Local Excitement in CP2K: eCSE 03-11 Technical Report

Sergey K. Chulkov, Matthew B. Watkins, and Iain Bethune

March 27, 2017

Abstract

An efficient and flexible time-dependent density functional theory module has been implemented in the CP2K atomistic simulation package. Excited state properties of a wide range of molecular systems up to few thousand atoms can now be accessed using local and non-local density functionals. A variety of techniques have been employed to reduce associated computational costs and achieve better parallel performance.

1 Introduction

CP2K [1] is an open source quantum chemical software package suitable for atomistic simulation of a wide range of systems including isolated molecules, liquid and solid state systems. A large variety of ground state methods is available including standard Hartree-Fock (HF) theory, Density Functional Theory (DFT), Random Phase Approximation (RPA) and many others. In contrast, only two options are available in CP2K to compute properties of a system in its electronic excited states which are the Real-Time propagation Time-Dependent DFT (RT-TDDFT) and a basic implementation of the Linear Response TDDFT (LR-TDDFT) within the Tamm-Dancoff approximation. The latter method is also known as the Time-Dependent Density Functional Perturbation Theory (TDDFPT).

Being in general very robust, the original implementation of TDDFPT has several limitations. The main one is the fact that it supports only Local Density Approximation (LDA) and Generalised Gradient Approximation (GGA) exchange-correlation (XC) functionals. However, such local and semi-local functionals have incorrect long-range behaviour [2], and – because of this – well known underestimation of charge transfer states. Besides, the original implementation is found to be too restrictive on the number of possible excitations in an active space, leading to slow convergence of high-lying excited states.

The main objectives of this eCSE project are the following. The first aim is to support hybrid functionals in the TDDFPT code due to importance of exact exchange for accurate simulation of excited states in many systems. The second aim is to implement computation of oscillator strengths as they are completely missed in the original implementation. The final aim is to extend the number of excited states methods available by implementing a Maximum Overlap Method [3] along with some post-processing tools for RT-TDDFT calculations.

2 General Theory

2.1 Maximum Overlap Method

The idea behind the Maximum Overlap Method (MOM) [3] is to access excited states properties using the model of non-interacting electrons – the underlying model of the Self-Consistent Field (SCF) method and Density Functional Theory (DFT). Within MOM, the excited state wave-function is just a single Slater determinant constructed from a ground-state determinant by replacing one or several occupied orbitals by the same number of virtual orbitals. This excited state wave-function is then optimised iteratively using an SCF-like algorithm. The structure of the Slater determinant is

revised at every iteration by using the orbitals which overlap most with the occupied orbitals from the previous iteration. This is achieved by computing an overlap matrix between old occupied (C_{old}) and new (C_{new}) molecular orbitals (MOs) in a given atomic basis set $\{\chi_\mu\}$:

$$O = C_{\text{old}}^T S C_{\text{new}}, \tag{1}$$

where S is the metric matrix:

$$S_{\mu\nu} = \int \chi_\mu^*(\mathbf{r})\chi_\nu(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{2}$$

Then the overlap of the new j-th MO with all the old occupied MOs is computed by summing up individual overlap matrix elements:

$$p_j^{\text{sum}} = \left| \sum_{i=1}^{\text{nocc.}} O_{ij} \right|, \tag{3}$$

and the orbitals with largest overlap components p_j are used to build the wave-function at the next SCF iteration. An alternative projection formula, which ignores the phase of MOs, has also been proposed [4]:

$$p_j^{\text{norm}} = \sum_{i=1}^{\text{nocc.}} O_{ij}^2. \tag{4}$$

This formula is found to be more stable than the original one in many cases.

Once converged, electron excitation energy can be obtained as the difference in total energies between excited and ground electronic states. For this reason MOM is referred to as one of the ΔSCF method. It should always be kept in mind that by design the scope of the method is limited by those excited states which are well described by a single Slater determinant. On the other hand, an obvious advantage of the method is that properties of the system in such excited states can be easily accessed using relevant ground-state methods once the single-determinant wave-function is constructed for that state. In particular, the potential energy surface (PES) of the molecule in such excited state can be observed by computing gradients or by running molecular dynamics simulation. This enables location of minima and saddle points on such excited state PES, computation of vibrational spectra and related properties. It also provides an interesting comparison to TDDFT, as it does not share the inherent weakness in the calculation of charge transfer states.

2.2 Time-Dependent Density Functional Perturbation Theory (TDDFPT)

The aim of the method is to access excited state properties of a molecular system by applying a weak time-dependent electric field to that system. The optical properties, such as electron transition energies and oscillator strength, are then obtained by computing a linear response of the system on perturbation caused by the applied field. In this section we just recall the most important aspects of the method and the working equations. We refer reader to an excellent review paper [5] for formal foundation and further details.

The TDDFPT equation can be cast in form of a non-Hermitian eigenproblem [5]:

$$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \begin{pmatrix} X_p \\ Y_p \end{pmatrix} = \omega_p \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X_p \\ Y_p \end{pmatrix}, \tag{5}$$

where (X_p, Y_p) is an eigenvector and ω_p is a corresponding transition energy. Doing a simple matrix–vector multiplication it can be shown that all eigenvectors can be organised in pairs. Thus, for any given eigenvector (X_p, Y_p) the vector (Y_p^*, X_p^*) is also the solution of the TDDFPT equation with the paired eigenvalue $-\omega_p$. By convention, eigenvectors with positive eigenvalues correspond to excitation transitions, while vectors with negative eigenvalues describe de-excitation transitions.

In terms of Kohn-Sham orbitals $\{\phi\}$, the elements of the matrices A and B can be written in form [2]:

$$\begin{aligned} A_{ia\sigma,jb\tau} &= A^{\mathrm{E}}_{ia\sigma,jb\tau} + A^{\mathrm{J}}_{ia\sigma,jb\tau} + A^{\mathrm{K}}_{ia\sigma,jb\tau} + A^{\mathrm{F}}_{ia\sigma,jb\tau} \\ &= \delta_{ij}\delta_{ab}\delta_{\sigma\tau}(\epsilon_{a\sigma} - \epsilon_{i\sigma}) + (i_\sigma a_\sigma|j_\tau b_\tau) - \delta_{\sigma\tau}c_{\mathrm{HF}}(i_\sigma j_\sigma|a_\tau b_\tau) + (1 - c_{\mathrm{HF}})(i_\sigma a_\sigma|f_{\sigma\tau}|j_\tau b_\tau), \qquad (6) \\ B_{ia\sigma,jb\tau} &= (i_\sigma a_\sigma|b_\tau j_\tau) - \delta_{\sigma\tau}c_{\mathrm{HF}}(i_\sigma b_\sigma|a_\tau j_\tau) + (1 - c_{\mathrm{HF}})(i_\sigma a_\sigma|f_{\sigma\tau}|b_\tau j_\tau). \qquad (7) \end{aligned}$$

Here A^{E}, A^{J}, A^{K}, A^{F} denote orbital energy difference, electron-hole Coulomb, exact-exchange, and exchange-correlation (XC) terms respectively, while notations $(i_\sigma a_\sigma|j_\tau b_\tau)$ and $(i_\sigma a_\sigma|f_{\mathrm{xc},\sigma\tau}|j_\tau b_\tau)$ stand for four-centre Electron Repulsion Integrals (ERIs) and XC-integrals over Kohn-Sham orbitals $\{\phi\}$:

$$(i_\sigma a_\sigma|j_\tau b_\tau) = \int \phi^*_{i\sigma}(\mathbf{r})\phi_{a\sigma}(\mathbf{r})\frac{1}{|\mathbf{r} - \mathbf{r}'|}\phi^*_{j\tau}(\mathbf{r}')\phi_{b\tau}(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \qquad (8)$$

$$(i_\sigma a_\sigma|f_{\mathrm{xc},\sigma\tau}|j_\tau b_\tau) = \int \phi^*_{i\sigma}(\mathbf{r})\phi_{a\sigma}(\mathbf{r})f_{\mathrm{xc},\sigma\tau}(\mathbf{r},\mathbf{r}')\phi^*_{j\tau}(\mathbf{r}')\phi_{b\tau}(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \qquad (9)$$

We use the following index convention: i and j label occupied orbitals, a and b stands for virtual orbitals, and σ and τ refer to spin components. Besides, the quantity $\epsilon_{i\sigma}$ stands for the i-th Kohn-Sham orbital energy, and $f_{\sigma\tau}(\mathbf{r},\mathbf{r}')$ is an XC-kernel.

There are two approximations that are typically employed when solving TDDFPT equation. Within the adiabatic approximation, the XC-functional is simply one of the usual XC-functional used in ground state DFT. As such functionals do not depend on time explicitly, the XC-potential that electron feels is just a standard DFT XC-potential evaluated at the given time with the density at that time. As such, the expression for XC-kernel becomes relatively simple and reads:

$$f_{\mathrm{xc},\sigma\tau}(\mathbf{r},\mathbf{r}') = \frac{\mathrm{d}^2 E_{\mathrm{xc}}[\rho]}{\mathrm{d}\rho_\sigma(\mathbf{r})\mathrm{d}\rho_\sigma(\mathbf{r}')}. \qquad (10)$$

At the same time, within Tamm-Dancoff Approximation (TDA) [6] all the elements of the matrix B in Eq. (5) are assumed to be negligibly small. This obviously simplifies the TDDFPT equation by reducing it to a standard Hermitian eigenproblem:

$$AX = \omega X. \qquad (11)$$

Once being solved, the response function of the p-th transition Ψ_p can be constructed and used to compute the corresponding oscillator strength using dipole integrals in the "length" form:

$$f_p = \frac{2}{3}\omega_p \sum_{q=x,y,z} |\langle\Psi_0|q|\Psi_p\rangle|^2 . \qquad (12)$$

The above expression can be recast in terms of Kohn-Sham orbitals:

$$f_p = \frac{2}{3}\omega_p \sum_{q=x,y,z} \left| \sum_i^{\text{occ.}} \sum_a^{\text{virt.}} \sum_\sigma x_{ia\sigma,p}\langle\phi_{i\sigma}|q|\phi_{a\sigma}\rangle \right|^2 , \qquad (13)$$

where $x_{ia\sigma,p}$ are elements of the matrix X.

Using the commutation relation between Hamiltonian and position operators

$$[\hat{H}, \hat{\mathbf{r}}] = -\nabla, \qquad (14)$$

we can rewrite the Eq. (13) in the equivalent "velocity" form:

$$f_p = \frac{2}{3}\omega_p \sum_{q=x,y,z} \left| \sum_i^{\text{occ.}} \sum_a^{\text{virt.}} \sum_\sigma \frac{x_{ia\sigma,p}}{\epsilon_a - \epsilon_i} \left\langle \phi_{i\sigma} \left| \frac{\partial^2}{\partial q^2} \right| \phi_{a\sigma} \right\rangle \right|^2 . \qquad (15)$$

The last expression is best suited for calculations which employ periodic boundary conditions.

2.3 Auxiliary Density Matrix Method

The main bottleneck of (TD)DFT calculations using hybrid XC-functionals is evaluation of Electron Repulsion Integrals (ERIs) over Gaussian basis functions $\{\chi_\mu\}$:

$$(\mu\nu|\lambda\xi) = \int \chi^*_\mu(\mathbf{r})\chi_\nu(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}\chi^*_\lambda(\mathbf{r}')\chi_\xi(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{16}$$

The number of such integrals grows with the total number of atomic basis functions to the power of four. Not all of these integrals are indeed important. In particular, the product of two Gaussian functions centred on different atoms $\int \chi^*_\mu(\mathbf{r})\chi_\nu(\mathbf{r})\mathrm{d}\mathbf{r}$ decays with the distance between these atoms. This effectively means that the number of non-negligible ERIs scales quadratically with the system size for a given basis set. To take advantage of this fact, CP2K screens ERIs based on the Cauchy-Schwarz inequality:

$$(\mu\nu|\lambda\sigma) \leq \sqrt{(\mu\nu|\mu\nu)}\sqrt{(\lambda\xi|\lambda\xi)} \leq \text{threshold}, \tag{17}$$

and ignores the integrals which are less than the given threshold.

Even employing the efficient screening technique, evaluation of ERIs still remains a challenging task due to a large prefactor. This prefactor becomes even larger when the basis set is augmented with diffuse functions, because these functions have much slower long-range decay.

The main idea behind the Auxiliary Density Matrix Method (ADMM) is to evaluate ERIs using a small auxiliary basis set with rapidly decaying basis functions $\{\tilde{\chi_\mu}\}$. An approximate auxiliary density matrix $\tilde{P}$ is then constructed by fitting the density matrix in the primary basis set (P):

$$\tilde{P} = APA^T, \tag{18}$$

using a projector from the primary basis set onto the auxiliary basis set:

$$A = \tilde{S}^{-1}Q, \tag{19}$$

$$\tilde{S}_{\mu\nu} = \langle\tilde{\chi}_\mu|\tilde{\chi}_\nu\rangle, \quad Q_{\mu\nu} = \langle\tilde{\chi}_\mu|\chi_\nu\rangle. \tag{20}$$

Optionally, the auxiliary density matrix can be purified using a number of techniques [7] to ensure that all properties of a true density matrix are met.

In general, the exact-exchange energies computed using primary and auxiliary density matrices can be substantially different especially in case of significantly smaller auxiliary basis sets. ADMM assumes that this difference has mainly (semi-)local nature and can be well described using some reference LDA or GGA exchange functional ($E_\mathrm{x}^\mathrm{DFT}$). It naturally leads to the following simple expression for approximate exact-exchange energy in the primary basis set:

$$E^\mathrm{HFX}[P] \approx E^\mathrm{HFX}[\tilde{P}] + (E_\mathrm{x}^\mathrm{DFT}[P] - E_\mathrm{x}^\mathrm{DFT}[\tilde{P}]). \tag{21}$$

It can be seen that this correction requires to compute two extra (semi-)local exchange energies which introduces some performance penalty. Albeit being in general negligibly small, this penalty may outweigh the cost of ERIs evaluation in a primary basis set in case of small molecular systems (several atoms). On the other hand, in case of moderately sized and large systems ADMM is an extremely beneficial technique as it reduces computational costs of hybrid-functional simulations by few orders of magnitude by introducing only small discrepancies in total energies.

3 Implementation

The TDDFPT module in CP2K was completely rewritten for several reasons. One of them is to make the code more flexible and easy to use. In order to run TDDFPT calculation using the original implementation of the method, a series of keywords in different input sections should be set consistently. Thus, for efficiency reason the default behaviour of the CP2K's DFT engine (QUICKSTEP) [8] is to limit the size of the eigenproblem for the Kohn-Sham matrix to the occupied orbitals only. This is a very

reasonable optimisation because only occupied orbitals contribute towards the ground state electron density. However, this is not the case for excited state calculations: we do need virtual Kohn-Sham orbitals to construct an initial guess for response wave-functions. CP2K provides two mechanism to request these virtual orbitals. One of them is solely used by a number of linear-response methods (including TDDFPT) and it is activated via a keyword `FORCE_EVAL/DFT/EXCITATIONS`. The other method can be requested via a keyword `FORCE_EVAL/DFT/SCF/ADDED_MOS` and it is used mainly for smearing and visualisation purposes. Unfortunately, these mechanisms are in conflict with each other and cannot be activated simultaneously. This obviously introduces unnecessary level of complexity from the user's perspective, as they should be aware about implementation-specific details. The new implementation aims to be user friendly by hiding such details from end users. In particular, at the very beginning the TDDFT code checks whether or not all required virtual orbitals are available through one of these mechanisms, and – if not – computes them automatically by diagonalising the Kohn-Sham matrix one more time.

Another feature of the new implementation is that excited state calculation is now requested as a DFT property, not as a separate 'run type'. This greatly improves flexibility of the code: the TDDFPT calculation can now be performed not just at one single point in time, but at arbitrary points along an optimisation trajectory or during molecular dynamics simulation. The code rewriting also helped us to achieve some performance gain by improving the underlying algorithm and by using advanced parallelisation strategies. Moreover, the code is now organised in a much cleaner way to improve its readability and to ease its further maintenance.

To solve the TDDFPT equation in Tamm-Dancoff approximation, according to Eq. 11 we need to solve the eigenproblem for the matrix A with matrix elements expressed in the most general spin-polarised case as:

$$A_{ia\sigma,jb\tau} = \langle \Psi^a_{i,\sigma} | \hat{A} | \Psi^b_{j,\tau} \rangle. \tag{22}$$

Here, σ and τ denote spin components of two electrons, and Ψ^a_i is a singly excited wave-function, whose i-th occupied Molecular Orbital (MO) is replaced by the a-th virtual MO. The size of this matrix along each dimension is equal to the number of possible single excitations which is the number of occupied MOs times the number of virtual MOs. The number of matrix elements grows very rapidly with the system size as well as the number of basis functions, and quickly becomes too large to fit into main memory even for a relatively small molecular system (dozens of atoms). Fortunately, due to the fact that the matrix A itself is a diagonally dominant matrix and only few excited states are typically in interest, the block Davidson method [9] is an optimal choice to handle such an eigenproblem.

One of the most attractive feature of the Davidson method is that the entire matrix A does not need to be computed. Since only the action of the TDDFPT operator $\hat{A}$ on some trial vector – which is a linear combination of virtual MOs – is actually required, it is convenient to store the response wave-function in a contracted matrix form $C^{(1)}_\sigma$. The shape of this matrix is exactly the same as the shape of the ground state MO coefficients matrix $C^{(0)}_\sigma$. The number of rows and columns of both matrices are equal to the number of atomic basis functions in the given basis set and the number of occupied MOs for a given spin component σ respectively. However, in contrast with the matrix $C^{(0)}_\sigma$, which i-th column stores expansion coefficients of the i-th occupied MO $\phi_{i,\sigma}$ in terms of the atomic basis set $\{\chi_\mu\}_{\mu=1\ldots N}$:

$$\phi_{i,\sigma} = \sum_{\mu=1}^{N} c^{(0)}_{\mu i,\sigma} \chi_\mu,$$

the i-th column of the matrix $C^{(1)}_\sigma$ stores coefficients not just for a single virtual MO ϕ_a, but for a linear combination of such MOs $\sum_a c^a_i \phi_a$. This combination corresponds to an effective orbital that is occupied by an electron from the i-th occupied orbital during the excitation process. The squared expansion coefficient $(c^a_i)^2$ in this linear combination corresponds to probability of an electron being excited from the i-th occupied MO to the a-th virtual MO. The orthonormalisation condition imposed on response wave-functions in the most general spin-polarised case leads to the following relation

between expansion coefficients in the contracted matrix form:

$$\mathrm{Tr}\left(C^{(1),\mathrm{T}}_{p,\uparrow} S C^{(1)}_{q,\uparrow}\right) + \mathrm{Tr}\left(C^{(1),\mathrm{T}}_{p,\downarrow} S C^{(1)}_{q,\downarrow}\right) = \delta_{p,q}. \tag{23}$$

The central part of the TDDFPT algorithm is evaluation of the action of the response operator $\hat{A}$. In case of the p-th excited state described by a response wave-function $C^{(1)}_p$ the calculation is performed in the following steps.

1. For every spin component σ construct the response density matrix:

$$P^{(1)}_{p,\sigma} = \frac{1}{2}\left(C^{(0)}_{\sigma} C^{(1),\mathrm{T}}_{p,\sigma} + C^{(1)}_{p,\sigma} C^{(0),\mathrm{T}}_{\sigma}\right). \tag{24}$$

   Map the response density onto a real-space grid $P^{(1)}_{p,\sigma} \to \rho^{(1)}_{p,\sigma}(\mathbf{r})$ and the corresponding reciprocal-space grid by performing the fast Fourier transformation (FFT):

$$\rho^{(1)}_{p,\sigma}(\mathbf{G}) = \mathrm{FFT}[\rho^{(1)}_{p,\sigma}(\mathbf{r})].$$

2. Compute the energy difference term:

$$\hat{A}_E C^{(1)}_{p,\sigma} = F_\sigma C^{(1)}_{p,\sigma} - \epsilon_\sigma S C^{(1)}_{p,\sigma}, \tag{25}$$

   where F_σ and S are Kohn-Sham and metric matrices respectively, and ϵ_σ is a diagonal matrix of Kohn-Sham orbital energies.

3. Compute the Coulomb term by

   (a) solving the Poisson equation on the reciprocal grid: $\rho^{(1)}_{p,\sigma}(\mathbf{G}) \to v^{(1)}_{p,\sigma}(\mathbf{G})$;

   (b) calculating components of the electrostatic potential on the real-space grid using the inverse FFT: $v^{(1)}_{p,\sigma}(\mathbf{r}) = \mathrm{FFT}^{-1}[v^{(1)}_{p,\sigma}(\mathbf{G})]$;

   (c) transforming electrostatic potential from the grid representation into a matrix representation in the atomic basis set: $v^{(1)}_{p,\sigma}(\mathbf{r}) \to V^{(1)}_{p,\sigma}$ by evaluating expectation values:

$$\left(V^{(1)}_{p,\sigma}\right)_{\mu\nu} = \int \chi^*_\mu(\mathbf{r}) v^{(1)}_{p,\sigma}(\mathbf{r}) \chi_\nu(\mathbf{r}) \mathrm{d}\mathbf{r}; \tag{26}$$

   (d) evaluating the expression:

$$\hat{A}_J C^{(1)}_{p,\sigma} = \alpha V^{(1)}_{p,\sigma} C^{(1)}_{p,\sigma}, \tag{27}$$

   where the scaling factor α is equal to 2 and 0 for singlet and triplet excited states computed using spin-unpolarised electron density, or equal to 1 for spin-polarised calculations.

4. Compute the adiabatic kernel term by:

   (a) evaluating the integral

$$\bar{f}_{\mathrm{xc},p,\sigma\tau}(\mathbf{r}) = \int \frac{\mathrm{d}^2 E_{\mathrm{xc}}[\rho_{\uparrow\downarrow}]}{\mathrm{d}\rho_\sigma(\mathbf{r}) \mathrm{d}\rho_\tau(\mathbf{r}')} \rho^{(1)}_{p,\tau}(\mathbf{r}') \mathrm{d}\mathbf{r}' \tag{28}$$

   on the real-space grid for each spin component τ. In case of hybrid calculations, the exchange part of the XC-functional E_{xc} should be scaled by a factor of $(1 - c_{\mathrm{HF}})$, where c_{HF} is the amount of exact exchange;

   (b) transforming the above integrals from the real-space grid representation into a matrix representation in the atomic basis set: $\bar{f}_{\mathrm{XC},p,\sigma\tau}(\mathbf{r}) \to \bar{F}_{\mathrm{XC},p,\sigma\tau}$;

(c) evaluating the expression:

$$\hat{A}_F C^{(1)}_{p,\sigma} = \zeta \bar{F}_{\mathrm{XC},p,\sigma\uparrow} C^{(0)}_{\uparrow} + \eta \bar{F}_{\mathrm{XC},p,\sigma\downarrow} C^{(0)}_{\downarrow}, \tag{29}$$

where the scaling factors (ζ, η) are equal to $(2, 0)$ or $(1, -1)$ for singlet and triplet states computed using spin-unpolarised electron density, or $(1, 1)$ otherwise.

5. Using precomputed ERIs $(\mu\nu|\lambda\xi)$ (see Eq. 16), compute matrix elements of the exact exchange operator:

$$(K_{p,\sigma})_{\mu\lambda} = \sum_{\nu\xi} (\mu\nu|\lambda\xi) \left(P^{(1)}_{p,\sigma}\right)_{\nu\xi}. \tag{30}$$

The exact exchange term of the response operator is then calculated as a simple matrix product scaled by the amount of the exact exchange:

$$\hat{A}_K C^{(1)}_{p,\sigma} = c_{\mathrm{HF}} K_{p,\sigma} C^{(0)}_{p,\sigma}. \tag{31}$$

6. All components of the action are then summed up to obtain the final result:

$$\hat{A} C^{(1)}_{p,\sigma} = \hat{A}_E C^{(1)}_{p,\sigma} + \hat{A}_J C^{(1)}_{p,\sigma} + \hat{A}_F C^{(1)}_{p,\sigma} + \hat{A}_K C^{(1)}_{p,\sigma}. \tag{32}$$

Once we are able to compute the action of the response operator on an arbitrary response wavefunction, the block Davidson algorithm can be implemented in the following way. In case of the lowest k eigenstates:

1. sort possible single excitations according to the Kohn-Sham orbital energies. Construct the first k trial vectors in the Krylov space $G = \{C^{(1)}_{p,\sigma}\}_{p=1\ldots k}$ using the corresponding virtual orbitals with the lowest transition energies;

2. compute the action of the response operator on the trial vectors: $\{\hat{A} C^{(1)}_{p,\sigma}\}_{p=1\ldots k}$;

3. compute the reduced matrix with the matrix elements:

$$\tilde{A}_{pq} = \mathrm{Tr}\left(C^{(1),\mathrm{T}}_{p,\uparrow} \hat{A} C^{(1)}_{q,\uparrow}\right) + \mathrm{Tr}\left(C^{(1),\mathrm{T}}_{p,\downarrow} \hat{A} C^{(1)}_{q,\downarrow}\right). \tag{33}$$

4. solve the eigenproblem for the reduced matrix $\tilde{A}$ obtaining the lowest k Ritz values $(\{\theta_p\}_{p=1\ldots k})$ and the corresponding Ritz vectors $(\{u_{p,\sigma}\}_{p=1\ldots k})$ of the projected matrix A (see Eq. 22);

5. transform the Ritz vectors into the contracted matrix form:

$$w_{p,\sigma} = \sum_{q=1}^{k} (u_{p,\sigma})_q C^{(1)}_{q,\sigma}, \tag{34}$$

$$\hat{A} w_{p,\sigma} = \sum_{q=1}^{k} (u_{p,\sigma})_q \hat{A} C^{(1)}_{q,\sigma}; \tag{35}$$

6. compute the next k residual vectors using the formula

$$r_{p,\sigma} = C^{(0),T}_{\mathrm{virt.},\sigma} \left(\hat{A} w_{p,\sigma} - \theta_p S w_{p,\sigma}\right), \tag{36}$$

where $C^{(0)}_{\mathrm{virt.},\sigma}$ is a matrix of expansion coefficients of the virtual Kohn-Sham orbitals in terms of the atomic basis set;

7. expand the Krylov space by adding k new trial vectors:

$$t_{p,\sigma} = C^{(0)}_{\text{virt.}\sigma} \left(D_{p,\sigma} \circ r_{p,\sigma} \right) \tag{37}$$

orthogonalised against all the previous trial vectors. The symbol $\circ$ denotes the element-wise or Hadamard matrix product, while $D_{p,\sigma}$ is a precondition matrix in the contracted form with the following matrix element:

$$\left(D_{p,\sigma} \right)_{ai} = \left(\epsilon_a - \epsilon_i - \theta_p \right)^{-1} . \tag{38}$$

Here ϵ_i and ϵ_a are the corresponding occupied and virtual Kohn-Sham orbital energies respectively;

8. repeat the steps from 2 to 7 until the convergence of the Ritz values $\{\theta_p\}$.

By default, matrix elements of trial vectors are distributed across all available MPI processes in a block-cyclic fashion. Being in general very reasonable, this distribution scheme, however, is not optimal in case of relatively small number of atomic basis functions or a large number of processors as it leads to nonuniform distribution of the matrix blocks. In these extreme cases the scalability of the implementation can be improved by taking into account the fact that different trial vectors are almost independent from each other and can be computed simultaneously in parallel. To do so, all MPI processes were organised into parallel groups which compute their own subset of the trial vectors. The performance gain achieved using such parallelisation strategy is mainly due to the improved load balance and elimination of synchronisation points between processes which belong to different parallel groups during matrix-matrix multiplication operations and fast Fourier transformation. The Davidson algorithm contains two stages that cannot be parallelised in such a way, as they require all of the trial vectors in the Krylov space. These steps are obviously the diagonalisation of the reduced response matrix and orthogonalisation of the new trial vectors with respect to each other and the previous ones. The currently implemented strategy is to perform such operations using the global MPI communicator. This introduces unavoidable data shuffling between the processes, but is leads to significantly better scalability in case of a large number of parallel groups compared to alternative local strategies.

4 Results

The results presented in this report were obtained using CP2K 5.0 (revision 17830) on ARCHER – a Cray XC30 supercomputer with standard nodes containing two 2.7 GHz 12-core Intel Xeon E5-2697 v2 series processors and 64 GB of random access memory (RAM).

The first test system is a glycine molecule (NH_2CH_2COOH) – the smallest amino acid residue – in the gas phase. The structure of the molecule was optimised in CP2K using Perdew-Berke-Erzenhof (PBE) exchange-correlation functional [10] and aug-GTH-def2-QZVP basis set [11] (495 spherical basis functions), while core electrons were taken into account using Goedecker-Hutter-Tetter (GTH) pseudo-potentials [12]. A cubic unit cell with the size of 30 Å along each lattice vector and plane-wave cutoff of 400 Ry was employed with periodic boundary conditions applied in all three dimensions. The obtained structure was used to compute the lowest 8 excited singlet and triplet states within TDA-TDDFPT framework using the same setup. The maximum number of Krylov vectors was limited to 100. The calculation was continued until the maximum absolute deviation of the excited states energies in two consecutive iterations does not exceed 0.001 eV. A reference non-periodic TDA LR-TDDFT calculation was performed in NWChem [13] version 6.6 using the same optimised geometry and a similar quality all-electron basis set (aug-cc-pVQZ [14], 630 spherical basis functions).

The computed energies and oscillator strengths are summarised in Table 1. All the results, in general, are in good agreement with each other. The newly implemented TDDFPT solver, however, demonstrates better accuracy and convergence properties compared to the originally implemented algorithm. In particular, transition energies are around two times more accurate then the ones computed using the original solver and converge up to five times faster. The root mean square deviation (RMSD) of the transition energies computed against the reference all-electron calculation is equal to

Table 1: Energies and oscillator strengths of the lowest eight singlet and triplet excited states of a glycine molecule in the gas phase computed at the PBE level of theory. Maximum absolute deviation (MAD) and root mean square deviation (RMSD) are calculated against the reference NWChem results. Times are measured starting from a fully converged ground state electron density (one ground-state SCF iteration plus all TDDFPT iterations).

	Energy, eV						Oscillator strength, a.u.	
State	CP2K, new		CP2K, orig.		NWChem		CP2K, new	NWChem
number	singlet	triplet	singlet	triplet	singlet	triplet	singlet	singlet
1	4.795	4.689	4.812	4.704	4.794	4.678	0.011	0.014
2	5.271	4.877	5.284	4.888	5.249	4.836	0.002	0.001
3	5.896	5.804	5.942	5.839	5.892	5.796	0.018	0.017
4	5.958	5.831	5.993	5.898	5.950	5.810	0.007	0.007
5	6.153	5.876	6.168	5.908	6.155	5.863	0.008	0.012
6	6.185	5.994	6.227	6.009	6.180	5.974	0.007	0.004
7	6.225	6.102	6.247	6.143	6.215	6.098	0.027	0.016
8	6.495	6.161	6.508	6.174	6.483	6.167	0.000	0.000
MAD	0.022	0.041	0.050	0.088	—	—	0.011	—
RMSD	0.010	0.019	0.035	0.048	—	—	0.004	—
time, core-hours	1.81	0.97	5.02	4.77	7.56	5.66	—	—

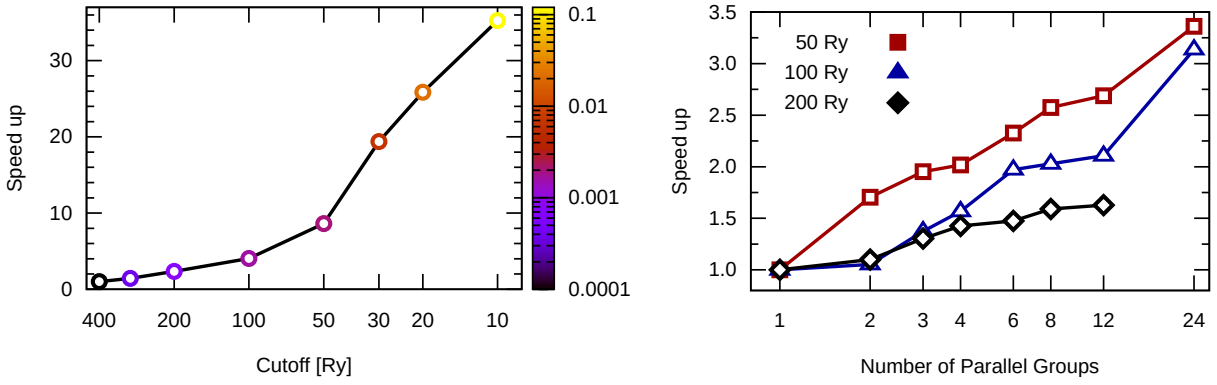
0.01 and 0.02 eV in case of singlet and triplet excited states respectively. The time spent to converge triplet states in spin-restricted case (0.97 core-hours) is two times smaller then the time spent to converge the same number of singlet states (1.81 core-hours). Two main factors contribute towards this performance gain. The first one is the fact that in case of triplet excited states we do not need to compute the electron-hole Coulomb interaction term due to opposite spins of electron and hole. Besides, the Davidson eigensolver requires only 5 iterations to converge eight triplet states whereas the same number of singlet states converges in 7 iterations.

The standard scenario of using the TDDFPT method is to compute excitation energies of several low-lying electronic states. These states typically involve transitions between a small number of the highest occupied and lowest unoccupied molecular orbitals which are usually relatively diffuse compared to the lower occupied orbitals. This fact helps to achieve further acceleration by using much smaller plain-wave cutoff in TDDFPT calculations compared to the ground state DFT. In order to demonstrate the computational benefits of this approach, the energy of eight singlet excited states of a glycine molecule in the gas phase were calculated at the PBE level of theory by varying the plane-wave cutoff from 400 Ry down to 10 Ry.

The left part of Figure 1 shows the achieved performance gain along with MAD as a function of the used plane-wave cutoff. It can be seen that reducing the plane-wave cutoff from 400 Ry down to 200 Ry accelerates TDDFPT calculation by a factor of 2 without noticeable lost in accuracy. In particular, MAD introduced by such reduction is equal to 0.0007 eV - smaller than the used convergence criteria (0.001 eV). A further speed up (by a factor of 19 in total) can be achieved by reducing cutoff down to 30 Ry that introduces an additional MAD error which is equal to 0.006 eV. Despite of the fact that this error exceeds the convergence criteria by a factor of 6, the RMSD of the transition energies computed against the all-electron calculation remains the same as in case of the non-reduced cutoff (0.01 eV, see Table 1).

The right part of Figure 1 demonstrates performance gain that can be achieved by reducing communication costs and improving load balancing. For the given test system, the response wave-function in contracted form is stored as a 495×15 matrix, which is the number of spherical atomic basis functions times the number of occupied molecular orbitals respectively. Being distributed in a block-cyclic fashion using the standard 32×32 blocks across 24 MPI cores, this response matrix occupies only 16 cores, while 8 cores remain idle. Using these idle cores to compute another excited state in parallel, we can expect additional performance gain by a factor of 1.5. According to Figure 1, this expectation is

Figure 1: Two approaches to improve computational efficiency of TDDFPT calculations. The left plot shows the performance gain achieved and the introduced maximum absolute deviation (MAD) error in TDDFPT energies of the lowest eight singlet excited states of a glycine molecule in the gas phase as the function of the plane-wave cutoff. The speed up and MAD are computed against the reference calculation which uses the optimal cutoff value for the ground state (400 Ry). The right plot shows the performance gain achieved by computing the lowest 24 singlet excited states in parallel as the function of the number of parallel groups. All calculations were performed using PBE XC-functional (see text).



in good agreement with the actual acceleration achieved using a fine real-space grid corresponding to the 200 Ry plane-wave cutoff. In case of coarse real-space grids with the cutoff 100 Ry and 50 Ry the performance gain is even larger (2.0 and 2.6 times respectively in case of 12 parallel groups) which reflects the additional improvement in load balancing of the plane-wave grids. Finally, extra acceleration can be achieved by using just 1 MPI process per excited states (24 parallel groups). The performance gain observed in this case is equal to 3.1 and 3.4 times in case of 100 Ry and 50 Ry real space grids respectively. Such significant improvement in performance is mainly due to lack of synchronisation barriers between MPI processes is this case as all the matrices related to some excited state become local to the corresponding process and not distributed among a number of processes.

To test the support of hybrid functionals, the lowest eight electron transition energies and corresponding oscillator strengths of a glycine molecule in the gas phase were also computed using PBE0 XC-functional [15]. The structure of the molecule was reoptimised beforehand using exactly the same setup as in case of the PBE XC-functional. All Electron Repulsion Integrals (ERIs) were screened using the Cauchy-Schwarz inequality (see Eq. 17) with the threshold equal to 1×10^{-8}, that is the default threshold setting used in NWChem 6.6. In addition, ERIs were truncated at the distance of 12 Å to avoid interactions between periodic images. To reduce computational costs, the non-zero ERIs were computed only once at the first SCF iteration, cached into the main memory and then reused when necessary. Other acceleration strategies, such as reduced plane-wave cutoff and parallel groups, were not employed. An extra calculation was also performed using ADMM with the aug-cFIT3 auxiliary basis set [7].

The results are summarised in Table 2. It can be seen that the results obtained using CP2K agree well with the reference all-electron calculation. In case of the standard algorithm, the RMSDs for excitation energies and oscillator strengths are equal to 0.037 eV and 0.005 a.u. respectively. The main computational bottleneck of the calculation using hybrid functionals is the evaluation of ERIs. Thus, for a given test system it takes 0.43 core-hours to compute 8.1×10^9 non-negligible ERIs, while all the other components – which are Coulomb term, exchange-correlation term, and four-index transformation which is needed to compute the exact-exchange term – require only 0.046 core-hours for each excited state per Davidson iteration.

By reducing the number of ERIs, ADMM accelerates calculations in two ways. The first way is straightforward: computational costs associated with evaluation of ERIs are in direct proportion with the number of ERIs. On the other hand, in case of the limited amount of spare memory, reduction

Table 2: Energies and oscillator strengths of the lowest eight singlet excited states of a glycine molecule in the gas phase computed at the PBE0 level of theory. Maximum Absolute Deviation (MAD) and Root Mean Square Deviation (RMSD) are calculated against the reference NWChem results. Total times are measured starting from a fully converged ground state electron density; TDDFPT calculations were converged in 9 (CP2K) and 7 (NWChem) Davidson iterations.

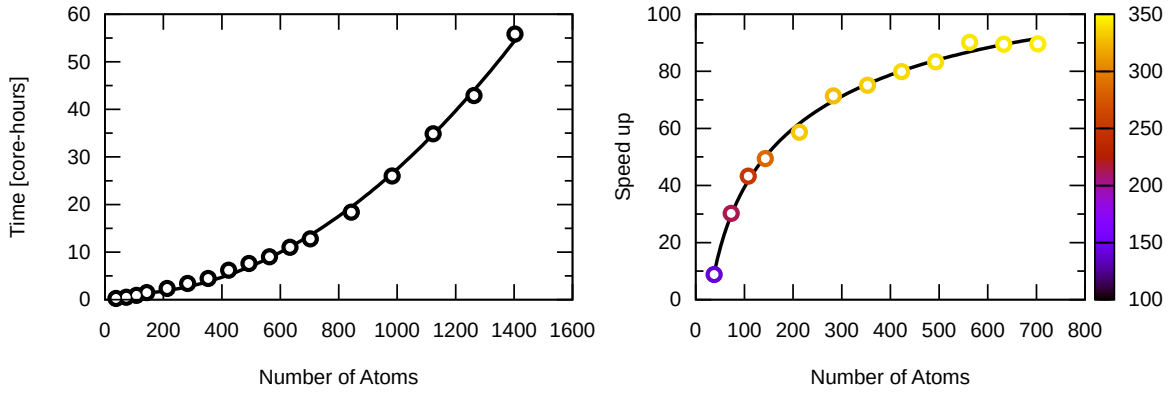
State number	Energy, eV			Oscillator strength, a.u.		
	CP2K ADMM	CP2K	NWChem	CP2K ADMM	CP2K	NWChem
1	5.592	5.624	5.704	0.001	0.001	0.002
2	5.920	5.906	5.938	0.005	0.005	0.010
3	6.905	6.873	6.892	0.016	0.018	0.029
4	6.940	6.911	6.927	0.022	0.020	0.028
5	7.084	7.051	7.092	0.003	0.004	0.004
6	7.236	7.187	7.212	0.002	0.002	0.003
7	7.395	7.393	7.416	0.002	0.002	0.002
8	7.582	7.585	7.601	0.000	0.000	0.001
MADE	0.112	0.080	—	0.013	0.011	—
RMSD	0.043	0.037	—	0.005	0.005	—
Number of sph. ERIs	5.212×10^6	8.064×10^9	3.957×10^{10}	—	—	—
Time (ERIs), core-hours	0.00	0.43	0.74	—	—	—
Time (total), core-hours	5.44	3.79	17.15[1]	—	—	—

[1] Semi-direct calculation: 54.6% of ERIs were computed once and cached into the main memory while the remaining ERIs were recomputed on-the-fly.

in the number of ERIs increases the percentage of cached integrals. When needed, these ERIs can be retrieved from memory instead of recomputing them on the fly, reducing the overall computational cost. Even for our small test system and tight Cauchy-Schwarz threshold, the amount of memory needed to store non-negligible ERIs in compressed form is equal to 15 GiB. Taking into account the fact that the amount of the required spare memory scales quadratically with the system size the situation of insufficient memory is likely to occur in actual simulations. Alternatively, in case of the aug-cFIT3 auxiliary basis set, ADMM reduces the time spent to compute ERIs by a factor of 1600 introducing a small additional error. In our test case, RMSD in the lowest eight transition energies computed against the all-electron calculation increases from 0.037 eV up to 0.043 eV, while RMSD evaluated against transition energies computed using primary basis set only is equal to 0.029 eV. At the same time it reduces the amount of memory needed to store precomputed ERIs down to 11 MiB, increasing the size of the system that can be simulated using the same hardware by a factor of 40. As an overhead, two reference exchange energies for primary and auxiliary density matrices have to be evaluated, which roughly doubles the computational costs associated with non-exact-exchange terms.

The computational benefit of using ADMM becomes more significant with increasing system size. To demonstrate the performance gain, the lowest eight excited states of a series of polyglycine peptides in alpha-helix conformation with different number of glycine fragments were computed at the PBE0 and PBE0/ADMM levels of theory. For this test case, the primary and auxiliary basis sets in use were TZV2P-MOLOPT-GTH and cFIT3 respectively for all atoms. The calculation was performed using the truncated potential with 12 Å cutoff radius and periodic boundary conditions applied in all directions. 20 Å of vacuum space was also added along each lattice vector. The number of Davidson iterations were limited to 3, and a tight convergence criteria was specified to ensure that all three iterations were actually performed. In order to minimise computational costs, the number of ARCHER compute nodes was kept as small as possible but still enough to cache all non-negligible ERIs in memory. The important ERIs were determined based on the Cauchy-Schwarz inequality with the threshold equal to 1×10^{-8}. In addition, the TDDFPT plane-wave cutoff was reduced down to 200 Ry and the excited states were processed simultaneously using 4 parallel groups.

Figure 2: Computational performance of TDDFPT/ADMM with respect to the system size. The left plot shows the time spent to perform 3 TDDFPT iterations, while the right plot visualises the achieved performance gain compared to the primary basis set calculation. The ratio between the number of spherical non-negligible electron repulsion integrals in primary and auxiliary atomic basis sets is encoded in colour. See text for the actual setup used.



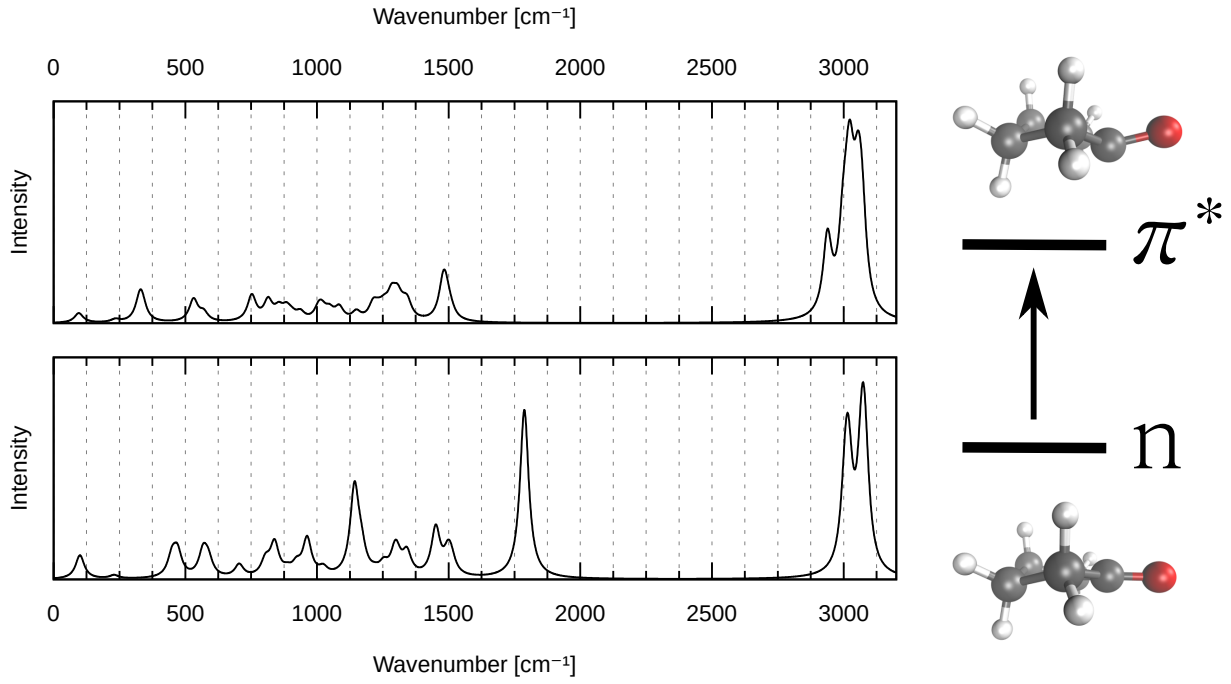
It should be also noted that diffuse basis functions are indeed quite important to obtain correct excitation energies. On the other hand, as the number of atoms becomes very large, the ground state density becomes challenging to converge due to nearly linear dependence of basis function. As the main purpose of this test is to determine the achieved performance gain, not the accuracy with respect to the experimental data, diffuse functions were not been used.

The obtained results are visualised in Figure 2. It can be seen that the achieved performance gain by using ADMM rapidly increases by a factor of 58 as the total number of atoms becomes equal to 200. The trend is in direct proportion to the ratio between the number of functions in primary and auxiliary atomic basis sets. This ratio is rapidly increases with the system size and becomes equal to 330 in case of a polyglycine peptide composed from 213 atoms. With further increase in system size this ratio then slowly increases up to 340 as the number of atoms increases up to 703. However even despite of this fact the actual ADMM-related speedup increases in half (up to a factor of 91). In our opinion this happens due to better load balance in case of ADMM calculation as it requires significantly smaller amount of memory to cache non-negligible ERIs compared to non-ADMM calculation. As each computational node equipped with a limited amount of RAM, this higher demand in memory requires significantly large number of MPI processes to fit the ERIs. In particular, in case of a polyglycine peptide composed from 703 atoms the number of such processes are equal to 144 and 9984 for ADMM and no-ADMM calculations respectively.

As a final test case, the structure of a cyclopentanone molecule in the gas phase in the first singlet electronic excited state $S_1(n, \pi^*)$ was optimised using the Maximum Overlap Method (MOM). In order to do so, the optimal structure of the molecule in its electronic ground state S_0 was obtained using B3LYP XC-functional [16] and the aug-GTH-def2-QZVP basis set along with GTH pseudo-potentials. A fully periodic calculation was performed using a cubic unit cell with size 30 Å along each lattice vector and a plane-wave cutoff of 400 Ry. Starting from the optimal geometry, one electron was promoted from the highest occupied molecular orbital to the lowest unoccupied molecular orbitals, and the MOM procedure was switched on.

To validate the results, infra-red spectra at the equilibrium points on the ground state (S_0) and $S_1(n, \pi^*)$ potential energy surfaces were computed (see Figure 3). Table 3 compares six theoretical fundamental frequencies with experimentally recorded data. It can be seen that the computed frequencies are in very good agreement with the experimental data. According to the experimental data, C=O in-plane and out-of-plane wagging, ring bending, and C=O stretching modes are red-shifted by 125, 137, 170, and 540 cm^{-1} respectively. The computed harmonic fundamental frequencies demonstrate similar shifts which are equal to 143, 113, 173, and 533 cm^{-1} respectively. C=O wagging modes computed within the harmonic approximation appear in inverted order. We assume that this is

Figure 3: Infrared spectra of a cyclopentanone molecule in the gas phase in the ground (bottom) and the first singlet excited (top) states. The spectra have been computed at the B3LYP level of theory within the harmonic approximation.



mainly due to missed anharmonic corrections, as anharmonic coupling terms between C=O stretching and wagging modes seem to be substantial. Nevertheless, the maximum absolute deviation between the available experimental and harmonic transition frequencies of the cyclopentanone molecule in its $S_1(n, \pi^*)$ state is equal to 27 cm^{-1} which is within the accuracy of the B3LYP XC-functional.

5 Conclusions

We demonstrate that the new implementation of TDDFPT within the CP2K package is working correctly and shows good efficiency.

The combination of the linear-scaling algorithm originally proposed by Chaissang and Hutter with the ADMM method produces a low scaling algorithm for the calculation of low lying excited states in the Casida formalism. Indeed, as the system size increases, the overhead for the hybrid calculation is fairly minor.

Table 3: Fundamental transition frequencies of a cyclopentanone molecule in the gas phase in the ground [S_0] and the first singlet excited [$S_1(n, \pi^*)$] states. Experimental frequencies are given according to Ref. [17]. All values are in cm^{-1}.

		B3LYP			Experiment		
State	Description	Ground	Excited	$-\Delta\nu$	Ground	Excited	$-\Delta\nu$
ν_3	C=O stretch	1788	1255	533	1770	1230	540
ν_{11}	Ring angle bend	704	531	173	705	532	173
ν_{18}	Ring twist	230	235	-5	238	238	0
ν_{25}	C=O out-of-plane wag	449	336	113	446	309	137
ν_{26}	Ring twist	100	96	4	95	91	4
ν_{36}	C=O in-plane wag	470	327	143	467	342	125

The reduced scaling (linear in system size) and smaller prefactor than previous implementations will allow qualitatively correct excited state calculations to be performed more routinely, for larger systems, in a time resolved manner along molecular dynamics trajectories and will be suitable for high-throughput applications.

Acknowledgments

This work was funded under the embedded CSE programme of the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk). We also gratefully thank the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), for computational resources on the ARCHER UK National Supercomputing Service.

References

[1] *CP2K developers home page*, http://cp2k.org/.

[2] A. Dreuw, J. L. Weisman, and M. Head-Gordon, J. Chem. Phys. **119**, 2943 (2003).

[3] A. T. B. Gilbert, N. A. Besley, and P. M. W. Gill, J. Phys. Chem. A **112**, 13164 (2008).

[4] O. V. Ershova and N. A. Besley, Chem. Phys. Lett. **513**, 179 (2011).

[5] M. E. Casida, J. Mol. Struct.: THEOCHEM **914**, 3 (2009).

[6] S. Hirata and M. Head-Gordon, Chem. Phys. Lett. **314**, 291 (1999).

[7] M. Guidon, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **6**, 2348 (2010).

[8] J. VandeVondele et al., Computer Physics Communications **167**, 103 (2005).

[9] M. Crouzeix, B. Philippe, and M. Sadkane, SIAM J. Sci. Comput. **15**, 62 (1994).

[10] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[11] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys **7**, 3297 (2005).

[12] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).

[13] M. Valiev et al., Comput. Phys. Commun. **181**, 1477 (2010).

[14] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[15] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[16] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[17] J. Zhang, W. Chiang, and J. Laane, J. Chem. Phys. **98**, 6129 (1992).