Supporting Information for Optical excitations of polyacenes and bacteriochlorophyll-A using a GW and Bethe-Salpeter equation approach in the Exciton code

C. H. Patterson*

School of Physics, Trinity College Dublin, Dublin 2, Ireland

E-mail: Charles.Patterson@tcd.ie

In Supporting Information computational details of calculations are given, including dimensions of active spaces and basis sets used. This is followed by a brief description of the implementation of the GW and BSE methods in the Exciton code and an analysis of scaling of times taken in various parts of BSE calculations. Finally, the algorithm implemented for solution of the BSE generalized eigenvalue problem is outlined.

1 Details of Calculations

1.1 Basis sets

Molecular geometries for all species used in this work were obtained by energy minimization using a B3LYP hybrid DFT method and a 6-311G* Gaussian orbital basis[5] and the Crystal code.[6] The cc-pVTZ basis sets for H and C in the acenes and the cc-pVDZ basis for H, C, N, O and Mg in BChla are listed in Ref. [7]. The performance of corresponding cc-pVDZ-RI and cc-pVTZ-RI resolution of the identity auxiliary basis sets is described in Ref. [4].

1.2 Excitation energy convergence with active space dimension

The size of the active space and the dimension of the Hamiltonian matrix in BSE-TDA calculations is the number of transitions, which is the number of occupied states times the number of virtual states included in the active space. In BSE calculations the dimension of the Hamiltonian matrix is twice the dimension of the active space. For relatively small systems the dimension of the active space remains reasonably small, even if all valence occupied and virtual states are included in the active space. For example, for tetracene with a cc-pVTZ basis set for C and H, the number of occupied valence states is 42 and the number of virtual states is 648, resulting in an active space of dimension 27216, which can be diagonalized in about 400 s on 48 cores using a call to *pdsyev* in Scalapack.[2]

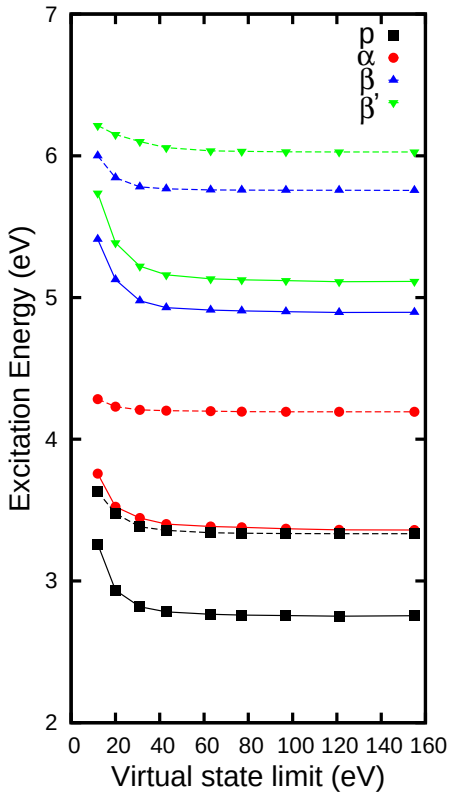

Figure 1: Lowest optical excitation energies (p, α, β, β') in tetracene as a function of virtual state cutoff energy using TDHF-TDA (dashed lines) and BSE-TDA (solid lines) methods.

For large systems such as BChla, however, the total number of occupied valence states and virtual states becomes very large so that limiting the dimension of the active space while

achieving convergence of excited state properties is essential. For BChla with a cc-pVDZ basis set for all atoms, there are 166 occupied valence states and 1040 virtual states, resulting in an active space dimension of 172640.

Setting up the Hamiltonian for TDHF-TDA calculations is fast, by comparison to diagonalization time and, especially, to calculating the screened interaction. Convergence of TDHF-TDA calculations with respect to active space dimension is similar to that for calculations which include expensive screened Coulomb interactions and hence they can be used to check whether low energy excitation energies have converged with respect to active space dimension.

Fig. 1 shows the dependence of excitation energies from BSE-TDA and TDHF-TDA calculations on tetracene on the cutoff energy used in selecting the number of virtual states in a calculation. A calculation for tetracene which includes all 42 valence states and 190 virtual states (cutoff energy around 40 eV) has a Hamiltonian dimension of 42 x 190 = 7980. Inclusion of all 648 virtual states in the cc-pVTZ basis for tetracene leads to a corresponding matrix dimension of 27216. Fig. 1 shows that excitation energies converge to within 10 meV of the full cc-pVTZ basis set value by including virtual states with energies up to 60 eV, with a corresponding matrix dimension of order 10000. TDHF-TDA and BSE-TDA excitation energies converge with a similar number of virtual states, but the variation in excitation energy with the number of virtual states for TDHF calculations is smaller than for BSE-TDA calculations.

For BChla calculations in this work, all 166 valence states and 504 virtual states were included in the active space, resulting in Hamiltonian dimensions of 83664. Virtual state energy cutoffs correspond to highest HF energy levels of 42 eV.

1.3 Excitation energy convergence with number of RPA modes

The most expensive single step in a BSE calculation is generation of matrix elements of the screened Coulomb interaction (Eq. 14). A significant saving in computer time can be

achieved by including only lower energy modes which contribute significantly to electron-hole screening by truncating the sum on index α in Eqs. 14 and 17. The maximum possible number of RPA modes is the number of virtual MO states times the number of occupied MO states included in the active space.

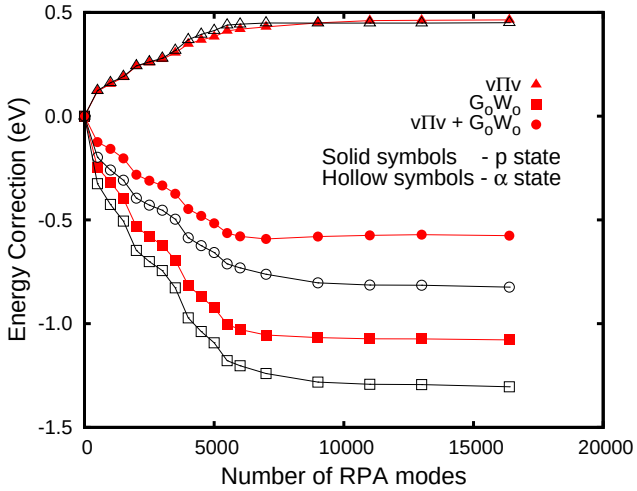

Figure 2: Lowest optical excitation energies (p, α) in tetracene as a function of the number of RPA modes included in the self-energy correction to single-particle excitation energies and screening of the electron-hole attraction. 42 occupied states and 390 virtual states were included in each calculation, corresponding to a Hamiltonian matrix dimension of 16380 and a virtual state limit of about 100 eV. Energy corrections are given relative to the predicted TDHF-TDA energies for that Hamiltonian matrix size.

A BSE-TDA calculation in which the dynamic part of the screened interaction, W_d, is switched off is equivalent to a TDHF-TDA calculation. Hence, by performing BSE-TDA calculations where the number of RPA modes included in the screened interaction is raised from zero to the full set of RPA modes, excitation energies will change from the TDHF-TDA excitation energies to full screened BSE-TDA excitation energies. Fig. 2 shows plots of BSE-TDA excitation energes relative to TDHF-TDA excitation energies for the p and α excitations of tetracene as a function of the number of RPA modes used. The screened interaction appears in both the self-energy (Eq. 17) and in Coulomb matrix elements in **A** and **B** blocks of the Hamiltonian (Table 1). In the former case, HF energy eigenvalue differences $(\epsilon_a - \epsilon_i)$ are replaced by G_oW_o energy eigenvalue differences, $(\widetilde{\epsilon_a} - \widetilde{\epsilon_i})$, which leads to the reduction in excitation energies shown as square symbols in Fig. 2, which exceed 1

eV when converged with respect to the number of RPA modes. In the latter case, including W_d in Coulomb matrix elements results in increases of excitation energy as bare electron-hole matrix elements in TDHF-TDA (ab|ji) become screened (ab|W_d|ji). These increases are shown as triangles in Fig. 2. p and α excitation energies shift upwards by just under 0.5 eV. Finally, the combined effects of changes in energy eigenvalue differences and changes in electron-hole matrix element screening are shown as circles. Convergence of the BSE-TDA excitation energies with number of RPA modes when both G_oW_o and $v\Pi^{RPA}v$ terms are included is achieved with around 10000 RPA modes, with an energy cutoff around 50 eV. Limiting the number of RPA modes and the dimension of the excitation space considerably reduces computer time and memory requirements.

2 Hamiltonian Assembly

2.1 Coulomb integrals via resolution of the identity

The approach to calculating Coulomb integrals known as 'resolution of the identity' (RI), or density fitting, introduces an auxiliary basis set, denoted $\{\chi\}$, in addition to the usual atomic orbital wave function basis set, $\{\phi\}$. A product of wave function basis functions, $\Phi_{mn} = \phi_m(\mathbf{r})\phi_n^*(\mathbf{r})$, is expanded in the auxiliary basis as,

$$\Phi_{mn}(\mathbf{r}) = c_\mu^{mn}\chi_\mu(\mathbf{r}). \tag{1}$$

If the expansion coefficients are constrained to minimize the Coulomb integral,

$$\int d\mathbf{r}d\mathbf{r}'(\Phi_{mn}^*(\mathbf{r}) - c_\mu^{*mn}\chi_\mu^*(\mathbf{r}))|\frac{1}{|\mathbf{r}-\mathbf{r}'|}|\ (\Phi_{rs}(\mathbf{r}') - c_\nu^{rs}\chi_\nu(\mathbf{r}')), \tag{2}$$

then the error in the integral,

$$\int d\mathbf{r}d\mathbf{r}'\Phi_{mn}^*(\mathbf{r})|\frac{1}{|\mathbf{r}-\mathbf{r}'|}|\Phi_{rs}(\mathbf{r}'), \tag{3}$$

is minimized and the expansion coefficients are,[4]

$$c_{\mu}^{*mn} = (mn|\nu)(\nu|\mu)^{-1} \tag{4}$$

where

$$\begin{aligned}(mn|\nu) &= \int d\mathbf{r}d\mathbf{r}'\Phi_{mn}^{*}(\mathbf{r})|\tfrac{1}{|\mathbf{r}-\mathbf{r}'|}|\chi_{\nu}(\mathbf{r}') \\ (\mu|\nu) &= \int d\mathbf{r}d\mathbf{r}'\chi_{\mu}^{*}(\mathbf{r})|\tfrac{1}{|\mathbf{r}-\mathbf{r}'|}|\chi_{\nu}(\mathbf{r}').\end{aligned}$$

Three-center Coulomb integrals over the atomic orbital wave function basis, (mn|μ), are transformed to integrals over the MO basis by multiplying (mn|μ) by MO expansion coefficients, c_m^i, (ia|μ) = $c_m^{i*}c_n^a$ (mn|μ). An (ia|jb) integral over MOs ψ_i, etc. is computed from $(ia|\mu)$ and $(\mu|\nu^{-1}(\nu|jb)$ factors by contracting over index μ,

$$(ia|jb) \approx (ia|\mu)(\mu|\nu)^{-1}(\nu|jb). \tag{5}$$

2.2 Three-center integral distribution and scaling

One of the advantages of a RI approach in many-body computations compared to a conventional four-center integral approach is that transformation of integrals from the contracted Gaussian orbital basis to the MO basis is reduced from N^4 to N^2 in the number of MOs. This comes at the expense of computing a large number of three-center Coulomb integrals over the auxiliary and wave function bases. The largest basis sets used in this work had of order 670 MOs and 5000 auxiliary basis functions, resulting in of order 2.2 x 10^9 (ia|jb) and (ij|ab) integrals, which occupy around 18 GB of memory and are distributed over cores. $(\mu|\nu)^{-1}(\nu|jb)$ factors occupy a further 18 GB of memory. In contrast, there would be 2.0 x 10^{11} transformed four center integrals occupying 1.6 TB of memory in the conventional four-center integral approach.

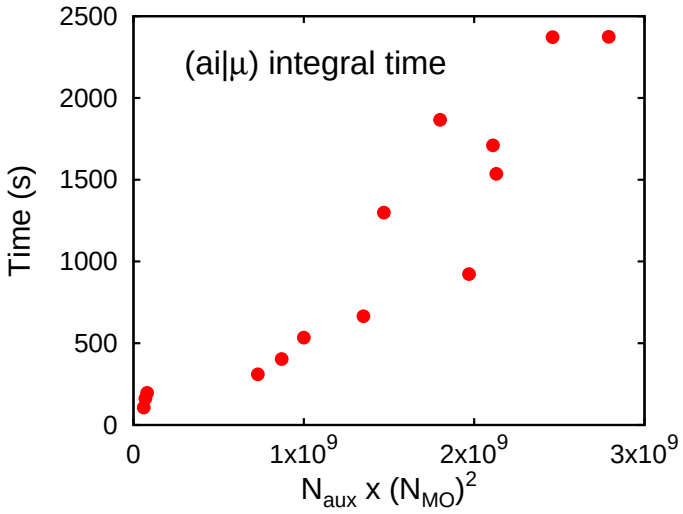

Figure 3: Time taken to compute 3-center Coulomb integrals and 3-center Coulomb integrals times the inverted Coulomb matrix in seconds over N_{aux} auxiliary basis functions and N_{MO} molecular orbitals.

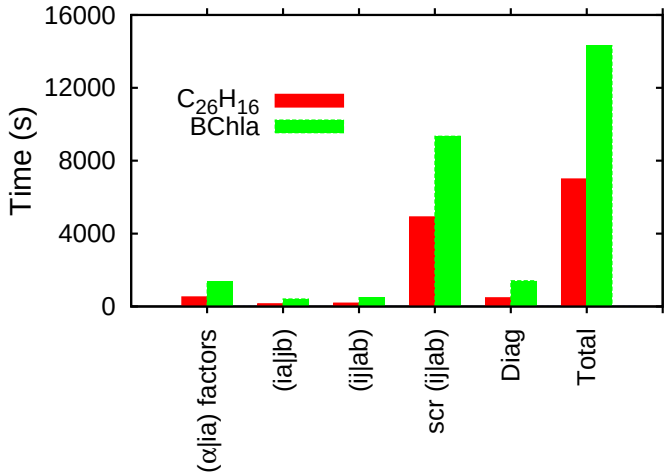

Figure 4: Time taken in various stages of BSE-TDA calculations on hexacene and BChla molecules. The columns correspond to the time take to compute $(ia|\mu)$ and $(\mu|\nu)^{-1}(\nu|jb)$ factors, times to build (ia|jb), (ij|ab) and screened (ij|ab) parts of the Hamiltonian, Scalapack diagonalization time and the total time taken for the calculation.

The (ia|jb) integral in Eq. 5 is computed from $(ia|\mu)$ and $(\mu|\nu)^{-1}(\nu|jb)$ factors, so that (ia|jb) is calculated by contracting over index μ. The relatively small two center Coulomb matrix, $(\mu|\nu)$, is computed and inverted on a single core before being broadcast to all cores. Three center integrals over Gaussian orbitals are distributed so that all integrals with an auxiliary basis function index located on a particular atomic center are generated and stored on the same core. Multiplication of $(\mu|\nu)^{-1}$ by $(\nu|\text{jb})$ requires message passing in forming the sum over index ν. Contraction over the index μ does not require any further message passing. If fewer cores than atomic centers are available then more than one index μ is stored on a given core. Fig. 3 shows scaling of the time taken to compute three-center integrals over MOs, including message passing for three-center integrals times the inverse of the two-center Coulomb matrix (i.e. the $(ia|\mu)$ and $(\mu|\nu)^{-1}(\nu|jb)$ factors in Eq. 5), for tetracene to hexacene and BChla.

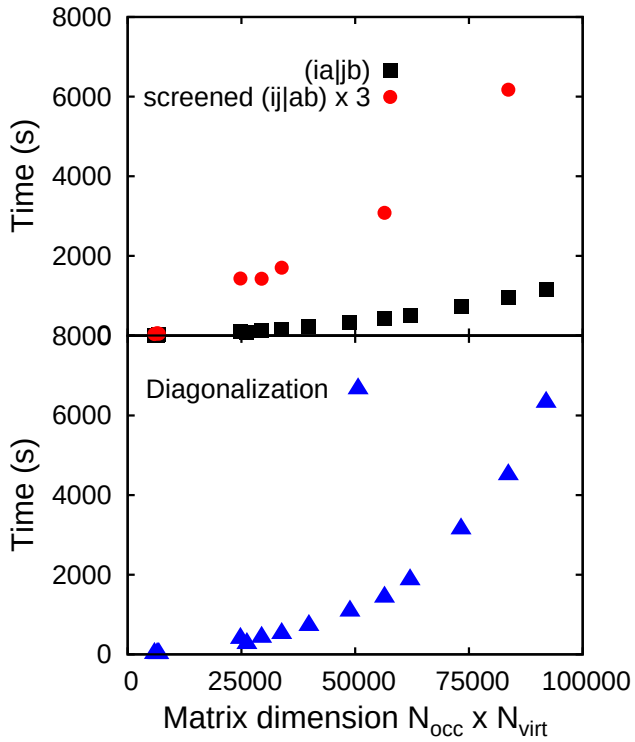

Figure 5: (upper panel) Time required to build various parts of the Hamiltonian in seconds for tetracene to hexacene and BChla examples. (lower panel) Time required to diagonalize the Hamiltonian in seconds and to write RPA calculation eigenvectors to an MPI file.

Each matrix element of the Hamiltonian requires summation of all contractions over

auxiliary basis function indices, μ, which are distributed over cores. For unscreened matrix elements, (ia|jb) and (ij|ab), this is achieved by contracting over all auxiliary basis functions on each core followed by an MPI_Reduce call in which matrix elements are summed in the Hamiltonian matrix. This is a relatively fast part of Hamiltonian building. Fig. 4 is a histogram showing breakdown of the total time taken in integral generation, Hamiltonian building and diagonalization using Scalapack in BSE-TDA calculations for hexacene and BChla. Assembling the unscreened (ia|jb) and (ij|ab) matrix elements is the fastest step, followed by three-center integral generation.

Fig. 4 also shows that the screened electron-hole interaction take by far the longest of any part of a BSE or BSE-TDA calculation. This is because of the dependence of the screened interaction on products of Coulomb integrals. Matrix elements w_{ai}^{α} in Eq. 15 are computed in batches over index α and require a call to MPI_Allreduce, so that each core has a copy of w_{ai}^{α}.

Scaling of the time taken to contract (ia|jb), (ij|ab) and screened (ij|ab) matrix elements over auxiliary basis functions is shown in the top panel of Fig. 5 using tetracene to hexacene and BChla as examples. The total number of matrix elements of each type to be calculated is the square of the number of transitions, N_{occ} x N_{virt}. The figure shows that both types of unscreened matrix elements scale linearly with matrix dimension for the examples chosen. Scaling of the diagonalization time using the *pdsyev* library routine in Scalapack for these example systems is shown in the lower panel of Fig. 5. The time taken is linear in the matrix dimension.

Scalapack parallel routines require matrices to be in block-cyclic format[2] and all matrix elements to be present (not only lower or upper triangular blocks of a symmetric or Hermitian matrix). Compute cores available to a particular job are divided into a m x n processor array. For the systems studied here 168 cores were in a 12 x 14 array. This meant that the auxiliary basis sets in the BChla system with 140 atoms could be arranged with at most one atomic center per core. Each core contains many small Hamiltonian blocks which are not contiguous

in the full matrix which is in conventional array index (linear) format. The block-cyclic array format is used to build the part of the Hamiltonian stored on each core before the call to Scalapack *pdsyev*. Eigenvectors are returned in the same block-cyclic format.

Block-cyclic format eigenvectors are written to an MPI file in linear format. They are used in this format in building the screened electron-hole exchange part of the Hamiltonian. The time taken to write this large file (of the order of 20 GB) is shown as I/O in the lower panel of Fig. 5. In a future version of the code, RPA eigenvectors could be distributed between nodes by message passing rather than MPI I/O, which would reduce the time taken by this step, but would require an additional 20GB of memory over all cores in the processor array.

3 Solution of generalized eigenvalue problem

Recently, Shao and coworkers[1] reported an implementation of parallel algorithms for solving the BSE using matrices of the dimension of the **A** block of the particle-hole Hamiltonian, **H**, i.e. half of the full BSE matrix dimension, using the Scalapack library.[2] The algorithm they describe for real Hamiltonians is also implemented in the Exciton code used in this work and it is briefly restated here. It is based on the observation that solving the generalized eigenvalue problem for the matrix $\mathbf{C}^{-1}\mathbf{H}$ is similar to the matrix

$$\begin{pmatrix} \mathbf{0} & \mathbf{A}+\mathbf{B} \\ \mathbf{A}-\mathbf{B} & \mathbf{0} \end{pmatrix}, \tag{6}$$

under an orthogonal similarity transformation, so that eigenvalues of $\mathbf{C}^{-1}\mathbf{H}$ are square roots of the eigenvalues of the matrix $(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B})$. In turn, $(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B})$ is similar to the symmetric matrix $\mathbf{L}_2^T(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B})\mathbf{L}_2^{-T} = \mathbf{L}_2^T\mathbf{L}_1(\mathbf{L}_2^T\mathbf{L}_1)^T$ where,

$$\mathbf{A}+\mathbf{B} = \mathbf{L}_1\mathbf{L}_1^T$$

$$\mathbf{A}-\mathbf{B}=\mathbf{L}_2\mathbf{L}_2^T$$

are the Cholesky decompositions of $(\mathbf{A}+\mathbf{B})$ and $(\mathbf{A}-\mathbf{B})$, respectively. The key to the reduction in dimension of the BSE generalized eigenvalue problem is that the spectral decomposition of $\mathbf{C}^{-1}\mathbf{H}=\xi^o\mathbf{\Omega}^o\xi^{o,-1}$ can be obtained from the singular value decomposition,

$$\mathbf{L}_2^T\mathbf{L}_1=\mathbf{U}\mathbf{\Omega}_+^o\mathbf{V}^T. \tag{7}$$

It can be shown[1] that the matrix ξ^o (Eq. 3) is given by,

$$\xi^o=\frac{1}{2}\begin{pmatrix}\mathbf{L}_2\mathbf{U}+\mathbf{L}_1\mathbf{V} & \mathbf{L}_2\mathbf{U}-\mathbf{L}_1\mathbf{V}\\ \mathbf{L}_2\mathbf{U}-\mathbf{L}_1\mathbf{V} & \mathbf{L}_2\mathbf{U}+\mathbf{L}_2\mathbf{U}\end{pmatrix}\begin{pmatrix}\mathbf{\Omega}_+^{o,-1/2} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Omega}_-^{o,-1/2}\end{pmatrix}, \tag{8}$$

allowing $\mathbf{X}$ and $\mathbf{Y}$ to be identified as,

$$\mathbf{X}=\frac{1}{2}(\mathbf{L}_2\mathbf{U}+\mathbf{L}_1\mathbf{V})\mathbf{\Omega}_+^{o,-1/2}$$
$$\mathbf{Y}=\frac{1}{2}(\mathbf{L}_2\mathbf{U}-\mathbf{L}_1\mathbf{V})\mathbf{\Omega}_-^{o,-1/2}$$

For all cases considered here (RPA, BSE and TDHF for finite systems), the $\mathbf{A}$ and $\mathbf{B}$ matrices are real and $\mathbf{A}-\mathbf{B}$ and $\mathbf{A}+\mathbf{B}$ are positive definite. For the special case of RPA where $\mathbf{A}-\mathbf{B}$ is diagonal, $\mathbf{L}_2=\mathbf{L}_2^T=(\mathbf{A}-\mathbf{B})^{1/2}$, and $\mathbf{X}$ and $\mathbf{Y}$ become,[3]

$$\mathbf{X}^\alpha=\frac{1}{2}\left[\mathbf{\Omega}_+^{\alpha,1/2}\left(\mathbf{A}-\mathbf{B}\right)^{-1/2}+\mathbf{\Omega}_+^{\alpha,-1/2}\left(\mathbf{A}-\mathbf{B}\right)^{1/2}\right]\mathbf{Z}^\alpha$$
$$\mathbf{Y}^\alpha=\frac{1}{2}\left[\mathbf{\Omega}_-^{\alpha,1/2}\left(\mathbf{A}-\mathbf{B}\right)^{-1/2}-\mathbf{\Omega}_-^{\alpha,-1/2}\left(\mathbf{A}-\mathbf{B}\right)^{1/2}\right]\mathbf{Z}^\alpha.$$

In this case the symmetric matrix eigenvalue problem,

$$\left(\mathbf{K} - \mathbf{\Omega}^{\alpha,2}\right)\mathbf{Z}^{\alpha} = 0, \tag{9}$$

is solved, where

$$\mathbf{K} = (\mathbf{A} - \mathbf{B})^{1/2}(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B})^{1/2}$$
$$\mathbf{Z}^{\alpha} = \mathbf{\Omega}_{+}^{\alpha,-1/2}(\mathbf{A} - \mathbf{B})^{1/2}(\mathbf{X}^{\alpha} - \mathbf{Y}^{\alpha})$$

References

(1) Shao, M.; da Journada, F. H.; Yang, C.; Deslippe, J.; Louie, S. G. *Linear Algebra Appl.* **2016**, *488*, 148.

(2) Blackford, L. S.; Choi., J.; Cleary, A.; D'Azevedo, E.; Demmel, J.; Dhillon, I.; Dongarra, J.; Hammarling, S.; Henry, G.; Petitet, A.; Stanley, K.; Walker, D.; Whaley, R. C. *ScaLAPACK Users' Guide*; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1997.

(3) Bruneval, F.; Rangel, T.; Hamed, S. M.; Shao, M.; Yang, C.; Neaton, J. B. *Computer Phys. Comm.* **2016**, *208*, 149.

(4) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.

(5) Krishnan, R.; Binkley, J.; Seeger, R.; Pople, J. *J. Chem. Phys.* **1980**, *72*, 650.

(6) Dovesi, R.; Saunders, V. R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Pascale, F.; Civalleri, B.; Doll, K.; Harrison, N. M.; Bush, I.; D'Arco, P.; Llunell, M. 2009.

(7) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.