Spooling up the solution of the Bethe-Salpeter equation by a double-grid method and Wannier interpolation

David Kammerlander,1,2,* Silvania Botti,2,3 Miguel A. L Marques,2 Andrea Marini,4 and Claudio Attaccalite1
1Institut Néel, Université Joseph Fourier and CNRS, F-38042 Grenoble Cedex 9, France
2Université de Lyon, F-69000 Lyon, France and LPMCN, CNRS UMR 5586, Université de Lyon 1, F-69622 Villeurbanne Cedex, France
3Laboratoire des Solides Irradiés and ETSF, École Polytechnique, CEA-DSM, CNRS, F-91128 Palaiseau, France
4Istituto di Struttura della Materia (ISM), Consiglio Nazionale delle Ricerche, Via Salaria Km 29.5, CP 10, I-00016 Monterotondo Stazione, Italy

(Received 1 June 2012; published 13 September 2012)

The Bethe-Salpeter equation is a widely used approach to describe optical excitations in bulk semiconductors. It leads to spectra that are in very good agreement with experiment, but the price to pay for such accuracy is a very high computational burden. One of the main bottlenecks is the large number of \( k \) points required to obtain converged spectra. In order to circumvent this problem we propose a strategy to solve the Bethe-Salpeter equation based on a double-grid technique coupled to a Wannier interpolation of the Kohn-Sham band structure. This strategy is then benchmarked for a particularly difficult case, the calculation of the absorption spectrum of GaAs, and for the well-studied case of Si. The considerable gains observed in these cases fully validate our approach, and open the way for the application of the Bethe-Salpeter equation to large and complex systems.

DOI: 10.1103/PhysRevB.86.125203 PACS number(s): 78.20.Bh

Optical spectra are important for the characterization and prediction of material properties, as optical excitations are at the core of, for example, light-emitting devices, laser technology, and photovoltaics. For extended systems many-body perturbation theory,1,2 a Green’s-function-based approach, is the most accurate method to calculate optical properties. Perhaps inevitably, it is also one of the most computationally costly methods available to the community. It involves the solution of an equation of motion for the two-particle Green’s function, the Bethe-Salpeter equation (BSE), that describes coupled and correlated electron-hole excitations.3

The standard numerical techniques used to solve the BSE are based on an expansion of the relevant quantities in electron-hole states (needing therefore both filled and empty states), and are based on an expansion of the relevant quantities in electron-perturbation theory,1,2 a Green’s-function-based approach, is the most accurate method to calculate optical properties. Perhaps inevitably, it is also one of the most computationally costly methods available to the community. It involves the solution of an equation of motion for the two-particle Green’s function, the Bethe-Salpeter equation (BSE), that describes coupled and correlated electron-hole excitations.3

The standard numerical techniques used to solve the BSE are based on an expansion of the relevant quantities in electron-hole states (needing therefore both filled and empty states), and require a very dense \( k \)-point sampling of the Brillouin zone (BZ). Typically, the number of electron-hole states used in the expansion can be relatively small if one is only interested in the visible spectra, but the number of \( k \) points can easily reach several thousands. Some approaches have been put forward to reduce the computational burden of the BSE. For example, the number of \( k \) points can be reduced by interpolating the interaction integrals in \( k \) space,4 while recent implementations allow for the complete exclusion of empty states.5

It is well known that optical spectra are very sensitive to the \( k \)-point sampling.3-7 A common approach to alleviate the problem is the use of arbitrarily shifted \( k \)-point grids, that often yield sufficient sampling of the Brillouin zone while keeping the number of \( k \) points manageable. Such a shifted grid, indeed, does not use the symmetries of the Brillouin zone and guarantees a maximum number of nonequivalent \( k \) points accelerating spectrum convergence.8 However, it might induce artificial splitting of normally degenerate states, and thus produce artifacts in the spectrum,7 such as the splitting of some peaks or even the appearance of spurious excitations in some directions. Of course, these artifacts (slowly) disappear with increasing density of \( k \) points,3,6 and consequent increase of the computational burden. In view of that a very dense \( k \)-point sampling is crucial to obtain an accurate lineshape, including the correct peak positions,6,10 but very hard to achieve in practical calculations.

In this article, we propose a new strategy to solve the BSE equation that alleviates the need for dense \( k \)-point grids. The independent-particle part of the BSE is first evaluated on a very dense \( k \) grid (40 \( \times 40 \times 40 \) in the example below) by making use of Wannier interpolation.11 The BSE is then solved in an unshifted coarse \( k \) grid (10 \( \times 10 \times 10 \) in the example below) using a double-grid technique to take into account the fast-changing independent-particle contribution. This approach is simple to implement, and leads to a considerable gain in computational time.

In the following, we start by presenting a short review of the theoretical ingredients for the description of optical spectra within the BSE.4,12 We then discuss our approach and prove its usefulness with a notoriously difficult example, the calculation of the optical absorption spectrum of the standard semiconductor GaAs. The subsequent discussion of bulk silicon concludes the benchmark.

The optical absorption spectrum is described by the imaginary part of the macroscopic dielectric function \( \epsilon_M(\omega) \) in the long wavelength limit, which in turn can be obtained from the two-point contraction of the reducible four-point polarizability \( L \),

\[
\epsilon_M(\omega) = 1 - \lim_{q \to 0} \nu(q) \lambda \int d\mathbf{r} d\mathbf{r}' e^{-i(q-\mathbf{r})} L(\mathbf{r}, \mathbf{r}', \mathbf{r}'; \omega), \tag{1}
\]

with the Coulomb potential \( \nu = 4\pi/(\mathbf{G} + \mathbf{q})^2 \), the transferred momentum \( \mathbf{q} \), and \( \lambda \) the direction of light polarization. The quantity \( L \) satisfies the BSE, a Dyson-like equation,

\[
L(1,2,3,4) = L^0(1,2,3,4) + \int d(5678) L^0(1,2,5,6) \Xi(5,6,7,8) L(7,8,3,4), \tag{2}
\]
with the abbreviation of space, spin, and time coordinates \((1) = (r_1, \sigma_1, t_1)\), and where \(L^0(1, 2, 3, 4) = iG(1, 3)G(4, 2)\) is the independent particle polarization, expressed as a product of single-particle Green’s functions. Equation (2) describes the effects of the electron-hole interaction mediated by the BSE kernel \(\Xi = \bar{V} - W\) that is composed of (i) a bare, repulsive short-range exchange term, that includes the microscopic components of the Coulomb interaction, and that,
\[\bar{V}_{G=0} = v_G; \bar{V}_{G=\alpha} = 0,\]
(ii) an attractive, static screened Coulomb potential \(W\), the \textit{direct term}, arising from the variation of the self-energy. Dynamical effects due to the self-energy influence both the (single) quasiparticle renormalization and the excitonic two-body interaction \(W\),

\[\Xi = \frac{1}{\omega - \epsilon_{c_1} - \epsilon_{c_2} - \omega_i - \alpha} \delta_{k_1, k_2},\]

is not true for metals.\(^{14}\) As the interaction is instantaneous, only two time variables of the initial four remain. And, due to the time-translation invariance, \(L\) and \(L^0\) depend only on the relative time difference. A time-energy Fourier transformation then turns the polarization into a function of a single frequency \(L(1, 2, 3, 4; \omega)\) where from now on \((1) = (r_1, \sigma_1)\) only.

By taking advantage of the two-particle nature of the BSE, all expressions are conveniently written on the basis of the electron-hole vertical transition space composed of \(N_c\) valence bands, \(N_v\) conduction bands, and \(N^BZ_k\) points in the whole BZ. The dimension of this basis is \(4 \times N_c \times N_v \times N^BZ_k\). In the case of vanishing spin-orbit coupling, the BSE can be separated into the two subspaces of singlet and triplet excitons,\(^{7}\) each of them with dimension \(2 \times N_c \times N_v \times N^BZ_k\). The basis sets that span these subspaces are constructed from pairs of single-particle states \(\phi_{n,k}\) with \(n\) as band index and \(k\) as point and spin variable, such that

\[\Phi_k(r_1, r_2) = \phi_{c,k}(r_1) \cdot \phi_{v,*}^{\sigma}(r_2),\]

with the short-hand notation \(\Phi = (c, v, k)\), and \(c\) and \(v\) running over indices of conduction and valence bands, respectively.

In this basis the polarizability is written as

\[L_{k_1, k_2}(\omega) = \int dr_1 dr_2 dr_3 dr_4 \Phi_k(r_1, r_2) \times L(1, 2, 3, 4; \omega) \Phi_k^*(r_3, r_4),\]

and \(L^0\), that is now diagonal, reads

\[L^0_{k_1, k_2}(\omega) = \frac{f_{c_1} - f_{v_1}}{\epsilon_{c_1,k_1} - \epsilon_{v_1,k_2} - \omega - i\eta} \delta_{k_1, k_2},\]

where \(f\) denotes the occupation number. The infinitesimal \(\eta\) shifts the pole \(\omega - \epsilon_{c_1,k_1} - \epsilon_{v_1,k_2}\) away from the real axis, and is thus responsible for a finite lifetime of the excitation.

In this basis, the BSE becomes a matrix equation,

\[L_{k_1, k_2} = L^0_{k_1, k_2} + L^0_{k_1, k_2} \Xi_{k_1, k_2},\]

where we used Einstein’s notation for summations over repeated indices in the tensor products, and omitted the explicit energy dependence for clarity. In the following we adopt the notation of \(O\) for the matrix representation in the electron-hole basis of an arbitrary operator \(O\).

It can be shown that the kernel \(\Xi\) couples pairs of excitations \((uc)\) with \((v'c')\), but also \((uc)\) with \((c'v)\), leading to the so-called resonant and coupling terms, respectively.\(^{4}\) Here we make use of the so-called Tamm-Dancoff approximation, and neglect the latter. We thus arrive at a Hilbert space of dimension \(N_c \times N_v \times N^BZ_k\), regardless of the symmetries of the \(k\) grid. Note that these are standard approximations for the solution of the BSE.

Equation (8) can be solved symbolically, yielding for each frequency \(\omega\),

\[L_{k_1, k_2} = [1 - L^0_{k_1, k_2}]^{-1} L^0_{k_1, k_2},\]

that is interrupted at convergence. If, however, a convergence is not attained, that is, the assumption of the expandability of Eq. (7) is falsified \textit{a posteriori}, we perform the full inversion.

The solution of Eq. (8) has two distinct bottlenecks in terms of computational cost. First, the calculation of matrix \(\Xi\) is very time consuming. Second, its storage needs large quantities of memory. In view of that, reducing the number of \(k\) points is a major issue. To this end, Rohlfling and Louie\(^{4}\) employed a double-grid technique where the kernel \(\Xi\) is calculated on a coarse grid with its subsequent interpolation onto a fine \(k\)-point mesh where the BSE is solved. This approach helps reducing the time necessary to compute \(\Xi\), but it is less helpful to save memory, since it requires the storage of the computed and interpolated matrix elements of the kernel. Its use is justified with the authors’ observation that \(\Xi\) varies little with respect to the \(k\) points, as the single-particle wave functions \(\phi_{n,k}\) are quite robust with respect to \(k\) (with the possible exception of sudden band crossings). In a similar spirit it has been proved\(^{15,16}\) for the random phase approximation (RPA) that \(L^0\) is a rapidly varying quantity and that could be correctly evaluated by performing additional Monte Carlo integrations on a large number of random \(k\) points. Note that \(L^0\) can in principle be easily calculated for a large number of \(k\) points and bands.

By taking into account these observations, we define in our approach two grids: (i) a coarse one with points \(k\) on which we calculate and store \(\Xi\) and solve Eq. (8), and (ii) a fine \(k\) grid with vectors \(\mathbf{k}\) on which we compute \(L^0\). The mapping of \(L^0_{k_1, k_2}\) to \(L^0_{\mathbf{k}_1, \mathbf{k}_2}\) is performed through a double-grid technique\(^{17}\) with a suitably chosen interpolation for the kernel. To simplify our approach we use the simplest zeroth-order interpolation that leads to averaging the finely resolved \(L^0\) in a neighborhood around each point of the coarse grid. Consequently, this technique is expected to work if the oscillator strengths and \(\Xi\) are smoothly varying functions of \(\mathbf{k}\). In practice, we define

\[L^0(\omega)_{\mathbf{k}_1, \mathbf{k}_2} = \frac{1}{N_k} \sum_{\mathbf{k} \in D_k} \frac{f_{c_1,k} - f_{v_1,k}}{\epsilon_{c_1,k} - \epsilon_{v_1,k} - \omega - i\eta} \delta_{\mathbf{k}_1, \mathbf{k}_2},\]
where $N_k$ is the number of $k$ points of the fine grid in the domain $D_k$ around $k$ of the coarse grid. We would like to note that we are averaging the polarization $L^0$ that has poles at the particle excitation energies, which is not equivalent to averaging the particle excitation energies that appear in the diagonal of the excitonic Hamiltonian. An arbitrary $k$-point resolution of $L^0$ is possible once the respective single-particle energies $\epsilon_{nk}$ are available.

In general the calculation of quasiparticle states on the $\mathbf{k}$ grid are not practical. Fortunately, there is a solution to this problem that relies on the interpolation of the (quasiparticle) electronic structure to a dense $k$ grid using maximally localized Wannier functions. In this method, $\epsilon_{nk}$ at an arbitrary $k$ point $\mathbf{k}$ is the result of (i) a rotation of the initial quasiparticle Hamiltonian into the Wannier basis, (ii) its Fourier interpolation to the fine grid of $k$ points, and (iii) the diagonalization of the resulting Hamiltonian. Note that, even if this procedure leads to an expression that has the form of a Slater-Koster tight-binding interpolation, the obtained single-particle energies are calculated directly from the underlying $ab\ initiali$ eigensystem rather than just fitted to it.

To illustrate the efficiency of our scheme we calculated the optical spectra of semiconducting GaAs, known for its slow convergence with respect to the $k$-point sampling. The Kohn-Sham band structure and wave functions were obtained with density functional theory (DFT) within the local density approximation using norm-conserving pseudopotentials with an energy cutoff of 14 Hartree and the experimental lattice constant of 10.68 Bohr.

For the DFT part we utilized the code ABINIT. To obtain the energy bands [used in Eq. (9)] we employ the code WANNIER90 to perform a Wannier interpolation to a $40 \times 40 \times 40$ regular $k$-point grid in the whole BZ. Finally, optical spectra were calculated using the code YAMBO that uses the DFT Kohn-Sham wave functions and the interpolated single-particle energies as input.

**FIG. 1.** (Color online) Calculated RPA absorption spectra starting from $GW$ corrected bands for bulk GaAs with $N_v=2$ and $N_c=3$ using several $k$ grids without (a) and with (b) double-grid method. The shaded area in (b) shows the spectrum corresponding to the $40 \times 40 \times 40 k$ grid of panel (a) for better comparison between the double-grid method and the standard calculations.

**FIG. 2.** (Color online) Calculated BSE absorption spectra for bulk GaAs with $N_v=2$ and $N_c=3$ using several $k$ grids without (a) and with (b) double-grid method. In (a) one spectrum is calculated on a symmetric $10 \times 10 \times 10 k$ grid, while the others are on shifted grids to accelerate convergence. The shaded area in both panels indicate the experimental spectrum at 22 K.

It is well known that self-energy corrections in GaAs can be simulated by a rigid shift of the conduction bands. We therefore applied a scissor operator of 0.9 eV that yields an overall agreement of the band dispersions within 0.1 eV with the $GW$ corrected bands, and a close agreement with experimental data as well.

For all spectra in Figs. 1 and 2 we included the two highest valence bands and the two (three for BSE) lowest conduction bands, considering only the resonant part of the BSE kernel $\Sigma$, and used a Lorentzian broadening of 0.1 eV. Furthermore, we neglected spin-orbit coupling. Omitting local field effects (LFE), the noninteracting RPA spectra in Fig. 1(a) are obtained with the $GW$ corrected bands, and a close agreement with experimental data as well.

For all spectra in Figs. 1 and 2 we included the two highest valence bands and the two (three for BSE) lowest conduction bands, considering only the resonant part of the BSE kernel $\Sigma$, and used a Lorentzian broadening of 0.1 eV. Furthermore, we neglected spin-orbit coupling. Omitting local field effects (LFE), the noninteracting RPA spectra in Fig. 1(a) are obtained with the $GW$ corrected bands, and a close agreement with experimental data as well.

**TABLE I.** Number of $k$ points in the (ir)reducible BZ $N_k^{IBZ}$ that are used for the calculation of the spectra in Figs. 1–3.

<table>
<thead>
<tr>
<th>Calc.</th>
<th>Figs.</th>
<th>$k$-point grid $N_k^{IBZ}$</th>
<th>$N_k^{IBZ}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPA</td>
<td>3(a)</td>
<td>$60 \times 60 \times 60$</td>
<td>5216</td>
</tr>
<tr>
<td></td>
<td>1(a)</td>
<td>$40 \times 40 \times 40$</td>
<td>1661</td>
</tr>
<tr>
<td>RPA, dgrid</td>
<td>1(b), 3(a)</td>
<td>$12 \times 12 \times 12$</td>
<td>72</td>
</tr>
<tr>
<td>BSE</td>
<td>2(a)</td>
<td>$10 \times 10 \times 10$ shifted</td>
<td>1000</td>
</tr>
<tr>
<td>BSE, dgrid</td>
<td>2(b), 3(b)</td>
<td>$10 \times 10 \times 10$</td>
<td>47</td>
</tr>
</tbody>
</table>

125203-3
seen by comparing the noninteracting and interacting results of Figs. 1 and 2, respectively.

Figure 2(a) illustrates that a symmetric $10 \times 10 \times 10$ grid alone does not provide enough independent sampling points for a converged BSE spectrum. Shifting this grid in a direction different from the high symmetry directions provides 1000 instead of only 47 nonequivalent sampling points (see Table I). This leads to a spectrum that is sufficiently compatible with experimental results of Ref. 20. Nevertheless, the low-energy region (peak at 1.9 eV) and the region between the two main transitions at 3.2 and 5.1 eV are still expected to change on a denser $k$ grid. With our double-grid technique the BSE spectrum is converged even on a symmetric grid of $10 \times 10 \times 10$ [see Fig. 2(b)]. Contrary to the spectrum on an equally dense, but shifted grid, our spectrum is smooth in between the main transitions at 3.2 and 5.1 eV. It is noteworthy to mention that, for our scheme, a shifted, coarse $k$ grid does not improve the convergence of the spectrum of GaAs.

In the same fashion we calculated the RPA and BSE absorption spectra of Si, shown in Fig. 3. To converge the ground-state KS energies and wave functions with DFT we used an energy cutoff of 15 Ha and a lattice constant of 10.2 Bohr obtained by crystal relaxation.\(^{13}\) Similarly to GaAs, we took advantage of the good approximation of self-energy corrections by a rigid shift of the conduction bands of 0.8 eV.\(^{1}\) For the spectra we employed the two highest valence bands and the two (three for BSE) lowest conduction bands, and we included LFE by a dielectric matrix of size $51 \times 51$. Furthermore, we diminished the Lorentzian broadening to 0.05 eV in order to better resolve the first peak of the BSE spectrum at 3.5 eV.

Consequently a high $k$-point resolution of $60 \times 60 \times 60$ was necessary to converge the RPA spectrum on a symmetric grid without the double-grid method. The shaded area in (b) shows the experimental spectrum (indicated by the shaded area).

A final remark for Si on the sampling of the dense $k$ points $\vec{k}$ in Eq. (9) is in order now. Instead of using a $40 \times 40 \times 40$ regular grid of the full BZ (as for GaAs), we found it favorable to resort to a set of four $40 \times 40 \times 40$ shifted MP grids that respects the face-centered cubic symmetry of the Si crystal. Note that the calculations of the spectra using the double-grid technique were then again performed on unshifted, symmetric MP grids.

Although a MP grid that respects the symmetries of the BZ does not reduce the dimension of the BSE kernel, its use is still advantageous for three reasons. Firstly, there is no artificial splitting of degenerate states.\(^{9}\) Secondly, no artificial crystal anisotropy is introduced that has to be compensated by averaging the computed spectra over the three spatial directions of light polarization.\(^{5,34}\) Finally, in the calculation of the exchange term $\Sigma$, symmetries of the BZ can be exploited, which translates in a strong reduction of computational time.

In conclusion, we presented a double-grid method to solve the BSE on a coarse $k$-point grid, where the average of the strongly varying, but easily obtainable, independent-particle polarization is used. Converged spectra are reached for relatively small symmetric $k$-point grids. This allows for a considerably faster calculation of the BSE kernel. The single-particle energy bands in a dense $k$-point grid, the basic ingredient of our method, are not calculated directly, but are obtained through Wannier interpolation of the electronic band structure. As examples, we discussed the convergence of the absorption spectra of GaAs and Si with respect to the number of $k$ points. The speed-up is considerable, and opens the way for the solution of the BSE equation in large, complex systems.

D.K. and C.A. are financially supported by the Joseph Fourier university funding program for research (pôle Smingue). D.K. and M.A.L.M. acknowledge financial support from the French ANR (Grant No. ANR-08-CEC8-008-01). Computational resources were provided by GENCI (Project No. x2011096017).

---

1kammerlander.david@gmail.com


SPEEDING UP THE SOLUTION OF THE BETHE-...