Numerically accurate linear response-properties in the configuration-interaction singles (CIS) approximation

Numerically accurate linear response-properties in the configuration-interaction singles (CIS) approximation

Jakob S. Kottmann,[1] Sebastian Höfener,[2, *] and Florian A. Bischoff[1, †]

[1]*Institut für Chemie, Humboldt-Universität zu Berlin,*

Unter den Linden 6, 10099 Berlin, Germany

[2]*Institut für Physikalische Chemie, Karlsruher Institut für Technologie, D-76131 Karlsruhe*

Abstract

In the present work, we report an efficient implementation of configuration interaction singles (CIS) excitation energies and oscillator strengths using the multi-resolution analysis (MRA) framework to address the basis-set convergence of excited state computations. In MRA (ground-state) orbitals and excited states are constructed adaptively guaranteeing an overall precision. Thus not only valence but also in particular low-lying Rydberg states can be computed with consistent quality at the basis set limit *a priori*, or without special treatments, which is demonstrated using a small test set of organic molecules, basis sets, and states. We find that the new implementation of MRA-CIS excitation energy calculations is competitive to conventional LCAO calculations when the basis-set limit of medium-sized molecules are sought, which requires large, diffuse basis sets. This becomes particularly important if accurate calculations of molecular electronic absorption spectra with respect to basis-set incompleteness is required, in which both valence as well as Rydberg excitations can contribute to the molecule's UV/VIS fingerprint.

---

* sebastian.hoefener2@kit.edu

† florian.bischoff@hu-berlin.de

I. INTRODUCTION

Electronically excited states of molecules play an essential role in e.g. photochemistry and photophysics, which have attracted much attention due to their key role in natural and artificial photosynthesis. The toolbox of quantum chemistry provides a number of methods to compute excited states, such as time-dependent DFT (TDDFT), time-dependent Hartree-Fock (HF), or linear-response (LR) coupled cluster (CC).[1–3] TDDFT is an efficient formulation and yields accurate results for valence excited states and properties, but can severely fail for charge-transfer states and Rydberg states. Coupled-cluster models provide a way to systematically decrease the method error leading to the exact solution of the Schrödinger equation, but the reduction of the method error leads inescapably to an increase in scaling with respect to the system size and thus significantly longer computation times.

In this article we use the formalism of multi-resolution analysis (MRA) to compute excitated states of small and medium-sized molecules. Unlike standard quantum chemistry, MRA needs to be formulated in the so-called first quantization, i.e. in the real-space representation of operators and states. In the original formulation by Schrödinger, observables are represented by operators and states by functions.[4] Since from all mathematical solutions of the Schrödinger equation only particular ones are suited as physical solutions due to certain constraints such as norm and differentiability, this formulation was denoted first quantization. The name second quantization was chosen based on distinct physical observations, for instance the creation or annihilation of photons requiring also a quantization of the electromagnetic field,[5] but for most quantum-chemical applications and for the present work, second quantization denotes pragmatically the introduction of a basis, whose basis functions are combined and rotated to form (occupied and unoccupied) molecular orbitals as used to construct approximate solutions of e.g. the Schrödinger equation.[6]

The fundamental ansatz to virtually all basis set-based methods is the pre-definition of shape and number of basis functions, leaving the task to find an approximate solution of an actual problem with a number of adjustable (linear) parameters as small as possible. In quantum chemistry, a lot of effort has been invested in finding optimal basis sets. The big advantage of such pre-parameterized basis sets is at the same time the

biggest drawback: experience is not only helpful, it is rather indispensably needed, and many basis sets have been developed which are not used anymore because of imbalance leading to high accuracy in some cases and large errors in other cases. New basis sets are reported in the literature on a regular basis, each optimized for a particular task, for instance explicit correlation[7], (new) relativistic Hamiltonians[8], or the need to investigate advanced properties such as polarizabilities.[9] For excited-state calculations, the accuracy of the model is often less important than the quality of the basis, because standard basis sets have been optimized for ground-state energies,[10, 11] and are often too compact for excited states. The arbitrariness that comes with the use of ground-state basis sets for excited states makes it hard to compute reliable, reproducible data when (excited-state) calculations are limited to a certain basis-set size.

Many problems of pre-defined basis sets are avoided by using multi-resolution analysis (MRA).[12–15] MRA functions are represented on an adaptive grid that is constructed during the solution of the equations. In principle this technique can be used for all methods and molecules, while guaranteeing that the results are accurate up to a requested threshold. This technique has been used before for similar purposes, such as for computing ground state energies,[12, 14] polarizabilities,[16, 17] for solving the time-dependent Schrödinger equation,[18, 19] or computing TDHF/TDDFT excitation energies.[13] The increase in accuracy and reliability is paid for by a larger prefactor in computational efficiency compared to traditional LCAO approaches. However, the computational scaling is significantly lower for MRA than for LCAO, so that for large molecules MRA can become more efficient.

In the present work, we address the basis-set convergence for selected linear-response properties at the hand of the configuration-interaction singles (CIS) method. In section II, we review the basic equations for first and second quantization-based CIS response properties. In section IV, we report our findings concerning the basis-set incompleteness error for a selection of molecules and basis sets. The article closes with a summary and a conclusion.

II. METHODOLOGY

In this section we repeat the linear response formalism for completeness, although it has been given before. The aim is to have a side-by-side comparison of the CIS working equations in first and second quantization. Contrary to the conventional CIS derivation the transition to density functional methods is straight forward within the linear response framework.

A. Linear response

The response of the molecular system with respect to an external, time-dependent perturbation $f(t) = fe^{i\omega t} + f^{\dagger}e^{-i\omega t}$ can be expressed as a Taylor series,

$$\rho(t) = \rho^{(0)} + \rho^{(1)}(t) + \mathcal{O}(\rho^{(2)}), \tag{1}$$

where $\rho(t)$ is the time-dependent density, and $\rho^{(0)}$ is the unperturbed ground state density. The perturbed Fock operator $F(t)$ depends on the perturbed density, and can also be expressed in a Taylor series,

$$F(t) = F^{(0)} + f(t) + F^{(1)}(t) + \mathcal{O}(F^{(2)}). \tag{2}$$

The linear response of the density is computed as the functional derivative of the corresponding zeroth-order equation, e.g. Hartree-Fock or Kohn-Sham equations, with respect to the density:

$$F^{(1)}(t) = \frac{\delta F}{\delta \rho^{(0)}} \circ \rho^{(1)}(t). \tag{3}$$

Inserting the expressions into the time-dependent Schrödinger equations and keeping terms up to first order yields:

$$i\frac{\partial}{\partial t}\rho(t) = [F(t), \rho(t)]$$
$$= [F^{(0)} + F^{(1)}(t) + f(t), \rho^{(0)} + \rho^{(1)}(t)] \tag{4}$$

Since the external perturbation oscillates with frequency ω, the perturbed density ρ and the perturbed Fock operator F adopt the same time-dependence:

$$\rho^{(1)}(t) = \tilde{\rho}e^{i\omega t} + \tilde{\rho}^{\dagger}e^{-i\omega t}. \tag{5}$$

4

Factoring out the explicit time dependence $e^{i\omega t}$ from the perturbed density, and letting the perturbation go to zero leads to:

$$\left[F^{(0)}, \tilde{\rho}\right] + \left[\frac{\delta F}{\delta \rho^{(0)}} \circ \tilde{\rho}, \rho^{(0)}\right] = \omega \tilde{\rho}. \tag{6}$$

This general equation defines the perturbed density $\rho^{(1)}$ and thus the linear response of the system. Depending on a formulation in first or second quantization, different working equations are obtained as discussed in the following.

B.   CIS excitation energies in second quantization

In second quantization the perturbed densities are expressed as rotations from the occupied to the virtual space [20]

$$\rho_{pq}^{(1)} = d_{pq}e^{i\omega t} + d_{qp}^{\dagger}e^{-i\omega t} \tag{7}$$

$$\tilde{\rho}_{pq} = d_{pq} \tag{8}$$

with only the occupied/virtual and virtual/occupied blocks of d_{pq} having non-zero entries, which are denoted **x** and **y** ("excitation" and "deexcitation") vectors

$$d_{ai} \equiv x_{ai}, \qquad d_{jb} \equiv y_{bj}. \tag{9}$$

The general SCF response equation Eq. (6) becomes

$$\left[F_{pq}^{(0)}, d_{qr}\right] + \left[\frac{\partial F_{pq}}{\partial P_{st}}d_{st}, P_{qr}\right] = \omega d_{pr}. \tag{10}$$

In the Tamm-Dancoff approximation the de-excitation coefficients **y** are neglected. Using the diagonal nature of the Fock and density matrices the following set of linear equations are obtained:

$$\epsilon_a x_{ai} - x_{ai}\epsilon_i + \frac{\partial F_{ai}}{\partial P_{bj}}x_{bj}P_{ii}^{(0)} = \omega x_{ai}, \tag{11}$$

where i, j denote (active) occupied and a, b (active) virtual orbitals, and ϵ denote Hartree-Fock orbital energies. These can be cast into a matrix eigenvalue problem with

$$\mathbf{A}\mathbf{x}^r = \omega^r \mathbf{x}^r, \tag{12}$$

where we have introduced an additional index r indicating the r-th excitation corresponding to the r-th eigenpair of the matrix $\mathbf{A}$. The closed-shell expression matrix reads for singlet excitations:[21]

$$A^s_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b) + 2(ia|r^{-1}_{12}|jb) - (ij|r^{-1}_{12}|ab), \tag{13}$$

and for triplet excitations:

$$A^t_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b) - (ij|r^{-1}_{12}|ab). \tag{14}$$

C. CIS excitation energies in first quantization

In first quantization the perturbed densities $\tilde{\rho}$ and $\tilde{\rho}^\dagger$ are also expressed as projections from the occupied to the virtual space and vice versa:

$$\tilde{\rho} = \sum_i |x_i\rangle\langle\varphi_i| + |\varphi_i\rangle\langle y_i| \tag{15}$$

$$\tilde{\rho}^\dagger = \sum_i |\varphi_i\rangle\langle x_i| + |y_i\rangle\langle\varphi_i| \tag{16}$$

The first-order responses $|x_i\rangle$ have the same orthogonality properties as the vector $\mathbf{x}$ in second quantization: [13, 16]

$$|x^r_i\rangle = \left(1 - \rho^{(0)}\right)|x^r_i\rangle = 1 - \sum_j^{n_{\text{occ}}} |\varphi_j\rangle\langle\varphi_j|x^r_i\rangle, \tag{17}$$

$$\langle\mathbf{x}^r|\mathbf{x}^p\rangle = \sum_i^{n_{\text{occ}}} \langle x^r_i|x^p_i\rangle = \delta_{rp}, \tag{18}$$

where the superscript r (p) again indicates the r-th (p-th) root of the response equation, i.e. the r-th (p-th) excitation energy, and $\rho^{(0)}$ denotes the ground-state density. While in second quantization the orthogonality comes with the hermiticity of the matrix $\mathbf{A}$ in case of CIS (but not for other models), in first quantization the orthogonality must be explicitly enforced even for CIS.

Starting from Eq. (6), by left-projecting on $1 - \rho^{(0)}$, right-projecting on $\rho^{(0)}$, using the orthogonality relations, and dropping the $|y_i\rangle$ for the CIS approximation, the three terms

in Eq. (6) become:

$$\left(1 - \rho^{(0)}\right)\left[F^{(0)}, \tilde{\rho}\right]\rho^{(0)} = \sum_i \left(F^{(0)} - \epsilon_i\right)|x_i\rangle\langle\varphi_i|, \tag{19}$$

$$\left(1 - \rho^{(0)}\right)\left[\frac{\delta F}{\delta\rho^{(0)}} \circ \tilde{\rho}, \rho^{(0)}\right]\rho^{(0)} = \left(1 - \rho^{(0)}\right)\left(\frac{\delta F}{\delta\rho^{(0)}} \circ \tilde{\rho}\right)\rho^{(0)}, \tag{20}$$

$$\left(1 - \rho^{(0)}\right)\omega\tilde{\rho}\rho^{(0)} = \omega\sum_i |x_i\rangle\langle\varphi_i|. \tag{21}$$

Recombination of these terms and dropping the trailing $\langle\varphi_i|$ yields

$$\left(F^{(0)} - \epsilon_i\right)|x_i\rangle + \left(1 - \rho^{(0)}\right)\left(\frac{\delta F}{\delta\rho^{(0)}} \circ \tilde{\rho}\right)|\varphi_i\rangle = \omega|x_i\rangle \tag{22}$$

which is reminiscent of the second-quantized formulation of Eq. (12). This can be rearranged to:

$$(T - \epsilon_i - \omega)|x_i\rangle = -\left(1 - \rho^{(0)}\right)\left(\frac{\delta F}{\delta\rho^{(0)}} \circ \tilde{\rho}\right)|\varphi_i\rangle - (J - K + V_{\text{nuc}})|x_i\rangle, \tag{23}$$

$$|x_i\rangle = -G_\omega\left[\left(1 - \rho^{(0)}\right)\left(\frac{\delta F}{\delta\rho^{(0)}} \circ \tilde{\rho}\right)|\varphi_i\rangle - (J - K + V_{\text{nuc}})|x_i\rangle\right]. \tag{24}$$

These equations need to be solved iteratively since the solution $|x_i\rangle$ is part of the right hand side through $\tilde{\rho}$, see Eq. (15). The Green's operator G is the inverse of the Helmholtz operator [22]

$$(T - \epsilon_i - \omega)^{-1} = G_\omega. \tag{25}$$

Given the perturbed density $\tilde{\rho}$ in real space,

$$\tilde{\rho}(\mathbf{r}) = \sum_i \varphi_i(\mathbf{r})x_i(\mathbf{r}), \tag{26}$$

the variation of the Fock operator $\frac{\delta F}{\delta\rho^{(0)}}$ can be computed as:

$$\frac{\delta F}{\delta\rho^{(0)}} \circ \tilde{\rho}(\mathbf{r}) = \frac{\delta J}{\delta\rho^{(0)}} \circ \tilde{\rho}(\mathbf{r}) - \frac{\delta K}{\delta\rho^{(0)}} \circ \tilde{\rho}(\mathbf{r}), \tag{27}$$

$$\frac{\delta J}{\delta\rho^{(0)}} \circ \tilde{\rho}(\mathbf{r}) = \int d^3r' \frac{\tilde{\rho}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \tag{28}$$

$$\frac{\delta K}{\delta\rho^{(0)}} \circ \tilde{\rho}(\mathbf{r})|\varphi_i\rangle = \sum_j x_j(\mathbf{r}) \int d^3r' \frac{\varphi_i(\mathbf{r}')\varphi_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{29}$$

D. Calculation of MRA excitation energies

1. Initial guess for the excitation vectors

Let $\hat{f}$ be an excitation operator which generates approximately the excitation orbitals $|x_i\rangle$ when acting on ground state orbitals $|\varphi_i\rangle$,

$$|x_i\rangle \approx \hat{f}\,|\varphi_i\rangle. \tag{30}$$

This operator is expanded into polynomials up to n-th order,

$$\hat{f} \approx \sum_{i=1}^{3} c_i r_i + \sum_{i<j=1}^{3} c_{ij} r_i r_j + \cdots + \sum_{i<j<\ldots<n=1}^{3} c_{ij\ldots n} r_i r_j \ldots r_n. \tag{31}$$

For symmetric molecules the polynomials are grouped into irreducible representations according to their transformational behavior. In case of low-order symmetry groups or molecules without symmetry, expansion coefficients in Eq. (31) for the initial trial vectors can be obtained from either an approximate higher order symmetry group, or a guess based on intuition, experimental or calculated data such as multipole transition moments.

2. Iterative procedure and orthogonalization

Eq. (23) has to be solved for each target excitation vectors $|x^r\rangle$ while the individual vectors are orthonormalized in each iteration. A simple Gram-Schmidt orthogonalization fails to converge quickly when trial vectors are containing contributions to several excited states. A way to disentangle these excitations is to compute the Fock matrix elements of the excitation vectors F_{pr} by projecting the working equations from the left on the excitations $\langle x^p|$,

$$F_{pr} = \sum_i \langle x_i^p | F^{(0)} | x_i^r\rangle + \sum_i \langle x_i^p | (1 - \rho^{(0)}) \left(\frac{\delta F}{\delta \rho^{(0)}} \circ \tilde{\rho} \right) | \varphi_i \rangle. \tag{32}$$

The amplitudes are finally determined by solving the generalized eigenvalue problem of the Fock matrix:

$$|x_i^p\rangle = \sum_r |x_i^r\rangle U_{rp}, \tag{33}$$

where $\mathbf{U}$ are the eigenvectors of the perturbed Fock matrix.

The excitation energies ω^r of the r-th root can be extracted from the diagonal matrix elements through:

$$\omega^r = \frac{F_{rr} - \sum_i \epsilon_i \langle x_i^r | x_i^r \rangle}{\sum_i \langle x_i^r | x_i^r \rangle}, \tag{34}$$

or through an updating scheme: [12]

$$\Delta \omega^r = \frac{\langle \mathbf{x}^r | G_\omega V \mathbf{x}^r \rangle}{||G_\omega V | \mathbf{x}^r \rangle||^2}. \tag{35}$$

Here $G_\omega V |x\rangle$ is a short-hand notation for the right hand side of Eq. (24). The updating procedure is correct to second order with respect to the error in the excitation functions $|\mathbf{x}^r\rangle$ and thus provides accurate excitation energies. Furthermore, this scheme tends to be more accurate numerically because no derivative operator is involved. In contrast, the perturbed Fock matrix scheme is more stable if the current excitation vectors and energies are far from convergence.

The complete iterative procedure for obtaining MRA excitation vectors at CIS level of theory reads:

1. Guess initial trial vectors according to Sec. II D 1.

2. Iterate Eq. (23) once for all vectors $|\mathbf{x}^r\rangle$.

3. Orthogonalize the excitations and compute the excitation energies ω^r for each vector $|\mathbf{x}^r\rangle$ either by the updating scheme in Eq. (35) or through Eq. (34).

4. If not converged, return to 2.

5. Calculate oscillator strengths.

In CIS, oscillator strengths are obtained as contractions of the ground-state and excitation vectors with an appropriate operator. Explicit expressions for the oscillator strengths read for the length gauge (superscript l) and velocity gauge (superscript v) [23]:

$$f^l = \frac{4}{3}\omega^r |\langle \mathbf{x}^r | \boldsymbol{\mu} | \mathbf{0} \rangle|^2 = \frac{4}{3}\omega^r |\sum_i \langle x_i^r | \boldsymbol{\mu} | \varphi_i \rangle|^2, \tag{36}$$

$$f^v = \frac{4}{3\omega^r} |\langle \mathbf{x}^r | \mathbf{p} | \mathbf{0} \rangle|^2 = \frac{4}{3\omega^r} |\sum_i \langle x_i^r | \mathbf{p} | \varphi_i \rangle|^2. \tag{37}$$

The MRA formalism grants gauge invariance with respect to basis set incompleteness, but the gauge dependence inherent in the CIS model still persists.

III.   COMPUTATIONAL DETAILS

All LCAO-based calculations were performed with the ricc2 (CCS and CCSD) and escf (CIS) modules of the Turbomole program package version 6.5.[24] Standard basis sets were taken from the Turbomole basis set library. [25–27] The Rydberg basis used was "CM2" taken from Ref. 28, consisting of $2s2p2d$ with exponents 0.01 and 0.0033, respectively, which are placed in the center of mass of the molecules.[29]

The MRA treatment has been implemented in a local version of the MADNESS library. [30] The precision threshold of all functions (orbitals, excitation vectors) was set to 10^{-5} in the response equation and 10^{-6} in the ground state calculation. The polynomial order was chosen to be $k = 8$. Calculations were considered converged when the norm of the excitation vectors would not change more than 10^{-3}.

IV.   RESULTS AND DISCUSSION

This section is organized as follows. We first analyze the use of different initial trial vectors and convergence behavior for MRA CIS excitation energies. Using the new MRA implementation, we review basis-set based results using different basis sets and augmentation strategies in terms of basis-set incompleteness error and convergence to the basis-set limit. Excitation energies are treated for singlet and triplet excitations, while oscillator strengths are only discussed for singlet excited states because triplet excitations are spin-forbidden in non-relativistic treatments thus lacking a physical meaning. Finally, we discuss the performance of the new implementation of MRA CIS excitation energies.

A.   Initial trial vectors and convergence in MRA CIS

Similar to LCAO calculations the quality of the initial trial vectors is crucial for recovering the lowest excited states if not all roots are to be determined. In LCAO methods this problem arises mainly in symmetric systems such as benzene and is not dominant in asymmetric molecules. However, in MRA the configuration space is much larger and thus the problem is present in virtually all calculations. While in LCAO methods the guess is based on orbital energy differences, such an approach is not available in MRA methods, because no virtual orbitals exist and therefore no orbital energies as well.
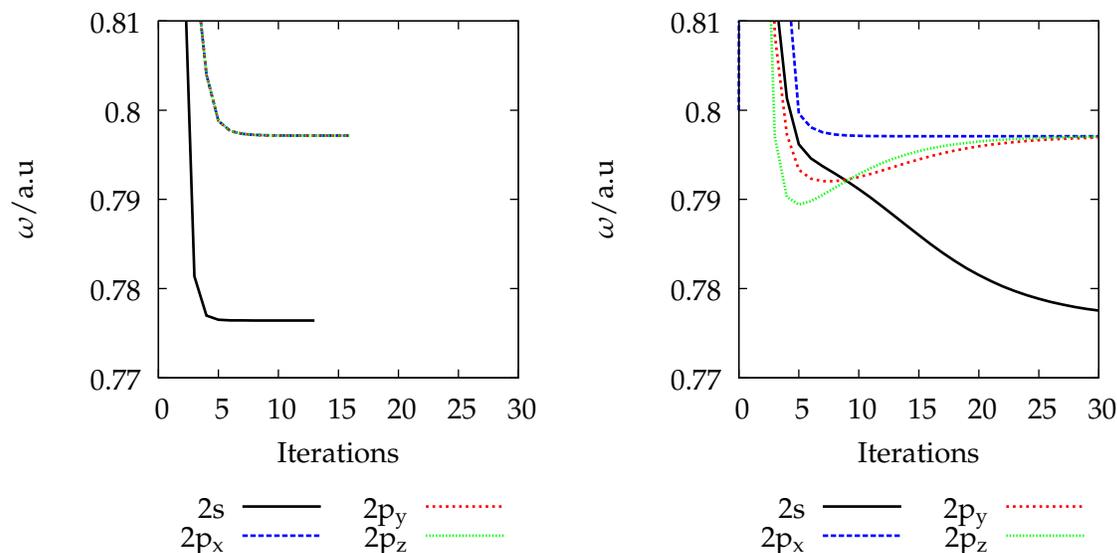
10

FIG. 1. Convergence behavior for s- and p-type excited states of the helium atom with different orthonormalization procedures: perturbed Fock matrix orthogonalization (left) and Gram-Schmidt orthogonalization (right). For the perturbed Fock matrix orthogonalization (left) all components converge smoothly within less than 10 iterations, while in the Gram-Schmidt orthogonalization (right) no convergence is achieved after 20 iterations.

Physically motivated initial trial vectors are obtained by multiplying the ground-state orbitals with a dipole operator, leading to a fast convergence for bright states, i.e. states with large oscillator strengths, but typically misses dark states, i.e. states with small oscillator strengths. Initial trial vectors for molecules with high symmetry can be obtained from symmetry-adapted guess functions, which converge quickly in all cases to the corresponding excited states. While the actual MRA calculations are carried out without applying point-group symmetry, the initial trial vectors may exhibit a particular symmetry. Trial vectors can also be based on a LCAO calculation in a small basis, mapping the ground-state orbitals using a Procrustes rotation.

In Fig. 1 the convergence of the MRA-CIS excitation energies of the helium atom is shown. The guess excitation vectors $|\mathbf{x}\rangle$ are excitations to the hybrid sp^3 orbitals. Depending the orthonormalization procedure these excitations can be quickly disentangled (perturbed Fock matrix diagonalization) or not (Gram-Schmidt orthogonalization). In addition, we used a Krylov subspace method[31] to accelerate the convergence of the

excitation vectors.

B.   CIS model error and basis set incompleteness

In Tab. I, two low-lying excitation energies of the ethylene molecule are listed, with 1^1B_{3u} being a valence excitation, and 2^1B_{3u} a Rydberg excitation. The reference MRA values were obtained with our new implementation. The basis set convergence for the valence excitation (1^1B_{3u}) is quite rapid, and already the smallest basis set, aug-cc-pVDZ, exhibits a basis-set incompleteness error of only about 25 meV, while the standard deviation concerning the method error of CCS can be assumed to be about 1.2 eV. [32] The Rydberg excitation (2^1B_{3u}), on the other hand, shows no rapid convergence with respect to the basis set, and the smallest basis, aug-cc-pVDZ, shows a basis-set incompleteness error of about 1.5 eV, which is about the same order of magnitude as the method error. Increasing the basis set with respect to the cardinal number, as done for ground-state energy calculations, decreases the error only slowly and leads to errors of about 1.1 eV and 0.8 eV for aug-cc-pVTZ and aug-cc-pVQZ, respectively, indicating that no higher l-quantum numbers are missing in the basis set. A severe problem is that such a behavior can easily be misinterpreted as "proper" convergence if only ground-state basis sets are employed, although it is only a pseudo convergence because the missing (diffuse) functions are not taken into account. A fairly rapid convergence with respect to the basis set is obtained if additional diffuse functions are added, which are less important in ground-state calculations. This illustrates how computing certain properties can change the basis-set requirements drastically, which requires "experience-based" development and application of basis sets.

In the present example the computation of excitation energies with balanced quality is only possible if the basis-set incompleteness error can be controlled. However, even with quite large basis sets there is no guarantee to achieve basis-set convergence, and, in addition very large basis sets often lead to linear dependencies, convergence problems, and decreased computational efficiency (*vide infra*).

TABLE I. $^1B_{3u}$ excitation energies of C_2H_4 in eV calculated using different basis sets and coupled-cluster methods. For excitation energies configuration-interation singles (CIS) and coupled-cluster singles (CCS) yield identical results. ($R_{CC} = 2.5187$ bohr, $R_{CH} = 2.0422$ bohr, $\Theta_{HCH} = 121.35^o$)

	1^1B_{3u} (valence)		2^1B_{3u} (Rydberg)	
	CCS	CCSD	CCS	CCSD
aug-cc-pVDZ	7.745	8.042	10.620	10.430
aug-cc-pVTZ	7.730	8.027	10.135	10.372
aug-cc-pVQZ	7.726	8.030	9.834	10.183
d-aug-cc-pVDZ	7.725	8.014	9.043	9.349
d-aug-cc-pVTZ	7.718	8.016	9.046	9.449
d-aug-cc-pVQZ	7.719	n.c.[a]	9.045	n.c.[a]
MRA	7.718	—	9.033	—

[a] Calculation did not converge.

C. Basis-set convergence

We computed singlet and triplet CIS excitation energies of a small test set of molecules with different basis sets and compared to the basis-set limit calculations from MRA, as obtained with our new implementation. The test set consists of the six lowest singlet and six lowest triplet excitations of water, ammonia, ethylene, benzene, norbornane, cyclopropane, and formaldehyde, totaling 84 states. The basis sets were selected to represent different cardinal numbers (X=D, T, Q) of commonly used basis set families, as well as a benchmarking basis set (d-aug-cc-pVQZ). The aim of this section is not to provide an extensive assessment of basis sets, but to put the MRA results into context.

In Tab. II we compare MRA CIS to widely-used basis set families, namely Dunning-type and Karlsruhe-type. The correlation-consistent basis sets pioneered by Dunning, [11] abbreviated cc-pVXZ, are primarily constructed to provide accurate ground-state energies for correlated methods. Additional diffuse functions are provided through augmentation functions, leading to aug-cc-pVXZ basis sets. Furche and Rappoport [9] constructed optimized basis sets for response properties for density-functional theory (DFT)

TABLE II. The first 7 singlet and triplet CIS excitation energies of the ethylene molecule in eV, as well as excitation 2^1B_{3u}.

State	Γ	def2-SVPD	SVPD+d	def2-TZVPD	def2-TZVPD+CM2	aug-cc-pVQZ	d-aug-cc-pVQZ	MRA
T_1	$^3B_{3u}$	3.621	3.614	3.619	3.617	3.617	3.617	3.616
T_2	$^3B_{1u}$	7.580	7.009	7.194	6.942	6.927	6.922	6.922
T_3	$^3B_{3g}$	8.565	7.700	8.452	7.759	7.652	7.638	7.639
T_4	^3A_g	10.930	7.838	10.114	7.974	7.912	7.771	7.770
T_5	$^3B_{2g}$	8.599	7.860	8.128	7.792	7.788	7.771	7.772
T_6	2^3B_{3g}	9.157	8.536	8.565	8.505	8.496	8.495	8.495
T_7	2^3B_{1u}	9.912	8.612	9.285	8.611	8.824	8.506	8.478
S_1	$^1B_{1u}$	7.904	7.218	7.489	7.156	7.149	7.140	7.141
S_2	$^1B_{3u}$	8.064	7.781	7.896	7.818	7.726	7.719	7.718
S_3	$^1B_{3g}$	9.238	7.782	8.666	7.836	7.742	7.724	7.723
S_4	$^1B_{2g}$	8.823	7.955	8.310	7.897	7.901	7.878	7.877
S_5	^1A_g	11.890	8.156	11.105	8.192	8.403	8.101	8.101
S_6	2^1B_{1u}	10.536	8.666	9.881	8.679	8.903	8.587	8.570
S_7	^1A_u	11.347	8.846	10.328	8.824	8.894	8.788	8.742
...	...	...	...	...	...	...	...	...
S_{12}	2^1B_{3u}	13.116	9.111	12.248	9.021	9.834	9.045	9.033

applications based on static polarizability calculations. These basis sets have at most one additional *spd* set of diffuse basis functions and thus correspond approximately to the aug-cc-pVXZ basis sets, while being more compact and consisting of a reduced number of basis functions. Consequently all of these sets are sufficient for describing valence excited states, but fail to describe Rydberg states accurately.

Up to doubly augmented triple zeta basis sets are needed to converge excitation energies (or excited-state properties) to the basis set limit. Alternatively, the atom-centered basis sets are supplemented with diffuse center-of-mass (CM) basis sets to describe the Rydberg states. This procedure works very well, but still the ordering of the states might be incorrect, e.g. the $^3B_{1g}$ and the ^3A_g states of ethylene in case of the basis def2-

TZVPD+CM2, see Tab. II. Furthermore, the excitation energies might still exhibit large errors if the CM basis functions are combined with small valence basis set.

Some statistical measures of the lowest excitations of the test set molecules are given in Fig. 2, all with respect to MRA calculations which have been considered the basis set limit. No standard deviation or variance are given because the distributions are not Gaussian shaped. It can be seen that the doubly augmented d-aug-cc-pVQZ basis set exhibits only small errors and are in good agreement with the MRA calculations. However, this basis is very large and contains diffuse functions, which leads to a low computational efficiency. The singly augmented aug-cc-pVQZ can yield significant deviations > 0.5eV for 15 of the 84 states, but can be considered to be in reasonable agreement with the basis-set limit. In contrast, the CM-augmented def2-TZVPD basis set seems more balanced, since only few states exhibit large deviations. For medium-sized basis sets such as def2-TZVPD little can be said about the quality of the individual states due to the almost equal distribution of errors, although the basis set has already been optimized for response properties.

An alternative *ad-hoc* correction of the singly augmented basis sets can be the addition of a second set of diffuse functions, taken from the Dunning family, resulting in "doubly augmented" basis sets and in significantly improved behavior, "def2-SVPD+d" and "def2-TZVPD+d", which are obtained by adding the second set of diffuse functions from the d-aug-cc-pVDZ and d-aug-cc-pVTZ, respectively. This additional set of diffuse functions are essential for the description of Rydberg states, for both Karlsruhe-type or Dunning-type. In addition it is quite problematic that neglecting the second set of diffuse basis functions might lead to pseudo-convergence.

Assuming that CIS exhibits a standard deviation of about 1.2 eV and maximum errors as large as 2.4 eV, [32], the number of states inside an error bar of 1.0 eV count as follows: 47 for SVPD (56 %), 81 for SVPD+d (96 %), 69 for TZVPD (82 %), 83 for TZVPD+CM2 (99 %), 81 for aug-cc-pVQZ (96 %), d-aug-cc-pVQZ (100 %). This means, for instance, that for the smallest basis set, def2-SVPD, 37 states (about 50 %) exhibit errors larger than the method error. However, when for this basis set additional diffuse functions are included, leading to the basis denoted SVPD+d here, 35 (out of 37) states drop below 1.0 eV error and only two states remain beyond 1.0 eV error. We thus estimate that 35 states have significant Rydberg character, while the accuracy of the two remaining states can be increased by using TZVPD indicating significant valence character. Leaving out

TABLE III. CIS excitation energies in eV and oscillator strengths for selected excitations of the water molecule.

Γ		def2-TZVPD	TZVPD+CM	d-aug-cc-pVQZ	MRA (k=8,ϵ=1.e-4)
1^1B_2	Excitation energy	8.908	8.806	8.778	8.778
	–length	5.0×10^{-2}	4.9×10^{-2}	4.8×10^{-2}	4.8×10^{-2}
	–velocity	7.2×10^{-2}	7.0×10^{-2}	7.1×10^{-2}	7.1×10^{-2}
1^1A_2	Excitation energy	10.627	10.503	10.443	10.443
	–length	0	0	0	0
	–velocity	0	0	0	0
1^1A_1	Excitation energy	11.165	11.010	10.966	10.967
	–length	1.0×10^{-1}	8.1×10^{-2}	7.1×10^{-2}	7.0×10^{-2}
	–velocity	8.7×10^{-2}	7.6×10^{-2}	6.5×10^{-2}	6.4×10^{-2}
2^1B_2	Excitation energy	11.791	11.274	11.190	11.187
	–length	1.3×10^{-2}	8.2×10^{-3}	1.1×10^{-2}	1.1×10^{-2}
	–velocity	1.6×10^{-2}	7.4×10^{-3}	1.0×10^{-2}	1.0×10^{-2}
2^1A_1	Excitation energy	12.182	11.577	11.505	11.495
	–length	1.6×10^{-2}	2.4×10^{-2}	3.4×10^{-2}	3.6×10^{-2}
	–velocity	9.3×10^{-3}	1.6×10^{-2}	2.4×10^{-2}	2.6×10^{-2}
3^1B_2	Excitation energy	13.742	11.954	11.937	11.922
	–length	2.4×10^{-3}	4.2×10^{-4}	1.3×10^{-4}	2.0×10^{-5}
	–velocity	3.0×10^{-3}	2.5×10^{-3}	1.7×10^{-3}	1.3×10^{-3}

Rydberg-type excitations, it means that the SVPD basis set is able to obtain 47 out of 49 valence excitations below 1.2 eV accuracy, i.e. the estimated method error.

Within the limited test set the SVPD basis thus seems to yield an acceptable accuracy for the lowest excitations in the CIS approximation. For more accurate methods such as CCSD this is not true any more, since the method error is significantly lowered. In the TDDFT framework the situation is inversed, since the small basis set is not able to describe Rydberg or charge-transfer states, but neither is TDDFT, so that the errors might be balanced again, albeit on a low level.

The oscillator strengths show a similar behavior compared to the excitation energies.

Exemplary results are presented for the water molecule in Tab. III. All oscillator strengths as obtained with the best basis set, d-aug-cc-pVQZ, yield only small differences compared to the reference MRA results. One exception is the oscillator strength in the velocity representation belonging to the 3^1B_2 excitation with an excitation energy of 11.921 eV, which has an error of one order of magnitude. The error is again found in the LCAO calculations, where the oscillator strengths converge slower than the corresponding excitation energy. For properties of the excited states an accurate description of the excited wave functions is even more important than for the excitation energies by themselves.

D.   Spatial extent of excited states

In Fig. 3, we have plotted the MRA solution vectors $|x_{HOMO}\rangle$ corresponding to the highest occupied molecular orbital (HOMO) of the ethylene molecule for excitations S_1 and S_{12} from Tab. II, where excitation S_1 corresponds to a valence state and S_{12} to a Rydberg state. It can be clearly seen that the Rydberg state extends significantly beyond the ethylene molecule, which illustrates why basis sets suited for ground-state calculations are not suited in general for describing such a state. The shape of the solution vectors $|x\rangle$ reveals that no simple mapping to the molecular geometry is observed, which supports the qualitative observation that center-of-mass (CM) basis sets lead to significantly improved accuracy for Rydberg states, while increasing the cardinal number of a given basis set often hardly yields improvements (*vide supra*).

E.   Computation times

In MRA-CIS the most time consuming step for each iteration step is the calculation of the exchange operator applied on the response vector $|x\rangle$, scaling quadratically with the number of occupied orbitals N_{occ} and linear with the number of excitations. CPU times for one iteration of one excitation vector $|x\rangle$ are plotted for 6 different organic molecules in Fig. 4. The computational time grows approximately quadratically with the number of occupied orbitals. Conventional CIS calculations scale with N^2 to N^4, where the actual measured scaling depends on applied approximations and the size of the calculation leading to "in-core" or "out-of-core" algorithms. [1] A quadratic scaling is reached only
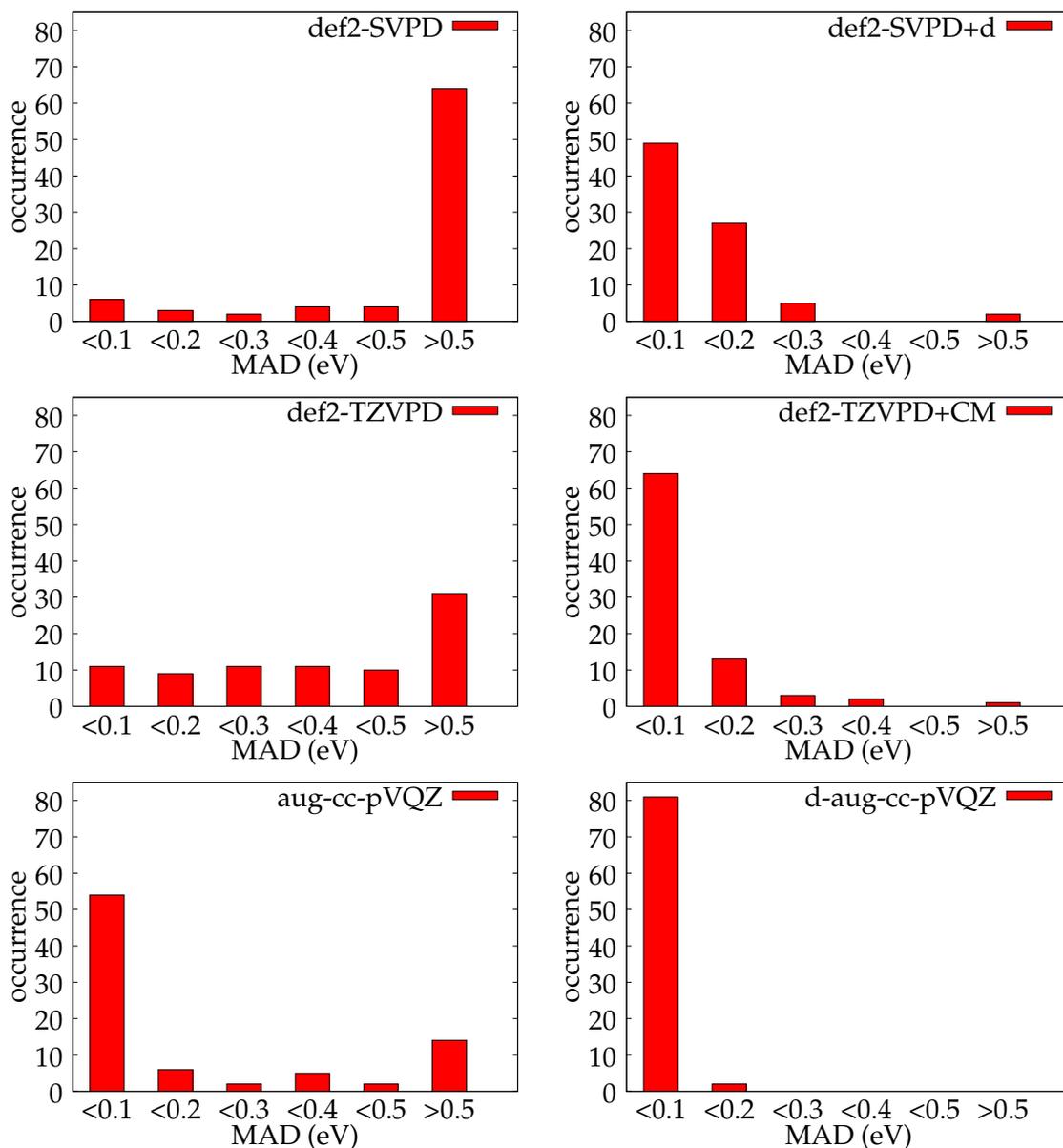
FIG. 2. Graphical representation of errors obtained using different basis sets compared to the MRA limit. Included are singlet and triplet states of the test set of 7 molecules. The total number of states is 84. Conventional basis sets are not able to account for Rydberg states and additional diffuse functions have to be used. MRA, on the other hand, is able to achieve accurate results for all states without special considerations.
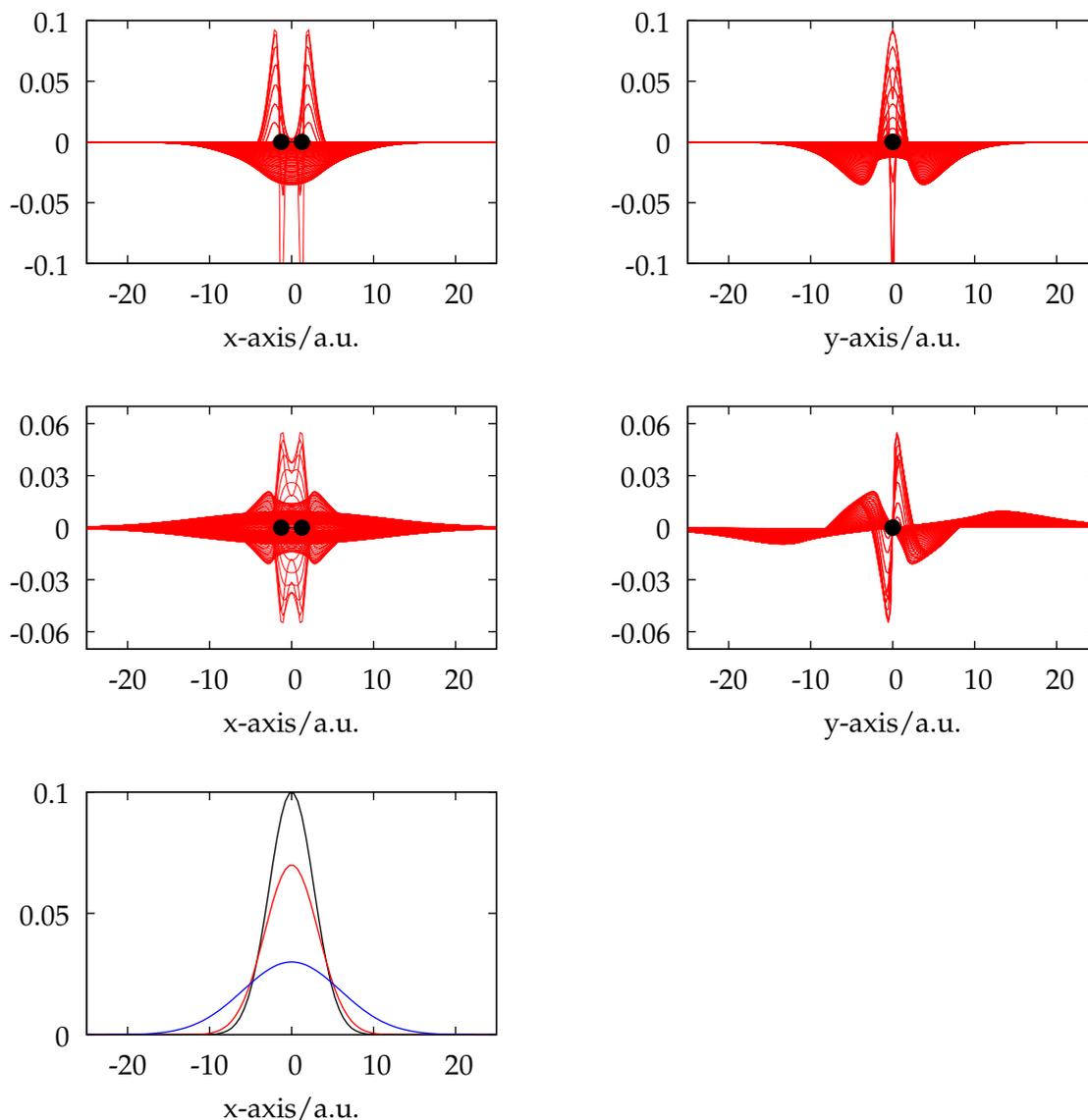
FIG. 3. Excited-state solution vectors corresponding to the HOMO for $^1B_{1u}$ (top) and 2^1B_{3u} (bottom) of the ethylene singlet excitations of Tab. II (left: view along y axis, right: view along x axis), respectively. The ethylene molecule is oriented along the x axis and the hydrogens are located in the xy plane. The graphical representation illustrates that the spatial extension of the valence excitation is significantly smaller compared to the Rydberg excitation. In the bottom plot the most diffuse s-functions from different basis sets are plotted. The exponents of the Gaussians are 0.067 (def2-SVPD, black), 0.047 (aug-cc-pVDZ, red) and 0.014 (d-aug-cc-pVDZ, blue).
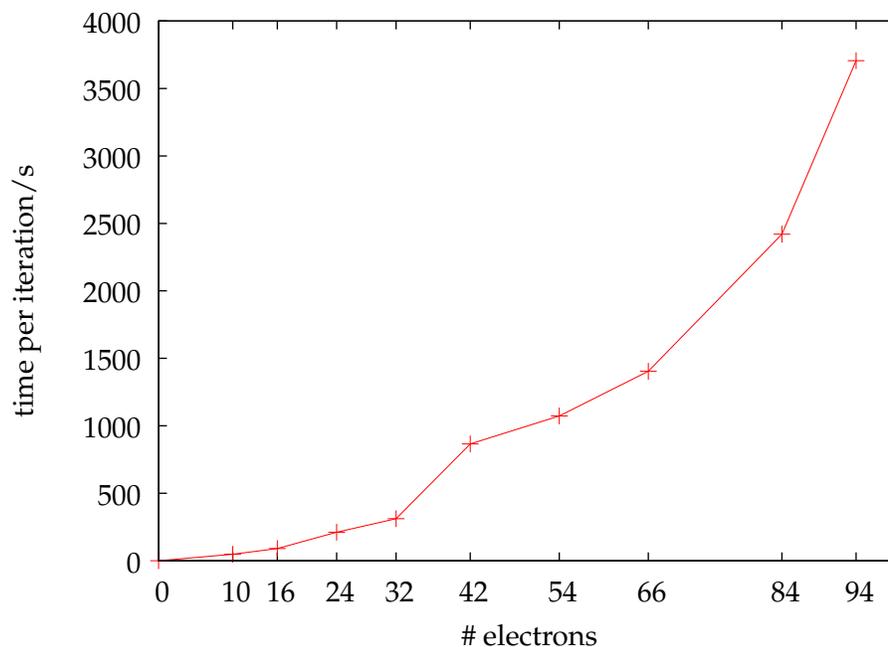
FIG. 4. Averaged CPU times for one iteration of one excitation vector in seconds, reveling a quadratic scaling of the MRA method as implemented. The molecules employed are water (10 electrons), formaldehyde (16 electrons), cyclopropane (24 electrons), acetone (32 electrons), benzene (42 electrons), norbornane (54 electrons), thymine (66 electrons), camphor (84 electrons) and phenantrene (94 electrons).

in calculations with small basis sets, while it usually cannot be achieved for diffuse basis sets.

While LCAO calculations have a fast initial convergence and take only seconds for small molecules and basis sets, MRA calculations have relatively large overhead due to the increased accuracy. This is shown in Fig. 5, where total CPU times for the calculation of the lowest 10 excitation energies of toluene are given.

V.   CONCLUSIONS

In the present work, we report an efficient implementation of CIS excitation energies using multi-resolution analysis (MRA). This method allows for an unbiased calculation of excitation energies guaranteeing the same accuracy for valence as well as Rydberg excited states *a priori*, i.e. without special treatment. While Gaussian basis sets are hard
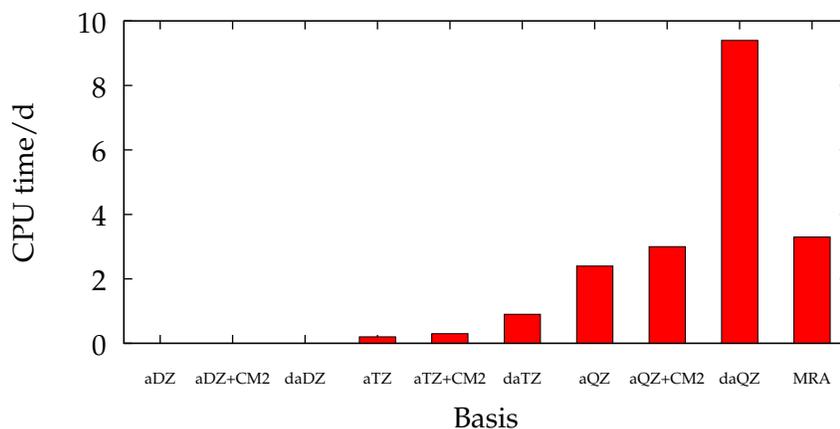
20

FIG. 5. Total CPU times in days for the calculation of the 10 lowest CIS singlet excitation energies of toluene (50 electrons) in C_1 symmetry with basis sets aug-cc-pVXZ (abbreviated aXZ) and d-aug-cc-pVXZ (abbreviated daXZ), as well as MRA.

to beat for lower accuracy, MRA enables the systematic exploration of the basis set limit.

The new MRA implementation was used to investigate the accuracy of excitation energies of selected compounds and basis sets. The analysis revealed a rather moderate performance for conventional basis sets when molecular absorption spectra are to be studied, because they are not able to yield accurate excitation energies for Rydberg states, which also occur among the lowest excitations. We found that only a very large and diffuse basis, d-aug-cc-pVQZ, can yield reliable excitation energies for both valence and Rydberg excitations. However, the (almost saturated) augmentation of larger basis sets does not seem desirable due to linear dependencies and resulting numerical instabilities. For instance, even for the simple ethylene molecule the CCSD calculation did not converge. This leads to the severe problem of pseudo convergence when, based on numerical issues, additional diffuse functions have to be excluded from further considerations in excited-state calculations of extended systems.

Based on the current implementation, the calculation of CIS excitation energies using MRA is comparably cheap and guarantees a numerically accurate treatment of excited states. The approach can be used as a benchmark for constructing or analyzing new basis sets. Due to the favorable scaling MRA CIS calculations can also be used directly to compute the UV/VIS spectra of small and medium-sized molecules without basis set artifacts.

21

The future prospects are twofold: First, the current implementation is massively parallel and can be used to compute the CIS excitation energies of large molecules, taking advantage of the quadratic scaling algorithm. Second, correlated wave-function models beyond CIS, such as CC2 or CCSD, will be developed for MRA to balance the error in the model. For these, the main obstacle is the higher dimensionality of the wave function, i.e. six dimensional (6D) for CC2 and CCSD instead of three dimensional (3D) in case of CIS, that needs to be overcome.[14] Work along these lines is in progress.

VI. ACKNOWLEDGMENTS

---

[1] F. Furche and D. Rappoport, Density functional methods for excited states: equilibrium structure and electronic spectra, in *Computational Photochemistry*, edited by M. Olivucci, pages 93–128, Elsevier: Amsterdam, 2005.

[2] A. Dreuw and M. Head-Gordon, Chem. Rev. **105**, 4009 (2005).

[3] O. Christiansen, P. Jørgensen, and C. Hättig, Int. J. Quant. Chem. **68**, 1 (1998).

[4] E. Schrödinger, Annalen der Physik **384**, 361 (1926).

[5] P. A. Dirac, Proc. Roy. Soc. A **114**, 243 (1927).

[6] E. Schrödinger, Annalen der Physik **384**, 734 (1926).

[7] K. A. Peterson, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).

[8] K. A. Peterson and K. E. Yousaf, J. Chem. Phys. **133**, 174116 (2010).

[9] D. Rappoport and F. Furche, J. Chem. Phys. **133**, 134105 (2010).

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

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

[12] R. J. Harrison, G. I. Fann, T. Yanai, Z. Gan, and G. Beylkin, J. Chem. Phys. **121**, 11587 (2004).

[13] T. Yanai, R. J. Harrison, and N. Handy, Mol. Phys. **103**, 413 (2005).

[14] F. A. Bischoff and E. F. Valeev, J. Chem. Phys. **139**, 114106 (2013).

[15] L. Frediani, E. Fossgaard, T. Flå, and K. Ruud, Mol. Phys. **111**, 1143 (2013).

[16] H. Sekino, Y. Maeda, T. Yanai, and R. J. Harrison, J. Chem. Phys. **129**, 034111 (2008).

[17] T. Kato, Y. Yokoi, and H. Sekino, Int. J. Quantum Chem. **113**, 286 (2012).

[18] S. Hamada and H. Sekino, Int. J. Quantum Chem. **111**, 1480 (2011).

[19] N. Vence, R. Harrison, and P. Krstić, Phys. Rev. A **85**, 033403 (2012).

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

[21] R. Bauernschmitt and R. Ahlrichs, Chem. Phys. Lett. **256**, 454 (1996).

[22] G. Beylkin and M. J. Mohlenkamp, Siam J. Sci. Comput. **26**, 2133 (2005).

[23] T. B. Pedersen and H. Koch, Chem. Phys. Lett. **293**, 251 (1998).

[24] TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

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

[26] R. A. Kendall, T. H. Dunning Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[27] D. E. Woon and T. H. Dunning, J. Chem. Phys. **100**, 2975 (1994).

[28] O. Christiansen et al., J. Chem. Phys. **105**, 6921 (1996).

[29] K. Kaufmann, W. Baumeister, and M. Jungen, J. Phys. B: At. Mol. Opt. Phys. **22**, 2223 (1999).

[30] MADNESS, Multiresolution Adaptive Numerical Environment for Scientific Simulation, http://code.google.com/p/m-a-d-n-e-s-s.

[31] R. J. Harrison, J. Comp. Chem. **25**, 328 (2002).

[32] H. Larsen, K. Hald, J. Olsen, and P. Jørgensen, J. Chem. Phys. **115**, 3015 (2001).