Accelerating Optical Absorption Spectra and Exciton Energy Computation via Interpolative Separable Density Fitting

Wei Hu[1,2], Meiyue Shao[1], Andrea Cepellotti[3,4], Felipe H. da Jornada[3,4], Lin Lin[5,1], Kyle Thicke[6], Chao Yang[1], and Steven G. Louie[3,4]

[1]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States
{whu,myshao,cyang}@lbl.gov
[2]Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui 230026, China
{whuustc}@ustc.edu.cn
[3]Department of Physics, University of California, Berkeley Berkeley, California 94720, United States
{andrea.cepellotti,jornada,sglouie}@berkeley.edu
[4]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States
{andrea.cepellotti,jornada,sglouie}@berkeley.edu
[5]Department of Mathematics, University of California, Berkeley Berkeley, California 94720, United States
linlin@math.berkeley.edu
[6]Department of Mathematics, Duke University, Durham, NC 27708, United States
kyle.thicke@duke.edu

Abstract. We present an efficient way to solve the Bethe–Salpeter equation (BSE), a method for the computation of optical absorption spectra in molecules and solids that includes electron–hole interactions. Standard approaches to construct and diagonalize the Bethe–Salpeter Hamiltonian require at least $\mathcal{O}(N_e^5)$ operations, where N_e is the number of electrons in the system, limiting its application to smaller systems. Our approach is based on the interpolative separable density fitting (ISDF) technique to construct low rank approximations to the bare exchange and screened direct operators associated with the BSE Hamiltonian. This approach reduces the complexity of the Hamiltonian construction to $\mathcal{O}(N_e^3)$ with a much smaller pre-constant, and allows for a faster solution of the BSE. Here, we implement the ISDF method for BSE calculations within the Tamm–Dancoff approximation (TDA) in the BerkeleyGW software package. We show that this novel approach accurately reproduces exciton energies and optical absorption spectra in molecules and solids with a significantly reduced computational cost.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

1 Introduction

Many-Body Perturbation Theory is a powerful tool to describe one-particle and two-particle excitations and to obtain exciton energies and absorption spectra in molecules and solids. In particular, Hedin's GW approximation [8] has been successfully used to compute quasi-particle (one-particle) excitation energies [10]. However, the Bethe–Salpeter equation (BSE) [22] is further needed to describe the excitations of an electron–hole pair (a two-particle excitation) in optical absorption in molecules and solids [21] and is often necessary to obtain a good agreement between theory and experiment. Solving the BSE problem requires constructing and diagonalizing a structured matrix Hamiltonian. In the context of optical absorption, the eigenvalues are the exciton energies and the corresponding eigenfunctions yield the exciton wavefunctions.

The Bethe–Salpeter Hamiltonian (BSH) consists of bare exchange and screened direct interaction kernels that depend on single-particle orbitals obtained from a quasiparticle (usually at the GW level) or mean-field calculation. The evaluation of these kernels requires at least $\mathcal{O}(N_e^5)$ operations in a conventional approach, which is very costly for large systems that contain hundreds or thousands of atoms. Recent efforts have actively explored methods to generate a reduced basis set, in order to decrease the high computational cost of BSE calculations [1, 11, 15, 18, 20].

In this paper, we present an efficient way to construct the BSH, which, when coupled to an iterative diagonalization scheme, allows for an efficient solution of the BSE. Our approach is based on the recently-developed Interpolative Separable Density Fitting (ISDF) decomposition [17]. The ISDF decomposition has been applied to accelerate a number of applications in computational chemistry and materials science, including the computation of two-electrons integrals [17], correlation energy in the random phase approximation [16], density functional perturbation theory [14], and hybrid density functional calculations [9]. In this scheme, a matrix consisting of products of single-particle orbital pairs is approximated as the product between a matrix built with a small number of auxiliary basis vectors and an expansion coefficient matrix [9]. This decomposition effectively allows us to construct low-rank approximations to the bare exchange and screened direct kernels. The construction of the ISDF-compressed BSE Hamiltonian matrix only requires $\mathcal{O}(N_e^3)$ operations when the rank of the numerical auxiliary basis is kept at $\mathcal{O}(N_e)$ and when the kernels are kept in a low-rank factored form, resulting in considerably faster computation than the $\mathcal{O}(N_e^5)$ complexity required in a conventional approach. By keeping the interaction kernel in a decomposed form, the matrix–vector multiplications required in the iterative diagonalization procedures of the Hamiltonian H_{BSE} can be performed efficiently. We can further use these efficient matrix–vector multiplications in a structure preserving Lanczos algorithm [23] to obtain an approximate absorption spectrum without an explicit diagonalization of the approximate H_{BSE}. We have implemented the ISDF-based BSH construction in the BerkeleyGW software package [3], and verifyied that this approach can reproduce accurate exciton energies and optical absorption spectra for molecules and solids, while

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

significanreducing the computational cost associated with the construction of the BSE Hamiltonian.

2 Bethe–Salpeter equation

The Bethe–Salpeter equation is an eigenvalue problem of the form

$$H_{\mathrm{BSE}}X = EX\ , \tag{1}$$

where X is the exciton wavefunction, E the corresponding exciton energy. The Bethe–Salpeter Hamiltonian H_{BSE} has the following block structure

$$H_{\mathrm{BSE}} = \begin{bmatrix} D + 2V_A - W_A & 2V_B - W_B \\ -2\overline{V}_B + \overline{W}_B & -D - 2\overline{V}_A + \overline{W}_A \end{bmatrix}, \tag{2}$$

where $D(i_v i_c, j_v j_c) = (\epsilon_{i_c} - \epsilon_{i_v})\delta_{i_v j_c}\delta_{i_c j_c}$ is an $(N_v N_c) \times (N_v N_c)$ diagonal matrix with $-\epsilon_{i_v}$, $i_v = 1, 2, \ldots, N_v$ the quasi-particle energies associated with valence bands and ϵ_{i_c}, $i_c = N_v + 1, N_v + 2, \ldots, N_v + N_c$ the quasi-particle energies associated with conduction bands. These quasi-particle energies are typically obtained from a GW calculation [21]. The V_A and V_B matrices represent the bare *exchange* interaction of electron–hole pairs, and the W_A and W_B matrices are referred to as the screened *direct* interaction of electron–hole pairs. These matrices are defined as follows:

$$\begin{aligned} V_A(i_v i_c, j_v j_c) &= \int \bar{\psi}_{i_c}(\mathbf{r})\psi_{i_v}(\mathbf{r})V(\mathbf{r},\mathbf{r}')\bar{\psi}_{j_v}(\mathbf{r}')\psi_{j_c}(\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}', \\ V_B(i_v i_c, j_v j_c) &= \int \bar{\psi}_{i_c}(\mathbf{r})\psi_{i_v}(\mathbf{r})V(\mathbf{r},\mathbf{r}')\bar{\psi}_{j_c}(\mathbf{r}')\psi_{j_v}(\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}', \\ W_A(i_v i_c, j_v j_c) &= \int \bar{\psi}_{i_c}(\mathbf{r})\psi_{j_c}(\mathbf{r})W(\mathbf{r},\mathbf{r}')\bar{\psi}_{j_v}(\mathbf{r}')\psi_{i_v}(\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}', \\ W_B(i_v i_c, j_v j_c) &= \int \bar{\psi}_{i_c}(\mathbf{r})\psi_{j_v}(\mathbf{r})W(\mathbf{r},\mathbf{r}')\bar{\psi}_{j_c}(\mathbf{r}')\psi_{i_v}(\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}', \end{aligned} \tag{3}$$

where ψ_{i_v} and ψ_{i_c} are the valence and conduction single-particle orbitals typically obtained from a Kohn–Sham density functional theory (KSDFT) calculation respectively, and $V(\mathbf{r},\mathbf{r}')$ and $W(\mathbf{r},\mathbf{r}')$ are the bare and screened Coulomb interactions. Both V_A and W_A are Hermitian, whereas V_B and W_B are complex symmetric. Within the so-called Tamm–Dancoff approximation (TDA) [19], both V_B and W_B are neglected in Equation (2). In this case, the H_{BSE} becomes Hermitian and we can focus on computing the upper left block of H_{BSE}.

Let $M_{cc}(\mathbf{r}) = \{\psi_{i_c}\bar{\psi}_{j_c}\}$, $M_{vc}(\mathbf{r}) = \{\psi_{i_c}\bar{\psi}_{i_v}\}$, and $M_{vv}(\mathbf{r}) = \{\psi_{i_v}\bar{\psi}_{j_v}\}$ be matrices built as the product between orbital pairs in real space, and $\hat{M}_{cc}(\mathbf{G})$, $\hat{M}_{vc}(\mathbf{G})$, $\hat{M}_{vv}(\mathbf{G})$ be the reciprocal space representation of these matrices. Equations (3) can then be written succinctly as

$$V_A = \hat{M}_{vc}^{*}\hat{V}\hat{M}_{vc}\ ,\quad W_A = \mathrm{reshape}(\hat{M}_{cc}^{*}\hat{W}\hat{M}_{vv})\ , \tag{4}$$

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

where $\hat{V}$ and $\hat{W}$ are reciprocal space representations of the operators V and W respectively, and the reshape function is used to map the $(i_c j_c, i_v j_v)$th element on the right-hand side of (4) to the $(i_c i_v, j_c j_v)$th element of W_A. While in this paper we will focus, for simplicity, on the TDA model, we note that a similar set of equations can be derived for V_B and W_B.

The reason to compute the right-hand sides of (4) in the reciprocal space is that $\hat{V}$ is diagonal and an energy cutoff is often adopted to limit the number of the Fourier components of ψ_i. As a result, the leading dimension of $\hat{M}_{cc}$, $\hat{M}_{vc}$ and $\hat{M}_{cc}$, denoted by N_g, is often much smaller than that of M_{cc}, M_{vc} and M_{vv}, which we denote by N_r.

In addition to performing $\mathcal{O}(N_e^2)$ Fast Fourier transforms (FFTs) to obtain $\hat{M}_{cc}$, $\hat{M}_{vc}$ and $\hat{M}_{vv}$ from M_{cc}, M_{vc} and M_{vv}, respectively, we need to perform at least $\mathcal{O}(N_g N_c^2 N_v^2)$ floating-point operations to obtain V_A and W_A using matrix–matrix multiplications.

Note that, in order to achieve high accuracy with a large basis set, such as that of plane-waves, N_g is typically much larger than N_c or N_v. The number of occupied bands is either N_e or $N_e/2$ depending on how spin is counted. The number of conduction bands N_c included in the calculation is typically a small multiple of N_v (the precise number being a free parameter to be converged), whereas N_g is often as large as $100 - 10000 \times N_e$ ($N_r \sim 10 \times N_g$).

3 Interpolative separable density fitting (ISDF) decomposition

In order to reduce the computational complexity, we seek to minimize the number of integrals in Equation (3). To this aim, we rewrite the matrix M_{ij}, where the labels i and j are indices of either valence or conducting orbitals, as the product of a matrix Θ_{ij} that contains a set of N_{ij}^t linearly independent auxiliary basis vectors with $N_{ij}^t \approx tN_e \ll \mathcal{O}(N_e^2)$ (t is a small constant referred as a rank truncation parameter) [9] and an expansion coefficient matrix C_{ij}. For large problems, the number of columns of M_{ij} (i.e. $\mathcal{O}(N_v N_c)$, or $\mathcal{O}(N_v^2)$, or $\mathcal{O}(N_c^2)$) is typically larger than the number of grid points N_r on which $\psi_n(\mathbf{r})$ is sampled, i.e., the number of rows in M_{ij}. As a result, N_{ij}^t is much smaller than the number of columns of M_{ij}. Even when a cutoff is used to limit the size of N_c or N_v so that the number of columns in M_{ij} is much less than N_g, we can still approximate M_{ij} by $\Theta_{ij} C_{ij}$ with a Θ_{ij} that has a smaller rank $N_{ij}^t \sim t\sqrt{N_i N_j}$.

To simplify our discussion, let us drop the subscript of M, Θ and C for the moment, and describe the basic idea of ISDF. The optimal low rank approximation of M can be obtained from a singular value decomposition. However, the complexity of this decomposition is at least $\mathcal{O}(N_r^2 N_e^2)$ or $\mathcal{O}(N_e^4)$. Recently, an alternative decomposition has been developed, which is close to optimal but with a more favorable complexity. This type of decomposition is called Interpolative Separable Density Fitting (ISDF) [9], which we describe below.

In ISDF, instead of computing Θ and C simultaneously, we first fix the coefficient matrix C, and determine the auxiliary basis matrix Θ by solving a

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

linear least squares problem

$$\min \|M - \Theta C\|_F^2 \,, \tag{5}$$

where each column of M is given by $\psi_i(\mathbf{r})\bar{\psi}_j(\mathbf{r})$ sampled on a dense real space grids $\{\mathbf{r}_i\}_{i=1}^{N_r}$, and $\Theta = [\zeta_1, \zeta_2, \ldots, \zeta_{N^t}]$ contains the auxiliary basis vectors to be determined, $\|\cdot\|_F$ denotes the Frobenius norm.

We choose C as a matrix consisting of $\psi_i(\mathbf{r})\bar{\psi}_j(\mathbf{r})$ evaluated on a subset of N^t carefully chosen real space grid points, with $N^t \ll N_r$ and $N^t \ll N_e^2$, such that the (i,j)th column of C is given by

$$[\psi_i(\hat{\mathbf{r}}_1)\bar{\psi}_j(\hat{\mathbf{r}}_1), \cdots, \psi_i(\hat{\mathbf{r}}_k)\bar{\psi}_j(\hat{\mathbf{r}}_k), \cdots, \psi_i(\hat{\mathbf{r}}_{N^t})\bar{\psi}_j(\hat{\mathbf{r}}_{N^t})]^{\mathsf{T}}. \tag{6}$$

The least squares minimizer is given by

$$\Theta = MC^*(CC^*)^{-1}. \tag{7}$$

Because both multiplications in (7) can be carried out in $\mathcal{O}(N_e^3)$ due to the separable structure of M and C [9], the computational complexity for computing the interpolation vectors is $\mathcal{O}(N_e^3)$.

The interpolating points required in (6) can be selected by a permutation produced from a QR factorization of M^{T} with Column Pivoting (QRCP) [2]. In QRCP, we choose a permutation Π such that the factorization

$$M^{\mathsf{T}}\Pi = QR \tag{8}$$

yields a unitary matrix Q and an upper triangular matrix R with decreasing matrix elements along the diagonal of R. The magnitude of each diagonal element R indicates how important the corresponding column of the permuted M^{T} is, and whether the corresponding grid point should be chosen as an interpolation point. The QRCP decomposition can be terminated when the (N^t+1)-st diagonal element of R becomes less than a predetermined threshold, obtaining N^t leading columns of the permuted M^{T} that are, within numerical accuracy, maximally linearly independent. The corresponding grid points are chosen as the interpolation points. The indices for the chosen interpolation points $\hat{\mathbf{r}}_{N^t}$ can be obtained from indices of the nonzero entries of the first N^t columns of the permutation matrix Π. Notice that the standard QRCP procedure has a high computational cost of $\mathcal{O}(N_e^2 N_r^2) \sim \mathcal{O}(N_e^4)$, however, this cost can be reduced to $\mathcal{O}(N_r N_e^2) \sim \mathcal{O}(N_e^3)$ when QRCP is combined with the randomized sampling method [17].

4 Low rank representations of bare and screened operators via ISDF

The ISDF decomposition applied to M_{cc}, M_{vc} and M_{vv} yields

$$M_{cc} \approx \Theta_{cc}C_{cc}, \quad M_{vc} \approx \Theta_{vc}C_{vc}, \quad M_{vv} \approx \Theta_{vv}C_{vv}. \tag{9}$$

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

It follows from Equations (3), (4) and (9) that the exchange and direct terms of the BSE Hamiltonian can be written as

$$V_A = C_{vc}^* \widetilde{V}_A C_{vc}, \quad W_A = \text{reshape}(C_{cc}^* \widetilde{W}_A C_{vv}), \tag{10}$$

where $\widetilde{V}_A = \hat{\Theta}_{vc}^* \hat{V} \hat{\Theta}_{vc}$ and $\widetilde{W}_A = \hat{\Theta}_{cc}^* \hat{W} \hat{\Theta}_{vv}$ are the *projected* exchange and direct terms under the auxiliary basis $\hat{\Theta}_{vc}$, $\hat{\Theta}_{cc}$ and $\hat{\Theta}_{vv}$. Here, $\hat{\Theta}_{vc}$, $\hat{\Theta}_{cc}$ and $\hat{\Theta}_{vv}$ are reciprocal space representations of Θ_{vc}, Θ_{cc} and Θ_{vv}, respectively, that can be obtained via FFTs. Note that the dimension of the matrix $C_{cc}^* \widetilde{W}_A C_{cc}$ on the right-hand side of Equation (10) is $N_c^2 \times N_v^2$. Therefore, it needs to be reshaped into a matrix of dimension $N_v N_c \times N_v N_c$ according to the mapping $W_A(i_c j_c, i_v j_v) \to W_A(i_v i_c, j_v j_c)$ before it can be used in the BSH together with the V_A matrix.

Once the ISDF approximations for M_{vc}, M_{cc} and M_{vv} are available, the cost for constructing a low-rank approximation to the exchange and direct terms reduces to that of computing the projected exchange and direct kernels $\hat{\Theta}_{vc}^* \hat{V} \hat{\Theta}_{vc}$ and $\hat{\Theta}_{cc}^* \hat{W} \hat{\Theta}_{vv}$, respectively. If the ranks of Θ_{vc}, Θ_{cc} and Θ_{vv} are N_{vc}^t, N_{cc}^t and N_{vv}^t, respectively, then the computational complexity for computing the compressed exchange and direct kernels is $\mathcal{O}(N_{vc}^t N_{vc}^t N_g + N_{cc}^t N_{vv}^t N_g + N_{vv}^t N_g^2)$, which is significantly lower than the complexity of the conventional approach, which is $\mathcal{O}(N_g N_c^2 N_v^2)$. When $N_{vc}^t \sim t\sqrt{N_v N_c}$, $N_{cc}^t \sim t\sqrt{N_c N_c}$ and $N_{vv}^t \sim t\sqrt{N_v N_v}$ are on the order of N_e, the complexity of constructing the compressed kernels is $\mathcal{O}(N_e^3)$.

5 Iterative diagonalization of the BSE Hamiltonian

In the conventional approach, exciton energies and wavefunctions can be computed by using the recently developed BSEPACK library [24, 25] to diagonalize the BSE Hamiltonian H_{BSE}.

When ISDF is used to construct low-rank approximations to the bare exchange and screened direct operators V_A and W_A, we should keep both matrices in the factored form given by Equation (10). We propose to use iterative methods to diagonalize the approximate BSH constructed via the ISDF decomposition.

Within the TDA, several iterative methods such as the Lanczos [13] and LOBPCG [12] algorithms can be used to compute a few desired eigenvalues of the H_{BSE}. For each iterative step, we need to multiply H_{BSE} with a vector x of size $N_v N_c$. When V_A is kept in the factored form given by (10), $V_A x$ can be evaluated as three matrix vector multiplications performed in sequence, i.e.,

$$V_A x \leftarrow C_{vc}^* \left[\widetilde{V}_A (C_{vc} x) \right]. \tag{11}$$

The complexity of these calculations is $\mathcal{O}(N_v N_c N_{vc}^t)$. If N_{vc}^t is on the order of N_e, then each $V_A x$ can be carried out in $\mathcal{O}(N_e^3)$ operations.

Because $C_{cc}^* \widetilde{W}_A C_{vv}$ cannot be multiplied with a vector x of size $N_v N_c$ before it is reshaped, a different multiplication scheme must be used. It follows from

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

the separable nature of C_{vv} and C_{cc} that this multiplication can be succinctly written as

$$W_A x = \text{reshape}\left[\Psi_c^*\left(\widetilde{W} \odot (\Psi_c X \Psi_v^*)\right)\Psi_v\right], \tag{12}$$

where X is a $N_c \times N_v$ matrix reshaped from the vector x, Ψ_c is a $N_{cc}^t \times N_c$ matrix containing $\psi_{i_c}(\hat{r}_k)$ as its elements, Ψ_v is a $N_{vv}^t \times N_v$ matrix containing $\psi_{i_v}(\hat{r}_k)$ as its elements, and $\odot$ denotes componentwise multiplication (Hadamard product). The reshape function is used to turn the $N_c \times N_v$ matrix–matrix product back into a size $N_v N_c$ vector. If N_{vv}^t and N_{cc}^t are on the order of N_e, then all matrix–matrix multiplications in Equation (12) can be carried out in $\mathcal{O}(N_e^3)$ operations. In this way, each step of the iterative method has a complexity $\mathcal{O}(N_e^3)$ and, if the number of iterative steps required to reach convergence is small, the iterative diagonalization can be solved in $\mathcal{O}(N_e^3)$ operations.

6 Estimating optical absorption spectra without diagonalization

The optical absorption spectrum can be readily computed from the eigenpairs of H_{BSE} as

$$\varepsilon_2(\omega) = \text{Im}\left[\frac{8\pi e^2}{\Omega} d_r^*\left((\omega - \mathrm{i}\eta)I - H_{\text{BSE}}\right)^{-1} d_l\right], \tag{13}$$

where Ω is the volume of the primitive cell, e is the elementary charge, d_r and d_l are the right and left optical transition vectors, and η is a broadening factor used to account for the exciton lifetime.

To observe the absorption spectrum and identify its main peaks, it is possible to use a structure preserving iterative method instead of explicitly computing all eigenpairs of H_{BSE}, which has also been implemented in the BSEPACK [25] library. When TDA is adopted, the structure preserving Lanczos reduces to a standard Lanczos algorithm.

7 Numerical results

In this section, we demonstrate the accuracy and efficiency of the ISDF method when it is used to compute exciton energies and optical absorption spectrum in the BSE framework. We implemented the ISDF based BSH construction in the BerkeleyGW software package [3]. We use the *ab initio* software package Quantum ESPRESSO (QE) [5] to compute the ground-state quantities required in the GW and BSE calculations. We use Hartwigsen–Goedecker–Hutter (HGH) norm-conserving pseudopotentials [7] and the LDA [6] exchange–correlation functional in Quantum ESPRESSO. We also check these calculations in the KSSLOV software [26], which is a MATLAB toolbox for solving the Kohn-Sham equations. All the calculations were carried out on a single core at the Cori[1] systems at the National Energy Research Scientific Computing Center (NERSC).

[1] https://www.nersc.gov/systems/cori/

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

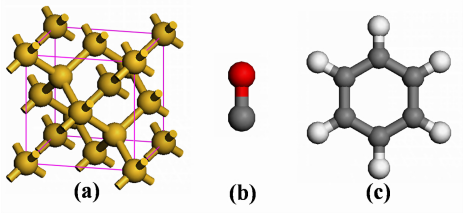

Fig. 1. Atomic structures of (a) a model silicon system Si_8, (b) carbon monoxide (CO) and (c) benzene (C_6H_6) molecules. The white, gray, red, and yellow balls denote hydrogen, carbon, oxygen, and silicon atoms, respectively.

Table 1. System size parameters for model silicon system Si_8, carbon monoxide (CO) and benzene (C_6H_6) molecules used for constructing corresponding BSE Hamiltonian H_{BSE}.

System	L (Bohr)	N_r	N_g	N_v	N_c	dim(H_{BSE})
Si_8	10.22	35937	2301	16	64	2048
CO	13.23	19683	1237	5	60	600
Benzene	22.67	91125	6235	15	60	1800

We performed calculations for three systems at the Gamma point. In particular, we choose a silicon Si_8 system as a typical model of bulk crystals (in the $\boldsymbol{k} = 0$ approximation, i.e. no sampling of the Brillouin zone) and two molecules: carbon monoxide (CO) and benzene (C_6H_6) as plotted in Fig. 1. All systems are closed shell systems, and the number of occupied bands is $N_v = N_e/2$, where N_e is the valence electrons in the system. We compute the quasiparticle energies and the dielectric function of CO and C_6H_6 in the BerkeleyGW [3], whereas for Si_8 in the KSSLOV [26].

7.1 Accuracy

We first measure the accuracy of the ISDF method by comparing the eigenvalues of the BSH computed with and without the ISDF decomposition.

In our test, we set the plane wave energy cutoff required in the QE calculations to $E_{\text{cut}} = 10$ Ha, which is relatively low. However, this is sufficient for assessing the effectiveness of ISDF. Such a choice of E_{cut} results in $N_r = 35937$ and $N_g = 2301$ for the Si_8 system in a cubic supercell of size 10.22 Bohr^3, $N_r = 19683$ and $N_g = 1237$ for the CO molecule ($N_v = 5$) in a cubic cell of size 13.23 Bohr, $N_r = 91125$ and $N_g = 6235$ for the benzene molecule in a cubic cell of size 22.67 Bohr. The number of active conduction bands (N_c) and valence bands (N_v), the number of reciprocal grids and the dimensions of the corresponding BSE Hamiltonian H_{BSE} for these three systems are listed in Table 1.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

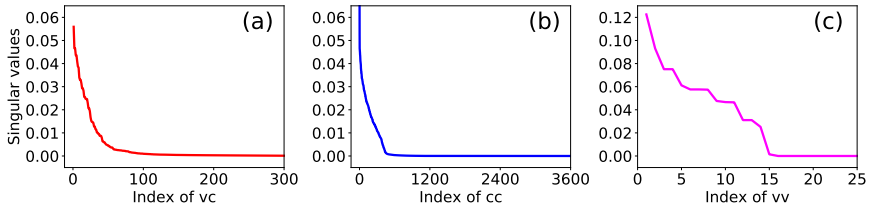

Fig. 2. The singular values of (a) $M_{vc}(\mathbf{r}) = \{\psi_{i_c}(\mathbf{r})\bar{\psi}_{i_v}(\mathbf{r})\}$ ($N_{vc} = 300$), (b) $M_{cc}(\mathbf{r}) = \{\psi_{i_c}(\mathbf{r})\bar{\psi}_{j_c}(\mathbf{r})\}$ ($N_{cc} = 3600$) and (c) $M_{vv}(\mathbf{r}) = \{\psi_{i_v}(\mathbf{r})\bar{\psi}_{j_v}(\mathbf{r})\}$ ($N_{vv} = 25$).

In Fig. 2, we plot the singular values of the matrices $M_{vc}(\mathbf{r}) = \{\psi_{i_c}(\mathbf{r})\bar{\psi}_{i_v}(\mathbf{r})\}$, $M_{cc}(\mathbf{r}) = \{\psi_{i_c}(\mathbf{r})\bar{\psi}_{j_c}(\mathbf{r})\}$ and $M_{vv}(\mathbf{r}) = \{\psi_{i_v}(\mathbf{r})\bar{\psi}_{j_v}(\mathbf{r})\}$ associated with the CO molecule. We observe that the singular values of these matrices decay rapidly. For example, the leading 500 (out of 3600) singular values of $M_{cc}(\mathbf{r})$ decreases rapidly towards zero. All other singular values are below 10^{-4}. Therefore, the numerical rank N_{cc}^t of M_{cc} is roughly 500 ($t = 8.3$), or roughly 15% of the number of columns in M_{cc}. Consequently, we expect that the rank of Θ_{cc} produced in ISDF decomposition can be set to 15% of N_c^2 without sacrificing the accuracy of the computed eigenvalues.

This prediction is confirmed in Fig. 3, where we plot the absolute difference between the lowest exciton energy of model silicon system Si_8 computed with and without using ISDF to construct H_{BSE}. To be specific, the error in the desired eigenvalue is computed as $\Delta E = E_{\text{ISDF}} - E_{\text{BGW}}$, where E_{ISDF} is computed from the H_{BSE} constructed with ISDF approximation, and E_{BGW} is computed from a standard H_{BSE} constructed without using ISDF. We first vary one of the ratios N_{cc}^t/N_{cc}, N_{vc}^t/N_{vc} and N_{vv}^t/N_{vv} while holding the others at a constant of 1. We observe that the error in the lowest exciton energy (positive eigenvalue) is around 10^{-3} Ha, when either N_{cc}^t/N_{cc} or N_{vc}^t/N_{vc} is set to 0.1 while the other ratios are held at 1. However, reducing N_{vv}^t/N_{vv} to 0.1 introduces a significant amount of error in the lowest exciton energy, likely because $N_v = 16$ is too small. We then hold N_{vv}^t/N_{vv} at 0.5 and let both N_{cc}^t/N_{cc} and N_{vc}^t/N_{vc} vary. The variation of ΔE with respect to these ratios is also plotted as in Fig. 3. We observe that the error in the lowest exciton energy is still around 10^{-3} Ha even when both N_{cc}^t/N_{cc} and N_{vc}^t/N_{vc} are set to 0.1.

We then check the absolute error ΔE (Ha) of all the exciton energies computed with the ISDF method by comparing them with the ones obtained from a conventional BSE calculation implemented in BerkeleyGW for the CO and benzene molecules. As we can see from Fig. 4, the errors associated with these eigenvalues are all below 0.002 Ha when N_{cc}^t/N_{cc} is 0.1.

7.2 Efficiency

At the moment, our preliminary implementation of the ISDF method within the BerkeleyGW software package is sequential. Therefore, our efficiency test is

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

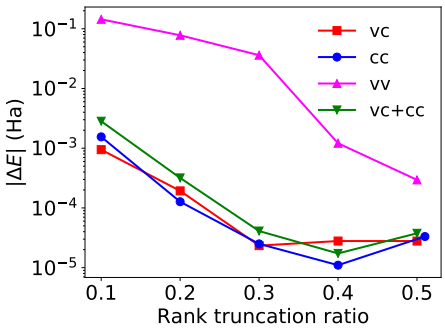

Fig. 3. The change of absolute error ΔE in the smallest eigenvalue of H_{BSE} associated with the Si_8 system with respect to different truncation levels used in ISDF approximation of M_{vc}, M_{cc} and M_{vv}. The curves labeled by 'vc', 'cc', 'vv' correspond to calculations in which only one of the ratios N_{vc}^t/N_{vc}, N_{cc}^t/N_{cc} and N_{vv}^t/N_{vv} changes while all other parameters are held constant. The curve labeled by 'vc + cc' corresponds to the calculation in which both N_{vc}^t/N_{vc} and N_{cc}^t/N_{cc} change at the same rate ($N_{vv}^t = N_{vv}$).

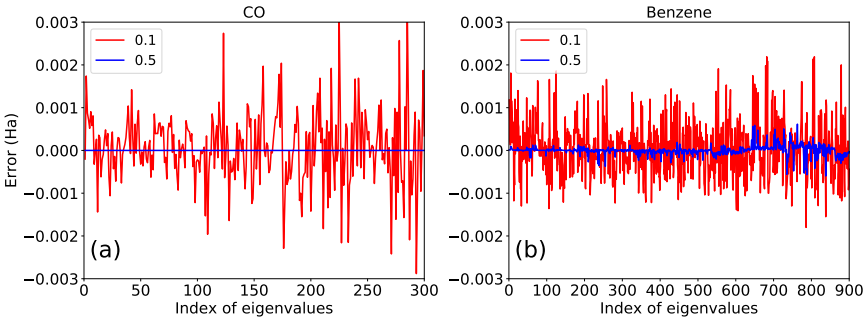

Fig. 4. Error in all eigenvalues of the BSH associated with the (a) CO and (b) benzene molecules. Two rank truncation ratios $N_{cc}^t/N_{cc} = 0.5$ ($t = 30.0$) and $N_{cc}^t/N_{cc} = 0.1$ ($t = 6.0$) are used in the tests.

limited by the size of the problem as well as the number of conducting bands (N_c) we can include in the bare and screened operators. As a result, our performance measurement does not fully reflect the computational complexity analysis presented in the previous sections. In particular, taking benzene as an example, $N_g = 6235$ is much larger than $N_v = 15$ and $N_c = 60$, therefore the computational cost of $N_g^2 N_v^2 \sim \mathcal{O}(N_e^4)$ term is much higher than the $N_g N_v^2 N_c^2 \sim \mathcal{O}(N_e^5)$ term in the conventional BSE calculations.

Nonetheless, in this section, we will demonstrate the benefit of using ISDF to reduce the cost for constructing the BSE Hamiltonian H_{BSE}. In Table 2, we focus on the benzene example and report the wall-clock time required to construct

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

Table 2. The variation of time required to carry out the ISDF decomposition of M_{vc}, M_{vv} and M_{cc} with respect to rank truncation ratio for the benzene molecule.

Rank truncation ratio			Time (s) for $M_{ij}(\mathbf{r})$		
N_{vc}^t/N_{vc}	N_{vv}^t/N_{vv}	N_{cc}^t/N_{cc}	M_{vc}	M_{vv}	M_{cc}
1.0	0.5	0.5	157.0	5.8	578.9
1.0	0.5	0.1	157.0	5.8	34.3
0.1	0.1	0.1	4.3	0.7	34.3

the ISDF approximations of the M_{vc}, M_{cc}, and M_{vv} matrices at different rank truncation levels. Without using ISDF, it takes 746.0 seconds to construct the reciprocal space representations of M_{vc}, M_{cc}, and M_{vv} in BerkeleyGW. Most of the time is spent in the several FFTs applied to M_{vc}, M_{cc}, and M_{vv}, in order to obtain the reciprocal space representation of these matrices. We can clearly see that by reducing N_{cc}^t/N_{cc} from 0.5 ($t = 30.0$) to 0.1 ($t = 6.0$), the wall-clock time used to construct the low-rank approximation to M_{cc} reduces from 578.9 to 34.3 seconds. Furthermore, the total cost of computing M_{vc}, M_{cc} and M_{vv} is reduced by a factor 19 when compared with the cost of a conventional approach (39.3 vs. 746.0 seconds) if N_{vc}^t/N_{vc}, N_{vv}^t/N_{vv} and N_{cc}^t/N_{cc} are all set to 0.1.

Since the ISDF decomposition is carried out on a real-space grid, most of the time is spent in performing the QRCP in real space. Even though QRCP with random sampling has $\mathcal{O}(N_e^3)$ complexity, it has a relatively large pre-constant compared to the size of the problem. This cost can be further reduced by using the recently proposed centroidal Voronoi tessellation (CVT) method [4].

In Table 3, we report the wall-clock time required to construct the projected exchange and direct matrices $\widetilde{V}_A$ and $\widetilde{W}_A$ that appear in Equation (10) from the ISDF approximations of M_{vc}, M_{vv}, and M_{cc}. The current implementation in BerkeleyGW requires 103,154 seconds (28.65 hours) in a serial run for the full construction of H_{BSE}. In the present reimplementation, without ISDF, it takes $1.574 + 4.198 = 5.772$ seconds to construct both W_A and V_A. Note that the original implementation in BerkeleyGW is much slower as it requires a complete integration over G vectors for each pair of bands. When N_{cc}^t/N_{cc} is set to 0.1, the cost for constructing the full W_A, which has the largest complexity, is reduced by a factor 2.8. Furthermore, if N_{vc}^t/N_{vc}, N_{vv}^t/N_{vv} and N_{cc}^t/N_{cc} are all set to 0.1, we reduce the cost for constructing $\widetilde{V}_A$ and $\widetilde{W}_A$ by a factor of 63.0 and 10.1 respectively.

7.3 Optical absorption spectra

One important application of BSE is to compute the optical absorption spectrum, which is determined by optical dielectric function in Equation (13). Fig. 5 plots the optical absorption spectra for both CO and benzene obtained from approximate H_{BSE} constructed with the ISDF method and the H_{BSE} constructed in a conventional approach implemented in BerkeleyGW. When the rank trun-

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

Table 3. The variation of time required to construct the projected bare and screened matrices $\widetilde{V}_A$ and $\widetilde{W}_A$ exhibited by the ISDF method respect to rank truncation ratio for the benzene molecule.

Rank truncation ratio			Time (s) for H_{BSE}	
N_{vc}^t/N_{vc}	N_{vv}^t/N_{vv}	N_{cc}^t/N_{cc}	$\widetilde{V}_A$	$\widetilde{W}_A$
1.0	1.0	1.0	1.574	4.198
1.0	0.5	0.1	1.574	1.474
0.1	0.1	0.1	0.025	0.414

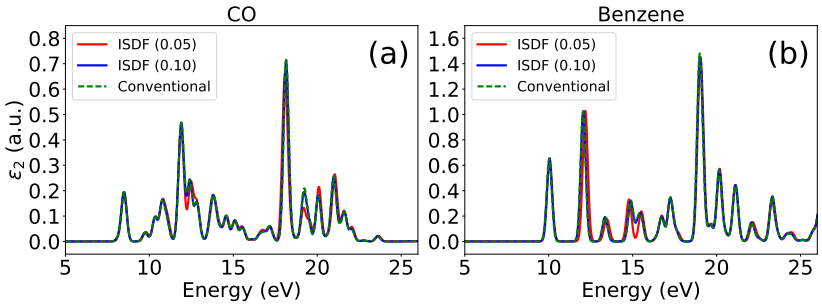

Fig. 5. Optical dielectric function (imaginary part ε_2) of (a) CO and (b) benzene molecules computed with the ISDF method (the rank ratio N_{cc}^t/N_{cc} is set to be 0.05 ($t = 3.0$) and 0.10 ($t = 6.0$)) compared to conventional BSE calculations in BerkeleyGW.

cation ratio N_{cc}^t/N_{cc} is set to be only 0.10 ($t = 6.0$), the absorption spectrum obtained from the ISDF approximate H_{BSE} is nearly indistinguishable from that produced from the conventional approach. When N_{cc}^t/N_{cc} is set to 0.05 ($t = 3.0$), the absorption spectrum obtained from ISDF approximate H_{BSE} still preserves the main features (peaks) of the absorption spectrum obtained in a conventional approach even though some of the peaks are slightly shifted, and the height of some peaks are slightly off.

8 Conclusion and outlook

In summary, we have demonstrated that the interpolative separable density fitting (ISDF) technique can be used to efficiently and accurately construct the Bethe–Salpeter Hamiltonian matrix. The ISDF method allows us to reduce the complexity of the Hamiltonian construction from $\mathcal{O}(N_e^5)$ to $\mathcal{O}(N_e^3)$ with a much smaller pre-constant. We show that the ISDF based BSE calculations in molecules and solids can efficiently produce accurate exciton energies and optical absorption spectrum in molecules and solids.

In the future, we plan to replace the costly QRCP procedure with the centroidal Voronoi tessellation (CVT) method [4] for selecting the interpolation

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

points in the ISDF method. The CVT method is expected to significantly reduce the computational cost for selecting interpolating point in the ISDF procedure for the BSE calculations.

The performance results reported here are based on a sequential implementation of the ISDF method. In the near future, we will implement a parallel version suitable for large-scale distributed memory parallel computers. Such an implementation will allow us to tackle much larger problems for which the favorable scaling of the ISDF approach will be more pronounced.

Acknowledgments

This work is supported by the Center for Computational Study of Excited-State Phenomena in Energy Materials (C2SEPEM) at the Lawrence Berkeley National Laboratory, which is funded by the U. S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract No. DE-AC02-05CH11231, as part of the Computational Materials Sciences Program, which provided support for developing, implementing and testing ISDF for BSE in BerkeleyGW. The Center for Applied Mathematics for Energy Research Applications (CAMERA) (L. L. and C. Y.) provided support for the algorithm development and mathematical analysis of ISDF. Finally, the authors acknowledge the computational resources of the National Energy Research Scientific Computing (NERSC) center.

References

1. P. Benner, S. Dolgov, V. Khoromskaia, and B. N. Khoromskij. Fast iterative solution of the Bethe–Salpeter eigenvalue problem using low-rank and QTT tensor approximation. *J. Comput. Phys.*, 334:221–239, 2017.
2. T. F. Chan and P. C. Hansen. Some applications of the rank revealing QR factorization. *SIAM J. Sci. Statist. Comput.*, 13:727–741, 1992.
3. J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie. BerkeleyGW: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures. *Comput. Phys. Commun.*, 183(6):1269–1289, 2012.
4. K. Dong, W. Hu, and L. Lin. Interpolative separable density fitting through centroidal Voronoi tessellation with applications to hybrid functional electronic structure calculations, 2017. arXiv:1711.01531.
5. P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502, 2009.
6. S. Goedecker, M. Teter, and J. Hutter. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B*, 54:1703, 1996.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48

7. C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space gaussian pseudopotentials from H to Rn. *Phys. Rev. B*, 58:3641, 1998.
8. L. Hedin. New method for calculating the one-particle Green's function with application to the electron–gas problem. *Phys. Rev.*, 139:A796, 1965.
9. W. Hu, L. Lin, and C. Yang. Interpolative separable density fitting decomposition for accelerating hybrid density functional calculations with applications to defects in silicon. *J. Chem. Theory Comput.*, 13(11):5420–5431, 2017.
10. M. S. Hybertsen and S. G. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B*, 34:5390, 1986.
11. P. B. V. Khoromskaia and B. N. Khoromskij. A reduced basis approach for calculation of the Bethe–Salpeter excitation energies by using low-rank tensor factorisations. *Mol. Phys.*, 114:1148–1161, 2016.
12. A. V. Knyazev. Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method. *SIAM J. Sci. Comput.*, 23(2):517–541, 2001.
13. C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards*, 45:255–282, 1950.
14. L. Lin, Z. Xu, and L. Ying. Adaptively compressed polarizability operator for accelerating large scale ab initio phonon calculations. *Multiscale Model. Simul.*, 15:29–55, 2017.
15. M. P. Ljungberg, P. Koval, F. Ferrari, D. Foerster, and D. Sánchez-Portal. Cubic-scaling iterative solution of the Bethe–Salpeter equation for finite systems. *Phys. Rev. B*, 92:075422, 2015.
16. J. Lu and K. Thicke. Cubic scaling algorithms for RPA correlation using interpolative separable density fitting. *J. Comput. Phys.*, 351:187–202, 2017.
17. J. Lu and L. Ying. Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *J. Comput. Phys.*, 302:329–335, 2015.
18. M. Marsili, E. Mosconi, F. D. Angelis, and P. Umari. Large-scale GW-BSE calculations with N^3 scaling: Excitonic effects in dye-sensitized solar cells. *Phys. Rev. B*, 95:075415, 2017.
19. G. Onida, L. Reining, and A. Rubio. Electronic excitations: Density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.*, 74:601, 2002.
20. D. Rocca, D. Lu, and G. Galli. Ab initio calculations of optical absorption spectra: Solution of the Bethe–Salpeter equation within density matrix perturbation theory. *J. Chem. Phys.*, 133:164109, 2010.
21. M. Rohlfing and S. G. Louie. Electron–hole excitations and optical spectra from first principles. *Phys. Rev. B*, 62:4927, 2000.
22. E. E. Salpeter and H. A. Bethe. A relativistic equation for bound-state problems. *Phys. Rev.*, 84:1232, 1951.
23. M. Shao, F. H. da Jornada, L. Lin, C. Yang, J. Deslippe, and S. G. Louie. A structure preserving Lanczos algorithm for computing the optical absorption spectrum. *SIAM J. Matrix. Anal. Appl.*, to appear.
24. M. Shao, F. H. da Jornada, C. Yang, J. Deslippe, and S. G. Louie. Structure preserving parallel algorithms for solving the Bethe–Salpeter eigenvalue problem. *Linear Algebra Appl.*, 488:148–167, 2016.
25. M. Shao and C. Yang. BSEPACK user's guide, 2016. https://sites.google.com/a/lbl.gov/bsepack/.
26. C. Yang, J. C. Meza, B. Lee, and L.-W. Wang. KSSOLV—a MATLAB toolbox for solving the Kohn–Sham equations. *ACM Trans. Math. Softw.*, 36:1–35, 2009.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_48