SCUOLA
NORMALE
SUPERIORE

Classe di Scienze

Corso di perfezionamento in
Metodi e modelli per le scienze molecolari

XXXIII ciclo

Time-dependent coupled-cluster for ultrafast spectroscopy

Settore Scientifico Disciplinare **CHIM/02**

Candidata
Dr.ssa Alice Balbi

Relatore

Prof. Henrik Koch

Anno accademico 2022−2023

Acknowledgements

When I embarked on my Ph.D. journey, I could have never imagined that a significant portion of it would be spent working from my bed, isolated from the outside world due to a global pandemic.

This situation made me question my life choices. In the face of all of that fragility, all of that people dying, with the awakening of the awareness that our time here can end anytime for anyone, even for me, was it really what I wanted to dedicate long years, day by day, in front of a computer, reading some theoretical chemistry paper or debugging some mind-breaking code?

Despite asking myself this question repeatedly, the answer remained unchanged: yes. Pursuing this path is not just something I enjoy, but it is an integral part of my identity. Without it, I wouldn't feel like myself anymore. Regardless of what may happen, I cannot envision a life without exploring the beauty of our universe, the joy of thinking, continuous learning, striving for understanding, and taking those infinitesimal yet precious steps forward in the grand puzzle that humankind is collectively unraveling.

I am grateful to myself for all the years of study that have led me to this point, but most importantly, for not giving up during the difficult times.

I am grateful to my supervisor, Henrik Koch, for the incredible opportunity he gave me by welcoming me into his group. I am thankful for everything I have learned during these years and for the privilege of working in such an interesting research field. Additionally, I would like to express my gratitude to all the members of his group, both at NTNU and SNS, for creating a stimulating and friendly environment. Special thanks go to Andreas, with whom I collaborated on all the papers included in this thesis.

I would also like to express my gratitude to the following significant persons:

To Andrea, my love, the bright spark that has given me the courage to always pursue what I truly desire, and to our non-human companions Tremor, Gremlin, and Chikito that made every day worth living.

To my dad, for his "Ph.D. student's pills" - daily motivational emails that, despite the physical distance, always made me feel his support.
To my mom, who never missed my phone calls and was always there to lend an ear whenever I needed to talk.
To my entire family, for their love and encouragement throughout my journey.

To Lucy, my best friend since we were three years old. Despite the distance, time constraints, and various challenges, she has always been an irreplaceable presence in my life, and I know she will always be.

To Skev, for all the memes and terrible jokes, for his invaluable assistance with anything related to computers, whether it was installing an operating system or helping me hack my favorite game, for always being a true friend and a wonderful person.

To my sweet puddings Matteo, Alberto, and Andrea that, while sharing the office, made me laugh too much and debate about probably anything the human mind can conceive and beyond.

To the friends I met in Pisa, to the wonderful people of "A code unite" who made me truly feel at home and filled all of my free time with something bigger than us, to Ginevra that was for me an inspiration of willpower, to all my fellow activists, to Martina, Patrizia, and Daniele for all the lives that we saved and for those that we lost, in which we will be forever bonded.

This thesis uses the Masters/Doctoral Thesis Class File Version 1.6 LaTeX template by Vel and Johannes Böttcher.

Contents

Chapter 1

Introduction

Chemical reactivity is governed by electronic motion, which occurs at an attosecond (1 as $= 10^{-18}$ s) time scale. Until the last century, it was not possible to observe it directly, as the shortest available laser pulses had duration in the order of femtoseconds (1 fs $= 10^{-15}$ s), which is the scale of nuclear motion [1, 2]. During the last two decades, big advances in laser technology, such as X-ray free electron lasers (XFELs) [3, 4], have led to the experimental realization of sub-femtosecond laser pulses, making possible real-time observation and control of electron dynamics in molecules [5–11]. This provided the potential for a deeper understanding of chemical reactions, including some fundamental biochemical mechanisms such as the early stages of photosynthesis [12,13] or DNA radiation damage [14]. In parallel to these technological advancements, it is fundamental to carry on theoretical and computational research that, apart from being interesting *per se*, is an important tool to design innovative experiments to be performed at laser facilities and to predict and interpret their outcomes.

The methods used to model the time evolution of electrons can be divided into two macrocategories: real-time electronic structure theory and frequency-domain response theory [15,16]. In real-time methods, the time-dependence of the system is considered explicitly by evolving the time-dependent Schrödinger equation (TDSE) in the time domain. Despite the methods to explicitly solve the TDSE began to be applied in the field of nuclear physics more than half century ago [17–21], until recently they didn't become very popular in the electronic structure community, mainly because of the lack of electronic correlation in the Hartree-Fock methods and the inaccessibly high computational cost for the correlated methods [22]. This changed with technological advancements that led to more powerful computers and efficient algorithms, which revived the research in real-time correlated methods like density functional theory [23,24], multiconfigurational self-consistent-field [25–27], configuration interaction [28–31], algebraic diagrammatic construction [32,33] and coupled-cluster [34–44]. When conducting computational chemistry studies on a system, it is essential to select an appropriate method in order to achieve a good balance between accuracy and computational cost. This evaluation depends not only on the nature of the system under study but also on the desired level of accuracy and the available computational resources. Computational chemistry methods are generally categorized as either ab initio or semiempirical. Unlike semiempirical methods, ab initio methods rely solely on physical principles in their calculations and do not incorporate experimental data. This characteristic makes them extendable to higher orders of accuracy at the cost of increased computational effort. On the other hand, while semiempirical methods offer the advantage of reduced computational cost, they are limited by their system-dependent accuracy.

Although it is impossible to declare one method as universally superior to another, in a specific situation, we can determine which method is more accurate, computally cost-effective, or suitable. When modeling a chemical system, it is often crucial to consider electronic correlation as an important aspect. It can be classified into static and dynamic correlation. Static correlation is linked to the presence of degenerate or quasi-degenerate states and, if not negligible, makes any description with a single determinant inadequate. On the other hand, dynamic correlation is directly related to the electronic motion. The Hartree-Fock (HF) method is the simplest ab initio method for electronic structure calculations. Despite the advantage of being computationally cheap, it has the significant drawback of neglecting completely the electronic correlation effects, except for the Pauli correlation. To account for dynamic correlation, post-HF methods such as coupled-cluster (CC), configuration interaction (CI), or Møller-Plesset perturbation theory (MP) are necessary. Coupled-cluster, while having a high computational cost and being nonvariational, has the advantage of being size-extensive at any truncation level and can provide results of the desired accuracy with sufficient computational resources. It recovers dynamical correlation well, but it cannot recover static correlation in its more common formulations, making it unsuitable for systems that cannot be effectively approximated by a single determinant wave function. However, by employing multi-reference formulations, it is possible to overcome this limitation and accurately capture static correlation.

In this thesis, the theoretical framework that we have chosen for the study of attochemical phenomena is that of time-dependent coupled-cluster theory [45], as it constitutes an accurate approach for the description of molecules in strong fields. In this method, the time-dependence is contained in the cluster amplitudes and multipliers. Their time evolution is described by means of first-order differential equations, that can be resolved using integrators like Runge-Kutta and Gauss-Legendre. As outlined in Publication I [46] of this thesis, we have successfully implemented these methods in $e^{\mathcal{T}}$, an open source electronic structure program with emphasis an on coupled-cluster techniques. In particular, this represented the first released implementation of a time-dependent coupled-cluster method. After the assessment of the accuracy of our integration procedures by comparison with results already present in the literature, we have used our code to calculate the electronic response to a pump-probe sequence of laser pulses, with the pump being valence-exciting and the probe being core-exciting. As higher level coupled-cluster methods computation cost scales rapidly with the size of the system, and quickly reaches the limits of practicability, we have chosen lithium hydride (LiH) for the convergence tests of parameters like the basis set, the size of the time steps, and the integration method. Then, we observed how the transient absorption spectrum of the lithium fluoride (LiF) molecule oscillates with the time delay between the pump pulse and the probe one. We attributed this effect to the quantum interference of states in the pump-induced superposition. The detailed results of our work on pump-probe transient absorption spectra of LiH and LiF can be found in Publication II [39] of this thesis.

We extended the aforementioned implementation to a time-dependent equation-of-motion coupled-cluster approach with the use of a reduced basis calculated

with an asymmetric band Lanczos algorithm, as detailed in Publication III [47] of this thesis. By increasing the band Lanczos chain length (i.e. the number of iterations) and including a sufficient number of relevant states, this approach converges to the same spectral features as the aforementioned time-dependent coupled-cluster method, but with significantly lower computational times, as we showed for lithium fluoride. We observed the limits of core-valence separation (CVS) approximation, showing how for the LiH molecule several peaks around the Li K-edge were not correctly retrieved. This was because for light atoms such as Li the energy separation between core and valence orbitals is small, so that pure valence excitation (which would be excluded by the core-only CVS projector) can fall within the core excitation region. Finally, we used the computational procedure to model attosecond transient absorption by the glycine molecule, which is a promising candidate for experimental investigations. We also modeled the electronic Raman effect [5, 48–51], in particular the electronic impulsive stimulated Raman scattering (ISXRS) [52] population transfer induced by an ultrashort laser pulse described as a semiclassical external electric field in dipole approximation by means of the time-depend equation-of-motion coupled-cluster (TD-EOM-CC) model. As described in Publication IV [53] of this thesis, we studied it for some small molecules, namely neon, carbon monoxide, pyrrole, and p-aminophenol, by performing some simulations in which the frequency of the ultrashort external electromagnetic pulse that irradiated the system corresponded to that of the lowest allowed core excitation of the considered system. We assessed the occurrence of the Raman effect by looking at the final population probabilities of the different states and noticing how some valence states become populated because of the de-excitation from core excited states. We simulated the real-time evolution of the electronic density of p-aminophenol interacting with an external electromagnetic field and visualized it through a movie showing isodensity surfaces of the difference between the time-dependent electronic and the ground state density. We were able to observe the movement of the electronic density and track its localization over time. The significance of the research work presented in this thesis lies in the importance of developing theoretical and computational tools aimed at improving our understanding, ability to predict phenomena, and possibility to design experiments in the field of attochemistry. Perhaps the most groundbreaking potential application of this research is the possibility to steer electrons through the use of external electric fields, thereby enabling direct control over chemical reactions. This would represent a major advance in our ability to manipulate chemical systems at the atomic scale and has enormous potential for driving technological advancements in the years to come.

1.1   List of publications included in the thesis

Paper I
Sarai D. Folkestad, Eirik F. Kjønstad, Rolf H. Myhre, Josefine H. Andersen, Alice Balbi, Sonia Coriani, Tommaso Giovannini, Linda Goletto, Tor S. Haugland, Anders Hutcheson, Ida-Marie Høyvik, Torsha Moitra, Alexander C. Paul, Marco Scavino, Andreas S. Skeidsvoll, Åsmund H. Tveten, and Henrik Koch
$e^{\mathcal{T}}$ 1.0: An open source electronic structure program with emphasis on coupled cluster and multilevel methods
J. Chem. Phys. **152,** 184103 (2020)

Paper II
Andreas S. Skeidsvoll, Alice Balbi, and Henrik Koch
Time-dependent coupled-cluster theory for ultrafast transient-absorption spectroscopy
Phys. Rev. A **102,** 023115 (2020)

Paper III
Andreas S. Skeidsvoll, Torsha Moitra, Alice Balbi, Alexander C. Paul, Sonia Coriani, and Henrik Koch
Simulating weak-field attosecond processes with a Lanczos reduced basis approach to time-dependent equation-of-motion coupled-cluster theory
Phys. Rev. A **105,** 023103 (2022)

Paper IV
Alice Balbi, Andreas S. Skeidsvoll, and Henrik Koch
Coupled cluster simulation of impulsive stimulated X-ray Raman scattering
Submitted | arXiv: 2305.19942 [physics.chem-ph]

1.2   Other publications

Other publications in the field not included in the thesis.

Iker Léon, Nicola Tasinato, Lorenzo Spada, Elena R. Alonso, Santiago Mata, Alice Balbi, Cristina Puzzarini, Jose L. Alonso, and Vincenzo Barone
Looking for the elusive imine tautomer of creatinine: different states of aggregation studied by quantum chemistry and molecular spectroscopy
ChemPlusChem **86,** 1374 (2021)

Zoi Salta, Nicola Tasinato, Jacopo Lupi, Rahma Boussessi, Alice Balbi, Cristina Puzzarini, and Vincenzo Barone
Exploring the Maze of $C_2N_2H_5$ Radicals and Their Fragments in the Interstellar Medium with the Help of Quantum-Chemical Computations
ACS Earth Space Chem. **4,** 774 (2020)

Chapter 2

Standard electronic structure methods

2.1 Second quantization

Second quantization is an alternative formulation of quantum mechanics that differs from the traditional first quantization approach. In the first quantization, observables are represented by operators, and states are represented by functions. However, in second quantization, states are also represented by operators, specifically, sequences of elementary creation and annihilation operators that act on the vacuum state. In the second quantization formulation, one of the essential concepts is the Fock space. This is a vector space in which each element is in a one-to-one correspondence with a Slater determinant with the orbitals in canonical order. The vectors $|\mathbf{k}\rangle$ of a fermionic Fock space are defined as

$$|\mathbf{k}\rangle = |k_1, k_2, ..., k_M\rangle \, , k_P = \begin{cases} 1 & \phi_p \text{ occupied,} \\ 0 & \phi_p \text{ unoccupied,} \end{cases} \tag{2.1}$$

where ϕ_p are elements of a basis of orthonormal spin orbitals. Creation and annihilation operators in the fermionic case are defined respectively as

$$a_P^\dagger |\mathbf{k}\rangle = \delta_{k_P 0} \Gamma_P^{\mathbf{k}} |k_1, ..., 1_p, ..., k_M\rangle \, , \tag{2.2}$$

$$a_P |\mathbf{k}\rangle = \delta_{k_P 1} \Gamma_P^{\mathbf{k}} |k_1, ..., 0_p, ..., k_M\rangle \, , \tag{2.3}$$

where $\Gamma_P^{\mathbf{k}}$ is the phase factor, defined as

$$\Gamma_P^{\mathbf{k}} = \sum_{Q=1}^{P-1} (-1)_Q^k. \tag{2.4}$$

The phase factor, $\Gamma_P^{\mathbf{k}}$, is introduced to ensure the correct antisymmetry of the fermionic wave function when applying the creation and annihilation operators. However, in practice, the anticommutation relations between these operators automatically take care of the phase factor, making any explicit reference to it unnecessary. The anticommutation relations between creation and annihilation operators are

$$\left[a_P^\dagger, a_Q^\dagger \right]_+ = 0, \tag{2.5}$$

$$\left[a_P, a_Q \right]_+ = 0, \tag{2.6}$$

$$\left[a_P^\dagger, a_Q\right]_+ = \delta_{PQ}. \tag{2.7}$$

In second quantization, the field- and spin-free molecular electronic Hamiltonian in nonrelativistic and Born-Oppenheimer approximation takes the following form

$$\hat{H} = \sum_{PQ} h_{PQ} a_P^\dagger a_Q + \frac{1}{2} \sum_{PQRS} g_{PQRS} a_P^\dagger a_R^\dagger a_S a_Q + h_{nuc}, \tag{2.8}$$

where

$$h_{PQ} = \int \phi_P^* \left(\mathbf{x}\right) \left(-\frac{1}{2}\nabla^2 - \sum_I \frac{Z_I}{r_I}\right) \phi_Q \left(\mathbf{x}\right) d\mathbf{x}, \tag{2.9}$$

$$g_{PQRS} = \int \int \frac{\phi_P^* \left(\mathbf{x}_1\right) \phi_R^* \left(\mathbf{x}_2\right) \phi_Q \left(\mathbf{x}_1\right) \phi_S \left(\mathbf{x}_2\right)}{r_{12}} d\mathbf{x}_1 d\mathbf{x}_2, \tag{2.10}$$

$$h_{nuc} = \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{R_{IJ}}. \tag{2.11}$$

As in the Born-Oppenheimer approximation the positions of the nuclei are fixed, h_{nuc} is a numerical constant for a given molecular geometry. It is interesting to notice how the terms of the Hamiltonian contain at most two electronic excitations: this will turn out to be useful in chapter 3.

2.2   The variational principle

The variational principle is a fundamental concept in quantum mechanics, which states that the solution of the Schrödinger equation $\hat{H}\left|0\right\rangle = E_0 \left|0\right\rangle$ is equivalent to finding the stationary points of the energy functional

$$E[\tilde{0}] = \frac{\langle\tilde{0}|\hat{H}|\tilde{0}\rangle}{\langle\tilde{0}|\tilde{0}\rangle}, \tag{2.12}$$

where $|\tilde{0}\rangle$ is a trial wave function, $\hat{H}$ is the Hamiltonian operator, and $E[\tilde{0}]$ is the energy associated with the trial wave function. In other words, it can be proven that the solutions of the Schrödinger equation are in a one-to-one correspondence with the stationary points of the energy functional $E[\tilde{0}]$. This principle manifests its practical utility through what is called *variational method*. In this method the wave function $|C\rangle$ is expressed in terms of a set of parameters C. The goal is to find which combination of parameters leads to the lowest energy possible, corresponding to a stationary point of the function

$$E(C) = \frac{\langle C|\hat{H}|C\rangle}{\langle C|C\rangle}. \tag{2.13}$$

The wave function corresponding to that energy is the best approximation of the ground state wave function that can be obtained from the variation of the

aforementioned parameters. The obtained energy represents an upper bound to the ground state energy.

2.3 Hellmann-Feynman theorem

The Hellmann-Feynman theorem provides a means to calculate the first-order change in energy resulting from a perturbation $\alpha\hat{V}$ of the Hamiltonian $\hat{H}$

$$\hat{H} \rightarrow \hat{H} + \alpha\hat{V}. \tag{2.14}$$

Specifically, according to the theorem, the derivative of the energy with respect to α evaluated at $\alpha = 0$ is equal to the expectation value of $\hat{V}$ with respect to the unperturbed wave function $|\Psi\rangle$

$$\left.\frac{dE(\alpha)}{d\alpha}\right|_{\alpha=0} = \langle\Psi|\hat{V}|\Psi\rangle. \tag{2.15}$$

This theorem can also be generalized to transition expectation values [45], as follows

$$\left.\frac{d}{d\alpha}\langle\Lambda(\alpha)|\hat{H}_0 + \alpha\hat{V}|\mathrm{CC}(\alpha)\rangle\right|_{\alpha=0} = \langle\Lambda|\hat{V}|\mathrm{CC}\rangle, \tag{2.16}$$

which will be explored further in section 3.2.

2.4 Hartee-Fock

The Hartree-Fock (HF) method serves as the basis for all post-HF models and is the simplest ab initio wave function model. Its main deficiency is its inability to describe electronic correlation except for the Fermi correlation. The basic idea behind the method is to approximate the true wave function, which exists only in the entire multi-electron Hilbert space, with a Slater determinant, which is the simplest representation for which the antisymmetry holds. In this representation, aside from Fermi correlation, each electron behaves as an independent particle. Therefore, we can formulate an effective one-electron Schrödinger equation for each electron in the system. The effective Hamiltonian of the one-electron Shrödinger equation is called the Fock operator and can be written as

$$\hat{f} = \sum_{pq} f_{pq}E_{pq}, \tag{2.17}$$

where the matrix $\mathbf{f}$, known as the Fock matrix, is symmetric. The eigenvalues of the Fock matrix correspond to the orbital energies. Koopmans' theorem states that in the case of closed-shell HF, the negative of the orbital energy of the highest occupied orbital corresponds to the ionization energy of the system. However, this only holds for closed-shell HF. The Hartree-Fock method is typically solved using an iterative technique known as the self-consistent field (SCF) method. The goal of the SCF method is to determine the best (i.e., lowest) energy achievable by varying

the orbital coefficients of the Slater determinant or, in second quantization, the configuration state function (CSF) that describes the system being studied. In second quantization, the SCF method starts with a reference occupation number vector as an initial guess. The Fock matrix is then constructed based on this guess and diagonalized to obtain new orbital coefficients. This process is repeated iteratively until the energy converges within a certain chosen threshold. These transformations are unitary so that the wave function can be written in the form

$$|\kappa\rangle = e^{-\hat{\kappa}} |0\rangle, \tag{2.18}$$

where $|0\rangle$ represents a reference single determinant state and $\hat{\kappa}$ is an anti-Hermitian operator. The Hartree-Fock energy in second quantization can be expressed as

$$E^{(0)} = \langle \text{HF}|\hat{H}|\text{HF}\rangle = \sum_{pq} D_{pq} h_{pq} + \frac{1}{2} \sum_{pqrs} d_{pqrs} g_{pqrs} + h_{nuc}, \tag{2.19}$$

where D_{pq} and d_{pqrs} are respectively the one- and two- density matrix elements and are defined as

$$D_{pq} = \langle \text{HF}|E_{pq}|\text{HF}\rangle, \tag{2.20}$$

$$d_{pqrs} = \langle \text{HF}|e_{pqrs}|\text{HF}\rangle. \tag{2.21}$$

where E_{pq} is the *singlet excitation operator*, defined as

$$E_{pq} = a_{p\alpha}^{\dagger} a_{q\alpha} + a_{p\beta}^{\dagger} a_{q\beta}, \tag{2.22}$$

and e_{pqrs} is the *two-electron excitation operator*, defined as

$$e_{pqrs} = E_{pq} E_{rs} - \delta_{qr} E_{ps}. \tag{2.23}$$

For a closed shell HF, the energy becomes

$$E^{(0)} = 2 \sum_{i} h_{ii} + \sum_{ij} \left(2 g_{iijj} - g_{ijji} \right) + h_{nuc}, \tag{2.24}$$

where the one-electron term represents the kinetic energy of electrons and their interaction with the nuclei, the two-electron term represents the interactions of the electrons among each other and h_{nuc} is a scalar term that represents the repulsion among the nuclei.

Chapter 3

The coupled-cluster model

Coupled-cluster (CC) is a powerful post-HF method that is especially well-suited for describing systems with significant dynamic correlation while it neglects static correlation, at least in its canonical formulations. One of the key advantages of CC is that it is size-extensive, as will be discussed in greater detail in Sec. 3.1. Coupled-cluster is generally viewed as a great compromise between accuracy and computational cost, to the point that its CCSD(T) (CC singles and doubles with perturbative triples) formulation is called "the gold standard". The idea behind coupled-cluster is to refine the HF description by adding contributions from electronic excitations involving transitions from occupied to virtual orbitals. By incorporating these additional contributions, the method is able to capture electronic correlation effects. The CC wave function can be written as

$$|\text{CC}\rangle = e^{\hat{T}} |\text{HF}\rangle , \qquad (3.1)$$

where $\hat{T}$ is the cluster operator, which is a sum of excitation operators $\hat{\tau}_\mu$ describing excitations from occupied to virtual orbitals and their corresponding amplitudes $t\mu$

$$\hat{T} = \sum_\mu t_\mu \hat{\tau}_\mu. \qquad (3.2)$$

The excitation operator $\hat{\tau}_\mu$ acting on the reference state $|\text{HF}\rangle$ creates an excited state $|\mu\rangle$ by promoting electrons from occupied to virtual orbitals, and the corresponding amplitude t_μ represents the weight of that virtual excitation in the description of the state.

An excitation μ is referred to as "single" if it involves the excitation of a single electron from an occupied orbital to a virtual orbital, "double" if it involves the excitation of two electrons from two occupied orbitals to two virtual orbitals, and so on. The operator $\hat{T}$ can be decomposed as a sum of operators, each corresponding to a specific order of excitation

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N, \qquad (3.3)$$

where $\hat{T}_1$ contains only single excitation operators, $\hat{T}_2$ contains only double excitation operators, and so on, with N being the number of electrons in the system. When performing an actual computation, typically only excitations of selected orders are included. For example, in CCS (coupled-cluster singles), only single excitations are included in the cluster operator, so that we have $\hat{T} = \hat{T}_1$, in CCSD (coupled-cluster singles and doubles) we have $\hat{T} = \hat{T}_1 + \hat{T}_2$, and so on. Unlike other methods like HF or configuration interaction (CI), the use of variational principle in order to

determine the CC wave function results in a highly complex set of equations. The computational cost of solving these equations is substantial, making the variational approach unusable for CC calculations. What is done instead is to multiply the Schrödinger equation from the left by $e^{-\hat{T}}$ and then left project it on the $\langle \text{HF}|$ state and on the excited states $\langle \mu|$, obtaining the following projected equations

$$\langle \text{HF}|\hat{H}|\text{CC}\rangle = E, \tag{3.4}$$

$$\langle \mu|\hat{H}^T|\text{HF}\rangle = 0, \tag{3.5}$$

where $\hat{H}^T$ is the similarity transformed Hamiltonian defined as $\hat{H}^T = e^{-\hat{T}}\hat{H}e^{\hat{T}}$. It's worth noting that $\hat{H}^T$ is not a Hermitian operator, even though the original Hamiltonian $\hat{H}$ is Hermitian. This is because the similarity transformed Hamiltonian $\hat{H}^T$ involves the cluster operator $\hat{T}$, which is not anti-Hermitian. The projection approach used in CC leads to a much more tractable set of equations compared to the variational principle approach. This is because the number of terms is limited by the bielectronic nature of the Hamiltonian operator, which causes terms containing bra and ket that differ by more than two excitations to be equal to zero. Equations 3.5 are solved iteratively by substituting at each iteration the value of energy obtained from 3.4 until convergence. It is important to note that, since $\hat{H}$ is a two-particle operator and due to the Brillouin theorem, the energy expression, even when considering the full expansion of $\hat{T}$, is given by

$$E = \langle \text{HF}|\hat{H}\left(1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right)|\text{HF}\rangle, \tag{3.6}$$

meaning that amplitudes higher than double order do not contribute directly to the energy, regardless of the applied level of truncation. Another important property of the coupled-cluster method is the *cluster-commutation condition*, that states that

$$k > 2s_A^- \implies [[\dots[[\hat{A}, \hat{T}_{n_1}], \hat{T}_{n_2}], \dots], \hat{T}_{n_k}] = 0. \tag{3.7}$$

Here s_A^- is the down rank of the operator $\hat{A}$ defined as

$$s_A^- = \frac{1}{2}\left(n_o^c + n_v^a\right), \tag{3.8}$$

where n_o^c is the number of creation operators in $\hat{A}$ for the occupied spin orbitals, while n_v^a is the number of the annihilation operators in $\hat{A}$ for the virtual spin orbitals. When $\hat{A}$ is the Hamiltonian $\hat{H}$, the highest down rank is 2, so its Baker–Campbell–Hausdorff (BCH) expansion will be at most quartic in the amplitudes

$$e^{-\hat{T}}\hat{H}e^{\hat{T}} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2}[[\hat{H}, \hat{T}], \hat{T}] + \frac{1}{6}[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}] + \frac{1}{24}[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}]. \tag{3.9}$$

In order to determine the coupled-cluster amplitudes, several iterative methods are available. We can view the left-hand side of Eq. 3.5 as a vector function $\Omega_\mu(\mathbf{t}) = \langle \mu|e^{-\hat{T}}\hat{H}e^{\hat{T}}|\text{HF}\rangle$. We can approximate the $\Omega_\mu(\mathbf{t})$ by expanding it up to

the linear term

$$\boldsymbol{\Omega}\left(\mathbf{t}^{(n)} + \Delta\mathbf{t}\right) \cong \boldsymbol{\Omega}^{(0)}\left(\mathbf{t}^{(n)}\right) + \boldsymbol{\Omega}^{(1)}\left(\mathbf{t}^{(n)}\right)\Delta\mathbf{t}, \tag{3.10}$$

where

$$\Omega_\mu^{(0)}\left(\mathbf{t}^{(n)}\right) = \langle\mu|e^{-\hat{T}^{(n)}}\hat{H}e^{\hat{T}^{(n)}}|\mathrm{HF}\rangle, \tag{3.11}$$

is the *vector function* at the nth iteration while

$$\Omega_{\mu\nu}^{(1)}\left(\mathbf{t}^{(n)}\right) = \langle\mu|e^{-\hat{T}^{(n)}}\left[\hat{H},\hat{\tau}_\nu\right]e^{\hat{T}^{(n)}}|\mathrm{HF}\rangle, \tag{3.12}$$

is the *Jacobian matrix*. We can put $\boldsymbol{\Omega}^{(0)}\left(\mathbf{t}^{(n)}\right) + \boldsymbol{\Omega}^{(1)}\left(\mathbf{t}^{(n)}\right)\Delta\mathbf{t} = 0$ neglecting all the terms that have a higher order than linear dependence on $\Delta\mathbf{t}$. From this, we obtain the relation

$$\boldsymbol{\Omega}^{(1)}\left(\mathbf{t}^{(n)}\right)\Delta\mathbf{t}^{(n)} = -\boldsymbol{\Omega}^{(0)}\left(\mathbf{t}^{(n)}\right), \tag{3.13}$$

that, together with the relation

$$\mathbf{t}^{(n+1)} = \mathbf{t}^{(n)} + \Delta\mathbf{t}^{(n)}, \tag{3.14}$$

gives to equations that can be iterated until convergence in order to obtain the coupled-cluster amplitudes. The major problem with this procedure, called *Newton's method*, is that it is very computationally demanding as it requires solving the set of linear equations 3.13 in every iteration. For this reason, other methods that avoid the need for solving the set of linear equations are more commonly used to obtain the coupled-cluster amplitudes. One such example is the algorithm called *perturbation-based quasi-Newton* methods, which assumes the Jacobian $\Omega_{\mu\nu}^{(1)}\left(\mathbf{t}^{(n)}\right)$ to be diagonally dominant, which is supported by the results of the Møller-Plesset theory. Thus it can be approximated by putting all off-diagonal terms to zero, so that from equation 3.13 we can write

$$\Delta t_\mu^{(n)} = -\epsilon_\mu^{-1}\Omega_\mu^{(0)}\left(\mathbf{t}^{(n)}\right), \tag{3.15}$$

where the term ϵ_μ represents the difference between the total energy of unoccupied orbitals and the total energy of occupied orbitals for the spin orbitals in $\hat{\tau}_\mu$. The convergence of this method can be improved significantly by the acceleration method known as *direct inversion of the iterative subspace* (DIIS). In this framework, the amplitudes are no longer obtained as in 3.14, but through the relation

$$\mathbf{t}^{(n)} = \sum_{k=1}^{n-1} w_k\mathbf{t}^{(k)} + w_n\left(\mathbf{t}^{(n-1)} + \Delta\mathbf{t}^{(n-1)}\right), \tag{3.16}$$

where the w_k are the interpolation weights. The w_k are determined by minimizing the norm of the averaged error vector in a least-squares sense

$$\Delta\mathbf{t}^{ave} = \sum_{k=1}^{n} w_k\Delta\mathbf{t}^{(k-1)}. \tag{3.17}$$

The interpolation weights must sum to unity

$$\sum_{k=1}^{n} w_k = 1. \tag{3.18}$$

As DIIS significantly reduces the iterations needed for convergence, it is used by default in many quantum chemistry programs, including the $e^{\mathcal{T}}$ program [46].

3.1   Size-consistency and extensivity in coupled-cluster

When assessing the validity of a computational method, one characteristic of interest is whether or not it has good scaling properties, i.e. whether or not it provides results of consistent quality when applied to systems of different sizes. A correct scaling is particularly important to compare results obtained for systems with different sizes, which is necessary for example, to compute the heat of reaction from the heat of formation of the involved species [54]. As *size-consistency* and *size-extensivity* are both terms related to scaling properties, there can be some confusion between the two, to the point that they are sometimes, improperly, used interchangeably. A method is referred to as *size-consistent* if, given two non-interacting infinitely separated subsystems A and B of a system AB, the sum of the energies calculated for each subsystem separately is equal to the global energy calculated for the whole system AB

$$E_{AB} = E_A + E_B. \tag{3.19}$$

When equation 3.19 holds, the energy E_{AB} is said to be *additively separable*. We can choose to describe the system by means of a set of orbitals each localized either on the subsystem A or on the subsystem B. Since the two subsystems are non-interacting by hypothesis, the term of the Hamiltonian that accounts for the interaction between the two subsystems A and B is zero and we can write the Hamiltonian of the global system as

$$\hat{H}_{AB} = \hat{H}_A + \hat{H}_B. \tag{3.20}$$

In this framework, amplitudes associated with orbitals localized on distinct, non-interacting fragments are equal to zero, thereby the cluster operator can be written as

$$\hat{T}_{AB} = \hat{T}_A + \hat{T}_B. \tag{3.21}$$

It follows that we can write the wave function of the global system as

$$
\begin{aligned}
|\text{CC}_{AB}\rangle &= e^{\hat{T}_{AB}} |\text{HF}_{AB}\rangle \\
&= e^{\hat{T}_A + \hat{T}_B} |\text{HF}_A \text{HF}_B\rangle \\
&= e^{\hat{T}_A} e^{\hat{T}_B} |\text{HF}_A \text{HF}_B\rangle \\
&= e^{\hat{T}_A} |\text{HF}_A\rangle \, e^{\hat{T}_B} |\text{HF}_B\rangle \\
&= |\text{CC}_A\rangle |\text{CC}_B\rangle .
\end{aligned}
\tag{3.22}
$$

A wave function that can be expressed as the product of the wave functions of two non-interacting subsystems is referred to as *multiplicatively separable*. From the result in Eq. 3.22 it follows that

$$
\begin{aligned}
E_{AB} |\text{CC}_{AB}\rangle &= \hat{H}_{AB} |\text{CC}_{AB}\rangle \\
&= \left(\hat{H}_A + \hat{H}_B \right) |\text{CC}_A\rangle |\text{CC}_B\rangle \\
&= \left(\hat{H}_A |\text{CC}_A\rangle \right) |\text{CC}_B\rangle + |\text{CC}_A\rangle \left(\hat{H}_B |\text{CC}_B\rangle \right) \\
&= E_A |\text{CC}_A\rangle |\text{CC}_B\rangle + |\text{CC}_A\rangle E_B |\text{CC}_B\rangle \\
&= (E_A + E_B) |\text{CC}_{AB}\rangle .
\end{aligned}
\tag{3.23}
$$

This implies that the energy is *additively separable* as defined in Eq. 3.19, and therefore that the coupled-cluster method is size-consistent. It is worth noting that this property holds for any truncation level of the cluster operator. To check for size-consistency in practice, one can perform an energy calculation for the global system and compare the result to the sum of the energies obtained for each subsystem separately.

On the other hand, a method is considered *size-extensive* if its energy scales linearly with N, where N is large and indicates the number of repeating units of the system under investigation, such as nodes on a lattice. Note that to assess size-extensivity it is not relevant whether or not the considered subsystems are interacting, while when speaking on size-consistency they have to be non-interacting. Additionally, size-consistency refers only to situations in which the two considered subsystems are at infinite distance, while size-extensivity has well-defined meaning for any separation between the considered N subsystems [55]. The concept of size-extensivity is not obviously extendable to non-periodic systems [56], but it can be shown that connected diagrams scale proportionally to N. Therefore, any method that satisfies the Brueckner-Goldstone linked-diagram theorem [57] - as is the case for coupled-cluster - is size-extensive [58]. As size-extensivity lacks a natural operational definition, it is difficult to demonstrate that this property holds for methods that do not support a diagrammatic representation, such as density functional theory (DFT). The coupled-cluster method is both size-consistent and size-extensive at any truncation level. This characteristic, which is missing in some other important methods such as CI, is desirable as it allows to retain a good scaling behavior at lower computational cost [59], which is important in many cases,

such as for calculations regarding bigger molecules or dissociation reactions.

3.2   Lagrangian coupled-cluster

Numerous properties in quantum chemistry can be expressed in terms of energy derivatives. When the energy is determined variationally (see Sec. 2.2), its derivative with respect to a parameter can be expressed by means of the Hellmann-Feynman theorem (see Sec. 2.3). Unfortunately, this is not the case for coupled-cluster energies, that are determined by a nonvariational projection technique. However, by paying the price of considering an additional set of parameters $\bar{t}_\nu$ called multipliers, a variational reformulation of coupled-cluster energy can be done. In the *Lagrangian reformulation of coupled-cluster theory* [60,61], this is accomplished by writing the coupled-cluster Lagrangian as

$$L\left(\boldsymbol{t}, \bar{\boldsymbol{t}}\right) = \langle\Lambda|\hat{H}e^{\hat{T}}|\mathrm{HF}\rangle, \qquad (3.24)$$

where the dual state $\langle\Lambda|$ is

$$\langle\Lambda| = \langle\mathrm{HF}| + \sum_{\nu>0}\bar{t}_\nu\,\langle\nu|\,e^{-\hat{T}}. \qquad (3.25)$$

The amplitudes and multipliers can then be determined variationally from the stationarity conditions for the Lagrangian

$$0 = \frac{\partial L}{\partial\bar{t}_\mu} = \langle\mu|\hat{H}^T|\mathrm{HF}\rangle, \qquad (3.26)$$

$$0 = \frac{\partial L}{\partial t_\mu} = \langle\mathrm{HF}|\hat{H}^T|\mu\rangle + \sum_\nu\bar{t}_\nu\,\langle\nu|\left[\hat{H}^T, \tau_\mu\right]|\mathrm{HF}\rangle. \qquad (3.27)$$

The second stationarity condition is often expressed as

$$\boldsymbol{0} = \eta^T + \bar{\boldsymbol{t}}\mathbf{A}, \qquad (3.28)$$

where the elements of the vector η are

$$\eta_\mu = \langle\mathrm{HF}|\hat{H}^T|\mu\rangle, \qquad (3.29)$$

while the elements of $\mathbf{A}$, called *Jacobian* matrix, are given by

$$A_{\mu\nu} = \langle\mu|\left[\hat{H}^T, \hat{\tau}_\nu\right]|\mathrm{HF}\rangle. \qquad (3.30)$$

When considering a perturbed Hamiltonian $\hat{H} = \hat{H}_0 + \alpha\hat{V}$, we can write the coupled-cluster Lagrangian as

$$L(\alpha, \mathbf{t}, \bar{\mathbf{t}}) = \langle\mathrm{HF}|(\hat{H} + \alpha\hat{V})e^{\hat{T}}|\mathrm{HF}\rangle + \sum_\nu\bar{t}_\nu\,\langle\nu|e^{-\hat{T}}\left(\hat{H} + \alpha\hat{V}e^{\hat{T}}\right)|\mathrm{HF}\rangle. \qquad (3.31)$$

This is a variational expression that allows us to exploit the Hellmann-Feynman theorem to obtain the first-order change in energy

$$\frac{dE}{d\alpha}\Big|_{\alpha=0} = \frac{\partial L}{\partial \alpha}\Big|_{\alpha=0} = \langle \Lambda|\hat{V}|\mathrm{CC}\rangle. \tag{3.32}$$

Note that this generalization of Hellmann-Feynman theorem to coupled-cluster does not depend on the perturbed amplitudes and gives size-extensive first-order properties.

3.3 The equation-of-motion coupled-cluster model

When dealing with excited states in the coupled-cluster framework, one might initially consider treating each excited state separately with a calculation similar to that done for the ground state, but with a different choice of a zero-order determinant for each state of interest. However, this approach would be very expensive and selecting an appropriate zero-order determinant for excited states would often be impossible. The equation-of-motion coupled-cluster (EOM-CC) method [62–64] is an approach that can be used to describe CC excited states. In this framework, the right excited states are parametrized as

$$|\Psi_i\rangle = e^{\hat{T}}\hat{R}_i|\mathrm{HF}\rangle, \tag{3.33}$$

while the left excited states are parametrized as

$$\langle\widetilde{\Psi}_i| = \langle\mathrm{HF}|\,\hat{L}_i e^{-\hat{T}}, \tag{3.34}$$

where $\hat{R}_i$ and $\hat{L}_i$ are the right excitation operator and left excitation operator for the ith state respectively. The right and left operators $\hat{R}_i$ and $\hat{L}_i$ can be expressed as linear expansions in a finite set of operators $\hat{\tau}_\kappa, \hat{\tau}_\kappa^\dagger$,

$$\hat{R}_i = \sum_\kappa \hat{\tau}_\kappa r_{\kappa i}, \quad \hat{L}_i = \sum_\kappa l_{i\kappa}\hat{\tau}_\kappa^\dagger, \tag{3.35}$$

where the operator with index 0 is the unit operator,

$$\hat{\tau}_0 = \hat{\tau}_0^\dagger = 1. \tag{3.36}$$

In this model, the exponential $e^{\hat{T}}$ describes the correlation common to both the ground state and all of the excited states, while the parameters $r_{\kappa i}$ and $l_{i\kappa}$ reflect the differences between the ground state and the considered excited state. However, this distinction is rough as the cluster operator $\hat{T}$ doesn't provide an optimal description of the correlation of the excited states as it was optimized for the ground state. Therefore, some of the effects of the correlation will inevitably be contained in the expansion coefficients. We assume that the EOM-CC states are biorthonormal, which means that

$$\left(\widetilde{\Psi}_i|\Psi_j\right) = \delta_{ij}, \tag{3.37}$$

where we introduced the notation $|\cdot) = e^{\hat{T}} |\cdot\rangle$ and $(\cdot| = \langle\cdot| e^{-\hat{T}}$.

Additionally, the determinants that can be generated from the operators $\hat{\tau}_\mu$ and $\hat{\tau}_\mu^\dagger$ are defined to be biorthonormal

$$(\mu|\nu) = \langle\mu|\nu\rangle = \delta_{ij}. \tag{3.38}$$

The reason for using biorthonormality rather than orthonormality in the usual sense is that the latter would lead to cumbersome calculations, as there is no commutation between the operators contained in $\hat{T}$ and those in $\hat{T}^\dagger$. After some mathematical manipulation, we can see that the excitation energies can be obtained as eigenvalues of the following equations

$$\mathbf{A}\mathbf{R}_j = \mathbf{R}_j \Delta E_j, \tag{3.39}$$

$$\mathbf{L}_j^T \mathbf{A} = \Delta E_j \mathbf{L}_j^T, \tag{3.40}$$

where $\mathbf{A}$ is the Jacobian matrix as defined in Eq. 3.30, $\Delta E_j = E_j - E_0$, $L_{i\mu} = l_{i\mu}$ and $R_{\nu j} = r_{\nu j}$ for $\mu > 0$ and $\nu > 0$.

Chapter 4

Electron dynamics using coupled-cluster theory

The time evolution of quantum systems is described by the time-dependent Schrödinger equation, which, in atomic units, can be written as

$$i\frac{\partial\Psi(t)}{\partial t} = \hat{H}(t)\Psi(t). \tag{4.1}$$

In order to describe time-dependent phenomena by means of the coupled-cluster method, it is necessary to introduce time-dependence into the parametrization of the coupled-cluster wave function. Specifically, in the real-time time-dependent coupled-cluster (TDCC) approach, this is done by allowing the coupled-cluster amplitudes t_μ, multipliers $\bar{t}_\mu$, and phase factor ϵ as functions of time, leading to the following TDCC wave function [45]

$$|\text{CC}(t)\rangle = e^{\hat{T}(t)}|\text{HF}\rangle e^{i\epsilon(t)}, \tag{4.2}$$

where $\hat{T}(t) = \sum_\mu t_\mu(t)\hat{\tau}_\mu$. Consequently, the time-dependent Schrödinger equation (TDSE) in the coupled-cluster formulation can be written as

$$i\frac{d}{dt}|\text{CC}(t)\rangle = \hat{H}(t)|\text{CC}(t)\rangle. \tag{4.3}$$

We left-project Eq. 4.3 onto $\langle\text{HF}|$, obtaining

$$i\langle\text{HF}|\frac{d}{dt}|\text{CC}(t)\rangle = \langle\text{HF}|\hat{H}|\text{CC}(t)\rangle, \tag{4.4}$$

that can be written as

$$i\langle\text{HF}|\sum_\mu \dot{t}_\mu(t)\hat{\tau}_\mu|\text{CC}(t)\rangle - \dot{\epsilon}(t)\langle\text{HF}|e^{\hat{T}(t)}|\text{HF}\rangle e^{i\epsilon(t)} = \langle\text{HF}|\hat{H}e^{\hat{T}(t)}|\text{HF}\rangle e^{i\epsilon(t)}, \tag{4.5}$$

from which, using that $\langle\text{HF}|\hat{\tau}_\mu = 0$ and $\langle\text{HF}|\text{CC}\rangle = 1$, we obtain the **time derivative of the phase factor** $\epsilon(t)$

$$\dot{\epsilon}(t) = -\langle\text{HF}|\hat{H}(t)e^{\hat{T}(t)}|\text{HF}\rangle. \tag{4.6}$$

Note that the phase factor $\epsilon(t)$ is irrelevant for the calculation of physical observables as $e^{i\epsilon(t)}$ cancels out, but the knowledge of the expression for $\dot{\epsilon}(t)$ will be useful in the following derivations. To obtain the differential equations describing the time

evolution of each amplitude t_μ, we can left-multiply the coupled-cluster TDSE by $e^{-\hat{T}(t)}$ on both sides and then left-project it onto the $\langle\mu|$ corresponding to the amplitude of interest, obtaining

$$i\,\langle\mu|e^{-\hat{T}(t)}\frac{d}{dt}|\mathrm{CC}(t)\rangle = \langle\mu|e^{-\hat{T}(t)}\hat{H}(t)|\mathrm{CC}(t)\rangle, \qquad (4.7)$$

that can be written as

$$i\,\langle\mu|e^{-\hat{T}(t)}\left(\sum_\nu \dot{t}_\nu(t)\hat{\tau}_\nu\right)e^{\hat{T}(t)}|\mathrm{HF}\rangle\,e^{i\epsilon(t)} - \dot{\epsilon}(t)\,\langle\mu|e^{-\hat{T}(t)}e^{\hat{T}(t)}|\mathrm{HF}\rangle\,e^{i\epsilon(t)} = \\ \langle\mu|e^{-\hat{T}(t)}\hat{H}(t)e^{\hat{T}(t)}|\mathrm{HF}\rangle\,e^{i\epsilon(t)}. \qquad (4.8)$$

After simplifying the $e^{-\hat{T}(t)}$ operator with the $e^{\hat{T}(t)}$ factor where possible, and eliding the $e^{i\epsilon(t)}$ factors as they are present in all of the terms of the equation, we obtain

$$i\sum_\nu \dot{t}_\nu(t)\,\langle\mu|\nu\rangle - \dot{\epsilon}(t)\,\langle\mu|\mathrm{HF}\rangle = \langle\mu|e^{-\hat{T}(t)}\hat{H}(t)e^{\hat{T}(t)}|\mathrm{HF}\rangle. \qquad (4.9)$$

Considering that $\langle\mu|\nu\rangle = \delta_{\mu\nu}$ and $\langle\mu|\mathrm{HF}\rangle = 0$, we obtain the **time derivative of the cluster amplitude** $t_\mu(t)$

$$\dot{t}_\mu(t) = -i\,\langle\mu|e^{-\hat{T}(t)}\hat{H}(t)e^{\hat{T}(t)}|\mathrm{HF}\rangle. \qquad (4.10)$$

Analogously, we can parametrize the time-dependence of the coupled-cluster dual state as

$$\langle\Lambda(t)| = \left(\langle\mathrm{HF}| + \sum_\mu \bar{t}_\mu(t)\,\langle\mu|\,e^{-\hat{T}(t)}\right)e^{-i\epsilon(t)}, \qquad (4.11)$$

and obtain the differential equations describing the time evolution of the multipliers from the TDSE for the dual state by multiplying it by $e^{\hat{T}(t)}$ on the right and right-projecting it on $|\nu\rangle$ obtaining

$$\frac{d}{dt}\left(\langle\Lambda(t)|\right)e^{\hat{T}(t)}|\nu\rangle = i\,\langle\Lambda(t)|\hat{H}(t)e^{\hat{T}(t)}|\nu\rangle. \qquad (4.12)$$

We can write the left-hand side (LHS) of Eq. 4.12 as

$$\mathrm{LHS} = \sum_\mu \dot{\bar{t}}_\mu(t)\,\langle\mu|e^{-\hat{T}(t)}e^{\hat{T}(t)}|\nu\rangle\,e^{-i\epsilon(t)} - \\ \sum_\mu \bar{t}_\mu(t)\,\langle\mu|\left(\sum_\lambda \dot{t}_\lambda(t)\hat{\tau}_\lambda\right)e^{-\hat{T}(t)}e^{\hat{T}(t)}|\nu\rangle\,e^{-i\epsilon(t)} - \\ i\dot{\epsilon}(t)e^{-i\epsilon(t)}\,\langle\mathrm{HF}|e^{\hat{T}(t)}|\nu\rangle - i\dot{\epsilon}(t)\sum_\mu \bar{t}_\mu(t)\,\langle\mu|e^{-\hat{T}(t)}e^{\hat{T}(t)}|\nu\rangle\,e^{-i\epsilon(t)}. \qquad (4.13)$$

We simplify the $e^{-i\epsilon(t)}$ factor as it is present in every term, elide $e^{-\hat{T}(t)}$ with $e^{\hat{T}(t)}$, and note that $\langle \text{HF}|e^{\hat{T}(t)}|\nu\rangle = 0$ so that we obtain

$$\dot{\bar{t}}_\nu(t) - \sum_{\mu\lambda} \bar{t}_\mu(t)\dot{t}_\lambda(t) \langle \mu|\hat{\tau}_\lambda|\nu\rangle - i\dot{\epsilon}(t)\bar{t}_\nu(t) = i\left\langle \widetilde{\Lambda}(t)\middle| \hat{H}(t)e^{\hat{T}(t)}\middle|\nu\right\rangle, \qquad (4.14)$$

where $\left\langle \widetilde{\Lambda}(t)\right| = \langle HF| + \sum_\mu \bar{t}_\mu(t) \langle \mu| e^{-\hat{T}(t)}$. We substitute the time derivatives of amplitudes and phase factor with their values obtained in Eqs. 4.6 and 4.10

$$\dot{\bar{t}}_\nu(t) + i\sum_{\mu\lambda} \bar{t}_\mu(t) \langle \lambda|e^{-\hat{T}(t)}\hat{H}(t)e^{\hat{T}(t)}|\text{HF}\rangle \langle \mu|\hat{\tau}_\lambda|\nu\rangle + i\langle \text{HF}|\hat{H}(t)e^{\hat{T}(t)}|\text{HF}\rangle \bar{t}_\nu(t) =$$

$$= i\left\langle \widetilde{\Lambda}(t)\middle| \hat{H}(t)e^{\hat{T}(t)}\middle|\nu\right\rangle. \tag{4.15}$$

We define $\left|\widetilde{CC}(t)\right\rangle = e^{\hat{T}(t)}|HF\rangle$ and introduce a commutator in the right-hand side (RHS) of Eq. 4.15

$$\begin{aligned} \text{RHS} = {} & i\left\langle \widetilde{\Lambda}(t)\middle| \hat{H}(t)\hat{\tau}_\nu\middle|\widetilde{CC}(t)\right\rangle = \\ & i\left\langle \widetilde{\Lambda}(t)\middle| \left[\hat{H}(t), \hat{\tau}_\nu\right]\middle|\widetilde{CC}(t)\right\rangle + i\left\langle \widetilde{\Lambda}(t)\middle| \hat{\tau}_\nu\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle = \\ & i\left\langle \widetilde{\Lambda(t)}\middle| \left[\hat{H}(t), \hat{\tau}_\nu\right]\middle|\widetilde{CC}(t)\right\rangle + i\sum_\mu \bar{t}_\mu(t) \left\langle \mu\middle|\hat{\tau}_\nu e^{-\hat{T}(t)}\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle, \end{aligned} \tag{4.16}$$

so that we obtain

$$\begin{aligned} \dot{\bar{t}}_\nu(t) = {} & -i\sum_{\mu\lambda} \bar{t}_\mu(t) \langle \lambda|e^{-\hat{T}(t)}\hat{H}(t)e^{\hat{T}(t)}|\text{HF}\rangle \langle \mu|\hat{\tau}_\lambda|\nu\rangle - i\langle \text{HF}|\hat{H}(t)e^{\hat{T}(t)}|\text{HF}\rangle \bar{t}_\nu(t) + \\ & i\left\langle \widetilde{\Lambda}(t)\middle| \left[\hat{H}(t), \hat{\tau}_\nu\right]\middle|\widetilde{CC}(t)\right\rangle + i\sum_\mu \bar{t}_\mu(t) \left\langle \mu\middle|\hat{\tau}_\nu e^{-\hat{T}(t)}\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle. \end{aligned} \tag{4.17}$$

We insert the resolution of identity $\hat{\mathbf{1}} = |\text{HF}\rangle \langle \text{HF}| + \sum_\eta |\eta\rangle \langle \eta|$ between $\hat{\tau}_\nu$ and $e^{-\hat{T}(t)}$ in the last term of the RHS of Eq. 4.17

$$i\sum_\mu \bar{t}_\mu(t) \left\langle \mu\middle|\hat{\tau}_\nu e^{-\hat{T}(t)}\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle =$$

$$i\sum_\mu \bar{t}_\mu(t) \langle \mu|\nu\rangle \left\langle \text{HF}\middle|\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle + i\sum_{\mu\eta} \bar{t}_\mu(t) \langle \mu|\hat{\tau}_\nu|\eta\rangle \left\langle \eta\middle|e^{-\hat{T}(t)}\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle =$$

$$i\bar{t}_\nu(t) \left\langle \text{HF}\middle|\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle + i\sum_{\mu\eta} \bar{t}_\mu(t) \langle \mu|\hat{\tau}_\nu|\eta\rangle \left\langle \eta\middle|e^{-\hat{T}(t)}\hat{H}(t)\middle|\widetilde{CC}(t)\right\rangle. \tag{4.18}$$

So that inserting this result into Eq. 4.17 we obtain

$$
\begin{aligned}
\dot{\bar{t}}_\nu(t) = &-i \sum_{\mu\lambda} \bar{t}_\mu(t) \left\langle \lambda | e^{-\hat{T}(t)} \hat{H}(t) e^{\hat{T}(t)} | \mathrm{HF} \right\rangle \left\langle \mu | \hat{\tau}_\lambda | \nu \right\rangle - i \left\langle \mathrm{HF} | \hat{H}(t) e^{\hat{T}(t)} | \mathrm{HF} \right\rangle \bar{t}_\nu(t) + \\
&i \left\langle \widetilde{\Lambda}(t) \middle| \left[\hat{H}(t), \hat{\tau}_\nu \right] \middle| \widetilde{\mathrm{CC}}(t) \right\rangle + i\bar{t}_\nu(t) \left\langle \mathrm{HF} \middle| \hat{H}(t) \middle| \widetilde{\mathrm{CC}}(t) \right\rangle + \\
&i \sum_{\mu\eta} \bar{t}_\mu(t) \left\langle \mu | \hat{\tau}_\nu | \eta \right\rangle \left\langle \eta \middle| e^{-\hat{T}(t)} \hat{H}(t) \middle| \widetilde{\mathrm{CC}}(t) \right\rangle,
\end{aligned}
$$
$$(4.19)$$

that, by cancelling out terms with opposite sign, leads to the equation describing the **time derivative of the multipliers**, $\dot{\bar{t}}_\nu(t)$

$$
\dot{\bar{t}}_\nu(t) = i \left\langle \widetilde{\Lambda}(t) \middle| \left[\hat{H}(t), \hat{\tau}_\nu \right] \middle| \widetilde{\mathrm{CC}}(t) \right\rangle. \tag{4.20}
$$

The differential equations 4.10 and 4.20 can be solved using the methods presented in Sec. 5.1.

4.1 The time-dependent equation-of-motion coupled-cluster model

In the time-dependent equation-of-motion coupled-cluster model (TD-EOM-CC) [30, 34, 65–67], which is the time-dependent version of the equation-of-motion coupled-cluster (EOM-CC) model described in Sec. 3.3, the time-dependence is introduced in the coefficients of the linear expansion of the left and right wave functions

$$
|\Psi(t)\rangle = \sum_i |\psi_i\rangle s_i(t), \tag{4.21}
$$

$$
\left\langle \widetilde{\Psi}(t) \right| = \sum_i k_i(t) \left\langle \tilde{\psi}_i \right|. \tag{4.22}
$$

Here the time-independent bra and ket are the ones defined in Eqs. 3.34 and 3.33 and the index i runs over the time-independent EOM-CC states, including the ground state, that are parametrized as in Eqs. 3.34 and 3.33. The time derivative of the coefficients $s_i(t)$ and $k_i(t)$ can be obtained by projecting the ket and bra time-dependent Schrödinger equations respectively onto the bra and the ket of EOM-CC state, obtaining

$$
i\frac{\partial s_m(t)}{\partial t} = \sum_j H_{mn}(t)s_n(t), \tag{4.23}
$$

$$
-i\frac{\partial k_n(t)}{\partial t} = \sum_m k_m(t)H_{mn}(t), \tag{4.24}
$$

where $H_{ij}(t) = \langle \widetilde{\psi}_i | \hat{H}(t) | \psi_j \rangle$. Once the equations 4.23 and 4.24 are solved, they can be used to calculate the time-dependent expectation value of an operator $\hat{X}$

$$\langle \hat{X} \rangle(t) = \sum_{ij} k_i(t) X_{ij} s_j(t). \tag{4.25}$$

The matrix element X_{ij} can be rewritten as

$$
\begin{aligned}
X_{ij} &= \langle \widetilde{\psi}_i | X | \psi_j \rangle \\
&= \sum_{\kappa\lambda} l_{i\kappa} \langle \kappa | \hat{X}^T | \lambda \rangle r_{\lambda j} \\
&= \sum_{\nu} l_{i0} \langle \mathrm{HF} | \hat{X}^T | \nu \rangle r_{\nu j} + \sum_{\mu\nu} l_{i\mu} \langle \mu | \hat{X}^T | \nu \rangle r_{\nu j} + \left(l_{i0} X_{00} + \sum_{\mu} l_{i\mu} \xi_{\mu}^X \right) r_{0j} \\
&= \sum_{\nu} l_{i0} \langle \mathrm{HF} | \left[\hat{X}^T, \hat{\tau}_\nu \right] | \mathrm{HF} \rangle r_{\nu j} + \sum_{\mu\nu} l_{i\mu} \left({}^{\mathrm{LR}}A_{\mu\nu}^X + \langle \mu | \hat{\tau}_\nu \hat{X}^T | \mathrm{HF} \rangle \right) r_{\nu j} \\
&\quad + \left(l_{i0} X_{00} + \sum_{\mu} l_{i\mu} \xi_{\mu}^X \right) r_{0j},
\end{aligned}
\tag{4.26}
$$

where

$$\hat{X}^T = e^{-\hat{T}} \hat{X} e^{\hat{T}}, \tag{4.27}$$

$$ {}^{\mathrm{LR}}A_{\mu\nu}^X = \langle \mu | \left[\hat{X}^T, \hat{\tau}_\nu \right] | \mathrm{HF} \rangle , \tag{4.28}$$

$$X_{00} = \langle \mathrm{HF} | \hat{X}^T | \mathrm{HF} \rangle , \tag{4.29}$$

$$\xi_{\mu}^X = \langle \mu | \hat{X}^T | \mathrm{HF} \rangle . \tag{4.30}$$

We can compare this expression with that for the coupled-cluster response theory [68]. Let's consider the matrix element X_{ij} in the specific case where the left vector corresponds to an excited state ($i = m$) and the right vector corresponds to the ground state ($j = 0$). By applying these condition to Eq. 4.26, we obtain

$$X_{m0} = \sum_{\mu} l_{m\mu} \xi_{\mu}^X, \tag{4.31}$$

which is linear in the excited determinant component $l_{m\mu}$. When, conversely, we consider the situation in which the left vector corresponds to the ground state ($i = 0$) and the right vector corresponds to an excited state ($j = n$), we obtain

$$X_{0n} = \sum_{\nu} \left({}^{\mathrm{LR}}\eta_{\nu}^X + \sum_{\mu} \bar{t}_{\mu} \langle \mu | \hat{\tau}_\nu \hat{X}^T | \mathrm{HF} \rangle \right) r_{\nu n} - \left(X_{00} + \sum_{\mu} \bar{t}_{\mu} \xi_{\mu}^X \right) \sum_{\nu} \bar{t}_{\nu} r_{\nu n}, \tag{4.32}$$

where

$$ {}^{\mathrm{LR}}\eta_{\nu}^X = \langle \mathrm{HF} | \left[\hat{X}^T, \hat{\tau}_\nu \right] | \mathrm{HF} \rangle + \sum_{\mu} \bar{t}_{\mu} \, {}^{\mathrm{LR}}A_{\mu\nu}^X. \tag{4.33}$$

In Eq. 4.32, which is equivalent to Eq. (65) in Ref. [68], we can observe that, apart from the term $\sum_\nu {}^{\mathrm{LR}}\eta_\nu r_{\nu n}$ that is also present in the response theory [45], the other terms are specific to the EOM-CC framework. Finally, we consider the case where both left and right vectors correspond to excited states $(i = m, j = n)$. We obtain

$$
\begin{aligned}
X_{mn} &= \sum_{\mu\nu} l_{m\mu}\Big({}^{\mathrm{LR}}A^X_{\mu\nu} + \langle\mu|\,\hat\tau_\nu \hat X^T\,|\mathrm{HF}\rangle \Big) r_{\nu n} - \sum_\mu l_{m\mu}\xi^X_\mu \sum_\nu \bar t_\nu r_{\nu n} \\
&= \sum_{\mu\nu} l_{m\mu}\Big({}^{\mathrm{LR}}A^X_{\mu\nu} + \langle\mu|\,\hat\tau_\nu \hat X^T\,|\mathrm{HF}\rangle - \xi^X_\mu \bar t_\nu \Big) r_{\nu n} \\
&= \sum_{\mu\nu} l_{m\mu}\Big({}^{\mathrm{EOM}}A^X_{\mu\nu} + \delta_{\mu\nu} X_{00} - \xi^X_\mu \bar t_\nu \Big) r_{\nu n},
\end{aligned}
\tag{4.34}
$$

where

$$
\begin{aligned}
{}^{\mathrm{EOM}}A^X_{\mu\nu} &= \langle\mu|\,\hat X^T\,|\nu\rangle - \delta_{\mu\nu} X_{00} \\
&= {}^{\mathrm{LR}}A^X_{\mu\nu} + \langle\mu|\,\hat\tau_\nu \hat X^T\,|\mathrm{HF}\rangle - \delta_{\mu\nu} X_{00},
\end{aligned}
\tag{4.35}
$$

as defined in Eq. (18) of Ref. [69].

We employed the asymmetric band Lanczos algorithm described in Sec. 5.2 to construct a reduced basis for TD-EOM-CC. To assess the accuracy and reliability of this approach, we compared the obtained results with those from TDCC (time-dependent coupled-cluster) calculations, as discussed in detail in [47].

4.2   External electromagnetic field

In our studies, all the systems under investigation consisted of an atomic or molecular system interacting with an external electromagnetic field. We assumed that the only time-dependent term in the Hamiltonian was the one representing the interaction with the external electromagnetic field

$$
H\,(t) = H_0 + V\,(t). \tag{4.36}
$$

Here, H_0 represents the time-independent part of the Hamiltonian, while $V(t)$ describes the semiclassical time-dependent interaction with the external electromagnetic field written in the dipole approximation and length gauge

$$
V\,(t) = -\mathbf{d} \cdot \boldsymbol{\mathcal{E}}(t). \tag{4.37}
$$

In this expression, $\mathbf{d}$ is the electric dipole moment operator, which is time-independent, and $\boldsymbol{\mathcal{E}}(t)$ is the time-dependent external electric field. The electric field can be written as a combination of time-dependent electric fields associated with individual laser pulses

$$
\boldsymbol{\mathcal{E}}(t) = \sum_n \boldsymbol{\mathcal{E}}_{0n} \cos(\omega_{0n}(t - t_{0n})) f_n(t). \tag{4.38}
$$

For each laser pulse n, $\boldsymbol{\mathcal{E}}_{0n}$ represents the amplitude, ω_{0n} is the carrier frequency, t_{0n} is the central time of the pulse and $f_n(t)$ is the envelope function that determines the shape of the pulse. When choosing an envelope function for the time-dependent electric field, it is often desirable to use a shape that closely resembles the physical pulse. One commonly employed choice is the Gaussian envelope

$$f_n(t) = \begin{cases} e^{-(t-t_{0n})^2/(2\sigma_n^2)} & a_n \leq t \leq b_n, \\ 0 & \text{otherwise,} \end{cases} \tag{4.39}$$

where $a_n = t_{0n} - N\sigma_n$ and $b_n = t_{0n} + N\sigma_n$. The energy exchanged between the electromagnetic field and the molecule is given by

$$\Delta E = -\int_{-\infty}^{\infty} \mathbf{d}(t) \cdot \frac{\partial \boldsymbol{\mathcal{E}}(t)}{\partial t} \, dt \,. \tag{4.40}$$

This equation can be frequency resolved by taking the Fourier transforms of the dipole moment $\mathbf{d}(t)$ and electric field $\boldsymbol{\mathcal{E}}(t)$

$$\widetilde{d}_i(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d_i(t) e^{-i\omega t} \, dt \,, \tag{4.41}$$

$$\widetilde{\mathcal{E}}_i(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{E}_i(t) e^{-i\omega t} \, dt \,. \tag{4.42}$$

The response function $S(\omega)$ is defined through the expression

$$\Delta E = \int_0^{\infty} \omega S(\omega) \, d\omega \,. \tag{4.43}$$

By combining this definition, Eq. 4.40, the Fourier transforms 4.41, and 4.42, we can write the response function as [1]

$$S(\omega) = -2 \operatorname{Im} \left(\widetilde{\mathbf{d}}(\omega) \cdot \widetilde{\boldsymbol{\mathcal{E}}}^*(\omega) \right), \quad \omega > 0. \tag{4.44}$$

It represents the absorption per unit frequency at a given frequency: positive (negative) $\omega S(\omega)$ equals the amount of energy gained (lost) by the molecule per unit frequency at ω.

---

[1] Here we followed the procedure of [70] but with a different Fourier transform convention, so we ended up with a response function $S(\omega)$ with a different sign.

Chapter 5

Computational tools

This chapter serves to provide an overview of the essential computational tools required for our calculations. We will discuss various algorithms that are instrumental in solving differential equations, Lanczos algorithm, and core-valence separation (CVS) approximation.

5.1 Methods for the solution of ordinary differential equations

A differential equation represents a relationship between a function, which is the unknown, and its derivatives. In general, the unknown function can depend on multiple independent variables. However, in the case of an ordinary differential equation (ODE), the unknown function is dependent on only one independent variable. Solving an ODE to obtain a closed-form expression for the unknown function is often challenging or even impossible. However, for practical purposes, obtaining a numerical solution given an initial condition is usually sufficient. Numerical methods for ODE solving can be categorized as either *explicit* or *implicit*. An *explicit* numerical method allows us to approximate the value of the unknown function at a point $t + \Delta t$ using an expression of the form $f(t + \Delta t) = g(f(t))$, where f represents the unknown function and g is a function that relies on the specific numerical method employed. In contrast, an *implicit* numerical method involves solving an equation of the form $g(f(t), f(t + \Delta t)) = 0$ to determine the value of $f(t + \Delta t)$. Here, g is a function that depends on the chosen numerical method. The *order* of a numerical method is a parameter that establish how well the method approximates the solution. An nth-order method has a local truncation error (the error caused by one iteration) of the order of $\mathcal{O}(h^{n+1})$, while the global truncation error (the cumulative error caused by many iterations) is of the order of $\mathcal{O}(h^n)$, where h is the chosen step size. An important class of numerical methods for the solution of ODE in the form $\frac{df}{dt} = g(t, f)$ with $f(t_0) = f_0$ as initial condition, is that of *Runge-Kutta methods* that in general can be schematized as

$$f_{n+1} = f_n + h \sum_{i=1}^{s} b_i k_i, \qquad (5.1)$$

where the k_i are defined as

$$k_i = g\left(t_n + c_i h, f_n + h \sum_{j=1}^{m} a_{ij} k_j\right), \qquad (5.2)$$

where $m = i - 1$ for explicit methods while $m = s$ for implicit methods, h is the step, f_n is the approximation to $f(t_n)$ where t_n is $t_0 + nh$. The a_{ij}, b_i, and c_i parameters are elements of the Butcher tableau, that is defined for each method and, in general, has the following form

$$
\begin{array}{c|cccc}
c_1 & a_{11} & a_{12} & \dots & a_{1s} \\
c_2 & a_{21} & a_{22} & \dots & a_{2s} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_s & a_{s1} & a_{s2} & \dots & a_{ss} \\
\hline
 & b_1 & b_2 & \dots & b_s
\end{array}
\quad = \quad
\begin{array}{c|c}
\boldsymbol{c} & \boldsymbol{A} \\
\hline
 & \boldsymbol{b^T}
\end{array}
$$

It is worth noting that for explicit Runge-Kutta methods, the sum over j in the computation of the k_i values is limited to $i - 1$. Consequently, the coefficient matrix $\boldsymbol{A}$ in the Butcher tableau for explicit methods will be lower triangular. This characteristic ensures that the computation of k_i at each stage only depends on the previously computed stages. On the other hand, this lower triangular property does not hold for implicit Runge-Kutta methods. In implicit methods, the sum over j extends to s, resulting in a more general coefficient matrix $\boldsymbol{A}$ in the Butcher tableau. The inclusion of the additional terms allows implicit methods to capture a wider range of dynamics but at the cost of increased computational complexity since the stages are interdependent.

The *Euler method* serves as the simplest form of the Runge-Kutta methods. It is a first-order explicit method, and its Butcher tableau is given by

$$
\begin{array}{c|c}
0 & 0 \\
\hline
 & 1
\end{array}
$$

Despite its simplicity, the Euler method has a significant drawback. Being a first-order method, it often necessitates very small step sizes to achieve sufficient accuracy. Consequently, a large number of steps is required, resulting in a substantial computational cost. As a result, higher-order methods are typically preferred over the Euler method to strike a balance between accuracy and computational efficiency.

The most widely used and known Runge-Kutta method is the explicit fourth-order method commonly referred to as *the* Runge-Kutta method or RK4. It is characterized by the following Butcher tableau

$$
\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{2} & \frac{1}{2} & & & \\
\frac{1}{2} & 0 & \frac{1}{2} & & \\
1 & 0 & 0 & 1 & \\
\hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}
$$

The concepts and methods discussed previously for the one-dimensional case can be readily extended to the multidimensional case. The procedure is analogous regardless of the number of dimensions involved.

Gauss-Legendre methods belong to the class of implicit Runge-Kutta methods. The Butcher tableau of the fourth-order Gauss-Legendre (GL4) method is

$$
\begin{array}{c|cc}
\frac{1}{2} - \frac{1}{6}\sqrt{3} & \frac{1}{4} & \frac{1}{4} - \frac{1}{6}\sqrt{3} \\
\frac{1}{2} + \frac{1}{6}\sqrt{3} & \frac{1}{4} + \frac{1}{6}\sqrt{3} & \frac{1}{4} \\
\hline
& \frac{1}{2} & \frac{1}{2}
\end{array}
$$

while that of sixth-order Gauss-Legendre (GL6) method is

$$
\begin{array}{c|ccc}
\frac{1}{2} - \frac{1}{10}\sqrt{15} & \frac{5}{36} & \frac{2}{9} - \frac{1}{15}\sqrt{15} & \frac{5}{36} - \frac{1}{30}\sqrt{15} \\
\frac{1}{2} & \frac{5}{36} + \frac{1}{24}\sqrt{15} & \frac{2}{9} & \frac{5}{36} - \frac{1}{24}\sqrt{15} \\
\frac{1}{2} + \frac{1}{10}\sqrt{15} & \frac{5}{36} + \frac{1}{30}\sqrt{15} & \frac{2}{9} + \frac{1}{15}\sqrt{15} & \frac{5}{36} \\
\hline
& \frac{5}{18} & \frac{4}{9} & \frac{5}{18}
\end{array}
$$

In contrast to explicit methods, where the values of the different k_i are obtained directly, implicit methods yield a system of m equations in which the k_i serve as the unknowns. To solve this system, various algorithms can be employed, e.g. the iterative method called *fixed point iteration*. The number of iterations required to solve these equations cannot be determined in advance, as it depends on the specific problem and desired level of accuracy. Consequently, the number of evaluations needed for each time step cannot be known in advance. This characteristic often leads to a higher computational cost compared to explicit methods. However, the use of Gauss-Legendre methods can be justified by their symplecticity, meaning that they can reproduce many physical conservation laws (e.g. conservation of energy) with a high degree of accuracy [41].

5.2 Lanczos algorithm

The Lanczos algorithm is an iterative projection method that enables the transformation of a large Hermitian matrix $\mathbf{A}$ into a smaller tridiagonal form $\mathbf{T}$. This algorithm begins with the selection of a starting vector $\mathbf{v}$ which will be used to represent the Hermitian matrix $\mathbf{A}$ in the orthogonal basis for the Krylov subspace

$$\mathcal{K}_j(\mathbf{A}, \mathbf{v}) = \text{span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \dots, \mathbf{A}^{j-1}\mathbf{v}\}, \tag{5.3}$$

in which it will assume a tridiagonal form $\mathbf{T}$. At each step of the algorithm, apart from scalar product and additions of a multiple of a vector to another, only one matrix-vector multiplication is needed, and one row and column of the matrix $\mathbf{T}$ is obtained. When $\mathbf{A}$ is a non-Hermitian matrix, the left and right eigenvectors will be different

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \qquad \text{and} \qquad \mathbf{y}^*\mathbf{A} = \lambda\mathbf{y}^*. \tag{5.4}$$

In contrast to the case where $\mathbf{A}$ is Hermitian, when dealing with a non-Hermitian matrix $\mathbf{A}$ the Lanczos algorithm requires the use of two starting vectors: $\mathbf{v}$ and $\mathbf{w}$. These vectors are chosen such that $\mathbf{v}^\dagger \cdot \mathbf{w} = 1$, which will lead to two Krylov spaces, $\mathcal{K}_j(\mathbf{A}, \mathbf{v})$ and $\mathcal{K}_j(\mathbf{A}, \mathbf{w})$ respectively and to two sets of eigenvectors, left and

right. The asymmetric band Lanczos algorithm [71] is a variation of the standard asymmetric Lanczos algorithm that can be more advantageous in certain scenarios, such as when dealing with matrices containing degenerate or closely clustered eigenvalues. In this variation, instead of the starting left and right vectors, m right $(\mathbf{b}_1, \ldots, \mathbf{b}_m)$ and p left $(\mathbf{c}_1, \ldots, \mathbf{c}_p)$ starting vectors are employed. From them, the sequences

$$\mathcal{K}_i(\mathbf{M}, \mathbf{b}_1, \ldots, \mathbf{b}_m) = \mathrm{span} \left\{ \underbrace{\mathbf{b}_1, \ldots, \mathbf{b}_m, \mathbf{M}\mathbf{b}_1, \ldots, \mathbf{M}\mathbf{b}_m, \mathbf{M}^2\mathbf{b}_1, \ldots}_{i} \right\}, \qquad (5.5)$$

and

$$\mathcal{K}_j(\mathbf{M}^T, \mathbf{c}_1, \ldots, \mathbf{c}_p) = \mathrm{span} \left\{ \underbrace{\mathbf{c}_1, \ldots, \mathbf{c}_p, \mathbf{M}^T\mathbf{c}_1, \ldots, \mathbf{M}^T\mathbf{c}_p, (\mathbf{M}^T)^2\mathbf{c}_1, \ldots}_{j} \right\}, \quad (5.6)$$

are generated, with $i - n$ and $j - n$ redundant vectors in the sequence respectively. The goal of this algorithm is then to construct suitable left and right linearly independent Lanczos vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ that constitutes a basis for the subspaces spanned by the first j $\mathbf{v}_i$ and $\mathbf{w}_i$ vectors respectively.

5.3   Core-valence separation approximation

Core excitation energies are internal to the spectrum of the Hamiltonian. In general, when using exterior eigenvalue methods such as Davidson or Lanczos algorithms, one needs to calculate all of the valence excitations before being able to calculate the core ones. This is really computationally demanding. The core-valence separation (CVS) approximation allows to neglect the matrix elements arising from an interaction between valence only excited states and excited states containing reference to core excitation. As a result, the Hamiltonian matrix can be separated into two distinct blocks: one block corresponds to valence-only excitations, while the other block represents interactions between excited states containing at least one core excitation. By diagonalizing these two blocks separately, we can focus directly on computing the core excitations of interest. This approach significantly reduces the computational burden, as we can start the computation from the lowest excitation involving at least one core state. By circumventing the need to calculate all valence excitations first, a considerable amount of computational time is saved.

Chapter 6

Experimental generation of attosecond pulses

The electronic motion inside molecules occurs on a subfemtosecond time scale. Due to the technological advances of the last decades, the generation of pulses down to a few tens of attoseconds became possible [72]. This opened up the way towards the possibility to follow the electronic motion in real-time in atoms, molecules, nanostructures, and solids and to steer it in order to directly control chemical reactions. This is likely to have a tremendous impact in a lot of different fields [73, 74] from material science, to biology (photosynthesis [12, 13], radiation damage and repair [75–77], cellular respiration, electron tunneling through proteins [78]), to physics and, of course, chemistry (from new synthetic to new spectroscopic techniques). As the shortest possible pulse of a given wavelength is one cycle long, pulses lasting exactly 1 femtosecond (fs) have a maximum wavelength of 300 nm. Consequently, subfemtosecond pulses exhibit wavelengths that correspond to the ultraviolet (UV) and X-ray regions of the electromagnetic spectrum. Synchrotrons [79] have emerged as a prominent X-ray radiation source in modern times. Over the past three decades, this technology has played a vital role in scientific advancements, reaching its pinnacle with the development of third-generation synchrotrons, which rely on undulators. Undulators are devices constituted of a structure of dipole magnets with alternating poles that create a static magnetic field that forces the approaching electrons into a oscillating path with a period determined by the distance between the magnets. As a consequence of this motion, the electrons emit radiation along the undulator's axis. Usually the distance between the centers of the magnets constituting the undulator is in the order of centimeters but, considering that the speed of the electrons is close to the speed of light, it is perceived by the electrons as in the order of $10^{-10}m$ because of the Lorentz contraction and of the Doppler relativistic effect, so the emitted radiation is an X-ray one. The undulators find application also in X-ray free electron lasers (XFELs) [80, 81], a technology that, in the context of the generation of ultrashort laser pulses, represents a major advancement with respect to synchrotrons, as it allows to obtain shorter pulses with higher peak power (10^3 W for 3rd generation synchrotrons, 10^{10} W for XFELs). In XFELs, the electrons are first accelerated to almost the speed of light by a linear particle accelerator (linac). Then, they enter inside an undulator that makes them oscillate in the directions transverse to that of their propagation so that they start to emit radiation. In XFELs, when a bunch of electrons is injected into the undulator, it has uniform density distribution. As the emitted radiation travels faster than the electrons themselves, it interacts with them, slowing down some of them and accelerating others depending on the phase with respect to each

other until the electrons organize themselves converging, with longitudinal density modulation becoming more pronounced the more the electromagnetic field grows intense and vice versa, with the radiation power that grows exponentially with the distance along the undulator, in the so-called electron crystals: small clusters called *microbunches* regularly spaced apart by a distance corresponding to the wavelength of the emitted radiation. When the modulation reaches saturation so that the electron beam is completely bunched, the electrons emit quasi-coherent radiation almost in phase. This phenomenon, which does not occur in synchrotrons but only in XFELs, is called self-amplified spontaneous emission (SASE) [82]. It leads to laser-like quasi-coherent quasi-monochromatic radiation without the need for a seed radiation that, unlike that of free electron lasers working on different wavelengths, would be difficult to obtain for XFELs due to the lack of suitable lasers.

The ultrafast high-intensity X-ray pulses from XFELs have proven to be an invaluable tool for the detection of ultrafast charge migration with high temporal and spatial resolution. One example of a simple nonlinear spectroscopic technique [83] made possible by this new technology is the impulsive stimulated X-ray Raman scattering (ISXRS) which has already been successfully applied for the investigation of ultrafast charge dynamics in atomic and small molecular systems [52]. This spectroscopic technique is based on the Raman effect, which occurs when the interaction of electromagnetic radiation with a chemical system results in a change in the system's state and the frequency of the radiation. When the radiation is an X-ray one as in the case of ISXRS, the involved excited states are the electronic ones. In particular, in ISXRS, we can observe the excitation of the ground state to a core excited state followed by the decay of a valence electron in the core hole, leading to a final valence excited state. These processes are initiated and driven by the external electromagnetic field, which, due to its brief duration, possesses a broad bandwidth encompassing both the excitation and de-excitation frequencies of interest. As the pulse duration is short compared to the time scale of the evolution of the system and as the core excitations targeted by the X-rays are sensitive to the local electronic environment, this technique is well suited to follow the charge migration throughout the molecule [84].

Due to these characteristics, ultrafast X-ray pulses can also be used as probe pulses in pump-probe transient absorption spectroscopy. This technique allows to obtain snapshots of the valence electrons' dynamics in molecules induced by a valence-exciting pulse. All of these experimental achievements rely on an interplay with theoretical development, that provides an irreplaceable tool for the interpretation of obtained data and for the prediction and design of new experimental schemes, which are often very demanding both from an economical and practical point of view.

Chapter 7

Summary of publications

7.1 Publication I - resume: e$^{\mathcal{T}}$ 1.0: An open source electronic structure program with emphasis on coupled-cluster and multilevel methods

Sarai D. Folkestad, Eirik F. Kjønstad, Rolf H. Myhre, Josefine H. Andersen, Alice Balbi, Sonia Coriani, Tommaso Giovannini, Linda Goletto, Tor S. Haugland, Anders Hutcheson, Ida-Marie Høyvik, Torsha Moitra, Alexander C. Paul, Marco Scavino, Andreas S. Skeidsvoll, Åsmund H. Tveten, and Henrik Koch
J. Chem. Phys. **152,** 184103 (2020)

In this paper, we present e$^{\mathcal{T}}$ 1.0, an electronic structure program that we developed. e$^{\mathcal{T}}$ 1.0 is a software written in object oriented modern Fortran (2008 standard) language and released as an open source program licensed under the GNU General Public Licence 3 (GPL 3.0). Its objective is to provide an electronic structure code that is not only efficient but also readily understandable, extensibility, and modification, empowering users to optimize existing code or introduce novel functionalities.

The primary emphasis is placed on the utilization of spin adapted implementations of coupled-cluster techniques such as CCS, CC2, CCSD, CC3, and CCSD(T), as well with the possibility to apply these methods in quantum mechanics/molecular mechanics (QM/MM) calculations or in conjunction with the polarizable continuum model (PCM).

For all CC methods, it is possible to compute the energy of ground and excited singlet state, dipole and quadruple moments, and equation-of-motion (EOM) oscillation strengths with an exception for CCSD(T) for which is possible to compute only the ground state. Coupled-cluster equations are solved by means of:

- Davidson method for linear and eigenvalue equations,

- DIIS for nonlinear coupled-cluster equations,

- Asymmetric Lanczos for nonperturbative coupled-cluster methods.

The Hartree-Fock method is also available in different fashions: restricted (both closed shell RHF and open shell ROHF), unrestricted (both standard UHF and constrained CUHF), time-dependent TDHF, quantum electrodynamics Hartee-Fock (QED-HF) and multilevel HF (MLHF) are present. Additionally, second-order Møller-Plesset (MP2) and full configuration interaction (FCI) are implemented.

e$^{\mathcal{T}}$ 1.0 relies on the following external dependencies:

- Libint 2 [85], a library for the evaluation of molecular integrals,

- PCMSolver 1.2 [86] library for solvation modeling,

- runtest library [87] for software testing,

- CMake module from autocmake [88] to find and configure BLAS and LAPACK libraries.

Wherever possible, the routines are parallelized by means of OpenMP [89].

The release of e$^{\mathcal{T}}$ 1.0 represented the first published implementation of explicitly time-dependent coupled-cluster (TD-CCS and TD-CCSD), multilevel HF, and multilevel CC2, as well as the most efficient implementation (at the time it was released) of Cholesky decomposition of electronic repulsion integrals [90] and coupled-cluster singles, doubles and perturbative triples (CC3) [91]. In real-time formulation of time-dependent CCS and CCSD, the time-dependence is contained into the cluster amplitudes and multipliers, whose evolution in time is described through differential equations that, in e$^{\mathcal{T}}$ 1.0, can be solved by the following integrators:

- Euler,

- Gauss-Legendre 2nd-, 4th- and 6th-order (GL2, GL4 and GL6),

- 4th-order explicit Runge-Kutta (RK4).

As time-dependent amplitudes and multipliers are in general complex [92], the differential equations that need to be solved in real-time propagation are implemented with complex variables. In e$^{\mathcal{T}}$ 1.0, real-time calculation can provide as output the time evolution of amplitudes, multipliers, density matrix, energy, electric field, and dipole moment. It is also possible to obtain spectra by means of the fast Fourier transform (FFT) of the time-dependent dipole moment, implemented as a modified version of FFTPACK 5.1 [93]. Furthermore, time-dependent density can be printed as output in .plt or .cube files that are readable by visualization software (such as Chimera) in order to obtain a movie of the evolution of electronic density in time. The external electromagnetic field can be specified in input as a sum of any number of classical electromagnetic pulses within the length gauge and the dipole approximation.

7.2 Publication II - resume: Time-dependent coupled-cluster theory for ultrafast transient absorption spectroscopy

Andreas S. Skeidsvoll, Alice Balbi, and Henrik Koch
Phys. Rev. A **102,** 023115 (2020)

In this paper, we use our implementation of spin-adapted TDCC model present in the $e^{\mathcal{T}}$ program [46] to simulate the response of small molecular systems to a pump-probe sequence of ultrashort laser pulses.

As a first step, we assess the accuracy of our integration procedures by comparison with results already present in literature for He and Be atoms [41]. The goal is to study the response of the lithium fluoride dimer (LiF) to a pump-probe sequence of ultrafast laser pulses, with a valence-exciting pump and a core-exciting probe. As higher level coupled-cluster methods computation cost scales rapidly with the size of the system and quickly reaches the limits of practicability, we have chosen a smaller system with an analogous structure, the lithium hydride dimer (LiH), for the convergence tests of the absorption spectra for parameters like the level of theory, the basis set, the size of the time steps, and the integration method. From the performed tests we notice how, in our calculations, there is no need for the use of computationally expensive symplectic integrators such as Gauss-Legendre (that we test at second, fourth, and sixth-order) as the deviations with respect to an integrator with a lower computational cost such as explicit fourth-order Runge-Kutta are negligible. We also observe the deviations related to the time step size, which in general doesn't appear to have a major impact on the results. The changes in the basis set, on the other hand, give rise to very relevant differences. We examine how the outcomes vary as we increase the size of the Dunning basis sets: cc-pVDZ, cc-p(C)VDZ, aug-cc-pVDZ, aug-cc-p(C)VDZ plus the corresponding triple zeta basis sets. We observe how the inclusion of core correlation functions affects substantially the position of probe absorption peaks while they leave unaffected the pump absorption peaks. Also, triple zeta functions have an impact on the position of the peaks in the probe absorption region, but their use increases substantially the computational cost of the calculation. We also compare the results obtained from TDCCS and TDCCSD calculation and note the very relevant difference in position and intensity of the peaks in both of the regions of the spectrum. Considering that the cost of the calculation increases quickly with the growth of the basis set and the level of theory, while it scales linearly with the time step, and keeping in mind that explicit Runge-Kutta methods require fewer evaluations than the Gauss-Legendre ones while providing comparable results, given the results obtained for the LiH dimer, we determine the TDCCSD/aug-cc-p(C)VDZ integrated with RK4 with 0.005 a.u. time steps to be the best compromise between accuracy and computational cost to use for the study of the LiF dimer.

In order to determine the frequency to use for the pump and probe pulses acting on the LiF dimer, we characterize the first eight valence excited states and the

first eight core excited states of this system at the EOM-CCSD level of theory combined with an aug-cc-p(C)VDZ basis set. For the calculations relative to the core excited states, the core-valence separation approximation is applied. We use the obtained values of energy to tune the frequency of the pump pulse to the value corresponding to the energy difference between the first valence excited state and the ground state, and the frequency of the probe pulse corresponding to the energy difference between the first core excited state and the ground state. Both pulses are polarized along the internuclear axis of the system. We repeat the calculation for various temporal distances between the central times of the pump and probe pulses. We obtain the reference absorption spectrum in which we observe two regions of absorption: one corresponding to the energy range of valence excitations and one to the core ones. We compute the normalized transient absorption as

$$
\begin{aligned}
\Delta S'(\omega, \tau) &= \mathcal{N}_{\text{probe}} \Delta S(\omega, \tau) \\
&= \mathcal{N}_{\text{probe}} \Big(S(\omega, \tau) - S_{\text{probe}}(\omega) \Big),
\end{aligned}
\tag{7.1}
$$

where S is the response function defined as

$$
S(\omega) = -2 \operatorname{Im} \Big(\widetilde{\mathbf{d}}(\omega) \cdot \widetilde{\boldsymbol{\mathcal{E}}}^{*}(\omega) \Big), \quad \omega > 0.
\tag{7.2}
$$

We note the presence of a peak in the core region with energy lower than the lowest ground state core excitation energy. This peak is the result of the creation of a valence-hole state created by the pump pulse that is in turn excited by the core-exciting probe. We observe how the peaks of the transient probe absorption spectrum of the lithium fluoride dimer oscillate with the time delay between the pump pulse and the probe one. Some of the oscillations are rapidly dumped for pump-probe delays lower than 40 a.u., which we attribute to the decreasing overlap between pump and probe pulses. For pump-probe delays higher than 40 a.u., which present a negligible pump-probe overlap, no dumping of the oscillations is observed. We select the five peaks for which the biggest amplitude of oscillation is observed and categorize the states involved in the transitions. Two of the transitions involve the $A_c{}^1\Sigma^+$ state, one starting from the ground state $X^1\Sigma^+$ and one from the valence excite state $B_v{}^1\Sigma^+$. We observe how the oscillation of the peaks relative to these two transitions can be fitted, through the least-square fitting procedure, by a sinusoidal function of the form $A\sin(\omega_A t + \phi_A) + C$ with ω_A corresponding to the energy difference between the ground state $X^1\Sigma^+$ and valence excited state $B_v{}^1\Sigma^+$. We interpreted this oscillation as due to the quantum interference between these two states [94]. Also, the oscillations of the other peaks that we took into consideration can be explained in terms of quantum interference between one of the states involved in the excitation with other states of the system.

7.3 Publication III - resume: Simulating weak-field attosecond processes with a Lanczos reduced basis approach to time-dependent equation-of-motion coupled-cluster theory

Andreas S. Skeidsvoll, Torsha Moitra, Alice Balbi, Alexander C. Paul, Sonia Coriani, and Henrik Koch
Phys. Rev. A **105,** 023103 (2022)

In this paper, we extend the implementation of TDCC discussed in [46] and [39] to a TD-EOM-CC approach with the use of a reduced basis calculated with an asymmetric band Lanczos algorithm. The goal is to provide a method that, thanks to its lower computational cost, enables the study of larger systems.

We use the asymmetric band Lanczos algorithm to generate approximate eigenvalues $\tilde{\omega}$ and the corresponding left $\widetilde{\mathbf{L}}_n$ and right $\widetilde{\mathbf{R}}_n$ approximate eigenvectors of the asymmetric Jacobian matrix $\mathbf{A}^{(0)}$. These eigenvectors represent the coefficients of respectively the left and right excitation operators of EOM-CC needed for obtaining the valence and core excited states to use in the differential equations that need to be solved to obtain a time-dependent description of the system.

When using the Lanczos algorithm, we first calculate the valence states, using starting vectors based on the ground state. Then, we calculate the core states, constructing starting vectors based on the previously obtained valence states together with the ground state. The choice of starting vectors is important as it can simplify the calculation of operator matrix elements, it can direct the algorithm towards states that are useful to represent the interactions and allow an affordable calculation of the transition strengths, that in turn is used, together with the excitation energies, to automatically select the reduced basis.

We assess the reliability of this method by a comparison of the TD-EOM-CC absorption spectrum of lithium fluoride dimer (LiF) calculated with the Dormand-Prince 5(4) [95] integration scheme at various Lanczos chain lengths (i.e. number of iterations) with the results already obtained for lithium fluoride in [39]. The core states are calculated within the CVS approximation while the valence states are calculated in full projection space.

By increasing the band Lanczos chain length and taking a sufficient number of relevant states into account, the TD-EOM-CC with Lanczos approach converges to the same spectral features as TDCC but with much lower computational times. In these calculations, only the states accessible by absorption of one photon from the pump pulse and one photon from the probe pulse were considered. This means that spectral features due to the absorption of two photons from the pump pulse like the smaller peaks below 6.9 eV, as claimed by us in [39] and confirmed later on in [43], cannot be reproduced. We show that by including in the calculation as well the valence states accessible by the absorption of two pump protons and the core states accessible by the interaction with a further probe photon, with sufficiently high chain length a spectrum indistinguishable from the TDCC one is obtained, confirming that the process is in fact due to a two pump photon absorption and

shows how the TD-EOM-CC with reduced basis can provide accurate results even in case of nonlinear interactions.

We show how a speedup in the convergence can be obtained by the use of the valence-only CVS projector to calculate the valence states. Thanks to the reduction in dimension, a smaller Lanczos chain length is sufficient to retrieve all of the spectral features as for the case of valence states calculated in the full projections space.

We observe the limits of CVS approximation, showing how for the LiH molecule several peaks around the Li K-edge are not correctly retrieved. This is due to the fact that for light atoms such as Li the energy separation between core and valence orbitals is small so that pure valence excitation (which would be excluded by the core-only CVS projector) can fall within the core excitation region. In the LiF calculations, this problem is not encountered as the pulse is tuned with the F K edge so the core and valence orbitals have a better separation.

When calculating the TD-EOM-CC spectrum of ethylene with a z dipole operator, we notice how a peak present in the TDCC spectrum is missing in the TD-EOM-CC one. We attribute this peak to a two photon absorption process and we manage to obtain its presence in a TD-EOM-CC spectrum using a procedure that emulates the physical process in play: we include a starting vector constructed from the z^2 quadrupole operator in the valence state-calculation. We capture a two photon absorption phenomenon for ethylene using a procedure that emulates the physical process in play.

Finally, we use the computational procedure to model attosecond transient absorption by the glycine molecule, which is a good candidate for experimental investigations. We get 17 converged valence excited states and 20 converged core excited states. We note that all of the converged valence states have an energy that is lower than double the carrier frequency, meaning that spectral features due to two-photon absorption may be not properly represented in the obtained spectra. We calculated different absorption spectra varying the value of the pump-probe time delay and then subtracted from them the ground state probe absorption spectra to obtain the difference spectra. In them, we note the presence of negative peaks, which indicates ground state bleaching. We also note that the peaks that are more sensitive to the variation of pump probe delay are those that are not energetically accessible from the ground state. We numerically integrated the probe absorption difference $\int \Delta S_C(\omega) d\omega$ as a function of pump probe delay and the dipole induced by the pump pulse in the direction from the center of mass to the N atom $\langle d \rangle_N$ as a function of time. We note how the dominant period of the integrated absorption and that of the generated dipole are very similar. This seems to indicate a correlation between the transient absorption and the charge migration induced by the pump, indicating that both quantities provide insight into the dominant time scale of the coherent superposition, as has been previously demonstrated for instantaneous pulses [96, 97].

7.4 Publication IV - resume: Coupled-cluster simulation of impulsive X-ray Raman scattering

Alice Balbi, Andreas S. Skeidsvoll, and Henrik Koch
Submitted | arXiv: 2305.19942 [physics.chem-ph]

Raman scattering is a light-matter interaction phenomenon that consists in the absorption of a photon by a molecular system, causing and excitation of the system, and subsequent emission of a photon with frequency different from that of the first one, causing a de-excitation of the system. When the involved states are the electronic ones, we speak about *electronic* Raman scattering [5, 48–51]. We used the TD-EOM-CC method to simulate the electronic Raman scattering induced by ultrashort laser pulses in small molecules. In particular, we focused on ISXRS, which is the extension to the impulsive limit of stimulated X-ray Raman scattering (SXRS). We directed our attention towards the scenario where the excitation involved is a core excitation that is transformed into a valence excited state when a valence electron decays into a core vacancy. Core excitations can be described as confined to a particular atomic site and are highly responsive to the nearby electronic surroundings, which makes them valuable for the local triggering of charge migration. For the study of core excitations, we use the core-valence separation scheme, that is detailed in section 5.3. Additionally, in our calculations, we use pulses with pulses large enough to include both the energy differences in play, so that a single pulse can be responsible for both the core-excitation and the stimulated emission to a valence excited state. [98] A similar procedure was recently performed experimentally for the first time at the Linac Coherent Light Source [52].

In order to follow the Raman process in our simulations, we calculated the population of an eigenstate $|\psi_i\rangle$ of the field-free Hamiltonian as

$$
\begin{aligned}
P_i(t) &= \langle \widetilde{\Psi}(t)|\psi_i\rangle \, \langle \widetilde{\psi_i}|\Psi(t)\rangle \\
&= b_i(t)c_i(t).
\end{aligned}
\tag{7.3}
$$

To avoid the occurrence non-stationary population that would make harder the interpretation of the TD-EOM-CC state, we diagonalize the Jacobian in the basis of all of the CVS and non-CVS (valence) states.

At first we performed some benchmark calculations choosing Ne as the system to investigate. This choice of the system is convenient as it is small, so the calculation maintain an acceptable computational cost even when the employed basis set is large, and has a spherical symmetry, so the results are independent from the direction chosen for the external electric field. We investigated the convergence of the final population of the $B_v\,^1D$ valence-excited state, the lowest-eigenvalue valence-excited state with a significant final population, for different choices of Dunning basis set and different levels of theory (CCS and CCSD). We observed how the choice of level of theory had a profound impact on the results and how the basis set convergence was reached with the aug-cc-pV5Z basis set. We also performed a series of calculations with different number of considered core excited

states with a fixed number of valence excited states and vice versa, to determine the number of states to include in order to reach convergence.

We then moved on to study ISXRS in carbon monoxide, a molecular system belonging to the $C_{\infty v}$ symmetry point group. As it is linear, it adds a further layer of complexity: now we expect the response of the system to depend on the direction of the electric field. From our calculations we can see how having the field perpendicular to the internuclear axis instead of along it, for symmetry reasons, switches of completely off the pathways that would populate the valence state under examination. For this system we also observe how the inclusion of about 30 core excited states is necessary to reach the convergence and how obviously the final populations are affected depending on whether the pump pulse is tuned with the C K-edge or with the O K-edge.

We further increased the complexity by moving on to study pyrrole, which has a geometry belonging to the C_{2v} symmetry point group. We can notice how when the direction of the pulse is perpendicular to the plane of the molecule only few states become populated as other are inaccessible due to symmetry reasons. When changing the direction of electric field to include also other symmetries, more states becomes accessible. We also show how the obtained final populations varies when changing the tuning of the pump pulse from N K-edge to C K-edge.

Finally, we studied a bigger system: the p-aminophenol, belonging to the C_s symmetry point group. We have chosen this system as we expect to observe the core excitation to be followed by a long range charge migration from one functional group to the other located at opposite side of the aromatic ring. This is indeed what we obtained from our simulations, which we made visually accessible by displaying two chosen isodensity surfaces, one positive and one negative.

These results will be submitted for publication in the near future. We believe that the toolkit that we developed will be useful for further research about the control of chemical reactivity by means of external electromagnetic field.

Chapter 8

Conclusion and outlook

All chemistry, at a fundamental level, is a consequence of electronic motion. Until recently, it wasn't possible to observe and control electron dynamics directly on its own time scale, as the best available temporal resolution was that of nuclear motion, that occurs on a femtosecond time scale. This changed with the development of ultrafast laser technology and spectroscopy, which led to the emergence of a new research field called attochemistry: the ability to steer the electrons in a molecular system opened up the possibility to indirectly control the force field experienced by the nuclei and thus the chemical reactivity itself. Such control over matter would have a tremendous impact on basically every research field linked with molecular sciences including biology, material science, and medicine. In particular, the most exciting perspective is to utilize these achievements to directly control chemical reactivity inside molecules and to understand deeply phenomena like photosynthesis or DNA electron damage, which was not possible to understand completely with the femtosecond temporal resolution. These groundbreaking technological achievements have stimulated the theoretical development of methods to describe these phenomena as accurate modeling is crucial for both planning experiments and interpreting the obtained results. Real-time simulations are now possible by evolving the time-dependent Schrödinger equation, which was previously impractical due to the limited computing power available for calculations involving electronic correlation effects. The advancements in computational power have enabled the utilization of correlated methods such as DFT, MC-SCF, CI, ADC, and CC in attochemistry, allowing for new insights and discoveries to be made in this field.

This thesis research focuses on implementing methods for real-time simulations of the behavior of molecular systems. Specifically, the study examines through time-dependent coupled-cluster theory how small- and medium-sized molecular systems react to ultrashort laser pulses modeled as external semiclassical electromagnetic fields in dipole approximation. To begin this research, we implemented in the e$^{\mathcal{T}}$ program package [46] the time-dependent coupled-cluster at the CCS and CCSD level of theory, as well as Gauss-Legendre and explicit Runge-Kutta integrators necessary for solving the differential equations for time-evolution of cluster amplitudes and multipliers. To validate this implementation, we compared the results for some small systems to those already established in the existing literature. To establish convergence, we performed calculations on the computationally inexpensive LiH system with different levels of theory, basis sets, choices of integrator, and integration steps. Using these results, we investigated pump-probe transient absorption [39], with a valence-exciting pump and core-exciting probe in LiF, a system with an analogous structure but a higher computational cost.

This calculation was performed at different pump-probe delays, enabling us to observe how the spectral peaks oscillated with the variation of this quantity. The oscillation frequency was found to be correlated with quantum interference among the involved states.

We further expanded our toolkit by including into the $e^{\mathcal{T}}$ program package an implementation of TD-EOM-CC and asymmetric band Lanczos, with the purpose of using them in combination [47]. First, we assessed the validity of this methodology by comparing its results for LiF with that already obtained through TDCC. We then observed how, when taking into account two-photon absorption, even the smaller features of the TDCC spectrum were reproduced correctly. We then proceeded to study LiH: as for light elements the energy difference between valence and core excited states is small, some of the peaks around the Li K-edge were not retrieved while using the core-only CVS as they were due to valence-only excitations.

We demonstrated the significance of selecting physically meaningful start vectors, highlighting the convergence of ethylene valence excited states, that are dark to one-photon transitions, when constructing starting vectors from both dipole and quadrupole operators. As a final application of this protocol, we modeled the transient absorption spectrum of the glycine molecule as a function of pump-probe delay. Our results show a correlation between the transient absorption and the charge migration induced by the pump pulse.

Finally, we employed TD-EOM-CC to model ISXRS. We started with an evaluation of convergence of the final population of neon valence states with respect to various calculation parameters such as the level of coupled-cluster theory, the basis set, and the total number of valence excited states and core excited states. We determined a set of satisfactory parameters and extended the analysis to two other systems, carbon monoxide, and pyrrole, evaluating the convergence of the final population of their states with respect to the number of included core-excited states. We also showed that the final populations of both carbon monoxide and pyrrole states are influenced by the polarization of the external electric field. Additionally, the impact of tuning the external electric field to the K-edge of different atoms was assessed. Lastly, we studied the time evolution of the electronic density of p-aminophenol during ISXRS and observed a rapid construction of a valence wavepacket, followed by subsequent charge migration in the molecule.

One possible improvement of our models is to facilitate calculations on bigger systems by the use of larger and adaptive time steps. The description can also be upgraded by including the representation of the electronic continuum in the basis set, e.g. through B-splines as demonstrated in [99]. This would enable us to model ionization and investigate additional phenomena that are relevant to attochemistry, such as the Auger effect. Another natural next step would be to include the nuclear motion in the model, which for instance can be done using the approach in [100].

In addition to the above-mentioned improvements, a crucial step would be to create experimental data for comparison with our theoretical results. This would help to validate our models and improve the understanding of the behavior of molecular systems subjected to ultrashort laser pulses. Moreover, we are convinced that the investigation of field-induced charge migration in molecular systems has tremendous potential for advancing our understanding and control of chemical

reactions. Consequently, future research endeavors could be directed toward applying our methods to increasingly larger and more complex molecular systems, exploring novel strategies for manipulating chemical reactions through external fields, and designing materials with tailored properties. These efforts could open up new frontiers in attochemistry, and pave the way for practical applications of our research findings.

Bibliography

[1] Ahmed H. Zewail. Femtochemistry: atomic-scale dynamics of the chemical bond. *The Journal of Physical Chemistry A*, 104(24):5660–5694, 2000.

[2] Ahmed H. Zewail. Femtochemistry. past, present, and future. *Pure and Applied Chemistry*, 72(12):2219–2231, 2000.

[3] C. Pellegrini, A. Marinelli, and S. Reiche. The physics of x-ray free-electron lasers. *Rev. Mod. Phys.*, 88:015006, Mar 2016.

[4] Claudio Pellegrini. The development of xfels. *Nat. Rev. Phys.*, 2:330–331, July 2020.

[5] Stephen R. Leone, C. William McCurdy, Joachim Burgdörfer, Lorenz S. Cederbaum, Zenghu Chang, Nirit Dudovich, Johannes Feist, Chris H. Greene, Misha Ivanov, Reinhard Kienberger, Ursula Keller, Matthias F. Kling, Zhi-Heng Loh, Thomas Pfeifer, Adrian N. Pfeiffer, Robin Santra, Kenneth Schafer, Albert Stolow, Uwe Thumm, and Marc J. J. Vrakking. What will it take to observe processes in 'real time'? *Nat. Photonics*, 8:162–166, Mar 2014.

[6] A. Baltuška, Th Udem, M. Ulberacker, M. Hentschel, E. Goulielmakis, Ch. Gohle, R. Holzwarth, V. S. Yakovlev, A. Scrinzi, T. W. Hänsch, and F. Krausz. Attosecond control of electronic processes by intense light fields. *Nature*, 421:611, Feb 2003.

[7] Joseph Duris, Siqi Li, Taran Driver, Elio G. Champenois, James P. MacArthur, Alberto A. Lutman, Zhen Zhang, Philipp Rosenberger, Jeff W. Aldrich, Ryan Coffee, Giacomo Coslovich, Franz-Josef Decker, James M. Glownia, Gregor Hartmann, Wolfram Helml, Andrei Kamalov, Jonas Knurr, Jacek Krzywinski, Ming-Fu Lin, Jon P. Marangos, Megan Nantel, Adi Natan, Jordan T. O'Neal, Niranjan Shivaram, Peter Walter, Anna Li Wang, James J. Welch, Joseph Z. Wolf, Thomas J. A. Xu, Matthias F. Kling, Philip H. Bucksbaum, Alexander Zholents, Zhirong Huang, James P. Cryan, and Agostino Marinelli. Tunable isolated attosecond x-ray pulses with gigawatt peak power from a free-electron laser. *Nat. Photonics*, 14:30–36, 2020.

[8] Markus Drescher, Michael Hentschel, Reinhard Kienberger, Gabriel Tempea, Christian Spielmann, Georg A. Reider, Paul B. Corkum, and Ferenc Krausz. X-ray pulses approaching the attosecond frontier. *Science*, 291(5510):1923–1927, 2001.

[9] Siqi Li, Taran Driver, Philipp Rosenberger, Elio G. Champenois, Joseph Duris, Andre Al-Haddad, Vitali Averbukh, Jonathan C. T. Barnard, Nora Berrah, Christoph Bostedt, Philip H. Bucksbaum, Ryan N. Coffee, Louis F. DiMauro, Li Fang, Douglas Garratt, Averell Gatton, Zhaoheng Guo, Gregor

Hartmann, Daniel Haxton, Wolfram Helml, Zhirong Huang, Aaron C. LaForge, Andrei Kamalov, Jonas Knurr, Ming-Fu Lin, Alberto A. Lutman, James P. MacArthur, Jon P. Marangos, Megan Nantel, Adi Natan, Razib Obaid, Jordan T. O'Neal, Niranjan H. Shivaram, Aviad Schori, Peter Walter, Anna Li Wang, Thomas J. A. Wolf, Zhen Zhang, Matthias F. Kling, Agostino Marinelli, and James P. Cryan. Attosecond coherent electron motion in auger-meitner decay. *Science*, 375(6578):285–290, 2022.

[10] F. Calegari, D. Ayuso, A. Trabattoni, L. Belshaw, S. De Camillis, S. Anumula, F. Frassetto, L. Poletto, A. Palacios, P. Decleva, J. B. Greenwood, F. Martín, and M. Nisoli. Ultrafast electron dynamics in phenylalanine initiated by attosecond pulses. *Science*, 346(6207):336–339, 2014.

[11] P. M. Kraus, B. Mignolet, D. Baykusheva, A. Rupenyan, L. Horný, E. F. Penka, G. Grassi, O. I. Tolstikhin, J. Schneider, F. Jensen, L. B. Madsen, A. D. Bandrauk, F. Remacle, and H. J. Wörner. Measurement and laser control of attosecond charge migration in ionized iodoacetylene. *Science*, 350(6262):790–795, 2015.

[12] Adriana Marais, Betony Adams, Andrew K Ringsmuth, Marco Ferretti, J Michael Gruber, Ruud Hendrikx, Maria Schuld, Samuel L Smith, Ilya Sinayskiy, Tjaart PJ Krüger, et al. The future of quantum biology. *Journal of the Royal Society Interface*, 15(148):20180640, 2018.

[13] Thomas Renger, Volkhard May, and Oliver Kühn. Ultrafast excitation energy transfer dynamics in photosynthetic pigment–protein complexes. *Physics Reports*, 343(3):137–254, 2001.

[14] Jenny Nguyen, Yuhan Ma, Ting Luo, Robert G Bristow, David A Jaffray, and Qing-Bin Lu. Direct observation of ultrafast-electron-transfer reactions unravels high effectiveness of reductive dna damage. *Proceedings of the National Academy of Sciences*, 108(29):11778–11783, 2011.

[15] Jeppe Olsen and Poul Jo/rgensen. Linear and nonlinear response functions for an exact state and for an mcscf state. *The Journal of chemical physics*, 82(7):3235–3264, 1985.

[16] Trygve Helgaker, Sonia Coriani, Poul Jørgensen, Kasper Kristensen, Jeppe Olsen, and Kenneth Ruud. Recent advances in wave function-based methods of molecular-property calculations. *Chemical reviews*, 112(1):543–631, 2012.

[17] SE Koonin, KTR Davies, V Maruhn-Rezwani, H Feldmeier, SJ Krieger, and JW Negele. Time-dependent hartree-fock calculations for o 16+ o 16 and ca 40+ ca 40 reactions. *Physical Review C*, 15(4):1359, 1977.

[18] H Flocard, SE Koonin, and MS Weiss. Three-dimensional time-dependent hartree-fock calculations: Application to o 16+ o 16 collisions. *Physical Review C*, 17(5):1682, 1978.

[19] KTR Davies, KR Sandhya Devi, and MR Strayer. Time-dependent hartree-fock fusion calculations for kr 86+ la 139 and kr 84+ bi 209 collisions. *Physical Review Letters*, 44(1):23, 1980.

[20] KR Sandhya Devi and SE Koonin. Mean-field approximation to p+ he scattering. *Physical Review Letters*, 47(1):27, 1981.

[21] Kenneth C Kulander, KR Sandhya Devi, and SE Koonin. Time-dependent hartree-fock theory of charge exchange: Application to he 2++ he. *Physical Review A*, 25(6):2968, 1982.

[22] Xiaosong Li, Niranjan Govind, Christine Isborn, A Eugene DePrince III, and Kenneth Lopata. Real-time time-dependent electronic structure theory. *Chemical Reviews*, 120(18):9951–9993, 2020.

[23] Makenzie R. Provorse and Christine M. Isborn. Electron dynamics with real-time time-dependent density functional theory. *Int. J. Quantum Chem.*, 116(10):739–749, 2016.

[24] Chao Lian, Mengxue Guan, Shiqi Hu, Jin Zhang, and Sheng Meng. Photoexcitation in solids: First-principles quantum simulations by real-time tddft. *Adv. Theory Simul.*, 1(8):1800055, 2018.

[25] Haruhide Miyagi and Lars Bojer Madsen. Time-dependent restricted-active-space self-consistent-field theory for laser-driven many-electron dynamics. *Phys. Rev. A*, 87:062511, Jun 2013.

[26] Takeshi Sato and Kenichi L. Ishikawa. Time-dependent complete-active-space self-consistent-field method for multielectron dynamics in intense laser fields. *Phys. Rev. A*, 88:023402, Aug 2013.

[27] Haruhide Miyagi and Lars Bojer Madsen. Time-dependent restricted-active-space self-consistent-field theory for laser-driven many-electron dynamics. ii. extended formulation and numerical analysis. *Phys. Rev. A*, 89:063416, Jun 2014.

[28] Pascal Krause, Tillmann Klamroth, and Peter Saalfrank. Time-dependent configuration-interaction calculations of laser-pulse-driven many-electron dynamics: Controlled dipole switching in lithium cyanide. *J. Chem. Phys.*, 123(7):074105, 2005.

[29] Jason A. Sonk, Marco Caricato, and H. Bernhard Schlegel. Td-ci simulation of the electronic optical response of molecules in intense fields: Comparison of rpa, cis, cis(d), and eom-ccsd. *J. Phys. Chem. A*, 115(18):4678–4690, 2011.

[30] Eleonora Luppi and Martin Head-Gordon. Computation of high-harmonic generation spectra of h2 and n2 in intense laser pulses using quantum chemistry methods and time-dependent density functional theory. *Molecular Physics*, 110(9-10):909–923, 2012.

[31] A. Eugene DePrince, Matthew Pelton, Jeffrey R. Guest, and Stephen K. Gray. Emergence of excited-state plasmon modes in linear hydrogen chains from time-dependent quantum mechanical methods. *Phys. Rev. Lett.*, 107:196806, Nov 2011.

[32] L.S. Cederbaum and J. Zobeley. Ultrafast charge migration by electron correlation. *Chem. Phys. Lett.*, 307(3):205–210, 1999.

[33] M. Ruberti. Quantum electronic coherences by attosecond transient absorption spectroscopy: ab initio b-spline rcs-adc study. *Faraday Discuss.*, 228:286–311, 2021.

[34] Daniel R. Nascimento and A. Eugene DePrince. Simulation of near-edge x-ray absorption fine structure with time-dependent equation-of-motion coupled-cluster theory. *J. Phys. Chem. Lett.*, 8(13):2951–2957, 2017.

[35] Lauren N. Koulias, David B. Williams-Young, Daniel R. Nascimento, A. Eugene DePrince, and Xiaosong Li. Relativistic real-time time-dependent equation-of-motion coupled-cluster. *J. Chem. Theory Comput.*, 15(12):6617–6624, 2019.

[36] Simen Kvaal. Ab initio quantum dynamics using coupled-cluster. *J. Chem. Phys.*, 136(19):194109, 2012.

[37] Christian Huber and Tillmann Klamroth. Explicitly time-dependent coupled cluster singles doubles calculations of laser-driven many-electron dynamics. *J. Chem. Phys.*, 134(5):054113, 2011.

[38] D. A. Pigg, G. Hagen, H. Nam, and T. Papenbrock. Time-dependent coupled-cluster method for atomic nuclei. *Phys. Rev. C*, 86:014308, Jul 2012.

[39] Andreas S. Skeidsvoll, Alice Balbi, and Henrik Koch. Time-dependent coupled-cluster theory for ultrafast transient-absorption spectroscopy. *Phys. Rev. A*, 102:023115, Aug 2020.

[40] Young Choon Park, Ajith Perera, and Rodney J. Bartlett. Equation of motion coupled-cluster for core excitation spectra: Two complementary approaches. *J. Chem. Phys.*, 151(16):164117, 2019.

[41] Thomas Bondo Pedersen and Simen Kvaal. Symplectic integration and physical interpretation of time-dependent coupled-cluster theory. *J. Chem. Phys.*, 150(14):144106, 2019.

[42] Håkon Emil Kristiansen, Øyvind Sigmundson Schøyen, Simen Kvaal, and Thomas Bondo Pedersen. Numerical stability of time-dependent coupled-cluster methods for many-electron dynamics in intense laser pulses. *J. Chem. Phys.*, 152(7):071102, 2020.

[43] Thomas Bondo Pedersen, Håkon Emil Kristiansen, Tilmann Bodenstein, Simen Kvaal, and Øyvind Sigmundson Schøyen. Interpretation of coupled-cluster many-electron dynamics in terms of stationary states. *Journal of Chemical Theory and Computation*, 17(1):388–404, 2021. PMID: 33337895.

[44] Young Choon Park, Ajith Perera, and Rodney J. Bartlett. Equation of motion coupled-cluster study of core excitation spectra ii: Beyond the dipole approximation. *J. Chem. Phys.*, 155(9):094103, 2021.

[45] Henrik Koch and Poul Jørgensen. Coupled cluster response functions. *J. Chem. Phys.*, 93(5):3333–3344, 1990.

[46] Sarai D. Folkestad, Eirik F. Kjønstad, Rolf H. Myhre, Josefine H. Andersen, Alice Balbi, Sonia Coriani, Tommaso Giovannini, Linda Goletto, Tor S. Haugland, Anders Hutcheson, Ida-Marie Høyvik, Torsha Moitra, Alexander C. Paul, Marco Scavino, Andreas S. Skeidsvoll, Åsmund H. Tveten, and Henrik Koch. et 1.0: An open source electronic structure program with emphasis on coupled cluster and multilevel methods. *The Journal of Chemical Physics*, 152(18):184103, 2020.

[47] Andreas S. Skeidsvoll, Torsha Moitra, Alice Balbi, Alexander C. Paul, Sonia Coriani, and Henrik Koch. Simulating weak-field attosecond processes with a lanczos reduced basis approach to time-dependent equation-of-motion coupled-cluster theory. *Phys. Rev. A*, 105:023103, Feb 2022.

[48] Jason D. Biggs, Yu Zhang, Daniel Healion, and Shaul Mukamel. Watching energy transfer in metalloporphyrin heterodimers using stimulated x-ray raman spectroscopy. *Proc. Natl. Acad. Sci.*, 110(39):15597–15601, 2013.

[49] Satoshi Tanaka and Shaul Mukamel. Probing exciton dynamics using raman resonances in femtosecond x-ray four-wave mixing. *Phys. Rev. A*, 67:033818, Mar 2003.

[50] Nina Rohringer. X-ray raman scattering: a building block for nonlinear spectroscopy. *Philos. Trans. Royal Soc. A: Math. Phys. Eng. Sci.*, 377(2145):20170471, 2019.

[51] Jason D. Biggs, Yu Zhang, Daniel Healion, and Shaul Mukamel. Two-dimensional stimulated resonance raman spectroscopy of molecules with broadband x-ray pulses. *J. Chem. Phys.*, 136(17):174117, 2012.

[52] Jordan T. O'Neal, Elio G. Champenois, Solène Oberli, Razib Obaid, Andre Al-Haddad, Jonathan Barnard, Nora Berrah, Ryan Coffee, Joseph Duris, Gediminas Galinis, Douglas Garratt, James M. Glownia, Daniel Haxton, Phay Ho, Siqi Li, Xiang Li, James MacArthur, Jon P Marangos, Adi Natan, Niranjan Shivaram, Daniel S. Slaughter, Peter Walter, Scott Wandel, Linda Young, Christoph Bostedt, Philip H. Bucksbaum, Antonio Picón, Agostino Marinelli, and James P. Cryan. Electronic population transfer via impulsive stimulated x-ray raman scattering with attosecond soft-x-ray pulses. *Phys. Rev. Lett.*, 125:073203, Aug 2020.

[53] Alice Balbi, Andreas S. Skeidsvoll, and Henrik Koch. Coupled cluster simulation of impulsive stimulated x-ray raman scattering, 2023.

[54] Rodney J Bartlett and George D Purvis III. Molecular applications of coupled cluster and many-body perturbation methods. *Physica Scripta*, 21(3-4):255, 1980.

[55] Rodney J Bartlett and George D Purvis. Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem. *International Journal of Quantum Chemistry*, 14(5):561–581, 1978.

[56] Marcel Nooijen, K. R. Shamasundar, and Debashis Mukherjee. Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory. *Molecular Physics*, 103(15-16):2277–2298, 2005.

[57] Jeffrey Goldstone. Derivation of the brueckner many-body theory. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 239(1217):267–279, 1957.

[58] Rodney J Bartlett. Many-body perturbation theory and coupled cluster theory for electron correlation in molecules. *Annual review of physical chemistry*, 32(1):359–401, 1981.

[59] T Daniel Crawford and Henry F Schaefer III. An introduction to coupled cluster theory for computational chemists. *Reviews in computational chemistry*, 14:33–136, 2007.

[60] Henrik Koch, Hans Jo Jensen, Trygve Helgaker, Gustavo E Scuseria, Henry F Schaefer, et al. Coupled cluster energy derivatives. analytic hessian for the closed-shell coupled cluster singles and doubles wave function: Theory and applications. *The Journal of chemical physics*, 92(8):4924–4940, 1990.

[61] Trygve Helgaker and Poul Jørgensen. Calculation of geometrical derivatives in molecular electronic structure theory. *Methods in Computational Molecular Physics*, pages 353–421, 1992.

[62] D Mukherjee and PK Mukherjee. A response-function approach to the direct calculation of the transition-energy in a multiple-cluster expansion formalism. *Chemical Physics*, 39(3):325–335, 1979.

[63] K Emrich. An extension of the coupled cluster formalism to excited states (i). *Nuclear Physics A*, 351(3):379–396, 1981.

[64] John F Stanton and Rodney J Bartlett. The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *The Journal of chemical physics*, 98(9):7029–7039, 1993.

[65] Daniel R Nascimento and A Eugene DePrince III. Linear absorption spectra from explicitly time-dependent equation-of-motion coupled-cluster theory. *Journal of chemical theory and computation*, 12(12):5834–5840, 2016.

[66] H Bernhard Schlegel, Stanley M Smith, and Xiaosong Li. Electronic optical response of molecules in intense fields: Comparison of td-hf, td-cis, and td-cis (d) approaches. *The Journal of chemical physics*, 126(24):244110, 2007.

[67] Daniel R Nascimento and A Eugene DePrince III. A general time-domain formulation of equation-of-motion coupled-cluster theory for linear spectroscopy. *The Journal of Chemical Physics*, 151(20):204107, 2019.

[68] Henrik Koch, Rika Kobayashi, Alfredo Sanchez de Merás, and Poul Jo/rgensen. Calculation of size-intensive transition moments from the coupled cluster singles and doubles linear response function. *The Journal of chemical physics*, 100(6):4393–4400, 1994.

[69] Rasmus Faber and Sonia Coriani. Resonant inelastic x-ray scattering and nonesonant x-ray emission spectra from coupled-cluster (damped) response theory. *Journal of Chemical Theory and Computation*, 15(1):520–528, 2018.

[70] Mengxi Wu, Shaohao Chen, Seth Camp, Kenneth J Schafer, and Mette B Gaarde. Theory of strong-field attosecond transient absorption. *J. Phys. B At. Mol. Opt. Phys.*, 49(6):62003, feb 2016.

[71] Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk van der Vorst. *Templates for the solution of algebraic eigenvalue problems: a practical guide*. SIAM, 2000.

[72] Thomas Gaumnitz, Arohi Jain, Yoann Pertot, Martin Huppert, Inga Jordan, Fernando Ardana-Lamas, and Hans Jakob Wörner. Streaking of 43-attosecond soft-x-ray pulses generated by a passively cep-stable mid-infrared driver. *Opt. Express*, 25(22):27506–27518, Oct 2017.

[73] Mauro Nisoli, Piero Decleva, Francesca Calegari, Alicia Palacios, and Fernando Martín. Attosecond electron dynamics in molecules. *Chemical reviews*, 117(16):10760–10825, 2017.

[74] Alicia Palacios and Fernando Martín. The quantum chemistry of attosecond molecular science. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(1):e1430, 2020.

[75] Stefan Weber. Light-driven enzymatic catalysis of dna repair: a review of recent biophysical studies on photolyase. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1707(1):1–23, 2005.

[76] Jenny Nguyen, Yuhan Ma, Ting Luo, Robert G Bristow, David A Jaffray, and Qing-Bin Lu. Direct observation of ultrafast-electron-transfer reactions unravels high effectiveness of reductive dna damage. *Proceedings of the National Academy of Sciences*, 108(29):11778–11783, 2011.

[77] Chuang Tan, Zheyun Liu, Jiang Li, Xunmin Guo, Lijuan Wang, Aziz Sancar, and Dongping Zhong. The molecular origin of high dna-repair efficiency by photolyase. *Nature communications*, 6(1):7302, 2015.

[78] Harry B Gray and Jay R Winkler. Electron tunneling through proteins. *Quarterly reviews of biophysics*, 36(3):341–372, 2003.

[79] Settimio Mobilio, Federico Boscherini, and Carlo Meneghini. *Synchrotron Radiation*. Springer, 2016.

[80] Uwe Bergman, Vittal K Yachandra, and Junko Yano. *X-ray free electron lasers: applications in materials, chemistry and biology*, volume 18. Royal Society of Chemistry, 2017.

[81] Nanshun Huang, Haixiao Deng, Bo Liu, Dong Wang, and Zhentang Zhao. Features and futures of x-ray free-electron lasers. *The Innovation*, 2(2):100097, 2021.

[82] R Bonifacio, C Pellegrini, and LM Narducci. Collective instabilities and high-gain regime free electron laser. In *AIP conference proceedings*, volume 118, pages 236–259. American Institute of Physics, 1984.

[83] Peter Hamm. Principles of nonlinear optical spectroscopy: A practical approach or: Mukamel for dummies. *University of Zurich*, 41(5):77, 2005.

[84] Kai Siegbahn. Electron spectroscopy for atoms, molecules, and condensed matter. *Rev. Mod. Phys.*, 54:709–728, Jul 1982.

[85] E. F. Valeev. Libint: A library for the evaluation of molecular integrals of many-body operators over gaussian functions. http://libint.valeyev.net/, 2017. version 2.7.0.

[86] Roberto Di Remigio, Arnfinn Hykkerud Steindal, Krzysztof Mozgawa, Ville Weijo, Hui Cao, and Luca Frediani. Pcmsolver: An open-source library for solvation modeling. *International Journal of Quantum Chemistry*, 119(1):e25685, 2019.

[87] Radovan Bast. Runtest, September 2018.

[88] Radovan Bast, Roberto Di Remigio, and Jonas Juselius. Autocmake, February 2020.

[89] Rohit Chandra, Leo Dagum, David Kohr, Ramesh Menon, Dror Maydan, and Jeff McDonald. *Parallel programming in OpenMP*. Morgan kaufmann, 2001.

[90] Sarai D. Folkestad, Eirik F. Kjønstad, and Henrik Koch. An efficient algorithm for cholesky decomposition of electron repulsion integrals. *The Journal of Chemical Physics*, 150(19):194112, 2019.

[91] Rolf H Myhre and Henrik Koch. The multilevel cc3 coupled cluster model. *The Journal of chemical physics*, 145(4):044111, 2016.

[92] Thomas Bondo Pedersen and Henrik Koch. Coupled cluster response functions revisited. *The Journal of chemical physics*, 106(19):8059–8072, 1997.

[93] Paul N Swarztrauber. Fft algorithms for vector computers. *Parallel Computing*, 1(1):45–63, 1984.

[94] Eleftherios Goulielmakis, Zhi-Heng Loh, Adrian Wirth, Robin Santra, Nina Rohringer, Vladislav S Yakovlev, Sergey Zherebtsov, Thomas Pfeifer, Abdallah M Azzeer, Matthias F Kling, Stephen R Leone, and Ferenc Krausz. Real-time observation of valence electron motion. *Nature*, 466(7307):739–743, 2010.

[95] J.R. Dormand and P.J. Prince. A family of embedded runge-kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, 1980.

[96] Anthony D. Dutoi, Kirill Gokhberg, and Lorenz S. Cederbaum. Time-resolved pump-probe spectroscopy to follow valence electronic motion in molecules: Theory. *Phys. Rev. A*, 88:013419, Jul 2013.

[97] Anthony D. Dutoi and Lorenz S. Cederbaum. Time-resolved pump-probe spectroscopy to follow valence electronic motion in molecules: Application. *Phys. Rev. A*, 90:023414, Aug 2014.

[98] Clemens Weninger, Michael Purvis, Duncan Ryan, Richard A. London, John D. Bozek, Christoph Bostedt, Alexander Graf, Gregory Brown, Jorge J. Rocca, and Nina Rohringer. Stimulated electronic x-ray raman scattering. *Phys. Rev. Lett.*, 111:233902, Dec 2013.

[99] Daniele Toffoli and Piero Decleva. A multichannel least-squares b-spline approach to molecular photoionization: Theory, implementation, and applications within the configuration–interaction singles approximation. *Journal of Chemical Theory and Computation*, 12(10):4996–5008, 2016.

[100] Thomas Bondo Pedersen and Henrik Koch. On the time-dependent lagrangian approach in quantum chemistry. *The Journal of chemical physics*, 108(13):5194–5204, 1998.

Sarai D. Folkestad [iD], Eirik F. Kjønstad [iD], Rolf H. Myhre [iD], Josefine H. Andersen [iD], Alice Balbi [iD], Sonia Coriani [iD], Tommaso Giovannini [iD], Linda Goletto [iD], Tor S. Haugland [iD], Anders Hutcheson [iD], Ida-Marie Høyvik [iD], Torsha Moitra [iD], Alexander C. Paul [iD], Marco Scavino [iD], Andreas S. Skeidsvoll [iD], Åsmund H. Tveten [iD], and Henrik Koch [iD]

COLLECTIONS

Paper published as part of the special topic on Electronic Structure Software

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

NWChem: Past, present, and future
The Journal of Chemical Physics **152**, 184102 (2020); https://doi.org/10.1063/5.0004997

PSI4 1.4: Open-source software for high-throughput quantum chemistry
The Journal of Chemical Physics **152**, 184108 (2020); https://doi.org/10.1063/5.0006002

TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations
The Journal of Chemical Physics **152**, 184107 (2020); https://doi.org/10.1063/5.0004635

The Journal of Chemical Physics

AIP Publishing

e^T 1.0: An open source electronic structure program with emphasis on coupled cluster and multilevel methods

View Online   Export Citation   CrossMark

Sarai D. Folkestad,[1] Eirik F. Kjønstad,[1] Rolf H. Myhre,[1] Josefine H. Andersen,[2] Alice Balbi,[3] Sonia Coriani,[2] Tommaso Giovannini,[1] Linda Goletto,[1] Tor S. Haugland,[1] Anders Hutcheson,[1] Ida-Marie Høyvik,[1] Torsha Moitra,[2] Alexander C. Paul,[1] Marco Scavino,[3] Andreas S. Skeidsvoll,[1] Åsmund H. Tveten,[1] and Henrik Koch[1,3,a]

AFFILIATIONS

[1] Department of Chemistry, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
[2] DTU Chemistry—Department of Chemistry, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[3] Scuola Normale Superiore, Piazza dei Cavalieri, 7, IT-56126 Pisa, PI, Italy

Note: This article is part of the JCP Special Topic on Electronic Structure Software.
[a] Author to whom correspondence should be addressed: henrik.koch@sns.it

ABSTRACT

The e^T program is an open source electronic structure package with emphasis on coupled cluster and multilevel methods. It includes efficient spin adapted implementations of ground and excited singlet states, as well as equation of motion oscillator strengths, for CCS, CC2, CCSD, and CC3. Furthermore, e^T provides unique capabilities such as multilevel Hartree–Fock and multilevel CC2, real-time propagation for CCS and CCSD, and efficient CC3 oscillator strengths. With a coupled cluster code based on an efficient Cholesky decomposition algorithm for the electronic repulsion integrals, e^T has similar advantages as codes using density fitting, but with strict error control. Here, we present the main features of the program and demonstrate its performance through example calculations. Because of its availability, performance, and unique capabilities, we expect e^T to become a valuable resource to the electronic structure community.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0004713

I. INTRODUCTION

During the last five decades, a wide variety of models and algorithms have been developed within the field of electronic structure theory and many program packages are now available to the community.[1] Programs with extensive coupled cluster functionality include CFOUR,[2] Dalton,[3] GAMESS,[4] Gaussian,[5] Molcas,[6] Molpro,[7] NWChem,[8] ORCA,[9] PSI4,[10] QChem,[11] and TURBOMOLE.[12] Although these are all general purpose quantum chemistry programs, each code is particularly feature rich or efficient in specific areas. For instance, a large variety of response properties[13] have been implemented in Dalton, CFOUR is particularly suited for gradients[14,15] and geometry optimization, and QChem is leading in equation of motion[16,17] (EOM) features. However, due to the long

history of many of these programs, it can be challenging to modify and optimize the existing features or to integrate new methods and algorithms.

In 2016, we began developing a coupled cluster code based on Cholesky decomposed electron repulsion integrals.[18,19] While starting anew, we have drawn inspiration from Dalton[3] and used it extensively for testing purposes. Our goal is to create an efficient, flexible, and easily extendable foundation upon which coupled cluster methods and features—both established and new—can be developed. That code has now evolved beyond a coupled cluster code into a freestanding electronic structure program. It is named e^T after the expression for the coupled cluster ground state wave function,[20]

$$|\Psi\rangle = e^T|R\rangle, \tag{1}$$

and released as an open source program licensed under the GNU General Public License 3 (GPL 3.0).

The first version of e^T offers an optimized Hartree–Fock (HF) code and a wide range of standard coupled cluster methods. It includes the most efficient published implementations of Cholesky decomposition of the electron repulsion integrals[21] and of coupled cluster singles, doubles, and perturbative triples[22,23] (CC3). Furthermore, e^T features the first released implementations of multilevel HF[24] (MLHF), multilevel coupled cluster singles and perturbative doubles[25,26] (MLCC2), and explicitly time-dependent coupled cluster singles (TD-CCS), and singles and doubles (TD-CCSD) theory. All coupled cluster models can be used in quantum mechanics/molecular mechanics[27,28] (QM/MM) calculations or be combined with the polarizable continuum model[29,30] (PCM).

e^T is primarily written in modern Fortran using the Fortran 2008 standard. The current version of the code is interfaced to two external libraries: Libint 2[31] for the atomic orbital integrals and PCMSolver 1.2[32] for PCM embedding. In addition, e^T applies the runtest library[33] for testing and a CMake module from autocmake[34] to locate and configure BLAS and LAPACK.

With the introduction of the 2003 and 2008 standards, Fortran has become an object oriented programming language. We have exploited this to make e^T modular, readable, and easy to extend. Throughout the program, we use OpenMP[35] to parallelize computationally intensive loops and BLAS and LAPACK routines wherever possible. In order to preserve code quality, extensive code review and enforcement of a consistent standard have been prioritized from the outset. While this requires extra effort from both developers and maintainers, it pays dividends in code readability and flexibility.

II. PROGRAM FEATURES

A. Coupled cluster methods

The e^T program features all standard coupled cluster methods up to perturbative triples: singles (CCS), singles with perturbative doubles[36] (CC2), singles and doubles[37] (CCSD), singles and doubles with non-iterative perturbative triples[38] [CCSD(T)], and singles and doubles with perturbative triples[22] (CC3). At the CCSD(T) level of theory, only ground state energies can be computed. For all other methods, efficient spin adapted implementations of ground and excited singlet states are available. Moreover, dipole and quadrupole moments, as well as EOM oscillator strengths, can be calculated. Equation of motion polarizabilities are available at the CCS, CC2, and CCSD levels of theory.

A number of algorithms are implemented to solve the coupled cluster equations. For linear and eigenvalue equations, we have implemented the Davidson method.[39] This algorithm is used to solve the ground state multiplier equations, response equations, and excited state equations. To handle nonlinear coupled cluster equations, we have implemented algorithms that use direct inversion of the iterative subspace[40,41] (DIIS) to accelerate convergence. The ground state amplitude equations can be solved using DIIS combined with the standard[1,42] quasi-Newton algorithm or exact Newton–Raphson. We also use a DIIS-accelerated algorithm[43] for

the nonlinear excited state equations in CC2 and CC3. Our implementation of DIIS incorporates the option to use the related conjugate residual with optimal trial vectors[44,45] (CROP) method for acceleration. For the nonperturbative coupled cluster methods, the asymmetric Lanczos algorithm is also available.[46,47]

The time-dependent coupled cluster equations can be explicitly solved for CCS and CCSD[48,49] using Euler, Runge–Kutta 4 (RK4), or Gauss–Legendre (GL2, GL4, and GL6) integrators. This requires implementations of the amplitude and multiplier equations with complex variables. Any number of classical electromagnetic pulses can be specified in the length gauge, assuming that the dipole approximation is valid. A modified version of the fast Fourier transform library FFTPACK 5.1[50] is used to extract frequency domain information.

B. Cholesky decomposition for the electronic repulsion integrals

Cholesky decomposition is an efficient method to obtain a compact factorization of the rank deficient electron repulsion integral matrix.[18,19,51] All post-HF methods in e^T rely on the Cholesky vectors to construct the electron repulsion integrals. One advantage of factorization is the reduced storage requirements; the size of the Cholesky vectors scales as $\mathcal{O}(n_{AO}^3)$, while the full integral matrix scales as $\mathcal{O}(n_{AO}^4)$. The Cholesky vectors are kept in memory when possible but are otherwise stored on disk. Another advantage is that they allow for an efficient construction and transformation of subsets of the integrals. The Cholesky decomposition in e^T is highly efficient, consisting of a two-step procedure that reduces both storage requirements and computational cost compared to earlier algorithms. For a description of the algorithm and performance comparisons to Molcas,[6] see Ref. 21.

C. Hartree–Fock

The restricted HF (RHF) and unrestricted HF (UHF) models are implemented in e^T. The implementations are integral direct and exploit Coloumb and exchange screening and permutation symmetry. We use a superposition of atomic densities[52] (SAD) initial guess constructed from spherically averaged UHF calculations on the constituent atoms. The Hartree–Fock equations are solved using a Roothan–Hall self-consistent field (SCF) algorithm accelerated by either DIIS or CROP. To improve the screening and reduce the number of integrals that must be evaluated, density differences are used to construct the Fock matrix.

D. Multilevel and multiscale methods

In MLHF, a region of the molecular system is defined as active. A set of active occupied orbitals are obtained through a restricted, partial Cholesky decomposition of an initial idempotent AO density matrix.[53] The active virtual orbitals are obtained by constructing projected atomic orbitals[54,55] (PAOs) centered on the active atoms. The PAOs are orthonormalized through the canonical orthonormalization procedure.[56] The MLHF equations are solved using a DIIS accelerated, MO based, Roothan–Hall SCF algorithm. Only the active MOs are optimized.[57]

The most expensive step of an MLHF calculation is the construction of the inactive two-electron contribution to the Fock

matrix. As the inactive orbitals are frozen, it is only necessary to calculate this term once. The iterative cost in MLHF is dominated by the construction of the active two-electron contribution to the Fock matrix. An additional Coulomb and exchange screening, which targets accuracy of the matrix in the active MO basis, reduces the cost. The active orbitals are localized, and consequently, the elements of the AO Fock matrix that correspond to AOs distant from the active atoms will not significantly contribute to the active MO Fock matrix. This is similar to the screening used in MLHF specific Cholesky decomposition of the electron repulsion integrals.[21]

In MLCC2,[23,25,26,58] an active orbital space is treated at the CC2 level of theory, while the remaining inactive orbitals are treated at the CCS level of theory. MLCC2 excitation energies are implemented in e^T. The active space is constructed using the approximated correlated natural transition orbitals,[59,60] Cholesky orbitals, or Cholesky occupied orbitals and PAOs spanning the virtual space.

Frozen orbitals are implemented for all coupled cluster methods in e^T. In addition to the standard frozen core (FC) approximation, reduced space coupled cluster calculations can be performed using semi-localized orbitals. This type of calculation is suited to describe localized properties. In reduced space calculations, the occupied space is constructed from Cholesky orbitals, and PAOs are used to generate the virtual space.

Two QM/MM approaches are available in e^T: electrostatic QM/MM embedding[61] and the polarizable QM/Fluctuating Charge[62] (QM/FQ) model. In the former, the QM density interacts with a set of fixed charges placed in the MM part of the system.[61] In QM/FQ, the QM and MM parts mutually polarize. Each atom in the MM part has a charge that varies as a response to differences in atomic electronegativities and the QM potential.[62] These charges enter the QM Hamiltonian through a term that is nonlinear in the QM density.[63]

PCM embedding can be used in e^T for an implicit description of the external environment. A solute is described at the QM level and is placed in a molecule shaped cavity. The environment is described in terms of an infinite, homogeneous, continuum dielectric that mutually polarizes with the QM part, as in QM/FQ.[64]

In the QM/PCM and QM/FQ implementations, additional terms are only added to the Fock matrix. Additional terms at the coupled cluster level can also be considered.[65–69]

E. Spectroscopic properties and response methods

Coupled cluster is one of the most accurate methods for modeling spectroscopic properties, and both ultraviolet-visible (UV/vis) and x-ray absorption spectra can be modeled in e^T. Core excitations are obtained through the core valence separation (CVS) approximation.[70] CVS is implemented as a projection[71,72] for CCS, CC2, MLCC2, and CCSD. For CC3, amplitudes and excitation vector elements that do not contribute are not calculated. This reduces the scaling of the iterative computational cost for excited states from $\mathcal{O}(n_{MO}^7)$ to $\mathcal{O}(n_{MO}^6)$.

Intensities are obtained from EOM oscillator strengths,[16,17] which are available for CCS, CC2, CCSD, and CC3. In addition, linear response[48] (LR) oscillator strengths can be calculated at the CCS level of theory. The asymmetric Lanczos algorithm[46,47] can be used to directly obtain both energies and EOM oscillator strengths

for CCS, CC2, and CCSD. It can also be combined with the CVS approximation.

Real-time propagation offers a nonperturbative approach to model absorption spectra. Following an initial pulse that excites the system, the dipole moment from the subsequent time evolution can be Fourier transformed to extract the excitation energies and intensities.

Valence ionization potentials are implemented for CCS, CC2, and CCSD. A bath orbital that does not interact with the system is added to the calculation. Excitation vector components not involving this orbital are projected out in an approach similar to the projection in CVS.[71,72]

III. ILLUSTRATIVE APPLICATIONS AND PERFORMANCE TESTS

In this section, we will demonstrate some of the capabilities of e^T with example calculations. Energy thresholds refer to the change in energy from the previous iteration. The maximum norm of the gradient vector is used in Hartree–Fock calculations. For coupled cluster calculations in e^T and Dalton, residual thresholds refer to the L^2 norm of the residual vectors. Finally, the Cholesky decomposition threshold refers to the largest absolute error on the diagonal of the electron repulsion integral matrix. This threshold gives an upper bound to the error of all matrix elements. Coupled cluster calculations were performed with either Cholesky vectors or electron repulsion integrals in memory. All geometries are available from Ref. 73.

A. Coupled cluster methods

The CC2 method is known to yield excitation energies with errors of about 0.1–0.4 eV for valence states with single excitation character.[74–76] The iterative cost of CC2 scales as $\mathcal{O}(n_{MO}^5)$, and it may be implemented with an $\mathcal{O}(n_{MO}^2)$ memory requirement. In Table I, we report the lowest FC-CC2/aug-cc-pVDZ excitation energy of the antibiotic rifampicin[77] (chemical formula $C_{43}H_{58}N_4O_{12}$, see Fig. 1). The calculated excitation energy is 2.58 eV, which is consistent with the orange color of the compound. The ground state was converged to a residual threshold of 10^{-6}, and the excited state was converged to residual and energy thresholds of 10^{-3} and 10^{-6}, respectively. We used a Cholesky decomposition threshold of 10^{-2}, which is sufficient to ensure accuracy of excitation energies in CC2 and CCSD (see Table IV). The calculation was performed on two Intel Xeon Gold 6138 processors using 40 threads and 360 GB shared memory. The average iteration time for the ground state equations was 73 min, and the average iteration time for the excited state equations was 9 h.

TABLE I. The lowest FC-CC2/aug-cc-pVDZ excitation energy (ω) of rifampicin. n_{frozen} is the number of frozen core orbitals.

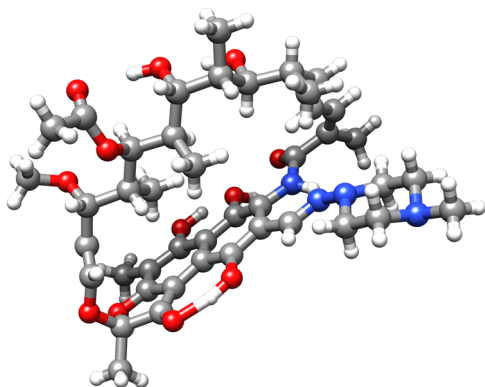| n_{AO}	n_{MO}	n_{frozen}	ω
1879	1865	59	2.579 eV

FIG. 1. Rifampicin.

At the CCSD level of theory, we report calculations for the amino acid tryptophan[78] (chemical formula $C_{11}H_{12}N_2O_2$) and excitation energies for the psychoactive agent lysergic acid diethylamide (LSD)[79] (chemical formula $C_{20}H_{25}N_3O$). Tryptophan and LSD are depicted in Fig. 2.

For tryptophan, we have determined the four lowest excitation energies and the corresponding oscillator strengths at the CCSD/aug-cc-pVDZ level of theory ($n_{MO} = 453$). Energies and oscillator strengths are reported in Table II. Timings for e^T 1.0, Dalton 2018, and QChem 5.0 are given in Table III. Thresholds in e^T were set to target an energy convergence of 10^{-6}: the residuals were converged to 10^{-6} for the ground state and 10^{-3} for the excited states (assuming quadratic errors for the energy). In QChem, thresholds for ground and excited states were set to 10^{-6}. We report the total wall time for each calculation. The excited state timing includes the time to converge ground state and excited state equations. The oscillator strength timing also includes the time to solve the multiplier and the left excited state equations. e^T and QChem are equally efficient for the CCSD ground state, while Dalton is considerably slower. For the CCSD excited state calculation, QChem reduced the wall time by a factor of 1.6 compared to e^T and a factor of 5.6 compared to Dalton. For the oscillator strength calculations, QChem reduced the wall time by a factor of 2.7 compared to e^T. The superior performance of QChem for oscillator strengths is primarily due to an efficient starting guess for the left excitation vectors, which are restarted from the right vectors: only 27 transformations are needed to converge all four roots. In e^T 1.0, orbital differences are used as the starting guess for both left and right states, which explains the poorer performance for oscillator strengths.
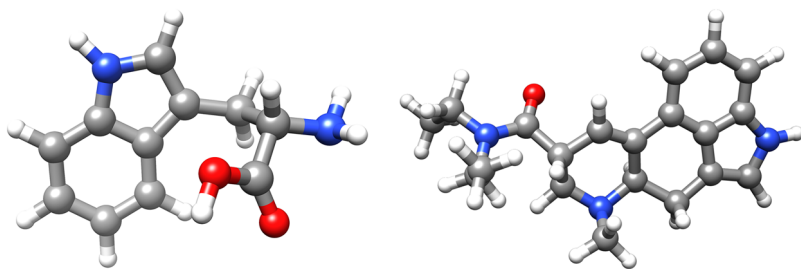
TABLE II. CCSD/aug-cc-pVDZ excitation energies (ω) and oscillator strengths (f_ω) for tryptophan.

	ω (eV)	f_ω
S_1	4.806	0.032
S_2	4.821	0.001
S_3	4.972	0.088
S_4	5.364	0.001

TABLE III. Total calculation times for CCSD/aug-cc-pVDZ ground state (t_{gs}), excitation energy (t_ω), and oscillator strength (t_{f_ω}) calculations for tryptophan. n_{calls}^{gs} is the number of calculations of the residual vector for the ground state. n_{calls}^R and n_{calls}^L are the number of calls to the Jacobian and Jacobian transpose transformations, respectively. The calculations were performed on an Intel Xeon E5-2699 v4 using 44 threads and 1.5 TB shared memory.

	t_{gs} (min)	t_ω (h)	t_{f_ω} (h)	n_{calls}^{gs}	n_{calls}^R	n_{calls}^L
Dalton 2018	1409	84	...	18	88	...
e^T 1.0	201	24	53	16	79	81
QChem 5.0	196	15	20	18	90	27

TABLE IV. The FC-CCSD/aug-cc-pVDZ correlation energy (E_{corr}) and lowest excitation energy (ω) of LSD. A set of decomposition thresholds (τ) for the Cholesky decomposition of the electron repulsion integral matrix were used. Both the ground and excited state equations are converged to within a residual threshold of 10^{-6}. Deviations in the correlation and excitation energies (ΔE_{corr} and $\Delta\omega$) are relative to $\tau = 10^{-8}$.

τ	E_{corr} (E_h)	ΔE_{corr} (E_h)	ω (E_h)	$\Delta\omega$ (E_h)
10^{-2}	−3.649 673 3	2.3×10^{-2}	0.165 734 3	7.1×10^{-4}
10^{-3}	−3.672 021 8	2.3×10^{-4}	0.165 037 0	7.7×10^{-6}
10^{-4}	−3.672 342 1	-9.2×10^{-5}	0.165 027 9	-1.4×10^{-6}
10^{-6}	−3.672 254 2	-3.6×10^{-6}	0.165 029 4	1.1×10^{-7}
10^{-8}	−3.672 250 6	...	0.165 029 3	...

We have performed FC-CCSD/aug-cc-pVDZ calculations on LSD ($n_{MO} = 777$, $n_{frozen} = 24$). To demonstrate the effect of integral approximation through Cholesky decomposition, we consider a range of decomposition thresholds. The correlation energy and the lowest excitation energy are given in Table IV. Both ground

FIG. 2. Tryptophan (left) and LSD (right).

and excited state residual thresholds are 10^{-6}. With a decomposition threshold of 10^{-2}, the error in the excitation energy ($\Delta\omega$) is less than $10^{-3} E_h$, well within the expected accuracy of FC-CCSD.[74–76]

The CC3 model can be used to obtain highly accurate excitation energies. However, an iterative cost that scales as $\mathcal{O}(n_v^4 n_o^3)$ severely limits system size. To the best of our knowledge, e^T 1.0 includes the fastest available implementation of CC3. A ground and excited state calculation on glycine (chemical formula $C_2H_5NO_2$) using the aug-cc-pVDZ basis set took 33 min with e^T 1.0. Comparable numbers for the new[23] and old[80] CC3 implementations in Dalton 2018[3] were 73 min and 1279 min, respectively.

We have calculated valence and core excitation energies and EOM oscillator strengths for the nucleobase uracil (chemical formula $C_4H_4N_2O_2$, see Fig. 3). The geometry was optimized at the CCSD(T)/aug-cc-pVDZ level using CFOUR.[2] One valence excitation energy was calculated at the FC-CCSD/aug-cc-pVTZ and FC-CC3/aug-cc-pVTZ levels of theory ($n_{MO} = 452$). Two core excited states were calculated for each of the oxygen atoms (O1 and O2, see Fig. 3) at the CCSD and CC3 levels. The aug-cc-pCVTZ basis was used on the oxygen being excited and aug-cc-pVDZ on the remaining atoms ($n_{MO} = 256$). The results are given in Table V. The total timings for the uracil calculations are presented in Table VI. In Table VII, we present the averaged timings from the CVS calculations. They clearly demonstrate the reduced computational cost of the CVS implementation for CC3. The ground state calculation was about four times more expensive per iteration than the right excited state. Without the CVS approximation, the computational cost of the excited states scales as $4n_v^4 n_o^3$ per iteration, while the ground state scales as $2n_v^4 n_o^3$. Using CVS, the excited state scaling is reduced to $4n_v^4 n_o^2$.

In Table VIII, we compare the timings for solving the ground and right excited state equations of glycine with aug-cc-pVDZ for different number of threads. All calculations were run on similar nodes, and all 40 cores on each node were reserved for the jobs to
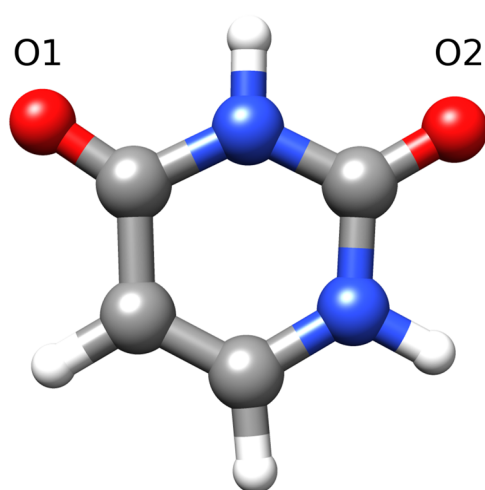

FIG. 3. Uracil with labels on the oxygens.

TABLE V. CC3 valence and core (oxygen edge) excitation energies (ω) and EOM oscillator strengths (f_ω) for uracil. Valence excitations were calculated with the aug-cc-pVTZ basis on all atoms and the frozen core approximation. Core excitations were calculated using the CVS approximation with the aug-cc-pCVTZ basis on the oxygen atom being excited and the aug-cc-pVDZ basis on the remaining atoms.

	CCSD		CC3	
	ω (eV)	f_ω	ω (eV)	f_ω
Valence	5.08	2.24×10^{-8}	4.81	2.23×10^{-6}
Core O1	536.04	3.35×10^{-2}	533.64	1.95×10^{-2}
	539.60	3.23×10^{-4}	535.66	2.24×10^{-4}
Core O2	536.98	3.13×10^{-2}	534.64	1.32×10^{-2}
	539.44	1.47×10^{-4}	535.75	1.34×10^{-4}

TABLE VI. Total wall times for CC3 on uracil. The valence calculation was performed on a node with two Intel Xeon Gold 6138 processors using 40 threads and 320 GB shared memory. The CVS calculations were performed on a node with two Intel Xeon Gold 6138 processors using 40 threads and 150 GB shared memory. n_o and n_v are the number of occupied and virtual orbitals, respectively.

Calculation	Basis set	t (h)	n_o	n_v
Valence excitation	aug-cc-pVTZ	147	21	431
CVS O1	aug-cc-pV(CT)Z	36	29	227
CVS O2	aug-cc-pV(CT)Z	38	29	227

TABLE VII. Average wall time per function call for both CC3 core excitation calculations on uracil. n_{calls} is the total number of routine calls in the two calculations.

Contributions	t (min)	n_{calls}
Ground state amplitudes	14	28
Ground state multipliers	23	30
Right excited states	4	195
Left excited states	7	244

minimize variation. Increasing the number of threads results in significant reductions in time, even for a relatively small system such as glycine with 20 occupied and 140 virtual orbitals. Intermediates are currently stored on disk, resulting in overhead that can be reduced by placing them in memory when possible. In addition to more adaptive memory usage, we are working on improving the coupled cluster algorithms for better parallelization.

B. Cholesky decomposition

We have determined the Cholesky basis for the transmembrane ion channel gramicidin A (chemical formula $C_{198}H_{276}N_{40}O_{34}$, see Fig. 4). The geometry is taken from the supplementary material of Ref. 81. Decomposition times are given in Table IX for the cc-pVDZ and aug-cc-pVDZ basis sets and a range of

TABLE VIII. Time to solve the ground and right excited state equations of glycine for CC3 and CCSD with different numbers of threads in seconds using a development version of e^T 1.1. Factor is the improvement compared to the row above. With perfect parallelization, the factors would be 5, 2, and 2. The calculations were performed on nodes with two Intel Xeon Gold 6138 processors and 150 GB shared memory.

Threads	CC3 GS		CC3 ES		CCSD GS		CCSD ES	
	Time (s)	Factor	Time (s)	Factor	Time (s)	Factor	Time (s)	Factor
1	6048	. . .	15 617	. . .	639	. . .	654	. . .
5	1681	3.60	4 418	3.53	150	4.27	238	2.75
10	923	1.82	2 308	1.91	76	1.96	126	1.89
20	675	1.37	1 482	1.56	41	1.88	72	1.76
40	532	1.27	1 252	1.18	30	1.34	53	1.34

decomposition thresholds. These are compared to the time of one HF iteration. Except when using cc-pVDZ with the tightest threshold, the decomposition time is small or negligible compared to one Fock matrix construction.
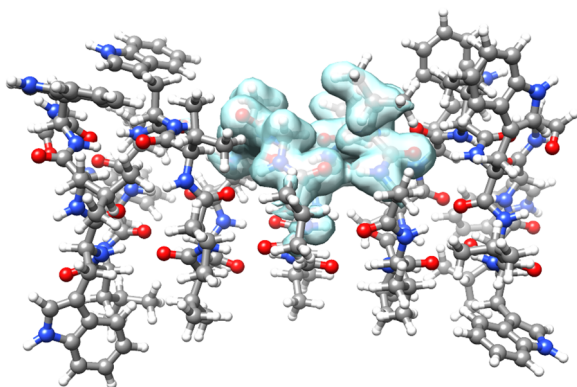
FIG. 4. Gramicidin A. The active MLHF/cc-pVDZ density is shown.

TABLE IX. Cholesky decomposition wall times (t) for gramicidin. τ is the decomposition threshold, and n_J is the number of Cholesky vectors. For reference, we include the time (t_{it}^{HF}) for one full Hartree–Fock iteration. All calculations were performed on an Intel Xeon E5-2699 v4 using 44 threads and 1.5 TB shared memory.

Basis	n_{AO}	τ	n_J	t (min)	t_{it}^{HF} (min)
cc-pVDZ	5188	10^{-2}	11 574	3	35
		10^{-3}	16 368	6	
		10^{-4}	24 652	12	
		10^{-8}	75 446	125	
aug-cc-pVDZ	8740	10^{-2}	12 813	8	1191
		10^{-3}	18 587	27	
		10^{-4}	29 818	61	
		10^{-8}	90 656	645	

C. Hartree–Fock

Systems with several hundred atoms are easily modeled in e^T using Hartree–Fock. In Table X, we present the wall times for calculations on gramicidin A (see Fig. 4) and an amylose chain with 16 glucose units (chemical formula $C_{96}H_{162}O_{81}$, see Fig. 5). The amylose geometry is taken from Ref. 24. We compare the results and timings from e^T 1.0 and QChem 5.0.[11] This comparison is complicated because the accuracy depends on several thresholds apart from the gradient and energy thresholds, e.g., screening thresholds and integral accuracy. Therefore, we list the energies and absolute energy differences along with the timings in Table X. QChem 5.0 outperforms e^T by about a factor of 2. The energies converge to slightly different results in the two programs. In the case of amylose, we find a 2×10^{-7} E_h energy difference using the tightest thresholds ($\tau_{SCF} = 10^{-10}$). Since QChem is a closed source program, we do not know the reason for the deviation. However, we are able to reproduce the e^T results for amylose to all digits using tight thresholds in LSDalton 2018.[3]

D. Multilevel and multiscale methods

To demonstrate the efficacy of multilevel methods for excitation energies, we consider a system of sulfur dioxide with 21 water molecules (see Fig. 6). In Table XI, we present different flavors of multilevel calculations to approximate the two lowest FC-CC2 excitation energies for this system. Three sets of active atoms are defined. The first set contains sulfur dioxide and nine water molecules; these atoms determine the active orbitals of the MLHF calculation. The second set contains sulfur dioxide and five water molecules; these atoms determine the reduced space coupled cluster calculations. The third set contains only sulfur dioxide and determines the CC2 active space in the MLCC2 calculations. The reduced space FC-CC2 calculations are denoted FC-CC2-in-HF and FC-CC2-in-MLHF and similarly for the reduced space FC-MLCC2 calculations. The orbital spaces are partitioned using the Cholesky occupied orbitals and PAOs for the virtual orbitals. In all calculations, the deviation with respect to full FC-CC2 is within the expected error of CC2.[75,76]

In order to assess the performance of the MLHF implementation, we compare full HF and MLHF for gramicidin A and amylose. The active electron densities from the MLHF calculations are shown

TABLE X. Hartree–Fock/cc-pVDZ calculations on amylose and gramicidin. The total wall time is denoted by t, and τ_{SCF} is the Hartree–Fock convergence threshold. We present timings for e^T and QChem along with the computed Hartree–Fock energies (E) and absolute energy differences ($|\Delta E|$) with respect to the calculation with the tightest threshold. Calculations were performed on two Intel Xeon E5-2699 v4 processors using 44 threads and 1.5 TB shared memory.

	e^T			QChem			
	τ_{SCF}	E (E_h)	$|\Delta E|$ (E_h)	t (min)	τ_{SCF}	E (E_h)	$|\Delta E|$ (E_h)	t (min)
Amylose	10^{-3}	−9792.085 129 90	4×10^{-5}	21	10^{-5}	−9792.085 350 39	2×10^{-4}	9
	10^{-4}	−9792.085 178 33	5×10^{-6}	31	10^{-6}	−9792.085 180 84	7×10^{-6}	14
	10^{-5}	−9792.085 174 42	7×10^{-7}	42	10^{-7}	−9792.085 171 19	2×10^{-6}	19
	10^{-6}	−9792.085 173 77	1×10^{-8}	60	10^{-8}	−9792.085 173 23	4×10^{-7}	26
	10^{-7}	−9792.085 173 76	$<1 \times 10^{-8}$	78	10^{-9}	−9792.085 173 61	3×10^{-8}	33
	10^{-10}	−9792.085 173 76	. . .	153	10^{-10}	−9792.085 173 58	. . .	46
Gramicidin	10^{-4}	−12 383.458 832 54	4×10^{-6}	130	10^{-6}	−12 383.458 825 13	1×10^{-5}	50
	10^{-5}	−12 383.458 836 34	7×10^{-8}	198	10^{-7}	−12 383.458 827 10	1×10^{-5}	77
	10^{-6}	−12 383.458 836 27	. . .	280	10^{-8}	−12 383.458 836 77	. . .	111
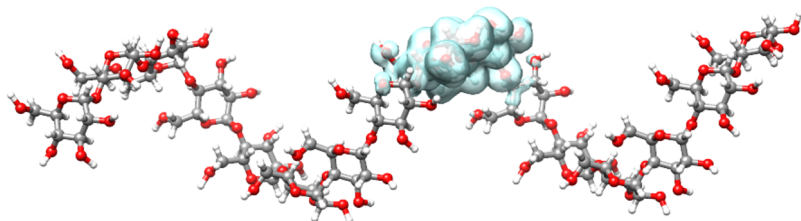
FIG. 5. Amylose chain of 16 glucose units. The active MLHF/cc-pVDZ density is shown.
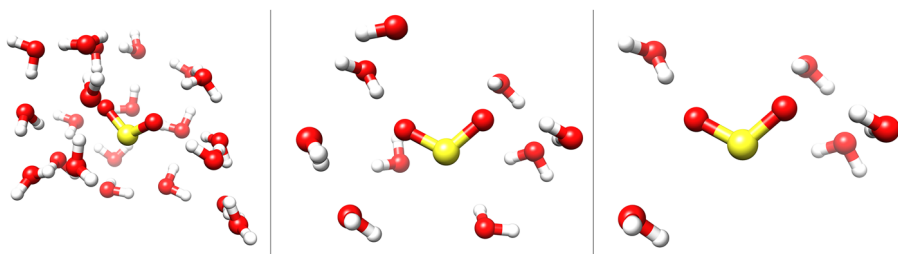
FIG. 6. SO_2 and water. (Left) SO_2 and 21 water molecules. (Middle) SO_2 and nine water molecules; these are the HF active atoms in the MLHF calculations. (Right) SO_2 and five water molecules; these are the CC active atoms. In the MLCC2 calculations, only SO_2 is treated at the CC2 level of theory.

TABLE XI. The two lowest excitation energies (ω_1 and ω_2) of SO_2 with 21 water molecules, calculated with full and reduced space FC-CC2 and FC-MLCC2 using HF and MLHF reference wave functions. The deviation from full FC-CC2 ($\Delta\omega_i = \omega_i - \omega_i^{FC\text{-}CC2}$) is given. We also list the number of occupied (n_o) and virtual (n_v) orbitals treated at the different levels of theory. There are a total of 121 occupied orbitals and 813 virtual orbitals in the system.

	HF		CCS		CC2		ω_1	$\Delta\omega_1$	ω_2	$\Delta\omega_1$
Calculation	n_o	n_v	n_o	n_v	n_o	n_v	(eV)	(eV)	(eV)	(eV)
FC-CC2	121	813	. . .	. . .	93	813	3.11	. . .	3.39	. . .
FC-CC2-in-HF	121	813	. . .	. . .	40	266	3.14	0.03	3.43	0.04
FC-CC2-in-MLHF	75	426	. . .	. . .	40	266	3.16	0.05	3.44	0.05
FC-MLCC2	121	813	93	813	14	67	3.18	0.07	3.45	0.06
FC-MLCC2-in-HF	121	813	40	266	14	66	3.18	0.07	3.45	0.06
FC-MLCC2-in-MLHF	75	426	40	266	15	66	3.20	0.09	3.47	0.08

TABLE XII. Multilevel Hartree–Fock wall times for amylose and gramicidin. t_{it} is the wall time to construct the Fock matrix. For the calculations with (aug)-cc-pVDZ, aug-cc-pVDZ is used on the active atoms and cc-pVDZ for the rest. The total number of AOs and the active MOs is labeled n_{AO} and n_{MO}^{active}, respectively. Thresholds for Coulomb and exchange are set to 10^{-12} and 10^{-10}, respectively, and the integral cutoff is set to 10^{-12}. Calculations were performed on two Intel Xeon E5-2699 v4 processors using 44 threads and 1.5 TB shared memory.

		HF		MLHF	
	Basis	n_{AO}	t_{it} (min)	n_{MO}^{active}	t_{it} (min)
Amylose	cc-pVDZ	3288	8	335	1
	(aug)-cc-pVDZ	3480	11	552	4
Gramicidin	cc-pVDZ	5188	35	546	11
	(aug)-cc-pVDZ	5506	69	942	50

in Figs. 4 and 5. The plots were generated using UCSF Chimera.[82] Cholesky orbitals were used to partition the occupied space, and PAOs were used for the virtual space. We present the timings in Table XII. For amylose, the iteration times are reduced significantly with MLHF: by a factor of 8 when cc-pVDZ is used on all atoms

and a factor of 3 when aug-cc-pVDZ is used on the active atoms. In contrast, only a factor of 3 was reported by Sæther *et al.*[24] in the cc-pVDZ case. The iteration time is also reduced by a factor of 8 for amylose/cc-pVDZ ($t_{iteration}$ = 1 m, n_{MO}^{active} = 318) when using Cholesky virtuals (as in Ref. 24) instead of PAOs. The savings for amylose reflect the small active region as well as the linear structure of the chain. Savings are less significant for the gramicidin system, where the MLHF iteration time is a third of the HF iteration time for cc-pVDZ, but only about two thirds when the active atoms are described using aug-cc-pVDZ. The smaller savings reflect the relatively large active region and the more compact shape of the gramicidin system.

For systems in solution, electronic spectra can be calculated using QM/MM or QM/PCM. Paranitroaniline (PNA) has an experimental vacuum-to-water solvatochromism of about 1 eV.[87] For QM/PCM, we use two different atomic radii, UFF[85] (QM/PCM^c) and Bondi[86] (QM/PCM^d), and the dielectric permittivity of water was set to ε = 78.39. For QM/MM, 64 snapshots were extracted from a classical molecular dynamics simulation[88] [see Fig. 7(a) for an example structure]. The UV/vis spectra were then computed by treating PNA at the CC2/aug-cc-pVDZ level and modeling the water using an FQ force field. Here, we present results using two different FQ parameterizations: QM/FQ^a from Ref. 83 and QM/FQ^b from
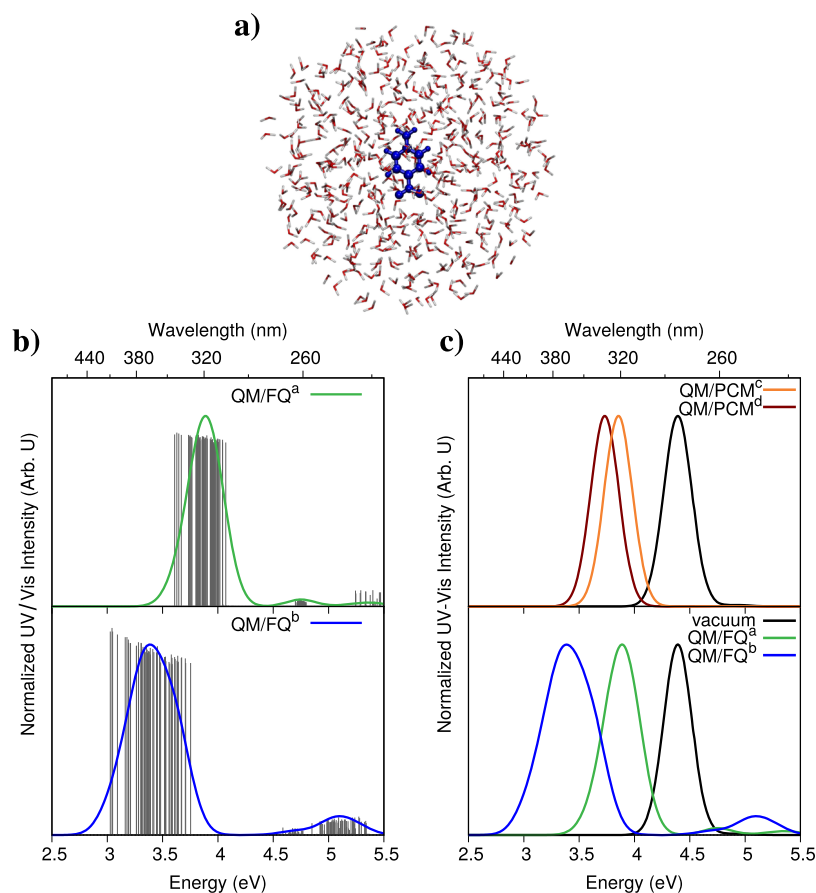
FIG. 7. (a) Schematic representation of a random snapshot of PNA in aqueous solution. (b) and (c) UV/vis spectra of PNA calculated at the CC2/aug-cc-pVDZ level of theory with an aqueous solution described at the PCM or FQ level of theory. (b) QM/FQ raw data (sticks) together with their Gaussian convolution (FWHM = 0.3 eV). (c) QM/PCM (top) and QM/FQ (bottom) spectra in aqueous solution. A gas phase CC2/aug-cc-pVDZ reference spectrum is also reported (black). For QM/FQ^a, the FQ parameterization is from Ref. 83, and for QM/FQ^b, the parameterization is from Ref. 84. In QM/FQ^c, the PCM cavity is constructed using the UFF radii,[85] and in QM/FQ^d, it is constructed using the Bondi radii.[86]

TABLE XIII. The first vertical excitation energy of PNA in vacuum (ω_v) and in aqueous solution (ω_s), as well as water-to-vacuum solvatochromic shifts ($\Delta\omega$). PNA is treated at the CC2/aug-cc-pVDZ level of theory, and the solution is described with PCM or FQ. 68% confidence intervals for excitation energies are also reported for QM/FQ, calculated as $\sigma/\sqrt{N}$, where σ is the standard deviation and N is the number of the snapshots used to obtain the average property. Experimental data are taken from Ref. 87.

	ω_v (eV)	ω_s (eV)	$\Delta\omega$ (eV)
CC2	4.38	. . .	. . .
CC2/FQ[a]	. . .	3.88±0.01	0.50±0.01
CC2/FQ[b]	. . .	3.38±0.01	1.00±0.01
CC2/PCM[c]	. . .	3.86	0.52
CC2/PCM[d]	. . .	3.76	0.62
Expt.[e]	4.25	3.26	0.99

[a]FQ parameterization taken from Ref. 83.
[b]FQ parameterization taken from Ref. 84.
[c]PCM cavity constructed by exploiting UFF radii.[85]
[d]PCM cavity constructed by exploiting the Bondi radii.[86]
[e]Reference 87.

Ref. 84 (see the supplementary material for additional computational details).

The spectra calculated using QM/FQ are presented in Fig. 7(b). The results for individual snapshots are presented as sticks together with their Gaussian convolution. As can be seen from Fig. 7, QM/FQ[b] results in a greater spread in the excitation energies. This

is probably due to the larger molecular dipole moments of the water molecules in this parameterization.[88,89]

In Fig. 7(c), the convoluted spectra calculated using QM/PCM[c] and QM/PCM[d] (top), and QM/FQ[a] and QM/FQ[b] (bottom), are presented with their vacuum counterparts. The excitation energies are also given in Table XIII together with the experimental data from Ref. 87. For QM/FQ, we also report 68% confidence intervals for the calculated excitation energies. QM/FQ[b] reproduces the experimental solvatochromism, while the other approaches give errors of 40%–50%.

E. Modeling spectroscopies

The spectroscopic properties can also be modeled with the Lanczos method or with real-time propagation of the coupled cluster wave function. In Fig. 8, we show CCSD/aug-cc-pCVDZ UV/vis absorption spectra of H_2O,[90] calculated using the Davidson (top) and asymmetric Lanczos (bottom) algorithms. Note that we have artificially extended the spectra beyond the ionization potential (12.3 eV IP-CCSD/aug-cc-pCVDZ) to illustrate convergence behavior. With the Lanczos algorithm, the low energy part of the spectrum converges with a smaller reduced space than the high energy part.[47]

We have also generated oxygen edge x-ray absorption spectra using the Davidson and Lanczos algorithms with CVS projection (see Fig. 9. We see the same overall behavior as in Fig. 8. Absorption spectra can also be obtained from real-time propagation of the coupled cluster wave function (see Fig. 10 for UV/vis and oxygen edge x-ray absorption spectra; see the
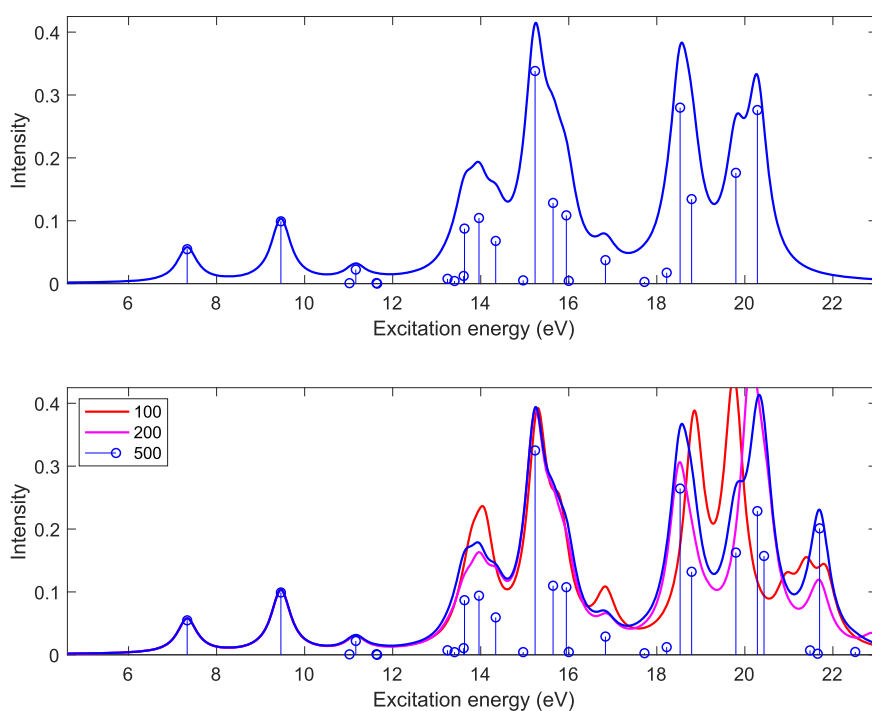
FIG. 8. Water CCSD/aug-cc-pCVDZ UV/vis absorption spectrum. Lorentzian broadening (0.02 E_h FWHM) has been applied to the stick spectra. The top plot shows the spectrum obtained using the Davidson. The spectrum in the bottom plot is from Lanczos calculations with chain lengths 100 (red), 200 (magenta), and 500 (blue).
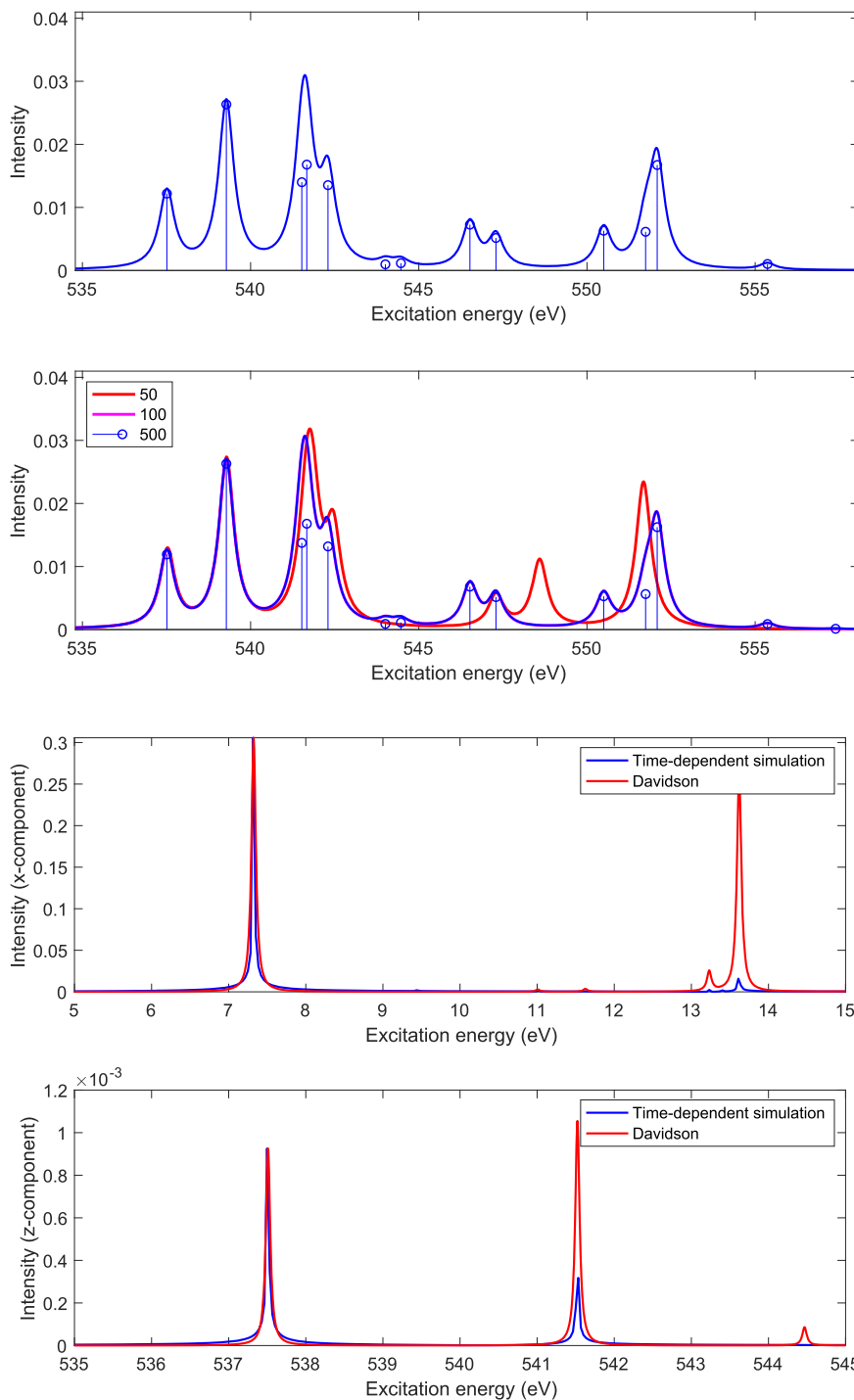
FIG. 9. Water CCSD/CVS/aug-cc-pCVDZ x-ray absorption spectrum. Lorentzian broadening (0.02 E_h FWHM) has been applied to the stick spectra. The top plot shows the spectrum obtained using the Davidson. The spectrum in the bottom plot is from Lanczos calculations with chain lengths 50 (red), 100 (magenta), and 500 (blue).

FIG. 10. Water UV and x-ray CCSD absorption spectra obtained using Davidson (CVS/aug-cc-pCVDZ for x ray and aug-cc-pVDZ for UV) and real-time propagation (aug-cc-pCVDZ for x ray and aug-cc-pVDZ for UV). The top and bottom plots show the simulated UV and x-ray spectra, respectively. The Davidson spectra were produced by applying Lorentzian broadening to the stick spectra (0.0025 E_h FWHM). Intensities from the time-dependent simulation have been scaled so that the intensity of the first peak matches the EOM oscillator strength.

supplementary material for computational details). The first peak in both plots has been scaled to match the intensity obtained using Davidson. The position of the peaks are the same with both approaches, but the intensities differ because we specified pulses with frequency distributions centered on the first excitation energy.

IV. CONCLUDING REMARKS

e^T 1.0 is an optimized open source electronic structure program. Several features are worth emphasizing. To the best of our knowledge, our CC3 implementation is the fastest for calculating

ground and excited state energies and EOM oscillator strengths. The low memory CC2 code has memory and disk requirements of order $\mathcal{O}(n_{\text{MO}}^2)$ and $\mathcal{O}(n_{\text{MO}}^3)$, respectively, allowing us to treat systems with thousands of basis functions. At the core of our program is the Cholesky decomposition of the electron repulsion integral matrix; our implementation is faster and less storage intensive than that of any other program. Exciting new developments are also part of e^T. It features the only spin adapted closed shell implementation of time-dependent coupled cluster theory. Furthermore, the MLHF and MLCC2 methods extend the treatable system size without sacrificing accuracy for intensive properties such as excitation energies.

The e^T source code is written in modern object oriented Fortran, making it easy to expand and contribute to the program. It is freely available on GitLab,[91] and the manual can be found at www.etprogram.org. We will continue to expand the capabilities of e^T, focusing on molecular properties and multilevel methods. We believe that the program will be useful for the quantum chemistry community, both as a development platform and for production calculations.

SUPPLEMENTARY MATERIAL

See the supplementary material for details regarding QM/MM calculations as well as specifications for the time-dependent CCSD propagation calculations.

AUTHOR'S CONTRIBUTIONS

S.D.F, E.F.K., and R.H.M contributed equally to this work.

ACKNOWLEDGMENTS

DATA AVAILABILITY

The e^T source code is publicly available on Gitlab (see Ref. 91). The geometries used in Sec. III are available at https://doi.org/10.5281/zenodo.3666109 (Ref. 73). Input files are available from the corresponding author upon request.

REFERENCES

[1] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley & Sons, 2014).

[2] J. Stanton, J. Gauss, M. Harding, P. Szalay, A. Auer, R. Bartlett, U. Benedikt, C. Berger, D. Bernholdt, Y. Bomble, *et al.*, "CFOUR, coupled-cluster techniques for computational chemistry, a quantum-chemical program package," current version available at http://www.cfour.de, 2010.

[3] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenaes, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjaergaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnaes, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren, "The Dalton quantum chemistry program system," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 269–284 (2014).

[4] M. S. Gordon and M. W. Schmidt, "Advances in electronic structure theory: GAMESS a decade later," in *Theory and Applications of Computational Chemistry*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005), Chap. 41, pp. 1167–1189.

[5] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford, CT, 2016.

[6] F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, "MOLCAS 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table," J. Comput. Chem. **37**, 506–541 (2016).

[7] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "MOLPRO: A general-purpose quantum chemistry program package," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242–253 (2012).

[8] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations," Comput. Phys. Commun. **181**, 1477–1489 (2010).

[9] F. Neese, "The ORCA program system," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 73–78 (2012).

[10] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, "PSI4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability," J. Chem. Theor. Comput. **13**, 3185–3197 (2017).

[11] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," Mol. Phys. **113**, 184–215 (2015).

[12] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, "Turbomole," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 91–100 (2014).

[13] T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, "Recent advances in wave function-based methods of molecular-property calculations," Chem. Rev. **112**, 543–631 (2012).

[14] H. Koch, H. J. A. Jensen, P. Jørgensen, T. Helgaker, G. E. Scuseria, and H. F. Schaefer, "Coupled cluster energy derivatives. Analytic hessian for the closed-shell coupled cluster singles and doubles wave function: Theory and applications," J. Chem. Phys. **92**, 4924–4940 (1990).

[15] J. F. Stanton, "Many-body methods for excited state potential energy surfaces. I. General theory of energy gradients for the equation-of-motion coupled-cluster method," J. Chem. Phys. **99**, 8840–8847 (1993).

[16] J. F. Stanton and R. J. Bartlett, "The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties," J. Chem. Phys. **98**, 7029–7039 (1993).

[17] A. I. Krylov, "Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The Hitchhiker's guide to Fock space," Ann. Rev. Phys. Chem. **59**, 433–462 (2008).

[18] N. H. F. Beebe and J. Linderberg, "Simplifications in the generation and transformation of two-electron integrals in molecular calculations," Int. J. Quantum Chem. **12**, 683–705 (1977).

[19] H. Koch, A. Sánchez de Merás, and T. B. Pedersen, "Reduced scaling in electronic structure calculations using Cholesky decompositions," J. Chem. Phys. **118**, 9481–9484 (2003).

[20] F. Coester and H. Kümmel, "Short-range correlations in nuclear wave functions," Nucl. Phys. **17**, 477–485 (1960).

[21] S. D. Folkestad, E. F. Kjønstad, and H. Koch, "An efficient algorithm for Cholesky decomposition of electron repulsion integrals," J. Chem. Phys. **150**, 194112 (2019).

[22] H. Koch, O. Christiansen, P. Jørgensen, A. M. Sanchez de Merás, and T. Helgaker, "The CC3 model: An iterative coupled cluster approach including connected triples," J. Chem. Phys. **106**, 1808–1818 (1997).

[23] R. H. Myhre and H. Koch, "The multilevel CC3 coupled cluster model," J. Chem. Phys. **145**, 044111 (2016).

[24] S. Sæther, T. Kjærgaard, H. Koch, and I.-M. Høyvik, "Density-based multilevel Hartree–Fock model," J. Chem. Theor. Comput. **13**, 5282–5290 (2017).

[25] R. H. Myhre, A. M. J. Sánchez de Merás, and H. Koch, "Multi-level coupled cluster theory," J. Chem. Phys. **141**, 224105 (2014).

[26] S. D. Folkestad and H. Koch, "The multilevel CC2 and CCSD methods with correlated natural transition orbitals," J. Chem. Theor. Comput. **16**, 179–189 (2019).

[27] A. Warshel and M. Karplus, "Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization," J. Am. Chem. Soc. **94**, 5612–5625 (1972).

[28] M. Levitt and A. Warshel, "Computer simulation of protein folding," Nature **253**, 694–698 (1975).

[29] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," Chem. Rev. **105**, 2999–3094 (2005).

[30] B. Mennucci, "Polarizable continuum model," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 386–404 (2012).

[31] E. Valeev, "Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions, Version 2.7.0" (2017), available at http://libint.valeyev.net/.

[32] R. Di Remigio, A. H. Steindal, K. Mozgawa, V. Weijo, H. Cao, and L. Frediani, "PCMSolver: An open-source library for solvation modeling," Int. J. Quantum Chem. **119**, e25685 (2019).

[33] R. Bast (2018). "Runtest," Zenodo. https://doi.org/10.5281/zenodo.1434751

[34] R. Bast, R. Di Remigio, and J. Juselius (2020). "Autocmake," Zenodo. https://doi.org/10.5281/zenodo.3634941

[35] See https://www.openmp.org/spec-html/5.0/openmp.html for technical specifications for OMP.

[36] O. Christiansen, H. Koch, and P. Jørgensen, "The second-order approximate coupled cluster singles and doubles model CC2," Chem. Phys. Lett. **243**, 409–418 (1995).

[37] G. D. Purvis and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," J. Chem. Phys. **76**, 1910–1918 (1982).

[38] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," Chem. Phys. Lett. **157**, 479–483 (1989).

[39] E. R. Davidson, "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices," J. Comput. Phys. **17**, 87–94 (1975).

[40] P. Pulay, "Improved SCF convergence acceleration," J. Comput. Chem. **3**, 556–560 (1982).

[41] P. Pulay, "Convergence acceleration of iterative sequences. the case of SCF iteration," Chem. Phys. Lett. **73**, 393–398 (1980).

[42] G. E. Scuseria, T. J. Lee, and H. F. Schaefer, "Accelerating the convergence of the coupled-cluster approach: The use of the DIIS method," Chem. Phys. Lett. **130**, 236–239 (1986).

[43] C. Hättig and F. Weigend, "CC2 excitation energy calculations on large molecules using the resolution of the identity approximation," J. Chem. Phys. **113**, 5154–5161 (2000).

[44] M. Ziółkowski, V. Weijo, P. Jørgensen, and J. Olsen, "An efficient algorithm for solving nonlinear equations with a minimal number of trial vectors: Applications to atomic-orbital based coupled-cluster theory," J. Chem. Phys. **128**, 204105 (2008).

[45] P. Ettenhuber and P. Jørgensen, "Discarding information from previous iterations in an optimal way to solve the coupled cluster amplitude equations," J. Chem. Theor. Comput. **11**, 1518–1524 (2015).

[46] C. Lanczos, *An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators* (United States Government Press Office, Los Angeles, CA, 1950).

[47] S. Coriani, T. Fransson, O. Christiansen, and P. Norman, "Asymmetric-Lanczos-chain-driven implementation of electronic resonance convergent coupled-cluster linear response theory," J. Chem. Theor. Comput. **8**, 1616–1628 (2012).

[48] H. Koch and P. Jørgensen, "Coupled cluster response functions," J. Chem. Phys. **93**, 3333–3344 (1990).

[49] T. B. Pedersen and S. Kvaal, "Symplectic integration and physical interpretation of time-dependent coupled-cluster theory," J. Chem. Phys. **150**, 144106 (2019).

[50] P. N. Swarztrauber, "FFT algorithms for vector computers," Parallel Comput. **1**, 45–63 (1984).

[51] F. Aquilante, L. Boman, J. Boström, H. Koch, R. Lindh, A. S. de Merás, and T. B. Pedersen, "Cholesky decomposition techniques in electronic structure theory," in *Linear-Scaling Techniques in Computational Chemistry and Physics* (Springer, 2011), pp. 301–343.

[52] J. H. Van Lenthe, R. Zwaans, H. J. J. Van Dam, and M. F. Guest, "Starting SCF calculations by superposition of atomic densities," J. Comput. Chem. **27**, 926–932 (2006).

[53] A. M. J. Sánchez de Merás, H. Koch, I. G. Cuesta, and L. Boman, "Cholesky decomposition-based definition of atomic subsystems in electronic structure calculations," J. Chem. Phys. **132**, 204105 (2010).

[54] P. Pulay, "Localizability of dynamic electron correlation," Chem. Phys. Lett. **100**, 151–154 (1983).

[55] S. Saebø and P. Pulay, "Local treatment of electron correlation," Ann. Rev. Phys. Chem. **44**, 213–236 (1993).

[56] P.-O. Löwdin, "On the nonorthogonality problem," in *Advances in Quantum Chemistry* (Elsevier, 1970), Vol. 5, pp. 185–199.

[57] I.-M. Høyvik, "Convergence acceleration for the multilevel Hartree–Fock model," Mol. Phys. **118**, 1626929 (2020).

[58] R. H. Myhre, A. M. J. Sánches de Merás, and H. Koch, "The extended CC2 model ECC2," Mol. Phys. **111**, 1109–1118 (2013).

[59] I.-M. Høyvik, R. H. Myhre, and H. Koch, "Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models," J. Chem. Phys. **146**, 144109 (2017).

[60] P. Baudin and K. Kristensen, "Correlated natural transition orbital framework for low-scaling excitation energy calculations (CorNFLEx)," J. Chem. Phys. **146**, 214114 (2017).

[61] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems," Angew. Chem., Int. Ed. **48**, 1198–1229 (2009).

[62] C. Cappelli, "Integrated QM/polarizable MM/continuum approaches to model chiroptical properties of strongly interacting solute-solvent systems," Int. J. Quantum Chem. **116**, 1532–1542 (2016).

[63] F. Lipparini, C. Cappelli, and V. Barone, "Linear response theory and electronic transition energies for a fully polarizable QM/classical Hamiltonian," J. Chem. Theory Comput. **8**, 4153–4165 (2012).

[64] R. Di Remigio, T. Giovannini, M. Ambrosetti, C. Cappelli, and L. Frediani, "Fully polarizable QM/fluctuating charge approach to two-photon absorption of aqueous solutions," J. Chem. Theory Comput. **15**, 4056–4068 (2019).

[65] M. Caricato, "Coupled cluster theory with the polarizable continuum model of solvation," Int. J. Quantum Chem. **119**, e25710 (2019).

[66] R. Cammi, "Quantum cluster theory for the polarizable continuum model. I. The CCSD level with analytical first and second derivatives," J. Chem. Phys. **131**, 164104 (2009).

[67] M. Caricato, "Absorption and emission spectra of solvated molecules with the EOM–CCSD–PCM method," J. Chem. Theory Comput. **8**, 4494–4502 (2012).

[68] M. Caricato, F. Lipparini, G. Scalmani, C. Cappelli, and V. Barone, "Vertical electronic excitations in solution with the EOM–CCSD method combined with a polarizable explicit/implicit solvent model," J. Chem. Theory Comput. **9**, 3035–3042 (2013).

[69] S. Ren, F. Lipparini, B. Mennucci, and M. Caricato, "Coupled cluster theory with induced dipole polarizable embedding for ground and excited states," J. Chem. Theory Comput. **15**, 4485–4496 (2019).

[70] L. S. Cederbaum, W. Domcke, and J. Schirmer, "Many-body theory of core holes," Phys. Rev. A **22**, 206–222 (1980).

[71] S. Coriani and H. Koch, "Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework," J. Chem. Phys. **143**, 181103 (2015).

[72] S. Coriani and H. Koch, "Erratum: "Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework" [J. Chem. Phys. 143, 181103 (2015)]," J. Chem. Phys. **145**, 149901 (2016).

[73] J. H. Andersen, A. Balbi, S. Coriani, S. D. Folkestad, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, E. F. Kjønstad, T. Moitra, R. H. Myhre, A. C. Paul, M. Scavino, A. S. Skeidsvoll, Å. H. Tveten, and H. Koch (2020). "Geometries e^T 1.0 paper," Zenodo. https://doi.org/10.5281/zenodo.3666109

[74] O. Christiansen, H. Koch, P. Jørgensen, and J. Olsen, "Excitation energies of H_2O, N_2 and C_2 in full configuration interaction and coupled cluster theory," Chem. Phys. Lett. **256**, 185–194 (1996).

[75] D. Kánnár and P. G. Szalay, "Benchmarking coupled cluster methods on valence singlet excited states," J. Chem. Theor. Comput. **10**, 3757–3765 (2014).

[76] D. Kánnár, A. Tajti, and P. G. Szalay, "Accuracy of coupled cluster excitation energies in diffuse basis sets," J. Chem. Theor. Comput. **13**, 202–209 (2016).

[77] See https://pubchem.ncbi.nlm.nih.gov/compound/Rifampicin for Rifampicin, 2020.

[78] See https://pubchem.ncbi.nlm.nih.gov/compound/Tryptophan for Tryptophan, 2020.

[79] See https://pubchem.ncbi.nlm.nih.gov/compound/Lysergide for Lysergide, 2020.

[80] K. Hald, P. Jørgensen, O. Christiansen, and H. Koch, "Implementation of electronic ground states and singlet and triplet excitation energies in coupled cluster theory with approximate triples corrections," J. Chem. Phys. **116**, 5963–5970 (2002).

[81] S. Li, W. Li, and T. Fang, "An efficient fragment-based approach for predicting the ground-state energies and structures of large molecules," J. Am. Chem. Soc. **127**, 7215–7226 (2005).

[82] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera–A visualization system for exploratory research and analysis," J. Comput. Chem. **25**, 1605–1612 (2004).

[83] S. W. Rick, S. J. Stuart, and B. J. Berne, "Dynamical fluctuating charge force fields: Application to liquid water," J. Chem. Phys. **101**, 6141–6156 (1994).

[84] T. Giovannini, P. Lafiosca, B. Chandramouli, V. Barone, and C. Cappelli, "Effective yet reliable computation of hyperfine coupling constants in solution by a QM/MM approach: Interplay between electrostatics and non-electrostatic effects," J. Chem. Phys. **150**, 124102 (2019).

[85] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "UFF: A full periodic table force field for molecular mechanics and molecular dynamics simulations," J. Am. Chem. Soc. **114**, 10024–10035 (1992).

[86] A. Bondi, "van der Waals volumes and radii," J. Phys. Chem. **68**, 441–451 (1964).

[87] S. A. Kovalenko, R. Schanz, V. M. Farztdinov, H. Hennig, and N. P. Ernsting, "Femtosecond relaxation of photoexcited para-nitroaniline: Solvation, charge transfer, internal conversion and cooling," Chem. Phys. Lett. **323**, 312–322 (2000).

[88] T. Giovannini, R. R. Riso, M. Ambrosetti, A. Puglisi, and C. Cappelli, "Electronic transitions for a fully QM/MM approach based on fluctuating charges and fluctuating dipoles: Linear and corrected linear response regimes," J. Chem. Phys. **151**, 174104 (2019).

[89] T. Giovannini, A. Puglisi, M. Ambrosetti, and C. Cappelli, "Polarizable QM/MM approach with fluctuating charges and fluctuating dipoles: The QM/FQFμ model," J. Chem. Theory Comput. **15**, 2233–2245 (2019).

[90] J. Olsen, P. Jørgensen, H. Koch, A. Balkova, and R. J. Bartlett, "Full configuration–interaction and state of the art correlation calculations on water in a valence double-zeta basis with polarization functions," J. Chem. Phys. **104**, 8007–8015 (1996).

[91] S. D. Folkestad, E. F. Kjønstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, T. Moitra, A. C. Paul, M. Scavino, A. S. Skeidsvoll, Å. H. Tveten, and H. Koch, Gitlab, 2020, https://gitlab.com/eT-program/eT.

Time-dependent coupled-cluster theory for ultrafast transient-absorption spectroscopy

Andreas S. Skeidsvoll [1],* Alice Balbi [2],* and Henrik Koch [1,2],†

[1]*Department of Chemistry, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*
[2]*Scuola Normale Superiore, Piazza dei Cavalieri, 7, IT-56126 Pisa, Italy*

We present a spin-adapted time-dependent coupled-cluster singles and doubles model for the molecular response to a sequence of ultrashort laser pulses. The implementation is used to calculate the electronic response to a valence-exciting pump pulse, and a subsequent core-exciting probe pulse. We assess the accuracy of the integration procedures used in solving the dynamic coupled-cluster equations, in order to find a compromise between computational cost and accuracy. The transient absorption spectrum of lithium fluoride is calculated for various delays of the probe pulse with respect to the pump pulse. We observe that the transient probe absorption oscillates with the pump-probe delay, an effect that is attributed to the interference of states in the pump-induced superposition.

I. INTRODUCTION

Recent advances in the field of ultrafast pulse shaping have enabled the generation of broadband few- to subfemtosecond laser pulses in the near-infrared to vacuum ultraviolet spectral ranges [1–3]. These ultrashort pulses open the possibility to study valence electron dynamics of molecules, on time scales shorter than times characteristic for nuclear dynamics. Also, the generation of intense isolated soft-x-ray free-electron laser pulses with subfemtosecond temporal widths has recently been achieved [4]. This paves the way for attosecond-resolved core-level spectroscopy at high intensities and repetition rates.

Core excitations are typically local to specific atoms, and are sensitive to their electronic environment [5]. The associated attosecond-resolved transient absorption can thus be used to observe superpositions of valence-excited states from the point of view of a specific atomic site, provided that the superposition is of a certain degree of coherence [6]. In the short-pulse limit, the energy-integrated absorption of a core-exciting pulse is indicative of the electronic hole density in the valence region around the nucleus of the specific atom [7,8]. For subfemtosecond pulses outside this limit, the relationship between the pump-induced charge migration and the resultant transient absorption of the probe pulse is more complex. Thus more complete theoretical models are necessary for guiding the pump-probe experiments and for interpreting ensuing results.

Provided that the transient absorption of a probe pulse can be modeled and understood, the valence-level pump and subsequent core-level probe by ultrafast pulses can then be used to investigate the valence electron response of molecules [7,8]. A refined conceptual understanding of this response will shed light on processes occurring in nature, such as photosynthesis and eyesight, and be used for the advance-ment of technological applications, such as photovoltaics and photocatalysis.

Nonperturbative modeling of electron dynamics for ultrafast laser-matter interactions offers certain advantages: the models are applicable for a large range of field intensities [9], and the interaction between a molecule and ultrashort pulses resembles experimental setups in a more natural way.

Electron correlation is often important for a qualitative and quantitative description of many-electron systems. The full configuration interaction (FCI) model is computationally impracticable in most situations [10], and thus we advocate the use of coupled-cluster theory in this paper. Other methods have been used to describe electron dynamics, such as real-time density-functional theory (DFT) [11,12]. However, DFT methods are limited by the accuracy of the exchange-correlation functionals, and thus could lead to misinterpretations [13]. Several implementations of real-time coupled-cluster models have been developed in the past, including approaches based on the time-dependent coupled-cluster (TDCC) equations derived by Koch and Jørgensen [14–17], and approaches based on equation of motion (EOM) theory [9,18–24]. These models offer an accurate description of dynamic correlation, and static correlation in excited states. Needless to say, the coupled-cluster models are also inherently size extensive and intensive [25]. This while keeping the polynomial scaling of the computational costs with respect to system size.

A spin-unrestricted time-dependent coupled-cluster singles and doubles (TDCCSD) model was recently implemented by Pedersen and Kvaal, and used to calculate the absorption spectra of helium and beryllium irradiated by ultrashort pulses at various intensities [26]. Even above the perturbative limit, the TDCCSD spectra show promising correspondence with spectra calculated with time-dependent FCI. The authors also noted that the Lagrangian time-dependent equations have a Hamiltonian structure, well suited for the use of symplectic integrators.

In this work, we will continue the discussion of TDCC models, by presenting a spin-adapted TDCC model of

---

ultrafast transient absorption spectroscopy. Applied to closed-shell molecules interacting with laser pulses within the dipole approximation, this model offers equivalent results as its spin-unrestricted counterparts, with lower computational costs. The reduced cost implies that larger molecules can be studied within this model, making progress towards the accurate modeling of correlated dynamics in interesting photoactive molecules.

This paper is organized as follows. In Sec. II we present the theory underlying the TDCC model and discuss a generalization of the Ehrenfest theorem in this framework. We also describe how absorption spectra are calculated. In Sec. III, we optimize the different parameters used in TDCC calculations, and illustrate this for the LiH molecule. The model is applied to transient absorption of the LiF molecule. Final remarks are given in Sec. IV.

II. THEORY

A. Spin-adapted coupled-cluster method

An accurate account of the electron correlation in molecules is offered by coupled-cluster models, in which the time-independent wave function can be written as

$$|CC\rangle = e^T |HF\rangle, \tag{1}$$

where $|HF\rangle$ is the closed-shell Hartree-Fock reference determinant and T is the spin-adapted cluster operator. The cluster operator is defined as a linear combination of singlet excitation operators τ_μ,

$$T = \sum_{\mu > 0} t_\mu \tau_\mu. \tag{2}$$

The expansion coefficients t_μ are referred to as the amplitudes. The operator T is usually truncated at a given level of excitation, for instance, after single excitations gives the coupled-cluster singles (CCS) model, after double excitations gives the coupled-cluster singles and doubles model (CCSD), and so on.

In the Lagrangian formulation of coupled-cluster theory, which satisfies the Hellman-Feynman theorem, the dual state corresponding to the $|CC\rangle$ state is [27]

$$\langle \Lambda | = \left(\langle HF| + \sum_{\nu > 0} \bar{t}_\nu \langle \nu | \right) e^{-T}, \tag{3}$$

where the linear expansion coefficients $\bar{t}_\nu$ will be referred to as the (Lagrange) multipliers. The level of excitations is truncated at the same level as the excitations in the cluster operator. We note that the $|CC\rangle$ state and its dual state $\langle \Lambda |$ are biorthonormal, $\langle \Lambda | CC\rangle = 1$.

In this formulation, the expectation values of operators are given as

$$\langle A \rangle = \langle \Lambda | A | CC \rangle$$

$$= \left(\langle HF| + \sum_{\nu > 0} \bar{t}_\nu \langle \nu | \right) \bar{A} | HF \rangle, \tag{4}$$

where the similarity transformed operator is defined as

$$\bar{A} = e^{-T} A \, e^T. \tag{5}$$

The amplitudes and multipliers that parametrize the ground state are determined from [28]

$$\langle \mu | \bar{H} | HF \rangle = 0, \tag{6}$$

$$\langle \Lambda | [H, \tau_\mu] | CC \rangle = 0, \tag{7}$$

and the corresponding ground-state energy E_{CC} is given by

$$E_{CC} = \langle \Lambda | H | CC \rangle$$

$$= \langle HF | H | CC \rangle, \tag{8}$$

where we have used Eq. (6) to eliminate the multiplier contribution.

B. Time-dependent coupled-cluster methods

In order to allow for time dependence in the description, the coupled-cluster state is parametrized as [14]

$$|CC(t)\rangle = e^{T(t)} |HF\rangle e^{i\epsilon(t)}, \tag{9}$$

and the corresponding dual state as

$$\langle \Lambda(t)| = \left(\langle HF| + \sum_{\nu > 0} \bar{t}_\nu(t) \langle \nu | \right) e^{-T(t)} e^{-i\epsilon(t)}. \tag{10}$$

The amplitudes t_μ and multipliers $\bar{t}_\mu$ now explicitly depend on time, while the excitation operators τ_μ are still time independent. An overall time-dependent phase $\epsilon(t)$ has also been introduced.

The equation describing the time evolution of the amplitudes $t_\mu(t)$ is obtained from the time-dependent Schrödinger equation for the $|CC\rangle$ state, by projecting onto the corresponding excited determinant $\langle \mu |$. This gives the differential equation

$$\frac{dt_\mu(t)}{dt} = -i \langle \mu | \bar{H}(t) | HF \rangle. \tag{11}$$

The equation describing the time evolution of the multipliers $\bar{t}_\mu(t)$ is obtained by projecting the time-dependent Schrödinger equation for the dual state $\langle \Lambda(t)|$ onto the excited determinants $|\nu\rangle$, giving the differential equation

$$\frac{d\bar{t}_\nu(t)}{dt} = i \left(\langle HF| + \sum_{\mu > 0} \bar{t}_\mu(t) \langle \mu | \right) [\bar{H}(t), \tau_\nu] | HF \rangle. \tag{12}$$

The equation for the phase $\epsilon(t)$ is determined by projection onto the $|HF\rangle$ state

$$\frac{d\epsilon(t)}{dt} = -\langle HF | \bar{H}(t) | HF \rangle. \tag{13}$$

Detailed derivations can be found in Ref. [14]. In this framework, the time-dependent expectation value of a generic operator $A(t)$ is defined as

$$\langle A(t) \rangle = \langle \Lambda(t) | A(t) | CC(t) \rangle, \tag{14}$$

where $\langle \Lambda(t) | CC(t) \rangle = 1$.

C. Generalized Ehrenfest theorem and conserved quantities in TDCC

For ease of notation, we suppress the explicit time dependence in this section. Ideally, observables calculated in

truncated TDCC should have the same properties as in the untruncated case, in order to give a faithful representation of the physical system. In this context, we derive a generalized Ehrenfest theorem for truncated TDCC (the detailed derivation is given in Appendix A). We obtain the equation

$$\frac{d}{dt}\langle \Lambda'|A|\mathrm{CC}\rangle = i\langle \Lambda'|H\,e^{T'}P_n e^{-T'}A|\mathrm{CC}\rangle$$

$$- i\langle \Lambda'|A\,e^{T}P_n e^{-T}H|\mathrm{CC}\rangle + \langle \Lambda'|\frac{\partial A}{\partial t}|\mathrm{CC}\rangle, \tag{15}$$

where the left $\langle \Lambda'|$ state and the right $|\mathrm{CC}\rangle$ state are independent solutions to the projected time-dependent Schrödinger equation. The projection operator P_n of maximum excitation level n is defined as

$$P_n = |\mathrm{HF}\rangle\langle \mathrm{HF}| + \sum_{\mu>0}^{n} |\mu\rangle\langle \mu|, \tag{16}$$

and in untruncated TDCC, $P_n = \mathbb{1}$. From Eq. (15) we can see that, in untruncated TDCC,

$$\frac{d}{dt}\langle \Lambda'|A|\mathrm{CC}\rangle = i\langle \Lambda'|[H,A]|\mathrm{CC}\rangle + \langle \Lambda'|\frac{\partial A}{\partial t}|\mathrm{CC}\rangle, \tag{17}$$

regardless of the initial values of the amplitudes, multipliers, and phases.

In truncated TDCC, the projection operator cannot in general be replaced by the identity operator, and hence Eq. (15) cannot be simplified further. Still, some conservation laws from untruncated TDCC apply under certain constraints: we see from (15) that the Hamiltonian matrix element $\langle \Lambda'|H|\mathrm{CC}\rangle$ is conserved for a time-independent Hamiltonian operator as long as $T' = T$, regardless of the initial values of the multipliers and phases. The overlap matrix element $\langle \Lambda'|\mathbb{1}|\mathrm{CC}\rangle$ is also conserved for $T' = T$, since $\exp(T)P_n \exp(-T)\mathbb{1}|\mathrm{CC}\rangle = |\mathrm{CC}\rangle$ and $\langle \Lambda'|_{T'=T}\mathbb{1}\exp(T)P_n\exp(-T) = \langle \Lambda'|_{T'=T}$. In conclusion, we note the energy and overlap conservation for a time-independent Hamiltonian in untruncated TDCC, and in truncated TDCC for $T' = T$.

D. Interaction with an external electromagnetic field

In the semiclassical approximation, the electronic Hamiltonian for a molecule interacting with an external electromagnetic field can be written as

$$H(t) = H_0 + V(t), \tag{18}$$

where H_0 is the time-independent electronic Hamiltonian and $V(t)$ is the operator describing the interaction with the external field. We choose to express the interaction in the length gauge and dipole approximation, meaning that the electromagnetic field is represented by an electric field,

$$V(t) = -\mathbf{d} \cdot \boldsymbol{\mathcal{E}}(t), \tag{19}$$

where $\mathbf{d}$ is the electric dipole moment operator. Since this operator is a one-electron operator, it can also be expressed in terms of the molecular-orbital (MO) dipole moment integrals d_{pq} and one-electron singlet excitation operators E_{pq},

$$\mathbf{d} = \sum_{pq} \mathbf{d}_{pq} E_{pq}. \tag{20}$$

Since electric fields are additive, the external electric field $\boldsymbol{\mathcal{E}}(t)$ can be written as a linear combination of individual laser pulses,

$$\boldsymbol{\mathcal{E}}(t) = \sum_{n} \boldsymbol{\mathcal{E}}_{0n} \cos[\omega_{0n}(t - t_{0n})]f_n(t), \tag{21}$$

where $\boldsymbol{\mathcal{E}}_{0n}$ is the peak electric field of pulse n in its polarization direction, ω_{0n} the carrier (angular) frequency and t_{0n} the central time of the pulse, and $f_n(t)$ an envelope function that determines its shape. A commonly used family of envelopes $f_n(t)$, that resemble physical laser intensity profiles, are the Gaussian functions. Since Gaussian functions have infinite support, we choose to set them to zero at a finite number N of root-mean-square (rms) widths σ_n outside the central time, i.e.,

$$f_n(t) = \begin{cases} e^{-(t-t_{0n})^2/(2\sigma_n^2)}, & a_n \leqslant t \leqslant b_n, \\ 0, & \text{otherwise}, \end{cases} \tag{22}$$

where $a_n = t_{0n} - N\sigma_n$ and $b_n = t_{0n} + N\sigma_n$. In addition to resembling physical intensity profiles, a useful feature of Gaussian envelopes is that they give pulses with Gaussian frequency distributions. Hence these pulses can offer a good compromise between temporal precision and spectral narrowness. This is useful for producing temporally precise electronic transitions within the molecule, while keeping the probability of ionization low.

E. Frequency-resolved transient absorption

Following the procedure of [29], the energy absorbed during the interaction with the external electromagnetic field can be given by

$$\Delta E = \int_{-\infty}^{\infty} \frac{dE(t)}{dt} dt. \tag{23}$$

The time derivative of the expectation value of the Hamiltonian in Eq. (18) can be found through Eq. (15):

$$\frac{dE(t)}{dt} = \frac{d}{dt}\langle \Lambda(t)|H(t)|\mathrm{CC}(t)\rangle$$

$$= \langle \Lambda(t)|\frac{\partial H(t)}{\partial t}|\mathrm{CC}(t)\rangle = -\mathbf{d}(t) \cdot \frac{\partial \boldsymbol{\mathcal{E}}(t)}{\partial t}, \tag{24}$$

where the TDCC dipole moment expectation value is given by

$$\mathbf{d}(t) = \langle \Lambda(t)|\mathbf{d}|\mathrm{CC}(t)\rangle. \tag{25}$$

The energy exchanged between the electromagnetic field and the molecule is thus given by

$$\Delta E = -\int_{-\infty}^{\infty} \mathbf{d}(t) \cdot \frac{\partial \boldsymbol{\mathcal{E}}(t)}{\partial t} dt. \tag{26}$$

Equation (26) can be frequency resolved by inserting the relations between the components $d_i(t)$ and $\mathcal{E}_i(t)$ and their Fourier transforms, $\tilde{d}_i(\omega)$ and $\tilde{\mathcal{E}}_i(\omega)$. We use the following convention:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(\omega)e^{i\omega t} d\omega, \tag{27}$$

$$\tilde{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt. \tag{28}$$

After inserting the relations, the expression

$$\Delta E = \int_0^\infty \omega S(\omega) d\omega \tag{29}$$

is obtained, where

$$S(\omega) = -2 \, \mathrm{Im}[\widetilde{\mathbf{d}}(\omega) \cdot \widetilde{\boldsymbol{\mathcal{E}}}^*(\omega)], \quad \omega > 0. \tag{30}$$

The response function $S(\omega)$ has the opposite sign as in [29], due to different Fourier transform conventions. It represents the absorption per unit frequency at a given frequency, so that positive (negative) $\omega S(\omega)$ equals the amount of energy gained (lost) by the molecule per unit frequency at ω [29].

The TDCC dipole moment $\mathbf{d}(t)$ can be found from Eq. (14)

$$\begin{aligned}
\mathbf{d}(t) &= \sum_{pq} \langle \Lambda(t) | E_{pq} | \mathrm{CC}(t) \rangle \mathbf{d}_{pq} \\
&= \left(\langle \mathrm{HF} | + \sum_{\mu > 0} \bar{t}_\mu(t) \langle \mu | \right) \bar{E}_{pq}(t) | \mathrm{HF} \rangle \mathrm{d}_{pq} \\
&= \sum_{pq} D_{pq}(t) \mathrm{d}_{pq},
\end{aligned} \tag{31}$$

where $D_{pq}(t)$ is an element of the standard coupled-cluster one-electron density matrix, which can be calculated given the time-dependent amplitudes and multipliers.

F. Initial value problem

In order to calculate the time-dependent amplitudes and multipliers for the system represented by the Hamiltonian in Eq. (18), the system is prepared in the ground state at $t = -T$ (before the interaction). The time-dependent amplitudes and multipliers are then propagated by integration of Eqs. (11) and (12), until $t = T$ (after the interaction). This is done using Runge-Kutta methods (a general introduction to these methods is given in Appendix B). Once the time-dependent amplitudes and multipliers are calculated, they can be used to calculate evenly sampled values of the TDCC dipole moment with Eq. (31).

The main Runge-Kutta method used for integration is the explicit Runge-Kutta (ERK) method known as RK4, and referred to as "the best-known fourth-order four-stage ERK method" in [30]. In many cases, this method gives a good compromise between accuracy and the number of evaluations for each time step.

The performance of two methods in the family of ν-stage 2νth-order implicit Runge-Kutta (IRK) methods, known as Gauss-Legendre methods, is also assessed. An interesting property of these methods is that they are symplectic, meaning that they often perform well with regards to preserving the energy expectation value of noninteracting Hamiltonian systems. The application of these methods to TDCC methods was discussed in greater detail in the work by Pedersen and Kvaal [26]. The Gauss-Legendre methods that will be considered here are the two-stage fourth-order Gauss-Legendre method (GL4) and the three-stage sixth-order Gauss-Legendre method (GL6).

G. Discrete Fourier transformation of TDCC dipole moment and electric field

After the dipole moment and electric field have been calculated in $[-T, T]$, a discrete approximation of $\widetilde{d}_i(\omega)$ and $\widetilde{\mathcal{E}}_i(\omega)$ can be found from doing the discrete Fourier transform of the time series.

Assuming that the finite and discrete time series are sampled from infinitely extending analytic dipole moment and electric-field functions, the time series can equally be represented as the analytic functions modulated by the rectangular window function,

$$f_{w_R}(t) = f(t) w_R(t), \tag{32}$$

sampled in $[-T, T]$, where the rectangular window function

$$w_R(t) = \begin{cases} 1, & |t| \leqslant T, \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

Since the Fourier transform of a windowed function is equal to the convolution of the Fourier transform of the function with the Fourier transform of the window function [31],

$$\widetilde{f}_w(\omega) = \widetilde{f}(\omega) * \widetilde{w}(\omega), \tag{34}$$

the spectral leakage of the peaks in the finite Fourier spectrum will be related to the Fourier transform of the rectangular window function. In order to reduce the intensity of sidelobes of peaks in the Fourier spectrum [31], the rectangular window can be replaced with a Hann window, by multiplying the sampled values with the Hann function,

$$w_H(t) = \cos^2\left(\frac{\pi t}{T}\right), \tag{35}$$

before doing the discrete Fourier transform.

III. RESULTS AND DISCUSSION

A. Convergence of LiH pump-probe absorption spectra

In the following, we investigate the convergence properties of the spin-adapted TDCC model of molecular ultrafast pump-probe absorption. The convergence will be assessed with respect to the individual variation of several parameters: the basis set, the size of the time steps, and the integration method. The TDCC method was implemented in the recently released e^T program [32]. This program is used for all reported computations. Unit conversions are done from Hartree atomic units using the 2018 CODATA recommended values [33]. All reported calculations are run on a two-socked node equipped with Intel Xeon-Gold 615 22.1 GHz processors and 1.5 TB of memory.

The higher level coupled-cluster methods scale rapidly with the size of the system, and quickly reach the limits of practicability. Therefore, we have chosen lithium hydride (LiH) for the convergence studies. This serves as an elementary example of a closed-shell molecule with atoms of different core excitation energies. The electronic charge can migrate between the two atoms, making it an interesting case for examination by pump-probe spectroscopy.

The lithium atom is placed at the origin, and the hydrogen atom at $-1.594\,913\,18$ Å along the z axis, corresponding to the experimentally measured equilibrium bond length of LiH

TABLE I. LiH and LiF pump and probe pulse parameters. Gaussian rms width σ, carrier frequency ω_0, peak electric-field strength $|\mathcal{E}_0|$, and the nonperturbative intensity parameter $z_0 = |\mathcal{E}_0|^2/(4\omega_0^3)$ [34] for the carrier waves. A Gaussian rms width of 20 a.u. corresponds to a field strength (intensity) full width at half maximum of 1.139 fs (805.5 as) and 10 a.u to a full width at half maximum of 569.6 as (402.8 as). From the relation $S_0 = |\mathcal{E}_0|^2/Z_0$ (SI units), where Z_0 is the impedance of free space, a peak electric-field strength of 0.01 a.u. corresponds to a peak intensity S_0 of 7.019×10^{12} W/cm^2, and 0.1 a.u. to a peak intensity of 7.019×10^{14} W/cm^2. The perturbation limit can be taken to be the intensity where $z_0 = 1$ for a given carrier frequency [34]. Note that $z_0 \ll 1$ for all pulses, indicating that the interactions also could be described with perturbative approaches.

	σ (a.u.)	ω_0 (eV)	$|\mathcal{E}_0|$ (a.u.)	z_0
LiH pump	20	3.552 47	0.01	1.12×10^{-2}
LiH probe	10	57.6527	0.1	2.63×10^{-4}
LiF pump	20	6.448 01	0.01	1.88×10^{-3}
LiF probe	10	688.018	0.1	1.55×10^{-7}

[35]. Gaussian envelopes are used for the pump and probe pulses, which are polarized in the z direction. The electric fields of each pulse are temporally truncated at eight rms widths σ from the central time, and thus nonzero only inside this interval [see Eq. (22)]. The carrier frequency of the pump pulse corresponds to the first LiH valence excitation energy and the carrier frequency of the probe pulse to the first LiH K-edge excitation energy. These excitation energies are calculated using EOM-CCSD. The core excitations are obtained within the core-valence separation (CVS) approximation [36]. The parameters of the pulses are shown in Table I. As the Gaussian envelopes give the pulses a frequency content distributed around the central frequencies, the pump and probe pulses will induce excitations in the valence and core regions, respectively.

The pump pulse is given a central time of $t = -40$ a.u. and the probe a central time of $t = 0$ a.u. The time-dependent dipole moment and electric field are calculated every 0.1 a.u. in the $[-5000$ a.u., 5000 a.u.$]$ interval. Since the system remains in the ground state until the onset of the truncated pump pulse—with the ground-state dipole moment—the interaction with the pulses only needs to be calculated in $[-200$ a.u., 5000 a.u.$]$. Subsequently, the Hann windowed components of the dipole moment and electric field are discrete Fourier transformed, and the transient absorption is calculated using Eq. (30).

We use the correlation-consistent basis sets of Dunning et al. (cc-pVXZ, X = D, T) [37] that are suitable for describing valence correlation effects in molecules. In some of the calculations, the basis sets are augmented by diffuse functions (denoted by aug-) and/or functions describing core correlation (denoted by C) [38]. From now on, we will use a C in round brackets to indicate that core correlation functions are added to the basis set of the heaviest atom in the molecule.

The individual variation of the calculation parameters is done with respect to a common reference: TDCCSD/ aug-cc-p(C)VDZ, and integrated with RK4 with 0.005 a.u. time steps. This basis set gives two occupied and 34 virtual
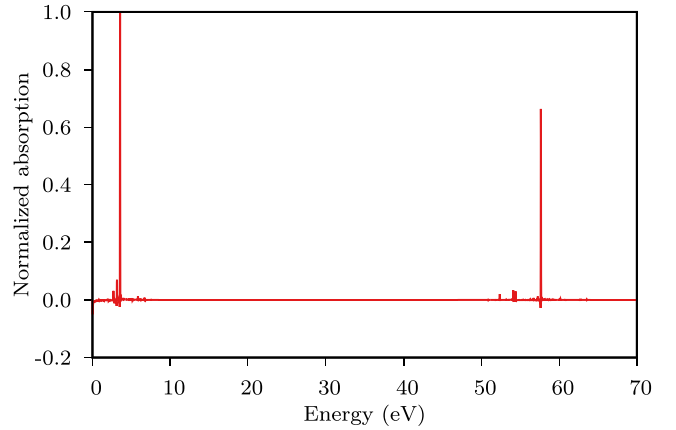

FIG. 1. Normalized reference LiH pump-probe absorption, $S'(\omega)$, as a function of energy. The time-dependent dipole moment is calculated using TDCCSD/aug-cc-p(C)VDZ, and integrated with RK4 with 0.005 a.u. time steps.

MOs, and hence 4828 time-dependent parameters. The reported calculation uses around 1 GB of memory and 0.12 s wall time per time step on eight cores, and 0.15 s per time step on four cores. The unnormalized reference absorption $S_{\mathrm{ref}}(\omega)$ is used to calculate the normalization factor

$$\mathcal{N}_{\mathrm{ref}} = \frac{1}{\max_\omega |S_{\mathrm{ref}}(\omega)|}. \tag{36}$$

This factor is used to normalize all the absorption spectra of the following LiH calculations, by means of

$$S'(\omega) = \mathcal{N}_{\mathrm{ref}} S(\omega), \tag{37}$$

where $S(\omega)$ is calculated with the parameters in question. The normalized deviation of $S'(\omega)$ from a more accurate result $S'_{\mathrm{acc}}(\omega)$ is calculated as

$$D'(\omega) = |S'(\omega) - S'_{\mathrm{acc}}(\omega)|. \tag{38}$$

The reference absorption spectrum, normalized according to Eq. (37), is shown in Fig. 1. We observe absorption in two energy regions: one corresponding to the valence-exciting pump pulse and the other to the core-exciting probe pulse.

1. TDCCS and TDCCSD

In Fig. 2, the normalized reference TDCCSD spectrum is shown together with the normalized time-dependent CCS (TDCCS) spectrum. The two spectra display substantial differences in intensities and positions of the peaks in both the pump and the probe absorption regions. Since TDCCSD includes double excitations, while TDCCS does not, this demonstrates that higher-order excitations are needed to obtain qualitatively correct results for the LiH model system.

2. Basis set

In Fig. 3, the normalized reference spectrum is shown together with normalized spectra calculated using cc-pVDZ, cc-p(C)VDZ, and aug-cc-pVDZ. The inclusion of diffuse functions in the basis sets seems important for representing the dynamics properly. Increasing the basis set from cc-pVDZ to aug-cc-pVDZ shifts the peaks in both the pump and the probe
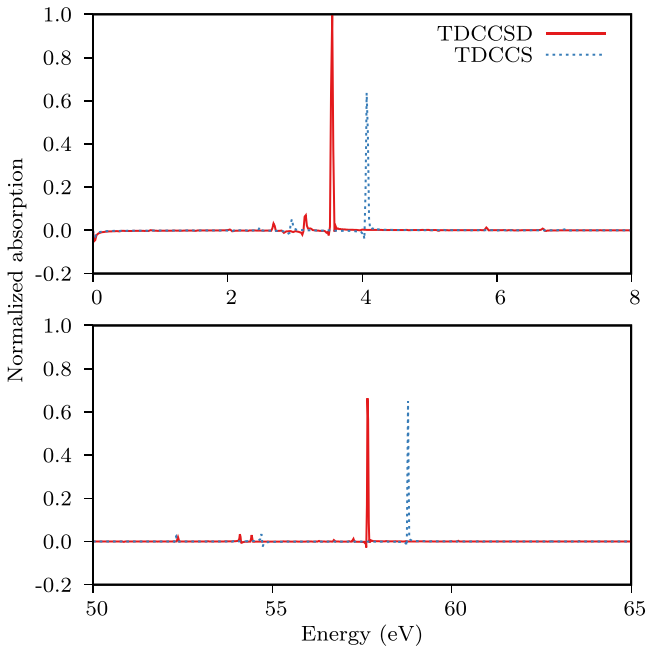
FIG. 2. Normalized TDCCSD and TDCCS LiH pump and probe absorption, $S'(\omega)$, as a function of energy. Time-dependent dipole moments are calculated using aug-cc-p(C)VDZ, and integrated with RK4 with 0.005 a.u. time steps.
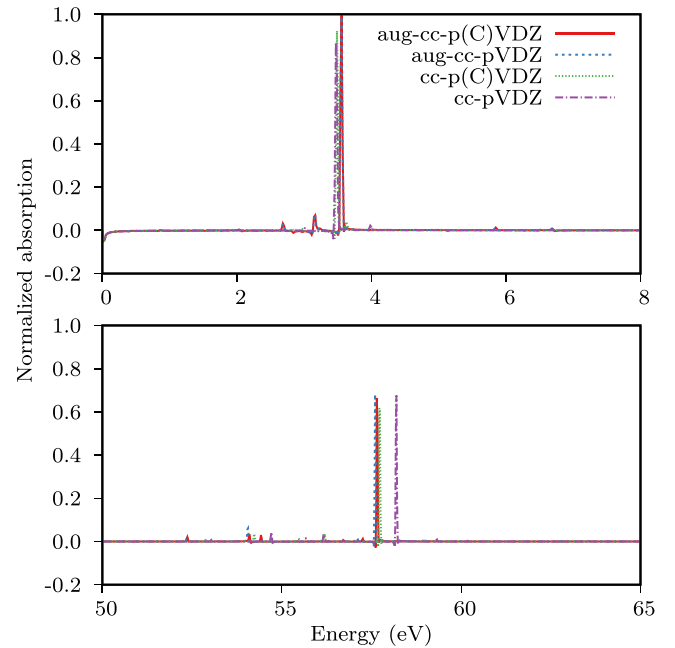

FIG. 3. Normalized aug-cc-p(C)VDZ, aug-cc-pVDZ, cc-p(C)VDZ, and cc-pVDZ LiH pump and probe absorption, $S'(\omega)$, as a function of energy. Time-dependent dipole moments are calculated using TDCCSD, and integrated with RK4 with 0.005 a.u. time steps.

absorption regions. This is consistent with the concept of the pump pulse forcing electrons to the outer valence regions of the molecule, which is better represented with diffuse functions.

Furthermore, comparing cc-p(C)VDZ and cc-pVDZ spectra in Fig. 3, we see the importance of the added core correlation functions. As expected, they cause a substantial shift in the probe absorption peaks, while they are not important for the pump absorption.

We also performed calculations with cc-pVTZ, cc-p(C)VTZ, aug-cc-pVTZ, and aug-cc-p(C)VTZ basis sets. Note that, for the aug-cc-pVTZ and aug-cc-p(C)VTZ spectra, the time-dependent dipole moments are only calculated in the $[-2500 \text{ a.u.}, 2500 \text{ a.u.}]$ interval, in order to reduce computational time. Thus these spectra have a lower resolution than the others. The normalized spectra are shown together with the normalized reference spectrum in Fig. 4. Here we observe that triple zeta functions change the position of the peaks in the probe absorption region. This indicates that basis sets larger than aug-cc-p(C)VDZ should be used if precise peak positions are required, bringing about a substantial increase in the computational costs. The aug-cc-p(C)VDZ basis set is used as the reference for the other LiH calculations, as the larger basis sets are too computationally expensive for practical purposes.

Note that the pulses are not strong enough to induce considerable multiphoton absorption (see Table I). The electrons should thus primarily be confined to low angular momentum bound states, which are fairly well described with the aforementioned basis sets. At higher intensities, the results obtained with these basis sets should deviate further from the complete basis set limit, as the representation of Rydberg

states and the continuum will be more important [39]. The results can then be improved by adding suitable functions to the basis set, for instance, Gaussians optimized for the representation of the continuum [39,40].

3. Integration

We calculated normalized spectra for 0.125 a.u., 0.025 a.u., and 0.001 a.u. time steps. The deviations from the 0.001 a.u. time step are calculated according to Eq. (38). The results are shown in Fig. 5. The deviations decrease with the time step size, indicating that the spectra approach a time step limit.

We further calculated normalized spectra with GL4 and GL6. The deviations of the RK4 (reference) and GL4 spectra from the GL6 spectrum are shown in Fig. 6. Although the TDCC equations have a Hamiltonian structure, the use of symplectic integrators does not seem to be necessary to calculate accurate spectra for this system, with the applied field strength. As the three integration methods give comparable results, we will use RK4 for the other calculations, as this generally requires fewer evaluations of the TDCC equations per time step.

B. LiF transient absorption

In this section, variations in molecular absorption caused by ultrafast charge migration are modeled in the described pump-probe framework. We consider the lithium fluoride (LiF) molecule, where the fluorine atom is placed at the origin and the lithium atom at $-1.563\,864\,13$ Å along the z axis. This corresponds to the experimentally measured equilibrium bond length of LiF [35]. In order to classify some of the transitions
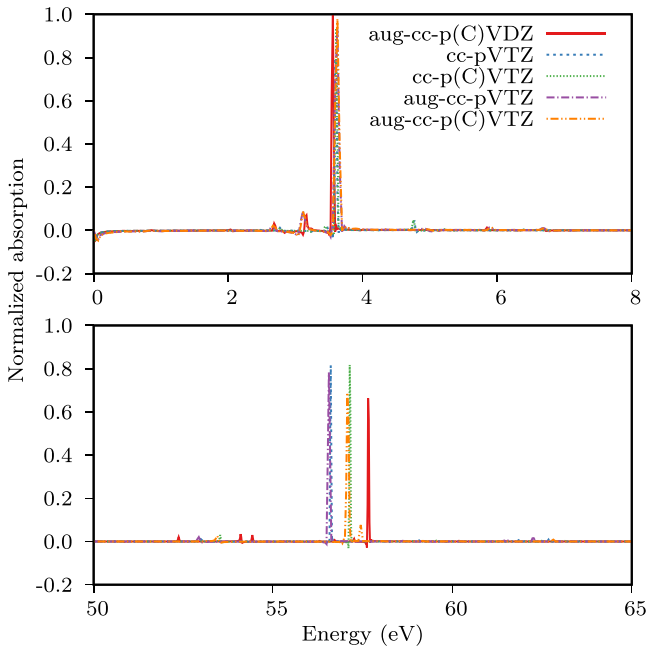
FIG. 4. Normalized aug-cc-p(C)VDZ, cc-pVTZ, cc-p(C)VTZ, aug-cc-pVTZ, and aug-cc-p(C)VTZ LiH pump-probe absorption, $S'(\omega)$, as a function of energy. Time-dependent dipole moments are calculated using TDCCSD, and integrated with RK4 with 0.005 a.u. time steps.

involved in the molecular absorption, the first eight valence-excited and the first eight core-excited states are calculated using EOM-CCSD/aug-cc-p(C)VDZ. The core excitations are obtained within the CVS approximation. The molecular term symbols and excitation energies are given in Table II.

In the TDCC calculations, all probe pulses are z polarized, and have carrier frequencies corresponding to the first LiF core excitation energy (see Table II). Central times are chosen to be 0 a.u., to minimize the effect of the windowing on
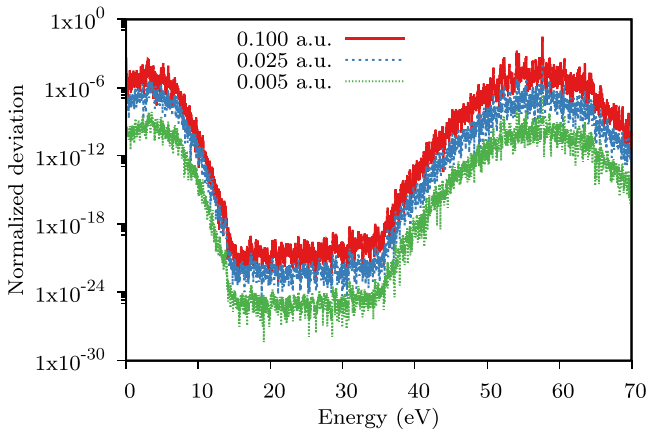
FIG. 5. LiH pump-probe absorption. Normalized deviation of the 0.100 a.u., 0.025 a.u., and 0.005 a.u. time step spectra from the 0.001 a.u. time step spectrum, $D'(\omega)$, as a function of energy. Time-dependent dipole moments are calculated using TDCCSD/aug-cc-p(C)VDZ, and integrated with RK4.
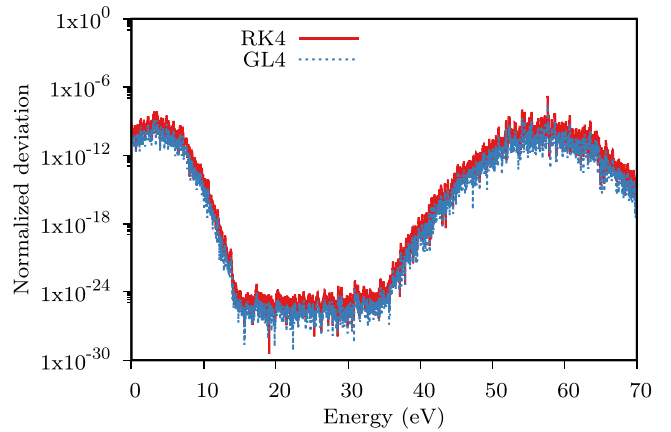
FIG. 6. LiH pump-probe absorption. Normalized deviation of the RK4 and GL4 spectra from the GL6 spectrum, $D'(\omega)$, as a function of energy. Time-dependent dipole moments are calculated using TDCCSD/aug-cc-p(C)VDZ, and integrated with 0.005 a.u. time steps.

the probe absorption. The pump pulses are also z polarized, and have carrier frequencies corresponding to the first LiF valence excitation energy (see Table II). The pump pulses have different central times with respect to the probe pulses, corresponding to probe delays from 0 a.u. to 240 a.u., in increments of 5 a.u. Other parameters of the pump and probe pulses are given in Table I. As for the LiH calculations, the electric fields of each pulse are temporally truncated at eight rms widths σ from the central time, and thus nonzero only inside this interval [see Eq. (22)].

The parameters used for the LiH reference calculation offered a compromise between computational cost and accuracy. For pragmatic reasons, we also use the parameters for all LiF calculations. The calculations in this section are thus done using TDCCSD/aug-cc-p(C)VDZ, and integrated with RK4 with 0.005 a.u. time steps. The basis set gives six occupied and 44 virtual MOs, and hence 70488 time-dependent parameters. The time-dependent dipole moments and electric fields are calculated every 0.1 a.u. in the [−5000 a.u., 5000 a.u.] interval, where the external field interactions are only calculated after the onset of the temporally truncated pump pulses. The

TABLE II. Molecular term symbols and ground-state excitation energies ΔE of some excited states of LiF, calculated with the EOM-CCSD method. Valence-excited states are denoted by a subscript v. Core-excited states, calculated within the CVS approximation, are denoted by a subscript c.

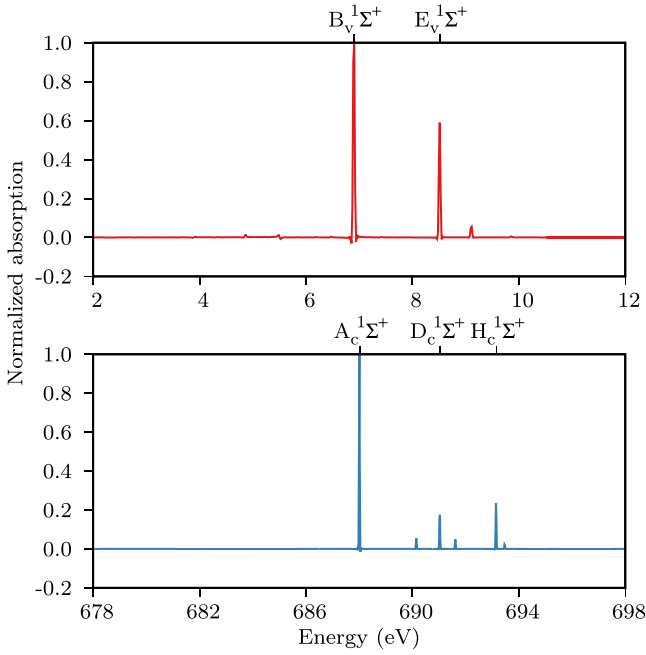| State	ΔE (eV)	State	ΔE (eV)
$A_v\ {}^1\Pi$	6.448 01	$A_c\ {}^1\Sigma^+$	688.018
$B_v\ {}^1\Sigma^+$	6.899 82	$B_c\ {}^1\Pi$	689.462
$C_v\ {}^1\Delta$	8.104 63	$C_c\ {}^1\Sigma^+$	690.159
$D_v\ {}^1\Sigma^-$	8.140 74	$D_c\ {}^1\Sigma^+$	691.039
$E_v\ {}^1\Sigma^+$	8.511 16	$E_c\ {}^1\Pi$	691.435
$F_v\ {}^1\Pi$	8.589 43	$F_c\ {}^1\Sigma^+$	691.625
$G_v\ {}^1\Pi$	8.625 89	$G_c\ {}^1\Pi$	692.917
$H_v\ {}^1\Sigma^+$	9.106 55	$H_c\ {}^1\Sigma^+$	693.154

FIG. 7. Normalized LiF pump and probe absorption, $S'_{\text{pump}}(\omega)$ (top) and $S'_{\text{probe}}(\omega)$ (bottom), as a function of energy. The most dominant peaks are identified with ground-state transitions to EOM-CCSD valence- and core-excited states. Time-dependent dipole moments are calculated with TDCCSD/aug-cc-p(C)VDZ, integrated with RK4 with 0.005 a.u. time steps.

reported calculations use around 2 GB of memory and 0.41 s wall time per time step on 16 cores, and 0.50 s per time step on eight cores.

In order to assess the relative occupation of the states in the pump-induced superposition [see Eq. (29)], the normalized absorption of the pump pulse, centered at 0 a.u., is calculated using

$$S'_{\text{pump}}(\omega) = \mathcal{N}_{\text{pump}} S_{\text{pump}}(\omega), \qquad (39)$$

where

$$\mathcal{N}_{\text{pump}} = \frac{1}{\max_{\omega} |S_{\text{pump}}(\omega)|}. \qquad (40)$$

An analogous procedure is used to obtain the normalized probe spectrum $S'_{\text{probe}}(\omega)$.

The normalized absorption of the pump pulse and of the probe pulse are plotted in Fig. 7, where the most dominant absorption peaks are identified using the calculated EOM-CCSD states (see Table II). The small pump absorption peaks that lie below the ground-state valence excitation energy gap are presumably caused by two-photon absorption. The positions of the other visible peaks in the two spectra fit well with single-photon EOM-CCSD transitions allowed by symmetry.

The pump-probe absorption $S(\omega, \tau)$ is calculated as a function of the energy, ω, and the delay of the probe pulse with respect to the pump pulse, τ. In order to directly assess the change in absorption caused by the interaction with the pump pulse, the normalized transient absorption

$$\Delta S'(\omega, \tau) = \mathcal{N}_{\text{probe}} \Delta S(\omega, \tau)$$
$$= \mathcal{N}_{\text{probe}}[S(\omega, \tau) - S_{\text{probe}}(\omega)] \qquad (41)$$

is calculated for all delays, where $\mathcal{N}_{\text{probe}}$ is the normalization factor for the probe spectrum. The normalized transient absorption in the probe absorption region is shown in Fig. 8. The spectrum features several constant energy peaks that oscillate with the pump-probe delay. The five peaks that oscillate the
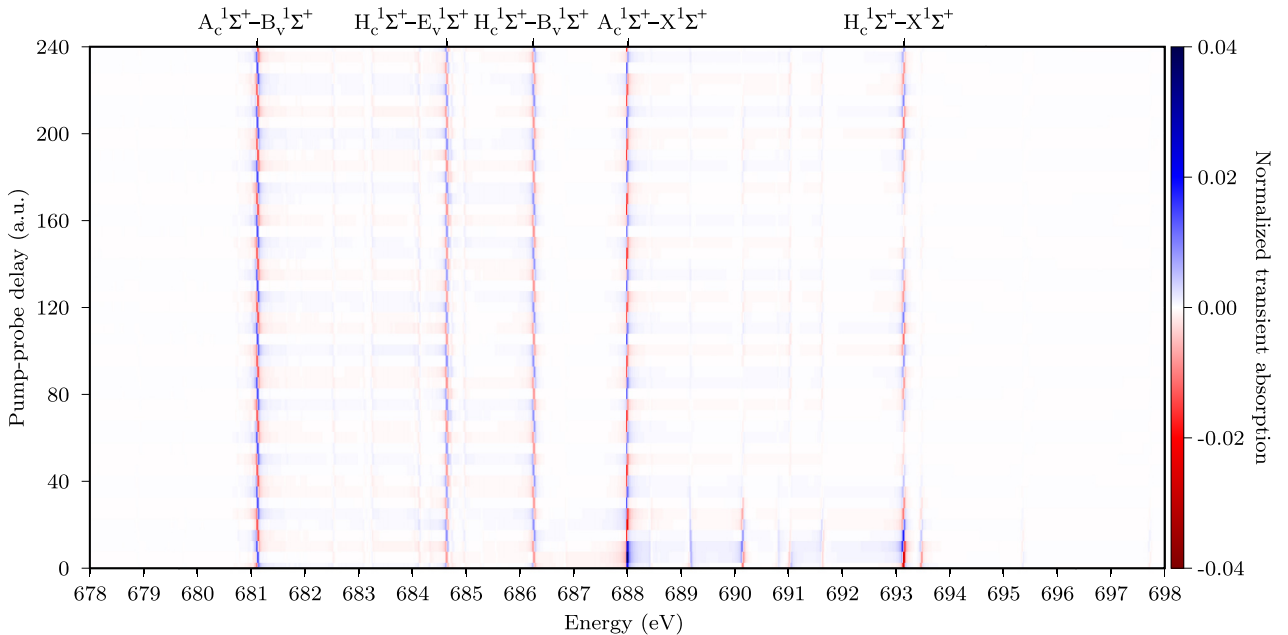

FIG. 8. Normalized LiF transient absorption $\Delta S'(\omega, \tau)$, as a function of energy and pump-probe delay. The five peaks oscillating with the largest amplitude are identified with EOM-CCSD transitions. Time-dependent dipole moments are calculated using TDCCSD/aug-cc-p(C)VDZ basis set, integrated with RK4 with 0.005 a.u. time steps.
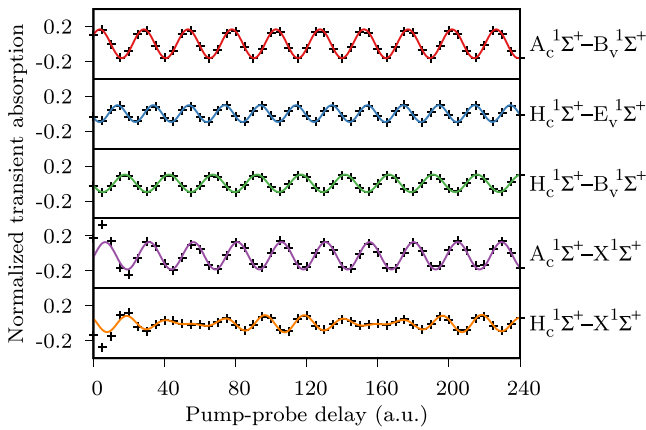
FIG. 9. Normalized LiF transient absorption $\Delta S'(\omega, \tau)$ (black crosses) as a function of pump-probe delay, given at the discrete Fourier transform energies closest to the energies of the transitions shown to the right. The colored functions in the four topmost panels are found from least-squares fitting $A \sin(\omega_A t + \phi_A) + C$, with fixed values of ω_A, to the absorption, in the domain [40 a.u., 240 a.u.]. The values of ω_A are 6.899 82 eV (red) 8.511 16 eV (blue), 6.899 82 eV (green), and 6.899 82 eV (purple). The orange function in the bottom panel is found from least-squares fitting $A \sin(\omega_A t + \phi_A) + B \sin(\omega_B t + \phi_B) + C$, with $\omega_A = $ 6.899 82 eV and $\omega_B = $ 8.511 16 eV, to the absorption, in the domain [40 a.u., 240 a.u.].

most with respect to the pump-probe delay are identified using the states in Table II. Note that, for pump-probe delays shorter than about 40 a.u., the oscillations of some of the peaks are rapidly damped as a function of increasing delays. This effect can be attributed to the decreasing overlap between the pump and probe pulses. For pump-probe delays longer than about 40 a.u., where the overlap of the pulses can be neglected, the damping of the oscillations is also negligible.

We note that the excitation by the pump pulse enables new transitions in the probe absorption region. An illustrative example is the oscillating peak at around 681.1 eV in Fig. 8. The energy corresponding to this peak is lower than the lowest ground-state core excitation energy of 688.018 eV. This peak is identified as the $A_c \, ^1\Sigma^+ - B_v \, ^1\Sigma^+$ transition. Its occurrence indicates that the pump has generated an electronic hole in a previously occupied region of the molecule, allowing a lower-energy core excitation to take place.

In Fig. 9, the normalized transient absorption of the five peaks identified in Fig. 8 are plotted at the nearest discrete Fourier-transform energies. Two of these peaks describe transitions involving the $A_c \, ^1\Sigma^+$ state. Beyond the pump-probe overlap region, the oscillations of these peaks correlate with the quantum interference of the two probed states, as expected for the ultrafast high-energy probing of two states in a coherent superposition [6]. This since both oscillations can be fitted with sinusoids with the frequency corresponding to the $B_v \, ^1\Sigma^+$ and $X \, ^1\Sigma^+$ energy difference.

Three peaks in Fig. 9 correspond to transitions involving the $H_c \, ^1\Sigma^+$ state. The oscillation of the $H_c \, ^1\Sigma^+ - B_v \, ^1\Sigma^+$ peak correlates well with the quantum interference of the $B_v \, ^1\Sigma^+$ and $X \, ^1\Sigma^+$ states, as the oscillations are well fitted with a sinusoid with the frequency corresponding to the energy

difference of these two states. Similarly, the oscillation of the $H_c \, ^1\Sigma^+ - E_v \, ^1\Sigma^+$ peak correlates with the quantum interference of the $E_v \, ^1\Sigma^+$ and $X \, ^1\Sigma^+$ states. Note that the oscillations of the two peaks are slightly phase shifted with respect to each other, an effect that may be caused by the difference in spectral phase of the two corresponding frequencies in the probe pulse.

The linear combination of two sinusoids is needed to give a good fit with the oscillation of the $H_c \, ^1\Sigma^+ - X \, ^1\Sigma^+$ peak: one corresponding to the $B_v \, ^1\Sigma^+$ and $X \, ^1\Sigma^+$ energy difference and the other corresponding to the $E_v \, ^1\Sigma^+$ and $X \, ^1\Sigma^+$ energy difference. Hence the ground state $X \, ^1\Sigma^+$ seems to have a similar probability of interfering with the $B_v \, ^1\Sigma^+$ and $E_v \, ^1\Sigma^+$ states. This is reasonable, considering that most of the population will be left in the ground state after the interaction with the pump pulse.

IV. CONCLUSION

In this work, a time-dependent coupled-cluster model of ultrafast pump-probe absorption spectroscopy has been presented. First, we investigated the convergence of LiH pump-probe absorption spectra with respect to different calculation parameters. The deviations related to the integration parameters (integration method and time step size) were small in comparison to other parameter-dependent deviations. As the computational costs scaled linearly with the time step size, we chose a time step size that gave a small deviation, 0.005 a.u. In future works, calculations on larger systems can be facilitated by the use of larger and adaptive time steps, as the maximum normalized deviation of the absorption calculated with 0.025 a.u. time steps was only on the order of 1×10^{-4}. The use of symplectic integrators did not seem to be necessary; hence RK4 was used. Changes in the basis set had a big impact on the results. As the computational cost scales steeply with respect to the basis set, TDCCSD/aug-cc-p(C)VDZ was chosen as a compromise between accuracy and computational cost.

After using the time-dependent coupled-cluster model to assess the convergence of LiH spectra, we used the model to calculate the ultrafast transient absorption in LiF, using the same parameters. The transient absorption displayed peaks that oscillate with respect to pump-probe delay, and the oscillation frequencies were correlated with the quantum interference of different states in the pump-induced superposition.

Note that nuclear motion, which has been neglected in the model, will cause broadening of the spectral peaks [41]. A natural next step would be to include the nuclear motion to the model, which for instance can be done using the approach in [42].

APPENDIX A: DERIVATION OF GENERALIZED EHRENFEST THEOREM IN TRUNCATED TDCC

For ease of notation, the time dependence is not written explicitly in this section. The derivation of Eq. (15) in truncated TDCC is given here. It makes use of the identity resolution

$$\mathbb{1} = |\text{HF}\rangle\langle\text{HF}| + \sum_{\mu>0} |\mu\rangle\langle\mu|, \tag{A1}$$

where the summation is over all the excited determinants. Sums that are restricted to the excited determinants in the projection space will be denoted by the upper summation limit n.

Consider a generic operator A with no parametric time dependence and two independent solutions to the projected time-dependent Schrödinger equation, $|\text{CC}\rangle$ and $\langle\Lambda'|$. The time derivative of the matrix element $\langle\Lambda'|A|\text{CC}\rangle$ is

$$\frac{d}{dt}\langle\Lambda'|A|\text{CC}\rangle = \left(\frac{d}{dt}\langle\Lambda'|\right)A|\text{CC}\rangle + \langle\Lambda'|\frac{\partial A}{\partial t}|\text{CC}\rangle$$
$$+ \langle\Lambda'|A\left(\frac{d}{dt}|\text{CC}\rangle\right). \tag{A2}$$

Equations (11) and (13) can be used to rewrite the term containing the time derivative of the $|\text{CC}\rangle$ state,

$$\langle\Lambda'|A\left(\frac{d}{dt}|\text{CC}\rangle\right) = \sum_{\mu>0}^{n}\langle\Lambda'|A\tau_\mu|\text{CC}\rangle\frac{dt_\mu}{dt} + i\langle\Lambda'|A|\text{CC}\rangle\frac{d\epsilon}{dt}$$
$$= -i\langle\Lambda'|A\,e^{T}P_n e^{-T}H|\text{CC}\rangle. \tag{A3}$$

Equations (11), (12), and (13) can be used to rewrite the term containing the time derivative of the $\langle\Lambda'|$ state,

$$\left(\frac{d}{dt}\langle\Lambda'|\right)A|\text{CC}\rangle = \sum_{\mu>0}^{n}\frac{d\bar{t}'_\mu}{dt}\langle\mu|e^{-T'}e^{-i\epsilon'}A|\text{CC}\rangle$$
$$- \sum_{\mu>0}^{n}\langle\Lambda'|\tau_\mu A|\text{CC}\rangle\frac{dt'_\mu}{dt} - i\langle\Lambda'|A|\text{CC}\rangle\frac{d\epsilon'}{dt}$$
$$= \sum_{\mu>0}^{n}i\langle\Lambda'|H\,e^{T'}|\mu\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$- \sum_{\mu>0}^{n}i\langle\Lambda'|e^{T'}\tau_\mu\bar{H}'|\text{HF}\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$+ \sum_{\mu>0}^{n}i\langle\Lambda'|\tau_\mu A|\text{CC}\rangle\langle\mu|\bar{H}'|\text{HF}\rangle$$
$$+ i\langle\Lambda'|A|\text{CC}\rangle\langle\text{HF}|\bar{H}'|\text{HF}\rangle. \tag{A4}$$

The right-hand side of Eq. (A1) is inserted between τ_μ and $\bar{H}'$ in the second term, giving

$$\left(\frac{d}{dt}\langle\Lambda'|\right)A|\text{CC}\rangle$$
$$= \sum_{\mu>0}^{n}i\langle\Lambda'|H\,e^{T'}|\mu\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$- \sum_{\mu>0}^{n}i\langle\Lambda'|e^{T'}|\mu\rangle\langle\text{HF}|\bar{H}'|\text{HF}\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$- \sum_{\mu>0}^{n}\sum_{\nu>0}i\langle\Lambda'|e^{T'}\tau_\mu|\nu\rangle\langle\nu|\bar{H}'|\text{HF}\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$+ \sum_{\mu>0}^{n}i\langle\Lambda'|\tau_\mu A|\text{CC}\rangle\langle\mu|\bar{H}'|\text{HF}\rangle$$
$$+ i\langle\Lambda'|A|\text{CC}\rangle\langle\text{HF}|\bar{H}'|\text{HF}\rangle$$
$$= \sum_{\mu>0}^{n}i\langle\Lambda'|H\,e^{T'}|\mu\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$- \sum_{\mu>0}i\langle\Lambda'|e^{T'}|\mu\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle\langle\text{HF}|\bar{H}'|\text{HF}\rangle$$
$$- \sum_{\nu>0}^{n}\sum_{\mu>0}i\langle\Lambda'|e^{T'}\tau_\nu|\mu\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle\langle\nu|\bar{H}'|\text{HF}\rangle$$
$$+ \sum_{\mu>0}^{n}i\langle\Lambda'|\tau_\mu A|\text{CC}\rangle\langle\mu|\bar{H}'|\text{HF}\rangle$$
$$+ i\langle\Lambda'|A|\text{CC}\rangle\langle\text{HF}|\bar{H}'|\text{HF}\rangle. \tag{A5}$$

The factors $\sum_{\mu>0}|\mu\rangle\langle\mu|$ in the second and third terms are replaced by using Eq. (A1), with $|\text{HF}\rangle\langle\text{HF}|$ subtracted from both sides of the equation, giving

$$\left(\frac{d}{dt}\langle\Lambda'|\right)A|\text{CC}\rangle$$
$$= \sum_{\mu>0}^{n}i\langle\Lambda'|He^{T'}|\mu\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$+ i\langle\Lambda'|e^{T'}|\text{HF}\rangle\langle\text{HF}|e^{-T'}A|\text{CC}\rangle\langle\text{HF}|\bar{H}'|\text{HF}\rangle$$
$$+ \sum_{\nu>0}^{n}i\langle\Lambda'|e^{T'}|\nu\rangle\langle\text{HF}|e^{-T'}A|\text{CC}\rangle\langle\nu|\bar{H}'|\text{HF}\rangle$$
$$= \sum_{\mu>0}^{n}i\langle\Lambda'|H\,e^{T'}|\mu\rangle\langle\mu|e^{-T'}A|\text{CC}\rangle$$
$$+ i\langle\Lambda'|e^{T'}|\text{HF}\rangle\langle\text{HF}|\bar{H}'|\text{HF}\rangle\langle\text{HF}|e^{-T'}A|\text{CC}\rangle$$
$$+ \sum_{\nu>0}i\langle\Lambda'|e^{T'}|\nu\rangle\langle\nu|\bar{H}'|\text{HF}\rangle\langle\text{HF}|e^{-T'}A|\text{CC}\rangle$$
$$= i\langle\Lambda'|H\,e^{T'}P_n e^{-T'}A|\text{CC}\rangle, \tag{A6}$$

where the definition of P_n is given in Eq. (16). Equations (A3) and (A6) are inserted into Eq. (A2), giving the desired result:

$$\frac{d}{dt}\langle\Lambda'|A|CC\rangle = i\langle\Lambda'|H\,e^{T'}P_n e^{-T'}A|CC\rangle$$
$$- i\langle\Lambda'|A\,e^{T}P_n e^{-T}H|CC\rangle$$
$$+ \langle\Lambda'|\frac{\partial A}{\partial t}|CC\rangle. \tag{A7}$$

APPENDIX B: RUNGE-KUTTA METHODS

The commonly used one-step integration methods known as Runge-Kutta methods are introduced below in the notation of [30].

Given the following Cauchy problem:

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{f}(t, \mathbf{y}(t)), \quad t \geqslant t_0, \ \mathbf{y}(t_0) = \mathbf{y}_0, \tag{B1}$$

we can find a numerical approximation of the solution $\mathbf{y}(t)$ by the use of a ν-stage Runge-Kutta method, which can be written in the form

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h\sum_{j=1}^{\nu} b_j \mathbf{f}(t_n + c_j h, \xi_j), \tag{B2}$$

where

$$\xi_j = \mathbf{y}_n + h\sum_{i=1}^{\nu} a_{ji}\mathbf{f}(t_n + c_i h, \xi_i), \quad j = 1, \ldots, \nu. \tag{B3}$$

Here, a_{ji}, b_j, and c_j are method specific coefficients, where a_{ji} and c_j need to satisfy the condition

$$\sum_{j=1}^{\nu} a_{ji} = c_j, \quad j = 1, \ldots, \nu \tag{B4}$$

to obtain nontrivial orders of integration. In explicit Runge-Kutta (ERK) methods, the matrix $A = (a_{ji})_{j,i=1,\ldots,\nu}$ is strictly lower triangular. In these methods, ξ_j are explicitly given as a function of $\xi_{j-1}, \ldots, \xi_1$.

In the cases where the matrix A is not strictly lower triangular, ξ_j may also depend on $\xi_j, \ldots, \xi_\nu$, which in practice means that a system of equations have to be solved at each time step. These methods are known as implicit Runge-Kutta (IRK) methods, and in many cases offer greater stability than their explicit counterparts. Since IRK methods involve the solution of a set of equations at each time step, it is hard to give an *a priori* estimate of the number of function evaluations needed at each time step. This number is usually higher than for ERK methods, leading in general to higher computational costs.

[1] M. T. Hassan, T. T. Luu, A. Moulet, O. Raskazovskaya, P. Zhokhov, M. Garg, N. Karpowicz, A. M. Zheltikov, V. Pervak, F. Krausz, and E. Goulielmakis, Nature (London) **530**, 66 (2016).

[2] M. Galli, V. Wanie, D. P. Lopes, E. P. Månsson, A. Trabattoni, L. Colaizzi, K. Saraswathula, A. Cartella, F. Frassetto, L. Poletto, F. Légaré, S. Stagira, M. Nisoli, R. M. Vázquez, R. Osellame, and F. Calegari, Opt. Lett. **44**, 1308 (2019).

[3] D. Fabris, T. Witting, W. A. Okell, D. J. Walke, P. Matia-Hernando, J. Henkel, T. R. Barillot, M. Lein, J. P. Marangos, and J. W. G. Tisch, Nat. Photon. **9**, 383 (2015).

[4] J. Duris, S. Li, T. Driver, E. G. Champenois, J. P. MacArthur, A. A. Lutman, Z. Zhang, P. Rosenberger, J. W. Aldrich, R. Coffee, G. Coslovich, F.-J. Decker, J. M. Glownia, G. Hartmann, W. Helml, A. Kamalov, J. Knurr, J. Krzywinski, M.-F. Lin, J. P. Marangos, M. Nantel, A. Natan, J. T. O'Neal, N. Shivaram, P. Walter, A. L. Wang, J. J. Welch, T. J. A. Wolf, J. Z. Xu, M. F. Kling, P. H. Bucksbaum, A. Zholents, Z. Huang, J. P. Cryan, and A. Marinelli, Nat. Photon. **14**, 30 (2020).

[5] T. J. A. Wolf, R. H. Myhre, J. P. Cryan, S. Coriani, R. J. Squibb, A. Battistoni, N. Berrah, C. Bostedt, P. Bucksbaum, G. Coslovich, R. Feifel, K. J. Gaffney, J. Grilj, T. J. Martinez, S. Miyabe, S. P. Moeller, M. Mucke, A. Natan, R. Obaid, T. Osipov, O. Plekan, S. Wang, H. Koch, and M. Gühr, Nat. Commun. **8**, 29 (2017).

[6] E. Goulielmakis, Z.-H. Loh, A. Wirth, R. Santra, N. Rohringer, V. S. Yakovlev, S. Zherebtsov, T. Pfeifer, A. M. Azzeer, M. F. Kling, S. R. Leone, and F. Krausz, Nature (London) **466**, 739 (2010).

[7] A. D. Dutoi, K. Gokhberg, and L. S. Cederbaum, Phys. Rev. A **88**, 013419 (2013).

[8] A. D. Dutoi and L. S. Cederbaum, Phys. Rev. A **90**, 023414 (2014).

[9] D. R. Nascimento and A. E. DePrince, J. Chem. Theory Comput. **12**, 5834 (2016).

[10] J. Olsen, P. Jørgensen, and J. Simons, Chem. Phys. Lett. **169**, 463 (1990).

[11] P. Hessler, N. T. Maitra, and K. Burke, J. Chem. Phys. **117**, 72 (2002).

[12] N. T. Maitra, J. Chem. Phys. **144**, 220901 (2016).

[13] A. Bruner, S. Hernandez, F. Mauger, P. M. Abanador, D. J. LaMaster, M. B. Gaarde, K. J. Schafer, and K. Lopata, J. Phys. Chem. Lett. **8**, 3991 (2017).

[14] H. Koch and P. Jørgensen, J. Chem. Phys. **93**, 3333 (1990).

[15] S. Kvaal, J. Chem. Phys. **136**, 194109 (2012).

[16] D. A. Pigg, G. Hagen, H. Nam, and T. Papenbrock, Phys. Rev. C **86**, 014308 (2012).

[17] H. E. Kristiansen, Ø. S. Schøyen, S. Kvaal, and T. B. Pedersen, J. Chem. Phys. **152**, 71102 (2020).

[18] J. A. Sonk and H. B. Schlegel, J. Phys. Chem. A **115**, 11832 (2011).

[19] J. A. Sonk, M. Caricato, and H. B. Schlegel, J. Phys. Chem. A **115**, 4678 (2011).

[20] E. Luppi and M. Head-Gordon, Mol. Phys. **110**, 909 (2012).

[21] D. R. Nascimento and A. E. DePrince, J. Phys. Chem. Lett. **8**, 2951 (2017).

[22] L. N. Koulias, D. B. Williams-Young, D. R. Nascimento, A. E. DePrince, and X. Li, J. Chem. Theory Comput. **15**, 6617 (2019).

[23] D. R. Nascimento and A. E. DePrince, J. Chem. Phys. **151**, 204107 (2019).

[24] Y. C. Park, A. Perera, and R. J. Bartlett, J. Chem. Phys. **151**, 164117 (2019).

[25] H. Koch, H. J. A. Jensen, P. Jørgensen, and T. Helgaker, J. Chem. Phys. **93**, 3345 (1990).

[26] T. B. Pedersen and S. Kvaal, J. Chem. Phys. **150**, 144106 (2019).

[27] H. Koch, H. J. A. Jensen, P. Jørgensen, T. Helgaker, G. E. Scuseria, and H. F. Schaefer, J. Chem. Phys. **92**, 4924 (1990).

[28] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley & Sons, Ltd., New York, 2014), Chap. 13, pp. 648–723.

[29] M. Wu, S. Chen, S. Camp, K. J. Schafer, and M. B. Gaarde, J. Phys. B: At., Mol., Opt. Phys. **49**, 62003 (2016).

[30] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, 2nd ed., Cambridge Texts in Applied Mathematics (Cambridge University Press, Cambridge, UK, 2008).

[31] F. J. Harris, Proc. IEEE **66**, 51 (1978).

[32] S. D. Folkestad, E. F. Kjønstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, T. Moitra, A. C. Paul, M. Scavino, A. S. Skeidsvoll, Å. H. Tveten, and H. Koch, J. Chem. Phys. **152**, 184103 (2020).

[33] E. Tiesinga, P. J. Mohr, D. B. Newell, and B. N. Taylor, The 2018 CODATA recommended values of the fundamental physical constants, NIST standard reference database 121, web version 8.1, 2020.

[34] H. R. Reiss, Prog. Quantum Electron. **16**, 1 (1992).

[35] R. D. Johnson III, NIST computational chemistry comparison and benchmark database, NIST standard reference database 101, release 20, 2019.

[36] S. Coriani and H. Koch, J. Chem. Phys. **143**, 181103 (2015).

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

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

[39] E. Coccia, B. Mussard, M. Labeye, J. Caillat, R. Taïeb, J. Toulouse, and E. Luppi, Int. J. Quantum Chem. **116**, 1120 (2016).

[40] M. Labeye, F. Zapata, E. Coccia, V. Véniard, J. Toulouse, J. Caillat, R. Taïeb, and E. Luppi, J. Chem. Theory Comput. **14**, 5846 (2018).

[41] M. Bixon and J. Jortner, J. Chem. Phys. **48**, 715 (1968).

[42] T. B. Pedersen and H. Koch, J. Chem. Phys. **108**, 5194 (1998).

Simulating weak-field attosecond processes with a Lanczos reduced basis approach to time-dependent equation-of-motion coupled-cluster theory

Andreas S. Skeidsvoll [1,*] Torsha Moitra [2,*] Alice Balbi [3] Alexander C. Paul [1]
Sonia Coriani [2,1,†] and Henrik Koch [1,3,‡]

[1]*Department of Chemistry, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*
[2]*Department of Chemistry, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*
[3]*Scuola Normale Superiore, Piazza dei Cavalieri 7, IT-56126 Pisa, Italy*

A time-dependent equation-of-motion coupled-cluster singles and doubles (TD-EOM-CCSD) method is implemented, which uses a reduced basis calculated with the asymmetric band Lanczos algorithm. The approach is used to study weak-field processes in small molecules induced by ultrashort valence pump and core probe pulses. We assess the reliability of the procedure by comparing TD-EOM-CCSD absorption spectra to spectra obtained from the time-dependent coupled-cluster singles and doubles method, and observe that spectral features can be reproduced for several molecules, at much lower computational times. We discuss how multiphoton absorption and symmetry can be handled in the method, and general features of the core-valence separation projection technique. We also model the transient absorption of an attosecond x-ray probe pulse by the glycine molecule.

I. INTRODUCTION

Stimulated by the recent experimental realization of various laser pulses with durations on the attosecond (1×10^{-18} s) time scale [1–5], capable of monitoring electronic motion, the theoretical simulation of coherent electron dynamics is currently an active field of research [6].

Real-time electronic structure theory considers the explicit time dependence of the electronic system by evolving the time-dependent Schrödinger equation in the time domain [6]. Explicitly time-dependent methods can directly provide the time-domain evolution of electronic wave functions together with nuclear motion, representing a versatile way of tracking ultrafast phenomena in both perturbative and nonperturbative regimes [7,8].

The development of real-time methods commenced in the late 1970s and early 1980s in the field of nuclear physics [9–11]. Despite these early endeavors, real-time methods did not become practical at that time due to the lack of electron correlation effects at the Hartree-Fock level and the high computational cost associated with propagation of correlated wave functions. However, decades of steady advancements in computing power and numerical algorithms have led to a renewed interest in explicit time propagation in correlated methods like density-functional theory [12,13], multiconfigurational self-consistent-field [14–16], configuration-interaction [17–20], algebraic diagrammatic construction [21,22], and coupled-cluster [23–33] methods.

In this paper, we present an implementation and representative case studies of the time-dependent equation-of-motion coupled-cluster (TD-EOM-CC) model for simulating weak-field attosecond valence pump–core probe processes. In conjunction with a reduced-space band Lanczos algorithm for obtaining the valence and core excited states, this model offers results similar to its time-dependent coupled-cluster (TDCC) counterpart in weak fields, at significantly lower computational costs. The reduction in cost enables the study of larger systems.

The paper is organized as follows. In Sec. II we detail the theory behind TD-EOM-CC and the asymmetric band Lanczos algorithm. Here, we also discuss a strategy used in order to guide the reduced space solver to directly obtain the transitions between excited states. The computational procedure used is detailed in Sec. III. In Sec. IV, simulations for various molecular systems are presented. First a benchmark study is presented for LiF, validating our proposed method. Second, the applicability of the core-valence separation (CVS) scheme is tested for LiH. Then, a two-photon absorption phenomenon has been captured using a stepwise procedure emulating the actual physical process for C_2H_4. Finally, we put forward a theoretical assessment of pump-probe absorption for the glycine molecule, which is deemed suitable for further experimental investigations. The findings are summarized in Sec. V.

II. THEORY

A. System

We model the system, composed of a molecule and its interaction with laser pulses, with the Hamiltonian

$$H(t) = H^{(0)} + V(t), \qquad (1)$$

*These authors contributed equally to this work.
†soco@kemi.dtu.dk
‡henrik.koch@sns.it

where the field-free Hamiltonian $H^{(0)}$ describes the molecule with fixed nuclei and without interactions with the external electromagnetic field. The semiclassical time-dependent interaction term, written in the dipole approximation and length gauge, is

$$V(t) = -\boldsymbol{d} \cdot \boldsymbol{\mathcal{E}}(t) \qquad (2)$$

and describes the interaction between the molecular electrons and the external electromagnetic field. The latter is represented by the electric field $\boldsymbol{\mathcal{E}}(t) = [\mathcal{E}_x(t)\, \mathcal{E}_y(t)\, \mathcal{E}_z(t)]^T$ and electronic dipole operator vectors, $\boldsymbol{d} = [d_x\, d_y\, d_z]^T$. We assume that the molecule is initially in the ground state of the field-free Hamiltonian, and take the electric field to be a linear combination of the electric fields of any number of laser pulses:

$$\boldsymbol{\mathcal{E}}(t) = \sum_n \boldsymbol{\mathcal{E}}_{0n} \cos[\omega_{0n}(t - t_{0n}) + \phi_n] f_n(t). \qquad (3)$$

The field of laser pulse n has an associated carrier frequency ω_{0n}, peak strength $|\boldsymbol{\mathcal{E}}_{0n}|$, polarization $\boldsymbol{\mathcal{E}}_{0n}/|\boldsymbol{\mathcal{E}}_{0n}|$, and an $8\sigma_n$-truncated Gaussian envelope function

$$f_n(t) = \begin{cases} e^{-(t-t_{0n})^2/(2\sigma_n^2)} & \text{if } |t - t_{0n}| \leqslant 8\sigma_n \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

with duration specified by σ_n, the temporal rms width. It is also specified by the central time t_{0n} and the carrier-envelope phase ϕ_n. We assume the carrier-envelope phase to be zero for all pulses, meaning that the maximum values of the envelope and cosine carrier functions belonging to pulse n coincide at t_{0n}.

The energy absorbed during the interaction can be given by [28,34]

$$\Delta E = \int_0^\infty \omega S(\omega) d\omega, \qquad (5)$$

where $S(\omega)$ is the response function

$$S(\omega) = -2\text{Im}[\langle \widetilde{\mathbf{d}} \rangle(\omega) \cdot \widetilde{\boldsymbol{\mathcal{E}}}^*(\omega)] \quad \omega > 0. \qquad (6)$$

The vectors $\langle \widetilde{\mathbf{d}} \rangle(\omega)$ and $\widetilde{\boldsymbol{\mathcal{E}}}(\omega)$ are the Fourier transforms of the time-dependent dipole moment expectation value and electric-field vectors, respectively, and the asterisk denotes complex conjugation. A positive or negative value of the function $S(\omega)$ describes the probability of absorption or emission of light with frequency ω, respectively [34].

B. TD-EOM-CC states

The time-dependent ket and bra of a TD-EOM-CC state can be expressed as

$$|\Psi(t)\rangle = \sum_j |\psi_j\rangle s_j(t), \quad \langle \widetilde{\Psi}(t)| = \sum_i k_i(t) \langle \widetilde{\psi}_i|, \qquad (7)$$

where the italic indices i and j are used to denote general equation-of-motion coupled-cluster (EOM-CC) states, including the ground state with index zero. The time-independent EOM-CC kets and bras are given by

$$|\psi_j\rangle = e^T R_j |\text{HF}\rangle, \quad \langle \widetilde{\psi}_i| = \langle \text{HF}| L_i e^{-T}. \qquad (8)$$

We assume that the EOM-CC states are biorthonormal:

$$\langle \widetilde{\psi}_i | \psi_j \rangle = \delta_{ij}. \qquad (9)$$

In the following, we let the indices κ and λ denote general determinants in the projection space, including the reference Hartree-Fock determinant with index zero. We use the indices μ and ν, on the other hand, to denote excited determinants.

The cluster operator T and the right and left operators R_j and L_i of Eq. (8) can be expressed as linear expansions in a finite set of operators τ_λ and $\tau_\kappa^\dagger$,

$$T = \sum_\nu \tau_\nu t_\nu, \quad R_j = \sum_\lambda \tau_\lambda r_{\lambda j}, \quad L_i = \sum_\kappa l_{i\kappa} \tau_\kappa^\dagger, \qquad (10)$$

where the operator with index zero is the unit operator,

$$\tau_0 = \tau_0^\dagger = 1, \qquad (11)$$

and the τ_ν and $\tau_\mu^\dagger$ operators generate excited determinants from the ket and bra reference Hartree-Fock determinants, respectively:

$$\tau_\nu |\text{HF}\rangle = |\nu\rangle, \quad \langle \text{HF}|\tau_\mu^\dagger = \langle \mu|, \qquad (12)$$

$$\tau_\mu^\dagger |\text{HF}\rangle = 0, \qquad \langle \text{HF}|\tau_\nu = 0. \qquad (13)$$

We assume that the determinants are biorthogonal:

$$\langle \kappa | \lambda \rangle = \delta_{\kappa\lambda}. \qquad (14)$$

If all possible electronic excitations are included in the summations in Eq. (10), the method is equivalent to full configuration interaction. The sum can also be restricted to given excitation levels, giving approximate methods that scale polynomially with the system size. This includes the coupled-cluster singles and doubles method, where summation is only done over single and double excitations. We do not explicitly state the excitation levels included in the following expressions, since they hold for both restricted and unrestricted summation.

The cluster amplitudes t_ν in Eq. (10) can be found from solving equations involving the similarity-transformed field-free Hamiltonian operator $\bar{H}^{(0)}$ projected onto the right reference and left excited determinants

$$\langle \mu | \bar{H}^{(0)} | \text{HF} \rangle = 0, \qquad (15)$$

where the similarity transformation of an operator X is denoted by an overbar:

$$\bar{X} = e^{-T} X e^T. \qquad (16)$$

After the optimal cluster amplitudes t_ν have been determined, the right and left vectors of EOM-CC state i, with components $r_{\lambda j}$ and $l_{i\kappa}$, can be found as right and left eigenvectors of the field-free Hamiltonian matrix, with elements

$$H_{\kappa\lambda}^{(0)} = \langle \kappa | \bar{H}^{(0)} | \lambda \rangle. \qquad (17)$$

The right and left eigenvectors of the matrix in Eq. (17) with the lowest eigenvalue, specifying the ground EOM-CC state with index zero, have the following structure:

$$r_{00} = 1, \quad r_{\nu 0} = 0, \qquad (18)$$

$$l_{00} = 1, \quad l_{0\mu} = \bar{t}_\mu. \qquad (19)$$

The multipliers $\bar{t}_\mu$ are solutions to the equations

$$\langle \text{HF} | \bar{H}^{(0)} | \nu \rangle + \sum_\mu \bar{t}_\mu A_{\mu\nu}^{(0)} = 0, \qquad (20)$$

where elements of the field-free coupled-cluster Jacobian matrix $\boldsymbol{A}^{(0)}$ are given by

$$A^{(0)}_{\mu\nu} = \langle \mu | [\bar{H}^{(0)}, \tau_\nu] | \text{HF} \rangle. \tag{21}$$

The other right and left eigenvectors of the matrix in Eq. (17) correspond to excited EOM-CC states, denoted by the italic indices m and n. The eigenvectors have the following reference determinant components:

$$r_{0n} = -\sum_\nu \bar{t}_\nu r_{\nu n}, \tag{22}$$

$$l_{m0} = 0. \tag{23}$$

These components enforce the biorthogonality between the ground and excited states, in accordance with Eq. (9). The vectors $\boldsymbol{R}_m$ and $\boldsymbol{L}_m$, containing the components $r_{\nu m}$ and $l_{m\mu}$ of excited EOM-CC state m, are right and left eigenvectors of $\boldsymbol{A}^{(0)}$ with eigenvalue ω_m.

C. Derivation of TD-EOM-CC equations

The time derivative of the coefficients of the TD-EOM-CC ket can be found from projecting the ket time-dependent Schrödinger equation (TDSE)

$$\iota \frac{\partial}{\partial t} |\Psi(t)\rangle = H(t) |\Psi(t)\rangle, \tag{24}$$

where ι denotes the imaginary unit, onto the bra of EOM-CC state i, giving

$$\iota \frac{\partial s_i(t)}{\partial t} = \sum_j H_{ij}(t) s_j(t), \tag{25}$$

where the matrix elements of an operator $X(t)$ are given by

$$X_{ij}(t) = \langle \widetilde{\psi}_i | X(t) | \psi_j \rangle. \tag{26}$$

Likewise, the time derivative of the coefficients of the TD-EOM-CC bra can be found from projecting the bra TDSE,

$$-\iota \frac{\partial}{\partial t} \langle \widetilde{\Psi}(t)| = \langle \widetilde{\Psi}(t)| H(t), \tag{27}$$

onto the ket of EOM-CC state j, giving

$$-\iota \frac{\partial k_j(t)}{\partial t} = \sum_i k_i(t) H_{ij}(t). \tag{28}$$

The TD-EOM-CC equations (25) and (28) were to our knowledge first presented in Ref. [18], and have also been used in Ref. [19]. In those works, the matrix elements d_{ij} of the dipole moment operator, entering in the time-dependent Hamiltonian, are approximated by discarding non-Hermitian components. This was achieved by using the Hermitian $(d_{ij} + d^*_{ji})/2$ instead of the d_{ij} given by Eq. (26). In the present paper, however, the full non-Hermitian d_{ij} are used in the solution of Eqs. (25) and (28). After the time-dependent coefficients $k_i(t)$ and $s_i(t)$ have been obtained, the time-dependent expectation value of a time-independent operator X can be calculated according to

$$\langle X \rangle(t) = \sum_{ij} k_i(t) X_{ij} s_j(t). \tag{29}$$

D. Asymmetric band Lanczos algorithm

We use the asymmetric band Lanczos algorithm to generate approximate eigenvalues $\widetilde{\omega}_n$ and right $\widetilde{\boldsymbol{R}}_n$ and left $\widetilde{\boldsymbol{L}}_n$ eigenvectors of the field-free Jacobian matrix $\boldsymbol{A}^{(0)}$. As outlined in Sec. II F, the approximate eigenvectors are used as a reduced basis for solving Eqs. (25) and (28).

The algorithm is a generalization of the simple asymmetric Lanczos algorithm, employing m right $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$ and p left $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p)$ starting vectors instead of single ones [35–37]. A sequence of right vectors is constructed by transforming the right starting vectors by increasing powers of a given square asymmetric matrix $\boldsymbol{M}$. For our purpose, $\boldsymbol{M} = \boldsymbol{A}^{(0)}$. The i first vectors in the sequence, which can be linearly dependent, span the n-dimensional right band Krylov subspace,

$$\mathcal{K}_i(\boldsymbol{M}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$$
$$= \text{span}\{\underbrace{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m, \boldsymbol{M}\boldsymbol{b}_1, \ldots, \boldsymbol{M}\boldsymbol{b}_m, \boldsymbol{M}^2\boldsymbol{b}_1, \ldots}_{i}\}, \tag{30}$$

where $i - n$ is the number of redundant vectors in the sequence. Likewise, a sequence of left vectors is constructed by transforming the left starting vectors by increasing powers of the transpose matrix $\boldsymbol{M}^T$. The j first vectors in the sequence, which can be linearly dependent, span the n-dimensional left band Krylov subspace,

$$\mathcal{K}_j(\boldsymbol{M}^T, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_p)$$
$$= \text{span}\{\underbrace{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p, \boldsymbol{M}^T\boldsymbol{c}_1, \ldots, \boldsymbol{M}^T\boldsymbol{c}_p, (\boldsymbol{M}^T)^2\boldsymbol{c}_1, \ldots}_{j}\}, \tag{31}$$

where $j - n$ is the number of redundant vectors in the sequence. Note that the subspaces can be regarded as block Krylov subspaces [37] whenever i and j are multiples of the number of starting vectors. The n right $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ and left $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ Lanczos vectors, which form respective bases for $\mathcal{K}_i(\boldsymbol{M}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$ and $\mathcal{K}_j(\boldsymbol{M}^T, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_p)$, are obtained by discarding redundant vectors of the sequences in Eqs. (30) and (31). A $n \times n$-dimensional band Krylov subspace approximation of $\boldsymbol{M}$ can then be obtained by expressing the matrix in the Lanczos vector bases.

The right and left Lanczos vectors can be generated iteratively with the recurrence relations [35]

$$\boldsymbol{M}\boldsymbol{V}_n = \boldsymbol{V}_n \boldsymbol{T}_n + \hat{\boldsymbol{V}}^C_n + \hat{\boldsymbol{V}}^D_n, \tag{32}$$

$$\boldsymbol{M}^T \boldsymbol{W}_n = \boldsymbol{W}_n \widetilde{\boldsymbol{T}}_n + \hat{\boldsymbol{W}}^C_n + \hat{\boldsymbol{W}}^D_n, \tag{33}$$

where the right and left Lanczos vectors form the matrices $\boldsymbol{V}_n = [\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_n]$ and $\boldsymbol{W}_n = [\boldsymbol{w}_1 \ \cdots \ \boldsymbol{w}_n]$, respectively. The un-normalized vectors that form the nonzero columns of $\hat{\boldsymbol{V}}^C_n = [\boldsymbol{0} \ \cdots \ \boldsymbol{0} \ \hat{\boldsymbol{v}}_{n+1} \ \cdots \ \hat{\boldsymbol{v}}_{n+m_c}]$ and $\hat{\boldsymbol{W}}^C_n = [\boldsymbol{0} \ \cdots \ \boldsymbol{0} \ \hat{\boldsymbol{w}}_{n+1} \ \cdots \ \hat{\boldsymbol{w}}_{n+p_c}]$ serve as candidates for the next right and left Lanczos vectors, respectively. The sparse matrices $\hat{\boldsymbol{V}}^D_n$ and $\hat{\boldsymbol{W}}^D_n$ contain un-normalized candidates from previous iterations that have been deflated (i.e., discarded) due to linear dependence on already accepted right and left Lanczos vectors, respectively. Finally, the nonzero elements of $\boldsymbol{T}_n$ and $\widetilde{\boldsymbol{T}}_n$ are used to enforce the biorthogonality

between the $m + p + 1$ vectors that can overlap in exact arithmetic at each iteration [35].

In numerical implementations, vectors are usually deflated when linear independence is below a given threshold, since inexact arithmetic prevents the description of exact linear dependence. The numbers m_c and p_c, initially equal to the number of right and left starting vectors m and p, give the current number of vectors available for deflation. We say that the sequence in Eqs. (30) and (31) is fully exhausted when m or p deflations have occurred, respectively. The iterative procedure is then terminated, giving equal numbers of right and left Lanczos vectors.

The iterative solution of Eqs. (32) and (33) is done in accordance with Algorithm 5.1 of Ref. [37], with two exceptions. The following biorthogonalization step is added at the beginning of step 1:

> **if** $n > 1$ **then**
>   **for** $k = 1$ **to** $\max\{1, n - p_c - 1\}$ **do**
> $\hat{v}_n \leftarrow \hat{v}_n - v_k(w_k^T \hat{v}_n)$
>   **end for**
> **end if**

and the following at the beginning of step 2:

> **if** $n > 1$ **then**
>   **for** $k = 1$ **to** $\max\{1, n - m_c - 1\}$ **do**
> $\hat{w}_n \leftarrow \hat{w}_n - (\hat{w}_n^T v_k)w_k$
>   **end for**
> **end if**

These additions lead to an algorithm that enforces the biorthogonality between all Lanczos vectors in inexact arithmetic,

$$W_n^T V_n = \Delta_n = \mathrm{diag}(\delta_1, \delta_2, \dots, \delta_m), \qquad (34)$$

and not just between the vectors that can overlap in exact arithmetic. We observe that this modification of the algorithm is important for numerical stability when the number of iterations becomes large, but the modification also makes the number of vector operations substantially higher. The number of operations can potentially be reduced in future implementations, e.g., by formulating a restarted asymmetric band Lanczos algorithm, based on existing approaches [38].

The iterative procedure continues until a given maximum chain length (i.e., number of iterations) $n = n^{\max}$ is reached, unless the procedure is terminated at a lower n because of full exhaustion of the sequence in Eq. (30) or (31). The algorithm generates the $n \times n$ matrices

$$\begin{aligned} T_n^{\mathrm{P}} &= \Delta_n^{-1} W_n^T M V_n \\ &= T_n + \Delta_n^{-1} W_n^T V_n^D, \end{aligned} \qquad (35)$$

$$\begin{aligned} \widetilde{T}_n^{\mathrm{P}} &= \left(W_n^T M V_n \Delta_n^{-1}\right)^T \\ &= \widetilde{T}_n + \left[\left(W_n^D\right)^T V_n \Delta_n^{-1}\right]^T, \end{aligned} \qquad (36)$$

in accordance with Algorithm 5.1 of Ref. [37]. The matrices T_n^{P} and $\widetilde{T}_n^{\mathrm{P}}$ are related by

$$\Delta_n T_n^{\mathrm{P}} = \left(\widetilde{T}_n^{\mathrm{P}}\right)^T \Delta_n, \qquad (37)$$

and are banded when no deflations have occurred [35].

The matrix T_n^{P} can be viewed as the oblique projection of M onto the n-dimensional $\mathcal{K}_i(M, b_1, \dots, b_m)$ and orthogonally to the n-dimensional $\mathcal{K}_j(M^T, c_1, \dots, c_p)$ [37]. Diagonalization of the matrix yields n eigenvalues, which approximate the eigenvalues of M, and associated right and left eigenvectors. The right eigenvectors can be transformed to approximate eigenvectors of M by premultiplication by V_n, and the left eigenvectors to approximate eigenvectors of M^T by premultiplication by $W_n^T \Delta_n^{-1}$ [35]. Approximate eigenvectors with dominant (low- and high-lying) eigenvalues are typically better converged than the ones in the middle [39–41].

E. Choice of starting vectors

In Appendix A, we demonstrate that operator matrix elements involving excited state n are linear in the right and left vector components r_{vn} and $l_{n\mu}$. We assume that n is one of the states targeted by the band Lanczos algorithm, and state i is a previously calculated ground or excited state. Matrix elements between n and i can thus be written as the product of two vectors. The first vector is the right or left vector of state n. The second vector, which we take to be the starting vector of the algorithm, is based on state i and operator X. This choice of starting vectors simplifies the calculation of operator matrix elements, as will be shown in the following. Analogous arguments have been used in previous work [42–45] to guide the choice of start vectors for the simple Lanczos algorithm for coupled-cluster response [42,43] and EOM-CC theories [44,45].

The starting vectors based on the ground state $i = 0$, which are similar to the starting vectors used in coupled-cluster linear response theory [42,43], are given by

$$b_{\mu 0}^X = \xi_\mu^X, \qquad (38)$$

$$c_{0v}^X = {}^{\mathrm{EOM}}\eta_v^X - X_{00} \bar{t}_v, \qquad (39)$$

and the starting vectors based on excited EOM-CC state $i = m$ are given by [44]

$$b_{\mu m}^X = \sum_v \left({}^{\mathrm{EOM}}A_{\mu v}^X + \delta_{\mu v} X_{00} - \xi_\mu^X \bar{t}_v\right) r_{vm}, \qquad (40)$$

$$c_{mv}^X = \sum_\mu l_{m\mu} \left({}^{\mathrm{EOM}}A_{\mu v}^X + \delta_{\mu v} X_{00} - \xi_\mu^X \bar{t}_v\right). \qquad (41)$$

The specification of ξ_μ^X, ${}^{\mathrm{EOM}}\eta_v^X$, X_{00}, and ${}^{\mathrm{EOM}}A_{\mu v}^X$ is given in Appendix A.

The starting vectors can be expressed in the Lanczos basis by inserting the resolution of identity in terms of the biorthonormal Lanczos vectors:

$$\begin{aligned} b_i^X &= \sum_j v_j \left(w_j^T b_i^X\right) \\ &= \sum_{j=1}^m v_j b_{ji}^X, \end{aligned} \qquad (42)$$

$$\begin{aligned} \left(c_i^X\right)^T &= \sum_j \left[\left(c_i^X\right)^T v_j\right] w_j^T \\ &= \sum_{j=1}^p c_{ij}^X w_j^T. \end{aligned} \qquad (43)$$

The sum in Eq. (42) is restricted since $\boldsymbol{b}_i^X \in \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ while each $(\boldsymbol{w}_{m+1}, \ldots)$ is biorthogonal to all $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m)$. Likewise, the sum in Eq. (43) is restricted since $\boldsymbol{c}_i^X \in \text{span}\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}$ while each $(\boldsymbol{v}_{p+1}, \ldots)$ is biorthogonal to all $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p)$.

Thus, the transition moments involving excited states can be obtained by contracting the starting vectors with the vectors of excited state n:

$$X_{in} = \left(\boldsymbol{c}_i^X\right)^T \boldsymbol{R}_n$$

$$= \sum_{j=1}^{p} c_{ij}^X \left(\boldsymbol{w}_j^T \boldsymbol{R}_n\right)$$

$$= \sum_{j=1}^{p} c_{ij}^X R_{jn}, \tag{44}$$

$$X_{ni} = \boldsymbol{L}_n^T \boldsymbol{b}_i^X$$

$$= \sum_{j=1}^{m} \left(\boldsymbol{L}_n^T \boldsymbol{v}_j\right) b_{ji}^X$$

$$= \sum_{j=1}^{m} L_{nj} b_{ji}^X. \tag{45}$$

R_{jn} and L_{nj} are simply the components of the right and left eigenvectors of $\boldsymbol{T}_n^{\mathrm{P}}$, respectively, and $b_{ji}^X = \boldsymbol{w}_j^T \boldsymbol{b}_i^X$ and $c_{ij}^X = (\boldsymbol{c}_i^X)^T \boldsymbol{v}_j$ are products of starting vectors and biorthonormal Lanczos vectors.

F. Generation of a reduced basis

The iterative process that is used to calculate sets of excited EOM-CC states is given below. The set J_c contains the indices of already calculated EOM-CC states. At the beginning of the procedure, where $c = 0$, only the ground state has been calculated, and $J_0 = \{0\}$. The iterative procedure for the cth EOM-CC state calculation is as follows.

(1) Choose a subset of the state indices from previous calculations, $I_c \subseteq J_{c-1}$, and a set of operators X_c based on the final states that can be accessed by the operators (see Sec. IV). Also choose a maximum chain length $n_c^{\max}$, and maximum eigenvalue $\omega_c^{\max}$ and minimum transition strength $S_c^{\min}$ values.

(2) Sequences of right and left starting vectors, $(\boldsymbol{b}_i^X)_{i \in I_c, X \in X_c}$ and $(\boldsymbol{c}_i^X)_{i \in I_c, X \in X_c}$, are constructed in accordance with Eqs. (38)–(41).

(3) The band Lanczos algorithm described in Sec. II D is run with the maximum chain length $n_c^{\max}$, the field-free Jacobian matrix $\boldsymbol{A}^{(0)}$, and the sequences of starting vectors. The algorithm terminates at $n_c \leqslant n_c^{\max}$, constructing the matrix $\boldsymbol{T}_{n_c}^{\mathrm{P}}$.

(4) The eigenvalues and corresponding right and left eigenvectors of $\boldsymbol{T}_{n_c}^{\mathrm{P}}$ are calculated. Together, these determine a set of n_c approximate EOM-CC states indexed by N_c.

(5) States $n \in N_c$ with approximate eigenvalues $\widetilde{\omega}_n$ of $\boldsymbol{A}^{(0)}$ that are greater than $\omega_c^{\max}$ are discarded.

(6) Matrix elements for all operators $X \in X_c$ and combinations of final $n \in N_c$ and initial $i \in I_c$ states are calculated in accordance with Eqs. (44) and (45). States $n \in N_c$ with
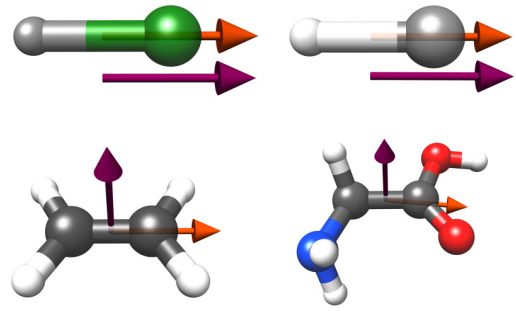

FIG. 1. Illustration of the structures of lithium fluoride (top left), lithium hydride (top right), ethylene (bottom left), and glycine (bottom right), together with the polarization of the valence-exciting pump (orange arrow) and core-exciting probe (purple arrow) pulses.

transition strengths $S_{in}^X = X_{in} X_{ni}$ that are smaller than $S_c^{\min}$ for all operators and initial states are discarded.

(7) For each nondiscarded state $n \in N_c$, the right and left eigenvectors of $\boldsymbol{T}_{n_c}^{\mathrm{P}}$ are transformed to approximate right and left eigenvectors $\widetilde{\boldsymbol{R}}_n$ and $\widetilde{\boldsymbol{L}}_n$ of $\boldsymbol{A}^{(0)}$ by premultiplication by $\boldsymbol{V}_n$ and $\boldsymbol{W}_n^T \boldsymbol{\Delta}_n^{-1}$, respectively.

(8) If an assessment of the convergence of the vectors is requested, residual norms of all approximate right and left vectors are calculated. All states with a left or right residual norm exceeding a given threshold are discarded.

(9) Finally, nondiscarded states with vectors linearly independent of previously calculated vectors are stored, and indexed by $N_c' \subseteq N_c$. The indices are added to the previous index set $J_c = J_{c-1} \cup N_c'$.

The iterative procedure is repeated if states of higher excitation levels are desired. Afterwards, the excited-state Jacobian and overlap matrices, with elements $\widetilde{A}_{mn}^{(0)} = \widetilde{\boldsymbol{L}}_m^T \boldsymbol{A}^{(0)} \widetilde{\boldsymbol{R}}_n$ and $\widetilde{S}_{mn} = \widetilde{\boldsymbol{L}}_m^T \widetilde{\boldsymbol{R}}_n$, are constructed in the reduced basis of the approximate eigenvectors. The right and left generalized eigenvalue problems are solved, giving new sets of right and left eigenvectors of $\widetilde{\boldsymbol{A}}^{(0)}$. The dipole and field-free Hamiltonian matrices are then calculated in the basis of both the ground and the newly generated excited states, in accordance with Eq. (26), and used in solving the time-dependent problems defined by Eqs. (25) and (28).

III. COMPUTATIONAL DETAILS

Experimental geometries from the NIST database [46] are used for LiH, LiF, and C_2H_4. An optimized geometry from the same database is used for glycine, obtained with the MP2 method with all electrons correlated and the cc-pVTZ basis set. The linear molecules LiH and LiF are aligned along the z axis, as done in Ref. [28]. The ethylene molecule is placed in the xy plane, with the C-C bond along the x axis. The glycine molecule is of C_s symmetry for the chosen geometry, with the xy plane as the mirror plane. An illustration of the structures of the molecules, together with the polarizations of the valence-exciting pump and core-exciting probe pulses, is shown in Fig. 1.

In all following calculations, the aug-cc-pCVDZ basis set [47,48] is used for atoms targeted by the core-exciting pulses; the aug-cc-pVDZ [47] basis set is adopted for the remaining
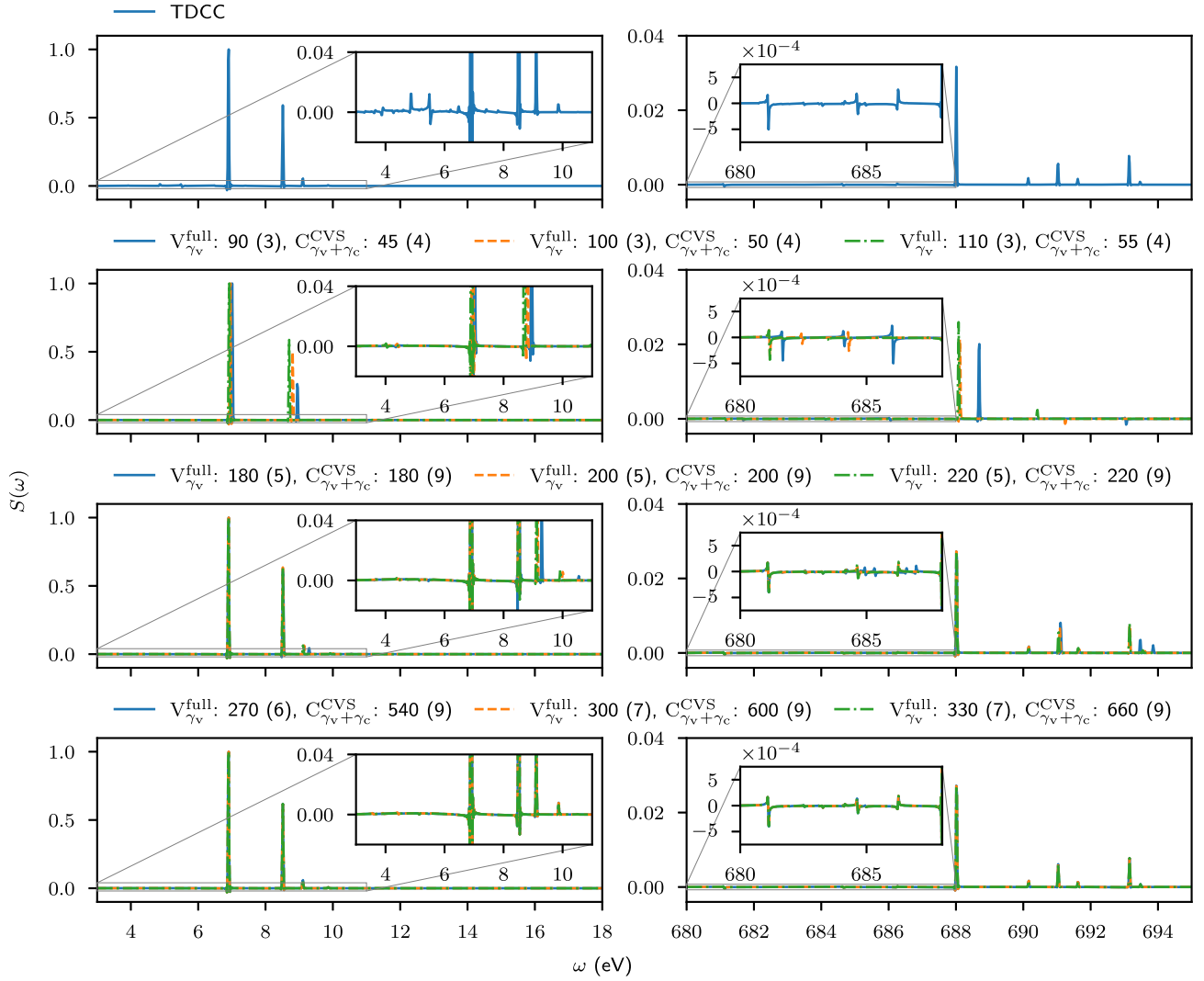
FIG. 2. LiF pump-probe absorption $S(\omega)$ as a function of frequency ω in the valence and core regions, normalized by the tallest peaks in the spectra. The TDCC results are shown in the top left and right panels. TD-EOM-CC results, calculated at different band Lanczos chain lengths, are shown in the lower panels. EOM-CC valence (V) states are calculated in the full projection space, while the core (C) states are calculated within the CVS approximation. Valence states energetically inaccessible by a single pump photon γ_v are discarded, and so are core states energetically inaccessible by subsequent absorption of a probe photon $\gamma_v + \gamma_c$. The chain lengths of the calculations are given, together with the number of converged states (in brackets).

atoms in the molecules. Valence and core states are obtained with the asymmetric band Lanczos algorithm with varying chain lengths as specified in Sec. IV.

Lanczos vectors with Euclidian norms of less than 1×10^{-9} are deflated, but this did not occur in any of the calculations. Final excited states that do not have a minimum transition strength of at least 1×10^{-7} to any initial state, for any of the operators used to construct the starting vectors, are discarded. This is done to only keep states that give a non-negligible contribution to the dynamics. Also, states with excitation energies above $\omega^{\text{max}} = \sum_i^{n_\gamma} E_{\gamma_i}^{\text{max}}$ are discarded, where $E_{\gamma_i}^{\text{max}} = \omega_i + 8\sigma_i^\omega$ is an estimate of the maximum energy of photon i involved in the n_γ-photon transition to the desired excited states. The carrier frequency ω_i and the frequency rms width $\sigma_i^\omega = 1/(2\sigma_i^t)$ are parameters of the pulse providing photon i (see Sec. II A).

CVS [49–51] projectors were used to calculate core states. A "core-only" CVS projector is applied to remove excitations that originate exclusively from valence orbitals. This is done by zeroing out all right and left vector elements that only involve molecular orbitals with energies greater than the energy of the lowest core molecular orbital of a given atom. This yields a Lanczos spectrum starting at the lowest core excitation energy of the chosen edge. A complementary "valence-only" CVS projector is used to obtain valence excited states that are orthogonal to the core excited states.

Except for the spectra presented in Fig. 2, only sufficiently converged valence and core band Lanczos vectors are used for calculating stationary states and corresponding Hamiltonian and transition moment matrices. This is done by discarding states with either right or left residual norms greater than 1×10^{-2} for valence states and 1×10^{-1} for core states.

In all calculations, valence states are calculated first, with starting vectors based on the ground state. All accepted valence states are, together with the ground state, used to construct starting vectors for the core state calculations (see Sec. II F). The energies and maximum transition strengths for all accepted valence and core states are given in the Supplemental Material [52].

A fixed pump-probe delay of 40 a.u. (about 0.968 fs) is used for lithium fluoride, lithium hydride, and ethylene. The delay is varied for glycine, in order to calculate the transient absorption of the molecule. In all calculations the central time of the probe pulse is set to zero and the negative central times of the pump pulses are set accordingly.

Unless otherwise stated, integration of the TD-EOM-CC and TDCC equations is done using a Dormand-Prince 5(4) integration scheme [53] with a maximum time step of 0.1 a.u., and maximum and minimum local errors of 1×10^{-7} and 1×10^{-9}, respectively (see Appendix B). Each component of the time-dependent dipole moment expectation value and electric-field vectors is multiplied with the Hann window before Fourier transformation.

All calculations are performed using a development version of the e^T program [54].

IV. RESULTS AND DISCUSSION

A. Lithium fluoride: Convergence and nonlinear pump interaction

When discussing the applicability of the band Lanczos algorithm for modeling attosecond pump-probe processes, a key question is how spectra are affected by the chain length used. With this in mind, a single TDCC LiF pump-probe absorption spectrum, calculated with the RK4 integrator and fixed time steps of 5×10^{-3} a.u., is in Fig. 2 compared to TD-EOM-CC spectra calculated with the Dormand-Prince 5(4) integration scheme and various band Lanczos chain lengths. In all calculations, the pulses have the parameters used for the LiF spectra in Sec. III B of Ref. [28], where the F K edge is targeted by the probe pulse. Figure 1 illustrates the polarization of the pulses relative to the orientation of the molecule. All states with energies inaccessible by the absorption of one photon from each pulse are discarded from the TD-EOM-CC calculations.

For lower chain lengths, the peaks of the band Lanczos spectra shown in Fig. 2 both shift and scale significantly with variations in the chain length, indicating that excitation energies and dipole matrix elements are badly converged. The convergence generally improves with the chain length, and low-energy high-amplitude peaks seem to converge first. Higher chain lengths are needed for good convergence of high-energy low-amplitude peaks, as expected from the convergence behavior of Lanczos algorithms.

As demonstrated, the inclusion of badly converged states can give spectral peaks with incorrect positions and amplitudes. In addition, these states can also increase the cost of matrix element calculation and propagation, decrease the convergence rate of consecutive band Lanczos calculations, and cause serious numerical instabilities during propagation. In order to avoid these adverse effects, states with badly con-

verged right or left vector residual norms will be discarded in the following band Lanczos calculations.

In Fig. 3, the aforementioned TDCC LiF pump-probe absorption spectrum is compared to TD-EOM-CC spectra from converged states only. Note that the three most dominant peaks in the TDCC spectrum are present in the green spectrum, which is calculated with a valence chain length of 300, but a chain length of 400 is needed in order to converge the short peak at around 9.9 eV. The low amplitude peaks below and around the tall peak at around 6.9 eV are missing.

In an earlier work [28], we speculated that the smaller peaks below 6.9 eV in the pump-only LiF spectrum could originate from two-photon absorption. This claim was later discussed by Pedersen *et al.* [32], where the TDCC state of LiF interacting with the pump pulse was analyzed in terms of stationary state populations. Their analysis supports the interpretation that two photons are absorbed from the pump pulse.

In order to take two-photon absorption into account, spectra are recalculated with the inclusion of valence states energetically accessible by two pump photons and core states accessible by an additional probe photon. The corresponding results obtained with chain lengths of 300 and 400 are shown in purple and red in Fig. 3, respectively. Note that the 300 valence chain length spectrum still lacks the smaller features of the TDCC spectrum, but the 400 valence chain length spectrum is practically indistinguishable from the TDCC one. This similarity corroborates the claim that two photons are absorbed from the pump pulse. Furthermore, the results demonstrate that reduced-basis TD-EOM-CC can faithfully reproduce TDCC results in particular systems, even when nonlinear interactions are involved. The embedded Dormand-Prince 5(4) integrator is seen to perform well for TD-EOM-CC.

The bottom panel of Fig. 3 demonstrates the use of the valence-only CVS projector to calculate the valence states. The approximation seems to improve the rate of convergence with respect to chain length, as a length of 300 is enough to retrieve all the features of the TDCC spectrum while a higher number is necessary in the nonprojected case. This improved convergence can be explained by the reduction in dimension from projecting out transitions from core orbitals. Moreover, since the approximation does not seem to lead to significant scaling or shifting of the valence peaks, it is adopted in the following calculations.

B. Lithium hydride: Applicability of the CVS projectors

To further assess the performance of the proposed procedure, as well as the applicability of the CVS projectors, we use the TD-EOM-CC procedure to model the interaction of the lithium hydride molecule with the pump-probe pair described in Sec. III A of Ref. [28]. Figure 1 illustrates the polarization of the pulses relative to the orientation of the molecule. Li K-edge spectra are notoriously difficult to describe accurately due to the small energy separation between the valence and core excitation regions. This can be considered a challenging test case for the applicability of the core-valence separation scheme.

A comparison between TD-EOM-CC and TDCC spectra is given in Fig. 4, where the latter is taken from Ref. [28]. In all core state calculations a fixed band Lanczos chain length of
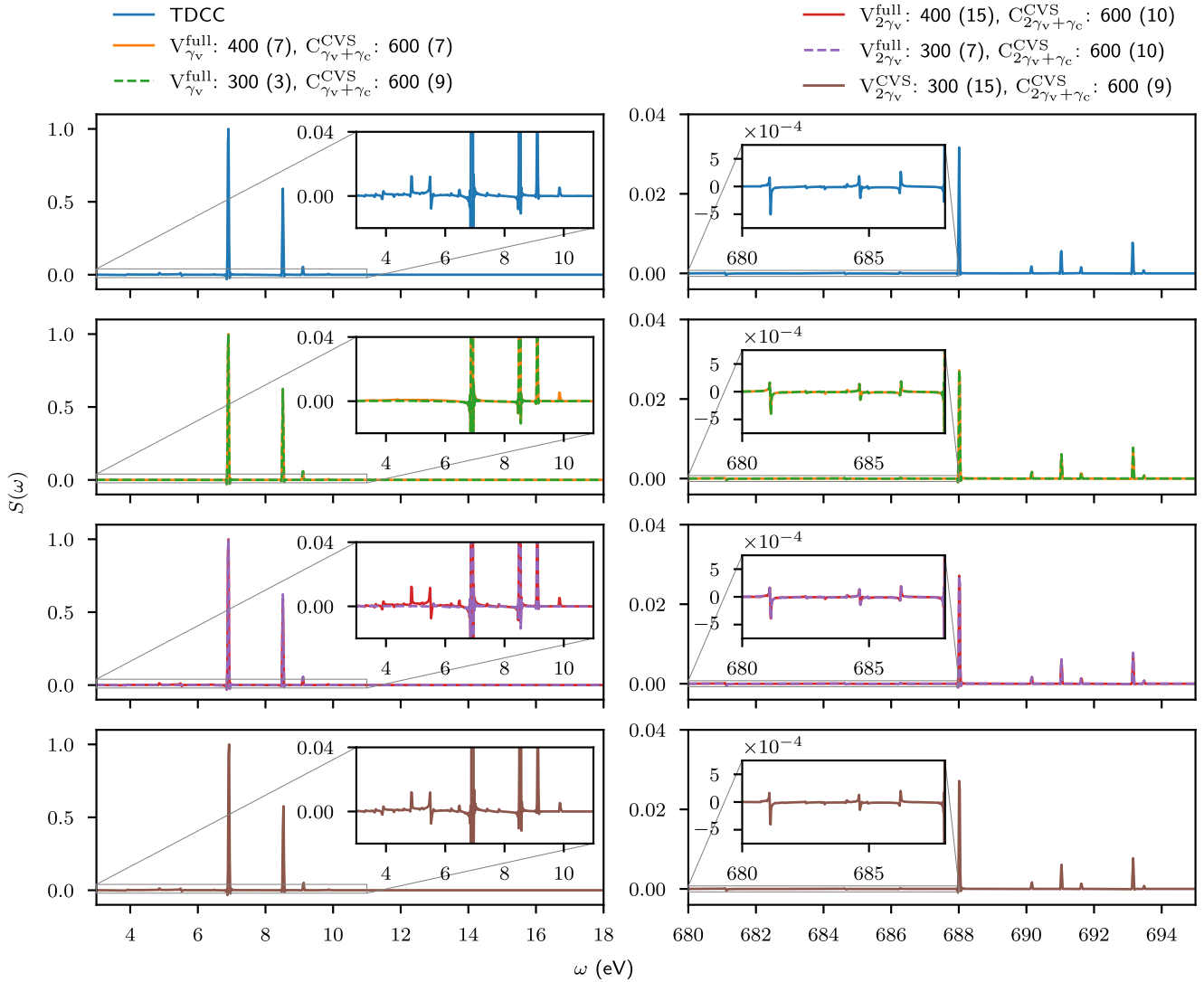
FIG. 3. LiF pump-probe absorption $S(\omega)$ as a function of frequency ω in the valence and core regions, normalized by the tallest peaks in the spectra. The TDCC results are shown in the top left and right panels. TD-EOM-CC results, calculated at different band Lanczos chain lengths, are shown in the lower panels. EOM-CC valence (V) states are calculated in the full projection space (middle panels) or within the CVS approximation (bottom panels), while the core (C) states are calculated within the CVS approximation only. Valence states energetically inaccessible by a single or two pump photons, γ_v or $2\gamma_v$, are discarded, and so are core states energetically inaccessible by subsequent absorption of a probe photon, $\gamma_v + \gamma_c$ or $2\gamma_v + \gamma_c$. The chain lengths of the calculations are given, together with the number of converged states (in brackets).

400 is used. However, the number of converged core states, given in brackets, differs due to the different starting vectors employed.

Since TD-EOM-CC with energy-limited valence and core states successfully reproduced the TDCC spectrum of LiF in Sec. IV A, a similar procedure is attempted for calculating the TD-EOM-CC LiH spectrum. That is, valence states inaccessible by two pump photons and core states inaccessible by an additional probe photon are discarded. The results are shown in orange in the second topmost panels of Fig. 4. The spectrum lacks some of the weaker features in the valence excitation energy region, and, more notably, many of the dominant features in the core excitation region. In other words, a characteristic of the LiH molecule seemingly prevents us from reproducing the TDCC spectrum using the procedure in

the previous section. In the following, we argue that the Li *K* edge in LiH involves states that cannot be obtained with the core-only CVS projector alone, since they do not correspond to core excitations.

TD-EOM-CC spectra calculated with states obtained with the valence-only CVS projector, energetically accessible by two pump and one probe photons, are shown in the three middle rows of panels in Fig. 4. The valence chain lengths used are 100 (green), 150 (red), and 200 (purple). The number of converged states in the valence region increases with the chain length. Remarkably, increasing the valence chain length also leads to additional peaks in the core region, illustrated in the right panels. This demonstrates that, apart from the two intense peaks obtained at about 54.1 eV and 57.7 eV, the other peaks are of pure valence excitation character.
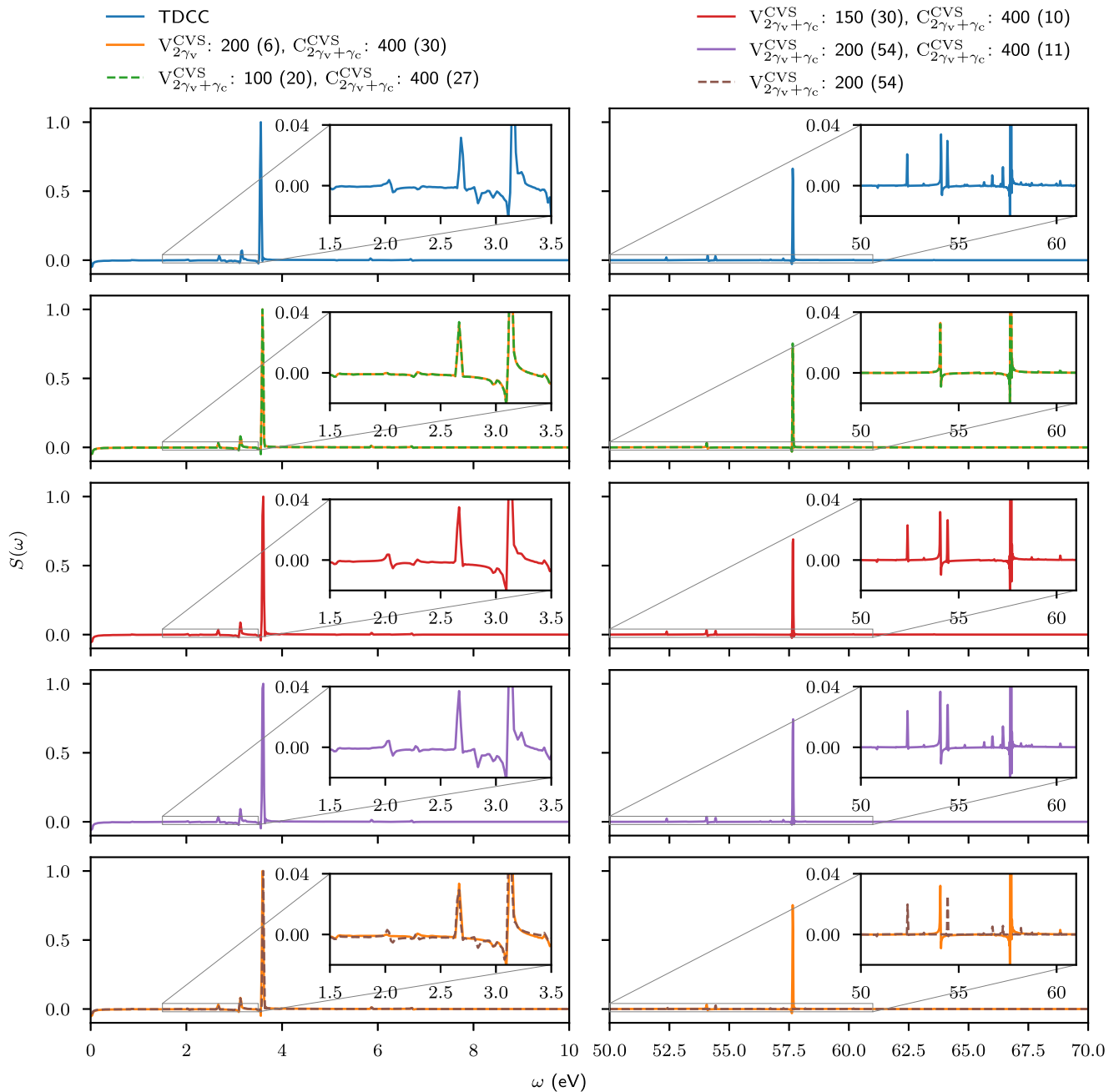
FIG. 4. LiH pump-probe absorption $S(\omega)$ as a function of frequency ω in the valence and core regions, normalized by the tallest peaks in the spectra. The TDCC results are shown in the top left and right panels. TD-EOM-CC results, calculated at different band Lanczos chain lengths, are shown in the lower panels. All EOM-CC valence (V) and core (C) states are calculated within the CVS approximation. For the CVS valence calculation for the results shown in orange, states energetically inaccessible by two pump photons, $2\gamma_v$ are discarded. For all other band Lanczos calculations, only the states inaccessible by two pump photons and one probe photon $2\gamma_v + \gamma_c$ are discarded. The results shown in brown are calculated from CVS valence states only. The chain lengths of the calculations are given, together with the number of converged states (in brackets).

The necessity to include high-energy states calculated with the valence-only CVS projector is further validated by superimposing a spectrum exclusively from valence-only CVS states (brown) with the spectrum calculated from energy-limited valence and core states (orange), shown in the bottom panels of Fig. 4. The peaks of the composite spectrum are in good agreement with the TDCC ones. Therefore, we assert that use of both the core-only and the complementary valence-only CVS projectors is necessary in order to accurately capture the spectral features around the Li K edge in LiH. Note that this should not be taken as a failure of the CVS projectors, but as a consequence of the peculiar electronic structure of LiH. In fact, the high-energy states of pure valence character can be more easily calculated in the dimension reduced by the valence-only CVS projector. One may still question whether the corresponding peaks will be as
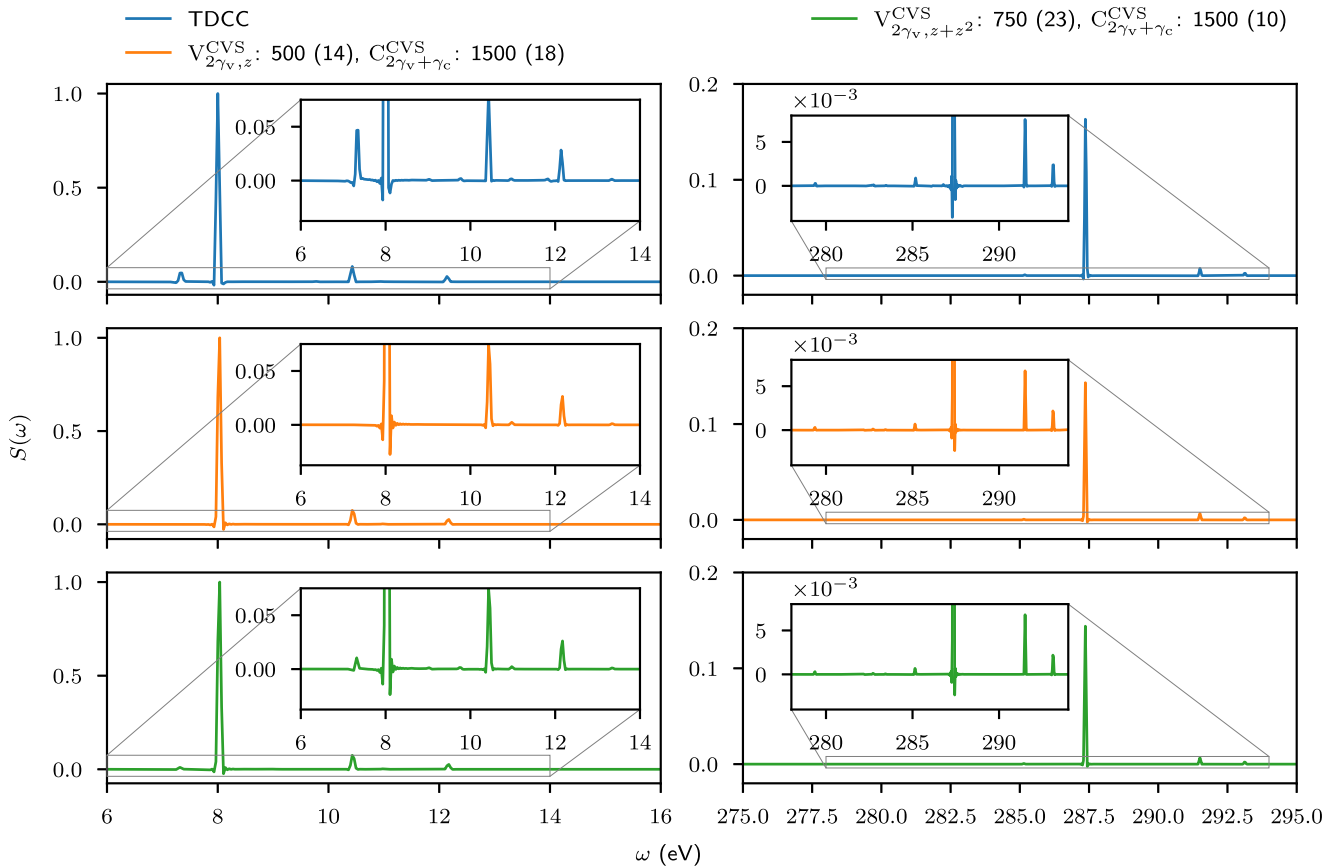
FIG. 5. Ethylene pump-probe absorption $S(\omega)$ as a function of frequency ω in the valence and core regions, normalized by the tallest peaks in the spectra. The TDCC results are shown in the top left and right panels, and TD-EOM-CC results are shown in the lower panels. In the middle panels, a ground-state z dipole operator starting vector is used for calculating the valence (V) states. In the bottom panels, both ground-state z dipole operator and z^2 quadrupole operator starting vectors are used. All valence and core (C) states are calculated within the CVS approximation. The chain lengths of the calculations are given, together with the number of converged states (in brackets).

prominent in experimental spectra, since continuum electrons are very crudely represented in the chosen basis set.

C. Ethylene: Nonlinear pump interaction for a different symmetry group

For ethylene, rms widths of the z-polarized pump and x-polarized probe pulses are set to 10 a.u. and 5 a.u., corresponding to intensity full width at half maximum (FWHM) durations of about 403 as and 201 as, respectively. Figure 1 illustrates the polarization of the pulses relative to the orientation of the molecule. The carrier frequency of the pump pulse is set to 0.294 114 89 a.u. (about 8.0 eV), and the probe pulse is set to 10.495 830 66 a.u. (about 285.6 eV, C K edge). The field strengths of the pump and probe pulses are set to 0.01 a.u. and 0.1 a.u., respectively. The time-dependent state is propagated with the Dormand Prince 5(4) integration scheme, from -2500 a.u. to 2500 a.u. of time. The TDCC spectrum, shown in the top panel Fig. 5, is characterized by four dominant peaks in the valence excitation region. A low amplitude peak at around 7.3 eV is present in the TDCC spectrum, but missing in the TD-EOM-CC spectrum calculated with a z dipole operator starting vector, shown in the middle panels. In accordance with the interpretation of the spectrum of LiF in Sec. IV B, we attribute the missing peak to a two-photon excitation process,

even though valence states energetically accessible by two photons are included. Note that quadratic functions of the z dipole operator belong to the A_g representation of D_{2h}, the point group of ethylene for the chosen geometry. Hence, we should not expect the single starting vector, belonging to the B_{1u} representation, to facilitate the convergence of the two-photon peaks.

In order to mimic the two pump photon absorption process, we include a starting vector constructed from the z^2 quadrupole operator in the valence-state calculation. The results, shown in the bottom panels of Fig. 5, now capture the two-photon peak at around 7.3 eV. The amplitude yielded by TD-EOM-CC is, however, underestimated compared to the TDCC one, which might indicate that more secondary valence excited states should be included in the computation. It might also hint at differences in two-photon absorption as described by TD-EOM-CC and TDCC.

D. Glycine: Transient absorption

As a final example, we use the computational procedure to model attosecond transient absorption by the glycine molecule. The polarization of the pump pulse is set to the polarization of the EOM-CC transition dipole moment between the ground and first dipole allowed valence excited
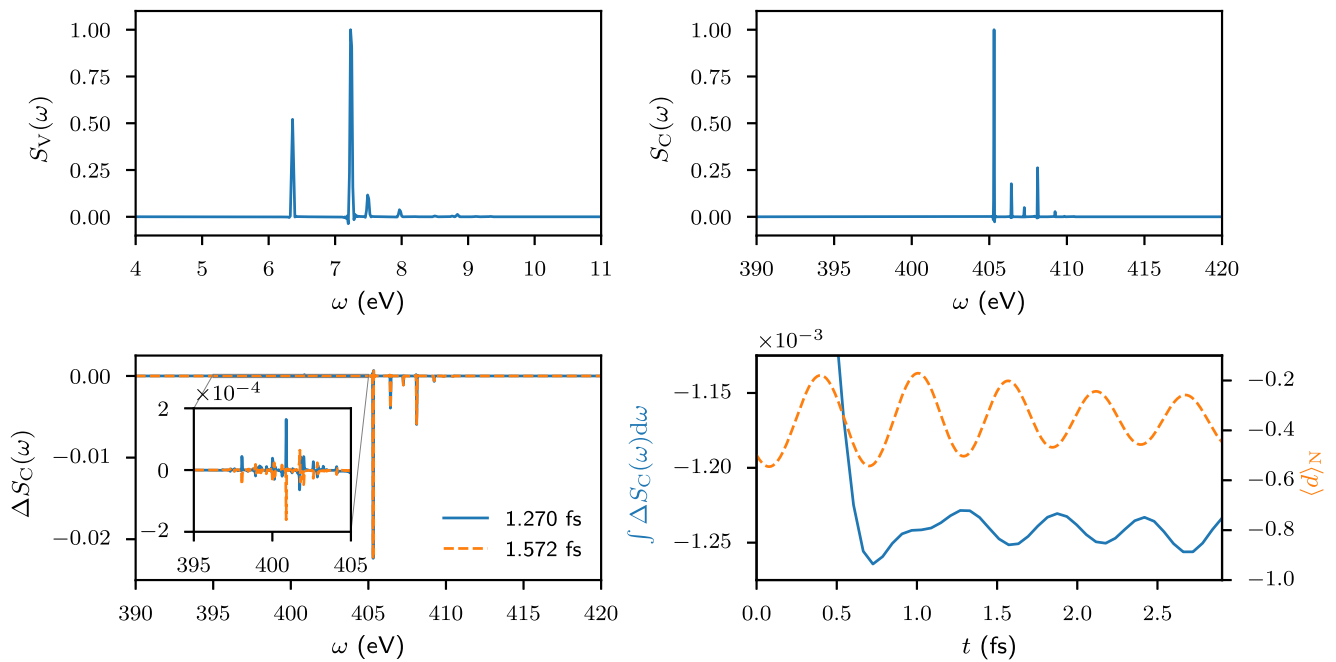
FIG. 6. Glycine. Top left: Pump pulse absorption from the ground state $S_V(\omega)$, normalized by the tallest peak in the spectrum. Top right: Probe pulse absorption from the ground state $S_C(\omega)$, normalized by the tallest peak in the spectrum. Bottom left: Pump-probe absorption minus the probe absorption from the ground state $\Delta S_C(\omega) = S(\omega) - S_C(\omega)$, normalized by the tallest peak in the $S_C(\omega)$ spectrum. The results for two different pump-probe delays are shown. Bottom right: In blue, the numerically integrated probe absorption difference $\int \Delta S_C(\omega) d\omega$ is shown as a function of pump-probe delay. In orange, the dipole induced by the pump pulse in the direction from the center of mass to the N atom, $\langle d \rangle_N$, is shown as a function of time after the center of the pump pulse.

state, $(0.490\,072x + 0.871\,682y)$, which is in the mirror plane of the molecule (xy plane). The probe is z polarized. Figure 1 illustrates the polarization of the pulses relative to the orientation of the molecule. The rms widths of the pump and probe pulses are set to 20 a.u. and 10 a.u., corresponding to intensity FWHM durations of about 806 as and 403 as, respectively. The carrier frequency of the pump pulse is set to 0.233 683 25 a.u. (about 6.4 eV), and the probe pulse is set to 14.894 573 19 a.u. (about 405.3 eV, N K edge). The field strengths of the pump and probe pulses are set to 0.01 a.u. and 0.1 a.u., respectively. The time-dependent state is propagated with the Dormand Prince 5(4) integration scheme, from −5000 a.u. to 5000 a.u. of time.

A single band Lanczos calculation is used for constructing the valence states, where states energetically inaccessible by two-photon transitions are discarded. Note that we do not need to use quadrupole operators in order to get two-photon valence states of glycine in the reduced basis, as otherwise done for the ethylene valence states, since both linear and quadratic functions of x and y dipole operators belong to the A' representation of C$_s$. A valence-state calculation, with ground-state starting vectors and a chain length of 1500, gives 17 converged states. A subsequent core state calculation, with ground- and valence-state starting vectors and a chain length of 3000, gives 20 converged states.

As a note of caution, all the converged valence states have energies below 10.5 eV, which is below the double frequency of the carrier photons. This indicates that two-photon absorption is not properly accounted for by the reduced basis, and

nonlinear features involving higher-energy valence states may be missing in the spectra.

In the top left panel of Fig. 6, the absorption of the pump pulse is shown as a function of frequency. Even though the glycine molecule is substantially larger than the other molecules considered, the spectrum is still dominated by a small number of peaks. The number of dominant peaks is also smaller than the number of converged states in the basis (=17). The spectrum of the absorption of the probe pulse by the ground state, shown in the top right panel, also has fewer dominant peaks than the number of converged states (=20).

In order to calculate the transient absorption of the probe pulse by the glycine molecule, absorption spectra are calculated with the pump-probe setup used for the other molecules, with pump-probe delays varying from 0 a.u. to 120 a.u. (about 2.903 fs) in intervals of 2.5 a.u. The reduced basis energies and dipole matrix elements do not have to be recalculated between the different TD-EOM-CC calculations, since these are independent of the pump-probe delay. Difference spectra are then calculated by subtracting the ground-state probe absorption spectrum from the pump-probe absorption spectra, before normalizing by the tallest peak in the ground-state probe spectrum. In the bottom left panel, the difference spectra at 52.5 a.u. and 65 a.u. (about 1.270 fs and 1.572 fs) are shown in blue and orange, respectively. Both spectra are dominated by negative peaks, indicative of ground-state bleaching. In addition, the spectra vary slightly with pump-probe delay, which is particularly visible for the peaks that are not energetically accessible from the ground state, e.g., in the energy range from 395 eV to 405 eV shown in the inset.

In order to quantify the delay-dependent difference in the absorption of the full probe pulse, each of the 49 pump-probe difference spectra are numerically integrated from 390 eV to 420 eV using the trapezoidal rule. The results are shown as function of pump-probe delay in the bottom right panel of Fig. 6, in blue. Note that the absorption difference is smaller for short pump-probe delays, which can be explained by the fact that ground-state bleaching happens gradually during the pump pulse interaction. A shorter pump-probe delay implies that the molecule is probed while bleaching still occurs, which can lead to a smaller difference between the pump-probe absorption and the ground-state probe absorption.

We have also calculated the pump-induced time-dependent dipole moment in the direction from the center of mass to the center of the N atom, as a way of quantifying the migration of charge between the end containing the N atom and the opposite end of the molecule. The dipole moment is shown in the bottom right panel of Fig. 6, in orange.

Note that the dominant periods of both the time-dependent dipole moment and the integrated absorption (after 1 fs), shown in the bottom right panel of Fig. 6, fall within 0.57(4) fs. This indicates that the pump-induced TD-EOM-CC state is a coherent superposition dominated by states with energy differences of 7.3(6) eV, which is in agreement with the ground-state pump absorption spectrum (top left). It also indicates that the dominant features of the time-dependent charge migration and the delay-dependent K-edge absorption are correlated and can be measured with phase-controlled pulses with finite duration, as has previously been demonstrated for instantaneous pulses [55,56].

V. CONCLUSION

We have demonstrated the use of the asymmetric band Lanczos algorithm to generate reduced TD-EOM-CC bases for various molecules, taking the characteristics of pulses suitable for probing attosecond phenomena into account. The specific starting vectors used in the calculations direct the band Lanczos algorithm towards states that are useful for representing the interactions. The starting vectors also allow for the affordable calculation of transition strengths, which are used, together with excitation energies, to automatically select the reduced basis. The basis is further reduced by removing unconverged states.

In Sec. IV A, we demonstrated how lithium fluoride spectral peaks can converge towards peaks calculated with TDCC by increasing the band Lanczos chain length and taking a sufficient number of relevant states into account. In particular, we showed that two-photon absorption has to be taken into account in order to reproduce the smaller features of the TDCC spectrum, as speculated in Ref. [28].

In Sec. IV B, we demonstrated that the core-only CVS projector eliminates several of the peaks around the K edge of lithium in lithium hydride. The missing peaks can be captured with the complementary valence-only CVS projector, which enabled us to target high-energy states of pure valence character. This observation indicates that care should be taken when the CVS scheme is used for light elements such as lithium, where the energy separation of the core and valence orbitals is small, so that pure valence excitations can fall within the region of core excitations.

In Sec. IV C, we used starting vectors constructed from both dipole and quadrupole operators, in order to converge ethylene valence states that are dark with respect to one-photon transitions from the ground state.

In Sec. IV D, pump and probe pulses with varying time delays were used to assess the transient absorption of a K-edge probe pulse as a function of pump-probe delay. We showed how the transient absorption seems to correlate with the migration of charge induced by the pump, and how both quantities seem to reveal the dominant timescale in the coherent superposition.

APPENDIX A: REWRITING EOM-CC MATRIX ELEMENT EXPRESSIONS

The matrix element X_{ij} of the operator X and the left and right vectors of EOM-CC states i and j, respectively, can be written as

$$
\begin{aligned}
X_{ij} &= \langle \widetilde{\psi_i} | X | \psi_j \rangle \\
&= \sum_{\kappa\lambda} l_{i\kappa} \langle \kappa | \bar{X} | \lambda \rangle r_{\lambda j} \\
&= \sum_{\nu} l_{i0} \langle \mathrm{HF} | \bar{X} | \nu \rangle r_{\nu j} + \sum_{\mu\nu} l_{i\mu} \langle \mu | \bar{X} | \nu \rangle r_{\nu j} \\
&\quad + \left(l_{i0} X_{00} + \sum_{\mu} l_{i\mu} \xi_{\mu}^{X} \right) r_{0j} \\
&= \sum_{\nu} l_{i0} \langle \mathrm{HF} | [\bar{X}, \tau_{\nu}] | \mathrm{HF} \rangle r_{\nu j}
\end{aligned}
$$

$$+ \sum_{\mu\nu} l_{i\mu}\big(^{\mathrm{LR}}A_{\mu\nu}^X + \langle\mu|\tau_\nu\bar{X}|\mathrm{HF}\rangle\big)r_{\nu j}$$
$$+ \Big(l_{i0}X_{00} + \sum_\mu l_{i\mu}\xi_\mu^X\Big)r_{0j}, \tag{A1}$$

where

$$^{\mathrm{LR}}A_{\mu\nu}^X = \langle\mu|[\bar{X},\tau_\nu]|\mathrm{HF}\rangle, \tag{A2}$$

$$X_{00} = \langle\mathrm{HF}|\bar{X}|\mathrm{HF}\rangle, \tag{A3}$$

$$\xi_\mu^X = \langle\mu|\bar{X}|\mathrm{HF}\rangle. \tag{A4}$$

1. Excited-state left and ground-state right vectors

The left vector of excited state m has the reference component $l_{m0} = 0$ [see Eq. (23)], while the right vector of the ground state has components $r_{00} = 1$ and $r_{\nu 0} = 0$ [see Eq. (18)]. Inserting this into Eq. (A1), we obtain

$$X_{m0} = \sum_\mu l_{m\mu}\xi_\mu^X, \tag{A5}$$

which is also the expression appearing in CC response theory [42]. Note that this expression is linear in the excited determinant components $l_{m\mu}$.

2. Ground-state left and excited-state right vectors

The left vector of the ground state has the components $l_{00} = 1$ and $l_{0\mu} = \bar{t}_\mu$ [see Eq. (19)], while the right vector of excited state n has the reference component $r_{0n} = -\sum_\nu \bar{t}_\nu r_{\nu n}$ [see Eq. (22)]. Inserting this into Eq. (A1), we obtain

$$X_{0n} = \sum_\nu \Big(^{\mathrm{LR}}\eta_\nu + \sum_\mu \bar{t}_\mu\langle\mu|\tau_\nu\bar{X}|\mathrm{HF}\rangle\Big)r_{\nu n}$$
$$- \Big(X_{00} + \sum_\mu \bar{t}_\mu\xi_\mu^X\Big)\sum_\nu \bar{t}_\nu r_{\nu n}, \tag{A6}$$

where

$$^{\mathrm{LR}}\eta_\nu = \langle\mathrm{HF}[\bar{X},\tau_\nu]\mathrm{HF}\rangle + \sum_\mu \bar{t}_\mu{}^{\mathrm{LR}}A_{\mu\nu}^X. \tag{A7}$$

The term $\sum_\nu {}^{\mathrm{LR}}\eta_\nu r_{\nu n}$ appears in CC response theory [42], and the other terms are specific to EOM-CC. Equation (A6) is equivalent to Eq. (65) in Ref. [57] and can also be written as

$$X_{0n} = \sum_\nu \big(^{\mathrm{EOM}}\eta_\nu^X - X_{00}\bar{t}_\nu\big)r_{\nu n}, \tag{A8}$$

where

$$^{\mathrm{EOM}}\eta_\nu^X = {}^{\mathrm{LR}}\eta_\nu^X + \sum_\mu \bar{t}_\mu\langle\mu|\tau_\nu\bar{X}|\mathrm{HF}\rangle - \Big(\sum_\mu \bar{t}_\mu\xi_\mu^X\Big)\bar{t}_\nu \tag{A9}$$

[see Eq. (18) of Ref. [58]]. Note that Eq. (A8) is linear in the excited determinant components $r_{\nu n}$.

3. Excited-state left and right vectors

The left vector of excited state m has the reference component $l_{m0} = 0$, while the right vector of excited state n has $r_{0n} = -\sum_\nu \bar{t}_\nu r_{\nu n}$ [see Eqs. (22) and (23)]. Inserting this into Eq. (A1), we obtain

$$X_{mn} = \sum_{\mu\nu} l_{m\mu}\big(^{\mathrm{LR}}A_{\mu\nu}^X + \langle\mu|\tau_\nu\bar{X}|\mathrm{HF}\rangle\big)r_{\nu n}$$
$$- \sum_\mu l_{m\mu}\xi_\mu^X \sum_\nu \bar{t}_\nu r_{\nu n}$$
$$= \sum_{\mu\nu} l_{m\mu}\big(^{\mathrm{LR}}A_{\mu\nu}^X + \langle\mu|\tau_\nu\bar{X}|\mathrm{HF}\rangle - \xi_\mu^X\bar{t}_\nu\big)r_{\nu n}$$
$$= \sum_{\mu\nu} l_{m\mu}\big(^{\mathrm{EOM}}A_{\mu\nu}^X + \delta_{\mu\nu}X_{00} - \xi_\mu^X\bar{t}_\nu\big)r_{\nu n}, \tag{A10}$$

where

$$^{\mathrm{EOM}}A_{\mu\nu}^X = \langle\mu|\bar{X}|\nu\rangle - \delta_{\mu\nu}X_{00}$$
$$= {}^{\mathrm{LR}}A_{\mu\nu}^X + \langle\mu|\tau_\nu\bar{X}|\mathrm{HF}\rangle - \delta_{\mu\nu}X_{00} \tag{A11}$$

[see Eq. (20) in Ref. [58]].

The term $\sum_{\mu\nu} l_{m\mu}{}^{\mathrm{LR}}A_{\mu\nu}^X r_{\nu n}$ appears in CC response theory, and the other terms are specific to EOM-CC [44,58]. Note that all matrix elements in Eq. (A10) are linear in both $l_{m\mu}$ and $r_{\nu n}$.

APPENDIX B: INTEGRATION SCHEME

In order to limit the error of the time-dependent results, the integration of the TDCC and TD-EOM-CC equations is done with the embedded Dormand-Prince method of order 5(4) [53]. This method yields both fourth- and fifth-order accurate solutions at each time step, and is specified by the Butcher tableau [53,59]

$$
\begin{array}{c|ccccccc}
0 & & & & & & & \\
\frac{1}{5} & \frac{1}{5} & & & & & & \\
\frac{3}{10} & \frac{3}{40} & \frac{9}{40} & & & & & \\
\frac{4}{5} & \frac{44}{45} & -\frac{56}{15} & \frac{32}{9} & & & & \\
\frac{8}{9} & \frac{19372}{6561} & -\frac{25360}{2187} & \frac{64448}{6561} & -\frac{212}{729} & & & \\
1 & \frac{9017}{3168} & -\frac{355}{33} & \frac{46732}{5247} & \frac{49}{176} & -\frac{5103}{18656} & & \\
1 & \frac{35}{384} & 0 & \frac{500}{1113} & \frac{125}{192} & -\frac{2187}{6784} & \frac{11}{84} & \\
\hline
 & \frac{35}{384} & 0 & \frac{500}{1113} & \frac{125}{192} & -\frac{2187}{6784} & \frac{11}{84} & 0 \\
 & \frac{5179}{57600} & 0 & \frac{7571}{16695} & \frac{393}{640} & -\frac{92097}{339200} & \frac{187}{2100} & \frac{1}{40}
\end{array}
\tag{B1}
$$

where the next-to-last and last rows give the coefficients of the fifth- and fourth-order solutions, respectively. Although the method has seven stages, its *first same as last* property assures that only six function evaluations are needed per time step.

The Euclidean distance between the solutions gives a fourth-order estimate of the local integration error:

$$\epsilon_{O(4)} = ||\mathbf{y}_{O(5)} - \mathbf{y}_{O(4)}||_2. \tag{B2}$$

This local error estimate is kept below a given maximum value by adapting the time step during the integration. The fifth-order solution is accepted as the solution at the beginning of the next step whenever the error estimate satisfies this condition.

The following adaptive time-stepping scheme was designed, implemented, and used together with the Dormand-Prince 5(4) method for the relevant calculations in Sec. IV. At the start of the integration, the step size is set to a given maximum value. During the integration, the variable step size is halved, and the integration step redone, whenever the error estimate exceeds the given maximum error. After a successful integration step, the step size is doubled whenever the error estimate is below a given minimum value, provided that the doubled step size is smaller than the maximum step size and also a submultiple of the elapsed time. This is in order to increase the efficiency of the integration while ensuring that the solution is evaluated at times corresponding to integer increments of the maximum time-step size. Evaluation of time-dependent observables is done using the solutions at these integer increments.

[1] M. T. Hassan, T. T. Luu, A. Moulet, O. Raskazovskaya, P. Zhokhov, M. Garg, N. Karpowicz, A. M. Zheltikov, V. Pervak, F. Krausz, and E. Goulielmakis, Nature (London) **530**, 66 (2016).

[2] M. Galli, V. Wanie, D. P. Lopes, E. P. Månsson, A. Trabattoni, L. Colaizzi, K. Saraswathula, A. Cartella, F. Frassetto, L. Poletto, F. Légaré, S. Stagira, M. Nisoli, R. M. Vázquez, R. Osellame, and F. Calegari, Opt. Lett. **44**, 1308 (2019).

[3] D. Fabris, T. Witting, W. A. Okell, D. J. Walke, P. Matia-Hernando, J. Henkel, T. R. Barillot, M. Lein, J. P. Marangos, and J. W. G. Tisch, Nat. Photon. **9**, 383 (2015).

[4] J. Duris, S. Li, T. Driver, E. G. Champenois, J. P. MacArthur, A. A. Lutman, Z. Zhang, P. Rosenberger, J. W. Aldrich, R. Coffee, G. Coslovich, F.-J. Decker, J. M. Glownia, G. Hartmann, W. Helml, A. Kamalov, J. Knurr, J. Krzywinski, M.-F. Lin, J. P. Marangos *et al.*, Nat. Photon. **14**, 30 (2020).

[5] A. Malyzhenkov, Y. P. Arbelo, P. Craievich, P. Dijkstal, E. Ferrari, S. Reiche, T. Schietinger, P. Juranić, and E. Prat, Phys. Rev. Research **2**, 042018(R) (2020).

[6] X. Li, N. Govind, C. Isborn, A. E. DePrince, and K. Lopata, Chem. Rev. **120**, 9951 (2020).

[7] B. F. E. Curchod and T. J. Martínez, Chem. Rev. **118**, 3305 (2018).

[8] E. F. Kjønstad and H. Koch, J. Chem. Theory Comput. **17**, 127 (2021).

[9] S. E. Koonin, K. T. R. Davies, V. Maruhn-Rezwani, H. Feldmeier, S. J. Krieger, and J. W. Negele, Phys. Rev. C **15**, 1359 (1977).

[10] P. Bonche, B. Grammaticos, and S. Koonin, Phys. Rev. C **17**, 1700 (1978).

[11] H. Flocard, S. E. Koonin, and M. S. Weiss, Phys. Rev. C **17**, 1682 (1978).

[12] M. R. Provorse and C. M. Isborn, Int. J. Quantum Chem. **116**, 739 (2016).

[13] C. Lian, M. Guan, S. Hu, J. Zhang, and S. Meng, Adv. Theory Simul. **1**, 1800055 (2018).

[14] H. Miyagi and L. B. Madsen, Phys. Rev. A **87**, 062511 (2013).

[15] T. Sato and K. L. Ishikawa, Phys. Rev. A **88**, 023402 (2013).

[16] H. Miyagi and L. B. Madsen, Phys. Rev. A **89**, 063416 (2014).

[17] P. Krause, T. Klamroth, and P. Saalfrank, J. Chem. Phys. **123**, 074105 (2005).

[18] J. A. Sonk, M. Caricato, and H. B. Schlegel, J. Phys. Chem. A **115**, 4678 (2011).

[19] E. Luppi and M. Head-Gordon, Mol. Phys. **110**, 909 (2012).

[20] A. E. DePrince, M. Pelton, J. R. Guest, and S. K. Gray, Phys. Rev. Lett. **107**, 196806 (2011).

[21] L. Cederbaum and J. Zobeley, Chem. Phys. Lett. **307**, 205 (1999).

[22] M. Ruberti, Faraday Discuss. **228**, 286 (2021).

[23] D. R. Nascimento and A. E. DePrince, J. Phys. Chem. Lett. **8**, 2951 (2017).

[24] L. N. Koulias, D. B. Williams-Young, D. R. Nascimento, A. E. DePrince, and X. Li, J. Chem. Theory Comput. **15**, 6617 (2019).

[25] S. Kvaal, J. Chem. Phys. **136**, 194109 (2012).

[26] C. Huber and T. Klamroth, J. Chem. Phys. **134**, 054113 (2011).

[27] D. A. Pigg, G. Hagen, H. Nam, and T. Papenbrock, Phys. Rev. C **86**, 014308 (2012).

[28] A. S. Skeidsvoll, A. Balbi, and H. Koch, Phys. Rev. A **102**, 023115 (2020).

[29] Y. C. Park, A. Perera, and R. J. Bartlett, J. Chem. Phys. **151**, 164117 (2019).

[30] T. B. Pedersen and S. Kvaal, J. Chem. Phys. **150**, 144106 (2019).

[31] H. E. Kristiansen, Ø. S. Schøyen, S. Kvaal, and T. B. Pedersen, J. Chem. Phys. **152**, 071102 (2020).

[32] T. B. Pedersen, H. E. Kristiansen, T. Bodenstein, S. Kvaal, and Ø. S. Schøyen, J. Chem. Theory Comput. **17**, 388 (2021).

[33] Y. C. Park, A. Perera, and R. J. Bartlett, J. Chem. Phys. **155**, 094103 (2021).

[34] M. Wu, S. Chen, S. Camp, K. J. Schafer, and M. B. Gaarde, J. Phys. B **49**, 062003 (2016).

[35] T. Chen, J. Demmel, M. Gu, Y. Saad, R. Lehoucq, D. Sorensen, K. Maschhoff, Z. Bai, D. Day, R. Freund, G. Sleijpen, H. van der Vorst, and R. Li, in *Templates for the Solution of Algebraic Eigenvalue Problems* (SIAM, Philadelphia, 2000), Chap. 7, pp. 149–231.

[36] I. H. Godtliebsen and O. Christiansen, Phys. Chem. Chem. Phys. **15**, 10035 (2013).

[37] R. W. Freund, Acta Numer. **12**, 267 (2003).

[38] N. Shimizu, T. Mizusaki, Y. Utsuno, and Y. Tsunoda, Comput. Phys. Commun. **244**, 372 (2019).

[39] J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. I: Theory* (SIAM, Philadelphia, 2002).

[40] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 4th ed. (Johns Hopkins University, Baltimore, MD, 2013).

[41] Y. Saad, *Iterative Methods for Sparse Linear Systems* (SIAM, Philadelphia, 2003).

[42] S. Coriani, T. Fransson, O. Christiansen, and P. Norman, J. Chem. Theory Comput. **8**, 1616 (2012).

[43] S. Coriani, O. Christiansen, T. Fransson, and P. Norman, Phys. Rev. A **85**, 022507 (2012).

[44] B. N. C. Tenorio, M. A. C. Nascimento, A. B. Rocha, and S. Coriani, J. Chem. Phys. **151**, 184106 (2019).

[45] T. Moitra, S. Coriani, and B. N. C. Tenorio, Mol. Phys. **119**, e1980235 (2021).

[46] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database No. 101 (Release 20 August 2019).

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

[48] D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572 (1995).

[49] L. S. Cederbaum, W. Domcke, and J. Schirmer, Phys. Rev. A **22**, 206 (1980).

[50] S. Coriani and H. Koch, J. Chem. Phys. **143**, 181103 (2015).

[51] B. N. C. Tenorio, T. Moitra, M. A. C. Nascimento, A. B. Rocha, and S. Coriani, J. Chem. Phys. **150**, 224104 (2019).

[52] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.105.023103 for the energies and maximum transition strengths for all EOM-CC states used in the TD-EOM-CC calculations in Sec. IV.

[53] J. Dormand and P. Prince, J. Comput. Appl. Math. **6**, 19 (1980).

[54] S. D. Folkestad, E. F. Kjønstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, T. Moitra, A. C. Paul, M. Scavino, A. S. Skeidsvoll, Å. H. Tveten, and H. Koch, J. Chem. Phys. **152**, 184103 (2020).

[55] A. D. Dutoi, K. Gokhberg, and L. S. Cederbaum, Phys. Rev. A **88**, 013419 (2013).

[56] A. D. Dutoi and L. S. Cederbaum, Phys. Rev. A **90**, 023414 (2014).

[57] H. Koch, R. Kobayashi, A. Sanchez de Merás, and P. Jørgensen, J. Chem. Phys. **100**, 4393 (1994).

[58] R. Faber and S. Coriani, J. Chem. Theory Comput. **15**, 520 (2019).

[59] A. Iserles, in *A First Course in the Numerical Analysis of Differential Equations*, 2nd ed., Cambridge Texts in Applied Mathematics (Cambridge University Press, New York, 2008), pp. 105–122.

Coupled cluster simulation of impulsive stimulated x-ray Raman scattering

Alice Balbi,[1, *] Andreas S. Skeidsvoll,[2, *] and Henrik Koch[1, 2, †]

[1]*Scuola Normale Superiore, Piazza dei Cavalieri, 7, I-56126, Pisa, Italy*

[2]*Department of Chemistry, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

(Dated: June 1, 2023)

Time-dependent equation-of-motion coupled cluster (TD-EOM-CC) is used to simulate impulsive stimulated x-ray Raman scattering (ISXRS) of ultrashort laser pulses by neon, carbon monoxide, pyrrole, and p-aminophenol. The TD-EOM-CC equations are expressed in the basis of field-free EOM-CC states, where the calculation of the core-excited states is simplified through the use of the core-valence separation (CVS) approximation. The transfer of electronic population from the ground state to the core- and valence-excited states is calculated for different numbers of included core- and valence-excited states, as well as for electric field pulses with different polarizations and carrier frequencies. The results indicate that Gaussian pulses can transfer significant electronic populations to the valence states through the Raman process. The sensitivity of this population transfer to the model parameters is analyzed. The time-dependent electronic density for p-aminophenol is also showcased, supporting the interpretation that ISXRS involves localized core excitations and can be used to rapidly generate valence wavepackets.

I. INTRODUCTION

The ability to experimentally generate short and intense x-ray laser pulses has been a subject of significant interest in the field of x-ray science. Recent technological advances, specifically the realization of x-ray free electron lasers (XFELs) [1, 2] and new approaches based on high harmonic generation (HHG) [3, 4], have made it possible to generate x-ray laser pulses with high intensities and pulse durations as short as a few hundred and even tens of attoseconds [5]. This progress has enabled the development of new experimental techniques with unprecedented temporal resolution, facilitating the imaging and control of atoms and molecules on the time scale of electronic motion. [6–12] An important phenomenon in this context is impulsive stimulated x-ray Raman scattering (ISXRS), which is the extension of stimulated x-ray Raman scattering (SXRS) to the impulsive limit, where the duration of the external field interaction is short compared to the time scales of the subsequent evolution of the system.

In general, Raman scattering is a light-matter interaction phenomenon in which photons trigger an excitation of an atomic or molecular system followed by a deexcitation to an energy level different from the initial one. In the context of x-ray Raman scattering, the involved transitions are electronic in character. [6, 13–16] We focus on the situation in which the electronic excitation in play is a core excitation, which is deexcited to a valence-excited state through the decay of a valence electron into a core vacancy, see Fig. 1. Core excitations are often localized on a specific atomic site and sensitive to the surrounding electronic environment, making them useful for the local initiation of charge migration. We treat the case where both the excitation and deexcitation are stimulated by an interaction with the same laser pulse. [17]
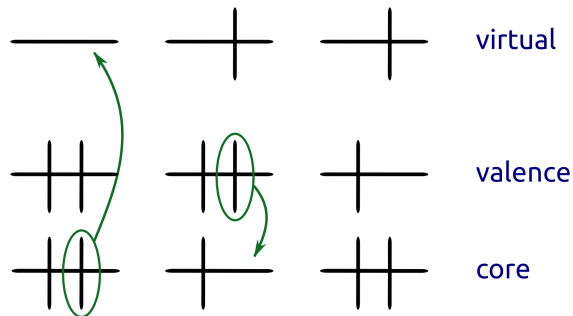

FIG. 1. Illustration of the steps in the ISXRS process. Initially, the molecule is in its ground state (left). An external x-ray pulse excites a core electron, leading to a core-excited state (middle). The same pulse can trigger the decay of a valence electron into the core vacancy, leading to a valence-excited state (right).

This is achievable by utilizing a pulse with sufficient bandwidth to encompass the energy differences between the ground state and the core-excited states of interest, as well as between these core-excited states and the final valence-excited states. The interaction with such pulses is similar to the interactions occurring in the first experimental demonstration of electronic population transfer via ISXRS, which was made for the NO molecule at the Linac Coherent Light Source as recently as in 2020. [18]

The progress in experimental techniques has stimulated the development of methods for modeling electron dynamics based on the time-dependent Schrödinger equation. Real-time methods, which involve solving this equation in the real time domain, offer a particularly suitable approach for analyzing ultrafast phenomena. [19] Among these methods, real-time coupled coupled cluster methods offer high accuracy and computational costs that scale polynomially with the system size.

---

* These authors contributed equally to this work.

† Electronic mail: henrik.koch@sns.it

The time evolution is described by differential equations that can be solved using standard numerical integration techniques such as Runge-Kutta methods.

A specific subcategory of real-time coupled cluster methods is the time-dependent coupled cluster (TDCC) methods, [20–28] where the time dependence is parametrized by cluster amplitudes and Lagrange multipliers. [29, 30] These methods offer the advantage of size-extensivity at all levels of truncation. Another subcategory, the time-dependent equation-of-motion coupled cluster (TD-EOM-CC) methods, [31–36] provides less potential for numerical issues compared to TDCC methods, [37] since the time dependence is parametrized by the linear coefficients used in EOM-CC methods and the cluster amplitudes remain fixed at their time-independent ground state values. [38–40]

In the basis of field-free EOM-CC states, the TD-EOM-CC method requires the predetermination of the excited states that are involved in the studied processes. Computationally, the exterior eigenvalue algorithms usually employed for calculating valence-excited states are inefficient for the calculation of the core-excited states often involved in x-ray interactions. This is because the core-excited states have large eigenvalues, and the states are embedded in an ionization (pseudo-)continuum. [41] A useful scheme for the study of core excitations is the core valence separation (CVS) scheme, which disregards all excitations that do not involve at least one core orbital. [42, 43] This allows for the approximate core-excited states to be calculated as the lowest energy states within the reduced excitation space.

In this article, we use the TD-EOM-CC method together with the CVS approximation to simulate the interaction of neon, carbon monoxide, pyrrole, and p-aminophenol with ultrashort laser pulses, and calculate the populations of the valence-excited states following ISXRS targeting molecular K-edges. The article is organized as follows. In Section II we briefly outline the theory behind the calculations. We provide details of the performed computations in Section III, and present and discuss the results in Section IV. Conclusions are presented in Section V.

II. THEORY

The time-dependent system is described by the Hamiltonian

$$H(t) = H^{(0)} + V(t), \qquad (1)$$

where $H^{(0)}$ is the electronic Hamiltonian of the molecule in the Born-Oppenheimer approximation. We describe the interaction with the external laser field $V(t)$ in the dipole approximation and length gauge,

$$V(t) = -\boldsymbol{d} \cdot \boldsymbol{\mathcal{E}}(t), \qquad (2)$$

where $\boldsymbol{d}$ is the vector of Cartesian dipole operators, and $\boldsymbol{\mathcal{E}}(t)$ the Cartesian electric field vector.

The eigenstates of the field-free Hamiltonian,

$$|\psi_j\rangle = \sum_\lambda e^T |\lambda\rangle \, r_{\lambda j} \qquad (3)$$

$$\langle \psi_i| = \sum_\kappa l_{i\kappa} \langle\kappa| \, e^{-T} \qquad (4)$$

can be found by first solving the ground state coupled cluster equations

$$\langle\mu| \, e^{-T} H^{(0)} e^T \, |\mathrm{HF}\rangle = 0, \qquad (5)$$

which determine the cluster amplitudes t_μ in the cluster operator,

$$T = \sum_\mu t_\mu \tau_\mu. \qquad (6)$$

Thereafter, the right and left vectors can be found as eigenvectors of the projected time-independent Schrödinger equation,

$$\sum_\lambda \langle\kappa| \, e^{-T} H^{(0)} e^T \, |\lambda\rangle \, r_{\lambda j} = r_{\kappa j} E_j, \qquad (7)$$

$$\sum_\kappa l_{i\kappa} \langle\kappa| \, e^{-T} H^{(0)} e^T \, |\lambda\rangle = E_i l_{i\lambda}. \qquad (8)$$

These equations lead to the following eigenvalue problems [44]

$$\mathbf{A}\mathbf{R}_j = \mathbf{R}_j \Delta E_j, \qquad (9)$$

$$\mathbf{L}_i^T \mathbf{A} = \Delta E_i \mathbf{L}_i^T, \qquad (10)$$

where $A_{\mu\nu} = \langle\mu| \, e^{-T} \big[H^{(0)}, \tau_\nu \big] e^T \, |\mathrm{HF}\rangle$, $L_{i\mu} = l_{i\mu}$ and $R_{\nu j} = r_{\nu j}$ for $\mu > 0$ and $\nu > 0$. The excitation energy $\Delta E_j = E_j - E_0$ is given as the difference between the excited state energy and the ground state energy

$$E_0 = \langle\mathrm{HF}| \, e^{-T} H e^T \, |\mathrm{HF}\rangle . \qquad (11)$$

The TD-EOM-CC ket and bra states can be expanded in the field-free EOM-CC kets and bras, $|\Psi(t)\rangle = \sum_j |\psi_j\rangle \, c_j(t)$ and $\langle\widetilde{\Psi}(t)| = \sum_i b_i(t) \langle\widetilde{\psi}_i|$. This gives the TD-EOM-CC equations [45]

$$i\frac{\mathrm{d}c_i(t)}{\mathrm{d}t} = \sum_j H_{ij}(t)c_j(t), \qquad (12)$$

$$-i\frac{\mathrm{d}b_j(t)}{\mathrm{d}t} = \sum_i b_i(t)H_{ij}(t), \qquad (13)$$

where $H_{ij}(t) = \langle\widetilde{\psi}_i| H(t) |\psi_j\rangle = \delta_{ij} E_j + \langle\widetilde{\psi}_i| V(t) |\psi_j\rangle$. The time-dependent population of EOM-CC state i in the TD-EOM-CC superposition state can be found as the product of the projections onto the ket and bra of the EOM-CC state,

$$\begin{aligned} P_i(t) &= \langle\widetilde{\Psi}(t)|\psi_i\rangle \, \langle\widetilde{\psi}_i|\Psi(t)\rangle \\ &= b_i(t)c_i(t). \end{aligned} \qquad (14)$$

The eigenvalues of core-excited states are interior to the spectrum of the molecular Hamiltonian, and often hard to reach using exterior eigenvalue methods like Davidson or Lanczos algorithms. The core-valence separation (CVS) approximation [42, 46] simplifies the calculation of these states by removing the valence-core and core-valence blocks of the Hamiltonian and has become a vital tool for the calculation of NEXAFS spectra. [41] Let I denote the set indexing the core orbitals. We invoke the CVS approximation through a projector $\mathcal{P}_I^{\mathrm{CVS}}$ that removes all vector elements that do not reference excitations from at least one core orbital, in each eigensolver iteration. [43] For the coupled cluster singles and doubles (CCSD) truncation level, this can be expressed in compact form as

$$\mathcal{P}_I^{\mathrm{CVS}} r_i^a = l_i^a \mathcal{P}_I^{\mathrm{CVS}} = 0, \quad i \notin I \tag{15}$$

$$\mathcal{P}_I^{\mathrm{CVS}} r_{ij}^{ab} = l_{ij}^{ab} \mathcal{P}_I^{\mathrm{CVS}} = 0, \quad i \notin I \wedge j \notin I. \tag{16}$$

This projection is effectively setting all elements of the valence-valence block of the full-space elementary basis EOM-CC Jacobian matrix $\mathbf{A}$ to zero, giving the CVS approximated Jacobian matrix, $\mathbf{A}^{\mathrm{CVS}}$. The core-excited EOM-CC states obtained in the CVS approximation can have a non-zero overlap with EOM-CC states obtained without invoking this approximation. The CVS states are in general also not eigenstates of the full field-free Jacobian, and can lead to TD-EOM-CC populations that are non-stationary, complicating the interpretation of the TD-EOM-CC state. To ensure that the populations are stationary, we diagonalize the Jacobian $\mathbf{A}$ in the basis of all the CVS and non-CVS (valence) states by first constructing the Jacobian and overlap matrices

$$A_{ij} = \boldsymbol{L}_i \mathbf{A} \boldsymbol{R}_j, \quad S_{ij} = \boldsymbol{L}_i \boldsymbol{R}_j. \tag{17}$$

respectively in the reduced space. Assuming linear independence of the vectors in the basis, the solution of the generalized eigenvalue problem defined by $\mathbf{A}$ and $\mathbf{S}$ gives a new set of right and left eigenvectors of $\mathbf{A}$, which preserve populations when there is no interaction with the external field.

III.   COMPUTATIONAL DETAILS

The electric field in Eq. (2) is represented as

$$\boldsymbol{\mathcal{E}}(t) = \boldsymbol{\mathcal{E}}_0 \cos(\omega_0(t - t_0) + \phi) f(t), \tag{18}$$

where $\boldsymbol{\mathcal{E}}_0$ is the peak electric field of the pulse in its polarization direction, ω_0 the carrier frequency and t_0 the central time of the pulse, and ϕ is the carrier-envelope phase. The envelope function $f(t)$ is chosen to have the Gaussian shape

$$f(t) = \begin{cases} e^{-(t-t_0)^2/(2\sigma^2)}, & -c \leq t \leq c, \\ 0, & \text{otherwise}, \end{cases} \tag{19}$$

where the RMS width is selected as $\sigma = 0.5$ and the envelope truncated at $c = 8\sigma$. In all calculations, we use the carrier-envelope phase $\phi = 0$ and the peak electric field strength $|\boldsymbol{\mathcal{E}}_0|$ of 10 a.u., which corresponds to the maximum intensity of $7.019 \times 10^{18}\,\mathrm{W\,cm^{-2}}$, calculated from the intensity relation $S_0 = |\boldsymbol{\mathcal{E}}_0|^2/Z_0$ where Z_0 is the impedance of free space.

All simulations are performed using a development version of the eT program [47] containing the TD-EOM-CC implementation described in Ref. [45]. The Runge-Kutta method known as RK4 is used to integrate Eq. (12) and Eq. (13), with time steps of 0.001 a.u. for neon, carbon monoxide, and p-aminophenol and 0.0001 a.u. for pyrrole.

IV.   RESULTS AND DISCUSSION

A.   Neon

In the following, the convergence properties of the final Raman-induced populations are investigated for the neon atom. This system is used for benchmarking purposes, as its small size allows for the use of larger basis sets. We focus on the convergence of the final population of the $B_v {}^1D$ valence-excited state, the lowest valence-excited state with a significant final population.

We first study the basis set convergence with respect to the cardinal number X of Dunning basis sets for CCS and CCSD levels of theory. The employed basis sets are cc-pVDZ, aug-cc-pVXZ (with X=D,...,6) and aug-cc-pCVXZ (with X=D,...,5). As the carrier frequency ω_0 of the electric field, we choose the average of two frequencies. The first frequency corresponds to the transition between the ground state $X {}^1S$ and the $B_c {}^1P$ core-excited state. The second frequency corresponds to the transition between the $B_c {}^1P$ core-excited state and the $B_v {}^1D$ valence-excited state. The $B_c {}^1P$ and $B_v {}^1D$ states are chosen as they are, respectively, the lowest core-excited and valence excited states that get significantly populated in the Raman process, except for the cc-pVDZ basis set, where the order of $A_c {}^1S$ and $B_c {}^1P$ energy levels is inverted. In these calculations, we include 4 core-excited states and 12 valence-excited states. The frequencies used for the different basis sets and levels of theory are given in the Supporting Information.

From Fig. 2, we can observe how the final populations calculated with CCS and CCSD are considerably different, implying that CCS is not accurate enough to provide an adequate description of the system. The addition of functions for describing core correlation (aug-cc-pCVXZ) leads to slightly lower final populations compared to the corresponding basis sets without these functions (aug-cc-pVXZ). For CCSD, the results for 5Z and 6Z are very similar, implying that basis-set convergence is reached for 5Z. Continuing, the convergence of the final population of the $B_v {}^1D$ state is explored with respect to the number of valence- and core-excited states included in the calculation. The total of the probabilites of all de-
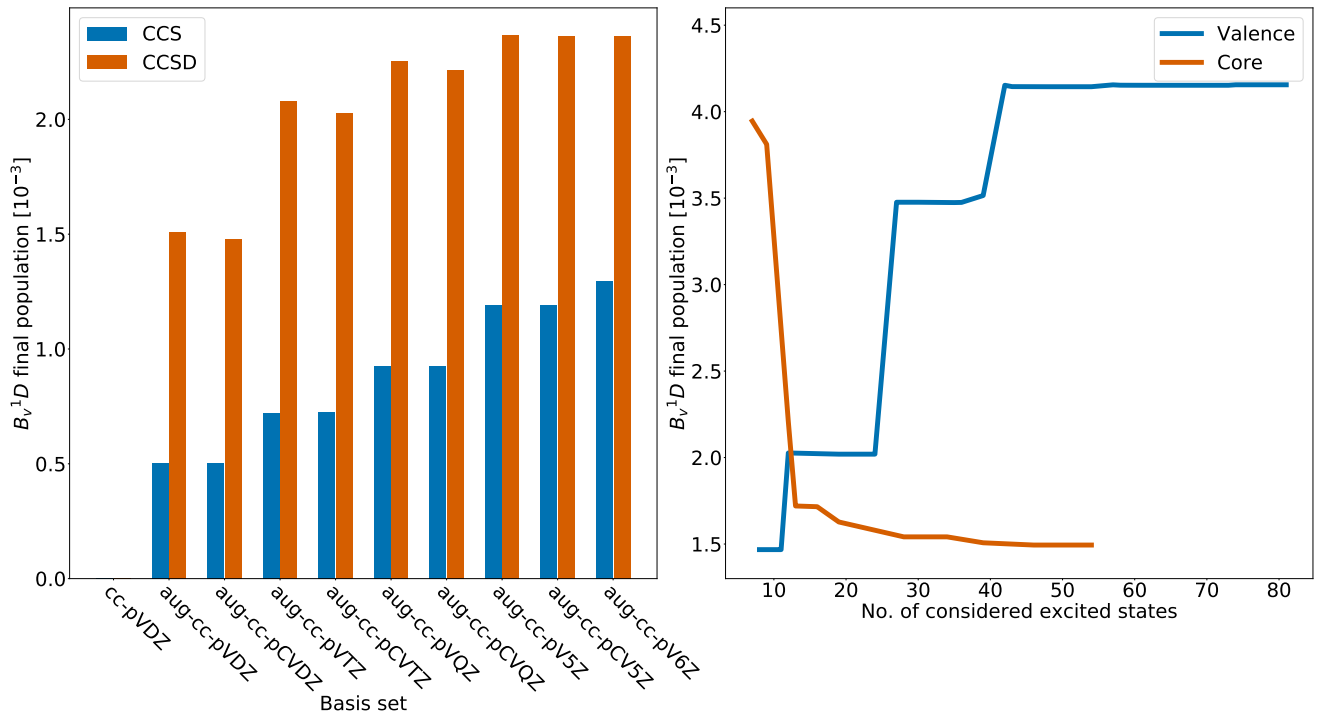
FIG. 2. The left panel shows the final population of the $B_v\,^1D$ states of neon for different choices of level of theory and basis set. The blue line in the right panel shows the final population of the same $B_v\,^1D$ states for different numbers of valence-excited states included in the simulation and the number of core-excited states fixed at 4, calculated with CCSD/aug-cc-pCVTZ. The red line in the right panel shows the final population of the same $B_v\,^1D$ states for different numbers of core-excited states included in the simulation and the number of valence-excited states fixed at 79, calculated with CCSD/aug-cc-pCVTZ.

generacies of a state is calculated, such that for instance the probabilities for the five degenerate states of D type are added together. We perform the calculations using the CCSD truncation level and the aug-cc-pCVTZ basis set. The right panel of Fig. 2 exhibits the convergence of the final population of the $B_v\,^1D$ states with respect to the number of valence-excited states included in the simulation, with the number of core-excited states fixed at 4. The results indicate that more than 40 valence-excited states are needed for convergence. An analogous procedure is performed, this time keeping the number of valence-excited states fixed while varying the number of core-excited states. In the right panel of Fig. 2, we can see how the final population of $B_v\,^1D$ starts to converge after around 15 core-excited states are included in the calculation.

B. CO

We continue by simulating ISXRS for the carbon monoxide molecule, which is linear and belongs to the $C_{\infty v}$ symmetry point group. Since the system is not centrally symmetric, results can differ depending on the polarization of the electric field. Theoretical and experimental studies of the core-excitation spectroscopy and ISXRS of this molecule have previously been car-

ried out. [48] In our simulations, the distance between the two nuclei is fixed at $1.128\,\text{Å}$, corresponding to the equilibrium bond length in the NIST database. [49]. The internuclear axis of the molecule is aligned along the z-axis and the carbon atom is placed at the origin of the coordinate system while the oxygen atom is placed at $1.128\,\text{Å}$ along the z-axis. The carrier frequency of the external electric field is again chosen as the average between two frequencies. The first is the transition frequency between the ground state and the first core-excited state, which is the lowest-energy core-excited state that gets significantly populated during the Raman process. The second is the frequency of transition between this core-excited state and the third valence-excited state, which is the lowest valence-excited state that gets significantly populated. For CCS/aug-cc-pCVTZ, the frequency is $20.029\,089\,E_{\text{h}}$, while for CCSD/aug-cc-pCVTZ it is $19.504\,022\,E_{\text{h}}$, corresponding to the O K-edge.

To investigate transitions at the C K-edge, we choose the lowest-energy molecular orbital localized on the carbon atom as the molecular orbital used in the CVS approximation. The carrier frequency of the electric field is again chosen as average of the transition frequencies between the ground state and the lowest core-excited state that is significantly populated, and that between that core-excited state and the lowest valence-excited state that is significantly populated, resulting in a carrier fre-
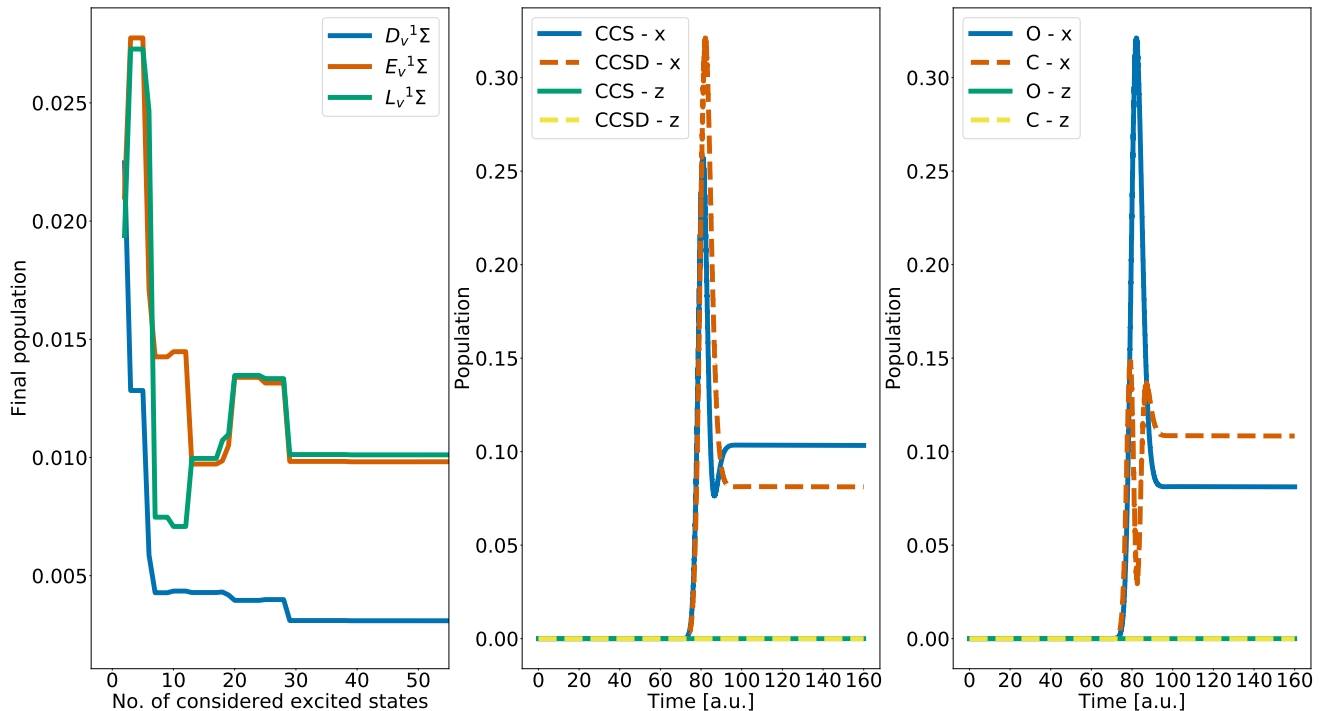
FIG. 3. The left panel shows the final population of the $D_v{}^1\Sigma$, $E_v{}^1\Sigma$, and $L_v{}^1\Sigma$ valence-excited state of carbon monoxide for different numbers of core-excited states included in the calculation, with the number of valence states fixed at 20 and the external electric field polarization in the positive z-direction. The central panel shows the time-dependent population of the third valence-excited state of carbon monoxide, calculated with the aug-cc-pCVTZ basis set and different levels of theory and electric field polarizations. The right panel shows the time-dependent population of the third valence-excited state for external electric fields tuned to different K-edges and with different polarizations, calculated with CCSD/aug-cc-pCVTZ.

quency of $10.402\,530\,E_\mathrm{h}$

In the carbon monoxide system, linearly polarized electric fields can be decomposed into two components: the polarization component parallel to the internuclear axis (along the z-axis) and the polarization component perpendicular to it (any direction in the xy-plane). As for neon, the convergence of the final population of certain valence-excited states is assessed with respect to the number of included core-excited states. The results are shown for the $D_v{}^1\Sigma$, $E_v{}^1\Sigma$ and $L_v{}^1\Sigma$ valence-excited states in the left panel of Fig. 3, demonstrating that convergence is attained by increasing the number of considered core-excited states. About 30 core-excited states are needed for convergence when the number of valence-excited states is fixed at 20. In the central panel of the figure, we can see how the time-dependent population of the third valence-excited state depends on the polarization of the electric field and level of theory, and also how the population is constant after the interaction with the field. The final population is exactly zero when the polarization is along the z-axis, as expected from the symmetry of the molecule and field. In the right panel, we can see how the time-dependent population of the third valence-excited state differs when the carrier frequency of the electric field is tuned to the K-edge of different elements (C or O). For the different tunings, the third

valence-excited state is reached through different transition pathways, involving other transition frequencies and transition moments. As for the results in the central panel, the population is exactly zero when the electric field is polarized along the z-axis, irrespective of the chosen frequency, for symmetry reasons. The populations are also constant after the interaction with the field

C. Pyrrole

We further increase the complexity of the modeled system by considering pyrrole, which belongs to the C_{2v} symmetry point group. The geometry of the molecule is obtained from the NIST database. [49], for which the molecule lies in the yz-plane and the symmetry axis is along the z-axis. The Supporting Information provides the geometry of the system, along with a figure that shows its orientation relative to the Cartesian coordinate axes. The final populations after the Raman process are assessed for the electric field polarization vector set equal to $(1, 0, 0)$ $(1, 1, 0)$ and $(1, 1, 1)$ in the chosen coordinate system. The Raman process involving the N K-edge is studied by performing calculations at the CCSD level of theory with aug-cc-pCVDZ for the nitrogen atom and aug-cc-pVDZ for the other atoms. The carrier frequency
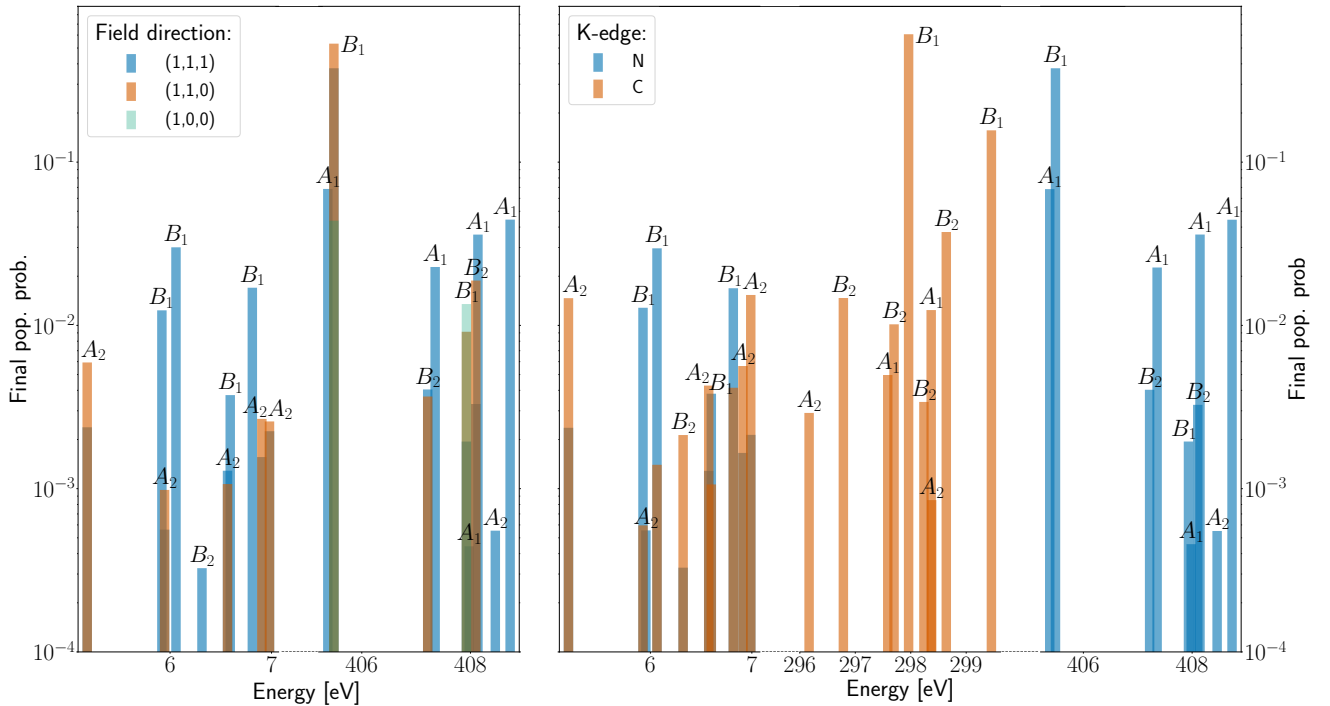
FIG. 4. The left panel displays the final populations of various excited states of pyrrole following ISXRS with different electric field polarizations, computed at the CCSD level of theory and with the aug-cc-pCVDZ basis set for the nitrogen atom and the aug-cc-pVDZ basis set for the other atoms. The right panel displays the final populations of different excited states of pyrrole for electric fields tuned to different K-edges, computed at the CCSD level of theory and with the aug-cc-pCVDZ basis set for the atom with the targeted K-edge shown in the inset and aug-cc-pVDZ basis set for the remaining atoms.

of the external electric field is chosen as the frequency of transition from the ground state to the most populated core-excited state, which is $14.901\,363\,E_\text{h}$. The Raman process involving the C K-edge is studied by performing calculations at the CCSD level of theory with aug-cc-pCVDZ for the carbon atoms and aug-cc-pVDZ for the other atoms. The core-excited states are calculated by using the CVS approximation restricted to the molecular orbital with the second-lowest energy. The carrier frequency of the external electric field is set to $10.949\,885\,E_\text{h}$, which is the transition frequency from the ground state to the fifth core-excited state, the lowest-energy core-excited state that is the most populated.

In the left panel of Fig. 4, we can see that new valence-excited states are populated as the polarization of the external electric field changes from $(1,0,0)$, to $(1,1,0)$, and to $(1,1,1)$. In particular, when the electric field is only polarized along the x-axis, there are no excitation to the valence-excited states. When the electric field has components along all three axes, all considered valence-excited states have a nonzero final population. An intermediate situation occurs when the electric field has components along both the x- and y-axes but not along the z-axis. This is since the different polarizations of the external electric field has components in different numbers of irreducible representation, enabling transitions to electronic states belonging to different irreducible repre-

sentations. In the right panel of Fig. 4, we can see how the final population of valence-excited states differs when the carrier frequency of the electric field is tuned to the N K-edge and C K-edge, calculated using the CVS approximation with the lowest- and next-to-lowest-energy molecular orbitals, respectively. In both cases, the polarization vector of the field is set to $(1,1,1)$. The valence-excited states that become populated are the same for the two K-edge frequencies, while the populations of the states are different.

D. p-aminophenol

Finally, we consider the planar p-aminophenol molecule. The molecule belongs to the C_s symmetry point group, which only contains the mirror plane and the identity as symmetry elements. This molecule is chosen in order to investigate if charge migration between the functional groups located at the opposite side of the aromatic ring can be observed, as the electronic charge can easily travel along the aromatic electron cloud. [50]

Compared to the systems analyzed previously, which offer only limited potential for charge migration due to their small sizes, the p-aminophenol molecule is a larger system containing two strongly electron donor substituents (amino and hydroxyl) on a benzene ring. [51].
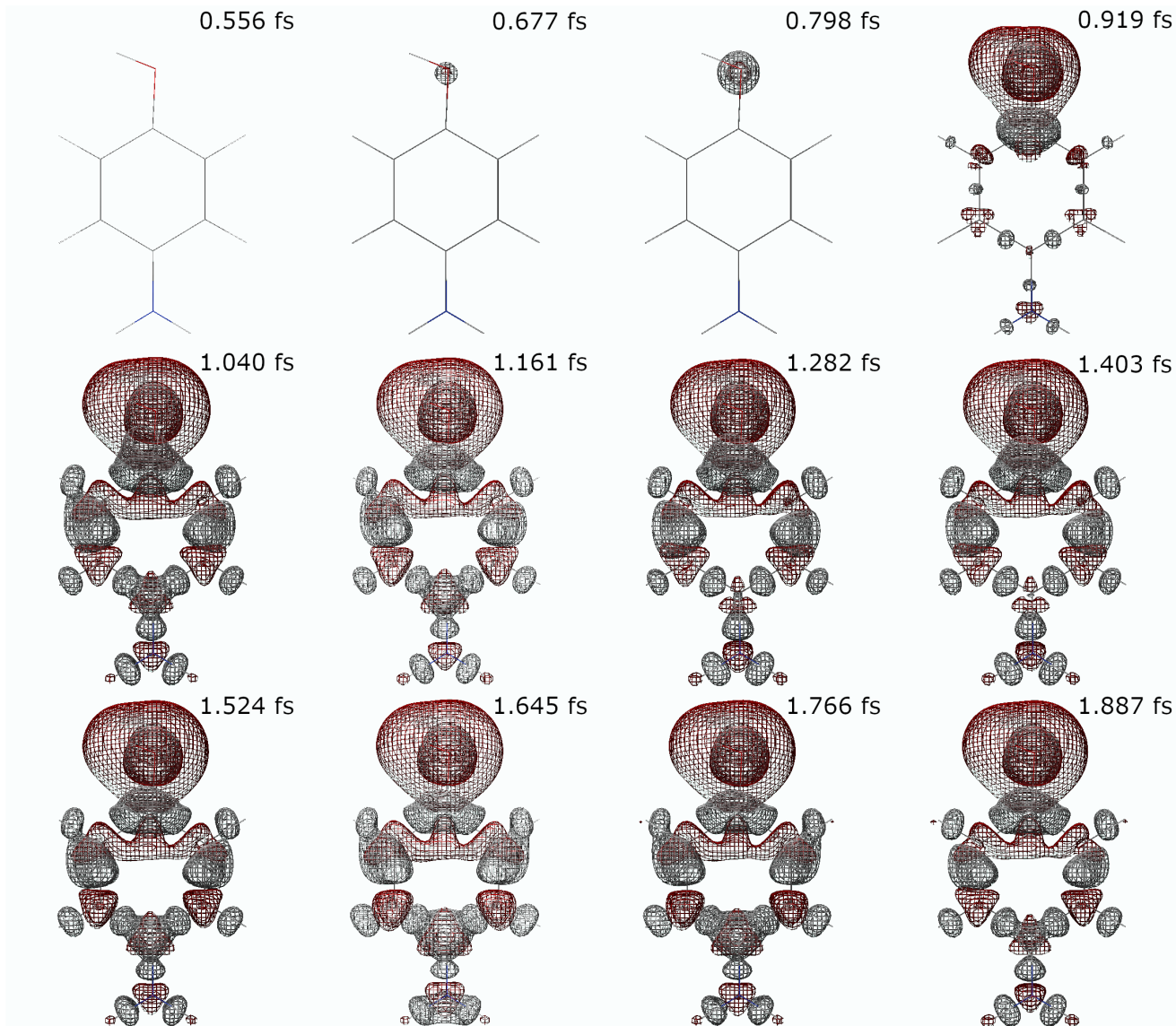
FIG. 5. Positive (gray) and negative (red) electronic isodensity surfaces of the time-dependent density after subtracting the ground state density of p-aminophenol, at the times specified at the top right corner of each subfigure. The structure of the p-aminophenol molecule is also shown in each subfigure.

We can thus expect a localized excitation to be followed by long-range charge migration.

The geometry of p-aminophenol is calculated at the B3LYP/aug-cc-pVDZ level of theory, and the molecule is placed in the xy-plane. The Supporting Information includes the geometry and a figure that illustrates the orientation of the molecule relative to the Cartesian coordinate axes. For the subsequent calculations, aug-cc-pCVDZ is used for the oxygen atom and aug-cc-pVDZ for all other atoms. The carrier frequency is chosen as $19.883\,479\,E_\mathrm{h}$, which corresponds to the frequency of transition from the ground state to the fourth core-excited state, which is the most populated state among the two lowest-energy core-excited states that have a non-

zero population after the Raman process.

In Fig. 5, the charge migration is illustrated through isodensity surfaces of the time-dependent density after subtracting the ground state density, calculated at different points in time. After the interaction with the external electromagnetic pulse, we can observe how the core excitation of the oxygen atom is reflected in a positive charge arising around that nucleus, enclosed in a negatively charged region at a bigger distance from the oxygen nucleus. This is followed by an alternating pattern of regions with increased or decreased electronic charge throughout the entire benzene ring up to the nitrogen atom of the amino group. In particular, the atoms of the ring gain some negative charge while the bonds be-

come more positively charged, and the bonds are thus expected to be weakened. Finally, we can observe how the nitrogen atom becomes negatively charged. As predicted, we observe a localized excitation at the hydroxyl substituent following oxygen K-edge excitation, followed by long-range charge migration, in accordance with what one could expect from a superposition of valence-excited states generated by ISXRS.

In the supplemental material we have included a movie that shows the temporal evolution of the electronic density depicted through isodensity surfaces of the time-dependent density difference, illustrating how the density oscillates after the interaction with the external electric field. The generation of electronic wavepackets with external laser pulses is interesting from an experimental point of view, as it represents the first step of controlling chemical reactions with laser pulses.

V. CONCLUSION

In this work, a time-dependent equation-of-motion coupled cluster model of ISXRS has been presented. First, we assessed the convergence of the final population of neon valence states with respect to different calculation parameters: the level of coupled cluster theory, the choice of basis set, and choices of the total number of valence- and core-excited states. We observed how the adequate description of the system required a proper representation of correlation and a sufficiently flexible basis set, since the CCS level of theory and basis sets without augmentation performed poorly. We also demonstrated that convergence of the population of a valence-excited state of neon was achieved when increasing the number of valence- and core-excited states for the given level of theory and basis set. Subsequently, the final populations of carbon monoxide states were assessed with respect to the number of included core-excited states. The results showed convergence for several valence-excited states for the given level of theory and basis set.

Furthermore, we demonstrated that the final populations of states of both carbon monoxide and pyrrole are significantly affected by the polarization of the external electric field, as symmetry can enable and forbid the transition to some of the excited states within the bandwidth of the pulse. We also assessed how the results were affected by tuning the external electric field to the K-edge of the different atoms, where the frequencies were calculated with the CVS approximation targeting the core molecular orbitals of the atoms. We observed how a different choice of K-edge led to changes in final populations as the final states were reached through different transition pathways.

After investigating ISXRS by neon, carbon monoxide, and pyrrole, we studied the time evolution of the electronic density of p-aminophenol. The ground-state density was subtracted from the time-dependent density, and the density difference was visualized through isodensity surfaces in real space. We observed the rapid formation of a valence wavepacket and subsequent charge migration in the molecule. Simulations of field-induced charge migration in molecular systems can be used to predict how chemical reactions can be controlled by external electric fields, which we believe will be a subject of further interest in the near future.

[1] C. Pellegrini, A. Marinelli, and S. Reiche, The physics of x-ray free-electron lasers, Rev. Mod. Phys. **88**, 015006 (2016).

[2] C. Pellegrini, The development of XFELs, Nat. Rev. Phys. **2**, 330 (2020).

[3] A. McPherson, G. Gibson, H. Jara, U. Johann, T. S. Luk, I. McIntyre, K. Boyer, and C. K. Rhodes, Studies of multiphoton production of vacuum-ultraviolet radiation in the rare gases, J. Opt. Soc. Am. B **4**, 595 (1987).

[4] M. Ferray, A. L'Huillier, X. Li, L. Lompre, G. Mainfray, and C. Manus, Multiple-harmonic conversion of 1064 nm radiation in rare gases, J. Phys. B: At. Mol. Opt. Phys. **21**, L31 (1988).

[5] T. Gaumnitz, A. Jain, Y. Pertot, M. Huppert, I. Jordan, F. Ardana-Lamas, and H. J. Wörner, Streaking of 43-attosecond soft-x-ray pulses generated by a passively cep-stable mid-infrared driver, Opt. Express **25**, 27506 (2017).

[6] S. R. Leone, C. W. McCurdy, J. Burgdörfer, L. S. Cederbaum, Z. Chang, N. Dudovich, J. Feist, C. H. Greene, M. Ivanov, R. Kienberger, U. Keller, M. F. Kling, Z.-H. Loh, T. Pfeifer, A. N. Pfeiffer, R. Santra, K. Schafer, A. Stolow, U. Thumm, and M. J. J. Vrakking, What will it take to observe processes in 'real time'?, Nat. Photonics **8**, 162 (2014).

[7] A. Baltuška, T. Udem, M. Ulberacker, M. Hentschel, E. Goulielmakis, C. Gohle, R. Holzwarth, V. S. Yakovlev, A. Scrinzi, T. W. Hänsch, and F. Krausz, Attosecond control of electronic processes by intense light fields, Nature **421**, 611 (2003).

[8] J. Duris, S. Li, T. Driver, E. G. Champenois, J. P. MacArthur, A. A. Lutman, Z. Zhang, P. Rosenberger, J. W. Aldrich, R. Coffee, G. Coslovich, F.-J. Decker, J. M. Glownia, G. Hartmann, W. Helml, A. Kamalov,

J. Knurr, J. Krzywinski, M.-F. Lin, J. P. Marangos, M. Nantel, A. Natan, J. T. O'Neal, N. Shivaram, P. Walter, A. L. Wang, J. J. Welch, J. Z. Wolf, Thomas J. A. Xu, M. F. Kling, P. H. Bucksbaum, A. Zholents, Z. Huang, J. P. Cryan, and A. Marinelli, Tunable isolated attosecond x-ray pulses with gigawatt peak power from a free-electron laser, Nat. Photonics **14**, 30 (2020).

[9] M. Drescher, M. Hentschel, R. Kienberger, G. Tempea, C. Spielmann, G. A. Reider, P. B. Corkum, and F. Krausz, X-ray pulses approaching the attosecond frontier, Science **291**, 1923 (2001).

[10] S. Li, T. Driver, P. Rosenberger, E. G. Champenois, J. Duris, A. Al-Haddad, V. Averbukh, J. C. T. Barnard, N. Berrah, C. Bostedt, P. H. Bucksbaum, R. N. Coffee, L. F. DiMauro, L. Fang, D. Garratt, A. Gatton, Z. Guo, G. Hartmann, D. Haxton, W. Helml, Z. Huang, A. C. LaForge, A. Kamalov, J. Knurr, M.-F. Lin, A. A. Lutman, J. P. MacArthur, J. P. Marangos, M. Nantel, A. Natan, R. Obaid, J. T. O'Neal, N. H. Shivaram, A. Schori, P. Walter, A. L. Wang, T. J. A. Wolf, Z. Zhang, M. F. Kling, A. Marinelli, and J. P. Cryan, Attosecond coherent electron motion in auger-meitner decay, Science **375**, 285 (2022).

[11] F. Calegari, D. Ayuso, A. Trabattoni, L. Belshaw, S. D. Camillis, S. Anumula, F. Frassetto, L. Poletto, A. Palacios, P. Decleva, J. B. Greenwood, F. Martín, and M. Nisoli, Ultrafast electron dynamics in phenylalanine initiated by attosecond pulses, Science **346**, 336 (2014).

[12] P. M. Kraus, B. Mignolet, D. Baykusheva, A. Rupenyan, L. Horný, E. F. Penka, G. Grassi, O. I. Tolstikhin, J. Schneider, F. Jensen, L. B. Madsen, A. D. Bandrauk, F. Remacle, and H. J. Wörner, Measurement and laser control of attosecond charge migration in ionized iodoacetylene, Science **350**, 790 (2015).

[13] J. D. Biggs, Y. Zhang, D. Healion, and S. Mukamel, Watching energy transfer in metalloporphyrin heterodimers using stimulated x-ray raman spectroscopy, Proc. Natl. Acad. Sci. **110**, 15597 (2013).

[14] S. Tanaka and S. Mukamel, Probing exciton dynamics using raman resonances in femtosecond x-ray four-wave mixing, Phys. Rev. A **67**, 033818 (2003).

[15] N. Rohringer, X-ray raman scattering: a building block for nonlinear spectroscopy, Philos. Trans. Royal Soc. A: Math. Phys. Eng. Sci. **377**, 20170471 (2019).

[16] J. D. Biggs, Y. Zhang, D. Healion, and S. Mukamel, Two-dimensional stimulated resonance raman spectroscopy of molecules with broadband x-ray pulses, J. Chem. Phys. **136**, 174117 (2012).

[17] C. Weninger, M. Purvis, D. Ryan, R. A. London, J. D. Bozek, C. Bostedt, A. Graf, G. Brown, J. J. Rocca, and N. Rohringer, Stimulated electronic x-ray raman scattering, Phys. Rev. Lett. **111**, 233902 (2013).

[18] J. T. O'Neal, E. G. Champenois, S. Oberli, R. Obaid, A. Al-Haddad, J. Barnard, N. Berrah, R. Coffee, J. Duris, G. Galinis, D. Garratt, J. M. Glownia, D. Haxton, P. Ho, S. Li, X. Li, J. MacArthur, J. P. Marangos, A. Natan, N. Shivaram, D. S. Slaughter, P. Walter, S. Wandel, L. Young, C. Bostedt, P. H. Bucksbaum, A. Picón, A. Marinelli, and J. P. Cryan, Electronic population transfer via impulsive stimulated x-ray raman scattering with attosecond soft-x-ray pulses, Phys. Rev. Lett. **125**, 073203 (2020).

[19] X. Li, N. Govind, C. Isborn, A. E. DePrince III, and K. Lopata, Real-time time-dependent electronic structure theory, Chem. Rev. **120**, 9951 (2020).

[20] S. Kvaal, Ab initio quantum dynamics using coupled-cluster, J. Chem. Phys. **136**, 194109 (2012).

[21] C. Huber and T. Klamroth, Explicitly time-dependent coupled cluster singles doubles calculations of laser-driven many-electron dynamics, J. Chem. Phys. **134**, 054113 (2011).

[22] D. A. Pigg, G. Hagen, H. Nam, and T. Papenbrock, Time-dependent coupled-cluster method for atomic nuclei, Phys. Rev. C **86**, 014308 (2012).

[23] A. S. Skeidsvoll, A. Balbi, and H. Koch, Time-dependent coupled-cluster theory for ultrafast transient-absorption spectroscopy, Phys. Rev. A **102**, 023115 (2020).

[24] Y. C. Park, A. Perera, and R. J. Bartlett, Equation of motion coupled-cluster for core excitation spectra: Two complementary approaches, J. Chem. Phys. **151**, 164117 (2019).

[25] H. E. Kristiansen, Ø. S. Schøyen, S. Kvaal, and T. B. Pedersen, Numerical stability of time-dependent coupled-cluster methods for many-electron dynamics in intense laser pulses, J. Chem. Phys. **152**, 071102 (2020).

[26] T. B. Pedersen, H. E. Kristiansen, T. Bodenstein, S. Kvaal, and Ø. S. Schøyen, Interpretation of coupled-cluster many-electron dynamics in terms of stationary states, J. Chem. Theory Comput. **17**, 388 (2021).

[27] Y. C. Park, A. Perera, and R. J. Bartlett, Equation of motion coupled-cluster study of core excitation spectra ii: Beyond the dipole approximation, J. Chem. Phys. **155**, 094103 (2021).

[28] T. B. Pedersen and S. Kvaal, Symplectic integration and physical interpretation of time-dependent coupled-cluster theory, J. Chem. Phys. **150**, 144106 (2019).

[29] H. Koch and P. Jørgensen, Coupled cluster response functions, J. Chem. Phys. **93**, 3333 (1990).

[30] T. B. Pedersen and H. Koch, Coupled cluster response functions revisited, J. Chem. Phys. **106**, 8059 (1997).

[31] D. R. Nascimento and A. E. DePrince III, Linear absorption spectra from explicitly time-dependent equation-of-motion coupled-cluster theory, J. Chem. Theory Comput. **12**, 5834 (2016).

[32] H. B. Schlegel, S. M. Smith, and X. Li, Electronic optical response of molecules in intense fields: Comparison of td-hf, td-cis, and td-cis (d) approaches, J. Chem. Phys. **126**, 244110 (2007).

[33] E. Luppi and M. Head-Gordon, Computation of high-harmonic generation spectra of h2 and n2 in intense laser pulses using quantum chemistry methods and time-dependent density functional theory, Mol. Phys. **110**, 909 (2012).

[34] D. R. Nascimento and A. E. DePrince, Simulation of near-edge x-ray absorption fine structure with time-dependent equation-of-motion coupled-cluster theory, J. Phys. Chem. Lett. **8**, 2951 (2017).

[35] D. R. Nascimento and A. E. DePrince III, A general time-domain formulation of equation-of-motion coupled-cluster theory for linear spectroscopy, J. Chem. Phys. **151**, 204107 (2019).

[36] L. N. Koulias, D. B. Williams-Young, D. R. Nascimento, A. E. DePrince, and X. Li, Relativistic real-time time-dependent equation-of-motion coupled-cluster, J. Chem. Theory Comput. **15**, 6617 (2019).

[37] A. S. Skeidsvoll and H. Koch, Comparing real-time coupled cluster methods through simulation of collective rabi oscillations (2023), arXiv:2301.05546 [physics.chem-ph].

[38] D. Mukherjee and P. Mukherjee, A response-function approach to the direct calculation of the transition-energy in a multiple-cluster expansion formalism, Chem. Phys. **39**, 325 (1979).

[39] K. Emrich, An extension of the coupled cluster formalism to excited states (i), Nucl. Phys. A **351**, 379 (1981).

[40] J. F. Stanton and R. J. Bartlett, The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties, J. Chem. Phys. **98**, 7029 (1993).

[41] P. Norman and A. Dreuw, Simulating x-ray spectroscopies and calculating core-excited states of molecules, Chem. Rev. **118**, 7208 (2018).

[42] L. S. Cederbaum, W. Domcke, and J. Schirmer, Many-body theory of core holes, Phys. Rev. A **22**, 206 (1980).

[43] S. Coriani and H. Koch, Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework, J. Chem. Phys. **143**, 181103 (2015).

[44] T. Helgaker, P. Jørgensen, and J. Olsen, Coupled-cluster theory, in *Molecular Electronic-Structure Theory* (John Wiley & Sons, Ltd, 2000) Chap. 13, pp. 648–723.

[45] A. S. Skeidsvoll, T. Moitra, A. Balbi, A. C. Paul, S. Coriani, and H. Koch, Simulating weak-field attosecond processes with a lanczos reduced basis approach to time-dependent equation-of-motion coupled-cluster theory, Phys. Rev. A **105**, 023103 (2022).

[46] A. Barth and L. S. Cederbaum, Many-body theory of core-valence excitations, Phys. Rev. A **23**, 1038 (1981).

[47] S. D. Folkestad, E. F. Kjønstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, T. Moitra, A. C. Paul, M. Scavino, A. S. Skeidsvoll, Å. H. Tveten, and H. Koch, et 1.0: An open source electronic structure program with emphasis on coupled cluster and multilevel methods, J. Chem. Phys. **152**, 184103 (2020).

[48] R. Püttner, I. Dominguez, T. J. Morgan, C. Cisneros, R. F. Fink, E. Rotenberg, T. Warwick, M. Domke, G. Kaindl, and A. S. Schlachter, Vibrationally resolved o $1s$ core-excitation spectra of co and no, Phys. Rev. A **59**, 3415 (1999).

[49] F. Lovas, E. Tiemann, J. Coursey, S. Kotochigova, J. Chang, K. Olsen, and R. Dragoset, Diatomic spectral database (2003).

[50] H. J. Wörner, C. A. Arrell, N. Banerji, A. Cannizzo, M. Chergui, A. K. Das, P. Hamm, U. Keller, P. M. Kraus, E. Liberatore, P. Lopez-Tarifa, M. Lucchini, M. Meuwly, C. Milne, J.-E. Moser, U. Rothlisberger, G. Smolentsev, J. Teuscher, J. A. van Bokhoven, and O. Wenger, Charge migration and charge transfer in molecular systems, Struct. Dyn. **4**, 061508 (2017).

[51] T. Brinck, M. Haeberlein, and M. Jonsson, A computational analysis of substituent effects on the o- h bond dissociation energy in phenols: polar versus radical effects, J. Am. Chem. Soc. **119**, 4239 (1997).