most commonly used approximation of TDDFT. There are drastic deviations between theory and experiment. First of all, the onset of absorption is redshifted in RPA and ALDA by about half an eV; this is not very surprising, and reflects the “band gap problem” of ground-state DFT: as we discussed in Section 3, the Kohn-Sham gap of standard local and semilocal xc functionals is smaller than the quasiparticle gap. One can correct for this error and shift the empty bands via a scissors operator [74, 75] or one can use other methods to obtain band structures with a better gap such as GW or hybrid functionals (see Section 3).

The second deviation is more problematic: both RPA and ALDA lack the first excitonic peak (labeled $E_1$ in the experimental data), and instead only have a weak shoulder. This discrepancy persists even if a better band structure (such as GW) is used as input to calculate the noninteracting response function $\chi_0$ [14]. This failure of the ALDA as well as the GGA xc functionals is by no means unique to silicon, but
systematically occurs in bulk insulators and semiconductors. The underlying reason, as we shall see, is the long-range behavior of the xc kernel: in reciprocal space, the head of \( f_{xc} \) should diverge as \( 1/q^2 \) for small \( q \), but semilocal approximations instead approach a constant. In the next subsections we explain this in more detail.

The main reason why excitons in TDDFT are a hard problem is thus that the standard local and semilocal xc kernels don’t work, and one needs to resort to non-standard kernels or even develop new ones. We will discuss this in Section 5.4 and show some results.

Another reason is of a more practical nature. Even if very good approximations for the xc kernel are available, the calculations can be numerically difficult since convergence in reciprocal space can be slow. This problem affects the standard many-body approaches (GW-BSE) as well [76, 77].

### 5.2 Formalism: direct calculation of exciton binding energies

The Kohn-Sham response function \( \chi \) has poles at the single-particle Kohn-Sham excitation energies, which can be clearly seen in Eqs. (25) and (29). On the other hand, the full many-body response function \( \chi \) has poles at the exact excitation energies of the system [1, 71]. This is true for any type of system, finite or extended. The Hartree and xc kernels in Eqs. (24) and (28) are responsible for transforming the Kohn-Sham excitation spectrum into the exact one; this includes the creation of excitations that have no counterpart in the Kohn-Sham spectrum, such as plasmons in metals or excitons in semiconductors and insulators.

It is convenient to describe electronic excitations as electronic eigenmodes of the system. The associated mode frequencies—the excitation energies of the system—are then obtained via the so-called Casida equation [78]:

\[
\begin{pmatrix}
A & B \\
B & A
\end{pmatrix}
\begin{pmatrix}
X \\
Y
\end{pmatrix}
=
\begin{pmatrix}
-1 & 0 \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
X \\
Y
\end{pmatrix},
\]

(32)

where the elements of the matrices \( A \) and \( B \) are

\[
A_{ia, jb'}(\omega) = (\varepsilon_i - \varepsilon_j)\delta_{ab} \delta_{\sigma\sigma'} + K_{iT, jb'}(\omega),
\]

(33)

\[
B_{ia, jb'}(\omega) = K_{iT, jb'}(\omega),
\]

(34)

with the Hartree-exchange-correlation (Hxc) matrix elements

\[
K_{iT, jb'}(\omega) = \int d^3r \int d^3r' \phi_{ia}(\mathbf{r})\phi_{ia}(\mathbf{r})\phi_{j\sigma'}(\mathbf{r})f_{Hxc}(\mathbf{r}, \mathbf{r'}, \omega)\phi_{j\sigma'}(\mathbf{r})\phi_{b'}(\mathbf{r'}). \quad (35)
\]

The indices \( i, j \) and \( a, b \) run over occupied and unoccupied Kohn-Sham orbitals, respectively. \( f_{Hxc} \) denotes the sum of the Hartree and xc kernels.

Most of the currently available xc kernels are frequency independent, in which case Eq. (32) becomes a (pseudo-)eigenvalue problem. The excitation frequencies
Excitons in TDDFT

of the system are explicitly given by the eigenvalues \( \Omega \). The eigenvector \( X \) together with \( Y \) describes how the Kohn-Sham excitations combine to form the excitation in the real system. The optical spectrum can be calculated with \( X \) and \( Y \).

The widely used Tamm-Dancoff approximation (TDA) sets the matrix \( B \) to zero and hence neglects the correlation between excitations and de-exitations. Within the TDA and using the adiabatic approximation for the \( xc \) kernel, Eq. (32) becomes

\[
\sum_{j \neq a'} \left[ \delta_{ij} \delta_{a'b} \delta_{\sigma' \sigma} (\varepsilon_{a\sigma} - \varepsilon_{i\sigma}) + K^{Hxc}_{ia'sj\sigma} \right] X_{j'\sigma'} = \Omega X_{i\sigma}. \tag{36}
\]

The real space representation of the \( Hxc \) kernel is related to the momentum space representation as

\[
f_{Hxc\sigma\sigma'}(\mathbf{r}, \mathbf{r}', \omega) = \frac{1}{V} \sum_{\mathbf{q} \in \text{FBZ}} \sum_{\mathbf{G}, \mathbf{G}'} e^{i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} f_{Hxc\sigma\sigma'}(\mathbf{q}, \mathbf{G}, \mathbf{G}', \omega) e^{-i(\mathbf{q} + \mathbf{G}') \cdot \mathbf{r}'.} \tag{37}
\]

With Eq. (37), the \( Hxc \) kernel in transition space, Eq. (35), becomes

\[
K^{Hxc}_{ia'sj\sigma} = \frac{1}{V} \sum_{\mathbf{q} \in \text{FBZ}} \sum_{\mathbf{G}, \mathbf{G}'} \langle \mathbf{k}_i \sigma | e^{i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} | \mathbf{a} \mathbf{k}_a \sigma \rangle f_{Hxc\sigma\sigma'}(\mathbf{q}, \mathbf{G}, \mathbf{G}')
\times \langle \mathbf{b} \mathbf{k}_j \sigma' | e^{-i(\mathbf{q} + \mathbf{G}')} \mathbf{r}' | \mathbf{j} \mathbf{k}_j \sigma' \rangle \delta_{\mathbf{k}_a - \mathbf{k}_i + \mathbf{G}, \mathbf{G}_0} \delta_{\mathbf{G}_b - \mathbf{k}_j + \mathbf{G}_0}, \tag{38}
\]

with the matrix elements defined as

\[
\langle \mathbf{k} \mathbf{G} | e^{i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} | \mathbf{a} \mathbf{k} \mathbf{G} \rangle \equiv \int d^3 r \varphi_{\mathbf{j} \mathbf{k} \sigma}(\mathbf{r}) e^{i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} \varphi_{\mathbf{a} \mathbf{k} \sigma}(\mathbf{r}), \tag{39}
\]

where the \( \mathbf{k}'s \) are the Bloch wavevectors of the corresponding wavefunctions, and \( \mathbf{G}_0, \mathbf{G}_b \) can be any reciprocal lattice vector. The Kronecker-\( \delta \)'s in Eq. (38) are a consequence of Bloch's theorem.

In the following, we will not consider any spin-dependent excitations (see Ref. [19] for a discussion of triplet excitons within TDDFT). Since we are interested in optical absorption, only vertical single-particle transitions need to be considered, so that \( \mathbf{k}_i = \mathbf{k}_a \) and \( \mathbf{k}_j = \mathbf{k}_b \), which implies \( \mathbf{q} = 0 \) in Eq. (38). Equation (36) then becomes, in reciprocal space,

\[
\sum_{\mathbf{j} \neq \mathbf{k}'} \left[ \delta_{\mathbf{j} \mathbf{k}_i} \delta_{\mathbf{j} \mathbf{k}'} \delta_{\mathbf{k} \mathbf{j} \mathbf{k}'} (\varepsilon_{\mathbf{k} \mathbf{G}} - \varepsilon_{\mathbf{j} \mathbf{G}}) + K^{Hxc}_{\mathbf{i} \mathbf{k}_a, \mathbf{j} \mathbf{k}'} \right] X_{j'\mathbf{k}'} = \Omega X_{i\mathbf{k}}. \tag{40}
\]

Notice that \( i, j \) and \( a, b \) denote occupied and unoccupied band indices, respectively. This includes, in principle, all empty bands including continuum states; in practice, however, only a limited number of valence and conduction bands in the vicinity of the band gap need to be included if Eq. (40) is to be solved numerically.

The Hartree part of the coupling matrix is given by

\[
K^{H}_{\mathbf{i} \mathbf{k}_a, \mathbf{j} \mathbf{k}'} = \frac{2}{V} \sum_{\mathbf{G} \neq 0} \frac{4\pi}{|\mathbf{G}|^2} \langle \mathbf{i} \mathbf{k} | e^{i\mathbf{G} \cdot \mathbf{r}} | \mathbf{a} \mathbf{k} \rangle \langle \mathbf{b} \mathbf{k}' | e^{-i\mathbf{G} \cdot \mathbf{r}} | \mathbf{j} \mathbf{k}' \rangle. \tag{41}
\]
The long-range part \((G = 0)\) of the Coulomb interaction is omitted so that the eigenvalues of Eq. (40) correspond to the poles in the macroscopic dielectric function, as we discussed in the end of Section 4.2. The xc part is given by

\[
K_{\text{xc},j\k}^{\text{sc}} = \frac{2}{V} \lim_{q \to 0} \sum_{GG'} f_{\text{xc},GG'}(q) \langle i\k | e^{i(q+G) \cdot r}| \k \rangle \langle j\k' | e^{-i(q+G') \cdot r}| j\k' \rangle .
\]

The solutions of Eq. (40) can be used to calculate the macroscopic dielectric function, using the following expression [\(l\) labels the \(l\)th eigenvalue of Eq. (40)]:

\[
\varepsilon_{\text{mac}}(\omega) = 1 - \lim_{q \to 0} \frac{4\pi}{q^2} \sum_l \left| \sum_{i\k} \langle i\k | e^{-i\mathbf{q} \cdot \mathbf{r}} | \k \rangle X^{(l)}_{i\k} \right|^2 \frac{\omega - \Omega_l + i\eta}{\omega} .
\]

**5.3 Why does ALDA fail?**

Now let us discuss the behavior of the head, wings, and body of the coupling matrix (42). For \(G = 0\), the matrix element \(\langle i\k | e^{i(q+G) \cdot r}| \k \rangle\) vanishes as \(q\) when \(q \to 0\), and similarly for the other matrix element, \(\langle j\k' | e^{-i(q+G') \cdot r}| j\k' \rangle\). This means that the head \((G = G' = 0)\) of \(K_{\text{xc},j\k}^{\text{sc}}\) will vanish unless the head of \(f_{\text{xc},GG'}(q)\) diverges at least as \(q^{-2}\). Likewise, the wings \((G = 0 \text{ and } G' \text{ finite})\) will vanish unless the wings of the xc kernel diverge at least as \(q^{-1}\).

All local and semilocal xc kernels (ALDA and adiabatic GGAs) remain finite for all \(G, G'\) and \(q\). This is easy to see for the ALDA, whose real-space form is

\[
f_{\text{xc}}^{\text{ALDA}}(r, r') = \frac{d^2 \varepsilon_{\text{xc}}(n)}{dn^2} \bigg|_{n = n_0(r)} \delta(r - r') ,
\]

where \(\varepsilon_{\text{xc}}(n)\) is the xc energy density of a homogeneous electron gas of uniform density \(n\) and \(n_0(r)\) is the ground-state density of the material. Carrying out the Fourier transform we find that the \(q\)-dependence simply drops out [79]:

\[
f_{\text{xc},GG'}^{\text{ALDA}}(q) = \frac{1}{V_{\text{cell}}} \int d^3r e^{-i(G-G') \cdot r} \frac{d^2 \varepsilon_{\text{xc}}(n)}{dn^2} \bigg|_{n = n_0(r)} ,
\]

where the integral runs over one unit cell with volume \(V_{\text{cell}}\). The adiabatic GGA xc kernels exhibit a similar behavior. If \(f_{\text{xc},GG'}^{\text{ALDA}}(q)\) is substituted into Eq. (42), then the contribution from the head and wings of \(f_{\text{xc}}\) to \(K_{\text{xc},j\k}^{\text{sc}}\) will vanish. For ALDA and GGA kernels, all changes to the Kohn-Sham spectrum can thus only come from the **body** of \(K_{\text{xc},j\k}^{\text{sc}}\) (where both \(G \neq 0\) and \(G' \neq 0\)), but these are not sufficiently strong to produce excitons.

The case \(q \to 0\) in reciprocal space corresponds to \(r \to \infty\) in real space. The long-range behavior of the xc kernel is relatively unimportant for low-lying excitations in finite systems such as atoms and molecules, which means that local and semilocal xc
kernels will work reasonably well (an exception to this statement are charge-transfer excitations [80]–[86]). However, for extended and periodic systems it is crucial to have xc kernels with the proper long-range behavior to obtain correct optical spectra [14, 73]. Gonze et al. [87, 88] pointed out that the head of $f_{xc}$ has to diverge as $q^{-2}$ for $q \to 0$ to correctly describe the polarization of periodic insulators. With the $q^{-2}$ divergence, the head of $f_{xc}$ contributes in the sum of Eq. (42), dominating the other parts of $f_{xc}$ (the wings and the body). Local and semilocal xc kernels do not have this long-range behavior, and there is no obvious and consistent way of modifying them to include the long-rangedness. Hence, a different class of approximate xc kernels—excitonic xc kernels—is needed.

### 5.4 Excitonic xc kernels

Since TDDFT is formally rigorous, it should in principle yield exact optical absorption spectra for insulators. But even if we start from an exact ground-state Kohn-Sham calculation (which would give the exact independent-particle spectrum), the xc kernel $f_{xc}^{GG'}(k, \omega)$ has to carry a heavy burden: it has to open the gap and shift the Kohn–Sham band edge to the true band edge, and it has to cause an effective electron–hole attraction, leading to excitonic features in the spectrum. Formally, the xc kernel can be separated into a quasiparticle and an excitonic part [89]–[91],

$$f_{xc} = f_{xc}^{qp} + f_{xc}^{ex}.$$  \hspace{1cm} (46)

The two parts are responsible for the opening of the gap and the excitonic effects, respectively. Further justification of Eq. (46) will be given below in Section 5.4.6. Let us now focus on the excitonic part and give some examples of how it can be approximated.

#### 5.4.1 Contact exciton

Let us begin with an apparent paradox, namely, the so-called contact exciton. Even though we have stressed that the proper long-range behavior of the xc kernel is crucial, an ultra-short-range xc kernel of the general form

$$f_{xc}^{cont}(r, r') = -A_{cont} \delta(r - r')$$  \hspace{1cm} (47)

can produce excitonic features if the constant $A_{cont}$ is properly chosen [15, 18, 92]. By the same token, an ad-hoc scaled ALDA, $\alpha f_{xc}^{ALDA}$, can in principle produce excitons, although the scaling factor $\alpha$ would have to be rather absurdly large (typically of order $\sim 10^3$).

The resolution of the contact-exciton paradox is that the contact kernel and the scaled ALDA work via the body of the coupling matrix $K_{ik, jk'}^{xc}$. In other words, the missing long-range behavior is, somewhat unphysically, compensated by an ul-
trastrong short-range electron-hole interaction. It is found [15, 90] that the contact xc kernel can be tuned to reproduce certain features of the optical spectrum (for instance, a bound-exciton peak) but at the cost of a poor description of other parts of the spectrum.

5.4.2 Long-range corrected kernel

Since we know, from Section 5.3, that the long-range $1/q^2$ behavior of the head of the xc kernel is the key to excitonic effects, it is straightforward to construct a simple ad-hoc approximation that captures the right physics for the right reason. The resulting, so-called long-range corrected (LRC) kernel has the following form:

$$f^\text{LRC}_{xc,GG'}(q) = \frac{A_{\text{LRC}}}{q + |G|^2} \delta_{GG'}.$$  \hspace{1cm} (48)

where $A_{\text{LRC}}$ is a system-dependent fitting parameter. Despite its simple form, LRC spectra (with properly chosen $A_{\text{LRC}}$) can be in good agreement with experiment [73, 92] since the head contribution of the kernel tends to dominate over the local-field effects that are contained in the contributions of the body of $K^\text{xc}_{iak,jbk}$. A simple connection with the high-frequency dielectric constant $\varepsilon_\infty$ has been suggested [73]:

$$A_{\text{LRC}} = 4.651 \varepsilon_\infty^{-1} - 0.213.$$  \hspace{1cm} (49)

The purpose of this empirical formula was to reproduce the continuum spectrum; hence, it cannot be expected to (and, in fact, does not) perform well for bound excitons (see Section 5.4.9).

5.4.3 Exact exchange

The frequency-dependent xc kernel can be formally constructed from many-body perturbation theory, using a diagrammatic expansion [93, 94]. The first-order term of this expansion is the exact exchange kernel $f_x(r,r',\omega)$, which can be represented as the sum of five diagrams (see Fig. 4a). Translated into formulas using the standard diagrammatic rules, one obtains [95, 96]:

$$\int d^3r_1 \int d^3r_2 \chi_i(r, r_1, \omega)f_x(r_1, r_2, \omega)\chi_i(r_2, r', \omega) = R_V(r, r', \omega) + R_S(r, r', \omega).$$  \hspace{1cm} (50)

$R_V$ is the first-order vertex diagram (the third one on the right-hand side in Fig. 4a),

$$R_V(r, r', \omega) = -2 \sum_{ijkl} \phi_i(r)\phi_j^*(r)\phi_k(r')\phi_l^*(r') \langle il|w|jk \rangle \frac{(f_i - f_j)(f_k - f_l)}{(z - \omega_i)(z - \omega_k)}. \hspace{1cm} (51)$$
where $z = \omega + i\theta^+$, and the $f_j$ are the usual occupation factors. $R^s_2$ denotes the sum of all the remaining four diagrams (the self-energy diagrams):

$$R^s_2 (r, r', \omega) = 4 \sum_{i j k} \frac{\phi_i (r) \phi^*_i (r') \phi_j (r) \phi^*_j (r') \langle j | \Delta | k \rangle}{\omega_{k j}} \left\{ \frac{(f_k - f_j) \omega_{k j}}{z^2 - \omega_{k j}^2} - \frac{(f_j - f_i) \omega_{i j}}{z^2 - \omega_{i j}^2} \right\} ,$$  

(52)

where $\Delta (r_1, r_2) = \Sigma_x (r_1, r_2) - v_x (r_1) \delta (r_1 - r_2)$. Here, $\Sigma_x$ is the exchange part of the self-energy, and $v_x$ is the exact exchange potential of DFT, which is defined as an orbital functional via the optimized effective potential (OEP) method [45, 97]. It is also possible to derive the exact exchange kernel directly as the functional derivative $f_x (\omega) = \frac{d}{\delta n (\omega)} [98, 99]$.

In periodic insulators, the exact-exchange kernel has the long-range behavior that is necessary for the formation of excitons [100, 101]. However, the resulting unscreened electron–hole interaction tends to lead to a dramatic overbinding of the excitons; in extreme cases, this causes a collapse of the optical spectra (i.e., the exciton would be so strongly bound that it falls below the valence band edge). This collapse can be prevented by a cutoff of the Coulomb singularity [100, 101]; this is equivalent to an evaluation of the xc kernel with a screened interaction [15, 102].

### 5.4.4 Hybrid functionals and meta GGAs

Hybrid xc functionals [46]–[59] replace a portion of the semilocal exchange energy with the exact exchange energy, so the long-range part of the corresponding xc kernel resembles a screened exact exchange kernel; as we saw above, this can produce bound excitons. In practice, however, hybrid functionals are used in a different manner [103]: the exact exchange part is treated nonlocally, similar to the time-dependent HF approach, instead of using the exact exchange kernel in Eq. (42).

The B3LYP hybrid functional has been used by Bernasconi et al. to calculate optical spectra in several semiconductor materials [104]–[106]; they achieve a generally good description of optical gaps, including excitonic features. Indeed, we have obtained some preliminary results [107] which confirm that B3LYP can be reasonably accurate for exciton binding energies in semiconductors, despite the fact that the 0.2 mixing parameter of the exact exchange is optimized for finite systems [46].

Range-separated hybrid functionals [50, 108, 109] are based on the idea of separating the Coulomb interaction into different spatial ranges, which are then treated differently, using either exact exchange or approximate semilocal exchange functionals. Recent applications of range-separated hybrids to solids have produced good quasiparticle gaps [51, 52, 54, 55, 57, 61]. In linear response, these functionals are again closely related to the exact exchange kernel. However, if the range-separated functional uses semilocal exchange for the long-range part, it cannot produce bound excitons for the same reason as in ALDA. Therefore, the popular HSE06 functional [50, 51, 110] cannot yield bound excitons, although it may still produce decent looking optical spectra of insulators [111].
Fig. 4 Diagrammatic representations of (a) the exact exchange kernel \( f_x \) [Eq. (50)] and (b) the excitonic xc kernel \( f_{ex} \) [Eq. (61)], the so-called nanoquanta kernel. Full lines represent noninteracting Kohn-Sham Green’s functions, and dashed lines represent quasiparticle Green’s functions. The thin wavy lines are bare Coulomb interactions; the thick wavy line is a screened interaction.

The so-called meta-GGA functionals [112]–[115] depend not just on the density and its gradients, but also on the kinetic-energy density, which is expressed in terms of Kohn-Sham orbitals and, hence, depends nonlocally on the density. This nonlocality produces good quasiparticle gaps in solids [60]–[64], and opens up the possibility of describing excitonic interactions with meta-GGAs. Nazarov and Vignale [116] tested two types of meta-GGAs, TPSS [112] and VS98 [113, 114]. They found TPSS to be unsuited for describing dielectric properties of solids; VS98, on the other hand, performed rather well. Further tests are needed; however, the implementation of meta-GGAs for the linear response in solids turns out to be technically rather difficult.

5.4.5 The PGG kernel

The exact exchange kernel can be approximated in various ways. The simplest approximation is known as the PGG kernel (after Petersilka, Gossmann and Gross [117, 118]). In real space, it is defined as

\[
 f_{x}^{PGG}(\mathbf{r}, \mathbf{r}') = -\frac{2}{|\mathbf{r} - \mathbf{r}'|} \frac{\sum_{i} |\phi_{i}^{\ast}(\mathbf{r}) \phi_{i}(\mathbf{r}')|^{2}}{\rho_{n}(\mathbf{r}) \rho_{n}(\mathbf{r}')},
\]  

(53)

where \( \rho_{n} \) is the ground-state electronic density. In this form, the PGG kernel has been successfully applied to calculate atomic and molecular excitation energies as well as plasmons in nanostructures [119]. Thus, one might be optimistic regarding its performance for excitons.

We convert the PGG kernel into reciprocal space, assuming that the Kohn-Sham orbitals have the form \( \phi_{\mathbf{k}}(\mathbf{r}) = N_{\text{cell}}^{-1/2} u_{\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}} \), where \( N_{\text{cell}} \) is the number of unit cells in the crystal and \( u_{\mathbf{k}}(\mathbf{r}) \) are Bloch functions. \( f_{x}^{PGG} \) can then be written as
Excitons in TDDFT 

\[ f_{\text{PGG}}^{\text{PGG}}(r,r') = - \sum_{\mathbf{k} \mathbf{k}'} \sum_{n m} 2 \delta_{\mathbf{k} \mathbf{k}'} e^{-i \mathbf{k} \cdot (r-r')} H_{\mathbf{k} \mathbf{k}'}(r,r'), \]  

(54)

where \( H_{\mathbf{k} \mathbf{k}'}(r,r') \) is periodic within one unit cell and defined as

\[ H_{\mathbf{k} \mathbf{k}'}(r,r') = \frac{u^*_{\mathbf{k}}(r) u_{\mathbf{k}'}(r') u_{\mathbf{k}'}(r) u^*_{\mathbf{m}'}(r')} {N_{\text{cell}}^2 n(r) n(r')} \]  

(55)

The Fourier transform of \( f_{\text{PGG}}^{\text{PGG}} \) yields

\[ f_{\text{PGG}}^{\text{PGG}}(\mathbf{q}, \mathbf{G}, \mathbf{G}') = -\frac{8 \pi}{V} \sum_{\mathbf{i} \mathbf{i}'} \sum_{\mathbf{m} \mathbf{m}'} \tilde{H}_{\mathbf{i} \mathbf{i}'}(\mathbf{q} - (\mathbf{k} - \mathbf{G} + \mathbf{G}' - \mathbf{G})), \]  

(56)

where \( \tilde{H} \) is obtained by numerical Fourier transform of expression (55) within one unit cell. For simplicity, we ignore the local-field effects and only use the head of the PGG kernel, which is given by

\[ f_{\text{PGG}}^{\text{PGG}}(\mathbf{q},0,0) = -\frac{8 \pi}{V} \sum_{\mathbf{i}, \mathbf{m}, \mathbf{k}} \tilde{H}_{\mathbf{i} \mathbf{k}}(0,0) \frac{1}{q^2}. \]  

(57)

Unfortunately, explicit calculations show that the performance of the PGG kernel is disappointing for solids: it does not produce any bound excitons at all, despite having a nonzero head contribution with the correct \( 1/q^2 \) behavior [19]. How can this be reconciled with the fact that the PGG kernel seems to work well in finite systems such as atoms and molecules? Periodic systems are dominated by the head of the xc kernel in reciprocal space; however, the situation is very different in finite systems, where the electron dynamics can be viewed as coming entirely from local-field effects. Thus, the strongly attractive nature of the PGG kernel in finite systems would at most translate into a strong body of the xc matrix in periodic systems (which, however, is irrelevant for excitons), but does not necessarily guarantee a strong head. Indeed, if one fits the head of the PGG kernel to the LRC kernel (48), one finds that the resulting constant \( A_{\text{PGG}}^{\text{LRC}} \) is orders of magnitude too weak [19].

The underlying reason for the failure of the PGG kernel for periodic insulators can be inferred from its real-space definition, Eq. (53), which can be written in the form

\[ f_{\text{PGG}}^{\text{PGG}} = -|\rho(r,r')|^2/2 |r-r'| n(r) n(r')|, \]  

where \( \rho(r,r') \) is the Kohn-Sham density matrix. For periodic solids, the long-range behavior of \( f_{\text{PGG}}^{\text{PGG}} \) is determined by both the Coulomb singularity as well as the density matrix. It is a well-known fact [120]–[124] that the one-particle density matrix in insulators decays exponentially as \( \rho(r,r') \sim \exp(-\gamma |r-r'|) \). This effectively cuts off the required long-range behavior and explains why the head of the PGG kernel is so weak.

### 5.4.6 The “nanoquanta” kernel

Let us introduce the so-called proper response function \( \overline{\chi} \) as
\[ \tilde{\chi} = \chi_s + \chi_s f_{\text{xc}} \tilde{\chi} \]  

(58)

(for simplicity, we drop all arguments and integrals). In the beginning of this Section, we defined the quasiparticle and excitonic parts of the xc kernel, see Eq. (46). It is then easy to write down the following relations for the two parts of \( f_{\text{xc}} \) \[91\]:

\[ \chi_{\text{qp}} = \chi_s + \chi_s f_{\text{qp}}^\text{qp} \chi_{\text{qp}}, \]  

(59)

\[ \tilde{\chi} = \chi_{\text{qp}} + \chi_{\text{qp}} f_{\text{ex}}^\text{ex} \tilde{\chi}. \]  

(60)

Here, \( \chi_{\text{qp}} \) is the quasiparticle response function, which uses quasiparticle states as input. Hence, \( \chi_{\text{qp}} \) has the quasiparticle gap built in by default, and the roles of \( f_{\text{qp}}^\text{qp} \) and \( f_{\text{ex}}^\text{ex} \) are clear from Eqs. (59) and (60). Our focus here is on the excitons: all we need to do, then, is start with a good approximation for \( \chi_{\text{qp}} \). Usually, one obtains it from the GW approach (see Section 3), but other approximations that yield good quasiparticle gaps (such as hybrids or the scissors operator) can be used as well.

The exact proper response function \( \tilde{\chi} \) is of course unknown. To construct an approximation for \( f_{\text{ex}}^\text{ex} \) from Eq. (60) one can proceed in two steps. First, replace \( \chi_{\text{qp}} f_{\text{ex}}^\text{ex} \tilde{\chi} \approx \chi_{\text{qp}} f_{\text{ex}}^\text{ex} \chi_{\text{qp}} \). Second, find a diagrammatic representation of \( \tilde{\chi} \) – \( \chi_{\text{qp}} \); the details of this representation are quite technical and will not be given here (in essence, it involves two-particle Green’s functions and four-point vertex functions \[1, 14, 15, 125\]). The key point is that this diagrammatic representation can be very easily approximated, and one ends up with the following expression:

\[ \int d3 \int d4 \chi_{\text{qp}}(1,3) f_{\text{ex}}^\text{ex}(3,4) \chi_{\text{qp}}(4,2) = \]  

\[ \int d3 \int d4 G_{\text{qp}}(1,3) G_{\text{qp}}(4,1) W(3,4) G_{\text{qp}}(3,2) G_{\text{qp}}(2,4). \]  

(61)

This is the xc kernel of Reining et al. \[89\] and many others (also known as the “nanoquanta” kernel \[126\]). Figure 4(b) shows its diagrammatic representation. The numbers in Eq. (61) represent space-time arguments, e.g., \(1 = (r_1, t_1)\). \( G_{\text{qp}} \) is a quasiparticle Green’s function, and \( W \) is a screened interaction, formally defined as

\[ W(1, 2) = w(1, 2) + \int d3 \int d4 w(1, 3) \tilde{\chi}(3,4) W(4,2), \]  

(62)

where \( w(1, 2) \) is the bare Coulomb interaction, and \( \tilde{\chi} \) is approximated by \( \chi_{\text{qp}} \).

The excitonic xc kernel \( f_{\text{ex}}^\text{ex} \) of Eq. (61) has been widely applied in a variety of systems \[14, 89, 90, 102],[127]–[132\]. Its performance is, in general, found to be excellent, at par with results obtained from solving the full BSE. This provided an important proof of concept that TDDFT is very well capable of capturing excitonic properties. The price to be paid, however, is that the many-body xc kernel (61) is not simple to implement and computationally costly (it is, essentially, as expensive as the BSE when it comes to calculating optical spectra, but somewhat more favorable if the full dielectric matrix is needed).
5.4.7 The “bootstrap” kernel

Compared to the exact exchange and nanoquanta kernels, the simplicity of the LRC kernel is desirable for practical use. The adjustable parameter $A_{\text{LRC}}$ requires prior knowledge to the system, however, and therefore the LRC kernel cannot be used as a black-box method. The bootstrap kernel proposed by Sharma et al. [133, 134] can be seen as an attempt to determine the $A_{\text{LRC}}$ parameter (which now depends on $q$, $G$, and $G'$) self-consistently while retaining the simplicity of the LRC kernel. The original definition [133] is written in terms of symmetrized quantities to avoid singularities [for example, $f_{\text{sym}}^\text{xc}(q, G, G') = v_G^{-1/2}(q)f_{\text{xc}}(q, G, G')v_{G'}^{-1/2}(q)$, and $\chi_{\text{sym}}^\text{xc}(q, G, G') = v_G^{1/2}(q)\chi(q, G, G')v_{G'}^{1/2}(q)$]. In terms of regular, non-symmetrized quantities, the kernel is defined as

$$f_{\text{boot}}^\text{xc}(q, G, G') = \frac{[q]^2}{|q + G|} \left[ \delta_{GG'} + v_G^{1/2}(q)v_{G'}^{1/2}(q)\chi(q, G, G') \right],$$

(63)

where $v_G(q) = 4\pi/|q + G|^2$ is the Coulomb potential, $\chi$ is the Kohn–Sham linear response function of Eq. (29), and $\chi$ is the TDDFT linear response function obtained via Eq. (28). Eqs. (63) and (28) are solved self-consistently for the xc kernel. The $q \to 0$ behavior of $\chi(q, 0, 0)$ is $O(q^2)$, and it is canceled by the $|q|^2$ in the numerator. Thus the bootstrap $f_{\text{xc}}$ has the correct $q \to 0$ behavior due to the presence of $|q + G| |q + G'|$ in the denominator.

The bootstrap kernel has been reported to yield good continuum spectra (including the enhancement of the band-edge spectra due to continuum excitons) for a wide range of materials [133]. Unfortunately, numerical applications of this xc kernel are plagued by its slow convergence with respect to the total number of bands included in $\chi_{\text{xc}}$. When convergence is finally achieved (which may require including dozens of unoccupied bands), the results for bound excitons tend to be disappointing, with exciton binding energies typically orders of magnitude smaller than the experimental values [19]. In fact, contrary to Ref. [133] the bootstrap kernel does not yield bound excitons for wide-gap insulators such as LiF and solid Ar (we will present numerical results in Section 5.4.9). Improving the performance of the bootstrap kernel by suitable modification is a subject of ongoing research.

5.4.8 The jellium-with-a-gap model

Trevisanutto et al. recently developed an xc kernel based on the jellium-with-a-gap model (JGM) [135]. Although the JGM kernel depends on the local density, it differs from local and semilocal xc kernels by having the correct $1/q^2$ and $1/q$ behavior of head and wings as $q \to 0$, and it can therefore in principle produce bound excitons. The JGM kernel is an empirical kernel since it requires the band gap as input. The kernel is defined as
Table 1 Exciton binding energies calculated with different TDDFT xc kernels, compared with experimental values (all energies in meV).

<table>
<thead>
<tr>
<th></th>
<th>GaAs</th>
<th>β-GaN</th>
<th>α-GaN</th>
<th>CdS</th>
<th>CdSe</th>
<th>Ar</th>
<th>Ne</th>
<th>LiF</th>
<th>AlN</th>
<th>ZnO</th>
<th>MgO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.</td>
<td>3.27</td>
<td>26.0</td>
<td>20.4</td>
<td>15.0</td>
<td>1900</td>
<td>4080</td>
<td>1600</td>
<td>75</td>
<td>60</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>LRC</td>
<td>0.858</td>
<td>0.514</td>
<td>0.513</td>
<td>1.40</td>
<td>0.304</td>
<td>0.127</td>
<td>1.40</td>
<td>0.304</td>
<td>0.127</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>boot</td>
<td>0.332</td>
<td>0.199</td>
<td>0.461</td>
<td>0.895</td>
<td>1.70</td>
<td>852</td>
<td>32.2</td>
<td>1.09</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JGM</td>
<td>0.833</td>
<td>0.382</td>
<td>0.741</td>
<td>1.42</td>
<td>41.0</td>
<td>0.593</td>
<td>993</td>
<td>4.45</td>
<td>1.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1

\[
f_{\text{xc}}^{JGM}(q,n,E_g) = \frac{4\pi [B(n) + E_g]}{q^2(1 + E_g)} \left( e^{-k(n,E_g)q^2} - 1 \right) - \frac{4\pi C(n)}{(3\pi^2 n)^{2/3}(1 + 1/q^2)(1 + E_g)},
\]

where \( C(n) = -\frac{\pi}{2(3\pi^2 n)^{1/3}} \frac{d[r_s \epsilon_c(r_s)]}{dr_s}, \) \( k(n,E_g) = -\frac{\alpha n^\beta}{B(n) + \frac{E_g^2}{4\pi q^2 n}[B(n) + E_g]}, \)

and \( B(n) = \frac{1 + a_1 r_s^{1/2} + a_2 r_s^{3/2}}{3 + b_1 r_s^{1/2} + b_2 r_s^{3/2}}. \)

Here, \( r_s = [3/(4\pi n)]^{1/3}, \) \( \epsilon_c \) is the LDA correlation energy per particle \( [138], \) and \( \alpha = -0.02552, \beta = -0.6916, a_1 = 2.15, a_2 = 0.435, b_1 = 1.57, b_2 = 0.409. \) The performance of the JGM kernel is similar to that of the bootstrap kernel for continuum spectra; unfortunately, it also severely underestimate exciton binding energies.

5.4.9 Some results

We now present some results for exciton binding energies, obtained using Eq. (40), to demonstrate the performance of several xc kernels that we have discussed. All TDDFT calculations are done on top of scissor-corrected \([74, 75]\) LDA band structures, so that only the excitonic part of the exact xc kernel is approximated by the functionals that we consider. All calculations include only the head of the xc coupling matrix \( K_{\text{xc},ij}^{\text{xc}}, \) \( i \neq j. \) Ignoring the wings and the body of the xc coupling matrix gives an estimated error of less than 5%. Other details specific to our numerical implementation are described in Ref. \([19]\).

Experimental and calculated exciton binding energies for several materials are collected in Table 1. We compare the performance of three different xc kernels: the LRC kernel (48), evaluated using Eq. (49), the bootstrap kernel (63) \([139]\), and the JGM kernel (64). Except for \( \alpha \)-GaN and AlN, all three TDDFT methods produce a bound exciton; however, the results are quantitatively not very accurate. By and large, the excitons are significantly underbound, sometimes by up to 2-3 orders of magnitude.
Excitons in TDDFT

Table 2

<table>
<thead>
<tr>
<th>Material</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
<th>A_{LRC}</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs</td>
<td>0.595</td>
<td>2.409</td>
<td>3.628</td>
<td>4.244</td>
<td>2.144</td>
<td>96.5</td>
<td>9.5</td>
<td>3.008</td>
<td>1.628</td>
<td>4.040</td>
<td>4.040</td>
</tr>
<tr>
<td>β-GaN</td>
<td>0.088</td>
<td>0.304</td>
<td>0.214</td>
<td>0.589</td>
<td>0.318</td>
<td>6.448</td>
<td>46.3</td>
<td>4.236</td>
<td>0.341</td>
<td>0.362</td>
<td>1.230</td>
</tr>
<tr>
<td>α-GaN</td>
<td>0.206</td>
<td>0.524</td>
<td>0.578</td>
<td>0.463</td>
<td>9.685</td>
<td>10.76</td>
<td>7.78</td>
<td>1.568</td>
<td>0.800</td>
<td>2.357</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Top row: fitted $A_{LRC}$ parameter which reproduces the experimental exciton binding energies with the LRC kernel. Second and third row: heads of the xc coupling matrix of the bootstrap and JGM kernels.

Fig. 5

Absorption spectra of solid Ar obtained from Eq. (43) (top panel) and from Eq. (30) (bottom panel, adapted from Ref. [129]). Reproduced with permission from APS from [129]. ©2007.

The LRC kernel contains the adjustable parameter $A_{LRC}$, which can be fitted to reproduce the experimental exciton binding energy. The fitted values of $A_{LRC}$ are given in the first row of Table 2. From the head of the bootstrap and the JGM xc coupling matrices, we can extract the corresponding $A_{LRC}$ parameters; the results are given in the second and third row of Table 2. Clearly, the bootstrap and JGM kernels produce heads that are significantly too weak compared to what would be needed to reproduce the experimental data. Clearly, the exciton binding energy depends crucially on the strength of the head of the xc coupling matrix.
Figure 5 shows optical absorption spectra of solid Ar. The spectra in the top panel were calculated from the solutions of Eq. (40) using Eq. (43). The bottom panel (adapted from Ref. [129]) was calculated via Eq. (30). The top panel compares RPA and LRC, both of them using a scissor-corrected LDA band structure as input. The LRC kernel clearly produces a very strong bound exciton, but the continuum part is too weak compared to the excitonic peak. The bottom panel compares experimental results with calculations using GW-BSE and GW-TDDFT, using the nanoquanta kernel. Clearly, BSE and the nanoquanta kernel are in excellent agreement, and both agree well with the experimental data (the splitting of the peaks is due to spin-orbit coupling, which is not included in the calculations). Notice that a second bound exciton is present, which is missing in the LRC calculation. For comparison, the bottom panel of Fig. 5 also shows ALDA results, based on an uncorrected LDA band structure. Clearly, the ALDA bears no resemblance whatsoever to experiment.

6 Comparison of TDDFT and the BSE

Exciton binding energies in TDDFT are determined via Eq. (40). The Hxc coupling matrix is given by

$$K_{iak,jbk'}^{Hxc} = K_{iak,jbk'}^{H} + K_{iak,jbk'}^{xc},$$

(68)

where the Hartree and xc matrices are defined in Eqs. (41) and (42). It turns out that the BSE leads to an equation for exciton binding energies that is formally identical to Eq. (40), except that it features the coupling matrix

$$K_{iak,jbk'}^{BSE} = K_{iak,jbk'}^{H} + K_{iak,jbk'}^{W},$$

(69)

where

$$K_{iak,jbk'}^{W} = \frac{1}{V} \sum_{GG'} W_{GG'}(q) \langle i | e^{i(q+G') r} | j \rangle \langle jk | e^{-i(q+G') r} | bk' \rangle.$$

(70)

The Hartree part is the same as in TDDFT, but the xc part is replaced by a coupling matrix featuring the screened interaction (62). In practice, one ignores the frequency dependence of $W$; explicitly, one finds

$$W_{GG'}(q, \omega = 0) = \frac{4\pi \varepsilon_{GG'}^{-1}(q, \omega = 0)}{|q + G'|^2}.$$

(71)

Let us now compare the two coupling matrices $K^{xc}$ and $K^{W}$. Two main differences become apparent: first, the order of the band indices $i, j, a, b$ is different; secondly, the xc matrix only depends on the long-range ($q = 0$) behavior, while the $W$ matrix also depends on other values of $q$.

Figure 6 shows contour plots of the xc and $W$ coupling matrices, calculated for a one-dimensional model insulator with a soft-Coulomb interaction [18]. The xc...
kernel is here the long-range corrected kernel $f^{LRC}_{xc}$ with a fitting parameter chosen such that the lowest exciton binding energy in TDDFT and BSE is the same.

The two coupling matrices shown in Fig. 6 are strikingly different. This is not surprising, since the screened interaction $W^{GG'}(q)$ has an extra degree of freedom over $f^{gg'}_{xc}(q = 0)$; hence, it cannot be expected that an adiabatic xc kernel can be found that reproduces the full BSE coupling matrix. One can only hope to reproduce a portion of the BSE coupling matrix, unless the xc kernel is made frequency-dependent so that at least some of the information from the $q$-dependence in the screened interaction is mapped into the frequency dependence of the xc kernel [89]. As a consequence, adiabatic xc kernels can be made to produce a single bound exciton, but not an excitonic Rydberg series (at least, not just with the head of the xc matrix) [18].

The question then arises how TDDFT can produce bound excitons at all, given the fact that the xc and BSE coupling matrices are so drastically different. To illustrate how this is possible, we now make a connection with the Wannier model that we discussed in Section 2, and ask: what is the TDDFT and BSE analog of the Wannier equation, Eq. (1)?

Let us consider, for simplicity, a two-band model in which there is only one filled valence band ($v$) and one empty conduction band ($c$). We define an effective two-body potential via the Fourier transform of the xc coupling matrix:

$$V^{xc}_{v-c}(R, R') = \sum_{k, k' \in BZ} e^{-ikR} K^{xc}_{vc,k} e^{ik' R'},$$

where $R, R'$ are direct lattice vectors. Since Wannier exciton radii extend over many lattice constants, one may replace $R$ by a continuous spatial variable $r$. Assuming, furthermore, parabolic valence and conduction bands, and using the effective-mass approximation, Eq. (40) becomes, after Fourier transformation,
where $E$ is the exciton binding energy, and the integration goes over all space. This shows that the TDDFT analog of the Wannier equation (1) is a nonlocal Schrödinger equation. With a proper choice of the xc kernel, the nonlocal effective electron-hole interaction potential $V_{e-h}^{xc}$ supports bound excitonic states.

Due to the formal similarity between BSE and TDDFT in the transition space representation, we can also define a BSE effective electron-hole interaction potential $V_{e-h}^{BSE}$ in analogy to Eq. (72), using the BSE kernel $K^W$. Figure 7 shows a comparison of $V_{e-h}^{xc}$ and $V_{e-h}^{BSE}$ for our one-dimensional model insulator. It can clearly be seen that the nonlocal potentials in both cases are dominated by the diagonal part; however, under further examination it turns out that $V_{e-h}^{xc}$ is more shallow than $V_{e-h}^{BSE}$, and hence is only able to sustain a single bound exciton.

7 Conclusions

There exist two alternative, complementary first-principles methods to calculate optical spectra and excitonic effects in extended periodic solids. The more traditional approach is based on many-body Green’s function techniques, exemplified by the GW-BSE method; the other approach is based on TDDFT. The Green’s function approach is formally straightforward, in the sense that excitonic particle-hole interactions are built in by construction; however, it is computationally costly. TDDFT, on the other hand, is computationally cheaper (provided that the approximate xc functionals are simple enough), but the price one has to pay is that intrinsically non-local electron-hole interaction effects have to be described via linearized local xc
Excitons in TDDFT

potentials; this is a somewhat unnatural way of dealing with excitonic interactions, which makes it non-straightforward to construct good approximations.

In this article, we have focused on the TDDFT approach for excitons, and tried to bring across the following points:

- Excitons in TDDFT are a difficult problem, because they require an xc kernel which has the long-range property \( f_{\text{xc},0,0}(q, \omega) \sim q^{-2} \) for \( q \to 0 \). The popular local and semilocal approximations such as ALDA and standard GGAs do not have this property: although they work well for finite systems such as atoms and molecules, they do not produce excitons in extended systems. New approximations are therefore required.

- There exist several approximate “excitonic” xc kernels, with various degrees of sophistication. Some kernels involve adjustable parameters, others don’t. A typical behavior of the simpler kernels is that they can reproduce some part of the optical spectrum reasonably well (e.g., a bound exciton, or the continuum part), but not all of it at the same time. The best excitonic xc kernel, the “nanoquanta” kernel, is computationally not much simpler than the BSE.

- Adiabatic (i.e., frequency-independent) xc kernels cannot produce an excitonic Rydberg series. At best, they can generate a single bound exciton (if only the head of the xc coupling matrix is used). The TDDFT analog of the excitonic Wannier equation features a nonlocal potential, which is too shallow to produce more than one bound level.

Despite the difficulties of developing accurate and efficient TDDFT approaches for excitons, much progress has been made over the past few years. Most likely, the path forward will bring TDDFT and traditional many-body theory even closer together. The outcome will be xc kernels that are functionals of occupied as well as unoccupied bands, similar in spirit to the nanoquanta and bootstrap kernels.

Hybrid xc kernels are another promising avenue forward; however, this means abandoning pure TDDFT and admitting nonlocal exchange. This has been an extremely successful strategy for finite systems: indeed, standard hybrid functionals are widely used for excitons in polymers and other nanoscale systems [140]–[143]. For periodic solids, the standard hybrid functionals may have to be suitably modified to describe excitonic properties; work along these lines is in progress [107].

Acknowledgments

This work was supported by NSF grant DMR-1005651. We thank Lucia Reining and Francesco Sottile for many helpful discussions. C. A. U. thanks the ETSF at the Ecole Polytechnique in Palaiseau for its hospitality.

References

27. Mott NF (1938) Trans Faraday Soc 34:500
We find that the convergence of the bootstrap kernel strongly depends on the number of bands used in the iterative calculation of the kernel. The results for solid Ar, solid Ne and LiF are obtained by calculating the bootstrap kernel with 30 bands. It turns out that some of our previous results reported in Ref. [19], where the bootstrap kernel seemed to work very well for Ar, Ne and LiF, were in fact not fully converged.