ISBN 978-82-326-1872-9 (printed ver.)
ISBN 978-82-326-1873-6 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2016:267

Rolf Heilemann Myhre

Multilevel coupled cluster theory

Development, implementation and application of multilevel coupled cluster theory for core and valence excited states

Rolf Heilemann Myhre
Doctoral thesis

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor
Faculty of Natural Sciences and Technology
Department of Chemistry

Doctoral theses at NTNU, 2016:267

NTNU

NTNU
Norwegian University of Science and Technology

NTNU
Norwegian University of Science and Technology

Rolf Heilemann Myhre

Multilevel coupled cluster theory

Development, implementation and application of multilevel coupled cluster theory for core and valence excited states

Thesis for the Degree of Philosophiae Doctor

Trondheim, October 2016

Norwegian University of Science and Technology
Faculty of Natural Sciences and Technology
Department of Chemistry

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Natural Sciences and Technology
Department of Chemistry

ISBN 978-82-326-1872-9 (printed ver.)
ISBN 978-82-326-1873-6 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2016:267

Printed by NTNU Grafisk senter

To my mother, Nina

Summary

This thesis consist of seven papers and an introduction describing the ongoing development and applications of multilevel coupled cluster theory. Coupled cluster theory is one of the most accurate, but also one of the most computationally demanding methods currently available in electronic structure theory. Multilevel coupled cluster theory can greatly reduce the cost by treating only a part of the molecule with a high level theory and the rest more approximately.

X-ray spectroscopy involving core excitations is becoming increasingly important in chemical investigations. Modelling core excitations is challenging and require high level methods. This thesis demonstrates how coupled cluster models can be combined with novel experimental techniques to follow the evolution of excited states in thymine.

Core excitations are highly local, being dominated by a single core electron. This makes them well suited for the multilevel coupled cluster approach. The method can be combined with a new type of orbitals, generated specifically for the excitation being investigated. We achieved almost two orders of magnitude reductions in computational cost with this approach.

Acknowledgements

First, I would like to thank my supervisor Henrik Koch. We have collaborated for four years now and he has given me opportunities to participate in projects that are truly unique. I would also like to thank him for his endless patience, encouragement and advice.

I would also like to thank my co-supervisor Svein Sunde and regret that we have not been able to publish anything together.

This thesis would not have been possible if not for my collaborators and I would like to thank James P. Cryan, Markus Gühr, Ida-Marie Høyvik, Alfredo M. J. Sanchéz de Merás, Richard Squibb Thomas J. A. Wolf, the other members of The Thymine Collaboration [1] and especially Sonia Coriani for their tireless work and support.

I would like to thank Todd J. Martínez and the rest of the Martínez group for hosting parts of this project at SLAC, California and Alfredo M. J. Sanchéz de Merás and the other members of ICMOL for hosting parts of this project in València, Spain.

Throughout this project, I have had many interesting and fruitful academic discussions and I would like to thank Basile F. E. Curchod, Eirik Hjertenæs, Eirik Kjønstad, Andrew Komornicki, Zheng Li, Shungo Miyabe, Robert M. Parrish, Nicolay Plotnikov, Kristin Marie Skjelbred, James W. Snyder Jr., Linda Stakvik and Xiaolei Zhu for their thoughts, perspectives and help.

A huge thank you goes to Andrew Dibbs and Eirik Kjønstad who volunteered to hunt down the worst typos and Norwegianisms in this text.

Working on this thesis has mostly been very enjoyable and for that I thank the students and employees at the Department of Chemistry, NTNU.

Finally, I would like to thank my family and friends who have supported and encouraged me throughout my studies.

Abbreviations

ADC ..... Algebraic Diagramatic Construction
AMO ..... Automated Molecular Orbital
AO ..... Atomic Orbital
CC ..... Coupled Cluster
CCS ..... Coupled Cluster Singles
CC2 ..... Coupled Cluster Singles and perturbative Doubles
CCSD ..... Coupled Cluster Singles and Doubles
CC3 ..... Coupled Cluster Singles, Doubles and perturbative Triples
CCSDT ..... Coupled Cluster Singles, Doubles and Triples
CI ..... Configuration Interaction
CISD ..... Configuration Interaction Singles and Doubles
CVS ..... Core-Valence Separation
DKH2 ..... Second order Douglas-Kroll-Hess correction
DUXAP ..... Delayed Ultrafast X-ray Auger Probing
ECC2 ..... Extended CC2 (Early name for MLCCSD)
FC ..... Franck-Condon
FEL ..... Free Electron Laser
HF ..... Hartree-Fock
HOMO ..... Highest Occupied Molecular Orbital
LCLS ..... Linac Coherent Light Source
LUMO ..... Lowest Unoccupied Molecular Orbital
MCSCF ..... Multiconfigurational Self-Consisten Field
MLCC ..... Multilevel Coupled Cluster

MP2 ..... Møller-Plesset second order perturbation theory

MBPT ..... Many-Body Perturbation Theory

MO ..... Molecular Orbital

NEXAFS ..... Near Edge X-ray Absorption Fine Structure

NTO ..... Natural Transition Orbitals

PAO ..... Projected Atomic Orbitals

SCF ..... Self Consistent Field

XAS ..... X-ray Absorption Spectroscopy

XPS ..... X-ray Photoelectron Spectroscopy

Contents

Chapter 1

Introduction

Coupled cluster (CC) theory is one of the most, if not the most, accurate method regularly employed in electronic structure theory. Unlike configuration interaction (CI), CC has a straightforward interpretation with electron interactions modelled as simultaneous excitations of two or more electrons from the reference state. An important property of CC is that it is size extensive, so the energy of two noninteracting molecules is equal to the sum of the energy of the two molecules. Another property of CC theory is that excitation energies and oscillator strengths are size intensive in the CC linear response (CCLR) formalism. When such properties are calculated for two identical molecules infinitely far apart, the results are identical to those of one molecule. In addition, CC methods form a hierarchy of increasingly accurate methods. This makes it possible to systematically improve the accuracy of calculations.

As always, there is no such thing as a free lunch. In the case of CC methods, the computational cost increases rapidly with system size and level of theory. For example, CC singles and double (CCSD) scales as N^6, where N is a measure of system size, while CCSD with perturbative triples (CC3) scales as N^7. For this reason, CCSD has been limited to medium size molecules, while CC3 has been mostly used for benchmarking.

X-ray spectroscopy is a rapidly developing spectroscopic method and new facilities with MHz repetition frequency are under construction at the European XFEL in Hamburg and SLAC National Accelerator Facility in California. In X-ray absorption spectroscopy (XAS) core electrons are excited, making it possible to target specific atoms in a molecule and obtain information about their local environment. From a theoretical perspective, core excitations bring new challenges not encountered when describing valence electron processes. In particular, the core hole created in core excitation processes results in a large charge transfer from the core to valence region. This again leads to large relaxation effects. Consequently,

high level methods and large basis sets are often required in order to reproduce experimental results quantitatively.

For small molecules it is possible to use CC3 and large basis sets to simulate accurate near edge X-ray fine structure (NEXAFS) spectra. However, this is too expensive for larger systems, requiring techniques to reduce the cost while maintaining the accuracy. This thesis will outline the development and implementation of multilevel coupled cluster theory (MLCC). By combining a high level and a less expensive CC method and restricting the high level method to a small part of the system, large savings in computational cost can be achieved. In this work, reductions up to two orders of magnitude are demonstrated.

1.1 Outline

Seven publications are included in this work labelled A to G. The first three papers, referred to as A, B, and C, describe the development and implementation of MLCC theory. The next paper, D, reports the application of multilevel CCSD (MLCCSD) to X-ray spectroscopy. In Papers E and F the accuracy of CC3 is assessed against experimental spectra. Finally, G describes the results of a time resolved NEXAFS experiment conducted at SLAC.

Chapter 2 briefly presents general electronic structure theory with an extra focus on CC theory. This chapter also presents the notation used in the rest of the thesis. Note that some of the naming conventions changed during the project, so it will not be consistent in all the papers. In particular, MLCCSD was initially referred to as ECC2.

The development of MLCC is discussed in Chapter 3, corresponding to Papers A, B and C. There is a particular focus on the treatment of excited states, as MLCC is best suited for size intensive properties like excitation energies. This chapter will also discuss how to generate an active space and the implementation of multilevel CC3 (MLCC3) in Dalton.

The next chapter, Chapter 4, describes the use of MLCC to simulate X-ray spectra, corresponding to Paper D. Example calculations using the newly developed automated molecular orbitals (AMOs) are also presented. These orbitals show great potential in combination with MLCC3.

Papers E and F assess the accuracy of CC theory, particularly CC3, for core excitations and will be discussed in Chapter 5. This chapter also deals with basis set requirements when calculating core excitations.

Chapter 6 presents the experiment described in Paper G with a focus on the theoretical methods. Calculating high level time dependent spectra of thymine required

development of new theory and considerable computational effort. This project is by far the largest of those presented in this thesis.

Finally, Chapter 7 contains the conclusion and plans for future work.

1.2 Publications

R. H. Myhre, Alfredo M. J. Sánchez de Merás and H. Koch
The extended CC2 model, ECC2
Mol. Phys. **111**, 1109 (2013)

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

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

R. H. Myhre, S. Coriani and H. Koch
Near-Edge X-ray Absorption Fine Structure within Multilevel Coupled Cluster Theory
J. Chem. Theory Comput. **12**, 2633 (2016)

R. H. Myhre, S. Coriani and H. Koch
An X-ray absorption study of glycine within Coupled Cluster Linear Response theory
Manuscript

R. H. Myhre, T. J. A. Wolf, S. Coriani, M. Gühr and H. Koch
A theoretical and experimental benchmark study of core excited states in N_2
Manuscript

T. J. A. Wolf, R. H. Myhre, J. P. Cryan, S. Coriani, R. Squibb, A. Battistoni, N. Berrah, C. Bostedt, P. Bucksbaum, R. Coffee, G. Coslovich, R. Feifel, K. Gaffney, J. Grilj, T. J. Martínez, S. Miyabe, S. P. Moeller, M. Mucke, A. Natan, R. Obaid, T. Osipov, O. Plekan, S. Wang, H. Koch and M. Gühr
Probing ultrafast $\pi\pi^*/n\pi^*$ internal conversion in organic chromophores via K-edge resonant absorption
Manuscript

Chapter 2

General theory

2.1 The Schrödinger equation

The goal in almost all electronic structure calculations is solving the Schrödinger equation [2].

$$H\Psi = i\hbar\frac{\partial}{\partial t}\Psi \tag{2.1}$$

$$H\Psi = E\Psi \tag{2.2}$$

Equation (2.1) is the time dependent version of the equation and describes the time evolution of the system with wave function Ψ. $\hbar$ is the reduced Planck constant, which is 1 in atomic units and usually omitted in derivations. The other equation, Eq. (2.2), is the time independent version and describes a system in a stationary state where Ψ is an eigenfunction of the Hamiltonian, H, with eigenvalue E, the energy. In principle, the Schrödinger equation can describe all spin-$\frac{1}{2}$ particles, i. e. fermions, relativistically by inserting the Dirac Hamiltonian [3]. However, the size of relativistic effects on molecular properties typically scales with atom number as

$$\delta P \sim \left(\frac{Z}{137}\right)^2 \tag{2.3}$$

and are usually neglected for light elements [4, 5]. The exception is processes involving core electrons because their high kinetic energy can give considerable relativistic effects. This will be discussed further in Section 5.3. For the rest of this thesis, the Hamiltonian refers to the non-relativistic version unless explicitly stated otherwise.

When neglecting relativistic effects, the Hamiltonian operator can conveniently be divided into five separate parts.

$$H = T_e + T_n + V_{ee} + V_{en} + V_{nn} \tag{2.4}$$

In Eq. (2.4), T refers to kinetic and V refers to potential energy terms, e refers to electrons and n refers to atomic cores. The term V_{nn} should be interpreted as the interaction between the cores, V_{ee} interactions between electrons and V_{en} interactions between electrons and cores. Atomic units are used in the following equations and capital indices refer to atomic cores while lower case indices refer to electrons.

$$T_e = -\frac{1}{2}\sum_i \nabla_i^2 \tag{2.5}$$

$$T_n = -\sum_I \frac{1}{2M_I}\nabla_I^2 \tag{2.6}$$

$$V_{ee} = \frac{1}{2}\sum_{i\neq j}\frac{1}{r_{ij}} \tag{2.7}$$

$$V_{ee} = -\sum_i\sum_I \frac{Z_I}{R_{iI}} \tag{2.8}$$

$$V_{ee} = \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{R_{Ij}} \tag{2.9}$$

Protons in the core are more than 1800 times more massive than electrons. This justifies the Born-Oppenheimer approximation [6], where the electron distribution is assumed to instantly adapt to changes in the atomic positions. This greatly simplifies the calculations because T_n and V_{nn} are reduced to constants while V_{en} becomes a simple one-electron operator. This assumption is usually valid, but there are exceptions such as conical intersections.

The hydrogen atom, H, and the dihydrogen cation, H_2^+, are the only chemical systems for which the non-relativistic Schrödinger equation can be solved analytically. For larger systems, the wave function must be parametrised in a finite dimensional basis set. In most models, this is done by expressing the wave function as a Slater determinant or a linear combination of them [7]. A Slater determinant is a determinant of N molecular spin orbitals, $\phi_p(\boldsymbol{x})$ where N is the number of electrons and $\boldsymbol{x}$ is a compound index with a spin and a spatial part.

$$|\phi_{p_1}\phi_{p_2}\dots\phi_{p_N}| = \frac{1}{\sqrt{N}}\begin{vmatrix} \phi_{p_1}(\boldsymbol{x_1}) & \phi_{p_2}(\boldsymbol{x_1}) & \cdots & \phi_{p_N}(\boldsymbol{x_1}) \\ \phi_{p_1}(\boldsymbol{x_2}) & \phi_{p_2}(\boldsymbol{x_2}) & \cdots & \phi_{p_N}(\boldsymbol{x_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p_1}(\boldsymbol{x_N}) & \phi_{p_2}(\boldsymbol{x_N}) & \cdots & \phi_{p_N}(\boldsymbol{x_N}) \end{vmatrix} \tag{2.10}$$

These determinants span a Hilbert space, called the Fock space, that contains the exact solution of the Schrödinger equation within the given basis set. By writing the basis vectors as determinants, we ensure that the wave function has the correct antisymmetry of fermions. Another advantage of formulating a Fock space is that its components, i.e. the determinants, can be expressed as strings of creation

operators, $a_p^\dagger$, operating on an empty vacuum state, similar to a Clifford algebra [8, 9]. The creation operator, $a_p^\dagger$ can be viewed as placing an electron in molecular spin orbital p. Correspondingly, the destruction operator, a_q, removes an electron from spin orbital q. The antisymmetry of the fermions are now preserved through anticommutation relations:

$$\{a_p^\dagger, a_q^\dagger\} = 0 \qquad \{a_p, a_q\} = 0 \qquad \{a_p^\dagger, a_q\} = \delta_{pq} \tag{2.11}$$

In this formalism and the spin orbital basis, the electronic Hamiltonian is written as

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} (pq|rs) a_p^\dagger a_r^\dagger a_s a_q \tag{2.12}$$

The first term in Eq 2.12 is a one-electron operator term with the scalars h_{pq} given by

$$h_{pq} = \int \phi_p^*(\boldsymbol{x}) \left(-\frac{1}{2}\nabla^2 - \sum_I \frac{Z_I}{r_I} \right) \phi_q(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{2.13}$$

The second term is a two-electron operator term and is needed to describe electron interactions. The repulsion integrals are given by

$$(pq|rs) = \int\int \frac{\phi_p^*(\boldsymbol{x_1}) \phi_r^*(\boldsymbol{x_2}) \phi_q(\boldsymbol{x_1}) \phi_s(\boldsymbol{x_2})}{r_{12}} \mathrm{d}\boldsymbol{x_1} \mathrm{d}\boldsymbol{x_2} \tag{2.14}$$

2.2 Hartree-Fock theory

Hartree-Fock (HF) theory [10, 11] is probably the most important method in electronic structure theory. Alone, it provides qualitative descriptions of ground state properties such as molecular geometries and can be extended with the random phase approximation [12] to describe excited state properties. It also serves as the starting point and reference for many higher level methods, including CC theory.

In HF theory, the single electron configuration that best approximates the wave function is determined. A physical interpretation is to view each electron as a particle moving in the potential of the nuclei and the average potential of the other electrons, but neglecting the instantaneous interactions between electrons. To achieve this, the Hamiltonian is divided into the Fock operator, F, and the fluctuation potential, U.

$$H = F + U \tag{2.15}$$

The Fock operator is an effective one-electron operator. It contains h_{pq} from Eq. 2.12 and the effective one-electron terms from the two-electron operator.

$$F_{pq} = \left(h_{pq} + \sum_i \left((pq|ii) - (pi|iq) \right) \right) a_p^\dagger a_q \tag{2.16}$$

Here, i refers to an occupied spin orbital and in this thesis the standard notation is used where $i, j, k, \ldots$ refer to occupied orbitals, $a, b, c, \ldots$ refer to virtual orbitals and $p, q, r, \ldots$ refer to general orbitals. The two terms in Eq. (2.16) correspond to Coulomb repulsion and exchange, respectively.

In practice, the HF equations are solved by minimising the energy. This corresponds to orbital rotations, creating a set of orthogonal molecular orbitals (MOs). The orbitals appear in the equation, so they have to be solved iteratively and the procedure is known as self-consistent field (SCF). The basis of MOs obtained from SCF are referred to as the canonical basis and the fact that the Fock matrix is diagonal simplifies the derivation of many methods, particularly perturbative models like CC3. However, the canonical solution is not a unique and the HF wave function is invariant to any orbital rotation as long as the occupied and virtual blocks of the Fock matrix remain noninteracting. The choice of orbital will be discussed further in Chapter 3. For closed shell systems, the spin symmetry of the solution is usually enforced. This is known as restricted HF. For open shell systems this is not possible and unrestricted HF is necessary. This carries the risk of obtaining a lower energy solution with wrong spin symmetry.

2.3 Configuration interaction

The difference between the HF energy and the exact solution within the basis set is referred to as the correlation energy. In order to calculate this, electronic exchange must be accounted for and several configuration vectors must be included in the wave function. The most straightforward procedure is to add all possible configurations and minimise the energy, known as full configuration interaction (FCI). This correspond to finding the eigenvalues of the Hamiltonian matrix. The drawback of this approach is that the number of Slater determinants that must be included scales combinatorially.

$$N_{det} = \binom{n}{N_\alpha}\binom{n}{N_\beta} \tag{2.17}$$

Above, n is the number of orbitals and N_α and N_β are the numbers of α- and β-electrons, respectively.

Clearly, such steep scaling limits the FCI approach to very small systems and approximations are required if we want to do more than benchmarking. There is little point in having the exact solution within the basis set if the basis set is too small. The Fock term usually dominates the other terms in the Hamiltonian. It is therefore a reasonable assumption that the HF wave function is a good reference that can be used as a starting point and reference for more accurate methods. Note that there are important cases, such as stretched geometries, where several configurations are required to describe the ground state. In such cases, the HF wave function is not a good reference.

Truncated configuration interaction (CI) is one such method. The starting point is to write the FCI wave function as an expansion from the HF wave function.

$$|\text{FCI}\rangle = \left(1 + \sum_{ai} Y_i^a + \sum_{a>b,i>j} Y_{ij}^{ab} + \dots\right) |\text{HF}\rangle \tag{2.18}$$

The operators Y_i^a and Y_{ij}^{ab} excite one and two electrons respectively from occupied to virtual orbitals and have an associated scalar weight, C.

$$Y_i^a |\text{HF}\rangle = C_i^a a_a^\dagger a_i |\text{HF}\rangle \tag{2.19}$$

$$Y_{ij}^{ab} |\text{HF}\rangle = C_{ij}^{ab} a_a^\dagger a_b^\dagger a_i a_j |\text{HF}\rangle \tag{2.20}$$

The number of determinants that has to be included is greatly reduced by truncating the expansion in Eq. (2.18) after some excitation level. For example, CI singles and doubles (CISD) is truncated after the double excitations and scales as N^6. This means that a doubling of system size will increase the computational cost 64 times, but CISD is much cheaper than FCI and, importantly, it scales polynomially. After the introduction of CCSD, CISD has become less popular because CCSD is more accurate and has the same scaling.

2.4 Coupled cluster theory

Coupled cluster theory was originally formulated in nuclear physics in the 1950s [13, 14], but was later adapted for quantum chemistry where it enjoyed greater success [15, 16]. Here we only discuss closed shell CC models.

In CC theory, the wave function is parametrised as the nonlinear product of excitation operators, τ_μ, acting on the HF reference state. Excitation operators are defined as the operators that excites the reference state to an excited state $|\mu\rangle$.

$$\tau_\mu |\text{HF}\rangle = |\mu\rangle \tag{2.21}$$

The cluster operator, X, is constructed by taking the sum of all excitation operators multiplied by their amplitudes, x_μ.

$$\begin{aligned} X &= X_1 + X_2 + X_3 + \dots \\ &= \sum_{\mu_1} x_{\mu_1} \tau_{\mu_1} + \sum_{\mu_2} x_{\mu_2} \tau_{\mu_2} + \sum_{\mu_3} x_{\mu_3} \tau_{\mu_3} + \dots \end{aligned} \tag{2.22}$$

The compound indices μ_1, μ_2 and μ_3 run over single, double and triple excited states, respectively, and in exact theory the cluster operator includes all excitations. While the Brillouin theorem ensures that the single excitations do not have a direct effect on the ground state energy, they act as orbital relaxation parameters and are

included in the standard CC formulation. In formulations with explicit orbital relaxation such as Brueckner CC theory [17, 18] or orbital adaptive CC [19], the single excitation amplitudes are zero and not included.

The coupled cluster wave function is obtained by letting the exponential of the cluster operator act on the HF reference state.

$$|\mathrm{CC}\rangle = \exp(X)\,|\mathrm{HF}\rangle \tag{2.23}$$

Note that the exponential in Eq. (2.23) has a finite expansion when acting on the reference state because there is a finite number of electrons to excite. Writing the wave function as an exponential is known as the exponential ansatz.

Including products of the excitation operators ensures that contributions from higher excited states are included that would be discarded in the truncated CI expansion in Eq. (2.18). The downside is that energy minimisation leads to a set of complicated nonlinear equations that are very difficult to solve [20]. Instead, the CC amplitudes and energy is determined using projection.

$$\langle \mathrm{HF}|\exp(-X)H\exp(X)\,|\mathrm{HF}\rangle = E \tag{2.24}$$

$$\langle \mu|\exp(-X)H\exp(X)\,|\mathrm{HF}\rangle = 0 \tag{2.25}$$

The untruncated CC wave function is equivalent to FCI up to a normalisation constant. To make the CC approach usable, the cluster operator must be truncated. The most common approach is to keep single and double excitations, resulting in the CCSD model [21]. When solving for a truncated cluster operator, the projection manifold is restricted to the same states as the excitations. For CCSD, the equations are only projected against the manifold of the singly and doubly excited states, resulting in amplitude equations that scale as N^6.

$$\Omega_{\mu_1} = \langle \mu_1|\,\hat{H} + [\hat{H}, X_2]\,|\mathrm{HF}\rangle = 0 \tag{2.26}$$

$$\Omega_{\mu_2} = \langle \mu_2|\,\hat{H} + [\hat{H}, X_2] + \frac{1}{2}[[\hat{H}, X_2], X_2]\,|\mathrm{HF}\rangle = 0 \tag{2.27}$$

In Eqs. (2.26) and (2.27), the X_1-transformed Hamilton operator,

$$\hat{H} = \exp(-X_1)H\exp(X_1) \tag{2.28}$$

is employed for simplicity. Note that the overlap between the CC and HF wave functions is always 1 and the cluster operator does not commute with its complex conjugate, so calculating the norm of the wave function, $\langle \mathrm{CC}|\mathrm{CC}\rangle$, scales as FCI.

Truncating the cluster operator at different levels results in a natural hierarchy of increasingly accurate and expensive models, starting with CC singles (CCS), CCSD and CC singles, doubles and triples (CCSDT) [22] which scale as N^4, N^6 and N^8, respectively. In addition, there are the CC perturbative models CC2 [23] and CC3 [24]. In these models, the Hamiltonian is split into the Fock operator, considered zero order, and the fluctuation potential considered first order. Similarly, double

excitations are considered first order, triple excitations second order and so on. Single excitations are treated like orbital relaxation parameters and included to infinite order. When solving for the double amplitudes in CC2, only terms up to first order are included and Eq. (2.27) is simplified, resulting in Eq. (2.29).

$$\Omega_{\mu_2} = \langle \mu_2 | \, \hat{H} + [F, X_2] \, | \mathrm{HF} \rangle = 0 \tag{2.29}$$

In canonical basis, Eq. (2.29) is particularly simple because F is diagonal. Consequently, the X_2 amplitudes can be separated and expressed purely in terms of the singles amplitudes and integrals. This eliminates the need for storage of the X_2 amplitudes and makes very efficient implementations possible. The standard implementation of CC2 scales as N^5, but it is possible to implement it scaling as N^4 [25].

In CC3, terms up to second order are included when solving for the triples and similarly to CC2, it is possible to separate out the triples in Eq. (2.30).

$$\Omega_{\mu_3} = \langle \mu_3 | \, [F, X_3] + [\hat{H}, X_2] \, | \mathrm{HF} \rangle = 0 \tag{2.30}$$

Avoiding storage of the triples is critical because the number of amplitudes scale as N^6. High storage demands would severely limit the applicability of the model which is already limited due to its N^7 scaling. The X_3-contributions to the singles and doubles amplitude equations can be viewed as corrections to the CCSD equations.

$$\Omega_{\mu_1} = \Omega_{\mu_1}^{\mathrm{CCSD}} + \langle \mu_1 | \, [H, X_3] \, | \mathrm{HF} \rangle = 0 \tag{2.31}$$

$$\Omega_{\mu_2} = \Omega_{\mu_2}^{\mathrm{CCSD}} + \langle \mu_2 | \, [\hat{H}, X_3] \, | \mathrm{HF} \rangle = 0 \tag{2.32}$$

The CC2 and CC3 models are considered intermediate in the CC hierarchy with CC2 between CCS and CCSD and CC3 between CCSD and CCSDT. Having a systematic hierarchy of models makes it possible to verify lower level results.

A model considered outside the hierarchy is the popular CCSD(T) model [26, 27]. In this model, the triples are included as a correction to the energy after solving the CCSD equations. Calculating the correction scales as N^7, but no further iterations are required. As the effect of the triples are only added after the fact, CCSD(T) cannot be used to calculate excited state properties in its standard formulation, however, it is very accurate for the ground state.

2.5 Coupled cluster response theory

Calculating time dependent properties of coupled cluster can be done with two different methods, response theory [28–32] and equation of motion (EOM) [33, 34]. These methods are similar and produce the same excitation energies, but other properties are somewhat different. This section will describe coupled cluster linear response, CCLR, and Section 2.6 will describe EOM.

In linear response, the system is assumed to be dominated by the time independent Hamiltonian in Eq. (2.12), H_0. A small time dependent perturbation, V^t, is added to the Hamiltonian.

$$H = H_0 + V^t \tag{2.33}$$

By expressing the perturbation as a Fourier series, we obtain its frequency components.

$$V^t = \int_{-\infty}^{\infty} \mathrm{d}\omega V^\omega e^{(-i\omega+\alpha)t} \tag{2.34}$$

The time dependent expectation value of a Hermitian operator, B, can be expanded in orders of the frequency with the first term being the time independent expectation value. Writing the time independent ground state as $|0\rangle$ and its orthogonal complement as $\{|k\rangle\}$, the time dependent expectation value can be expressed as an expansion in orders of the perturbation.

$$\begin{aligned}\langle B\rangle(t) = \langle 0|\, B\, |0\rangle &+ \int_{-\infty}^{\infty} \mathrm{d}\omega \langle\langle B; V^\omega\rangle\rangle_{\omega+i\alpha} e^{(-i\omega+\alpha)t} \\ &+ \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}\omega_1 \mathrm{d}\omega_2 \langle\langle B; V^{\omega_1}; V^{\omega_2}\rangle\rangle_{\omega_1+i\alpha,\omega_2+i\alpha} e^{(-i\omega_1-i\omega_2+2\alpha)t} \\ &+ \ldots\end{aligned} \tag{2.35}$$

In Eq (2.35), $\langle\langle B; V^\omega\rangle\rangle_{\omega+i\alpha}$ is the linear response function and the next term contains the quadratic response function. These can be viewed as the change in the observable B in response to the operator V^ω and operators V^{ω_1} and V^{ω_2}.

The exact linear response function is written as a sum over the orthogonal complement.

$$\langle\langle B; V^\omega\rangle\rangle_{\omega+i\alpha} = \sum_k \left(\frac{\langle 0|\, B\, |k\rangle \langle k|\, V^\omega\, |0\rangle}{\omega - \omega_k + i\alpha} - \frac{\langle 0|\, V^\omega\, |k\rangle \langle k|\, B\, |0\rangle}{\omega + \omega_k + i\alpha} \right) \tag{2.36}$$

Equation (2.36) diverges when ω approaches $\pm\omega_k$, corresponding to the excitation and deexcitation energies of the unperturbed system. The residues at these poles are the transition matrix elements.

As discussed in Section 2.4, CC theory is not variational and instead a bivariational Lagrangian must be constructed using the dual state, $\langle\Lambda|$.

$$\langle\Lambda| = \langle \mathrm{HF}| + \sum_\mu \bar{x}_\mu \langle\mu| \exp(-X) \tag{2.37}$$

Above, $\bar{x}$ are Lagrangian multipliers. The time independent expectation value is expressed in terms of $\langle\Lambda|$ and $|\mathrm{CC}\rangle$.

$$\langle B\rangle_0 = \langle\Lambda|\, B\, |\mathrm{CC}\rangle \tag{2.38}$$

Assuming the frequency components of the perturbation are time independent, we can express the linear response function relatively easily.

$$\begin{aligned}\langle\langle B;C\rangle\rangle_\omega = &\sum_\mu \langle\Lambda|\,[B,\tau_\mu]\,|\mathrm{CC}\rangle\, X^C_\mu(\omega) \\ &+ \sum_\mu \left(\langle\Lambda|\,[C,\tau_\mu]\,|\mathrm{CC}\rangle + \sum_\nu F_{\mu\nu} X^C_\nu(\omega)\right) X^B_\mu(-\omega)\end{aligned} \tag{2.39}$$

The vector $\boldsymbol{X}^B(\omega)$ is related to the Fourier transform of the time dependent CC amplitudes,

$$X^B_\mu(\omega) = \sum_\nu (-\boldsymbol{A} + \omega \boldsymbol{I})^{-1}_{\mu\nu} \langle\nu|\exp(-X)B\,|\mathrm{CC}\rangle \tag{2.40}$$

and the elements of the $\boldsymbol{F}$-matrix are the expectation values of the double commutator between the Hamiltonian and excitation operators.

$$F_{\mu\nu} = \langle\Lambda|\,[[H_0,\tau_\mu],\tau_\nu]\,|\mathrm{CC}\rangle \tag{2.41}$$

Finally, the elements of the Jacobian $\boldsymbol{A}$ in Eq. (2.40) are the derivatives of the Ω-vector with respect to the amplitudes.

$$A_{\mu\nu} = \frac{\partial \Omega_\mu}{\partial x_\nu} = \langle\mu|\exp(-X)[H_0,\tau_\nu]\exp(X)\,|\mathrm{CC}\rangle \tag{2.42}$$

In order to determine the pole structure of the response function, it is necessary to determine the eigenvalues of the Jacobian.

$$\boldsymbol{L}_k\boldsymbol{A} = \omega_k \boldsymbol{L}_k \qquad \boldsymbol{A}\boldsymbol{R}_l = \omega_l \boldsymbol{R}_l \qquad \boldsymbol{L}_k\boldsymbol{R}_l = \delta_{kl} \tag{2.43}$$

The Jacobian is not symmetric and the linear response function requires the eigenvectors from both left and right so both eigenproblems in Eq. (2.43) must be solved. In addition, the eigenvalues are not guaranteed to be real. This can become an issue with degenerate electronic states [35–37].

Deriving the response functions using the time dependent Schrödinger equation is the most intuitive and instructive approach. However, it is often easier to derive specific working equations using the quasienergy method [38, 39], especially for perturbative models where it is not obvious how to define a generalised Hellmann-Feynman theorem [40, 41].

Multiphoton processes requires higher order terms from the response expansion such as the quadratic and cubic response functions. These response functions can be derived similarly to the linear one, but the complexity of the equations grow quickly and they will not be discussed here [30, 31]. Transition moments between excited states can be derived from the quadratic response function and require the corresponding transition matrix elements $\Gamma^B_{k\to m}$.

$$\begin{aligned}\Gamma^B_{k\to m} = &\langle k|\exp(-X)[B,\tau_m]\,|\mathrm{CC}\rangle \\ &+ \sum_n X^B_n(\omega_k-\omega_m)\,\langle k|\exp(-X)[[H_0,\tau_n]\tau_m]\,|\mathrm{CC}\rangle \\ &+ \delta_{km}\,\langle\Lambda|\,B\,|\mathrm{CC}\rangle\end{aligned} \tag{2.44}$$

The Latin indices in Eq. (2.44) refer to excited CC states and the corresponding vectors are obtained by transforming to the basis that diagonalises the Jacobian.

$$\tau_m = \sum_\mu \tau_\mu R_{\mu m} \qquad \langle n| = \sum_\nu L_{n\nu} \langle \nu| \qquad X_n^B(\omega) = \sum_\nu L_{n\nu} X_\nu^B(\omega) \tag{2.45}$$

Solving equation (2.44) as it is written would require a full diagonalisation of the Jacobian which is obviously not possible. However, in actual implementations the terms are obtained in the primitive HF basis and transformed afterwards.

Solving for the eigenvalues of the Jacobian is usually the most expensive step in CCLR calculations. In most cases, we are only interested in a few electronic states and the corresponding eigenvalues and eigenvectors. The Davidson algorithm [42, 43] solves iteratively for the lowest eigenvalues and only requires matrix vector products. This is important because the Jacobian is too large to construct and store explicitly. More recently, the Lanczos algorithm [44–46] has been used to solve for the eigenvalues. It constructs a tridiagonal matrix which in principle makes it possible to compute full spectra. Unfortunately, it is numerically unstable and is limited by large storage and orthogonalisation requirements.

2.6 Coupled cluster equation of motion

The EOM method in CC [33, 34] is a similar to CCLR, but the equations are somewhat easier to solve. In EOM, the excited states are assumed to exist in a space spanned by a CI like expansion.

$$|\boldsymbol{c}\rangle = \sum_\mu c_\mu \tau_\mu |\mathrm{CC}\rangle = \exp(X) \sum_\mu c_\mu \tau_\mu |\mathrm{HF}\rangle \tag{2.46}$$

Note that the index, μ, in Eq. (2.46) runs over all states, including the reference state. The corresponding τ_0 operator is considered the identity operator. Similarly a biorthogonal state is defined for the left hand side.

$$\langle \bar{\boldsymbol{c}}| = \sum_\mu \bar{c}_\mu \langle \mathrm{HF}| \tau_\mu^\dagger \exp(-X) \tag{2.47}$$

Energies of the excited states can then be expressed as pseudo expectation values.

$$E(\boldsymbol{c}, \bar{\boldsymbol{c}}) = \frac{\langle \bar{\boldsymbol{c}}| H |\boldsymbol{c}\rangle}{\langle \bar{\boldsymbol{c}}|\boldsymbol{c}\rangle} \tag{2.48}$$

and determining the excited states reduces to an eigenproblem. The matrix to be diagonalised is nonsymmetric and is a similarity transform of the Hamiltonian.

$$H^X = \exp(-X) H \exp(X) \tag{2.49}$$

By partitioning the similarity transformed Hamiltonian, the close relation with CCLR becomes obvious.

$$\boldsymbol{H}^X = \begin{pmatrix} 0 & \boldsymbol{\eta}^T \\ \mathbf{0} & \boldsymbol{A} \end{pmatrix} + E_0 \boldsymbol{I} \tag{2.50}$$

The first row in the matrix is the vector $\boldsymbol{\eta}$,

$$\eta_\mu = \langle \mathrm{HF} | \, H \, | \mu \rangle \tag{2.51}$$

while $\boldsymbol{A}$ is the Jacobian from Section 2.5. For the ground state, the left and right solutions are the $\langle \Lambda |$ and $| \mathrm{CC} \rangle$ vectors from before and the excitation energies are the eigenvalues of $\boldsymbol{A}$. Consequently, CCLR and EOM-CC produce the same excited states and the same size intensives excitation energies. The two methods do differ, however, for other excited state properties.

For example, the ground state transition moments are calculated as pseudo expectation values. Because the Jacobian is not Hermitian, we must take the product of the left and right expectation values.

$$D = \langle \Lambda | \, \mu \, | \boldsymbol{c} \rangle \, \langle \bar{\boldsymbol{c}} | \, \mu \, | \mathrm{CC} \rangle \tag{2.52}$$

Similarly, transition moments between excited states can be calculated. Unlike the excitation energies, transition moments are not identical in EOM and CCLR and they are not size intensive in EOM [47]. The difference is generally small and the EOM transition moments are less computationally demanding than CCLR because there is one less equation to solve.

2.7 Other *ab initio* methods

This section will briefly mention some other important *ab initio* methods. Specifically, the multiconfigurational self-consistent field [48, 49] method (MCSCF), Møller-Plesset or many-body perturbation theory [50, 51] (MBPT) and the algebraic diagrammatic construction [52, 53] (ADC) approach.

A disadvantage of both CI and CC theory is that they rely on a good reference state to obtain accurate results. At equilibrium geometries, most systems are described well by a single reference and this is not a problem, but at stretched or twisted geometries the effects can be large. A classic example is ethene when the double bond is rotated 90°. At this geometry, two electron configurations become equally important and SCF breaks down. This is known as the multireference problem. The correlation energy missing due to multireference character is known as static correlation, as opposed to dynamic correlation arising from Coulomb repulsion.

Multiconfigurational SCF theory is one of the few methods that can describe multireference states. The basic idea of MCSCF is to optimise the MOs and the CI

coefficients simultaneously. For FCI, this is unnecessary because the coefficients already provide full flexibility. For truncated CI wave functions, on the other hand, it can provide important degrees of freedom. The MCSCF wave function is expressed as an orbital rotation operator, $\exp(-\kappa)$, acting on a CI wave function.

$$|\boldsymbol{\kappa}, \boldsymbol{C}\rangle = \exp(-\kappa) \sum_i C_i \, |i\rangle \tag{2.53}$$

As in CI, the wave function is constructed by minimising the energy expectation value.

$$E = \min_{\boldsymbol{\kappa}, \boldsymbol{C}} \frac{\langle \boldsymbol{\kappa}, \boldsymbol{C}| \, H \, |\boldsymbol{\kappa}, \boldsymbol{C}\rangle}{\langle \boldsymbol{\kappa}, \boldsymbol{C}|\boldsymbol{\kappa}, \boldsymbol{C}\rangle} \tag{2.54}$$

Unfortunately, this is a complicated nonlinear optimisation and can only be employed on relatively small systems. In order to restrict the number of configurations to include, MCSCF is usually employed on an active space. In the complete active space (CAS) approach the orbital space is partitioned into an inactive part where all the orbitals are doubly occupied, a secondary part with unoccupied orbitals and an active part where the occupation numbers are allowed to vary [49]. Such approaches can be implemented very efficiently but the size of the active space must be relatively small to avoid combinatorial explosion. This also limits the size of the basis set, making it difficult to get quantitative results.

One of the earliest post HF methods is Møller-Plesset perturbation theory, also known as many-body perturbation theory. In this approach, the Hamiltonian is assumed to be dominated by the Fock operator and the fluctuation potential, see Eq. (2.15), is viewed as a perturbation. It is possible to derive the equations of MBPT by inserting the definitions into the Schrödinger equation and removing terms of higher order. However, it is easier to use an exponential ansatz and CC and MBPT is closely related. In fact, MP2 and CC2 only differ in the inclusion of single excitations in CC2 and MBPT is size extensive.

A disadvantage with MBPT is that it is not suitable to describe excited states. In the perturbative framework, response theory will produce higher order poles and it is not possible to interpret eigenvalues as excitation energies [39]. Furthermore, the MBPT equations diverge for degenerate states.

More recently, the ADC method has been proposed as an excited state version of MBPT [52]. It was originally derived as a perturbation expansion of the polarisation propagator, but it is easier to derive using the intermediate state representation [53]. The starting point is an MBPT ground state, $|\Psi\rangle$, from which a set of excited states are generated.

$$\{|\mu\rangle\} = \{\tau_\mu \, |\Psi\rangle\} \tag{2.55}$$

By orthogonalising the excited states, the intermediate states, $\{|n\rangle\}$, are obtained. The ADC secular matrix $\boldsymbol{M}$ is then constructed using this basis.

$$M_{mn} = \langle m| \, H - E_0 \, |n\rangle \tag{2.56}$$

Excitation energies and excited states of the system are obtained by finding the eigenvalues and eigenvectors of $\boldsymbol{M}$. Transition moments and excited state properties can then be calculated as expectation values.

The ADC approach produces size intensive excitation energies and excited state properties. Furthermore the $\boldsymbol{M}$-matrix is Hermitian by construction, ensuring real eigenvalues. On the other hand, the approach relies on an MBPT ground state. Convergence of the Møller-Plesset perturbation expansion series is not guaranteed and may diverge in some cases [54]. Currently, ADC(2) and ADC(3) have been implemented as well as the extended ADC(2)-x version [55]. The approach has also been used for core excitations, which we will return to in Chapter 5.

2.8 Basis sets

In order to obtain accurate molecular properties, we require an appropriate basis set as well as an accurate correlation model. Using large basis sets is expensive because most models scale with some power of the total number of orbitals, if not worse. However, there is little point in using a high level method if the basis set does not have the flexibility to describe the relevant property. Almost all basis sets in use today are based on Gaussian functions and the two most popular types are Pople's split valence basis sets [56, 57] and Dunning's correlation consistent basis sets [58]. In calculations, a number of atomic orbitals (AOs) are usually centred to each atom. These are non-orthogonal and resemble the spherical harmonics that are the solution of the Schrödinger equation for the hydrogen atom. From these, the orthogonal MOs are constructed.

Split valence basis sets are constructed by optimising exponents and coefficients from HF atomic calculations. This produces basis sets well suited to describe the HF wave function, but may be less suited for post HF methods. In order to increase the flexibility of the basis set, diffuse and polarising primitive Gaussians can be added. In most applications, the core orbital energy will remain unchanged and it is not necessary to describe them with high accuracy. For this reason, core orbitals are usually described by fewer functions than the valence orbitals, hence the name.

The correlation consistent basis sets, referred to as cc-pVXZ where X is the cardinal number, are constructed specifically to describe electron correlation. The basis set is expanded by adding Gaussians that maximise the correlation energy, determined from atomic CISD calculations. Orbitals that make approximately the same contribution are added at the same time and it turns out that this corresponds to adding one Gaussian in each shell and one new shell for each level of the cardinal number.

Similarly to the split valence basis sets, the correlation consistent basis sets are not designed to recover correlation energy for the core electrons. If such effects

are important, extra core correlation functions must be added using the cc-pCVXZ basis sets [59]. Excited states and anionic states require additional diffuse functions, optimised for the correlation energy of anions [60].

Chapter 3

Multilevel coupled cluster

3.1 Scaling of coupled cluster

Coupled cluster is one of the most accurate methods in electronic structure theory, however, its high computational cost and steep scaling with system size has limited its applicability. The most well known model, CCSD [21], scales as V^4O^2 where V is the number of virtual and O is the number of occupied orbitals. A doubling of the system size will then lead to a 64-fold increase in computational cost. The available computing power is growing rapidly with technological development, but not fast enough to keep up with the demands for accurate calculations. Consequently, new and smarter ways to solve the CC equations are needed. Many methods have been proposed to reduce the scaling of CC models, far too many to mention here. For a more extensive overview, see Section B.1.

Perhaps the most straightforward way to reduce the computational cost is to remove linear dependence in the integrals. This can be accomplished using density fitting [61–65] or Cholesky decomposition [66–69]. In both methods, an auxiliary basis is constructed that contains the information of the two-electron integrals. In density fitting the auxiliary basis is provided beforehand and the integrals are fitted, while the two-electron integral matrix is decomposed in the Cholesky approach [70]. These methods greatly reduces the number of integral pairs to include, but does not reduce the overall scaling in the standard approach.

The starting point of most reduced scaling methods is to note that the integrals in Eq. (2.14) depend on a factor $\frac{1}{r_{12}}$. Consequently, the interactions between two electrons sufficiently far from each other must become negligible. In principle, it should be possible to obtain the FCI energy with a linearly scaling model for sufficiently large systems. Of course, such a model would be too expensive long before we reach systems in the linear scaling regime.

A more involved approach is to identify interactions between electrons that are far away from each other and discard them. The canonical MOs obtained from HF are highly delocalised, so this approach requires some kind of orbital localisation procedure. This approach was pioneered by Saebø and Pulay for MBPT with the projected atomic orbital method (PAO) [71, 72]. More recently, several methods has been developed for CC based on similar ideas, for example natural linear scaling CC [73], divide-expand-consolidate [74, 75], local pair natural orbital CC [76, 77] and orbital-specific-virtual local CC [78, 79].

All the methods mentioned in the previous paragraph treat the whole molecular system with the same level of theory. This is not the case with multilevel methods. Often, one is interested in a part of the system and would like to focus computational resources on it. Multilevel methods achieve this by treating the relevant parts with a high level method and the rest more approximately. The most well known of these methods are probably the ONIOM model [80, 81] and density embedding in density functional theory (DFT) [82–84]. Coupled cluster methods in CAS like active spaces have also been proposed [85, 86].

3.2 Multilevel coupled cluster equations

The starting point when deriving the multilevel CC equations is to divide the orbitals into two or more spaces. For the current derivation, the partition of the orbital space is arbitrary as long as the orbitals are orthogonal. How to generate the orbital spaces will be discussed in Section 3.4. We will begin with the two level MLCCSD method described in Papers A and B. The high level method in this model is CCSD while the rest is described by CC2. Note that the method is referred to as extended CC2 (ECC2) in these papers. After presenting the MLCCSD model, the method will be generalised to several spaces.

After partitioning the orbitals into an active and an inactive space, the cluster operator can also be partitioned. Excitation operators only referring to active orbitals, both occupied and virtual, are called internal while those only involving inactive orbitals are called external. Those that involve both active and inactive orbitals are referred to as semi-external. The cluster operator is then partitioned into an active operator, T, and an inactive operator, S.

$$X = T + S \tag{3.1}$$

Excitations that are internal in the active space are assigned to T, while the external and semi-external are assigned to S. In Paper A, we experimented with including the semi-external excitations in T. This would greatly reduce the computational savings of the model, undermining its purpose. Similarly to the cluster operator, the manifold of excited states can also be partitioned into states reached by internal excitations and states reached by external and semi-external excitations.

$$\{|\mu\rangle\} = \{|\mu^T\rangle\} \oplus \{|\mu^S\rangle\} \tag{3.2}$$

Singles are treated the same in CC2 and CCSD, so the singles are included in T.

The MLCCSD amplitude equations are obtained by inserting these definitions into the CCSD equations (2.26) and (2.27). Because the singles are treated the same in CC2 and CCSD, the singles amplitude equations are the same as in CCSD. Similarly, the equations for the T_2 amplitudes are unchanged, except that projection manifold is restricted to the T_2 states, $\{\langle\mu_2^T|\}$.

$$\langle\mu_2^T|\,\hat{H} + [\hat{H}, X_2] + \frac{1}{2}[[\hat{H}, X_2], X_2]\,|\mathrm{HF}\rangle = 0 \tag{3.3}$$

In the equations for the S_2 manifold there is an ambiguity. The CC2 equation, (2.29), only contains a term with the X_1-transformed Hamilton operator and a commutator between the Fock operator and the doubles cluster operator. All other terms are removed because they are higher order in the perturbation. In MLCCSD, there is a commutator term between the fluctuation potential operator and the active doubles T_2 operator.

$$\langle\mu_2^S|\,\hat{H} + [F, X_2] + [\hat{U}, T_2]\,|\mathrm{HF}\rangle = 0 \tag{3.4}$$

From a perturbation perspective, this term is second order and should not appear in the first order equations. On the other hand, the T_2 equations should be solved to infinite order. From a practical perspective, accuracy must be weighted against computational cost. The most expensive term to compute arising from the commutator is an index restricted version of the B term from CCSD.

$$\Omega^{\mathrm{B}}_{ABij} = \sum_{cd} t_{ij}^{cd}(Ac|Bd) \tag{3.5}$$

Capital indices are general and run over the whole orbital space, while the lowercase indices are restricted to the active space so the overall scaling is quadratic with system size. This is less than CC2 and the term should be included if it improves accuracy. Some testing was done in Paper A, but it was performed on size extensive properties that are poorly described by MLCCSD. Further benchmarking with size intensive properties should be performed. All MLCCSD results reported in Papers B and D were obtained with the version including the commutator.

The advantage of MLCCSD compared to full CCSD is the reduced cost. Theoretically, MLCCSD should scale as $V^2V_A^2O_A^2$, where V_A and O_A are the number of virtual and occupied active orbitals respectively. The overall scaling with system size then becomes quadratic for the CCSD contributions and the computational effort should be dominated by CC2 for sufficiently large systems.

The equations for the multilevel CC3 model are somewhat easier to derive than for MLCCSD because there are no ambiguities. In fact, the amplitude equations become the same as for full CC3, except that the indices are restricted to the active space, see Section C.2. There is, however, a subtlety with the Fock matrix. As stated in Section 2.4, the Fock operator must be diagonal in order to make the

triples amplitude equations separable. This cannot be assumed in MLCC3 because local orbitals are needed for a local active space, see Section 3.4. This would appear to be a problem, however, it turns out that it is sufficient that the Fock operator is diagonal in the active space. Off-diagonal elements in the Fock matrix will then be projected out of the T_3 amplitude equations and they become separable. The overall scaling of the CC3 part of the MLCC3 model becomes $VV_A^3O_A^3$, linear with respect to total system size.

Multilevel CCSD(T) is also possible with equations similar to MLCC3.

$$\varepsilon_{\mu_3^T} t_{\mu_3^T} = -\langle \mu_3^T | \left[U, X_2 \right] | \mathrm{HF} \rangle \tag{3.6}$$

$$E_{\mathrm{corr}}^{\mathrm{MLCCSD(T)}} = \sum_{\mu=\mu_1,\mu_2} x_\mu \langle \mu | \left[U, T_3 \right] | \mathrm{HF} \rangle \tag{3.7}$$

Note that Eqs. (3.6) and (3.7) differ from those given in Paper B. In the paper, the indices are restricted to the active space which results in a computationally size intensive correction while the one given here scales linearly. While the MLCCSD(T) model is easy to implement and is relatively cheap, it is unclear how useful it is. Total energy is a size extensive property that depends on the whole system. Furthermore, CCSD(T) is best for geometry optimisations, but it is challenging to have a consistent active space when the positions of the atoms change.

When expanding the MLCC model to more than two spaces, the cluster operator must be divided into three or more operators. The most active operator will still only contain internal excitations in the innermost orbital space and be labelled T. The next operator, labelled S will contain excitations internal in the second orbital space and those that are semi-external to the T space. Excitation that are semi-external to the next space, however, are not included. In this way, one constructs a hierarchy of increasingly smaller and higher level spaces. Extending the equations in such a scheme is relatively straightforward and an example is given in Section B.2.4. For sufficiently large systems, one could employ an outer space only described at the HF level or even with frozen density.

3.3 Multilevel coupled cluster response theory

As mentioned previously, multilevel coupled cluster theory is best suited for size intensive properties. This requires CC response theory. A detailed derivation of MLCCSD response is given in Section B.2.5 and the MLCC3 equations are given in Section C.2.3. The equations are quite involved, and take a lot of space, so they will not be repeated here. In order to derive the equations, the quasienergy approach was utilised as described in reference [87].

In general, the MLCC versions of the equations are very similar to the standard theory versions. The MLCCSD Jacobian is given in Eq. (B.51) and the first

rows corresponding to projection against $\langle \mu_1 |$ and $\langle \mu_2^T |$ are the same as in CCSD. Without the extra commutator in Eq. (3.4), the rows projecting against $\langle \mu_2^S |$ would be the same as in CC2. Due to the extra commutator, extra terms are added to the CC2 part. Note that the double-double part of the Jacobian is usually assumed to be diagonal in CC2. In general, the Fock matrix will not be diagonal in MLCCSD and the $A_{\mu_2^T,\nu_2^S}$ block will not be zero.

Several numerical examples of MLCCSD excitation energies are provided in Section B.3.2. In general, MLCCSD reproduces the CCSD results well, even when there are large discrepancies between CC2 and CCSD. All the active spaces used were generated with Cholesky orbitals [88], see Section 3.4, which may be problematic for diffuse basis sets. It is therefore encouraging that the errors with respect to CCSD do not increase when going from cc-pVDZ to aug-cc-pVDZ. Note that the excitation vectors, $\boldsymbol{R}_i$ in Eq. (2.43), have quite large contributions from semi-external excitations in some cases. This is a useful diagnostic for the accuracy compared to full space calculations and MLCCSD is relatively tolerant of semi-external contributions.

For MLCC3, there is only one type of triple excitations and the MLCC3 response equations reduce to the full space equations with the indices in the triples part restricted to the active space. As with the ground state amplitude equations, MLCC3 response scales as $VV_A^3O_A^3$.

A challenge when testing the MLCC3 model was finding systems with large differences between CCSD and CC3 that were small enough for a full CC3 calculation. The systems presented in Section C.4.2 demonstrate that the MLCC3 model is much more sensitive to the semi-external character of the excitation vector. Both CC2 and CCSD have double excitations, so any missing CCSD correlation can be partially recovered by the CC2 double excitations. In MLCC3, there are no triple excitations outside the active space and a good active space is crucial. In Section 3.4, various alternatives to Cholesky orbitals are presented. The results in Paper C demonstrate that it is possible to obtain CC3 accuracy with an almost two orders of magnitude reduction in computational cost. For core excitations, triples are required in order to obtain quantitative results, see Chapter 4.

3.4 Active orbital spaces

A critical component of MLCC is the choice of active space. The equations presented in Section 3.2 are valid for any choice of orbital space, but the active orbitals must be able to describe the property of interest. With a poor choice of active space we will not achieve full space accuracy and in some cases the results are worse than the low accuracy model. A related challenge is to verify that the results are accurate in the absence of full space results.

One way to choose an active space is using the HF orbital energy. In this approach,

we pick a number of the highest occupied MOs (HOMO) and lowest unoccupied MOs (LUMO). This is similar to the active spaces in CAS and earlier active space CC approaches [85, 86]. This approach is straightforward and the Fock matrix remains diagonal. However, the HF orbitals are not optimised for correlated calculations and display a very slow convergence to full space results as demonstrated in Paper A. Furthermore, it is difficult to choose a consistent active space if the geometry changes and the size of the active space scales with the size of the system.

Size intensive properties can be described in a localised active space. This requires localised molecular orbitals and we used Cholesky orbitals [88] for all the results reported in Papers B to D as well as some local geometry calculations in Paper A.

Cholesky orbitals are generated by decomposing the one-electron density matrix in the AO basis. The elements of the resulting Cholesky vectors can be interpreted as the orbital coefficients of a set of localised orbitals. Each of these vectors correspond to a diagonal element in the density matrix, which again refers to an atomic orbital. In order to construct a localised active space, a set of atoms are chosen before the calculation begins. When decomposing the density matrix, we start with the largest diagonal element corresponding to an AO on an active atom and the decomposition proceeds until there are no active diagonal elements left greater than a certain threshold. The orbitals in the inactive space are generated by decomposing the rest of the matrix. Similarly, a virtual density is constructed by subtracting the one-electron density from the inverse of the overlap matrix and virtual orbitals are generated by decomposing this matrix. See Section B.2.2 for a detailed description.

After the decomposition, the Fock matrix is diagonalised in each active space. The procedure is easily generalised to more than two spaces by assigning the atoms to the different spaces and decomposing each space consecutively. In order to ensure that the relevant density is included, the decomposition always starts with the most active space.

An advantage with Cholesky orbitals is that they are easy to generate as they only require an HF calculation and are generated non-iteratively. Another advantage is that a consistent active space can be chosen for different geometries. This is achieved by storing the ordering of the decomposed diagonal elements in the initial geometry and decomposing from the same elements for each geometry. As the diagonal elements correspond to AOs that are the same for each geometry, we obtain a smooth potential energy surface. This is difficult to do using other types of orbitals.

Unfortunately, it is difficult to generate a good active space for MLCC3 using Cholesky orbitals, see Paper C. Generally, excited states have a more diffuse electron density than the ground state and requires extra diffuse orbitals, as discussed in Section 2.8. In the HF one-electron density matrix, these orbitals will have very small elements that do not contribute to the Cholesky orbitals. This is demonstrated in Section C.4.2 and we propose a solution by explicitly adding extra diffuse orbitals. While this approach solves the problem, it is not very satisfactory because

it will make the active space scale with total system size.

One approach is to use a different type of localised orbitals. There are many types of localisation schemes [89–93]. By using a localisation scheme depending on the orbitals themselves and not the HF density, it is possible to generate a local active space with the required flexibility to describe the excited states. This approach does require additional criteria to decide which orbitals to include in the active space and there is no straightforward way to keep a consistent active space when changing the geometry. Different local orbitals have been combined with the MLCC3 approach for core excitations with some success [94].

An optimal active space would contains all the orbitals necessary to describe the process being studied, but no more than that in order to keep computational cost down. Lower level models are typically qualitatively correct in their descriptions. For example, the excitation vectors of CCSD and CC3 are usually very similar. This can be exploited to construct an active space well suited to describe a specific excitation and is the basis for the automated molecular orbitals, AMOs, recently developed by Høyvik [95].

Automated MOs are similar to natural transition orbitals, NTOs, [96–98] and are generated from the eigenvectors $\{ \boldsymbol{R}_i \}$ of the Jacobian, see Eq. (2.43). Unlike the NTOs, AMOs are generated from both the single and double excitation part of the excitation vector. This greatly increases the number of orbitals that can be generated and makes it possible to generate a sufficient active space.

The starting point when generating AMOs is to construct the occupied-occupied matrix $\boldsymbol{M}$ and virtual-virtual matrix $\boldsymbol{N}$.

$$M_{ij} = \sum_a R_{ai} R_{aj} + \sum_{abk} (1 + \delta_{ai,bk} \delta ij) R_{aibk} R_{ajbk} \tag{3.8}$$

$$N_{ab} = \sum_i R_{ai} R_{bi} + \sum_{ijc} (1 + \delta_{ai,cj} \delta ab) R_{aicj} R_{bicj} \tag{3.9}$$

Diagonalising M and N produces a set of eigenvalues and eigenvectors.

$$\boldsymbol{M}\boldsymbol{u}_k = \lambda_k^o \boldsymbol{u}_k \quad \boldsymbol{N}\boldsymbol{v}_k = \lambda_k^v \boldsymbol{v}_k \tag{3.10}$$

The eigenvectors form orthogonal transformation vectors, taking the canonical MOs to the AMOs. Because the excitation vectors are normalised, the sum of the eigenvalues is 1 and the magnitude of each eigenvalue corresponds to the importance of the corresponding AMO for the excitation. An active space can then be formed by including AMOs until the sum of the eigenvalues reaches some predefined threshold.

In an MLCC calculation, one would start with a full space calculation with the lower level method to obtain the excitation vectors, CCSD for example. After generating the active space and diagonalising the Fock matrix in each space, the excitation vectors are also transformed to the new set of MOs. These are then

used as start vectors for an MLCC3 calculation. If several excited states are to be determined, it is possible to generate the transformation vectors from the sum of the $\boldsymbol{M}$- and $\boldsymbol{N}$-matrices of each excitation, but results so far indicates that it is better to generate a separate active space for each excited state. If several levels of theory are desired, we can start with a full space calculation at the lowest level and then use the excitation vectors to generate the active space for the next level in an iterative procedure. This type of orbitals require very little input from the user, hence the name.

3.5 Multilevel coupled cluster implementation

This section will briefly present the implementation of closed shell MLCC3 in Dalton [99, 100] described in Section C.3. The MLCCSD code, including transition moments [101], is only implemented as a pilot code and is currently no faster than CCSD. Note that indices in the following equations refer to spatial orbitals.

The most expensive terms to calculate in MLCC3 are the triple amplitudes,

$$t_{ijk}^{abc} = -P_{ijk}^{abc} \frac{1}{\epsilon_{ijk}^{abc}} \left(\sum_D t_{ij}^{aD} (bD\hat{|}ck) - \sum_L t_{iL}^{ab} (Lj\hat{|}ck) \right) \tag{3.11}$$

and the virtual integral terms in the $\boldsymbol{\Omega}$ contribution,

$$\Omega_{ADij} = P_{ij}^{AD} \sum_{bck} \left(2t_{ijk}^{Abc} - t_{ijk}^{cbA} - t_{ijk}^{Acb} \right) (Db\hat{|}kc) \tag{3.12}$$

In the above equations, P is a permutation operator summing over the permutations of the indices and the triple amplitudes are referred to as t because they are only included in the active T operator. Note that the amplitude t_{ijk}^{Abc} is only nonzero if the general index is active. In Eq. (3.12), the amplitudes are written with a general index to match the permutation operator.

The MLCC3 algorithm is based on the CCSD(T) algorithm originally implemented by Rendell *et al.* [102]. This algorithm exploits the symmetry of the triple amplitudes to reduce the computational effort,

$$t_{ijk}^{abc} = t_{ikj}^{acb} = t_{jik}^{bac} = t_{kji}^{cba} = t_{jki}^{bca} = t_{kij}^{cab} \tag{3.13}$$

and is centred around a triple loop over the occupied indices. The loop is restricted to the indices $i \geq j \geq k$ and the minimum number of amplitudes are calculated. An important difference between the CCSD(T) energy correction and the CC3 contributions to the $\boldsymbol{\Omega}$ vectors from Eq. (2.32) is the X_1-transformed Hamiltonian. In CCSD(T), the single contributions are discarded and two-electron integrals, Eq. (2.14), are symmetric.

$$(pq|rs) = (qp|rs) = (pq|sr) = (qp|sr) \tag{3.14}$$

In CC3, the contributions from each of these integrals must be calculated separately, leading to an overall scaling of $2 \times VV_A^3O_A^3$. One N^7 term from Eq. (3.11) and one from Eq. (3.12). Multilevel CCSD(T) has also been implemented, scaling as $1 \times VV_A^3O_A^3$.

For sufficiently small active spaces, the dominant part of the MLCC3 calculation will be integral transformation. In Dalton, integrals are stored as AO integrals and they have to be transformed to MO basis which scales as N^4O_A where N is the total number of atomic orbitals. This bottleneck can be avoided by using Cholesky decomposed integrals [69]. On the other hand, if the CC3 part is dominated by the integrals, the overall calculation will be dominated by CCSD.

The current implementation has been parallelised with OpenMP [103] which results in considerable speedups compared to the serial version. OpenMP is limited to shared memory threading which limits the number of CPUs that can be employed. In order to utilise multinode architecture, MPI is required [104]. Because the MLCC3 code is structured as a loop where each iteration is independent of the others, it would be very well suited for an MPI implementation. The only subtlety would be how to distribute the integrals, but the code is already batching the integrals to reduce I/O and a similar procedure could be utilised for parallelisation. Unfortunately, the rest of the Dalton CC code is not parallelised or MPI safe, so such an implementation would require considerable programming effort. In the future, we wish to implement a new CCSD code in Dalton making it possible to incorporate MPI.

Chapter 4

X-ray spectroscopy in multilevel coupled cluster

4.1 The core-valence separation approximation

X-ray spectroscopy is a field in rapid development. Using soft X-rays in the 250-900 eV range it is possible to selectively excite the core electrons of specific elements in a compound. For example, the carbon K-edge is typically found between 280 eV and 300 eV while the nitrogen edge is found around 400 eV. Single functional groups can then be targeted and the method provides highly local chemical information. More recently, development of free electron lasers (FEL) has made time resolved X-ray absorption spectra (XAS) possible [105–113].

In electronic structure theory, core excited states poses a number of challenges not encountered when calculating valence excited states. When solving the eigenproblem in Eq. (2.43) starting with the lowest eigenvalue and building up is no longer an option because there will typically be hundreds of states lower than the first core excited state. In CC theory, there are two solutions to this problem. One is to use the Lanczos algorithm [44–46] to solve for the whole spectrum simultaneously. For small molecules, this is a valid approach, but as mentioned in Section 2.5, large storage requirements limits the applicability for larger systems. Nonetheless, the method is valuable as a benchmark for the other approach, core-valence separation (CVS) [114–116]

The CVS approximation is based on the observation that there is a large energy gap between core and valence excited states and their interactions will be negligible. Consequently, the coupling elements in the Jacobian can be discarded. In practice, this is achieved by projecting out the corresponding elements of the trial vectors and result vectors when solving Eq. (2.43).

Table 4.1: Comparison of core excitations, in eV, of oxygen for butanal using CCSD, MLCC3 and CC3. The active space was generated using AMOs.

	S_1	n_{occ}	n_{vir}	S_2	n_{occ}	n_{vir}
CCSD	534.71	20	171	539.44	20	171
MLCC3	533.01	10	115	535.89	12	142
CC3	533.00	20	171	535.86	20	171

The projection operator, $\mathcal{P}$ acts on a vector $\boldsymbol{b}$ so that all elements without at least one core orbital index are set to zero.

$$\mathcal{P}b_i^a = 0 \quad \forall\ i \neq \text{core} \tag{4.1}$$

$$\mathcal{P}b_{ij}^{ab} = 0 \quad \forall\ i \neq \text{core}\ \wedge\ j \neq \text{core} \tag{4.2}$$

The two methods can be combined if a full core excited spectrum is desired. In Paper D we use both to assess MLCCSD for core excited states.

4.2 Multilevel CCSD core excitations

Core electrons are described by highly localised 1s orbitals. Consequently, core excitations are also highly local. Figures D.6 and D.7 demonstrates the change in electron density between the ground state and the core excited states of ethanal, calculated at the CCSD level. The excitation leads to a large reduction in electron density in the core region and a corresponding increase around it. In Paper D we explore the use of MLCCSD for core excited states of carbon and oxygen in aldehydes.

In general, MLCCSD is able to reproduce the CCSD excited states very well. For the first excited states of oxygen, a Cholesky active space with only oxygen active is able to reproduce the CCSD position of the peak with an error of 0.01 eV for ethanal and 0.03 eV for propenal. The errors in the transition moments and the excitation energies of the higher Rydberg states are a bit larger, but including the adjacent carbon atoms make the CCSD and MLCCSD spectra almost indistinguishable. Comparing results from CVS-Davidson and full space Lanczos, the errors due to CVS are very small.

For the carbon excitations, the errors in MLCCSD are somewhat larger. When several atoms of the same element are bonded directly to each other, it is necessary to use larger active spaces to reproduce the full space spectra, compare Figures D.11 and D.12. Unfortunately, the CCSD results do not compare well with experimental results, see Tables D.7 and D.8.

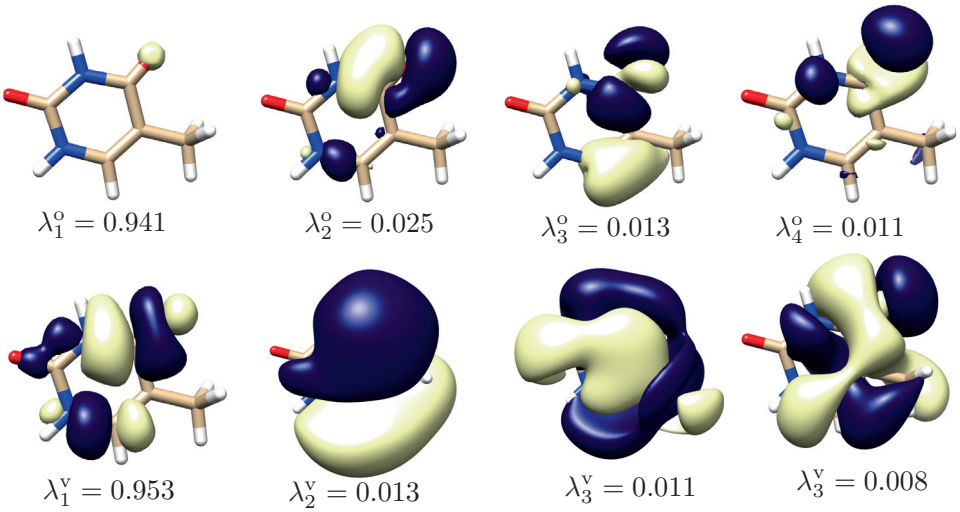


Figure 4.1: The four most important occupied (above) and virtual AMOs for the first core excitation of thymine. Basis aug-cc-pCVTZ/aug-cc-pVDZ

4.3 Multilevel CC3 core excitations

The MLCCSD results for core excited states were very encouraging, however, MLCC3 was initially disappointing [94]. Using Cholesky orbitals, the convergence towards full space results turned out to be much slower than for MLCCSD. Other types of localised orbitals converged faster, but the number of active orbitals were still too high to obtain large reductions in computational cost. The solution is to use the automated molecular orbitals developed by Høyvik [95] described in Section 3.4.

Butanal is a good test system because it is relatively small and the oxygen core excitations are localised at the end of the chain. Multilevel CCSD performed very well for this system, as shown in section D.3.4. It was also used to test MLCC3 with different types of localised orbitals with reasonably good results [94]. In Table 4.1, results calculated by Høyvik using MLCC3 and AMOs are compared with CCSD and CC3. The errors of MLCC3 are 0.01 and 0.03 eV for the two excitation energies. In comparison, the errors for CCSD are 1.71 and 3.58 eV. The basis set is aug-cc-pCVDZ on oxygen and aug-cc-pVDZ on the rest, resulting in a total of 20 occupied and 171 virtual orbitals. Note that this basis set is not large enough to reproduce experimental results, see Section 5.1.

In order to test the AMOs on a larger and more challenging system, we recalculated the core excitation energy of thymine from Paper G using aug-cc-pCVTZ on oxygen O(8), see Figure G.1, and aug-cc-pVDZ on the rest. Earlier attempts to calculate this excitation with MLCC3 and local orbitals were not very successful. Figure 4.1

shows the four most important occupied and virtual AMOs plotted with Chimera [117]. While the occupied AMOs are concentrated around O(8), the virtual AMOs appear to be completely delocalised. This explains why localised active spaces failed.

Table 4.2: Comparison of core excitations, in eV, of O(8) in thymine using CCSD, MLCC3 and CC3. The active space was generated using AMOs. Timings are given in seconds. Note that the calculations were performed on different systems and timings are not directly comparable.

	ΔE	n_{occ}	n_{vir}	t_{CPU}	n_{CPU}
CCSD	533.57	33	264	950	4
MLCC3	531.31	15	134	1370	4
CC3	531.19	33	264	122000	10

Table 4.2 contains the core excitation energies of thymine calculated at the CCSD, MLCC3 and CC3 level. The error in CCSD compared to CC3 is 2.38 eV, compared to just 0.12 eV with MLCC3. Timings for the calculations are given for one iteration of the excitation energy. They are not directly comparable because the full space calculation had to be run on the Stallo [118] cluster for a total of 491 hours, while the MLCC3 calculation could run on an iMAC at home. The theoretical speedup is

$$\left(\frac{264 \times 33}{134 \times 15}\right)^3 \approx 81 \tag{4.3}$$

With this active space, the time spent in the CCSD and CC3 part is about the same. Consequently, the CCSD part will become dominant if the active space is reduced further. This bottleneck can be avoided if several active spaces are used, as described in Section 3.4. A faster CCSD code than the one currently in Dalton would also be an improvement and this will be implemented with MLCCSD. Note that the full space calculation spent roughly 5% of the time on integral transformation while MLCC3 spent 20% of the time on this. Cholesky integral decomposition will reduce the computational cost of integral transformations [69].

Chapter 5

The accuracy of coupled cluster for core excitations

5.1 Basis set dependence

Figures 6.2, D.6 and D.7 demonstrate that core excitations involve large changes in the electron density compared to the ground state. Consequently, basis sets that describe the ground state well cannot automatically be assumed to be suitable for the excited state. A recent study of a number of small molecules [119] indicates that the basis set requirements depends on the type of excitation with Rydberg states requiring considerably larger and more diffuse basis sets.

Table 5.1 presents CC3 excitation energies computed with the aug-cc-pVXZ, aug-cc-pCVXZ, and d-aug-cc-pCVXZ basis set series, abbreviated aXZ, aCXZ, and daCXZ [59, 60]. Adding core polarising functions improves the description of the ground state. This results in higher excitation energies for all basis sets except double zeta. The aDZ basis contains only four contracted *s*-functions on oxygen, so the added core polarising function is used to describe the excited state and lowers the excitation energy. The shift induced by the core functions is more or less constant. Adding double augmentation has almost no effect on the first two excited states, but give large shifts for the higher excitations. Compared to experimental values, CC3 is a bit too low and the gap between the states is a bit too small.

In Figure 5.1, core excitation energies of water calculated with CC2, CCSD and CC3 are plotted against basis sets. All the basis set series shows a lowering of the excitation energy with basis set size. The gap between the first two excited states is described reasonably well with CCSD and CC2, but there are large discrepancies for the third excitation, particularly with smaller basis sets. From the CCSD results, the basis set converges faster when using doubly augmented basis sets. For the

Table 5.1: The first four core excitation energies (eV) of water calculated with the aug-cc-pVXZ, aug-cc-pCVXZ, and d-aug-cc-pCVXZ series of basis sets and the 6-311++G** basis set compared with experimental results by Schirmer *et al.* [120].

	S_1	S_2	S_3	S_4
aDZ	536.34	538.12	540.16	540.57
aTZ	533.43	535.20	536.84	537.13
aQZ	533.23	534.99	536.44	536.61
a5Z	533.16	534.92	536.15	536.23
aCDZ	535.59	537.37	539.40	539.81
aCTZ	533.81	535.58	537.23	537.51
aCQZ	533.65	535.42	536.87	537.04
aC5Z	533.60	535.37	536.60	536.68
daCDZ	535.57	537.33	538.15	538.20
daCTZ	533.80	535.56	536.39	536.45
daCQZ	533.65	535.41	536.25	536.30
6-311++G**	533.99	535.76	538.45	538.71
exp.	534.0	535.9	537.1	-

lowest state CCSD is too high, CC3 a bit too low, and CC2 close to experiment. Relativistic effects are not included, so we expect the accurate CC3 calculation to be a bit too low, see Section 5.3. Transition moments appear to be much less sensitive to the basis set than excitation energies [119].

For comparison, the split-valence basis set 6-311++G** is also included [121]. This basis set performs better than aCDZ, even though it is smaller. The largest exponential for the 1s orbitals is about 5485 for the Pople set compared to 11720 for aCDZ (and 164200 for aC5Z) and the ground state total energy is 0.25 eV lower for aCDZ. The Pople set contains one more contracted *s*-function and one less *d*-function and the extra *s*-function is required to describe the first excited state where 6-311++G** is 1.35 eV lower than aCDZ. For comparison, aCTZ is 2.55 and 2.73 eV lower than 6-311++G**for the two states.

Using large basis sets like d-aug-cc-pCVQZ is not feasible with CC3 for larger systems. For second row elements, this basis set includes 134 contracted basis functions per atom. Luckily, the core excitations are local, making it unnecessary to use the full basis set on the entire molecule.

Figure E.1 depicts the five most important conformers of glycine optimised using CCSD(T) and aug-cc-pVDZ basis in CFOUR [122, 123]. The first CC3 core excitation for each atom in the lowest energy conformer is given in Table E.3 calculated with various basis sets in the aug-cc-pCVXZ series. Basis sets with parentheses are

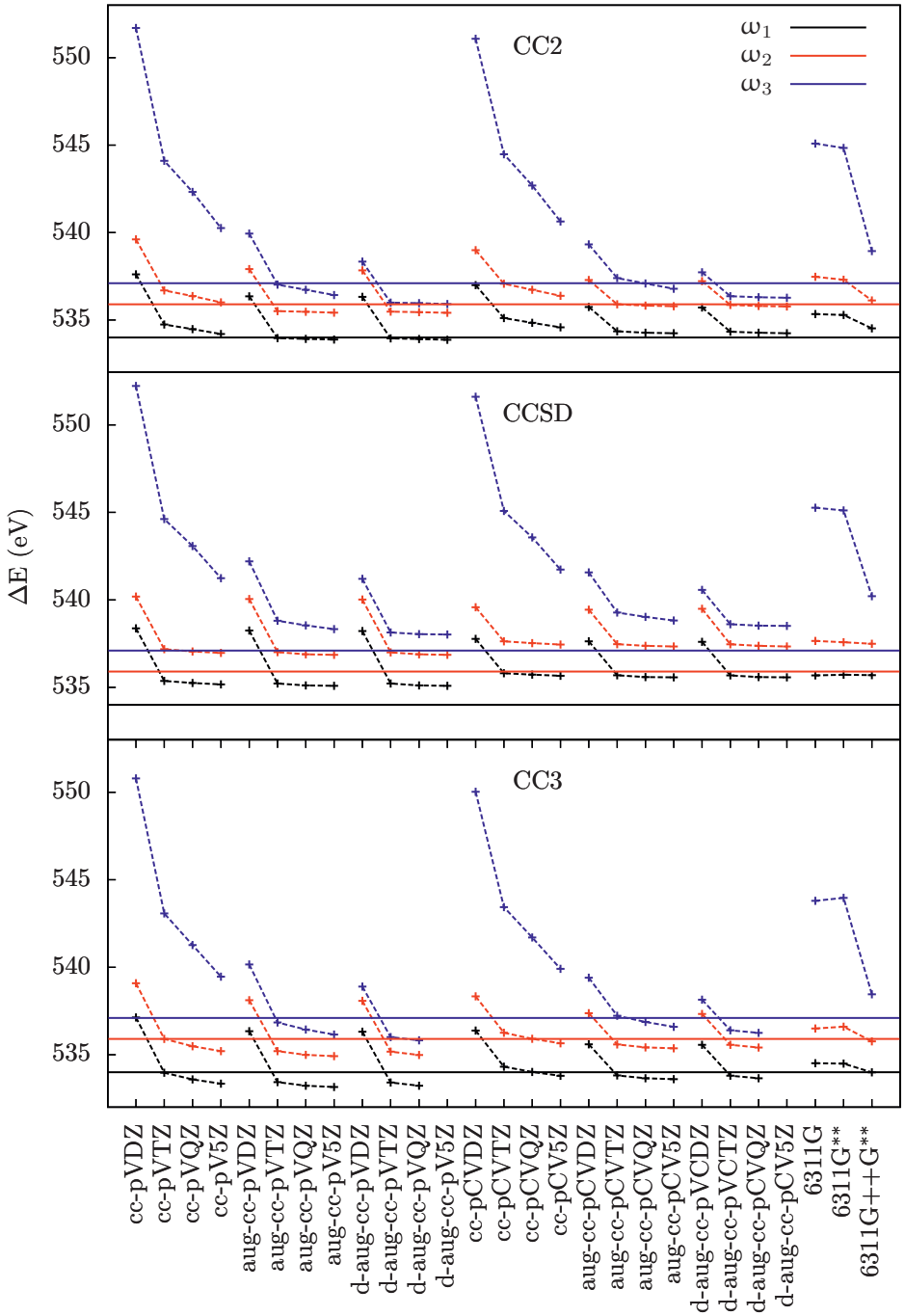


Figure 5.1: Plots of excitation energies of water against basis set using CC2, CCSD and CC3 compared to experimental results (solid lines) by Schirmer *et al.* [120].

mixed where the indicated zeta level is employed on the atom being excited and aug-cc-pVDZ is used on the rest. Comparing the excitation energies obtained with a(CT)Z and aCTZ, the largest difference is 0.15 eV while the difference between aCDZ and a(CT)Z is more than 1 eV for all atoms. With a(CQ)Z, the excitation energies are reduced another 0.02 to 0.16 eV.

Comparing to the experimental values obtained by Plekan *et al.* [124], the results vary depending on the element. Oxygen excitation energies are too low, nitrogen is very close and carbon is too high. This is similar to the results for molecular nitrogen, N_2, see Paper F, and thymine, see Paper G. A plausible explanation for this behaviour is that relativistic effects are not included, see Section 5.3. The results for C(1) is considerably worse than for the other atoms. This peak is hard to see because the other carbon has a very intense signal. In reference [124], the assignment was based on comparison to other amino acids and the SCF static-exchange method (STEX) [125]. Consequently, the experimental value is uncertain [126]. Glycine is a molecule with many conformers [127], however, calculations only showed small geometric effects on the core excitations, consistent with the experiment.

In Table E.4, calculated core ionisation energies are compared to experiment. In order to calculate the electron binding energy, a highly diffuse *s*-function is added to the basis set. By projecting out elements of the excitation vector not involving this function, the model is effectively removing an electron [116, 128]. The results are similar to the core excitations with aCDZ being more than an eV higher than a(CT)Z, small differences between a(CT)Z and aCTZ and a(CQ)Z being a bit lower. Unlike the excitation energies, the carbon ionisations are close to experiment while oxygen is more than 1 eV too low. This may be due to the combination of CVS and ionisation effectively fixing two indices, leaving only singles and approximate doubles to describe the relaxation. For nitrogen gas, N_2, the ionisation energy is well reproduced, see Paper F.

5.2 Vibrational effects

Another effect that may affect the experimental spectra is the vibrational levels of the excited state and changes in the zero point vibrational energy. For most molecules it is difficult to resolve any vibrational structure in NEXAFS spectra, though they can have large effects, particularly on intensity [129]. Nitrogen, N_2, displays a clear vibrational structure in both electron energy loss spectroscopy [130] and NEXAFS, see Paper F. Because N_2 only has one degree of freedom, it is possible to map the entire potential energy surface of both the ground and excited states. The vibrational levels and intensities of the vibrational transitions can then be obtained by fitting a Morse potential and solving for the vibrational levels [131–133].

In Figure F.1, the calculated CC3/d-aug-cc-pCVQZ NEXAFS spectrum is plotted together with an experimental spectrum obtained from the SOLEIL synchrotron in France [134]. The theoretical spectrum is plotted with Lorentzian line broadening and an empirical line broadening constant. Intensities in the spectra are normalised to the first vibrational peak. The error in the position of the first peak is 0.13 eV, while the error in the gap up to the first Rydberg state is 0.06 eV. Without double augmentation, the Rydberg states were shifted 0.4 eV up. In comparison, CCSD is 0.90 eV too high and ADC(2)-x is 1.66 eV too low compared to experiment.

Figure F.2 is a close up of the first peak in the spectrum and reveals that the distance between the vibrational levels is about 0.02 eV too high, see Table F.1. This happens because the CC3 potential surface is too deep. The ADC(2)-x potential is too shallow and the separation of the vibrational levels is too small. The equilibrium bond length of the first core excited state is 0.006 Å shorter than the one obtained in experiment [130] while the ground state is 0.002 Å too long [135]. Increasing the basis set and level of theory improves the ground state geometry [136].

For the Rydberg excited states, the vibrational levels are also too far apart, but the difference between the states is very good. Note that the second $^1\Sigma_u^+$ state is very weak and in between the vibrational levels of the second $^1\Pi_u$ state, so its exact position is uncertain, see Figure F.4. The intensities for these states are too low, but they were calculated at the CCSD level. For the first core excitation, CCSD is 0.77 eV too high compared to CC3 while it is about 1.1 eV too high for the Rydberg states. It is possible there are similar errors in the intensities. Calculations with aug-cc-pCV5Z and Lanczos calculations in aug-cc-pCVTZ indicate that basis set and CVS errors raises the excitation energy by about 50 meV. These sources of error are probably small compared to the largest known unknown, relativistic effects.

5.3 Relativistic effects

Relativistic effects are usually ignored in electronic structure calculations on light elements because they only affect the core electrons. Core electrons generally do not contribute much to molecular geometries, valence excitations and other "normal" calculations [4, 5]. Furthermore, full four component relativistic calculations require different basis set expansions for the so-called large and small components. While this does not affect the scaling of the model, it can increase the prefactor by several orders of magnitude [137].

The most important relativistic effects for closed shell systems are the mass-velocity effect that contracts the core orbitals, reducing the energy, and the Darwin effect which is related to the smearing of the electron charge and raises the energy [138]. These effects can be approximated for the one-electron term of the Hamiltonian by rescaling the one-electron integrals. This approach is known as the second order

Douglas-Kroll-Hess (DKH2) correction, [139, 140]. Using the DKH2 correction results in a more or less constant shift for the core excitation energies. Carbon excitations are shifted up by about 0.08 eV, Nitrogen by 0.20 eV and Oxygen by 0.34 eV.

Effects that are not accounted for by the DKH2 correction include spin-orbit coupling, two-electron mass-velocity and Darwin terms and the Gaunt term. For light elements, these effects are important and will generally decrease the excitation energy, particularly the Gaunt term [141]. As the relativistic effects mostly affect the core electron, they provide a constant shift for the whole spectrum and the DKH2 correction has not been employed in the calculations reported in this thesis. This is a subject that should be explored further in the future.

Chapter 6

Using coupled cluster to probe $\pi\pi^*$-nπ^* transitions in thymine

6.1 Excited states of thymine

Dimerisation of thymine induced by UV light is believed to be an important factor in the development of skin cancer [142]. As a result, the molecule has been extensively studied, both experimentally and theoretically [110, 112, 143–147]. However, the excited state dynamics of thymine is still debated and different studies have arrived at different conclusions. When exposed to UV light, the molecule will be excited to a $\pi\pi^*$ state with A$'$ symmetry. What happens next is unclear. Molecular dynamics with CAS indicates that the molecule is trapped in a local minimum in the $\pi\pi^*$ state [143] while X-ray ionisation experiments indicates a rapid transition to the lower nπ^* state with A$''$ symmetry [110].

Figure 6.1 shows simplified orbital diagrams for the relevant electronic states. The Koopmans' picture [148] described by the diagram is oversimplified. For example, the lowest singlet excited state is nπ^* in the ground state equilibrium geometry, even though HOMO has π-symmetry. Nevertheless, the diagrams are useful to describe the electronic states conceptually. When the molecule is excited with UV light an electron hole is formed, represented by half filled orbitals in the diagram. If the molecule is core excited, the 1s electron can fill the electron hole, resulting in a reduction of the core excitation energy. The final state will be the same as if the molecule was excited from the ground state so the reduction in excitation energy corresponds to the valence excitation energy.

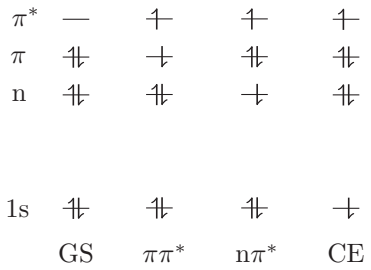


Figure 6.1: Simplified orbital diagram demonstrating the configuration of the ground state (GS), relevant valence excited states and the core excited state (CE).

Whether the core electron is excited to the electron hole depends on the transition moment between the excited states. The transition moments depends on the overlap integrals between the orbitals involved in the two states. Because the core electron is described by a highly local 1s orbital, the magnitude of the transition moments can be qualitatively predicted by the locality of the electron hole. Figure 6.2 shows the difference in the electron density between the ground state and the different excited states while Figure G.5 shows the difference between the core excited state and the other states. From the density plots, it is clear that the $n\pi^*$ electron hole is localised around oxygen 8 (O(8)), see Figure G.1. This makes it possible to distinguish between the two states. The same argument can be made based solely on the HF orbitals, see Figure G.1.

The core excited state is very high in energy, 530 eV above the ground state, and is consequently very unstable. Almost instantaneously, a valence electron will collapse into the core hole [149]. In order to get rid of the excess energy, the molecule will emit an Auger electron [150–152].

6.2 The experiment

The goal of the experiment described in Paper G is to demonstrate the capability of time resolved X-ray spectroscopy to distinguish between excited states of organic molecules. Currently, few facilities can perform such experiments, which was performed with the Linac Coherent Light Source (LCLS) free electron laser at SLAC National Accelerator Laboratory. In the experiment, thymine molecules were excited with a 267 nm (4.64 eV), UV pulse. The molecules were then probed after a varying time delay by a soft X-ray pulse. A monochromator was used to control the wavelength of the probe pulse. Measuring the absorption of X-rays is challenging, but the Auger electrons can be measured using a mass spectrometer. For details of the experiment, see Section G.2.1.

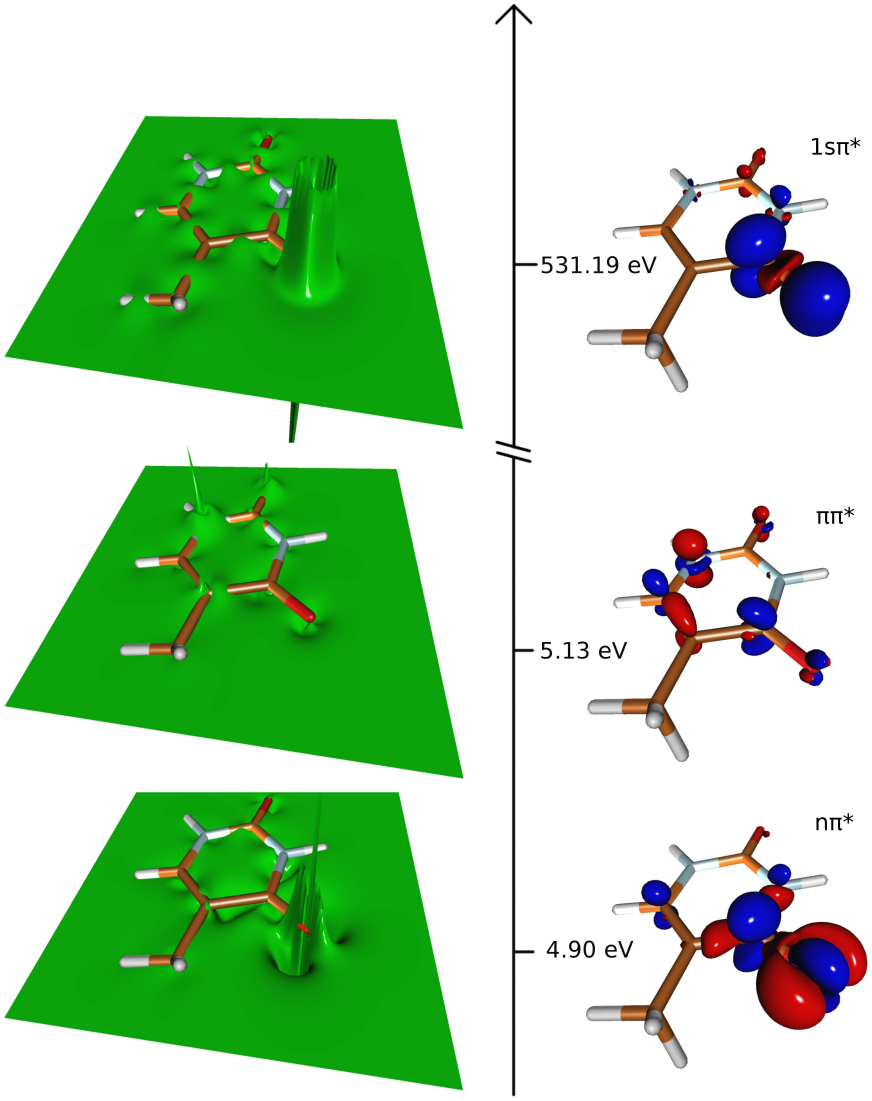


Figure 6.2: Change in electron density from the ground state to the $n\pi^*$, $\pi\pi^*$ and $1s\pi^*$ excited states of thymine calculated with CCSD and aug-cc-pCVDZ/aug-cc-pVDZ. Red isosurfaces indicate electron density reduction and blue indicate density increase, isovalue 0.01. Excitation energies are calculated with CC3 and aug-cc-pCVDZ/aug-cc-pVDZ (valence) and aug-cc-pCVTZ/aug-cc-pVDZ (core).

An earlier experiment using delayed ultrafast X-ray Auger probing (DUXAP) indicated a short lived excited state appearing immediately after the probe pulse with a signal around 508 eV and starting to disappear after 100 fs [110]. While the first signal faded, another signal started to build strength around 490 eV which lasted for more than 20 ps. This was interpreted as population being transferred from the $\pi\pi^*$ state to the $n\pi^*$. Because photoionisation experiments are sensitive to both electronic states and molecular geometry, NEXAFS should provide less ambiguous results.

The NEXAFS results are summarised in Figure G.2. A new feature appears in the spectrum at 526.4 eV, 5.0 eV lower than the ground state peak at 531.4 eV. The new feature appears about 60 fs after the bleach of the ground state signal and is still visible after 20 ps. It is natural to interpret this feature as buildup of population in the $n\pi^*$ state, however, accurate theoretical calculations are required to verify this.

6.3 Molecular geometries

The first step in predicting experimental results is obtaining good molecular geometries. Initially, we used geometries obtained using CAS from reference [143]. When comparing these geometries to geometries optimised using CC2, provided by Basile F. E. Curchod, qualitative differences were observed. In particular, the $n\pi^*$ state was always lower in energy than the $\pi\pi^*$ state when excitation energies were calculated with CCSD for the CAS geometries. With the CC2 geometries, the ordering changed and the $\pi\pi^*$ state was lower in the $\pi\pi^*$ stationary point.

To obtain more accurate geometries, new geometry optimisations were performed. The ground state geometry was optimised with CCSD(T) in CFOUR [122, 123] using aug-cc-pVDZ and is given in Table G.1. Excited state geometries were optimised in Q-Chem [153] using CCSD and aug-cc-pVDZ. Obtaining the excited state geometries was considerably more challenging than the ground state, particularly the $\pi\pi^*$ geometry. At the first attempt, the $n\pi^*$ state converged to a stationary point while the $\pi\pi^*$ state immediately passed through an intersection, making it necessary to stop the calculation and start again from the new geometry. This time, the energy minimisation did not converge but the molecule started bending and eventually hit an intersection with the ground state. At the third attempt the optimisation was started from an almost converged geometry and a $\pi\pi^*$ stationary point was obtained.

Vibrational frequencies were calculated using CFOUR for each of the stationary points to verify that they were in fact minima. The ground state was a minimum but neither of the excited states were, though the imaginary frequencies of the $n\pi^*$ geometry were very small. Slightly perturbed geometries were generated and the optimisations started again. The $n\pi^*$ optimisations moved around on a flat

potential energy surface dominated by rotations of the methyl group until it settled in a geometry distorted out of the plain. The final geometry, given in Table G.2, was less than 0.01 eV lower in energy than the planar stationary point but a frequency calculations confirmed it was a minimum. The $\pi\pi^*$ state did not converge, but went through an avoided crossing and ended up in the $n\pi^*$ minimum. Table G.3 contains the $\pi\pi^*$ saddle point geometry and the frequencies are given in Table G.4.

The main differences between the ground state and excited state geometries is a stretch of the C(5)-C(6) double bond for the $\pi\pi^*$ saddle point and a stretch of the O(8)-C(4) bond for the $n\pi^*$ minimum energy geometry. In addition, the $n\pi^*$ is distorted out of the plane.

6.4 Excited state calculations

Excitation energies, both core and valence, were calculated using Dalton [99] and all calculations except the CC3 core excitations used aug-cc-pCVDZ on the oxygens and aug-cc-pVDZ on the rest. The extra core correlation functions have almost no effect on the valence excitation, but were included for consistency. The new implementation described in Paper C were used for the CC3 calculations.

Figure G.1 B shows the calculated CCSD energies calculated for geometries interpolated between the various stationary points. There is a barrierless path from the Franck-Condon (FC) point [154, 155] to the conical intersection and the $\pi\pi^*$ saddle point. With CC3, the excitation energies of both states are reduced, but the effect varies so that the difference between the states is reduced for all three geometries, see Tables G.5 and G.6. This brings the intersection closer to the saddle point. Valence excitation energies were also calculated using aug-cc-pVTZ and CC2 and compared to aug-cc-pVDZ calculations. The largest difference between the two basis sets is 50 meV, indicating that aug-cc-pVDZ is sufficient for the valence excitations. Oscillator strengths were calculated with CCSD linear response in the length gauge and the oscillator strength to the $\pi\pi^*$ state is predicted to be several orders of magnitude larger than the $n\pi^*$ state. Two higher states were also calculated, but they do not intersect with the lower states and have very low oscillator strengths in the FC geometry.

The core excited states were calculated using the CVS approximation. Excitation energies were calculated using CC3 with auc-cc-pCVTZ on the oxygen being excited and aug-cc-pVDZ on the rest, see Table G.7. Oscillator strengths were calculated with CCSD linear response and are given in Tables G.8 and G.9.

To verify that excitations from the $n\pi^*$ state is indeed more intense than those from the $\pi\pi^*$, transition moments between excited states are required. In principle, transition moments from a valence excited state to a core excited can be calculated in the standard way using Eq. (2.44) which requires solving for $X^B_\mu(\omega_k \pm \omega_m)$. Solving for the plus combination works fine, but the minus combination corresponds to

an energy range with a large number of nearly degenerate Rydberg states. Consequently, it is very hard to converge the eigenproblem. It is possible that the expression in Eq. (2.44) can be rewritten to avoid solving for this vector, however, in EOM it does not appear at all. For this reason, the oscillator strengths given in Tables G.8 and G.9 between excited states were calculated using EOM-CCSD. The EOM-CCSD ground state oscillator strengths for core excited states were also calculated and the difference between CCSD-LR and EOM-CCSD is negligible.

6.5 Comparing results

Theoretical and experimental spectra are compared in Figure G.3. The double peak from the ground state is well reproduced with CC3 predicting excitation energies of 531.19 and 532.23 eV compared to experimental values of 531.4 and 532.3 eV. In the predicted spectrum, the first peak is somewhat more intense than the second while the opposite is observed in experiment. The reason for this discrepancy is most likely due to the beginning of the ionisation edge lending intensity to the second peak. Because the area of the theoretical spectrum is normalised to the experimental peaks, the predicted peaks appear too intense.

The predicted peak from the O(8) core in the nπ^* state is at 526.05 eV which fits well with the experimental feature at 526.4 eV, especially when considering that the ground state peak is also a bit too low. However, this is not enough to definitely assign the feature because the peaks corresponding to the $\pi\pi^*$ state in the FC and saddle point geometries are at 526.29 and 526.33 eV, respectively. Oscillator strengths, given in Table G.8 and G.9 are needed to distinguish the two states. The predicted oscillator strength from the O(8) core in the nπ^* state and minimum energy geometry is predicted to be more than an order of magnitude stronger than the $\pi\pi^*$ state in the FC or saddle point geometries. As a result, the predicted peaks from the $\pi\pi^*$ in Figure G.3 are barely visible. Note that the excitations from O(7) are predicted to be dark in the excited state.

Taken together, the experimental and theoretical results indicate that the relatively long lived feature is indeed the nπ^* state. With a lifetime of less than 100 fs, it appears unlikely that there is minimum in the $\pi\pi^*$ state and access to the intersection must be barrierless. The current results say nothing about the fate of the nπ^* state. Molecules in the minimum will be hot and it is possible they will eventually find a path on the potential energy surface leading to an intersection with the ground state. Another possibility is that an intersystem crossing takes place and the system enters a triplet state before reaching the ground state [147]. This is a subject that requires further study.

This project demonstrates that XAS can be used to distinguish between electronic states, even at femtosecond time scales. Combined with high level theoretical techniques, it is possible to assign signals with high precision and the evolution of

the states can be observed directly. While the results in the gas phase may not be directly transferable to thymine in DNA, it seems likely that the reactive $\pi\pi^*$ is short lived in the cells. We are not vampires and can walk around in sunlight.

Chapter 7

Conclusion and future work

This thesis has presented the ongoing development of multilevel coupled cluster theory. The code presented in Paper C is probably the fastest CC3 code currently available and has already been used extensively. For the rapidly developing field of X-ray spectroscopy, CC3 can reliably produce results accurate to within a few tenths of an eV.

Reaching basis set convergence is challenging for core excited states. In order to get a good description of the ground state, core correlation functions must be included. For double zeta basis, this is just one extra *s*- and *p*-function, while for triple zeta, the extra core functions are two *s*-, two *p*- and one *d*-function. For the larger sets, there are even more extra core functions. For the first core excited state, diffuse functions may not be required, but for Rydberg excitations diffuse functions are mandatory. Doubly augmented basis sets are recommended for such states.

For valence excitations, aug-cc-pVDZ can provide good results, at least for the lowest states. This is not the case for core excitations and even the smaller 6-311++G** set performs better than aug-cc-pCVDZ. The Pople set contains an extra relatively tight *s*-function, while aCDZ contains an extra diffuse *d*-function. The extra *s*-function appears to be important for the description of the "volcano" seen in for example Figure 6.2. For quantitative accuracy, larger sets are required and the (d-)aug-cc-pCVXZ series appears to converge steadily towards the basis set limit.

When comparing to experiment there appears to be a systematic discrepancy, as noted in Section 5.3. A likely explanation is relativistic effects, which is consistent with the fact that the discrepancy depends on the element and grows with nuclear charge. Another potentially important effect is coupling with vibrational states. Studying the effect of vibrations is complicated by the fact that it in most cases is not possible to discern vibrational levels in experimental spectra, nitrogen being an exception. In addition, sampling of the potential energy surface is required which

is very expensive. There have been some studies on the magnitude of these effects [129, 141], but their total effect is still unclear.

The downside to CC3 is, of course, its high computational cost. Currently MLCC3 ground state and excitation energies have been implemented properly, but lack of a good procedure to generate active spaces has so far limited the application of the multilevel part. Automated molecular orbitals have the potential to change this. As demonstrated in Section 3.4, these orbitals make it possible to construct an active space tailor made for the excitation in question with very little input from the user. While the steep scaling of CC3 limits the size of the systems that can be treated, it also means that a relatively modest reduction in the number of active orbitals can give large reductions in computational cost.

Cases where the excited states are nearly degenerate are potentially challenging, particularly if the ordering of the states is changed when going from CCSD to CC3. Highly diffuse excitations may also be complicated and a procedure to deal with such cases is required. A possible explanation is that the CVS approximation is not orbital invariant. Double zeta basis fails to describe the physics as demonstrated in Section 5.1 and may exacerbate the problem. Cases where MLCC3 fails can be identified quite easily by inspecting the excitation vector. If all the involved orbitals are active, the errors compared to the full space are very small. Currently, the contributions from the different spaces are printed automatically, making it easy to evaluate the quality of the results.

As mentioned in the end of Chapter 6, the fate of the $n\pi^*$ state in thymine is unknown. In order to investigate the possible role of triplet states, it is necessary to implement CC3 triplet excitations. This should be a relatively small modification of the singlet code [156]. Triplet excitations will also be needed in a proposed project at SLAC involving the excited states of nitrobenzene.

All the oscillator strengths reported in this thesis were computed using CCSD. The assumption is that CCSD intensities are similar to those of CC3, noting that one is usually only interested in relative oscillator strengths. For valence excitations, this is a reasonable assumption [157], however, as far as the author is aware, nobody has done a study of CC3 intensities for core excitations. Doing so requires the implementation of the left matrix vector product of the Jacobian and the $\boldsymbol{F}$-matrix, which will be the next implementation project after triplet states.

More long term, a proper implementation of MLCCSD is necessary. This is a large project, because such an implementation would ideally replace the ageing CCSD code in Dalton. Furthermore, it should be implemented with MPI, or at least be MPI safe, making it possible to implement MLCC3 in parallel.

Paper A

The Extended CC2 model ECC2

Mol. Phys. **111**, 1109 (2013)

The Extended CC2 model ECC2

Rolf H. Myhre[1], Alfredo M. J. Sánchez de Merás[2] and Henrik Koch[1*]

[1]*Department of Chemistry, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway*

[2]*Institute of Molecular Science, University of Valencia, ES-46071 Valencia, Spain*

Abstract

We present a size-extensive extension to the CC2 model that avoids the complications with quasi-degeneracies that are present in the CC2 model and related perturbation theory based approaches. The formulation also provides a consistent model for treating different parts of a molecular system at different levels of electron correlation. Such a subsystem approach leads to large reductions in the computational requirements without compromising the accuracy. In this initial study we focus on static molecular properties.

A.1 Introduction

Coupled cluster response theory [28, 29, 31, 32, 158–161] has proven a highly useful computational model in quantum chemistry. For instance, CCSD excitation energies can be calculated routinely for molecules with 25 atoms or more. For an explanation of the standard acronyms used in coupled cluster theory, see Ref. [160, 161]. The CC2 model [23] was originally developed with the purpose of calculating size-extensive and size-intensive dynamic molecular properties, such as frequency dependent polarizabilities, excitation energies and transition moments. The computational scaling of the CC2 model is N^5 compared to N^6 for CCSD, when the number of electrons increases with the total number of orbitals N. The method enters naturally in the hierarchy of coupled cluster [23] models CCS, CC2, CCSD, CC3 and CCSDT, etc that offer a systematic approach to the electron correlation problem. The accuracy of CC2 is expressed in orders of the fluctuation potential and the energy is correct to second order, just like the MP2 model. The CC2 linear response function is only correct to first order and the single replacement dominated excitations are correct to second order. However, due to the perturbation treatment of the double excitation cluster operator, the double replacement dominated electronic excitations are of zeroth order. For a recent overview of CC2 response theory implementations we refer to Hättig and coworkers [162].

The accuracy of the CC2 model has been extensively tested in the literature. Excitation energies have been shown to be significantly less accurate than there corresponding CCSD excitation energies [23]. In a recent study, Korona [163] showed that CC2 static average polarizabilities deviate around 6% from the CCSD values. In the same study, the anisotropy polarizabilities deviate as much as 18%. For some smaller molecules the CC2 shielding constants [164] show a difference of 10% for nitrogen comparing to CCSD. The optical rotatory power is a complicated property to calculate and 50% deviation between CC2 and CCSD is not unusual [165]. Finally, for quasi-degeneracy situations the CC2 model breaks down due to the perturbation theory treatment of the double excitations [166]. Although there are many cases where CC2 is more accurate than indicated above, we see a need to improve the accuracy of the CC2 model without increasing the computational scaling to the N^6 level.

In this paper we will explore a simple extension of the CC2 model that preserves the overall N^5 scaling. We will partition the double excitation space such that a small active set are treated to infinite order and the remainder are treated by perturbation theory. There are many ways to select active spaces either from a perturbation theory perspective or from a chemical perspective. We have recently demonstrated that subsystems of a molecular system can be efficiently defined using the Cholesky decomposition of the one-electron density matrix [88]. In this way electron correlation is introduced selectively in the molecule. The combination of this flexibility with the partitioning of the double excitations we are able to define models where one part is effectively described by CC2 and the other part by CCSD. The associated wave function is fully anti-symmetric across the borders between

the different levels of description. As we shall see the coupling between the different levels are consistently determined within the coupled cluster equations. The new method will be named extended CC2 and we will use the acronym ECC2.

The standard black box property of coupled cluster methods is lost when we introduce this partitioning and apply different levels of correlation treatment to different parts of the system. Previous experience with the selection of active spaces for multi-configuration methods has a long history of complications [167]. However, we propose to select active spaces at an atom level rather than at an orbital level, making the selection process more chemically intuitive. The advantage of selecting subsystems is that the computational scaling becomes independent of the size of the surrounding system. Even though density functional theory (DFT) is a fully black box method, we believe a subsystem approach will be able to compete with the scaling of DFT.

A.2 The ECC2 model

Initially we divide the cluster operator into two parts, T and S, that together form the complete cluster operator. In this notation the coupled cluster wave function is expressed as

$$|CC\rangle = \exp(T+S)|HF\rangle. \tag{A.1}$$

For each of the cluster operators T and S we associate a projection manifold that we denote $\{\langle\mu^T|\}$ and $\{\langle\mu^S|\}$ respectively. This **implies** the amplitude equations

$$\begin{aligned} \langle\mu^T|\exp(-T-S)H\exp(T+S)|HF\rangle &= 0 \\ \langle\mu^S|\exp(-T-S)H\exp(T+S)|HF\rangle &= 0 \end{aligned} \tag{A.2}$$

where we have assumed the projection manifolds are orthogonal to the Hartree-Fock reference $|HF\rangle$. The energy is determined from the standard expression

$$E = \langle HF|H\exp(T+S)|HF\rangle. \tag{A.3}$$

The strategy is now straightforward. The T operator will be treated to infinite order and the S part will be determined by perturbation theory. The standard CC2 model is obtained when $T = T_1$ (single excitations) and $S = S_2$ (double excitations). The ECC2 parametrisation opens the possibility for the T operator to include, besides the single excitations, a subset of double excitations and the S operator contains the remaining double excitations. To be more explicit we write the operators as

$$\begin{aligned} T = T_1 + T_2 &= \sum_{\mu_1} t_{\mu_1}\tau^\dagger_{\mu_1} + \sum_{\mu_2\epsilon T} t_{\mu_2}\tau^\dagger_{\mu_2} \\ S = S_2 &= \sum_{\mu_2\epsilon S} t_{\mu_2}\tau^\dagger_{\mu_2} \end{aligned} \tag{A.4}$$

and we also introduce the T_1 transformed Hamiltonian

$$\hat{H} = \exp(-T_1) H \exp(T_1). \tag{A.5}$$

The T-amplitude equations now take the following form

$$\begin{aligned} &\langle \mu_1 | \hat{H} + [\hat{H}, X_2] | HF \rangle = 0 \\ &\langle \mu_2^T | \hat{H} + [\hat{H}, X_2] + \frac{1}{2}[[\hat{H}, X_2], X_2] | HF \rangle = 0 \end{aligned} \tag{A.6}$$

that are identical to the standard CCSD equations with amplitudes $X_2 = T_2 + S_2$. However, due to the limited size of the doubles projection manifold the computational scaling is reduced as compared to CCSD. The scaling will depend on how the active double excitations are chosen and we now discuss these aspects.

We will only consider two types of active spaces (I and II) for the T operator. For the first type, we construct double excitations among a selection of occupied and virtual orbitals that we denote as active orbitals. This type of active space is made up of so-called internal excitations. The type II active space will, besides the type I excitations, also include excitations to and from external orbitals in such a way that at least one orbital index is active, denoted semi-external excitations. When T is a type II active space, then S-excitations are only among the external orbitals and we call them external excitations.

The computational scaling of the Eq.(A.6) for the double excitations can now be determined. When T is a type I active space we obtain by inspection of the coupled cluster equations a maximum scaling of $\mathrm{kO^2V^2}$, where O and V is the total number of occupied and virtual orbitals respectively. This is without counting the MO-transformation that scales as $\mathrm{kN^4}$. The k-factor only depends on the size of the active space. For instance, the computationally demanding B-term [168] that in full CCSD scales as $\mathrm{V^4O^2}$ now only scales as $\mathrm{kV^2}$. The prefactor k is for the B-term equal to $\mathrm{O_I^2V_I^2}$ where $\mathrm{O_I}$ is the number of active occupied orbitals in the type I active space and $\mathrm{V_I}$ is the corresponding number of virtual orbitals. We observe that for a fixed size of the active space, the added computational cost of ECC2 is of the order $\mathrm{O^2V^2}$, which is less than the CC2 model itself. When the active space is of type II, the computaional reduction is minimal. For example, the B-term scales as $\mathrm{kV^4O}$.

The S-amplitude equations will be solved by perturbation theory through first order in the fluctuation potential U. The Hamiltonian is partitioned into the Fock operator F and the U operator such that $H = F + U$, where we assume the orbitals are canonical and thus the Fock operator is diagonal [169]

$$F = \sum_p \varepsilon_p E_{pp}. \tag{A.7}$$

Using the identity

$$\langle \mu_2^S | [F, T] | HF \rangle = 0 \tag{A.8}$$

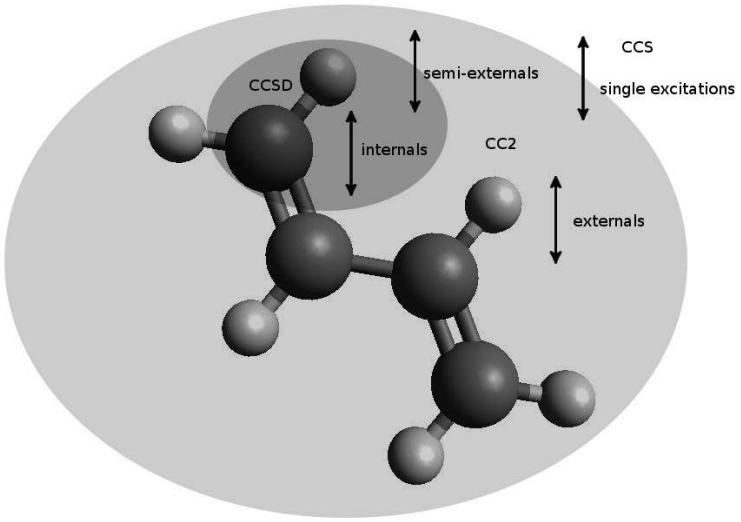


Figure A.1: The 1-3-butadiene molecule can be imagined to be part of a larger system that is described by a CCS wave function. All orbitals are allowed to relax through the T_1 cluster operator.

and that for type I active spaces we have the identity

$$\langle \mu_2^S | [[\hat{H}, T_2], T_2] | HF \rangle = 0, \tag{A.9}$$

we obtain the first-order equation for the S-amplitudes

$$\langle \mu_2^S | [F, S_2] + \hat{H} + [\hat{H}, T_2] | HF \rangle = 0. \tag{A.10}$$

This equation is valid for type I active spaces, but when considering type II active spaces, the term in Eq. (A.9) must be included when solving for the S-amplitudes. The zero'th-order condition, $\langle \mu_2^S | [F, S_2] | HF \rangle = 0$, is fulfilled when the amplitudes are zero and the first non-vanishing S_2 contribution is first order. The computational scaling of the commutator term in Eq. (A.10) for type I active spaces is similar to Eq. (A.6) and scales as kV^2O^2.

Equations (A.6) and (A.10) defines the ECC2 model that will be explored in the following sections. We may also consider models where the double excitations, both T_2 and S_2, are limited to a reduced orbital space and only T_1 includes excitations in the entire molecule. This types of model will be useful when considering a subsystem partitioning.

Before turning attention to some illustrative applications we will describe our subsystem approach. The idea of treating different parts of a molecular system by different level of accuracy was pioneered with Morokuma's ONIOM model [170]. Our approach is different as we use the Cholesky decomposition to define occupied

and virtual orbitals for subsets of atoms, see Ref. [88, 171] for details, and we can in this way define several levels of subsystems. Figure A.1 illustrates the principle. Initially we calculate the SCF one-electron density matrix in the atomic orbital basis. Then, based on a selection of active atoms, we decompose the density matrix only selecting diagonals that corresponds to active atoms. When the active diagonals are smaller than a given treshold we stop. The generated Cholesky vectors are now used as active occupied orbitals and they are orthogonal [171]. Subtracting the active density from the original density we can now repeat the process for the next subsystem. Virtual orbitals are generated in a similar manner starting from the virtual density [88]. The main difference between ECC2 and the ONIOM approach is that the ECC2 wave function is anti-symmetric across all levels of approximation and excitations between the different levels are included consistent with the coupled cluster equations. This is different from the ONIOM model that is not based on an approximate wave function for the entire system.

A.3 Illustrative applications

A.3.1 Quasi-degenerate cases

Describing molecular dissociation, and chemical reactions, is problematic with perturbation based theories. The problem arises when bond breaking leads to HOMO and LUMO becoming quasi-degenerate. Correction terms will then go towards singularities [160]. For second order theories, the highest order correction is negative, so the total energy goes towards negative infinity. In the following examples, we demonstrate that the ECC2 model remedies this problem by treating the HOMO and LUMO to infinite order. In Eq. (A.10) it is possible to discard the commutator term to reduce computational cost at the expense of accuracy. This is done in some of the following examples to assess the effect. In these cases, ECC2a refers to the model where the commutator terms are omitted, while they are included in ECC2b. ECC2 will refer to the model including the terms unless otherwise stated.

In figure A.2, the total energy of lithium hydride is plotted for a dissociation process. All models use aug-cc-pVDZ and the ECC2 models employ an active space including only HOMO-LUMO excitations. Not surprisingly, SCF performs poorly and far overestimates the dissociation energy because SCF wave functions dissociates into H^- and Li^+. The perturbative methods, MP2 and CC2, fail due to quasi-degeneracy. For such a small system, only four electrons, FCI, CCSD and ECC2-II are indistinguishable. At equilibrium distance, ECC2-I is no better than the other perturbative models. However, ECC2-I does not diverge at large internuclear distances and coincides with FCI when the distance is larger than 5 Å.

A similar behaviour is observed for the sodium dimer in Figure A.3. Only HOMO and LUMO are included in the active space and the basis is aug-cc-pVDZ. Again MP2 and CC2 fail to describe the correct behaviour, ECC2-II closely matches

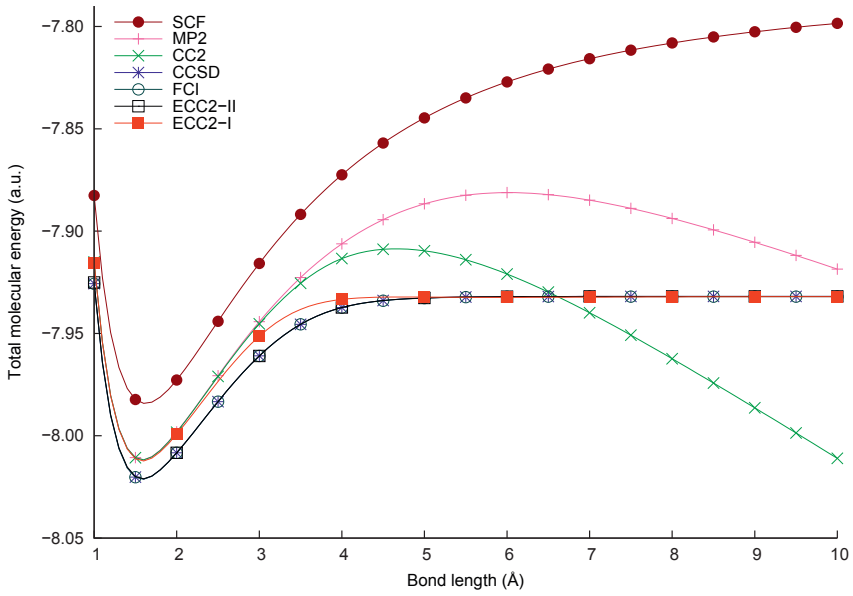


Figure A.2: Total energy curves of lithium hydride.

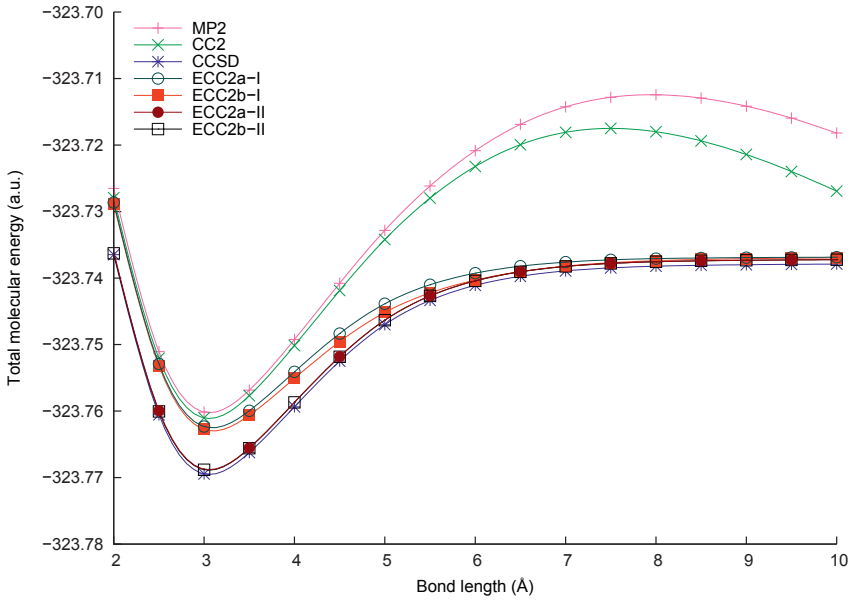


Figure A.3: Total energy curves of sodium dimer.

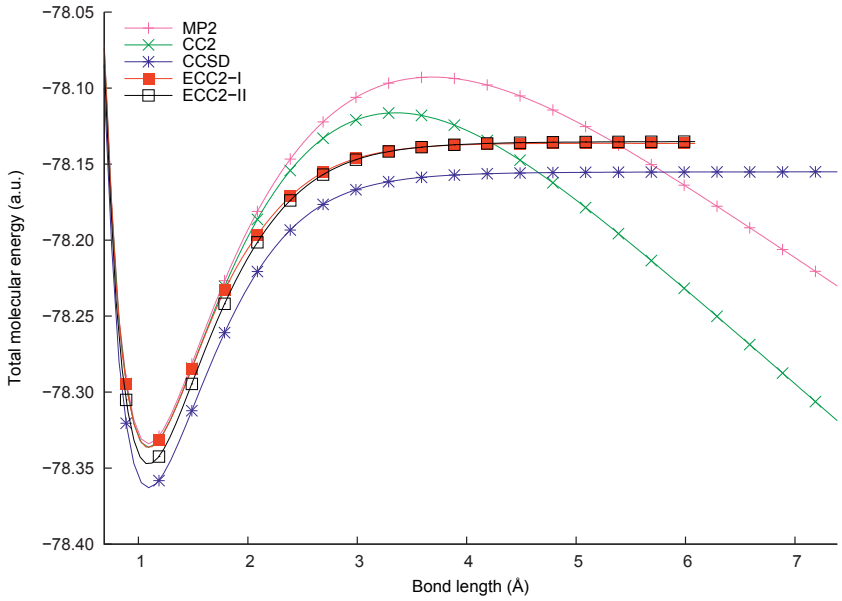


Figure A.4: Total energy curves for abstraction of one hydrogen atom from ethene.

CCSD while ECC2-I overestimates the equilibrium energy, but approaches CCSD at large distances. Curves for both ECC2a and ECC2b are presented in Figure A.3 and the commutator terms give a small improvement for both types of active spaces.

Figure A.4 shows the total energy of ethene during abstraction of one hydrogen atom using aug-cc-pVDZ. Initial geometries are given in Appendix A.4. For this reaction, an active space of only HOMO and LUMO gives an unphysical bump in the energy curve, as HOMO and LUMO changes nature along the abstraction coordinate. At short distances they are the π-orbitals of the ethene double bond while at longer distances they become the bonding and anti-bonding σ-orbitals of the carbon-hydrogen bond. This problem is solved by expanding the active space to include the two highest occupied and two lowest virtual orbitals. While the total ECC2 energy is very different from that of CCSD, the errors in the dissociation energy is about 4% for type I and 2% for type II active spaces. The difference in the total energy is due to correlation effects in the orbitals that are described at the CC2 level.

A.3.2 Static molecular properties

To assess the performance of ECC2 for electronic dipole moments and polarizabilities we investigate a number of molecules. We employed experimental geometries obtained from Ref. [172], except for 1-3-butadiene, where we used an optimized SCF geometry [173]. The ECC2 properties are calculated using finite differences and the difference between analytical and numerical results for CC2 are in the order of 0.01 a.u. for dipole moments and 0.001 a.u. for polarizabilities. In all ECC2 calculations we used type I active spaces and both models, a and b, are used to observe the effect of the commutator term in Eq. (A.10). We use unrelaxed Hartree-Fock orbitals to assess the performance of coupled cluster response theory.

In Table A.1, the results for hydrogen fluoride and ozone are presented. The CC2, CCSD and CC3 results are included for comparison. In these calculations we used an aug-cc-pVTZ basis. For HF, the difference between the CC2 and CCSD dipole moments is about 10%, and for the polarizability approximately 6%. Using the ECC2b model with an active space of four occupied and five virtual orbitals, the errors are reduced to 4% and 1%. In this model 9 orbitals enter CCSD space, compared to the 76 orbitals in the full CCSD model. Including more occupied orbitals in the active space is not expected to make much difference, as this would include the 1s orbital of fluorine. For ozone, large electron correlation effects are expected and the differences between CC2 and CCSD is indeed 21% and 8%. This is reduced to 3% and 1% for the best ECC2 model. This is very encouraging, as the error in polarizability is reduced by a factor of eight with respect to CCSD and only 16 out of a total of 138 orbitals are included in the active space. Ozone may be a special case, as we observe large changes when including only one virtual and one occupied orbital in the active space. For the other molecules studied, ECC2 give results that are intermediate between CC2 and CCSD and gradually approaches CCSD as the active space is expanded. For the dipole moment of ozone, CC2 underestimates the value and the smallest active space ECC2 model overestimates compared to CCSD.

Table A.2 contains the polarizabilities of ethyne and ethene, again using an aug-cc-pVTZ basis. For these systems, the error of CC2 with respect to CCSD is less, 3.3% and 4.4%. Using active spaces with 8 virtual and 5 and 6 occupied orbitals respectively, these errors are halved to 1.6% and 2.7%. Convergence is slow towards the CCSD values and we note that the active spaces did not include orbitals of all symmetry classes.

The polarizability is also calculated for the larger molecules benzene and 1-3-butadiene using aug-cc-pVDZ. These results are reported in Table A.3. The error is reduced from 4.8% to 2.7% for benzene and 7.5% to 2.6% for 1-3-butadiene. The initial convergence is rather slow for these systems, even worsening in some cases. However, when the seventh and eighth occupied and virtual orbitals are added, the error decrease rapidly for both systems. Inspection of the orbital energies reveals that, in both cases, orbitals with new symmetry classes are added to the active

Table A.1: Electronic dipole moment and polarizability in a.u. of hydrogen fluoride and ozone along the C_∞ and C_2 axis respectively. v is the number of virtual and o is the number of occupied orbitals.

		O3				HF			
Active space		Dipole moment		Polarizability		Dipole moment		Polarizability	
o	v	ECC2a	ECC2b	ECC2a	ECC2b	ECC2a	ECC2b	ECC2a	ECC2b
CC2		0.1474		14.71		0.1679		6.73	
1	1	0.2091	0.2037	14.17	13.99	0.1679	0.1679	6.72	6.73
2	2	0.2123	0.2044	14.14	13.99	0.1675	0.1685	6.72	6.74
3	3	0.2113	0.2047	14.08	13.96	0.1605	0.1590	6.54	6.50
4	4	0.1868	0.1866	14.06	13.97	0.1607	0.1592	6.49	6.45
4	5	-	-	-	-	0.1603	0.1580	6.46	6.42
5	5	0.1858	0.1837	13.99	13.92	-	-	-	-
6	6	0.1802	0.1821	13.84	13.82	-	-	-	-
7	7	0.1828	0.1838	13.74	13.76	-	-	-	-
8	8	0.1808	0.1824	13.74	13.76	-	-	-	-
CCSD		0.1876		13.63		0.1527		6.35	
CC3		0.1773		13.81		0.1559		6.33	

Table A.2: Polarizability in a.u. of ethyne and ethene along the C-C bonds. v is the number of virtual and o is the number of occupied orbitals.

Active space		C2H2		C2H4	
o	v	ECC2a	ECC2b	ECC2a	ECC2b
CC2		31.30		35.73	
1	1	31.30	31.30	35.72	35.72
2	2	31.29	31.30	35.72	35.73
3	3	31.22	31.22	35.70	35.71
4	4	31.17	31.11	35.69	35.70
5	5	31.14	31.08	35.34	35.29
5	6	31.00	31.06	-	-
5	7	31.07	30.99	-	-
5	8	30.91	30.79	-	-
6	6	-	-	35.27	35.19
6	7	-	-	35.26	35.17
6	8	-	-	35.24	35.15
CCSD		30.30		34.21	
CC3		30.42		34.03	

Table A.3: Polarizability in a.u. of benzene along a C_2 axis going through two hydrogen atoms and 1-3-butadiene along the C-C single bond. v is the number of virtual and o is the number of occupied orbitals.

Active space		C6H6		C4H6	
o	v	ECC2a	ECC2b	ECC2a	ECC2b
CC2		86.31		82.59	
1	1	86.32	86.32	82.56	82.56
2	2	86.35	86.31	82.55	82.55
3	3	86.33	86.30	82.53	82.55
4	4	86.30	86.30	82.51	82.54
5	5	86.29	86.29	82.46	82.52
6	6	86.26	86.25	82.44	82.49
7	7	85.77	85.44	80.38	80.17
8	8	85.18	84.94	80.04	80.84
9	9	85.07	84.82	79.97	79.73
10	10	84.92	84.62	79.34	78.85
11	11	-	-	79.34	78.85
CCSD		82.38		76.81	
CC3		81.52		76.29	

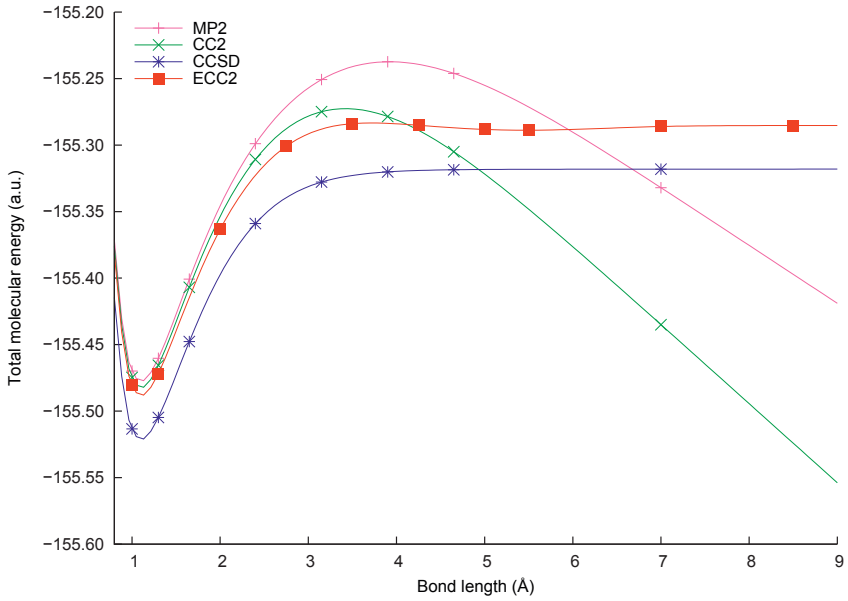


Figure A.5: Total energy curves for abstraction of one hydrogen atom from 1-3-butadiene using an active space of 11 occupied and 32 virtual orbitals.

space. Including orbitals from all symmetry classes may be a guiding criteria for choosing the active space. To test this, the polarizability of ethene is calculated using an active space with a total of ten orbitals, that include the lowest energy virtual orbitals from each symmetry. However, the calculated polarizability is 35.28 a.u., almost the same as the (5,5) model from Table A.2.

Comparing ECC2a and ECC2b, it is clear that including the commutator terms gives an improvement, both for total energy and static properties. As the computational scaling of the commutator term is relatively small, we recommend it to be included.

A.3.3 Subsystems

In this section we will demonstrate how the ECC2 model can be applied in connection with a subsystem approach. We use 1,3-butadiene as an example and consider the abstraction of a hydrogen atom along the CH bond. In Figure A.1 is an illustration of how the different approximation levels are selected. In the active CCSD

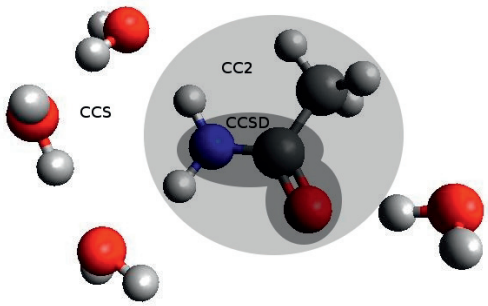


Figure A.6: Initial configuration and active spaces for ethanamide.

part we limit the Cholesky decomposition of the one-electron density matrix to generate only 5 occupied and 26 virtual orbitals. The remaining 10 occupied and 131 virtual orbitals are treated as the external orbital space. The entire external space is described at the CC2 level and thus there is no SCF part in this calculation. We employ an aug-cc-pVDZ basis and use a type I active space and initial geometries are given in Appendix A.4.

The results are given in Figure A.5 where we first observe the failure of CC2 and MP2 at large distances. The ECC2 curve displays the correct behavior in the limit, but the curve is shifted compared to CCSD. Evaluation of the dissociation energy reveals that the shift is almost constant. The full CCSD dissociation energy is 535.1 kJ/mol and for ECC2 it is 531.5 kJ/mol, a difference within chemical accuracy.

We may include more occupied and virtual orbitals in the active space by changing the threshold in the Cholesky decomposition to 2×10^{-4} and removing the limit on the number of active orbitals. With this procedure we get 11 occupied and 32 virtual orbitals and we obtain a dissociation energy of 536.1 kJ/mol. This is a clear improvement compared to the smaller active space.

We now consider a somewhat larger system that is illustrated in Figure A.6. Here, the polar parts of ethanamide are described at the full CCSD level while the unpolar methyl group is described at the CC2 level. The surrounding water molecules are described at the CCS level. This results in 12 occupied and 36 virtual orbitals being treated at the CCSD level, 4 occupied and 13 virtual at the CC2 level and 20 occupied and 92 virtual orbitals being treated at the CCS level. The initial geometry can be found in Appendix A.4

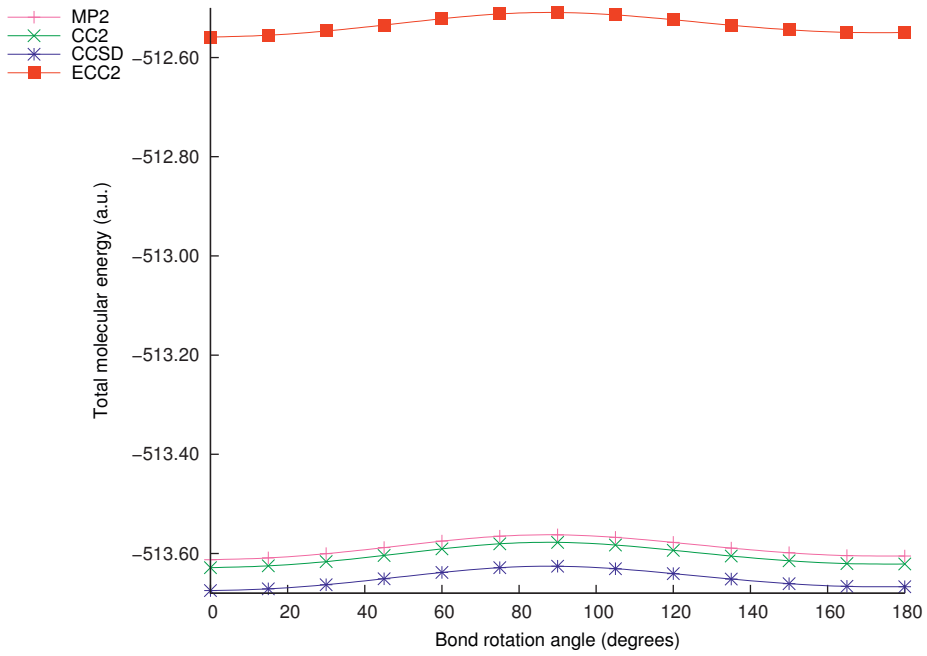


Figure A.7: Total energy curves for ethanamide as a function of the rotation angle around nitrogen-carbon bond in the presence of water.

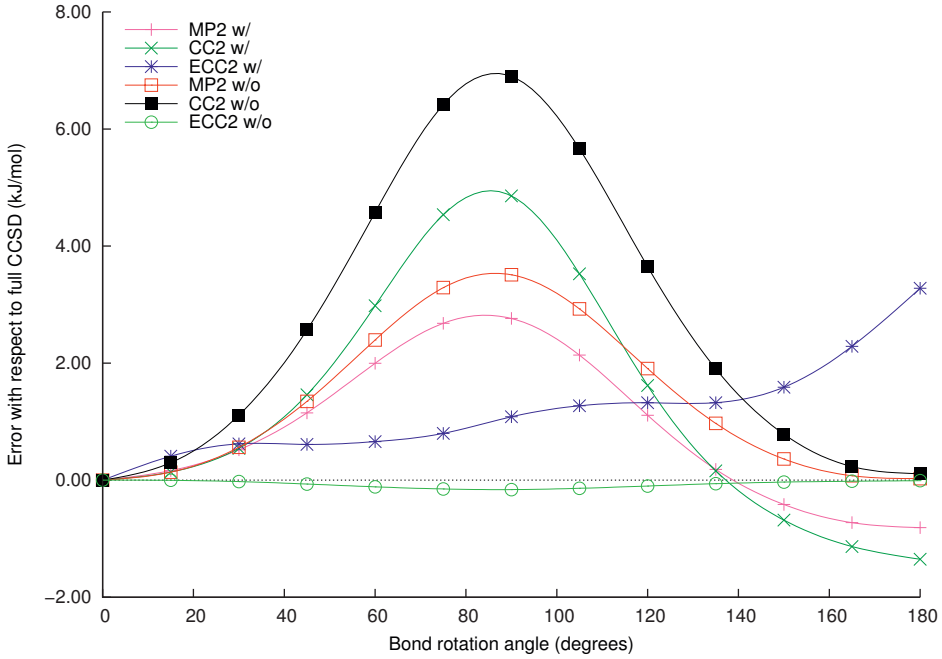


Figure A.8: Error in energy change as a function of the rotation angle of nitrogen-carbon bond with respect to full CCSD for ethanamide with and without water molecules present.

In Figure A.7 we have plotted the total energy as a function of the rotation angle of the amine group around the carbon-nitrogen bond in the presence of water molecules.

In Figure A.8 the error defined as $\Delta E_{\text{method}} - \Delta E_{\text{CCSD}}$ with $\Delta E = E(\phi) - E(0)$ where ϕ is the rotation angle from initial configuration. The error for ethanamide both with and without water molecules is presented. For both cases, the ECC2 error is within chemical accuracy and substantially more accurate than MP2 and CC2.

A.4 Conclusions

We have presented a proof of principle behind our proposed extension of the CC2 model that we call ECC2. We have shown that with little additional computational cost we may correct the divergence problems in CC2. Furthermore, static electric properties are improved in ECC2 and this provide an indication this will also be the case for dynamic properties.

We believe the real game changer is obtained with ECC2 in combination with

the subsystem partitioning of a molecular system. Electron correlation is a local property and it only seems logical to introduce this selectively in the parts of the molecule where it is needed to describe the chemistry of interest. The concept is best explained by Figures A.1 and A.6 and we have presented encouraging results for a very small system. However, this provides ample evidence that the method will work for much larger systems where it becomes really useful. We have previously documented the speedup of the subsystem approach based on Cholesky orbitals [88] and believe that for sufficiently large systems the method will be able to compete with the scaling of DFT as the correlation part of the calculation becomes constant with respect to size of the SCF and CCS parts.

In a forthcoming paper [174] we will develop the response theory associated with the ECC2 model and also provide more details about the full implementation of the method. The ideas behind ECC2 can of course be generalized to include other types of excitations such as triple excitations. This will be important in the future, especially when black box algorithms become computationally unfeasible and a more targeted application of electron correlation is needed.

Acknowledgements

We acknowledge support from the Spanish MINECO through project number CTQ2010-19738 and NOTUR for computer time.

Geometries

The initial geometries used in the calculations are presented in Tables A.4, A.5 and A.6.

To break a bond, the hydrogen atom were moved in the direction of the original equilibrium bond while the rest of the atoms were kept fixed. Similarly, under rotation, only the two hydrogen atoms bonded to the nitrogen were moved. The rotation axis was the carbon-nitrogen bond and the equilibrium dihedral angle were arbitrarily chosen as origin.

Table A.4: Initial geometry of ethene (Å).

atom	X	Y	Z
C	0.000000	0.000000	0.665647
C	0.000000	0.000000	−0.665647
H	0.000000	−0.923658	1.239711
H	0.000000	0.923658	−1.239711
H	0.000000	−0.923658	−1.239711
H	0.000000	0.923658	1.239711

Table A.5: Initial geometry of butadiene (Å).

atom	X	Y	Z
C	0.000000	0.000000	0.000000
C	−0.524786	3.619893	0.000000
C	−0.575773	1.160600	0.000000
C	0.050986	2.459292	0.000000
H	−1.729024	3.619893	0.000000
H	−0.394695	−0.938866	0.000000
H	−0.130090	4.558760	0.000000
H	1.138098	2.500263	0.000000
H	−1.662884	1.119629	0.000000
H	1.204238	0.000000	0.000000

Table A.6: Initial geometry of ethanamide and water (Å).

atom	X	Y	Z
C	0.000000	0.000000	0.000000
N	0.000000	0.000000	1.343933
H	0.000000	0.868414	1.889388
H	0.312840	−0.833192	1.840835
C	−0.376624	1.307903	−0.672187
O	0.253060	−1.003521	−0.691388
H	−1.363003	1.198198	−1.136253
H	0.339480	1.513937	−1.472421
H	−0.402903	2.149854	0.024644
H	0.754572	−0.846457	−2.313459
O	1.088567	−0.671047	−3.238879
H	1.886219	−1.216167	−3.310645
H	−0.004761	2.260462	3.901190
O	0.588784	2.125496	3.147711
H	1.289772	1.519432	3.497171
H	3.281178	0.406328	3.620628
O	2.411900	0.236049	4.012560
H	2.133887	−0.628957	3.630014
H	1.647040	−2.542963	1.997237
O	1.446729	−2.109176	2.878366
H	0.685755	−2.579973	3.247016

Paper B

Multi-level coupled cluster theory

J. Chem. Phys. **141**, 224105 (2014)

Multi-level coupled cluster theory

Rolf H. Myhre[1,2], Alfredo M. J. Sánchez de Merás[3] and Henrik Koch[1,2*]

[1] *Department of Chemistry, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway*

[2] *Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California 94305, USA*

[3] *Institute of Molecular Science, University of Valencia, ES-46071 Valencia, Spain*

Abstract

We present a general formalism where different levels of coupled cluster theory can be applied to different parts of the molecular system. The system is partitioned into subsystems by Cholesky decomposition of the one-electron Hartree-Fock density matrix. In this way the system can be divided across chemical bonds without discontinuities arising. The coupled cluster wave function is defined in terms of cluster operators for each part and these are determined from a set of coupled equations. The total wave function fulfills the Pauli-principle across all borders and levels of electron correlation. We develop the associated response theory for this multi-level coupled cluster theory and present proof of principle applications. The formalism is an essential tool in order to obtain size-intensive complexity in the calculation of local molecular properties.

B.1 Introduction

Wave function methods has reached a level of maturity where linear complexity algorithms have been developed for most of the standard electronic structure models including perturbation theory and coupled cluster (CC) theory. The locality of electron correlation is the fundamental basis for these methods [175] and can be exploited to reduce the scaling of electron correlation methods when orbitals are localized. For example, in the projected atomic orbital method the occupied orbital space is projected out from the atomic orbitals (AOs) giving a localized set of virtual orbitals [176–179]. By localizing the occupied orbitals, a domain in physical space can be defined and explicit correlation can be limited to electron pairs within the domain. This approach has been exended to CC theory and implemented as the local CCSD method [180–184]. The divide-expand-consolidate [185–187] (DEC) and frozen density embedding [84, 188, 189] (FDE) methods are based on fragmentation of the molecular system and calculating the electron correlation energy for each fragment. In DEC, the fragments are gradualy expanded until convergence while FDE considers the electron correlation for each fragment while keeping the electron density in the rest of the system frozen.

To calculate the total dynamic electron correlation energy for large systems the local pair natural orbital CCSD (LPNO-CCSD) [76, 190] and the orbital-specific-virtual local CCSD (OSV-LCCSD) [78, 191] methods are the most efficient currently available. In LPNO-CCSD, the occupied orbitals are localized and each electron pair is classified as strongly or weakly interacting using a Møller-Plesset second order perturbation (MP2) theory estimate. Only the strongly interacting pairs are treated explicitly with CC theory and the complexity is further limited by generating a truncated virtual set of orbitals for these electron pairs. Weakly interacting pairs are only included as MP2 estimates. The OSV-LCCSD method is similar, but in this case, a virtual basis is associated with each occupied orbital. This generally requires larger but fewer sets of virtual orbitals. Both these methods have recently been extended to perturbative triples [26] and are usually able to recover more than 90% of the triples correlation energy [77, 79] with total wall times reduced by several orders of magnitude for large systems compared to canonical calculations. Recently, CC2 excitation energies were implemented in a hybrid OSV-LPNO formulation [192]. Although this implementation is still a work in progress, reduced scaling led to improved timings compared to resolution of identity CC2 for a few hundred basis functions or more.

While the methods mentioned above make it possible to recover the correlation energy for large systems, their scaling is sensitive to the thresholds set in the program and how diffuse the basis functions are. An alternative approach to reduce the scaling is using approximate representations of the two-electron integral matrix. These methods have a long history in quantum chemistry and the most prominent are density fitting [61–65, 193, 194], Cholesky decomposition [66–68, 195, 196], pseudo spectral methods [197–205] and the more recent the tensor hypercontraction method [25, 206–208].

Many molecular properties are size intensive and only depend on a small part of the molecular system. In such cases, it is not necessary to compute the total correlation energy. Instead, one can use an accurate method on a small part of the system. In this paper we will explore such an alternative approach that can potentially lead to electron correlation methods with size-intensive complexity. In an earlier paper [209], we introduced the multi-level coupled cluster (MLCC) model denoted extended CC2 (ECC2) where an active space is treated with CCSD [21] and the rest using CC2 [23]. We were able to improve the description of abstraction processes were CC2 failed due to quasi degeneracies. In this communication, we discuss MLCC more generally with several levels of theory and introduce MLCC response theory. Using canonical orbitals, MLCC methods are reminiscent of the active space CC method by Olsen and Köhn [85, 210]. The major difference is that MLCC can employ several levels of theory, thus making it possible to gradually increase accuracy and the associated computational cost.

To fully exploit the locality of electron correlation, we combine MLCC with localized Cholesky orbitals [68, 88]. While any set of orthogonal orbitals can be employed, the Cholesky orbitals have the advantage that they can be computed without optimization. The orbitals are generated by decomposing the one-electron density matrix and this preserves the sparsity of the matrix. Carefully choosing the pivoting elements, the generated orbitals can be localized to a spatial part of the molecular system. Cholesky orbitals are less local than those of other methods [89–92], but they are highly convenient for partitioning the system into subsystems. The spatial division of the orbitals makes it possible to treat different parts of the molecular system with different levels of theory like in the ONIOM model [80, 81, 211].

This paper has been organized as follows. In the next section, we derive the working equations of MLCC with up to CC3 [24] level of theory and give a detailed account of the localization procedure. This is followed by a the derivation of the ECC2 linear response function. In Section B.3, we present proof of principle calculations and the last section contains our concluding remarks.

B.2 Theory

B.2.1 The multi-level coupled cluster model

In this section we present a detailed derivation of the MLCC model. The cluster operator X is divided into two separate contributions such that $X = T + S$. The coupled cluster wave function is written as

$$|\mathrm{CC}\rangle = \exp(X)|\mathrm{HF}\rangle = \exp(T+S)|\mathrm{HF}\rangle. \tag{B.1}$$

The projection manifold associated with each contribution to the cluster operator are denoted $\{\langle\mu^T|\}$ and $\{\langle\mu^S|\}$ and they are orthogonal to the Hartree-Fock (HF)

reference state and each other. Inserting these definitions into the Schrödinger equation and projecting onto the manifolds we obtain

$$\langle \mu^T | \exp(-T-S) H \exp(T+S) | \mathrm{HF} \rangle = 0 \tag{B.2a}$$
$$\langle \mu^S | \exp(-T-S) H \exp(T+S) | \mathrm{HF} \rangle = 0. \tag{B.2b}$$

The basic idea is to solve Eq. (B.2b) perturbatively for the amplitudes in the S operator while the amplitudes in T are determined from Eq. (B.2a) without any approximations. We divide the Hamiltonian into the Fock operator F and the fluctuation potential U as the perturbation

$$H = F + U. \tag{B.3}$$

The Fock operator is assumed to be diagonal such that $F = \sum_p \varepsilon_p a_p^\dagger a_p$ where p labels spin orbitals. We now expand the S in orders of the perturbation U

$$\begin{aligned} S =& S^{(0)} + S^{(1)} + S^{(2)} + \dots \\ =& \sum_{\mu \in \mu^S} s_\mu^{(0)} \tau_\mu^\dagger + \sum_{\mu \in \mu^S} s_\mu^{(1)} \tau_\mu^\dagger + \sum_{\mu \in \mu^S} s_\mu^{(2)} \tau_\mu^\dagger + \dots \end{aligned} \tag{B.4}$$

where $\tau_\mu^\dagger$ denote excitation operators. Inserting Eq. (B.4) into Eq. (B.2b) we obtain

$$\varepsilon_\mu s_\mu^{(0)} = 0 \tag{B.5}$$
$$\varepsilon_\mu s_\mu^{(1)} = -\langle \mu | \exp(-T) U \exp(T) | \mathrm{HF} \rangle \tag{B.6}$$
$$\varepsilon_\mu s_\mu^{(2)} = -\langle \mu | \exp(-T) [U, S^{(1)}] \exp(T) | \mathrm{HF} \rangle \tag{B.7}$$

for $\mu \in \mu^S$. In deriving Eqs. (B.5-B.7) we have used that $[F, \tau_\mu^\dagger] = \varepsilon_\mu \tau_\mu^\dagger$ where ε_μ are orbital energy differences. Equations (B.6) and (B.7) differs from standard CC perturbation theory [24] as the unperturbed part of the cluster operator is included. In this way, the T operator is treated to infinite order. Our model is now defined by the truncation of the unperturbed operator and the perturbation level of the perturbative part. For example, in ECC2 (section B.2.3), the T operator is truncated after double excitations and corresponds to CCSD while the S operator is included to first order in the perturbation U.

The expression for the total MLCC energy is the same as in standard CC theory and is given by

$$E = \langle \mathrm{HF} | H \exp(X) | \mathrm{HF} \rangle. \tag{B.8}$$

As in standard CC theory, there is no perturbative truncation in the energy expression, so the two operators are treated equally. Before introducing the ECC2 model, we now describe the scheme used to partition the molecular system.

B.2.2 Active spaces

We will use two different schemes to construct the active space. The first one is a straightforward selection based on Hartree-Fock orbital energies, typically around

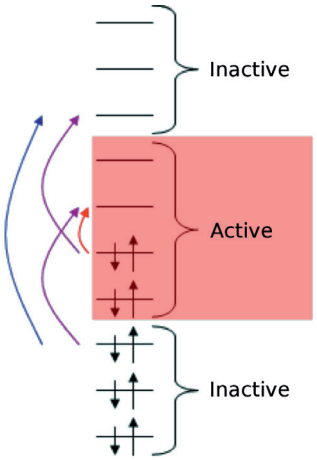


Figure B.1: Example of an active space based on orbital energies. Excitations from active to active orbitals (red) are included in T, while the semi external (purple), external (blue), and their combinations are included in S.

the HOMO-LUMO gap as illustrated in Figure B.1. This method was investigated in our previous paper [209] and will not be discussed further here.

The second method is based on localized Cholesky orbitals [171]. These orbitals are generated from Cholesky decomposing the AO Hartree-Fock density matrices

$$D_{\alpha\beta} = \sum_i C_{\alpha i} C_{\beta i} \tag{B.9}$$

$$D^V_{\alpha\beta} = \sum_a C_{\alpha a} C_{\beta a} \tag{B.10}$$

for occupied and virtual orbitals. The $\mathbf{C}$ matrix contains the molecular orbital (MO) coefficients where i and a label occupied and virtual MOs respectively and α and β label AOs. The sum of the occupied and virtual density matrices is the inverse of the AO overlap matrix $\mathbf{S}$,

$$\mathbf{D} + \mathbf{D}^V = \mathbf{S}^{-1}. \tag{B.11}$$

Aquilante et al. [171] have demonstrated that the Cholesky orbitals are orthogonal such that

$$D_{\alpha\beta} = \sum_i L_{\alpha i} L_{\beta i} \tag{B.12}$$

$$\sum_{\alpha\beta} L_{\alpha i} S_{\alpha\beta} L_{\beta j} = \delta_{ij} \tag{B.13}$$

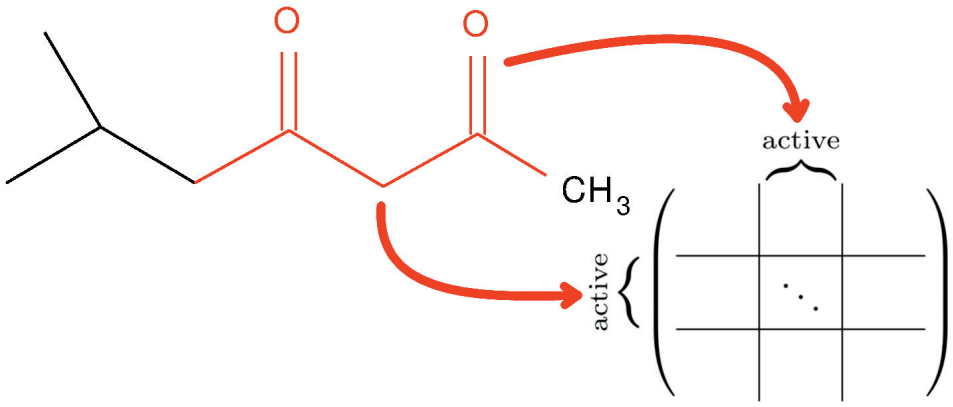


Figure B.2: Decompostition scheme for 6-methyl-2,4-heptadione. The AO one-electron Hartree-Fock density matrix is decomposed for diagonals corresponding to active atoms indicated in red.

where $L_{\alpha i}$ are the Cholesky orbital coefficients for the occupied part. Similar expressions hold for the virtual part and Eq. (B.11) shows that the virtual density matrix can be calculated without explicit knowledge of the virtual MO coefficients.

When partitioning the molecular system into two parts, we classify the atoms as active or inactive. In this way we obtain a set of active AOs that are centered on the active atoms and similarly for the inactive atoms. As illustrated in Figure B.2, the active Cholesky orbitals are determined by restricting the decomposition to the diagonal elements of $\mathbf{D}$ and $\mathbf{D}^V$ corresponding to active AOs. However, the active part is not fully decomposed, but only diagonal elements larger than a specific threshold are selected. We typically use the thresholds 0.2 and 0.3 for the occupied and virtual parts, respectively. This has been shown to be sufficient to reproduce the Lewis structure of the atoms [88]. After the active orbitals have been generated, the inactive parts are decomposed with thresholds that are one hundreth of the corresponding active thresholds. The described procedure generates two sets of orbitals that can be classified as either active or inactive and the total number is equal to the number of canonical orbitals.

Changing the molecular geometry will lead to changes in the diagonal elements of the density matrix. This may result in different elements being chosen in the decomposition which in turn can lead to discontinuities in the energy. However, if for each geometry we only decompose among the same diagonal elements denoted I and J in the following, the decomposition will be continuous. This restricted decomposition corresponds to decomposing the matrix $\hat{D}(R)$ where R denote the atomic coordinates

$$\hat{D}_{\alpha\beta}(R) = \sum_{IJ} D_{\alpha I}(R)\tilde{D}^{-1}_{IJ}(R)D_{J\beta}(R) \tag{B.14}$$

where

$$\tilde{D}_{IJ}(R) = D_{IJ}(R) \tag{B.15}$$

$$\hat{D}_{\alpha\beta}(R) = \sum_{I} L_{\alpha I}(R) L_{\beta I}(R) \tag{B.16}$$

since the Cholesky decomposition is equivalent to an inner projection [212, 213]. In Eqs.(B.14) and (B.15), $\tilde{\mathbf{D}}$ is the submatrix of $\mathbf{D}$ corresponding to the decomposed diagonal elements. The expression is numerically stable as long as $\tilde{\mathbf{D}}$ is positive definite. Furthermore, the expression can be used to calculate analytical molecular derivatives in a straightforward fashion [214].

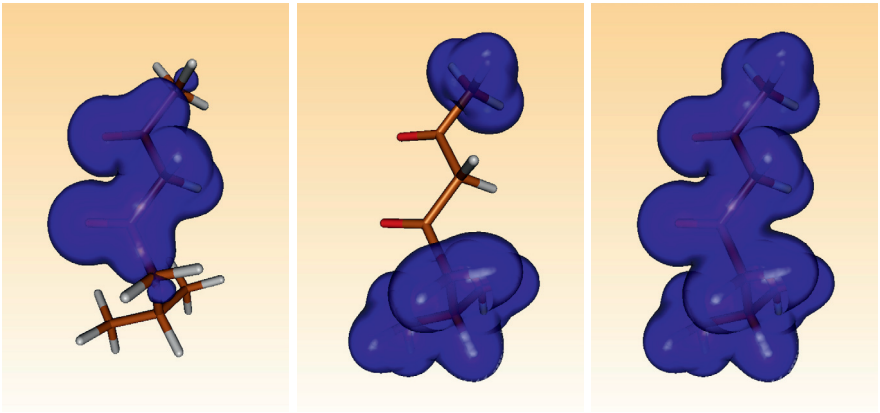
The above procedure can easily be generalized to more than two spaces by assigning the atoms in the system to more than two sets. The decomposition procedure then starts with the most acurate active set and decomposes $\mathbf{D}$ and $\mathbf{D}^{\mathbf{V}}$ to the thresholds. After the first orbital set is computed, the next set is calculated by decomposing the relevant diagonals. This is repeated for all sets except the last one, which is decomposed to one hundreth of the thresholds. In this way, any number of localized orbital spaces can be generated. To simplify the working equations, the Fock matrix is block diagonalized in each space. The resulting Fock matrix will no longer be diagonal, but there is no mixing between occupied and virtual parts.

After generating the orbital spaces, the projection manifolds can be assigned. The first projection manifold is restricted to the excitations internal to the first space as shown in Figure B.1 and only excitations from active to active orbitals are included. In the case of two spaces, all other excitations are included in the inactive manifold. If there are more orbital spaces, the second manifold includes excitations internal to the second space and those between the first and the second space. Each new space will then define a manifold that includes excitations internal in the space and excitations between the space and the preceeding ones.

In Figure B.3, the active and inactive Hartree-Fock one-electron densities are shown for 6-methyl-2,4-heptadione using cc-pVDZ and the scheme in Figure B.2. The electron density surface was generated with Molden [215]. Since the active orbitals are generated first, the active electron density extends into the inactive parts of the molecule, but normally not vice versa. This ensures that the electron density close to the active atoms is described by a high level theory. In passing, we note that the selection of active spaces can also be done using a density functional theory electron density.

B.2.3 The ECC2 model

Extended CC2 is a two-level CC model where the unperturbative part T is treated as CCSD and S as CC2. Single excitations are calculated the same way in both models and the singles part of the cluster operator is simply refered to as X_1. The



(a) Active (b) Inactive (c) Total

Figure B.3: Active, inactive and total electron densities for 6-methyl-2,4-heptadione. The basis set is cc-pVDZ and the surface corresponds to $\rho = 0.02$.

doubles part is split into two parts $X_2 = T_2 + S_2$. Since the T operator is evaluated to infinite order the working equations are the same as for standard CCSD

$$\langle \mu_1 | \hat{H} + [\hat{H}, X_2] | \mathrm{HF} \rangle = 0 \tag{B.17}$$

$$\langle \mu_2^T | \hat{H} + [\hat{H}, X_2] + \frac{1}{2}[[\hat{H}, X_2], X_2] | \mathrm{HF} \rangle = 0 \tag{B.18}$$

where we have used the X_1-transformed Hamiltonian.

$$\hat{H} = \exp(-X_1) H \exp(X_1). \tag{B.19}$$

When solving the S_2 equations, we first note that all excitations in the S manifold involve at least one inactive orbital and consequently

$$\langle \mu_2^S | [[\hat{H}, T_2], T_2] | \mathrm{HF} \rangle = 0. \tag{B.20}$$

Inserting the Hamiltonian from Eq. (B.3), we obtain

$$\begin{aligned} &\langle \mu_2^S | \, \hat{H} + [\hat{H}, T_2] + [F, S_2] + [\hat{U}, S_2] \, | \mathrm{HF} \rangle \\ &+ \langle \mu_2^S | \, [[\hat{U}, T_2], S_2] + \frac{1}{2}[[\hat{U}, S_2], S_2] \, | \mathrm{HF} \rangle = 0. \end{aligned} \tag{B.21}$$

As the last three commutators are at least second order in U, they are discarded. The $[\hat{H}, T_2]$ term in the first line would normally be considered first and second order. However, the Fock operator is not necessarily diagonal and the commutator with F must be included in order to satisfy the zero order condition. Furthermore, T is treated to infinite order, so the full $\hat{H}$ commutator is included. The working equations for the S_2 amplitudes then become

$$\langle \mu_2^S | [F, S_2] + \hat{H} + [\hat{H}, T_2] | \mathrm{HF} \rangle = 0. \tag{B.22}$$

To summarize, the ECC2 model is defined by the working equations (B.17), (B.18) and (B.22). We note that the ECC2 model is size extensive [169]. If two subsystems are treated at different levels of theory, CCSD for system A and CC2 for system B, the total energy is the sum of the model energies ($E = E_A^{CCSD} + E_B^{CC2}$). Thus for non-interacting subsystems the ECC2 equations reduce to the standard canonical equations for each subsystem.

The computational reductions compared to full CCSD is achieved through the reduced size of the projection manifold in Eq. (B.18). In full CCSD, equation (B.18) contains a term that scales as V^4O^2, where V is the number of virtual orbitals and O is the number of occupied orbitals. This particular term will scale as kV^2 in ECC2 where the k prefactor only depends on the number of active orbitals and scales as $V_A^2O_A^2$ where V_A and O_A denote the number of active orbitals. As the active space is expected to contain only a fraction of the total number of orbitals, this leads to a large reduction in computational cost. If the system size is increased while keeping the size of the active space fixed, the complexity is that of CC2. We have previously documented the performance of the ECC2 model for static properties [209]. In passing we note that the equation for the S_2 amplitudes resembles the equation for explicitly correlated amplitudes in the CCSD(F12) model [216].

B.2.4 Other MLCC models

For larger systems we may introduce an additional orbital space that is described by CCS. To achieve this, an additional Cholesky orbital space is calculated and all double excitation amplitudes involving orbitals from this space are discarded while the single excitation equations are the same. We may also define a static Hartree-Fock space beyond the CCS part. In this case, all excitations involving HF orbitals are not included.

To achieve higher accuracy, CCSD(T) [26] or CC3 [24] are natural extensions to the ECC2 model. In this case, we may assign up to five different orbital spaces that we denote P,Q,R,S and T, where P refers to static HF space. The splitting of the cluster operator X is summarized in the following equations.

$$X = X_1 + X_2 + X_3 \tag{B.23}$$
$$X_1 = Q_1 + R_1 + S_1 + T_1 \tag{B.24}$$
$$X_2 = R_2 + S_2 + T_2 \tag{B.25}$$
$$X_3 = T_3 \tag{B.26}$$

For CCSD(T) [24, 26] the amplitude equations are the same as for CCSD and a triples correction is calculated and added to the energy in the end. The correspond-

ing amplitude equations are summarized below

$$\begin{aligned} 0 &= \langle \mu_1 | \hat{H} + [\hat{H}, X_2] | \mathrm{HF} \rangle & \text{(B.27)} \\ 0 &= \langle \mu_2^R | \hat{H} + [F, R_2] + [\hat{H}, S_2 + T_2] | \mathrm{HF} \rangle & \text{(B.28)} \\ 0 &= \langle \mu_2^S | \hat{H} + [\hat{H}, X_2] + [[\hat{H}, X_2], X_2] | \mathrm{HF} \rangle & \text{(B.29)} \\ 0 &= \langle \mu_2^T | \hat{H} + [\hat{H}, X_2] + [[\hat{H}, X_2], X_2] | \mathrm{HF} \rangle. & \text{(B.30)} \end{aligned}$$

In the current implementation, the triples correction is only calculated from the T_2 amplitudes. When the orbital basis is non-cannonical the non-diagonal elements will appear when commuting the Fock operator and the triples cluster operator. However, this is not a complication as these terms will include excitations out of the active space and become zero when projected against the active manifold. The equation for the triples amplitudes becomes

$$\varepsilon_{\mu_3^T} t_{\mu_3^T} = -\langle \mu_3^T | [U, T_2] | \mathrm{HF} \rangle \tag{B.31}$$

giving the energy correction

$$E_{corr}^{CCSD(T)} = \sum_{\mu \in \mu_1^T, \mu_2^T} t_\mu \langle \mu | [U, T_3] | \mathrm{HF} \rangle . \tag{B.32}$$

As we will show in Section B.3, this approach can give large reductions in computational cost and still maintain the accuracy of the full model. The systems studied so far have been fairly limited and one may consider including X_2 in Eqs.(B.31) and (B.32), however our results indicate this is not necessary.

The CC3 model [24] includes the triple excitation in an iterative and perturbative fashion. Standard implementations scale as iterative N^7. Although this is a very steep scaling many systems require the triple excitations to achieve an accurate description. Implementing CC3 in an MLCC framework could make it possible to avoid the prohibitive computational cost. The equations for the amplitudes are given below

$$\begin{aligned} 0 &= \langle \mu_1 | \hat{H} + [\hat{H}, X_2] + [\hat{H}, T_3] | \mathrm{HF} \rangle & \text{(B.33)} \\ 0 &= \langle \mu_2^R | \hat{H} + [F, R_2] + [\hat{H}, S_2 + T_2] | \mathrm{HF} \rangle & \text{(B.34)} \\ 0 &= \langle \mu_2^S | \hat{H} + [\hat{H}, X_2 + T_3] + [[\hat{H}, X_2], X_2] | \mathrm{HF} \rangle & \text{(B.35)} \\ 0 &= \langle \mu_2^T | \hat{H} + [\hat{H}, X_2 + T_3] + [[\hat{H}, X_2], X_2] | \mathrm{HF} \rangle & \text{(B.36)} \\ 0 &= \langle \mu_3^T | [F, T_3] + [\hat{H}, X_2] | \mathrm{HF} \rangle & \text{(B.37)} \end{aligned}$$

Note that the T_3 commutator term in Eq. (B.33) is only non-zero when projecting against $\langle \mu_1^T |$ and that the doubles in the CCSD and CC3 spaces are both treated to infinite order in the perturbation thus, Eqs. (B.35) and (B.36) contain similar terms. The equations above have been derived using perturbation theory. However, these may not be the optimal working equations. In Eq. (B.37), a commutator

between the Hamiltonian and the full X_2 operator is included. This term will scale as $V_2 V_T^3 O_T^3$ where V_2 is the number of virtual orbitals in R, S and T. Depending on the size of the spaces, it is likely that the commutator term with R_2 is small. This may be true for the S_2 commutator term as well. Numerical testing will be necessary to decide whether these terms are needed or not. A similar argument can be made for the T_3 commutator term in Eq. (B.35), but this term will only scale linearly with the size of the CCSD space.

B.2.5 Response theory

We will now derive the linear response function for the ECC2 model using CC response theory [31, 32] and the quasienergy method [38, 39, 217]. This method assumes a periodic time-dependent perturbation, but is sufficient for our purposes. For application of the quasienergy method to CC2 and CC3 models, we refer the reader to Refs. [23, 87].

The starting point is the introduction of a periodic time dependent perturbation, V^t, in the Hamiltonian

$$H = F + U + V^t = H_0 + V^t. \tag{B.38}$$

We assume that V^t is a one-electron operator that can be expanded as finite a sum over frequencies

$$V^t = \sum_{j=-N}^{N} \sum_{A} A \epsilon_A(\omega_j) \exp(-i\omega_j t). \tag{B.39}$$

To ensure hermiticity, we require that $A = A^\dagger$ is a real frequency independent operator with $\omega_{-j} = -\omega_j$ and $(\epsilon_A(\omega_j))^* = \epsilon_A(-\omega_j)$.

Using the approach described in Ref. [87], we obtain the ECC2 quasienergy Lagrangian

$$\begin{aligned} L(t) = {} & \langle \mathrm{HF}| H \exp(X) |\mathrm{HF}\rangle \\ & + \sum_{\mu_1} \bar{t}_{\mu_1} \left(\langle \mu_1 | \hat{H} + [\hat{H}, X_2] |\mathrm{HF}\rangle - i \frac{\mathrm{d}t_{\mu_1}}{\mathrm{d}t} \right) \\ & + \sum_{\mu_2^T} \bar{t}_{\mu_2^T} \left(\langle \mu_2^T | \hat{H} + [\hat{H}, X_2] + \frac{1}{2}[[\hat{H}, X_2], X_2] |\mathrm{HF}\rangle - i \frac{\mathrm{d}t_{\mu_2^T}}{\mathrm{d}t} \right) \\ & + \sum_{\mu_2^S} \bar{t}_{\mu_2^S} \left(\langle \mu_2^S | [F + \hat{V}, S_2] + \hat{H} + [\hat{H}, T_2] |\mathrm{HF}\rangle - i \frac{\mathrm{d}t_{\mu_2^S}}{\mathrm{d}t} \right) \end{aligned} \tag{B.40}$$

Due to the periodicity of V^t the amplitudes can be expanded in the frequencies

$$t_\mu = t_\mu^{(0)} + t_\mu^{(1)} + t_\mu^{(2)} + ... \tag{B.41}$$

with

$$\begin{aligned} t_\mu^{(1)} &= \sum_{j=-N}^{N} t_\mu^{(1)}(\omega_j)\exp(-i\omega_j t) \\ &= \sum_{j=-N}^{N}\sum_{A} t_\mu^{A}(\omega_j)\epsilon_A(\omega_j)\exp(-i\omega_j t). \end{aligned} \tag{B.42}$$

The Lagrange multipliers can be similarly expanded. In this paper, we will only derive the linear response function and will not need higher than first order amplitudes due to the $2n+1$ and $2n+2$rules [169]. From the expansion of the amplitudes and multipliers, each term in the Lagrangian can also be assigned an order in the perturbation

$$L = L^{(0)} + L^{(1)} + L^{(2)} + ... \tag{B.43}$$

A stationary condition is obtained by taking the time average of the Lagrangian

$$\{L\}_{\mathcal{T}} = \int_0^{\mathcal{T}} L(t)\mathrm{d}t \tag{B.44}$$

where $\mathcal{T}$ is the period of the perturbation. The amplitudes and multipliers can then be determined using that

$$\frac{\partial}{\partial t_\mu^{(m)}}\{L^{(n)}\}_{\mathcal{T}} = \frac{\partial}{\partial \bar{t}_\mu^{(m)}}\{L^{(n)}\}_{\mathcal{T}} = 0. \tag{B.45}$$

for all $m \leq n$.

The zero order cluster amplitudes are the time independent amplitudes from Eqs. (B.17), (B.18) and (B.22). To find the corresponding multipliers, we use that

$$\frac{\partial\{L^{(0)}\}_{\mathcal{T}}}{\partial t_\mu^{(0)}} = 0 \tag{B.46}$$

resulting in

$$\bar{\mathbf{t}}^{(0)}\mathbf{A} = \boldsymbol{\eta}^{(0)}. \tag{B.47}$$

Where $\boldsymbol{\eta}^{(0)}$ is the same as in standard CCSD

$$\eta_\nu^{(0)} = -\langle \mathrm{HF}|[\hat{H}_0, \tau_\nu]|\mathrm{HF}\rangle \tag{B.48}$$

and $\mathbf{A}$ is the Jacobian matrix given below. Similarly, the first order stationary condition with respect to the multipliers, $\partial\{L^{(1)}\}_{\mathcal{T}}/\partial\bar{t}_\mu^{(1)} = 0$, is used to determine the first order amplitudes.

$$(\omega\mathbf{1} - \mathbf{A})\,\mathbf{t}^{(A)}(\omega) = \boldsymbol{\xi}^{(A)} \tag{B.49}$$

The right hand side is given by

$$\boldsymbol{\xi}^{(A)} = \begin{pmatrix} \langle\mu_1|\hat{A} + [\hat{A}, X_2^{(0)}]|\mathrm{HF}\rangle \\ \langle\mu_2^T|[\hat{A}, X_2^{(0)}]|\mathrm{HF}\rangle \\ \langle\mu_2^S|[\hat{A}, X_2^{(0)}]|\mathrm{HF}\rangle \end{pmatrix} \tag{B.50}$$

and is the same as in CCSD.

The singles part of the Jacobian matrix, **A**, is the same in CC2 and CCSD and the expressions are the same in ECC2. For the doubles, the matrix has two parts corresponding to $T_2$and S_2. The T_2 part is the same as for CCSD, while the S_2 part is similar to CC2, but differ due to the inclusion of the T_2 commutator term in Eq. (B.22). We have the following expression for the full ECC2 Jacobian

$$\boldsymbol{A} = \begin{pmatrix} \langle \mu_1 | [\hat{H}_0 + [\hat{H}_0, X_2^{(0)}], \tau_{\nu_1}] | \mathrm{HF} \rangle & \langle \mu_1 | [\hat{H}_0, \tau_{\nu_2^T}] | \mathrm{HF} \rangle & \langle \mu_1 | [\hat{H}_0, \tau_{\nu_2^S}] | \mathrm{HF} \rangle \\ \langle \mu_2^T | [\hat{H}_0 + [\hat{H}_0, X_2^{(0)}], \tau_{\nu_1}] | \mathrm{HF} \rangle & \langle \mu_2^T | [\hat{H}_0 + [\hat{H}_0, X_2^{(0)}], \tau_{\nu_2^T}] | \mathrm{HF} \rangle & \langle \mu_2^T | [\hat{H}_0 + [\hat{H}_0, X_2^{(0)}], \tau_{\nu_2^S}] | \mathrm{HF} \rangle \\ \langle \mu_2^S | [\hat{H}_0 + [\hat{H}_0, T_2^{(0)}], \tau_{\nu_1}] | \mathrm{HF} \rangle & \langle \mu_2^S | [\hat{H}_0, \tau_{\nu_2^T}] | \mathrm{HF} \rangle & \langle \mu_2^S | [F, \tau_{\nu_2^S}] | \mathrm{HF} \rangle \end{pmatrix} \tag{B.51}$$

After determining the zero order multipliers and first order amplitudes, the linear response function can be evaluated. The linear response function is given by [38]

$$\langle\langle A, B\rangle\rangle_{\omega_i} = \left.\frac{\partial\{L\}_{\mathcal{T}}}{\partial\epsilon_A(-\omega_i)\partial\epsilon_B(\omega_i)}\right|_{\epsilon=0}. \tag{B.52}$$

and inserting Eq. (B.40) into Eq. (B.52), we obtain the following expression

$$
\begin{aligned}
\langle\langle A,B\rangle\rangle_{\omega_i} = P(A(-\omega_i),B(\omega_i))\Big\{ & \langle \mathrm{HF}|\,[\hat{A}(-\omega_i),X_1^{(B)}(\omega_i)] + \frac{1}{2}[[\hat{H}_0,X_1^{(A)}(-\omega_i)],X_1^{(B)}(\omega_i)]\,|\mathrm{HF}\rangle \\
& + \sum_{\mu_1} \bar{t}^{(0)}_{\mu_1} \langle \mu_1|\,[\hat{A}(-\omega_i),X^{(B)}(\omega_i)] + \frac{1}{2}[[\hat{H}_0,X^{(A)}(-\omega_i)],X^{(B)}(\omega_i)]\,|\mathrm{HF}\rangle \\
& + \sum_{\mu_2^S} \bar{t}^{(0)}_{\mu_2^S} \langle \mu_2^S|\,[\hat{A}(-\omega_i),X_2^{(B)}(\omega_i)] + \frac{1}{2}[[\hat{H}_0,X_1^{(A)}(-\omega_i)],X_1^{(B)}(\omega_i)] + [[\hat{A}(-\omega_i),X_1^{(B)}(\omega_i)],X_2^{(0)}] \\
& + [[\hat{H}_0,X_1^{(A)}(-\omega_i)],T_2^{(B)}(\omega_i)] + \frac{1}{2}[[[\hat{H}_0,X_1^{(A)}(-\omega_i)],X_1^{(B)}(\omega_i)],T_2^{(0)}]\,|\mathrm{HF}\rangle \\
& + \sum_{\mu_2^T} \bar{t}^{(0)}_{\mu_2^T} \langle \mu_2^T|\,[\hat{A}(-\omega_i),X_2^{(B)}(\omega_i)] + \frac{1}{2}[[\hat{H}_0,X_1^{(A)}(-\omega_i)],X_1^{(B)}(\omega_i)] + [[\hat{A}(-\omega_i),X_1^{(B)}(\omega_i)],X_2^{(0)}] \\
& + [[\hat{H}_0,X_1^{(A)}(-\omega_i)],X_2^{(B)}(\omega_i)] + \frac{1}{2}[[[\hat{H}_0,X_1^{(A)}(-\omega_i)],X_1^{(B)}(\omega_i)],X_2^{(0)}] \\
& + \frac{1}{2}[[\hat{H}_0,X_2^{(A)}(-\omega_i)],X_2^{(B)}(\omega_i)]\,|\mathrm{HF}\rangle \Big\}
\end{aligned}
\tag{B.53}
$$

where

$$X^{(A)}(\omega_i) = \partial X^{(1)}(\omega_i)/\partial \epsilon_A(\omega_i) \tag{B.54}$$

is the derivative of the frequency dependent first order cluster operator obtained from Eq. (B.42) and

$$P(x,y)f(x,y) = f(x,y) + f(y,x). \tag{B.55}$$

As for the Jacobian matrix, the X_1 and T_2 parts of the response function are the same as for full CCSD , but the S_2 part differs from CC2. In particular, the terms on line four in Eq. (B.53) would not be included. These additional terms only include active double excitations and have lower scaling than CC2.

A direct implemention of the eigenvalue problem for the ECC2 Jacobian will require storing all doubles amplitudes on disk. As the number of amplitudes scales as O^2V^2, this may lead to extensive storage requirements for large systems. However, the S_2 part can be partitioned into the T part using a similar technique to the one used by Hättig and Weigend in RI-CC2 [218]. The number of double amplitudes can also be reduced by employing a CCS space. In a practical application we expect the CC2 part to be relatively small. Finally, increased use of solid state disks will make storage of amplitudes less unattractive due to the high I/O performance.

B.3 Numerical examples

Currently, ECC2 has been implemented as pilot code in the Dalton [99] software package. While the current implementation is too slow for production calculations, it is useful for proof of principle demonstrations and the examples in this section will illustrate the applicability of MLCC.

B.3.1 ECC2(T) calculations

To illustrate the efficiency of ECC2(T), we have chosen the model system depicted in Figure B.4. The complex consists of a neon dimer and three water molecules and our objective is to calculate the effect of the water molecules on the van der Waals interaction between the neon atoms. While the system geometry is highly artificial, it demonstrates the combination of van der Waals and dipole interactions that requires triples corrections. We have performed calculations with the standard CC models CC2, CCSD and CCSD(T), the second order perturbation model MP2 and two MLCC models. In both the MLCC models, the water molecules are treated with CCS, CC2 and CCSD with CCSD closest to neon and CCS furthest away. The neon atoms are treated with CCSD or CCSD(T).

We employ aug-cc-pVDZ basis and counterpoise correct when separating the system into two fragments. Midbond functions are placed between the neon atoms as

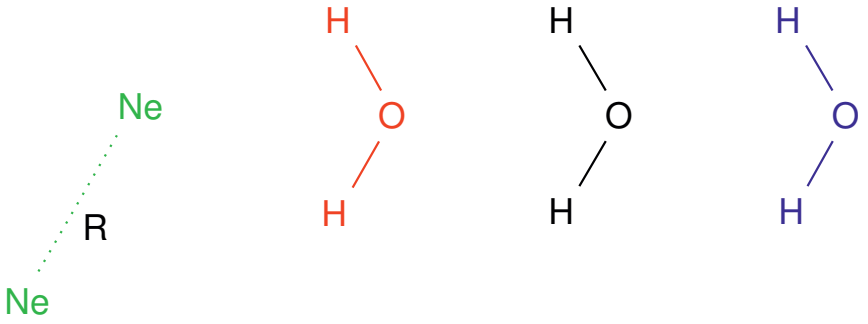


Figure B.4: Neon water complex. Green indicates CCSD(T), red CCSD, black CC2, and blue CCS. R is the distance between the neon atoms.

described in Refs. [219, 220] and the midbond center is considered active. One fragment contains the water molecules and the closest neon atom while the other only consists of the remaining neon atom. Ghost orbitals from the counterpoise correction are treated with the same level of theory as the corresponding atoms. To obtain consistent orbital spaces, the number of orbitals in each orbital space is kept the same. For example, there are in total 84 orbitals in the basis set on the neon atoms and the midbond functions, this implies that all three calculations have 84 orbitals in the CCSD(T) space.

The results of the calculations are presented in Figure B.5. Interaction energy increases with increasing accuracy of the electron correlation and the equilibrium distance contracts. The calculated interaction energies around the equilibrium distance are presented in Table B.1 and the differences between the MLCC models and the corresponding standard models are less than 0.15 cm^{-1}. Figure B.6 shows the interaction energy calculated using CCSD(T) with different numbers of water molecules. As expected, the water molecule closest to the neon dimer influence the interaction energy the most and the effect decreases with distance. Comparing Figures B.5 and B.6, the water molecules change the interaction energy by about 20%. This effect is well described by ECC2.

The computationally most expensive term in CCSD(T) [26] scales as V^4O^3. For the ECC2(T) model, the CCSD(T) space contained 10 occupied and 74 virtual orbitals compared to 25 occupied and 182 virtual orbitals in the full system. The theoretical reduction in computational time for this term is 572 fold. In practice, the CCSD(T) calculation for the whole system took 26823 seconds and only 159 seconds with ECC2(T), corresponding to a 168 fold reduction. For comparison, a CCSD iteration for the full system took 234 seconds.

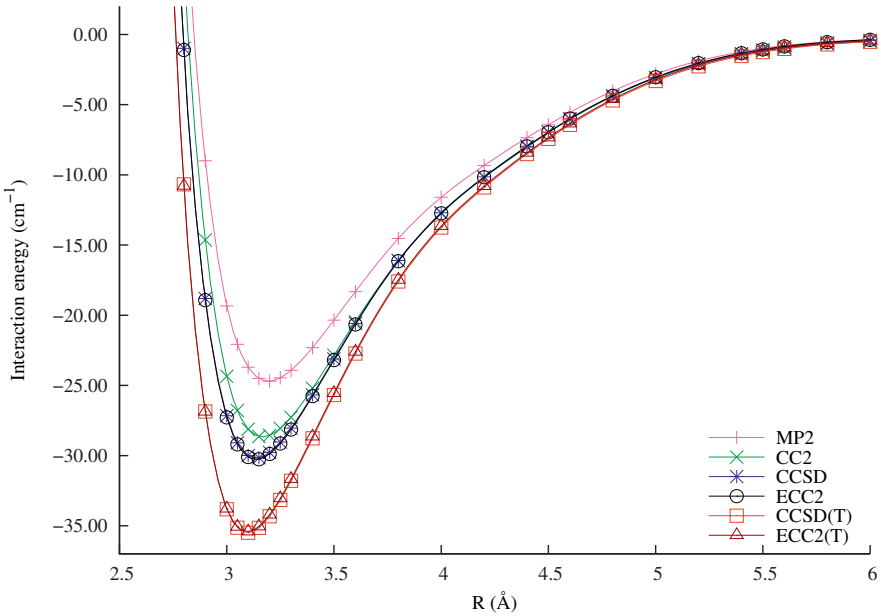


Figure B.5: Interaction energy for neon dimer with three water molecules calculated using counterpoise correction.

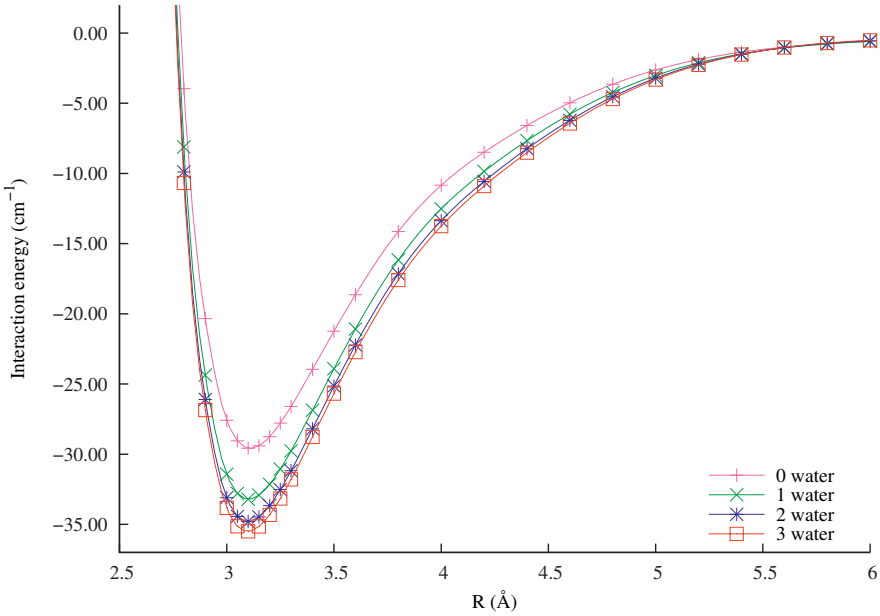


Figure B.6: Interaction energy calculated using CCSD(T) for neon dimer with zero, one, two and three water molecules.

Table B.1: Energy in cm^{-1} calculated around equilibrium for Neon water complex.

r(Å)	CCSD	ECC2	CCSD(T)	ECC2(T)
3.00	-27.16	-27.28	-33.83	-33.79
3.05	-29.08	-29.20	-35.14	-35.08
3.10	-29.99	-30.10	-35.50	-35.41
3.15	-30.15	-30.26	-35.15	-35.06
3.20	-29.78	-29.88	-34.32	-34.22

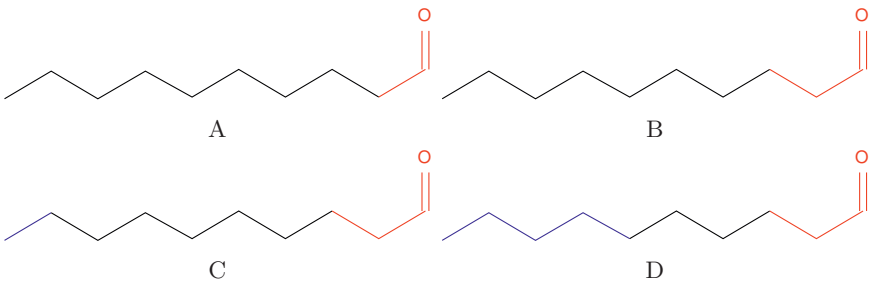


Figure B.7: Subsystems of decanal. Red atoms are treated with CCSD, black with CC2, and blue with CCS. Bonds are included in the most active space, so model C has two CCSD carbons and two CCS carbons. Hydrogen is treated the same as their parent carbon.

Table B.2: Excitation energies of decanal in eV using cc-pVDZ.

	CCSD	CC2	A	B	C	D
1	4.08	4.17	4.12	4.08	4.11	4.11
2	8.58	8.34	8.55	8.61	8.60	8.59

B.3.2 Excitation energies

Electronic excitations are often highly localized and therefore well suited for MLCC. Decanal is used as a model system and four different cases are presented in Figure B.7. We have used a modified geometry from Pubchem [221] (See supplementary material [222] for the geometry.). For case A and B we only employ CCSD and CC2 spaces, while for C and D we also introduce a CCS space. The results are presented in Table B.2. Note that all excitations reported in this section are singlet excitations.

For the first excitation, CC2 performs relatively well compared to CCSD with an error of 0.09 eV, but for the second excitation the difference is 0.24 eV. With an active space of only the oxygen atom, the closest carbon and the attached hydrogen, the error is greatly reduced, especially for the second exciation, with an error of just 0.03 eV. Including the next carbon and its hydrogen, case B, reduces the error of the first excitation to less than 10^{-3} eV while the second excitation is 0.03 eV too high.

For case C and D only slight changes from case B are observed, even though almost half the molecule is treated with CCS in model D. To get a better understanding of the behavior of the models, we analyze the single excitation vectors. Table B.3 contains the individual contributions from the different spaces to the excitation vectors given in percent. The T $\rightarrow$ T refer to internal to the CCSD space while S $\rightarrow$ S to internal to the CC2 space. The semi external excitations are divided into two groups. T $\rightarrow$ S, from CCSD to CC2, and S $\rightarrow$ T, the opposite. Excitations involving CCS orbitals are referred to as R.

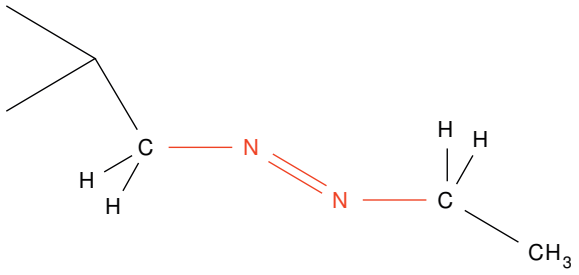
In cases B, C and D, the first excitation is more than 80% internal to the CCSD space. Consequently, the CCSD excitation energy is almost exactly reproduced. The second largest contribution is T $\rightarrow$ S and the error is generally not very sensitive to this type of excitation. For instance in case A, 45% of the first excitation is T $\rightarrow$ S and the error is still small. The second excitation is an even better example, because the T $\rightarrow$ T part is only 7%, but ECC2 is still much more accurate than CC2. Note that the CCS contribution is less than 1% for all excitations. Large CCS contributions tend to give too high excitation energies.

For all the systems presented in this paper, the character of the double excitation part of the excitations has little additional explanatory value. Most of the excitations have over 90% single excitation character and the distribution of the doubles excitations between spaces are more or less the same as for the single excitations.

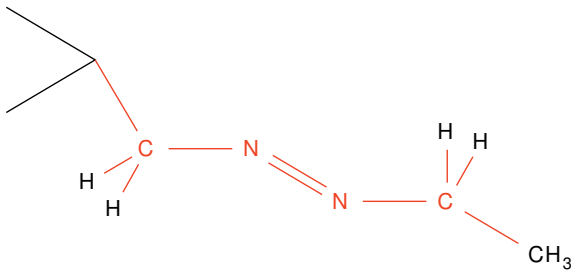
In Figure B.8 we present three different partitions of trans-ethyl-i-butyl-diazene with a geometry from Pubchem [223]. In case A, the nitrogen atoms are included in a CCSD space and the rest treated with CC2. For case B the active space is expanded to include the neighboring carbon atoms. For case C, the hydrogen atoms are included as well. Including the hydrogen atoms does not increase the number

Table B.3: Contributions in percent from the different types of single excitations for decanal using cc-pVDZ.

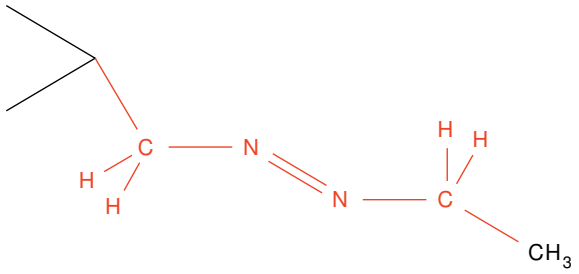
	Exci.	T → T	T → S	S → T	S → S	R	total
A	1	48.3	44.7	0.4	0.5	-	93.9
	2	6.7	57.0	0.7	29.8	-	94.3
B	1	83.3	9.7	0.4	0.1	-	93.5
	2	38.7	28.9	3.7	22.9	-	94.2
C	1	83.7	9.7	0.4	0.1	0.0	93.9
	2	46.9	29.3	3.6	14.2	0.1	94.1
D	1	83.8	9.6	0.4	0.1	0.0	93.9
	2	70.3	9.0	12.1	2.3	0.6	94.3



A



B



C

Figure B.8: Subsystems of trans-ethyl-i-butyl-diazene. See Fig B.7 for explanation.

Table B.4: Excitation energies of trans-ethyl-i-butyl-diazene in eV using cc-pVDZ above and aug-cc-pVDZ below.

	CCSD	CC2	A	B	C
1	3.40	3.38	3.41	3.41	3.41
2	7.41	7.48	7.43	7.40	7.40
3	8.14	7.94	8.14	8.10	8.13
4	8.31	8.02	8.24	8.29	8.28
5	8.46	8.03	8.47	8.44	8.45
1	3.39	3.36	3.41	3.40	3.40
2	6.24	5.71	6.15	6.20	6.21
3	6.55	6.02	6.46	6.52	6.53
4	6.60	6.06	6.51	6.57	6.58
5	6.88	6.28	6.75	6.83	6.83

of occupied active orbitals as the bonding electrons are included in the first space generated. However, the number of active virtual orbitals will increase.

Excitation energies were calculated using the basis sets cc-pVDZ and aug-cc-pVDZ and the results are reported in Table B.4. With the smallest basis set, CC2 performs well for the first excitation, but the error rapidly increases for the higher excitations. A minimal active space including only the nitrogen atoms is enough to reduce the error significantly. For case C the CCSD excitation energies are more or less reproduced. Table B.5 contains the character of the single excitations and we note that the internal contribution is not particularly high for the higher excitations.

With an augmented basis set, the CC2 results are much worse for all but the first excitation. The error also increases for the smallest active space, but to a much lesser extent. For cases B and C, the accuracy is almost as good as without augmentation, even though only a small part of the higher excitations is internal to the active space (See supplementary material [222] for contributions to the excitations.). Even for case C, at least 80% of these excitations are T $\rightarrow$ S.

An obvious application of MLCC is modeling molecules in solutions. The solute molecule can be described at high accuracy while the solution molecules are treated in a more approximate fashion. To explore these possibilities, we have chosen a model system consisting of one acetone and three water molecules shown in Figure B.9. The geometry was optimized with B3LYP [224] using Gaussian [225] (See supplementary material [222] for the geometry). In case A, acetone is treated with CCSD while the water molecules are treated with CC2. In the second case, the two water molecules further away from acetone are treated with CCS.

For case A, the first ECC2 excitation energy is very close to the CCSD value using cc-pVDZ. For the higher excitations, however, there are considerable errors. This is consistent with the percentages shown in Table B.7 as the first excitation is almost entirely internal in the active space while the higher are almost entirely

Table B.5: Contributions in percent from the different types of single excitations for trans-ethyl-i-butyl-diazene using cc-pVDZ.

	Exci.	T → T	T → S	S → T	S → S	total
	1	62.1	31.1	1.0	0.6	94.9
	2	40.4	21.0	16.2	11.3	89.0
A	3	23.4	18.7	29.3	23.9	95.3
	4	27.2	14.2	30.5	22.4	94.3
	5	17.8	72.5	1.1	2.9	94.4
	1	78.9	15.8	0.1	0.0	94.8
	2	70.3	15.0	2.2	0.6	88.0
B	3	66.4	19.2	7.3	2.0	94.9
	4	69.8	16.0	6.6	1.8	94.2
	5	21.4	71.8	0.3	0.6	94.0
	1	89.2	5.4	0.1	0.0	94.7
	2	80.6	5.3	1.8	0.2	87.9
C	3	81.6	6.5	6.0	0.6	94.7
	4	78.8	5.7	8.8	0.8	94.2
	5	56.6	36.6	0.4	0.4	94.0

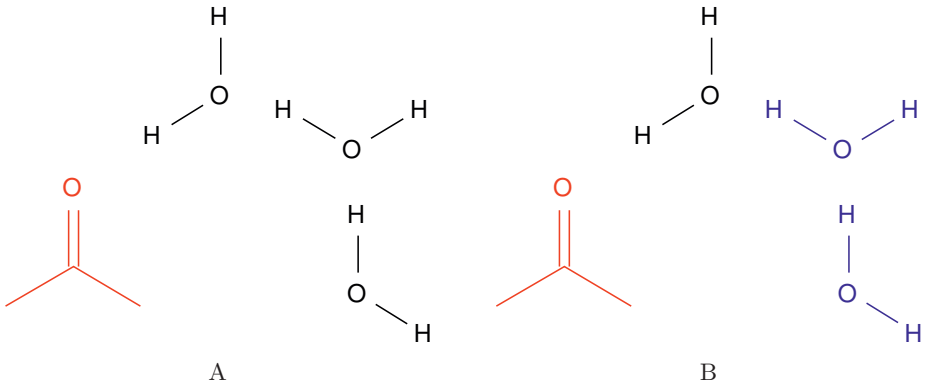


Figure B.9: Subsystems of acetone-water complex. See Fig B.7 for explanation.

Table B.6: Excitation energies of acetone with water in eV using cc-pVDZ above and aug-cc-pVDZ below.

	CCSD	CC2	A	B
1	4.81	4.87	4.81	4.91
2	8.40	8.24	8.28	8.90
3	8.75	8.31	8.64	9.20
1	4.76	4.76	4.77	4.93
2	6.96	6.31	6.95	7.36
3	7.51	7.04	7.12	8.09

Table B.7: Contributions in percent from the different types of single excitations for acetone with water using cc-pVDZ.

	Exci.	T → T	T → S	S → T	S → S	R	total
	1	84.1	0.7	8.9	0.1	-	93.8
A	2	0.4	1.0	5.3	89.4	-	96.0
	3	0.2	0.1	1.4	94.2	-	96.0
	1	85.8	0.1	0.4	0.0	8.0	94.3
B	2	0.1	0.2	1.6	93.1	1.6	96.7
	3	68.3	9.0	0.6	0.7	16.5	95.1

Table B.8: Contributions in percent from the different types of single excitations for acetone with water using aug-cc-pVDZ.

	Exci.	T → T	T → S	S → T	S → S	R	total
	1	76.7	4.9	11.4	0.7	-	93.7
A	2	41.6	34.4	9.6	8.8	-	94.4
	3	1.3	1.4	3.7	87.8	-	94.1
	1	79.6	1.1	0.3	0.0	13.8	94.9
B	2	55.3	18.2	0.5	0.2	21.4	95.7
	3	71.2	4.4	0.8	0.1	19.0	95.4

external. The first excitation of acetone is much lower than those of water and is almost entirely localized on acetone. The higher excitation energies, however, are about the same energy as those of water and have large external contributions. A similar pattern is observed with the aug-cc-pVDZ basis, except that now the second excitation is also mostly localized on acetone and is described accurately.

When including a CCS space the acetone excitation energies and errors increase. This is due to the considerable contributions from the CCS space. Case B gives solvent shifts of 0.45 eV and 0.49 eV with and without augmentation respectively for the first excitation. This is higher than the 0.35 eV and 0.32 eV from full CCSD, but more accurate than not including the water molecules. All three water molecules in this system are close to acetone, making the system too small to accommodate a CCS space. With enough water molecules to form a second or third solvation shell, it is likely that CCS can accurately describe the effect of the outer shells.

B.4 Conclusions

In this paper, we have expand upon the ECC2 model [209] and presented the detailed derivation of the equations, including how to incorporate triple excitations

in the MLCC framework. Initial tests of the ECC2(T) model using a proof of principle implementation are encouraging and demonstrate a two orders of magnitude reduction in computational requirements. We have also derived the ECC2 response function and calculated excitation energies for various test systems.

Our experience so far shows that multi-level coupled cluster theory is useful when treating the entire system with an accurate model is too expensive. The model is not black box and the user must make an appropriate choice of active spaces and verify that the results are accurate. For some systems and properties, such as the lowest excitation energies of decanal, it is obvious where the active space should be placed. In this case, verification can then be achieved by simultaneously performing several calculations with different numbers of active atoms and comparing the change in the excitation energies. In all CCSD/CC2 calculations we have performed so far, excitation energies with more than 80% internal character deviate 0.01 eV or less from the CCSD values. In several systems we have studied, ECC2 has produced excitation energies deviating less than 0.05 eV from the CCSD values, even with significant semi external character and poor performance by CC2.

Adding a CCS space makes it difficult to determine the accuracy of the results from the internal character of the excitation alone. This happens because the increased excitation energies in CCS can remove contributions to the excitation that would be included in a standard calculation. In some cases, acetone for example, this can even change the order of the excitation energies. To ensure that all relevant parts of the system are included in the active space, differences between ground- and excited state electron densities can be very useful. Initial results indicate that CCS gives a sufficient qualitative description to determine apropriate active space, even for more difficult cases such as conjugated and charge transfer systems. These findings will be published elsewhere.

For large systems, it is possible to use several levels of theory. Far away from the atoms of interest, the Hartree-Fock wave function can be employed. If the system size is increased even further, the computational complexity will scale as Hartree-Fock. This makes it possible to obtain accurate local properties without the prohibitive computational complexity encountered in standard CC theory.

Acknowledgments

We acknowledge support from the Spanish MINECO through project number CTQ2010-19738 and NOTUR for computer time. H.K. acknowledges financial support from the FP7-PEOPLE-2013-IOF funding scheme (Project 625321). H.K. and R.H.M. would like to thank Todd J. Martínez for hosting the project at Stanford University.

Paper C

The multilevel CC3 coupled cluster model

J. Chem. Phys. **145**, 044111 (2016)

The multilevel CC3 coupled cluster model

Rolf H. Myhre and Henrik Koch*

Department of Chemistry, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway
Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California 94305, USA

Abstract

We present an efficient implementation of the closed shell multilevel coupled cluster method where coupled cluster singles and doubles is used for the inactive orbital space and CCSD with perturbative triples is employed for the smaller active orbital space. Using Cholesky orbitals, the active space can be spatially localized and the computational cost is greatly reduced compared to full CC3 while retaining the accuracy of CC3 excitation energies. For the small organic molecules considered we achieve up to two orders of magnitude reduction in the computational requirements.

C.1 Introduction

Wave function methods are the most accurate methods available in contemporary electronic structure theory. They can be organized into systematic hierarchies of models with increasing accuracy. This property is important when developing systematic approaches to calculating molecular properties. The application range of wave function theory is limited by the steep scaling of computational requirements with system size. Consequently, the development of wave function methods for large systems is still a major challenge in quantum chemistry and the focus of intense research efforts.

When the entire system is described at the same level of theory we may obtain reductions in computational cost by discarding interactions between electrons that are far away from each other. The projected atomic orbital (PAO) method introduced by Pulay and Sæbø [71, 72] has been used to obtain reduced scaling algorithms for configuration interaction [176] and Møller-Plesset perturbation theory [177]. The use of PAO orbitals has also been explored for coupled cluster singles and doubles (CCSD) [226].

Among recently developed coupled cluster (CC) methods for local correlation we mention the local pair natural orbital CCSD method (LPNO-CCSD) [76, 190] and the orbital-specific-virtual local CCSD method [78, 191]. Riplinger et al. [77] used an implementation of LPNO-CCSD(T) to perform single point energy calculations of the protein crambin that contains 640 atoms. Schütz et al. [79] also performed single point CCSD(T) energy calculations on nucleobase dimers using orbital-specific virtual local triples corrections. Alternative ways to achieve reduced scaling is to divide the molecule into fragments and calculate interactions of the fragments using a screening protocol [73, 75, 185, 227, 228]. For response properties Helmich and Hättig [192] implemented CC2 excitation energies [23] using the PNO method combined with orbital specific virtuals.

An alternative approach to reduced scaling is to exploit the linear dependence of orbital products in the two-electron integral matrix. The most well known of these methods are density fitting [61–65, 229] and Cholesky decomposition [66–69]. These methods employ an auxiliary basis that is generally much smaller than the product basis. A more recent approach is tensor hyper-contraction [25, 206, 207]. In this approximation the integral parametrization is obtained by least squares fitting to exact or density fitted integrals. In this way an N^4 scaling CC2 algorithm has been implemented where N is a measure of system size.

Recently, we introduced the multilevel coupled cluster (MLCC) formalism [174] that makes it possible to apply different levels of coupled cluster theory to different parts of the molecular system. In this paper we describe the implementation of the multilevel CC3 (MLCC3) model that use CC3 [24, 87] to describe the active part of the molecule and CCSD [21] for the rest. The CC3 model scales as N^7 while CCSD scales as N^6. With an appropriate active space, the overall scaling

becomes N^6 while retaining the accuracy of CC3. We note that similar methods with energy based active spaces have been proposed [86, 230]. Even for relative small systems where most atoms are active the steep scaling of CC3 results in large reductions in computational requirements compared to full CC3. The framework may be extended with additional spaces with lower accuracy and scaling, but this will not be reported here.

Cannonical orbitals are highly delocalized and not suitable for partitioning the orbital space into sets corresponding to fragments of the molecule; thus localized orbitals must be used. In principle, any type of orbital localization procedure can be used and we employ Cholesky orbitals [88]. These orbitals are generated by Cholesky decomposing the HF density matrix. Each orbital corresponds to a diagonal element in the density matrix which again corresponds to an atomic orbital (AO). This makes it straightforward to assign an orbital space corresponding to the active atoms and the orbitals are generated non-iteratively. A possible disadvantage of Cholesky orbitals is that diffuse orbitals close to the HOMO-LUMO gap may be left out of the active space. We will demonstrate one way of overcoming this problem by explicitly including orbitals in the active space. We obtain standard CC3 when the entire molecule is included in the active space.

The above discussed methods describe the entire molecular system at the same level. This is contrary to multilevel approaches where the computational resources are focused on a part of the system. Unlike other multilevel approaches like ONIOM [80] and frozen density embedding [82, 83, 231], MLCC describes the whole system with a single CC wave function. In principle, size intensive scaling can be achieved by expanding the model with more spaces and describing the lowest space at the HF level or lower.

In Section C.2 we present the theory behind MLCC with a focus on the MLCC3 model with equations for the ground state energy and excitation energies. In Section C.3 we discuss the algorithm of our implementation and in Section C.4 we present example calculations and compare timings to full CC3. The Section C.5 contains our conclusion and future perspectives.

C.2 Theory

C.2.1 The MLCC3 model

We start our discussion by deriving the equations for the closed shell MLCC3 model. We refer to Ref. [174] for a general derivation with more than two spaces and a detailed discussion of how to generate the associated orbital spaces. The CC3 [24] model includes triple excitations as a perturbative correction to CCSD [21] and can be viewed as an intermediate between CCSD and full CCSDT. The number of T_3 amplitudes scales as V^3O^3, where V is the number of virtual orbitals

and O is the number of occupied orbitals. An implementation that explicitly stores the amplitudes will have limited applicability, even in an active space approach. However, the T_3 amplitudes do not enter the energy expression for CC3 and can be calculated independently of each other. Consequently, they can be calculated on the fly without explicit storage.

We define a general cluster operator X,

$$X = \sum_{\mu} x_\mu \tau_\mu, \tag{C.1}$$

where the index μ runs over the excitation manifold, $\{\langle\mu|\}$, x_μ is the corresponding amplitude and τ_μ an excitation operator from the reference to $|\mu\rangle$. In two level MLCC3, the cluster operator is divided into two separate operators, T and S. Each of the operators are associated with a projection manifold, $\{\langle\mu^T|\}$ and $\{\langle\mu^S|\}$. Excitations involving only active orbitals are called internal in the active space and are included in T. Excitations involving both active and inactive orbitals are called semi-external and those only involving inactive orbitals are called external. Both semi-external and external excitations are included in S. Inserting these definitions into the expression for the CC wave function and using the Hartree-Fock (HF) state as reference state we obtain

$$|\mathrm{CC}\rangle = \exp(X)\,|\mathrm{HF}\rangle = \exp(T+S)\,|\mathrm{HF}\rangle\,. \tag{C.2}$$

The X_1 and X_2 are treated the same way in CCSD and CC3 and no special notation is needed for these operators. All external and semi-external T_3 amplitudes are defined to be zero and the cluster operator is written as $X = X_1 + X_2 + T_3$ with

$$X_1 = \sum_{AI} x_I^A E_{AI} \tag{C.3}$$

$$X_2 = \frac{1}{2}\sum_{ABIJ} x_{IJ}^{AB} E_{AI} E_{BJ} \tag{C.4}$$

$$T_3 = \frac{1}{6}\sum_{abcijk} t_{ijk}^{abc} E_{ai} E_{bj} E_{ck}. \tag{C.5}$$

Capital letters refer to general indices that run over all orbitals, both active and inactive. $A, B, C...$ refer to virtual orbitals and $I, J, K...$ occupied orbitals. The indices $a, b, c...$ and $i, j, k...$ refer to virtual and occupied active orbitals respectively. The E_{AI} operator is a singlet excitation operator from orbital I to A. Similarly to the cluster operator, the projection manifold is written as

$$\{\langle\mu|\} = \{\langle\mu_1|\} \oplus \{\langle\mu_2|\} \oplus \{\langle\mu_3^T|\} \tag{C.6}$$

with no $\{\langle\mu_3^S|\}$ manifold.

Because the T_3 amplitudes do not appear in the energy expression, it is the same in CC3 and CCSD. The CC amplitude equations are solved by projection [21].

Inserting the above definitions into these equations gives [24]

$$\begin{aligned}\Omega_{\mu_1} &= \langle\mu_1|\exp(-X_2)\hat{H}\exp(X_2)\,|\mathrm{HF}\rangle + \langle\mu_1|\,[H,T_3]\,|\mathrm{HF}\rangle \\ &= \Omega_1^{\mathrm{CCSD}} + \langle\mu_1|\,[H,T_3]\,|\mathrm{HF}\rangle = 0\end{aligned} \tag{C.7}$$

$$\begin{aligned}\Omega_{\mu_2} &= \langle\mu_2|\exp(-X_2)\hat{H}\exp(X_2)\,|\mathrm{HF}\rangle + \langle\mu_2|\,[\hat{H},T_3]\,|\mathrm{HF}\rangle \\ &= \Omega_2^{\mathrm{CCSD}} + \langle\mu_2|\,[\hat{H},T_3]\,|\mathrm{HF}\rangle = 0\end{aligned} \tag{C.8}$$

$$\begin{aligned}\Omega_{\mu_3^T} &= \langle\mu_3^T|\,[\hat{H},T_3] + [\hat{H},X_2] \\ &+ [[\hat{H},X_2],X_2] + [[H,T_3],X_2]\,|\mathrm{HF}\rangle = 0.\end{aligned} \tag{C.9}$$

In Eqs. (C.7)-(C.9) we have used the X_1-transformed Hamilton operator,

$$\hat{H} = \exp(-X_1)H\exp(X_1) \tag{C.10}$$

Equation (C.9) corresponds to the amplitude equation for MLCCSDT. In CC3, the Hamiltonian is divided into the Fock operator F and the fluctuation potential U. The X_2 amplitudes are then considered first order in the perturbation, while the T_3 amplitudes are considered second order. Single excitations are used as relaxation parameters and included to infinite order. Removing all third order terms and higher from Eq. (C.9) results in the CC3 amplitude equations

$$\Omega_{\mu_3^T} = \langle\mu_3^T|\,[F,T_3] + [\hat{U},X_2]\,|\mathrm{HF}\rangle\,. \tag{C.11}$$

For canonical orbitals, the Fock operator is diagonal and equation (C.11) can be solved explicitly for the T_3 amplitudes. When using localized MOs, the Fock matrix is no longer diagonal, but can be block diagonalized for each of the orbital spaces. This ensures that any non-diagonal elements will be in the active-inactive block of the Fock matrix where the corresponding T_3 amplitudes are zero.

Projecting the commutator of the Fock operator and triple excitation operators against the excitation manifold results in the difference between the Fock diagonal elements. This corresponds to the energy differences between occupied and virtual orbitals.

$$\epsilon_{ijk}^{abc} = \epsilon_a + \epsilon_b + \epsilon_c - \epsilon_i - \epsilon_j - \epsilon_k. \tag{C.12}$$

Assuming a biorthogonal basis, the T_3 amplitudes can then be expressed as

$$t_{ijk}^{abc} = -\frac{1}{\epsilon_{ijk}^{abc}}\,\langle{}^{abc}_{ijk}|\,[\hat{U},X_2]\,|\mathrm{HF}\rangle \tag{C.13}$$

and Eq. (C.13) is inserted into Eqs. (C.7) and (C.8).

C.2.2 The working equations of MLCC3

The terms that need to be computed are most conveniently expressed using a biorthogonal basis. In this parametrization the single and double excited manifolds

are written as

$$\langle {}^{a}_{i}| = \frac{1}{2} \langle \mathrm{HF}| \, E_{ia} \tag{C.14}$$

$$\langle {}^{ab}_{ij}| = \frac{1}{6} \langle \mathrm{HF}| \, (2E_{jb}E_{ia} + E_{ib}E_{ja}). \tag{C.15}$$

Rewriting Eq. (C.13) in terms of integrals and amplitudes results in

$$t^{abc}_{ijk} = -P^{abc}_{ijk} \frac{1}{\epsilon^{abc}_{ijk}} \left\{ \sum_{D} t^{aD}_{ij} (bD\hat{|}ck) - \sum_{L} t^{ab}_{iL} (Lj\hat{|}ck) \right\}. \tag{C.16}$$

The integrals in Eq. (C.16) are the X_1-transformed integrals and P^{abc}_{ijk} is the permutation operator

$$P^{abc}_{ijk} O^{abc}_{ijk} = O^{abc}_{ijk} + O^{acb}_{ikj} + O^{bac}_{jik} + O^{cba}_{kji} + O^{cab}_{kij} + O^{bca}_{jki}. \tag{C.17}$$

Calculating the T_3 amplitudes will scale as $VV_A^3O_A^3$ where V_A is the number of active virtual and O_A the number of active occupied orbitals. If the size of the molecular system is increased without changing the active space, the cost of calculating the T_3 amplitudes will scale linearly with the total system size. The permutation symmetry in Eq. (C.16) over the occupied-virtual index pairs can be exploited to significantly reduce computational cost.

The triples contribution to $\mathbf{\Omega}_1$ becomes

$$\Omega^{\mathrm{A3}}_{ai} = \sum_{bcjk} (t^{abc}_{ijk} - t^{cba}_{ijk}) L_{jbkc} \tag{C.18}$$

where L_{jbkc} is defined as

$$L_{jbkc} = 2(jb|kc) - (kb|jc). \tag{C.19}$$

These terms scale as $V_A^3O_A^3$ and the scaling is size intensive with respect to total system size. However, their computational cost is typically much smaller than the cost of the $\mathbf{\Omega}_2$ terms.

Equation (C.8) produces three terms involving the triples, one Fock term, one occupied and one virtual term:

$$\Omega^{\mathrm{B3}}_{abij} = P^{ab}_{ij} \sum_{ck} \left(2t^{abc}_{ijk} - t^{cba}_{ijk}\right) \hat{F}_{kc} \tag{C.20}$$

$$\Omega^{\mathrm{C3}}_{abLI} = -P^{ab}_{LI} \sum_{cjk} \left(2t^{bac}_{Ijk} - t^{bca}_{Ijk} - t^{cab}_{Ijk}\right) (jL\hat{|}kc) \tag{C.21}$$

$$\Omega^{\mathrm{D3}}_{ADij} = P^{AD}_{ij} \sum_{bck} \left(2t^{Abc}_{ijk} - t^{cbA}_{ijk} - t^{Acb}_{ijk}\right) (Db\hat{|}kc) \tag{C.22}$$

The T_3 amplitudes with general indices appear in the C3 and D3 terms. These are zero if the index is inactive, but are written out to be consistent with the

permutation operator. Terms involving semi-external T_3 amplitudes are therefore omitted in the algorithm. The virtual term, D3, is the most computationally expensive and scales as $VV_A^3O_A^3$ while the C3 and B3 terms scale as $OV_A^3O_A^3$ and $V_A^3O_A^3$, respectively. The above expressions are identical to standard CC3 with six indices restricted to the active space. When including all orbitals in the active space we obtain standard CC3. The overall scaling of the T_3 contributions is linear with total system size.

A variation of MLCC3 is to restrict all indices to the active space. The scaling of this model is $V_A^4O_A^3$ and only depends on the size of the active space making the T_3 terms computationally size intensive. Yet another variant is to define two CCSD spaces and only include orbitals from one of them in the MLCC3 equations.

C.2.3 MLCC3 excitation energies

Within a multilevel framework the total correlation energy is not evaluated with high accuracy as the higher level correlation treatment is restricted to a part of the system. On the other hand, properties like excitation energies and transition moments are size intensive and only depend on the local environment. Therefore these properties are prime candidates for a multilevel approach. In CC theory the response functions can be derived from time dependent expectation values [31, 32, 87] and the quasi-energy (QE) [32, 38, 87].

The starting point for CCLR is a time dependent perturbation, V^t, that is added to the time independent Hamiltonian H_0.

$$H = F + U + V^t = H_0 + V^t \tag{C.23}$$

In the QE approach, V^t must be assumed to be periodic and is written as

$$V^t = \sum_{j=-N}^{N} \sum_{A} A\epsilon_A(\omega_j) \exp(-i\omega_j t). \tag{C.24}$$

Furthermore, we assume that A is a frequency independent one-electron operator. To ensure hermiticity, we require that A is Hermitian,

$$\omega_{-j} = \omega_j \tag{C.25}$$

and that $(\epsilon_A(\omega_j))^* = \epsilon_A(-\omega_j)$. The amplitudes can now be expanded in orders of the time dependent perturbation as

$$x_\mu(t) = x_\mu^{(0)} + x_\mu^{(1)}(t) + x_\mu^{(2)}(t) + ... \tag{C.26}$$

Expanding the first order amplitudes in the frequencies gives

$$\begin{aligned} x_\mu^{(1)}(t) &= \sum_{j=-N}^{N} x_\mu^{(1)}(\omega_j)\exp(-i\omega_j t) \\ &= \sum_{j=-N}^{N}\sum_{A} x_\mu^{A}(\omega_j)\epsilon_A(\omega_j)\exp(-i\omega_j t). \end{aligned} \tag{C.27}$$

Similarly, a set of Lagrangian multipliers, $\bar{x}$, is introduced. These can be expanded in the same way as the amplitudes.

Using the procedure described in reference [87], the following QE Lagrangian, $L(t)$, is obtained

$$\begin{aligned} L(t) &= \langle \mathrm{HF}|\, H\exp(X_1+X_2)\,|\mathrm{HF}\rangle \\ &+ \sum_{\mu_1}\bar{x}_{\mu_1}\left(\langle\mu_1|\,\hat{H}+[\hat{H},X_2+T_3]\,|\mathrm{HF}\rangle - i\frac{\partial x_{\mu_1}}{\partial t}\right) \\ &+ \sum_{\mu_2}\bar{x}_{\mu_2}\left(\langle\mu_2|\,\hat{H}+[\hat{H},X_2+T_3]+\frac{1}{2}[[\hat{H},X_2],X_2]\,|\mathrm{HF}\rangle - i\frac{\partial x_{\mu_2}}{\partial t}\right) \\ &+ \sum_{\mu_3^T}\bar{t}_{\mu_3^T}\left(\langle\mu_3^T|\,[F,T_3]+[\hat{H},X_2]+\frac{1}{2}[[\hat{V}^t,X_2],X_2]\,|\mathrm{HF}\rangle - i\frac{\partial t_{\mu_3^T}}{\partial t}\right) \end{aligned} \tag{C.28}$$

The QE Lagrangian is not variational, but its time average $\{L\}_\mathcal{T}$ must satisfy the stationary conditions

$$\frac{\partial}{\partial t_\mu^{(m)}}\{L^{(n)}\}_\mathcal{T} = \frac{\partial}{\partial \bar{t}_\mu^{(m)}}\{L^{(n)}\}_\mathcal{T} = 0. \tag{C.29}$$

for all $m \leq n$. In order to obtain first order quantities, zero order multipliers and zero and first order amplitudes are required. Equation (C.29) results in the same expressions for the time independent amplitudes as Eqs. (C.7),(C.8) and (C.11). The multipliers satisfy

$$\bar{\boldsymbol{x}}^{(0)}\boldsymbol{A} = \boldsymbol{\eta}^{(0)} \tag{C.30}$$

with the vector $\boldsymbol{\eta}^{(0)}$ given by

$$\eta_\nu^{(0)} = -\langle \mathrm{HF}|\,[\hat{H},\tau_\nu]\,|\mathrm{HF}\rangle \tag{C.31}$$

and the Jacobian matrix $\boldsymbol{A}$ given below. The first order amplitudes are found by solving

$$(\boldsymbol{A}-\omega\mathbf{1})\boldsymbol{x}^{(A)}(\omega_j) = -\boldsymbol{\xi}^A \tag{C.32}$$

with

$$\boldsymbol{\xi}^A = \begin{pmatrix} \langle\mu_1|\,\hat{A}+[\hat{A},X_2]\,|HF\rangle \\ \langle\mu_2|\,[\hat{A},X_2+T_3]\,|HF\rangle \\ \langle\mu_3^T|\,[\hat{A},T_3]+\frac{1}{2}[[\hat{A},X_2],X_2]\,|HF\rangle \end{pmatrix} \tag{C.33}$$

and the Jacobian matrix

$$\boldsymbol{A} = \begin{pmatrix} \langle\mu_1|\,[\hat{H}_0 + [\hat{H}_0, X_2], \tau_{\nu_1}]\,|\mathrm{HF}\rangle & \langle\mu_1|\,[\hat{H}_0, \tau_{\nu_2}]\,|\mathrm{HF}\rangle & \langle\mu_1|\,[\hat{H}_0, \tau_{\nu_3^T}]\,|\mathrm{HF}\rangle \\ \langle\mu_2|\,[\hat{H}_0 + [\hat{H}_0, X_2 + T_3], \tau_{\nu_1}]\,|\mathrm{HF}\rangle & \langle\mu_2|\,[\hat{H}_0 + [\hat{H}_0, X_2], \tau_{\nu_2}]\,|\mathrm{HF}\rangle & \langle\mu_2|\,[\hat{H}_0, \tau_{\nu_3^T}]\,|\mathrm{HF}\rangle \\ \langle\mu_3^T|\,[[\hat{H}_0, X_2], \tau_{\nu_1}]\,|\mathrm{HF}\rangle & \langle\mu_3^T|\,[\hat{H}_0, \tau_{\nu_2}]\,|\mathrm{HF}\rangle & \delta_{\mu_3^T \nu_3^T} \epsilon_{\mu_3^T} \end{pmatrix} \tag{C.34}$$

The above expressions are the same as in standard CC3 response theory with the indices in the triples restricted to the active space. Excitation energies correspond to the poles in the response function, i.e. when the matrix in Eq. (C.32) becomes singular.This occurs when ω is an eigenvalue of the Jacobian.

Solving the eigenvalue problem requires the computation of the matrix-vector products

$$\boldsymbol{\sigma} = \boldsymbol{AC}. \tag{C.35}$$

The vectors $\boldsymbol{C}_3$ and $\boldsymbol{\sigma}_3$ have the same dimension as T_3 and are too large to be stored. Instead, the Jacobian is partitioned and the triples part of $\boldsymbol{\sigma}$ is never explicitly calculated. Inspection of the matrix in Eq. (C.34), reveals that $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ can be expressed as the CCSD terms with some additional contributions from the triples

$$\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_1^{CCSD} + \langle \mu_1 | \left[H_0, C_3 \right] | \mathrm{HF} \rangle \tag{C.36}$$

$$\boldsymbol{\sigma}_2 = \boldsymbol{\sigma}_2^{CCSD} + \langle \mu_2 | \left[[\hat{H}_0, C_1], T_3 \right] | \mathrm{HF} \rangle + \langle \mu_2 | \left[\hat{H}_0, C_3 \right] | \mathrm{HF} \rangle . \tag{C.37}$$

The $\boldsymbol{\sigma}_3$ vector is expressed as

$$\begin{aligned} \boldsymbol{\sigma}_3^T = & \langle \mu_3^T | \left[F, C_3^T \right] | \mathrm{HF} \rangle \\ & + \langle \mu_3^T | \left[\hat{H}_0, C_2 \right] | \mathrm{HF} \rangle + \langle \mu_3^T | \left[[\hat{H}_0, C_1], X_2 \right] | \mathrm{HF} \rangle , \end{aligned} \tag{C.38}$$

and for a given eigenvalue, ω, the elements of $\boldsymbol{C}_3$ becomes

$$c_{\mu_3^T} = \frac{-1}{\epsilon_{\mu_3^T} - \omega} \Big(\langle \mu_3^T | \left[\hat{H}_0, C_2 \right] | \mathrm{HF} \rangle + \langle \mu_3^T | \left[[\hat{H}_0, C_1], X_2 \right] | \mathrm{HF} \rangle \Big). \tag{C.39}$$

Expressing $[\hat{H}_0, C_1]$ as a Hamiltonian $\tilde{H}_0$ with one-index transformed integrals, Eqs. (C.36),(C.37) and (C.39) have the same structure as the ground state equations, (C.7),(C.8) and (C.13). Partitioning the eigenvalue problem induces a dependence on ω in the matrix and the equations must be iterated until self consistent.

C.3 Implementation

We have implemented the closed shell MLCC3 model in a local version of the Dalton software package [99]. The ground state energy and singlet excitation energies have been implemented. In order to make the implementation useful for many-level MLCC theory, e.g. a hierarchy of methods like frozen density HF, HF, CCS, CC2, CCSD and CC3, the MLCC3 has been implemented as independent of the Dalton CC package as possible. Point group symmetry has not been utilized and the basic algorithms are developed to take maximal advantage of BLAS libraries for matrix and vector operations.

In Dalton the coupled cluster models are implemented with integrals in the AO basis. This approach is not suitable for MLCC because computational reductions are

Algorithm 1 MLCC3 algorithm

for i=1,O_A **do**
  for j=1,i **do**
    for k=1,j **do**
      $W^{ijk}(abc) = -P^{abc}_{ijk}\{ \sum_D t^{aD}_{ij}(bD\hat{|}ck) - \sum_L t^{ab}_{iL}(Lj\hat{|}ck)\}$
      $t^{abc}_{ijk} = \frac{1}{\epsilon^{abc}_{ijk}} W^{ijk}(abc)$
      for permutations **do**
        $\Omega^a_i = \sum_{bc}(t^{abc}_{ijk} - t^{cba}_{ijk})L_{jbkc}$
        $\Omega^{ab}_{ij} = \sum_c (t^{abc}_{ijk} - t^{cba}_{ijk})\hat{F}_{kc}$
      end for
      for permutations **do**
        $\Omega^{ab}_{Lj} = -\sum_c (2t^{abc}_{ijk} - t^{cba}_{ijk} - t^{acb}_{ijk})(iL\hat{|}kc)$
        $\Omega^{aD}_{ij} = \sum_{bc}(2t^{abc}_{ijk} - t^{cba}_{ijk} - t^{acb}_{ijk})(Db\hat{|}kc)$
      end for
    end for
  end for
end for
$\Omega^{AB}_{IJ} = \frac{1}{2}(\Omega^{AB}_{IJ} + \Omega^{BA}_{JI})$

achieved by restricting MO indices and MLCC3 is implemented with MO integrals. The required integrals are the ones in Eqs. (C.7), (C.8) and (C.13) and the corresponding C_1-transformed integrals used in response calculations. The integrals required have at most one inactive index in the MO basis and the most expensive integral transformation scales as N^4O_A, where N is the number of AOs. For small active spaces, the integral tranformation can be more expensive than the CC3 amplitude equations. The MO integrals are sorted to achieve optimal efficiency in the matrix-matrix multiplications.

The implemented algorithm is similar to the CCSD(T) implementation by Rendell *et al.* [102] and is outlined in Algorithm 1. The calculation is driven in an outer loop over the active occupied indices $i \geq j \geq k$. We utilize the symmetry of the T_3 amplitudes,

$$t_{ijk}^{abc} = t_{ikj}^{acb} = t_{jik}^{bac} = t_{kji}^{cba} = t_{kij}^{cab} = t_{jik}^{bca} \tag{C.40}$$

to reduce the cost of calculating amplitudes by approximately a factor six. Note that only V_A^3 amplitudes are available in memory at any time. Consequently, the contributions from these amplitudes must be added to all possible permutations of the Ω indices in each iteration of the loop.

The X_2 amplitudes are also sorted optimally for matrix multiplications. For a full space calculation, the amplitudes are ordered as A, B, J, I. With an active space, the occupied and virtual terms contribute to sets of double amplitudes with different dimensions and two arrays are necessary in memory. Similarly, one or two arrays are allocated for Ω_2. These arrays must be symmetrized and packed after the outer loop over occupied active orbitals. To avoid reading integrals inside the loop, the program checks if it can keep all of them in memory and will read them in if possible. If not, the program will read the integrals in batches over i,j and k in an outer loop.

In total, the terms in Eq. (C.13) require twelve matrix-matrix multiplications inside the outer loop, one virtual and one occupied term for each permutation of the indices. The virtual terms are the most expensive, scaling as VV_A^3. After each call to the matrix-matrix multiplier, the contributions must be reordered to the appropriate permutation. Reordering requires additional computational effort, but has been parallelized using OpenMP. After all the contributions are calculated they are divided by orbital energies and amplitudes with three identical indices are set to zero.

The Ω terms are expressed as products of the amplitudes multiplied by integrals or Fock matrix elements as shown in Algorithm 1. Terms involving the Fock matrix and Ω_1 contributions are multiplied by the same linear combinations of amplitudes and are therefore calculated together. Similarly, the virtual and occupied terms of Ω_2 are calculated together.

C.4 Results

C.4.1 Parallelization

We have tested the code on the Stallo cluster at University of Tromsø [118]. The cluster has two CPUs and 16 or 20 cores per node and we used 8 OpenMP threads for the calculations unless stated otherwise. This is more than the code can utilize for the smaller active spaces, but is appropriate for the larger calculations. To ensure that the measured timings were comparable, all timings given in the tables are from the same node. For this reason, only timings for ground state energy iterations are reported. The excitation energy iterations take about two and a half times longer.

Table C.1: CPU and wall time for the CC3 part of one iteration for Uracil in minutes. Speedup factors are given in parenthesis.

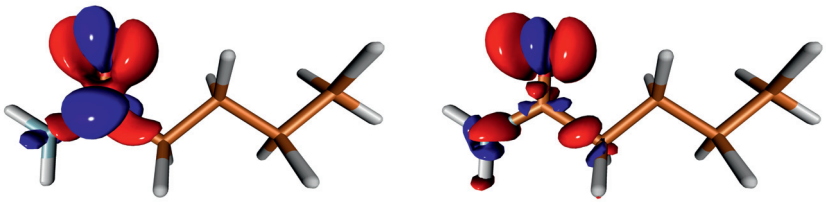
basis	orbitals		1	2	4	8
DZ	29 occ.	CPU	17	20	23	31
	103 virt.	Wall	16	10 (×1.6)	6 (×2.7)	4 (×4.0)
aDZ	29 occ.	CPU	163	202	215	287
	191 virt.	Wall	163	101 (×1.6)	54 (×3.0)	36 (×4.5)

Table C.1 contains timings for the triples part of one iteration of the full CC3 energy of Uracil with 1, 2, 4 and 8 threads. Timings for basis sets cc-pVDZ and aug-cc-pVDZ are reported. While the scaling is not linear with the number of threads, parallelization gives a substantial speedup. With 8 threads the calculations are about 4.5 times faster than with one thread for the largest basis set.It would be desirable to parallelize our CC3 code with MPI, but will require substantial changes to the CCSD code in Dalton.

C.4.2 Excitation energies

We explore the multilevel model on several different molecules using aug-cc-pVDZ basis sets [60] and different active spaces. The systems were chosen so that full CC3 is still feasible, but large enough to be divided into subsystems. The frozen core approximation is not used in any of the calculations.

Pentanamide [232] is an appropriate test system as the change in the electron density is localized around the amide group (see Fig C.1). In Fig. C.2 we show the three different active spaces used. The singlet excitation energies are given in Table C.2 and timings are reported in Table C.3. All calculations were carried out using sufficient memory to avoid batching over the integrals. In model **B**, all excitation



S_1 S_2

Figure C.1: Change in electron density between ground state and the first two singlet excited states of pentanamide calculated with CCSD. Volumes in blue have increased density and volumes in red have decreased density. Isovalue: 0.01

energies except for S_4 has errors less than 0.02 eV. The CCSD excitation energy for S_1 and S_3 differ less than 0.04 eV from CC3, but differences are larger for the other states, in particular the S_2 state that differ 0.17 eV. For model **C** the largest error is 0.013 eV.

Table C.2: Excitation energies in eV calculated for pentanamide with aug-cc-pVDZ. Error compared to Full CC3 in parenthesis.

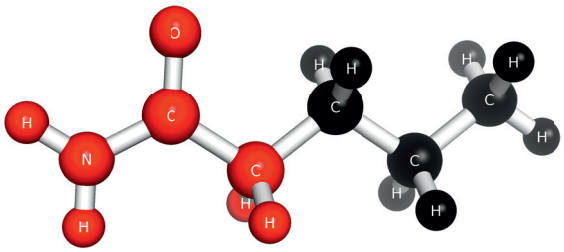
	CCSD	**A**	**B**	**C**	CC3
S_1	5.752 (0.016)	5.745 (0.009)	5.739 (0.003)	5.737 (0.001)	5.736
S_2	6.418 (0.171)	6.309 (0.062)	6.253 (0.006)	6.240 (-0.007)	6.247
S_3	6.565 (0.039)	6.533 (0.007)	6.513 (-0.013)	6.517 (-0.009)	6.526
S_4	7.223 (0.146)	7.180 (0.103)	7.140 (0.063)	7.090 (0.013)	7.077
S_5	7.293 (0.112)	7.269 (0.088)	7.202 (0.021)	7.193 (0.012)	7.181

The iteration time for model **B** is 4.4 times faster than full CC3. There are 28 occupied and 232 virtual orbitals in total. In model **B** there are 24 occupied and 160 virtual active orbitals implying a theoretical speedup of

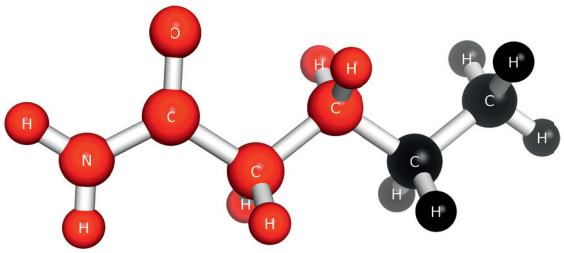
$$\frac{VV^3O^3}{VV_A^3O_A^3} = \left(\frac{28 \times 232}{24 \times 160}\right)^3 \approx 4.8. \tag{C.41}$$

In passing, we note that the integral transformation accounted for about 15% of the CPU time in **B** and about a third in **A**, but only 5% for full CC3. The integral transformation time can be significantly reduced using Cholesky decomposed two-electron integrals [68], however we have not implemented this in the new MLCC3 code.

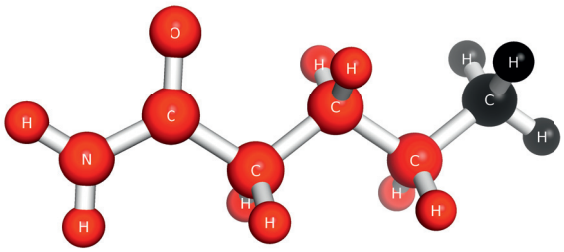
We now consider ethyl-butyl-diazene [223]. This molecule was previously used to investigate the performance of the CCSD/CC2 model (MLCCSD) [174] with excellent results. Based on the density plots in Fig. C.3 we choose the two active spaces shown in Fig. C.4.



A



B



C

Figure C.2: Subsystems of pentanamide. Atoms in red are included in the CC3 space.

Table C.3: Iteration time for ground state energy of pentanamide in minutes. For MLCC3 and CC3 we report time in the triples part. Speedup factors are given in parenthesis. The number of active orbitals in each model is given below.

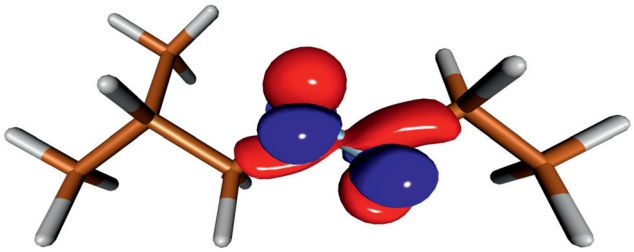
	CCSD	**A**	**B**	**C**	CC3
CPU	29	36	127	297	584
Wall	-	5.3 (×13.9)	17 (×4.4)	38 (×1.9)	74
O_A	28	19	24	27	28
V_A	232	124	160	196	232

After diagonalizing the active and inactive blocks of the Fock matrix we find several virtual orbitals in the inactive block that have lower energy than the active LUMO. This can make the model unbalanced because it will favor excitations out of the active space where the corresponding T_3 amplitudes are set to zero. Indeed, as seen in Table C.4, CCSD is more accurate than models **A** and **B**. Inspection of the excitation vectors confirms they are dominated by semi-external excitations. The vectors for states S_2 to S_5 in model **A** are less that 5% internal while those in **B** are less than 21% internal. The vectors for S_1 are 42% and 77% internal for **A** and **B** respectively. In comparison, the excitation vectors for pentanamide **B** are at least 88% internal except for S_4.

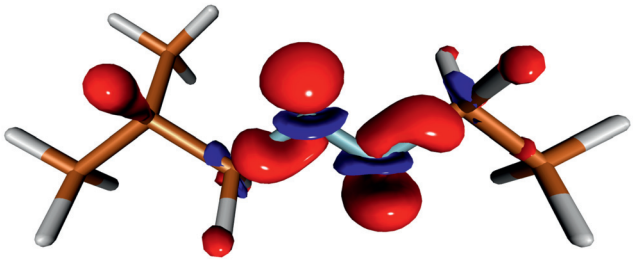
To avoid this problem we use a modified orbital partitioning. In the models labeled with + in Table C.4, virtual inactive Cholesky orbitals with lower energy than the active LUMO are included in the active space. Note this requires a new diagonalization of the active virtual Fock matrix. Including the additional orbitals reduces the error to about 0.03 eV or less for both **A**+ and **B**+. The excitation vector for S_1 in **A**+ is 79% internal while the vectors for the higher states are 90% internal. In **B**+ the vectors are 80% internal or more. We should stress there are several other ways of selecting orbitals in the active space. For instance by considering the diffuse orbitals separately in a virtual density matrix [174] or by well known orbital localization procedures [93]. A detailed analysis of orbital selection procedures will be published elsewhere.

Time per iteration is reported in Table C.5. The addition of virtual orbitals to the active space gives only a small increases in the observed computational time. Ideally, model **A**+ should be almost 800 times faster than full CC3. However, for such a small active space, the dominating part of the CC3 iteration is in the integral transformation. As a result, our calculation is only 90 times faster than full CC3. Excluding the integral transformation time the speedup factor is 900. This is probably due to better cache utilization for smaller active spaces and illustrates the potential improvements using Cholesky decomposed integrals. A similar behavior is observed for model **B**+.

Finally, we present results for hexadiene [233] using the active space indicated in

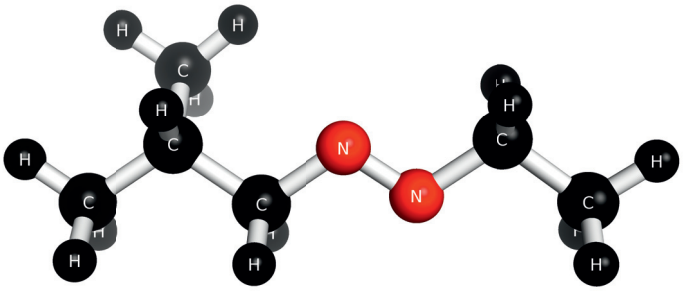


S_1

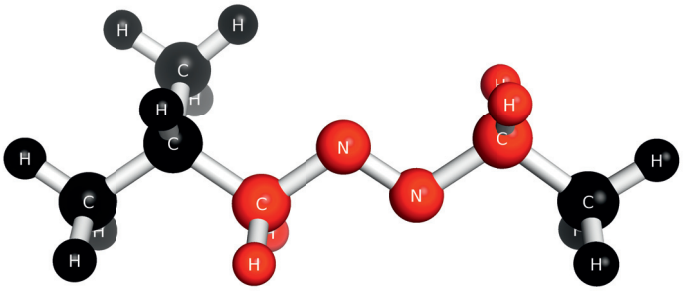


S_2

Figure C.3: Change in electron density between ground state and the first two singlet excited states of ethyl-butyl-diazene calculated with CCSD. Volumes in blue have increased density and volumes in red have decreased density. Isovalue: 0.01



A



B

Figure C.4: Subsystems of ethyl-butyl-diazene. Atoms in red are active.

Table C.4: Excitation energies calculated for ethyl-butyl-diazene, aug-cc-pVDZ. Error compared to Full CC3 in parenthesis.

	CCSD	**A**	**B**	**A+**	**B+**	CC3
S_1	3.394 (0.030)	3.420 (0.056)	3.385 (0.021)	3.387 (0.023)	3.382 (0.018)	3.364
S_2	6.236 (0.084)	6.274 (0.122)	6.298 (0.146)	6.159 (0.007)	6.161 (0.009)	6.152
S_3	6.552 (0.076)	6.593 (0.117)	6.619 (0.143)	6.473 (-0.003)	6.478 (0.002)	6.476
S_4	6.601 (0.087)	6.643 (0.129)	6.673 (0.159)	6.523 (0.009)	6.520 (0.006)	6.514
S_5	6.878 (0.130)	6.919 (0.171)	6.983 (0.235)	6.800 (0.052)	6.780 (0.032)	6.748

Table C.5: Iteration time for ground state energy of ethyl-butyl-diazene in minutes. For MLCC3 and CC3 we report time in the triples part. Speedup factors are given in parenthesis. The number of active orbitals in each model is given below.

	CCSD	**A**	**B**	**A**+	**B**+	CC3
CPU	60	16	43	20	50	1947
Wall	-	2.4 (×103)	6.1 (×40)	2.7 (×91)	6.6 (×37)	246
O_A	30	9	17	9	17	32
V_A	278	44	124	107	138	278

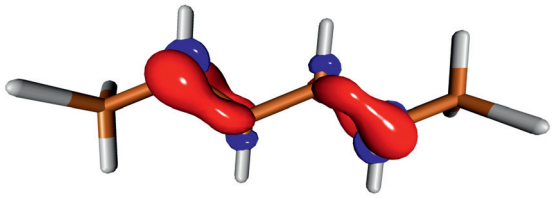
Fig. C.6. The selections of active space is based on the density plots in Fig. C.5. Six virtual orbitals are added to the active space using the same scheme as for ethyl-butyl-azene. This is a conjugated system which often requires triple excitations to accurately describe the excited states [157].

Our results are given in Table C.6 and iteration times are given in Table C.7. The MLCC3 model is about eight times faster than full CC3. In this case CCSD performs surprisingly well for the three lowest states with errors less than 0.1 eV. For these states the MLCC3 errors are reduced to less than 0.01 eV. For S_4 and S_5, the errors are larger for both CCSD and MLCC3, but MLCC3 improves on the CCSD numbers. Both models perform better for S_6 and S_7 and especially MLCC3 has small errors.

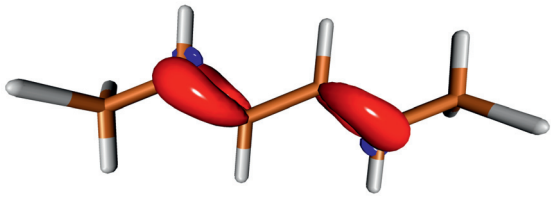
Table C.6: Excitation energies calculated for hexadiene, aug-cc-pVDZ. Error compared to Full CC3 in parenthesis.

	CCSD	MLCC3	CC3
S_1	5.664 (0.040)	5.620 (-0.004)	5.624
S_2	5.857 (0.040)	5.810 (-0.007)	5.817
S_3	6.128 (0.060)	6.058 (-0.010)	6.068
S_4	6.325 (0.137)	6.284 (0.096)	6.188
S_5	6.560 (0.225)	6.481 (0.146)	6.335
S_6	6.588 (0.039)	6.529 (-0.020)	6.549
S_7	6.740 (0.055)	6.701 (0.016)	6.685

The density plots in Fig. C.5 do not explain the difference in errors between S_1 and S_5. However, the reduced accuracy can be explained from analyzing the excitation vectors. The S_4 and S_5 states have the lowest internal character, in particular the internal single excitation components of the vectors is 69% and 67%, respectively. This could suggest the need for a larger space for these excitations. We note that MLCC3 performs better than CCSD for all states and provides the largest correction for the states with the largest errors in CCSD. For small active spaces, the MLCC3 method can be used to estimate the importance of triple excitations.



S_1



S_5

Figure C.5: Change in electron density between ground state and the first and fifth singlet excited states of hexadiene calculated with CCSD. Volumes in blue have increased density and volumes in red have decreased density. Isovalue: 0.01

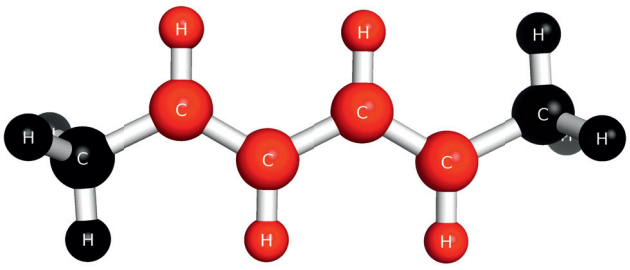


Figure C.6: Subsystem of hexadiene. Atoms in red are active.

Table C.7: Iteration time for ground state energy of hexadiene in minutes. For MLCC3 and CC3 we report time in the triples part. The number of active orbitals in each model is given below.

	CCSD	MLCC3	CC3
CPU	16	27	241
Wall	-	3.7 (8.1)	30
O_A	23	17	23
V_A	205	130	205

C.5 Conclusion

In this paper we have described an efficient implementation of the multilevel coupled cluster model MLCC3. The implemented algorithm utilizes permutation symmetry to reduce the cost of calculating the T_3 amplitudes by a factor of six. The code is parallelized using OpenMP and in future versions we will explore parallelization using MPI as well.

Using the multilevel approach we can obtain CC3 accuracy at a fraction of the cost. The user is required to select an active space relevant for the considered excitations. Maps of changes in electron density between ground and excited states, obtained from CCSD or a lower scaling method, can assist in selecting the active atoms. The active orbital space should be augmented with low energy virtual orbitals when using Cholesky orbitals and diffuse functions in the orbital basis. We will in the future explore other approaches to generate active orbitals. Analysis of the excitation vectors should be used to verify the quality of the results. The examples presented here suggest that an internal character of 80% or more is sufficient.

In general, CCSD with diffuse basis sets are very accurate [157], particularly for properties like oscillator strengths. For this reason, we have not yet implemented oscillator strengths in the current code. For core-valence excitations, CCSD shows larger discrepancies from CC3 and applications in this area are in progress.

Finally, we point out that the MLCC3 model can be generalized to include several levels of theory. In principle, any number of spaces can be added and for very large molecular systems, Hartree-Fock or even frozen densities may be used. This would essentially provide size intensive CC models.

Acknowledgments

We thank T. J. Martinez for hosting part of this project at Stanford University and acknowledge computer time from NOTUR through project nn2962k. H.K.

acknowledges financial support from the FP7-PEOPLE-2013-IOF funding scheme (Project No. 625321) and R.H.M. acknowledges financial support from the COST action "XUV/X-ray light and fast ions for ultrafast chemistry (XLIC)".

Paper D

Near-Edge X-ray Absorption Fine Structure within Multilevel Coupled Cluster Theory

J. Chem. Theory Comput. **12**, 2633-2643 (2016)

Near-Edge X-ray Absorption Fine Structure within Multilevel Coupled Cluster Theory

Rolf H. Myhre[1,2], Sonia Coriani[3,4] and Henrik Koch[1,2*]

[1]*Department of Chemistry, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway*
[2]*Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California 94305, USA*
[3]*Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Trieste, I-34127, Italy*
[4]*Aarhus Institute of Avanced Studies, University of Aarhus, DK-8000 Århus C, Denmark*

Abstract

Core excited states are challenging to calculate mainly because they are embedded in a manifold of high energy valence-excited states. Their locality, however, makes their determination ideal for local correlation methods. In this paper, we demonstrate the performance of multilevel coupled cluster theory in computing core spectra both within the core-valence separated and the asymmetric Lanczos implementations of coupled cluster linear response theory. We also propose a visualization tool to analyze the excitations using the difference between the ground and excited state electron densities.

D.1 Introduction

X-ray absorption spectroscopy (XAS) is a fundamental spectroscopic method for determining the electronic and structural properties of molecules, as well as their dynamic behavior [107–113]. In this type of spectroscopy, core electrons are excited, leaving behind a core hole. Since core orbitals are highly local, probing them yields important information about their local environment in the molecular system. An essential component for disclosing such information is the availability of computational methods to model the experimental spectra.

The calculation of core excitations using electronic structure theory encounters specific challenges not present for valence excitations. At the typical energies involved in X-ray spectroscopy, there is a high density of excited states with more loosely bounded electrons. These states form a continuum and the challenge is to obtain the core excited state within this continuum of states. Electronic structure programs usually employ subspace algorithms such as the Davidson algorithm [42, 43] to determine eigenvalues. These methods are biased towards the lowest excitation energies, which makes their common implementations impractical for finding the high-energy excited states in question.

Due to the significant reduction in the screening of the nucleus following core excitation, relaxation effects play an essential role and need to be accounted for in a reliable manner. These relaxation effects mainly affect the molecular system in two ways: a direct, attractive effect contracting the valence electron density and an indirect repulsive effect from the interaction between the excited electron and the valence electrons which increases polarization. Additional issues that must be considered include the choice of basis set, the effect of relativity, vibrational effects and spectral broadening schemes.

Despite such difficulties, many methods have been presented to compute core spectra. More than forty years ago, Slater proposed the transition state method [234, 235]. In this method, the molecular orbitals involved in the excitations are set to half occupation and the excitation energies are calculated as the difference between orbital energies. A somewhat similar method is the transition potential method where excitation energies are calculated from the derivative of the total energy with respect to the occupation number when the occupations are set to one half in the relevant orbitals [236, 237]. Another early approach is the multiple scattering or KKR [238–240] method. It describes electron propagation with a reference Green's function and a series of scattering events.

Density functional theory (DFT) based methods are the most common methods used to obtain core excitations and we will mention some of them here. One approach employs pseudopotentials, commonly used to describe core electrons and incorporate relativistic effects in DFT methods [241]. Pseudopotentials can be extended to include core holes, allowing the calculation of high energy excitations [242]. In the DFT variant of the ΔSCF method [243], the electron density corre-

sponding to Kohn-Sham orbitals is subtracted and added in each step of the self consistent iteration. Core excitations are then obtained by subtracting the density corresponding to a core hole [244]. Other methods are based on time-dependent DFT (TDDFT). Examples are the complex propagator approach (CPP) of Norman and coworkers [245, 246], and the TDDFT method with a restricted excitation manifold [247–249]. In the CPP approach [245, 246], the absorption cross section of the core excitation process is obtained directly from scanning the imaginary part of the complex dipole polarizability over the relevant frequency region. The restricted excitation manifold methods are similar in spirit to the core-valence separation (CVS) technique [114].

Core-valence separation is used in most wave-function-based methods for calculating X-ray excitations. Due to the large energy difference between the valence and core excited states, core-valence interaction parts of the Hamiltonian are very small and can be neglected [114]. The CVS approximation was first implemented within the second-order algebraic diagrammatic construction ADC(2) [52] method. Wenzel *et al.* [115, 250–252] recently proposed it for the ADC(2), ADC(2)-x and ADC(3) hierarchy of methods based on the intermediate state representation variant of the ADC formalism [53, 253]. Calculating excited states in ADC requires solving an eigenvalue problem for a secular matrix [53, 253]. The core-valence separation greatly simplifies the eigenvalue problem by making the core excitations extremal eigenvalues and significantly reducing the size of the vector space.

The calculation of X-ray absorption spectra has also been implemented within coupled cluster (CC) theory [45, 46, 116, 254–258]. As in ADC, computing excited states in CC theory requires solving an eigenvalue problem, specifically the eigenvalues of the Jacobian. To obtain the eigenvalues corresponding to core excited states, an asymmetric Lanczos algorithm was used in Refs. 45 and 46 to construct a truncated tridiagonal representation of the Jacobian matrix. Diagonalization of this matrix makes it possible to obtain a full spectrum which includes core excitations.

It is well known that the Lanczos algorithm is numerically unstable and may require explicit biorthogonalization [259]. This makes it necessary to store and read a large number of vectors from disk, which becomes prohibitive for systems with more than 300 orbitals. For this reason, the CVS approximation has recently been implemented for CC linear response (CCLR) theory both within the Lanczos algorithm, and for conventional CCLR [116]. Alternative algorithms targeting X-ray excitations within the CC formalism are the CC-CPP approach [257], and the energy-specific EOM-CC approach of Peng *et al.* [258] With the CVS approximation, the cost of calculating core excited states is approximately the same as the calculation of valence excited states and the bottleneck is the steep scaling of the CC methods with the dimension of the molecular system.

The locality of core excitations makes their calculation suitable for local methods such as the multilevel coupled cluster (MLCC) approach [174, 209]. The MLCC method treats a small part of the molecular system with a high accuracy CC

method and the rest of the system more approximately [174, 209]. In this paper we test the capability of the multilevel CCSD (MLCCSD) approach to reproduce full CCSD Near Edge X-ray Absorption Fine Structure (NEXAFS) spectra. We compare spectra obtained both with the full space or regular Lanczos algorithm and with the Davidson algorithm with the CVS approximation.

In order to visualize and investigate the local nature of the excitations, ground and excited state one-electron densities are calculated. The difference in the electron densities is then plotted and a visualization of the excitation process is obtained. This is similar to the visualization technique recently introduced by Wenzel and Dreuw [260]

The paper is organized as follows. In the next section we give a brief outline of how to compute core spectra within the MLCC approach. Section D.3 presents the results of some example calculations. The last section contains our concluding remarks, including a comparison to experiment for ethanal.

D.2 Theory

D.2.1 Computing spectra within CC linear response theory

The CC wave function ansatz for a closed-shell system is defined by the exponential parametrization

$$|\mathrm{CC}\rangle = \exp(X)\,|\mathrm{HF}\rangle \tag{D.1}$$

where $|\mathrm{HF}\rangle$ is the Hartree-Fock reference state and $X = \sum_\mu x_\mu \tau_\mu$ is the cluster operator with the cluster amplitudes x_μ and the corresponding excitation operators τ_μ. Note that in conventional CC theory the cluster operator and amplitudes are usually written as T and t_μ, respectively. Here we use a slightly different notation because the symbols T and t_μ are reserved for the active space cluster operator and amplitudes in MLCC theory [174, 209], see also Section D.2.3. The ground state energy and amplitudes are determined by projection of the Schrödinger equation on the reference state and a manifold of excitations

$$E = \langle \mathrm{HF}|\exp(-X)H\exp(X)\,|\mathrm{HF}\rangle \tag{D.2}$$

$$\Omega_\mu = \langle \mu|\exp(-X)H\exp(X)\,|\mathrm{HF}\rangle = 0\ . \tag{D.3}$$

In CC linear response theory, excitation energies, ω_k, and left, $\mathbf{L}_k$, and right, $\mathbf{R}_k$, excitation vectors are obtained solving the asymmetric eigenvalue equations

$$\mathbf{A}\mathbf{R}_k = \omega_k \mathbf{R}_k; \quad \mathbf{L}_k\mathbf{A} = \omega_k \mathbf{L}_k \tag{D.4}$$

with the biorthogonality condition $\mathbf{L}_i\mathbf{R}_k = \delta_{ik}$. The Jacobian matrix $\mathbf{A}$ in Eq. (D.4) is defined as the derivative

$$A_{\mu\nu} = \frac{\partial \Omega_\mu}{\partial x_\nu} = \langle \mu | \exp(-X)[H, \tau_\nu] \exp(X) | \mathrm{HF} \rangle \tag{D.5}$$

Transition strengths for dipole components A and B are determined from the single residues of the linear response function, and take the form

$$S^{AB}_{0\to j} = \frac{1}{2} \left\{ \mathcal{M}^A_{0j} \mathcal{M}^B_{j0} + (\mathcal{M}^B_{0j} \mathcal{M}^A_{j0})^* \right\} \tag{D.6}$$

where the left and right transition moments are given by

$$\mathcal{M}^A_{0j} = \boldsymbol{\eta}^A \mathbf{R}_j + \bar{\mathbf{M}}^j(\omega_j) \boldsymbol{\xi}^A; \quad \mathcal{M}^B_{j0} = \mathbf{L}_j \boldsymbol{\xi}^B \tag{D.7}$$

and the auxiliary Lagrangian multipliers $\bar{\mathbf{M}}^j(\omega_j)$ are obtained from the solution of the linear equation

$$\bar{\mathbf{M}}^j \left(\mathbf{A} + \omega_j \mathbf{I} \right) = -\mathbf{F}\mathbf{R}_j. \tag{D.8}$$

We refer to Refs. [31] and [159] for the definition of the remaining building blocks.

Equation (D.4) is most often solved iteratively via some generalization of the Davidson algorithm [42, 43]. The iterative procedure is initiated by selecting unit vectors corresponding to specific occupied to virtual orbital excitations, often based on Hartree-Fock orbital energy differences. The procedure will converge towards the lowest eigenvalues and eigenvectors even if the initial start vectors correspond to high energy excitations. This makes the procedure ill-suited for core excitations.

Another way to solve Eq. (D.4) is to build a tridiagonal representation, $\mathbf{T}$, of the Jacobian matrix $\mathbf{A}$ using an asymmetric Lanczos algorithm. The nonzero elements of the tridiagonal matrix $\mathbf{T} = \mathbf{P}^T\mathbf{A}\mathbf{Q}$ are given by

$$T_{ll} = \alpha_l = \mathbf{p}_l^T \mathbf{A} \mathbf{q}_l \tag{D.9}$$

$$T_{l+1,l} = \beta_l = \sqrt{\mathbf{p}_{l+l}^T \mathbf{q}_{l+1}} \tag{D.10}$$

$$T_{l,l+1} = \gamma_l = \mathrm{sgn}\{\mathbf{p}_{l+1}^T \mathbf{q}_{l+1}\} \beta_l \tag{D.11}$$

with the biorthogonal $\mathbf{p}_l$ and $\mathbf{q}_l$ vectors given by

$$\mathbf{q}_{l+1} = \beta_l^{-1} (\mathbf{A}\mathbf{q}_l - \gamma_{l-1}\mathbf{q}_{l-1} - \alpha_l \mathbf{q}_l) \tag{D.12}$$

$$\mathbf{p}_{l+1}^T = \gamma_l^{-1} (\mathbf{p}_l^T \mathbf{A} - \beta_{l-1}\mathbf{p}_{l-1}^T - \alpha_l \mathbf{p}_l^T). \tag{D.13}$$

It is neither necessary nor convenient to generate the full tridiagonal matrix $\mathbf{T}$, and the procedure can be truncated at some dimension $J \ll n$ where n is the dimension of the full Jacobian. The diagonalization of $\mathbf{T}$ produces an effective spectrum that covers the entire frequency range and converges from the top and bottom with increasing Lanczos chain length J [44, 46, 261].

A convenient choice of start vectors for the Lanczos algorithm is

$$\mathbf{q}_1 = u_A^{-1}\boldsymbol{\xi}^A = \frac{\boldsymbol{\xi}^A}{||\boldsymbol{\xi}^A||}; \qquad \mathbf{p}_1^T = v_A^{-1}\boldsymbol{\eta}^A = \frac{||\boldsymbol{\xi}^A||}{\boldsymbol{\eta}^A\boldsymbol{\xi}^A}\boldsymbol{\eta}^A. \tag{D.14}$$

With this choice, the eigenvectors $\mathbf{L}$ and $\mathbf{R}$ of $\mathbf{T}$ form the basis for an approximate diagonal representation of the (complex) linear response function [45, 46]. The absorption cross-section can then be computed from its imaginary component [45, 46]. Moreover, one may compute directly the transition strengths as

$$S_{0\to j}^{AA} = u_A v_A L_{j1} R_{1j} - v_A^2 \sum_l \frac{L_{j1} L_{l1} \mathcal{F}_{lj}}{(\omega_j + \omega_l)} \quad , \tag{D.15}$$

see Refs. 45 and 46 for further definitions and details.

D.2.2 The core-valence separation

The CVS approximation can be implemented within both the Davidson and the asymmetric Lanczos algorithms [116]. Initially, a set of one or more core orbitals is selected. In each iteration of the solver, a projector $\mathcal{P}^v$ is applied on the trial vectors, removing all elements not referencing at least one of the selected core orbitals. For a singles and doubles trial vector, $\mathbf{b}$, the effect of the projector is

$$\begin{cases} \mathcal{P}^v b_i^a = 0 & \forall\ i = \text{valence}, \\ \mathcal{P}^v b_{ij}^{ab} = 0 & \forall\ i,j = \text{valence} \end{cases} \tag{D.16}$$

For the Davidson algorithm, Eq. (D.4) becomes the projected eigenvalue equation

$$\mathcal{P}^v(\mathbf{A}\mathcal{P}^v\mathbf{R}_k) = \omega_k \mathcal{P}^v \mathbf{R}_k, \tag{D.17}$$

and similarly for the left eigenvectors. By applying the projector in each iteration during the solution of Eq. (D.8), the computation of CVS-CC transition moments and transition strengths is also easily obtained.

Within the Lanczos algorithm, the projector is applied during the iterative construction of the $\mathbf{T}$ matrix, i.e. to the $\mathbf{p}_l^T$ and $\mathbf{q}_l$ vectors and their linear transformations, $\mathbf{p}_l^T\mathbf{A}$ and $\mathbf{A}\mathbf{q}_l$. The resulting Lanczos eigenvectors, as well as the Lanczos trial vector bases, $\mathbf{P}^T$ and $\mathbf{Q}$, only contain excitations involving at least one core orbital. This effectively decouples them from excitations with contributions from occupied valence orbitals only. Diagonalization of the tridiagonal matrix yields the core excitations as lowest roots and quickly converges to the exact results with a significantly smaller Lanczos chain lengths. The oscillator strengths and cross sections are obtained without further modifications to the general procedure.

D.2.3 Multilevel coupled cluster theory for core spectra

Multilevel coupled cluster theory divides the molecular system into an active and an inactive part. By treating the active part with a highly accurate method and the inactive part more approximately, high accuracy results can be achieved at greatly reduced cost [174, 209]. For example, the most expensive term in CCSD [21] scales as V^4O^2 where V and O are the number of virtual and occupied orbitals respectively. By dividing the system into an active CCSD part and an inactive part described by CC2 [23], one obtains the MLCCSD model. In this model, the scaling of the most expensive CCSD term is reduced to $V^2V_A^2O_A^2$ where V_A and O_A are the number of active virtual and occupied orbitals respectively. For some of the other CCSD terms arising from the double commutator, see Eq. (D.19) below, the scaling reduction is less favorable, but no terms scale worse than V^2O^2. This is less than the scaling of CC2 which is V^3O^2, so the overall scaling will be that of CC2 for large systems.

To divide the system, a localized set of molecular orbitals (MO) is generated and each orbital is assigned to an atom. Orbitals corresponding to active atoms then form an active orbital space while those corresponding to inactive atoms form an inactive space. We require that the orbitals are orthogonal and that the blocks of the Fock matrix corresponding to each space are diagonal. In this work, we have used Cholesky orbitals [88], but other schemes are possible. After generating the orbital spaces, the cluster operator is divided into terms corresponding to each space, $X = T + S$, where T corresponds to the active space and S to the inactive. The approach can be generalized to several spaces.

The amplitude equations, Eq. D.3, become

$$\langle \mu_1 | \, \hat{H} + [\hat{H}, X_2] \, | \mathrm{HF} \rangle = 0 \tag{D.18}$$

$$\langle \mu_2^T | \, \hat{H} + [\hat{H}, X_2] + \frac{1}{2}[[\hat{H}, X_2], X_2] \, | \mathrm{HF} \rangle = 0 \tag{D.19}$$

$$\langle \mu_2^S | \, [F, S_2] + \hat{H} + [\hat{H}, T_2] \, | \mathrm{HF} \rangle = 0. \tag{D.20}$$

In Eqs. (D.18)-(D.20), $\hat{H}$ refers to the X_1-transformed Hamiltonian operator

$$\hat{H} = \exp(-X_1) H \exp(X_1), \tag{D.21}$$

and μ^T and μ^S denote the active and inactive excited state manifolds, respectively. Note that the equations for the single amplitudes, Eq. (D.18), and active double amplitudes, Eq. (D.19), are the same as for full CCSD, while the equations for the inactive amplitudes, Eq. (D.20), are similar to those of CC2, but contain some extra terms involving the T-operator to give a balanced description. Excitations only involving active orbitals are referred to as internal while those only involving inactive orbitals are referred to as external. Excitations between the spaces are called semi-external.

When calculating MLCCSD excitation energies, the equations are similar to those of full CCSD, however, the Jacobian matrix **A** is modified. The changes in the **A**-matrix blocks are schematically summarized in Eq. (D.22) below

$$\mathbf{A}^{\mathrm{MLCCSD}} = \begin{array}{c} \\ T_1 \\ T_2 \\ S_2 \end{array} \begin{array}{c} \begin{array}{ccc} T_1 & T_2 & S_2 \end{array} \\ \begin{pmatrix} \mathrm{CCSD} & \mathrm{CCSD} & \mathrm{CCSD} \\ \mathrm{CCSD} & \mathrm{CCSD} & \mathrm{CCSD} \\ \mathrm{CCX} & \mathrm{CCX} & \mathrm{CC2} \end{pmatrix} \end{array} \tag{D.22}$$

The first two block rows, projecting against the singles and active doubles, are the same as for standard CCSD, while the lower right block is the same as in CC2 and is diagonal. Ideally, the CC2 block will have the largest dimension, resulting in a large reduction in computational cost compared to CCSD. The blocks labeled CCX in Eq. (D.22) contain hybrid terms between CC2 and CCSD that scales as CC2. For more details on MLCCSD linear response, see Ref. 174.

Similarly, the intensities can be found using either Eqs. (D.6)-(D.7) or Eq. (D.15), both involving the use of a MLCC-modified **F** matrix [262]. The **F** matrix can not be as easily divided into blocks corresponding to the active and inactive space, but contains a mixture of terms that scale as CC2.

D.3 Results

D.3.1 General computational details

The MLCCSD approach for core excitation spectra has been implemented in a development version of the Dalton code [99, 100]. As the current MLCCSD implementation is not optimized, our study has been limited to the relatively small molecules ethanal (acetaldehyde), propenal (acrolein) and butanal (butyraldehyde). Core absorption spectra have been computed using MLCCSD and compared to the corresponding full CCSD and CC2 spectra. As core-valence polarized basis sets are usually required for an accurate treatment of core excitations, we use aug-cc-pCVDZ basis set [59] for the atoms of the specific edges in ethanal and butanal, and either the aug-cc-pVDZ, indicated by aug-cc-p(C)VDZ, or the aug-cc-pCVDZ basis for the remaining atoms. Our results indicate that the difference in the results between these basis sets is negligible. Due to the high computational cost associated with the regular Lanczos algorithm, we only used aug-cc-pVDZ in the case of propenal.

The active spaces used are described for each case in the following subsections. We use the standard IUPAC numbering when discussing specific atoms, e.g. carbon 1 (C1) will always refer to the carbon closest to the oxygen atom, carbon 2 (C2) to the next carbon in the chain, and so on. Spectra are plotted using calculated

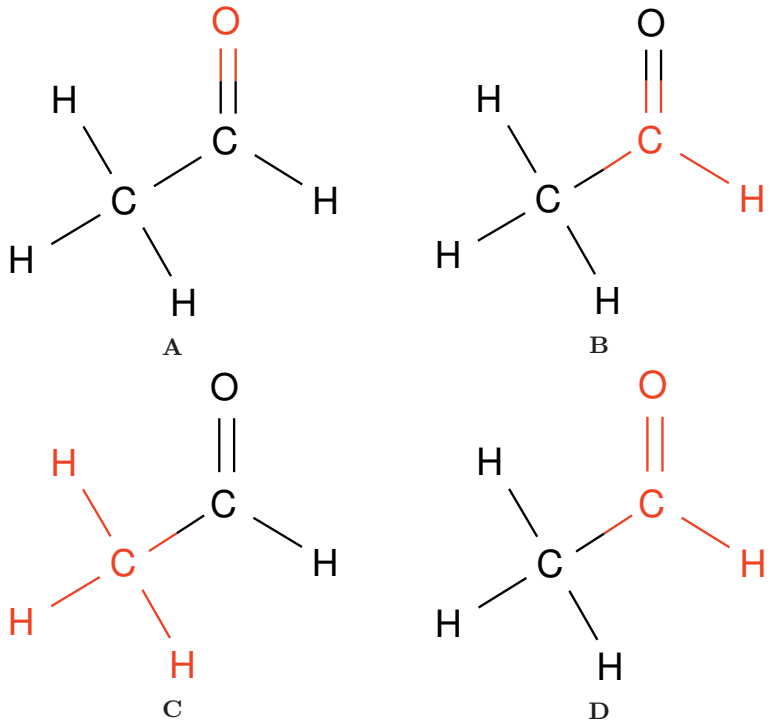


Figure D.1: Active spaces of ethanal indicated in red.

excitation energies and oscillator strengths in the length gauge with an empirical line broadening of 1000 cm^{-1} and a normalized Lorentzian line shape function.

D.3.2 Ethanal

In Figure D.1, the active atoms of ethanal are colored red. Spaces **A**, **B** and **C** only contain one second row atom each, and they are the smallest active spaces possible. Our results indicate that this is too small for some of the excitations we consider. Consequently we have also used space **D** which is the union of **A** and **B**.

The oxygen K-edge spectra calculated using the regular Lanczos algorithm with a chain length of 2000 are presented in Figure D.2. The plots show the excitation energies and intensities calculated using CC2, CCSD, and MLCCSD with spaces **A** and **D**, employing aug-cc-pCVDZ for all models except MLCCSD **A**, where we used aug-cc-p(C)VDZ instead. For the main edge, both MLCCSD models agree almost perfectly with full CCSD. Model **A** is not able to reproduce the higher energy fine structure, while **D** is reasonably close. Both MLCCSD models, unlike CC2, reproduce the large gap between the main edge and the fine structure.

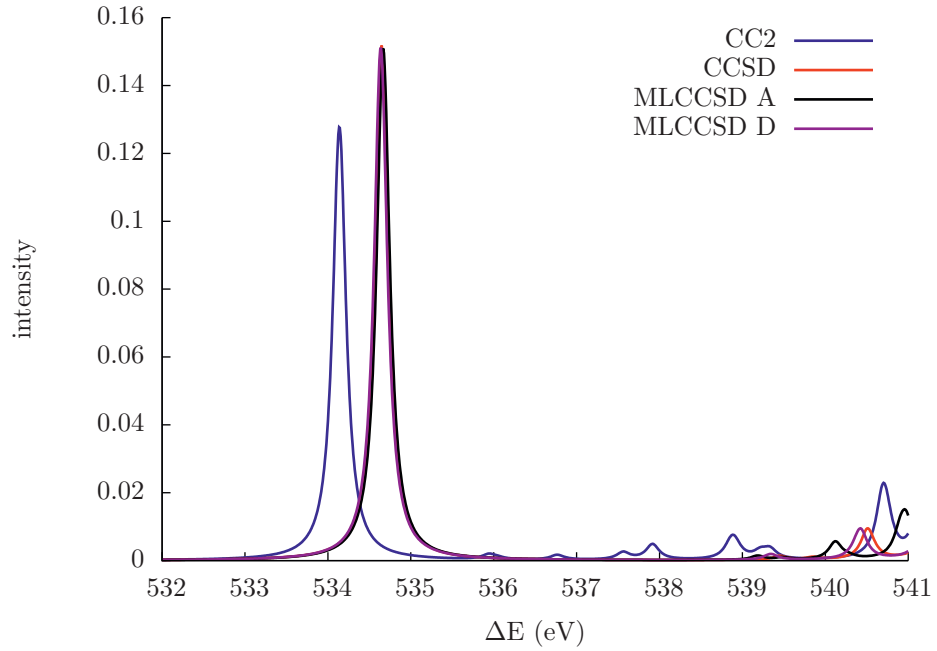


Figure D.2: Ethanal. Comparison of the oxygen K-edge spectra as obtained in CC2, CCSD and MLCCSD using the regular Lanczos algorithm. Basis sets: aug-cc-p(C)VDZ for MLCCSD **A** and aug-cc-pCVDZ for the rest.

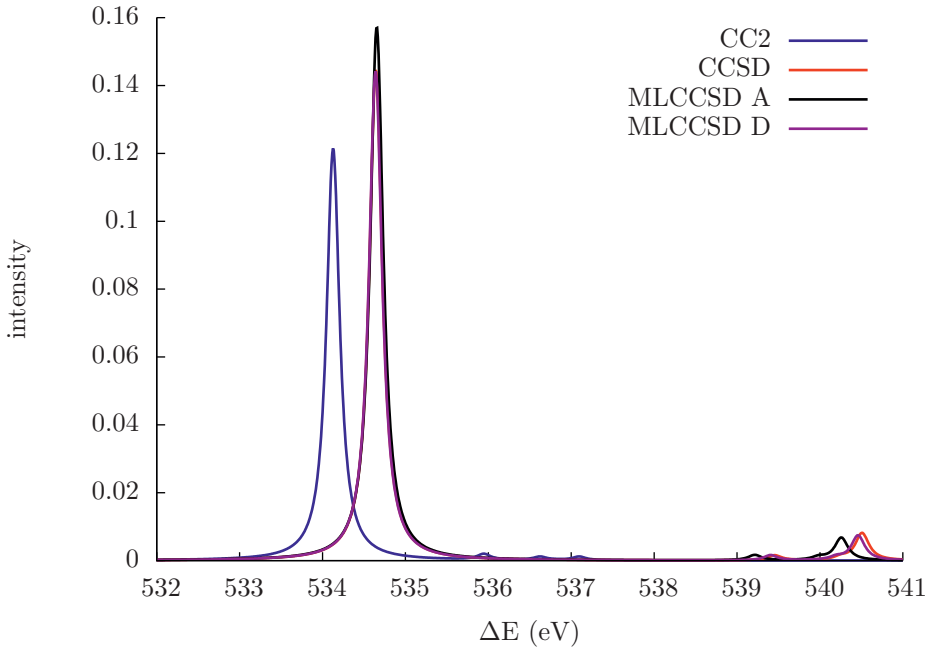


Figure D.3: Ethanal. Comparison of the oxygen K-edge spectra as obtained in CC2, CCSD and MLCCSD using both the Davidson algorithm with the CVS approximation. Basis sets: aug-cc-p(C)VDZ for MLCCSD **A** and aug-cc-pCVDZ for the rest.

Table D.1: Ethanal. Oxygen core excitation energies (in eV) and corresponding oscillator strengths, $f \times 100$ (dimensionless) computed using the CVS-Davidson algorithm. Basis sets: aug-cc-p(C)VDZ for MLCCSD **A** and aug-cc-pCVDZ for the rest.

	CC2		CCSD		MLCCSD **A**		MLCCSD **D**	
	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
S_1	534.13	3.64	534.64	4.34	534.65	4.72	534.64	4.32
S_2	535.95	0.05	539.46	0.04	539.21	0.05	539.40	0.05
S_3	536.62	0.03	540.25	0.03	539.99	0.03	540.20	0.03
S_4	536.83	0.00	540.50	0.22	540.26	0.18	540.46	0.20
S_5	537.10	0.03	540.56	0.02	540.28	0.02	540.49	0.02

Figure D.3 contains the equivalent spectra calculated using the CVS approximation within the standard implementation of CCLR. Note that only five excitations were calculated using each model. As a result, the spectra presented are not complete in the given energy range. This is particularly important for CC2 because the highest eigenvalue computed was at 537.10 eV. The excitation energies are collected in Table D.1. Similar to the Lanczos case, the MLCCSD results are very good for the main edge and reasonably good for the fine structure. Unlike in the Lanczos case, model **A** overestimates the oscillator strength of the main edge by about 10%.

The carbon K-edge spectra obtained using the regular Lanczos algorithm are plotted in Figure D.4. The main edge is well reproduced by MLCCSD **D**. For the higher excitations, however, MLCCSD seems to be closer to CC2. In should be noted that the Lanczos algorithm calculates all roots in this case, so the bands in Figure D.4 originate from both the active and inactive spaces. Furthermore, it is less straightforward to determine which excitations correspond to which individual core orbitals when using the Lanczos method without CVS.

These complications are lifted when applying, both in the Davidson and in Lanczos case, the CVS approximation as shown in Figure D.5. Note that the CC2 results are not included in the figure for clarity, but the values are reported in Table D.2 and they are very similar to the Lanczos CC2 results. As the core orbitals are selected beforehand, it is possible to choose an active space corresponding to the relevant atom or atoms. From Figure D.5, we observe that MLCCSD **C** reproduces the carbon excitations of the methyl group very well. Model **B** has a slightly shifted main edge, whereas **D** is almost indistinguishable from full CCSD.

Surprisingly, model **B** reproduces the peak just below 293 eV better than model **D**. Similar behavior is observed for the other molecules in this study. At these higher energies, the excitations have a higher double character and are more delocalized so the MLCCSD method is expected to be less successful. For this reason we believe the agreement with model **B** is a result of fortuitous error cancellation. Some of this may be due to the localized orbitals. Full CCSD is orbital invariant as long as there is no mixing of occupied and virtual orbitals. This invariance is broken by

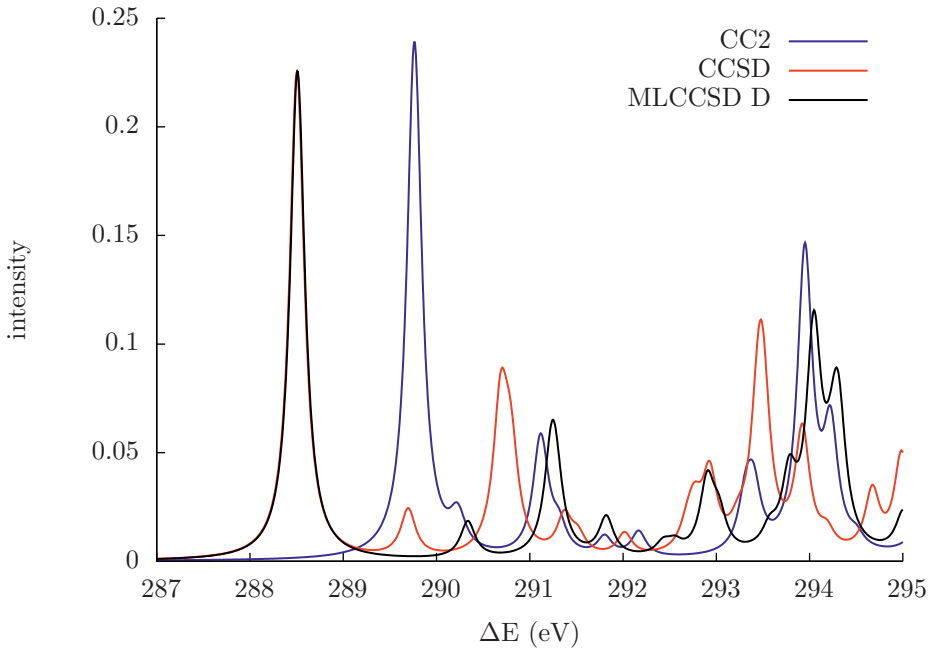


Figure D.4: Ethanal. Comparison of the carbon K-edge spectrum as obtained in CC2, CCSD and MLCCSD via the regular Lanczos algorithm. Basis set aug-cc-pCVDZ.

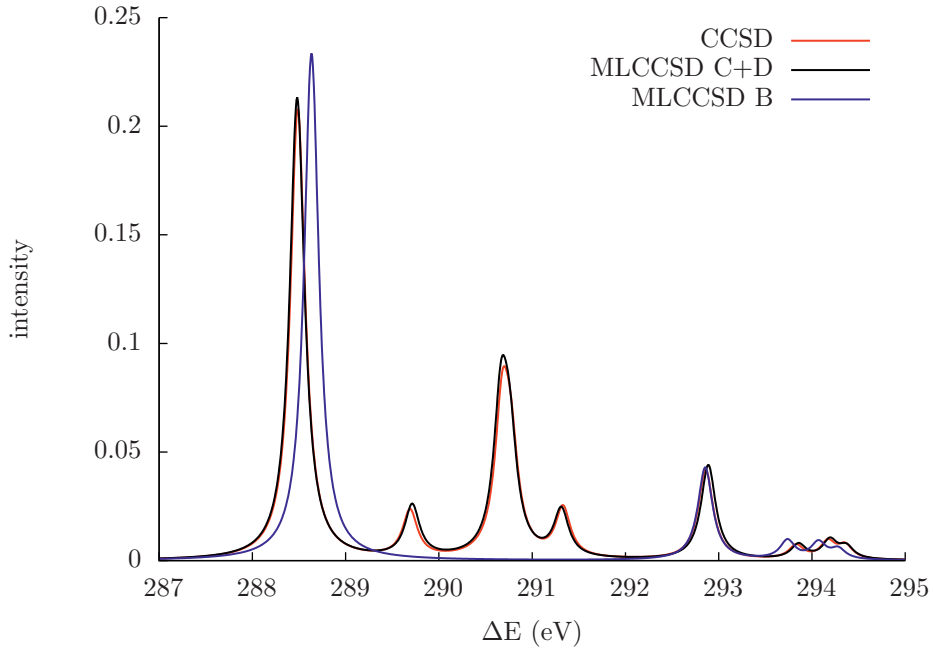


Figure D.5: Ethanal. Comparison of the carbon K-edge spectrum as obtained in CC2, CCSD and MLCCSD using the CVS-Davidson algorithm. Basis sets: aug-cc-p(C)VDZ for MLCCSD **B** and aug-cc-pCVDZ for the rest.

the CVS approximation and the Cholesky 1s orbital energy is about 0.04 eV higher than the canonical orbital energy in active space **B** while only 0.01 for active space **D**. These are small differences, but the effect is generally greater for smaller active spaces and basis sets.

To visualize the effect of the excitation process, we have plotted the change in the one-electron density between the ground and core excited states. This is achieved by calculating the orbital occupations in the two states and calculating the CCSD density difference using Molden [215]. In Figure D.6 the density difference in the molecular plane is plotted for the most intense peaks of each atom in ethanal. Immediately noticeable is the very large reduction in the core density of the atom being excited and a corresponding increase in the density around the core. Furthermore, the oxygen excitation is highly localized with almost all the change taking place on the oxygen atom, and only a small change on C1. The carbon excitations are also fairly localized, but less so than for oxygen, particularly for C1.

Figure D.7 shows the same excitations plotted as 3D isosurfaces. Blue surfaces correspond to increased densities and red to decreased densities. This reveals that most of the increase goes out of the molecular plane for C1 and oxygen. Analysis of the excitation vector reveals that the excitations are indeed 1s $\rightarrow \pi^*$ excitations. The excitation from C2 is also a 1s $\rightarrow \pi^*$ excitation, but it is less clear from the density change. The greater localization for oxygen and C2 can explain why MLCCSD **A** and **C** perform better for these atoms than MLCCSD **B** does for C1.

D.3.3 Propenal

Propenal is the smallest conjugated aldehyde and we will use it to study the performance of MLCCSD for a conjugated system. The active spaces adopted are summarized in Figure D.8. Spaces **A** to **D** only contain one second row element each, while **E** and **F** contain a double bond each.

The CVS Davidson results for oxygen are presented in Figure D.9 and Table D.3. As for ethanal, both spaces containing oxygen reproduce the main edge well, and **E** is better for the fine structure. Again, CC2 fails and yields too small a gap between the main edge and the next excitation.

Obtaining the carbon K-edge spectrum in propenal is more complicated than for ethanal because there are more carbon atoms. Combining the regular Lanczos algorithm with MLCCSD would produce a large number of peaks corresponding to inactive atoms and make the spectrum very difficult to interpret. In Figure D.10 we compare the CCSD and CC2 spectra, obtained with the CVS-Davidson algorithm, to the CCSD regular Lanczos spectrum. While CC2 performs better for carbon than for oxygen, there are still large discrepancies compared to CCSD. Comparing the Lanczos and CVS algorithms, we generally find a good agreement below 292 eV, though with some small differences. These differences may be due

Table D.2: Ethanal. Carbon core excitation energies (eV) and corresponding oscillator strengths $f \times 100$ (dimensionless) computed using the CVS-Davidson algorithm. Basis sets: aug-cc-p(C)VDZ for MLCCSD **B** and aug-cc-pCVDZ for the rest.

		CC2		CCSD		MLCCSD **C**+**D**		MLCCSD **B**	
		ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
C1	S_1	289.74	6.63	288.49	6.23	288.48	6.38	288.63	7.00
	S_2	293.48	0.84	292.86	1.24	292.89	1.31	292.86	1.28
	S_3	294.23	0.21	293.83	0.19	293.86	0.20	293.74	0.26
	S_4	294.54	0.14	294.17	0.24	294.19	0.26	294.07	0.23
	S_5	294.76	0.12	294.36	0.19	294.37	0.17	294.28	0.14
C2	S_1	290.15	0.01	289.55	0.00	289.52	0.00	-	-
	S_2	290.21	0.49	289.69	0.64	289.71	0.72	-	-
	S_3	291.10	1.23	290.68	1.95	290.67	2.03	-	-
	S_4	291.12	0.57	290.77	1.22	290.76	1.34	-	-
	S_5	291.66	0.52	291.33	0.68	291.31	0.65	-	-

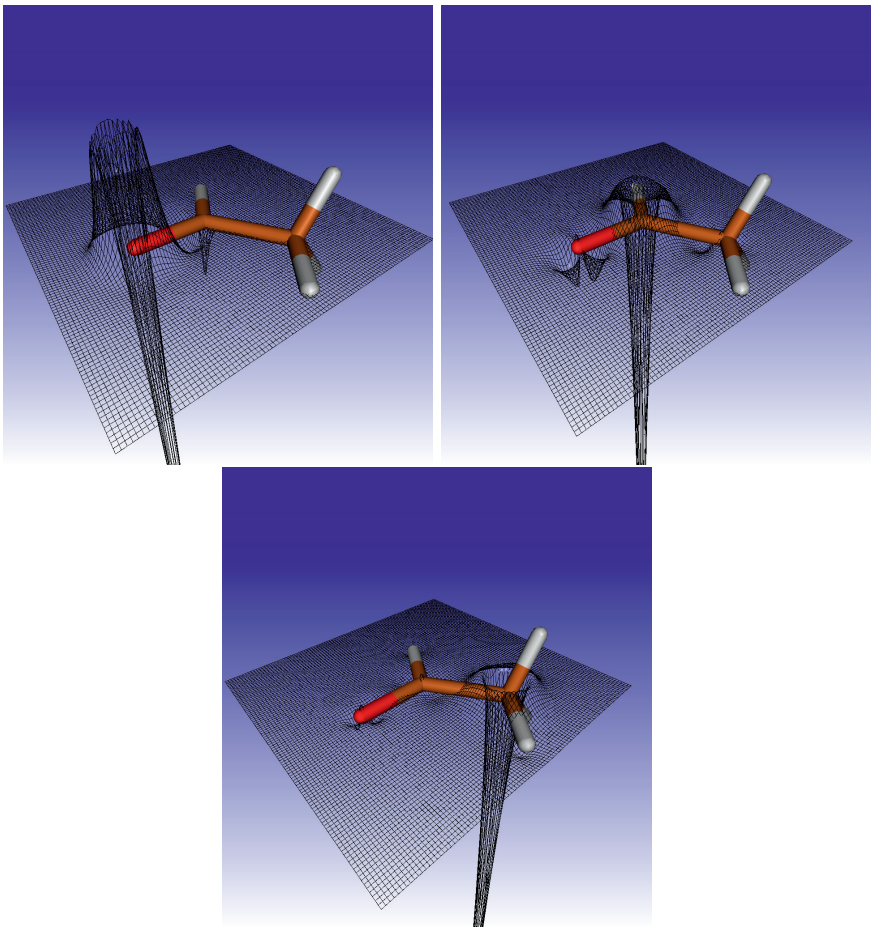


Figure D.6: Difference in excited and ground state electron density in the molecular plane for the most intense peaks of oxygen (left), C1 (middle) and C2 (right) calculated at the CCSD level.

Table D.3: Propenal. Oxygen CVS-Davidson core excitation energies (eV) and corresponding oscillator strengths $f \times 100$ (dimensionless). Basis set: aug-cc-pVDZ.

	CC2		CCSD		MLCCSD **A**		MLCCSD **E**	
	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
S_1	534.28	3.31	535.01	4.22	535.04	4.53	535.02	4.22
S_2	536.53	0.03	540.11	0.02	539.87	0.03	540.12	0.02
S_3	537.05	0.01	540.71	0.00	540.45	0.01	540.72	0.00
S_4	537.22	0.17	540.93	0.05	540.61	0.07	540.94	0.05
S_5	537.29	0.00	541.05	0.13	540.76	0.10	541.06	0.14

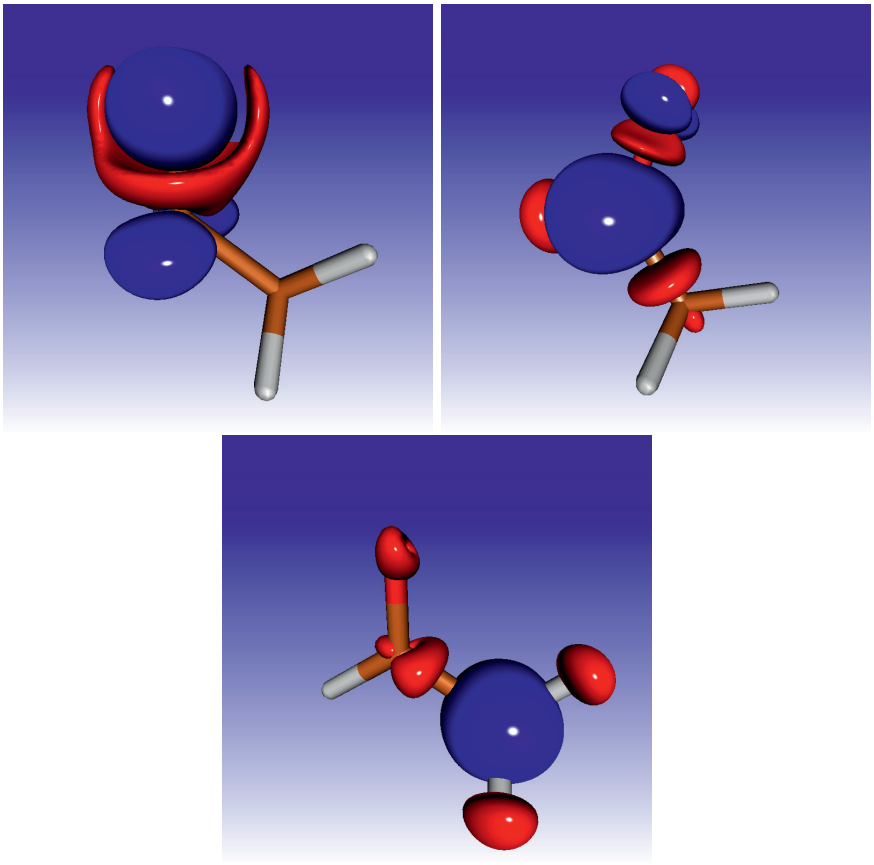


Figure D.7: Difference in excited and ground state electron density plotted as isosurfaces (0.01) for the most intense peaks of oxygen (left), C1 (middle) and C2 (right) calculated with CCSD. Blue corresponds to increased and red to decreased density.

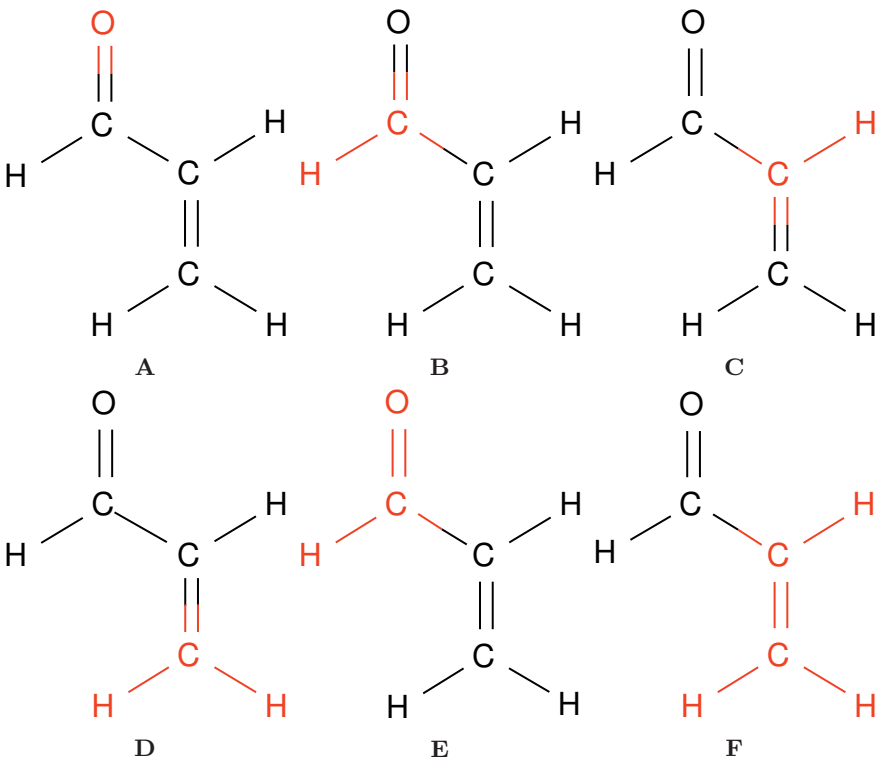


Figure D.8: Active spaces of propenal in red

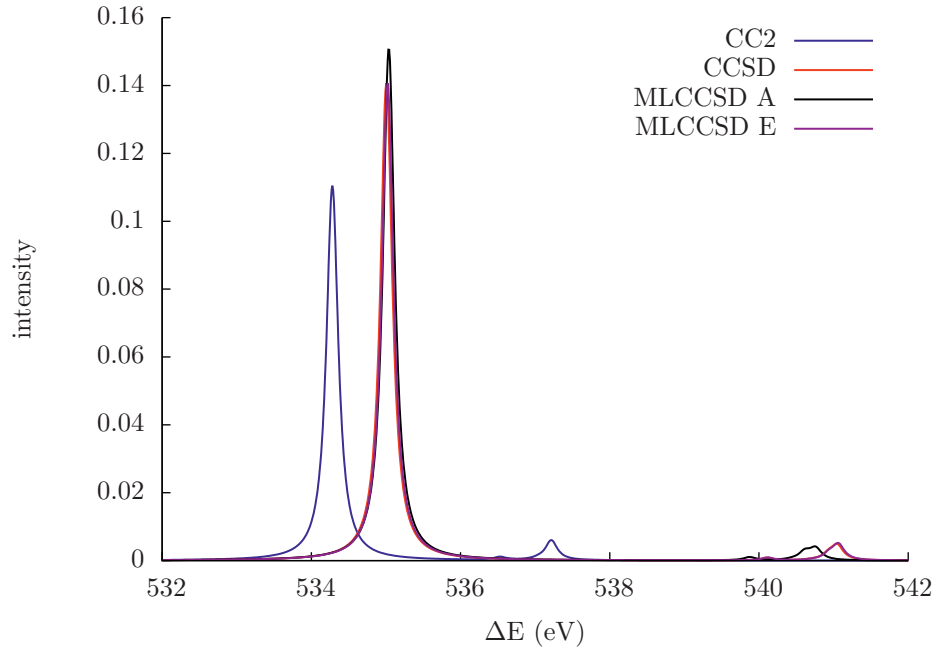


Figure D.9: Propenal. Comparison of the oxygen K-edge spectra obtained with CC2, CCSD, and MLCCSD **A** and **E** and the CVS-Davidson algorithm. Basis set: aug-cc-pVDZ.

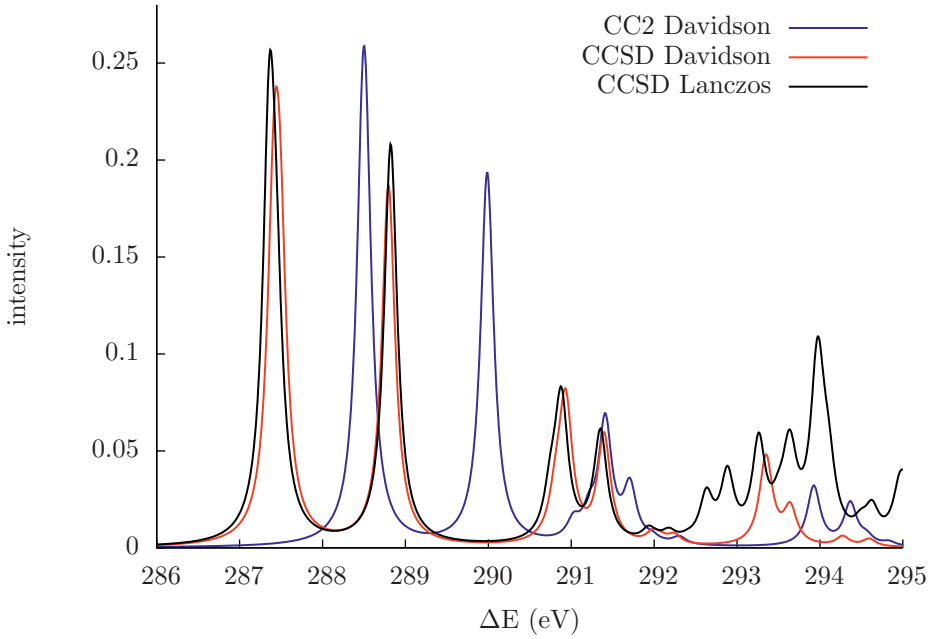


Figure D.10: Propenal. Comparison of the carbon K-edge spectra obtained with the CVS-Davidson and regular Lanczos algorithms. Basis set: aug-cc-pVDZ.

to the Lanczos algorithm not being fully converged for the chosen chain length of 3000. Above 292 eV, the two spectra differ significantly. This happens because there is a large number of states in this energy range and only five roots for each C atom were determined using the CVS-Davidson approach. Consequently, only few states were obtained within this energy range and it was somewhat arbitrary which states the Davidson algorithm converged to. Because CCSD and MLCCSD use different orbitals, the two methods will often converge to different roots in such cases. This complication can be avoided by calculating more states. The CVS results for CC2 and CCSD, as well as for two sets of MLCCSD active spaces, are reported in TableD.4.

In Figure D.11, the same CVS CCSD spectrum is compared to a MLCCSD spectrum computed using the minimal active spaces **B**, **C** and **D**. MLCCSD reproduces the main features of the spectrum and for most applications this will be sufficiently accurate. Figure D.12 is similar to Figure D.11, but a set of larger active spaces was used in the MLCCSD calculation. Spaces **E** and **F** both contain two second row atoms, **E** contains oxygen and C1 and **F** contains the remaining carbons. As can be seen in the figure, this gives excellent results for the excitations below 293 eV. Note that the peak at 287.5 eV appears to be a bit too intense. This happens because the peak is the sum of two excitations that are separated by 0.07 eV in the CCSD case and by 0.03 eV in MLCCSD, see Table D.4. The calculated intensities

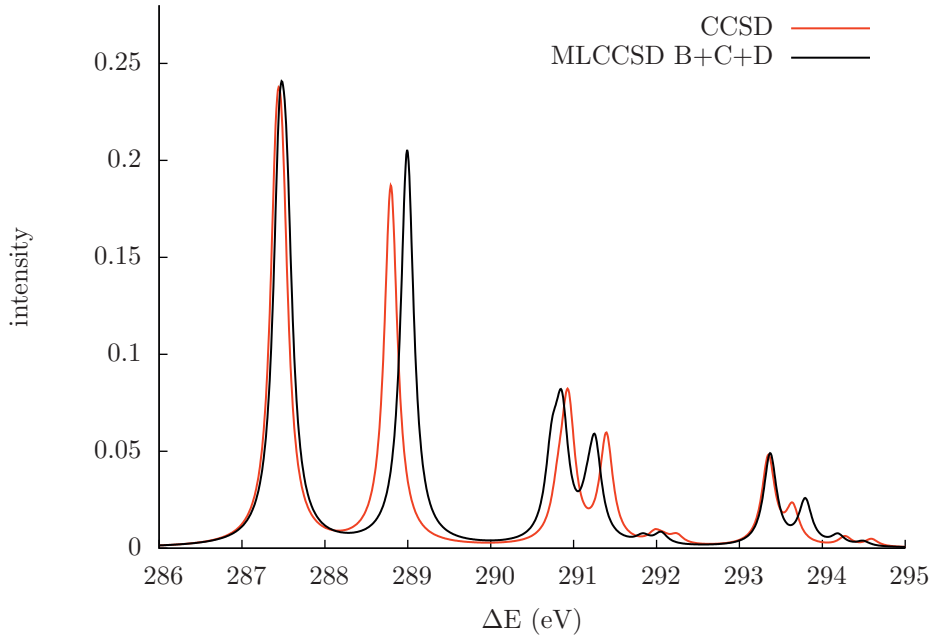


Figure D.11: Propenal. Comparison of the carbon K-edge spectra obtained with CCSD and MLCCSD **B**, **C** and **D** with the CVS-Davidson algorithm. Basis: set aug-cc-pVDZ.

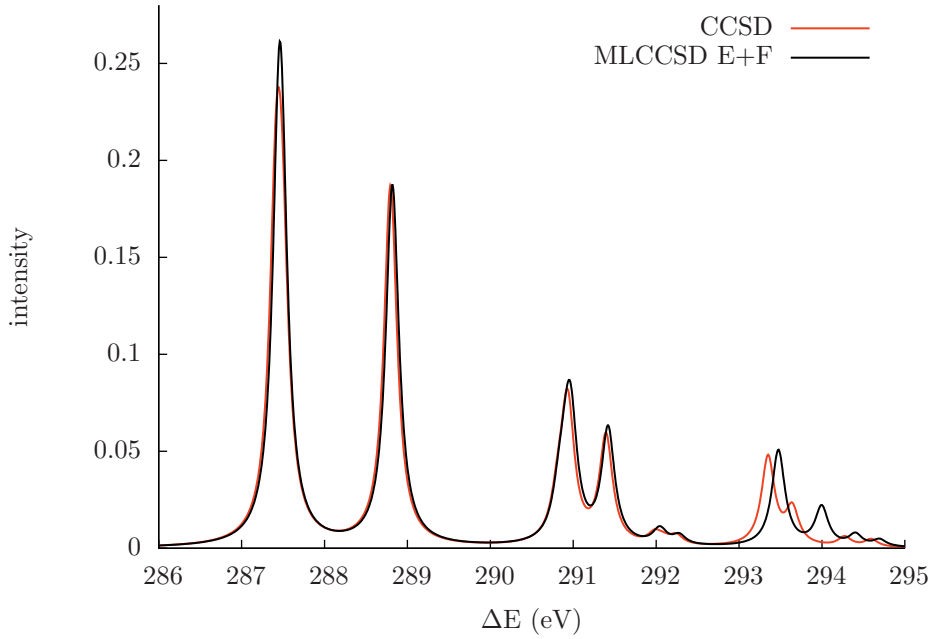


Figure D.12: Propenal. Comparison of the carbon K-edge spectra obtained with CCSD and MLCCSD **E** and **F** with the CVS-Davidson algorithm. Basis: set aug-cc-pVDZ.

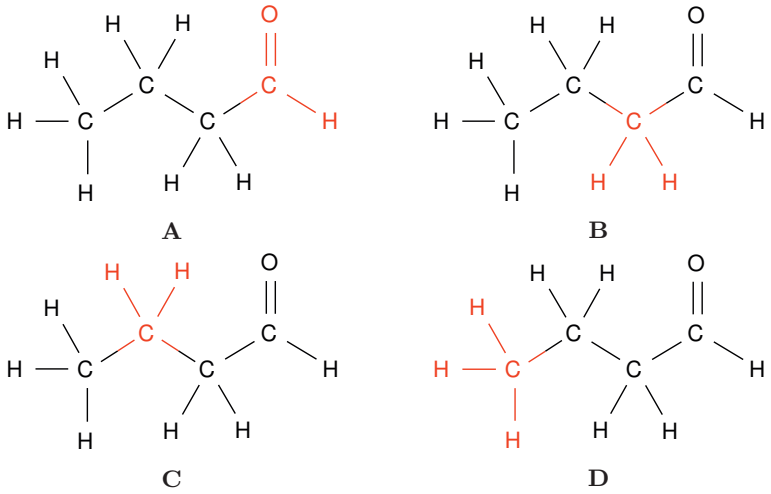


Figure D.13: Active spaces of butanal in red.

are actually slightly lower for MLCCSD.

D.3.4 Butanal

Finally, to investigate the behavior of single bonded carbon chains we calculated the core excitation energies of butanal [263]. We used the active spaces in Figure D.13. Active space **A** contains the oxygen atom and the attached methine group which proved sufficient for ethanal and propenal. The active spaces **B**, **C** and **D** only contain a methyl or a methylene group.

Figure D.14 shows the spectra at the oxygen K-edge of butanal calculated with CC2, CCSD and MLCCSD and the CVS-Davidson algorithm. Similarly to ethanal and propenal, the excitation energy of the main edge is too low with CC2 and the gap to the higher shake-up excitations is too small. The spectra for CCSD and MLCCSD are almost indistinguishable, even for the higher excitations.

Table D.5 collects the core excitation energies of oxygen in butanal with the corresponding oscillator strengths. None of the errors in the MLCCSD energies compared to CCSD are larger than 0.03 eV. The errors in the oscillator strengths are also small, whereas the relative errors vary a lot because the magnitude of the oscillator strengths vary by three orders of magnitude.

Obtaining the carbon K-edge spectrum requires a bit more care in the choice of active space, depending on the accuracy required. In Figure D.15, the peaks below 289 eV and above 292 eV correspond to C1 and were calculated using active space **A** in MLCCSD and these are well reproduced. The rest of the peaks are from the

Table D.4: Propenal. Carbon CVS-Davidson core excitation energies (eV) and corresponding oscillator strengths $f \times 100$ (dimensionless). Basis set: aug-cc-pVDZ.

		CC2		CCSD		MLCCSD **B**+**C**+**D**		MLCCSD **E**+**F**	
		ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
C1	S_1	289.99	5.76	288.80	5.56	288.99	6.11	288.82	5.58
	S_2	293.93	0.92	293.35	1.38	293.37	1.42	293.47	1.49
	S_3	294.37	0.65	293.64	0.56	293.79	0.69	294.00	0.60
	S_4	294.56	0.12	294.27	0.15	294.19	0.16	294.40	0.18
	S_5	294.83	0.06	294.59	0.11	294.49	0.07	294.70	0.10
C2	S_1	288.52	3.50	287.48	3.75	287.54	3.66	287.48	3.67
	S_2	291.03	0.29	290.91	0.71	290.74	0.65	290.91	0.74
	S_3	291.41	1.66	290.94	1.50	290.86	1.84	290.97	1.73
	S_4	291.41	0.21	291.41	0.16	291.17	0.36	291.41	0.24
	S_5	291.97	0.08	292.08	0.07	291.84	0.06	292.07	0.07
C3	S_1	288.49	4.40	287.41	4.32	287.45	4.85	287.45	4.31
	S_2	291.22	0.37	290.81	0.53	290.73	0.53	290.83	0.54
	S_3	291.71	0.85	291.39	1.51	291.26	1.41	291.42	1.53
	S_4	292.27	0.02	291.98	0.16	291.83	0.06	292.03	0.18
	S_5	292.31	0.10	292.24	0.14	292.05	0.18	292.28	0.14

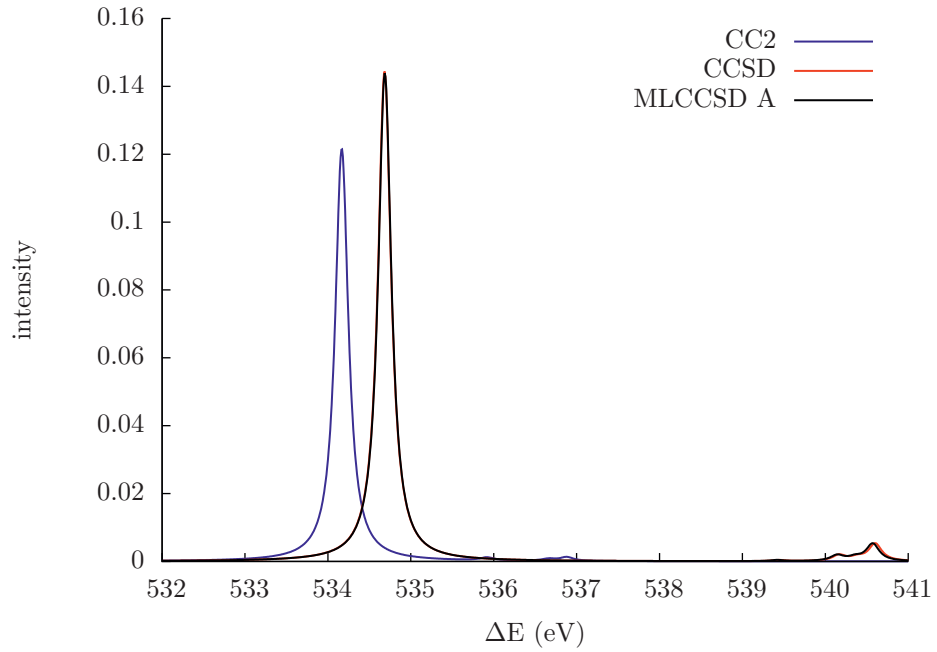


Figure D.14: Butanal. CVS-Davidson oxygen K-edge calculated with the Davidson algorithm. Basis set: aug-cc-pCVDZ

Table D.5: Butanal. Oxygen CVS-Davidson core excitation energies (eV) and corresponding oscillator strengths $f \times 100$ (dimensionless, in parentheses). Basis set: aug-cc-pCVDZ.

	CC2		CCSD		MLCCSD **A**	
	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
S_1	534.17	3.65	534.68	4.33	534.69	4.31
S_2	535.92	0.03	539.44	0.01	539.42	0.01
S_3	536.66	0.01	540.17	0.05	540.14	0.05
S_4	536.68	0.00	540.39	0.03	540.36	0.03
S_5	536.88	0.03	540.60	0.15	540.57	0.15

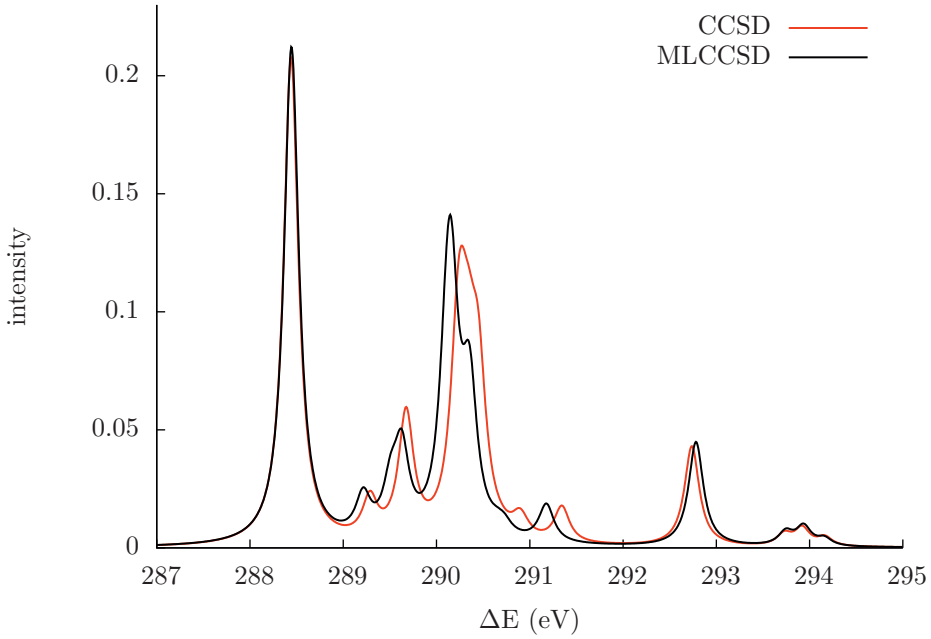


Figure D.15: Butanal. CVS-Davidson carbon K-edge calculated with the Davidson algorithm. Basis set: aug-cc-pCVDZ

atoms C2, C3 and C4. They are shifted down a bit compared to CCSD, but overall correspond reasonably well with CCSD. The CC2 spectrum is omitted from Figure D.15 for clarity, but it showed similar discrepancies from CCSD as seen in ethanal and propenal.

Comparing the excitation energies in Tab. D.6, the largest error for C1 is about 0.04 eV. For the other carbons, the errors are larger, but none are larger than 0.3 eV. We note that in some cases, the CC2 excitation energies are more accurate than the MLCCSD excitation energies. These are weak excitations that will be difficult to see in an experiment and oscillator strengths are better described by MLCCSD.

D.3.5 Comparison with experiment

In order to demonstrate the accuracy of CC2 and CCSD we compare with experimental values for ethanal. As large basis set effects are observed for core excitations we use aug-cc-pCVTZ[59] in this comparison. From our experience, the larger basis set is only required on the core being excited. As ethanal is rather small we employ the same basis for all atoms.

Table D.6: Butanal. Carbon CVS-Davidson core excitation energies (eV) and corresponding oscillator strengths $f \times 100$ (dimensionless). Basis set aug-cc-pCVDZ.

		CC2		CCSD		MLCCSD	
		ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
C1	S_1	289.70	6.59	288.44	6.21	288.44	6.33
	S_2	293.34	0.86	292.74	1.27	292.78	1.33
	S_3	294.16	0.14	293.73	0.15	293.74	0.17
	S_4	294.25	0.13	293.92	0.22	293.94	0.24
	S_5	294.43	0.09	294.16	0.11	294.16	0.09
C2	S_1	290.04	0.11	289.52	0.01	289.46	0.02
	S_2	290.08	0.62	289.68	1.01	289.63	1.05
	S_3	290.75	0.98	290.45	1.78	290.35	1.76
	S_4	290.96	0.08	290.88	0.17	290.73	0.13
	S_5	291.28	0.02	291.30	0.03	291.12	0.02
C3	S_1	289.93	0.35	289.66	0.60	289.51	0.55
	S_2	290.51	0.07	290.35	1.25	290.16	1.19
	S_3	290.52	0.55	290.54	0.06	290.28	0.10
	S_4	290.80	0.01	290.91	0.07	290.61	0.01
	S_5	290.84	0.05	290.93	0.05	290.65	0.08
C4	S_1	289.70	0.38	289.29	0.50	289.21	0.51
	S_2	290.48	0.78	290.22	1.45	290.11	1.44
	S_3	290.51	0.71	290.27	1.47	290.16	1.43
	S_4	290.80	0.00	290.68	0.02	290.55	0.01
	S_5	291.36	0.35	291.35	0.43	291.18	0.46

Table D.7: Ethanal. Oxygen core excitation energies (in eV) computed using the CVS-Davidson algorithm compared with experiment[264] Basis sets: aug-cc-pCVTZ.

	CC2	CCSD	MLCCSD **A**	MLCCSD **D**	Experiment
S_1	532.58	532.69	532.61	532.69	531.53
S_2	534.61	537.50	536.67	537.47	535.42
S_3	535.27	538.28	537.39	538.24	536.32
S_4	535.43	538.52	537.61	538.49	537.05

Table D.8: Ethanal. Relative oxygen core excitation energies (in eV) computed using the CVS-Davidson algorithm compared with experiment [264] Basis sets: aug-cc-pCVTZ.

	CC2	CCSD	Experiment
S_1	0.00	0.00	0.00
S_2	2.03	4.82	3.89
S_3	2.69	5.59	4.79
S_4	2.85	5.83	5.52

In Table D.7 we report oxygen K-edge excitation energies in ethanal and compare to the NEXAFS excitation energies obtained by Prince *et al.*[264] Both CCSD and CC2 is about 2 eV too high compared with experiment for the first core excitation and the difference is about 0.1 eV between them. The difference between the methods is larger for the higher states, as seen in Table D.8. Relative to the first excited state, CC2 is about 1.9 eV too low for S_2 and get progressively worse for the higher states. The relative CCSD excitation energy of S_2 is about 0.9 eV too high and the difference between the higher states is a bit too low. We note that MLCCSD **D** reproduces full CCSD well, while MLCCSD **A** has larger errors.

D.4 Conclusions

With the examples presented in this paper, we have demonstrated that it is possible to determine CCSD core excitation energies and oscillator strengths using MLCCSD with small active spaces. In many cases, the results obtained with the smallest active spaces are sufficiently accurate given that a linewidth of 1 eV or more is not uncommon in experiments [264]. Small expansions of the active spaces in ethanal and propenal were sufficient to obtain CCSD results with errors less than 0.01 eV.

The current implementation is not yet fully optimal for production calculations and no timings are therefore reported in this paper. However, the most expensive term in CCSD scales as V^4O^2, while CC2 scales as V^3O^2. Butanal is a quite small molecule and with active space **A**, the most expensive term to calculate in MLCCSD will be the same as in full CCSD. The theoretical time reduction in the CCSD part is

$$\frac{V^4O^2}{V^2V_A^2O_A^2} = \frac{187^2 \times 20^2}{61^2 \times 8^2} \approx 59 \tag{D.23}$$

and the overall theoretical scaling is that of CC2. Actual time reduction will depend on the implementation and size of the active part compared to the whole system, but results with multi-level CC3 [265] indicate that time reductions are close to the

theoretical value for sufficiently large systems. We should emphasize that MLCCSD is not a black box method as the active spaces need to be chosen. However, we are confident that with time, users will gain experience in selecting optimal spaces.

To further investigate the excitation processes, we have visualized the change in electronic density between states. For core excitations, a large reduction of electronic density is observed in the core and a corresponding increase in the valence region. This may be used to determine a suitable active space. If the electron change is delocalized with a low level method, a larger active space may be required. For small active spaces that only contain one second row atom, increasing the size will probably not appreciably affect the overall computational cost because the inactive part is relatively more expensive. Systematic studies with different spaces will give an indication on whether the active space is large enough.

Acknowledgments

H.K. and R.H.M. thank T. J. Martinez for hosting part of this project at Stanford University and we would like to thank Kristin Marie Skjelbred for providing a preliminary version of the transition moment code. H.K. acknowledges financial support from the FP7-PEOPLE-2013-IOF funding scheme (Project No. 625321). S.C. acknowledges financial support from the AIAS-COFUND program (Grant Agreement No. 609033). The COST Actions No. CM1002 “COnvergent Distributed Environment for Computational Spectroscopy (CODECS)”, MP1306 “Modern Tools for Spectroscopy on Advanced Materials (EUSPEC)” and CM1204 “XUV/X-ray light and fast ions for ultrafast chemistry (XLIC)” are also acknowledged.

Paper E

Core-level and UV spectra of glycine within Coupled Cluster Linear Response theory

Manuscript

Core-level and UV spectra of glycine within Coupled Cluster Linear Response theory

Rolf H. Myhre[1,2], Sonia Coriani[3,4] and Henrik Koch[1,2*]

[1]*Department of Chemistry, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway*
[2]*Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California 94305, USA*
[3]*Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Trieste, I-34127, Italy*
[4]*Aarhus Institute of Avanced Studies, University of Aarhus, DK-8000 Århus C, Denmark*

Abstract

The coupled cluster models CCSD and CC3 are used to investigate the (core) excited states and electron binding energies of glycine in the gas phase. Excited states and binding energies in the UV spectral range are calculated using standard coupled cluster linear response, while core-level excited states and ionization potentials are calculated using the core-valence separation approximation. The importance of different conformers at the temperature of the experimental measurement is also assessed.

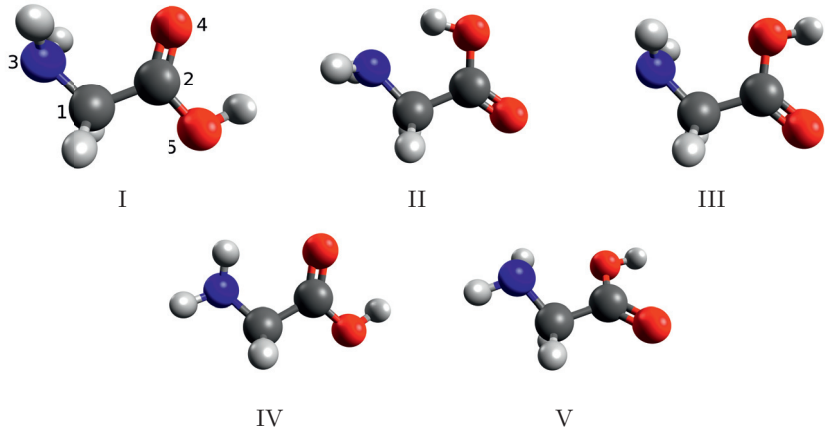


Figure E.1: Conformers of glycine.

E.1 Introduction

Glycine is the smallest amino acid found in proteins and is often used as a model system for studying properties of biological molecules [266]. It plays an essential role in living organisms and has even been found in space [267], supporting the hypotesis of panspermia. In the gas phase it exists as neutral species. In solution, it is a zwitterion and it is one of the smallest molecules with internal hydrogen bonding. Furthermore, hydrogen bonding results in a large number of predicted stable conformers [127, 268, 269].

A large number of studies have been performed on glycine in order to characterize the molecule, including several X-ray spectroscopy studies in gas phase [124, 270, 271], in solution [272] and even in solid state [273, 274]. In order to interpret the experimental results, accurate computational methods are required. The static exchange (STEX) method [125] has been used in many studies to simulate the spectra. While STEX can give good qualitative spectra, it requires shifts of several eV to agree with experimental peaks.

A much more advanced approach is the algebraic diagramatic construction (ADC) method [52, 275]. In this approach, an effective Hamiltonian is constructed based on perturbation theory. This results in a size extensive and explicitly Hermitian theory. By combining this method with the core-valence separation approximation, the ADC(2) approach has been often used to calculate X-ray absorption spectra. Wenzel *et al.* [250], for instance, were able to obtain spectra at the ADC(2) and UADC(2) level of medium to large size systems. The extended ADC(2) method, named ADC(2)-x, showed good agreement with experimental data, but fortuitous error cancelation probably plays an important role, and it is difficult to systematically improve the accuracy [252].

More recently, the CVS approximation has been implemented within the coupled cluster (CC) framework [116]. The CC ansatz provides a systematic hierarchy of models with increasing accuracy. This makes it possible to quantitatively predict molecular properties and spectra and to assess whether a specific approximation within the hierarchy is accurate enough. The calculation of spectroscopic observables related to the interaction of the molecular system with X-ray radiation displays strong dependence on the level of theory and size of the basis sets, making a systematic approach particularly important.

In this work, we present and compare spectroscopic data of glycine calculated at the CCSD (coupled cluster singles and doubles) [21] and CC3 (coupled cluster singles, doubles and approximate triples) [24, 265] levels of theory. Specifically, we have modelled absorption and ionization spectral parameters in the UV and X-ray energy ranges, and compare them to experimental values from the literature [124, 271]. As excited state properties show a strong dependence on the basis set, we have performed calculations with different sets in order to determine the best set with respect to both accuracy and cost. We also investigate the effect of different conformers.

The paper is organized as follows. In Section E.2, we present a brief summary of CC theory and CC linear response. In the Section E.3 we present our results and compare to experiments and Section E.4 contains our concluding remarks.

E.2 Theory

In coupled cluster theory, the wave function is typically written as the exponential of a cluster operator, X, working on a Hartree-Fock (HF) reference state

$$|\mathrm{CC}\rangle = \exp(X)\,|\mathrm{HF}\rangle \ . \tag{E.1}$$

The cluster operator is a sum of excitation operators each weighted by their corresponding amplitude, $X = \sum_\mu x_\mu \tau_\mu$, and the wave function is determined by projection

$$\begin{aligned} E &= \langle \mathrm{HF}|\exp(-X) H \exp(X)\,|\mathrm{CC}\rangle \\ \Omega_\mu &= \langle \mu|\exp(-X) H \exp(X)\,|\mathrm{CC}\rangle = 0 \ . \end{aligned} \tag{E.2}$$

Excitation energies and excited state properties are usually calculated using either CC linear response (CCLR) [31, 32] or equation of motion (EOM) CC [34]. The critical step in both methods is finding the eigenvalues of the Jacobian matrix, $\boldsymbol{A}$, with elements given by

$$A_{\mu\nu} = \frac{\partial \Omega_\mu}{\partial x_\nu} = \langle \mu|\exp(-X)[H, \tau_\nu]\exp(X)\,|\mathrm{CC}\rangle\,. \tag{E.3}$$

Because the Jacobian is not symmetric, both the right, $\boldsymbol{R}_i$, and left, $\boldsymbol{L}_j$ eigenvectors must be determined

$$\boldsymbol{A}\boldsymbol{R}_i = \omega_i \boldsymbol{R}_i; \quad \boldsymbol{L}_j \boldsymbol{A} = \omega_j \boldsymbol{L}_j. \tag{E.4}$$

The eigenvectors must satisfy the biorthogonality condition, $\boldsymbol{L}_j \boldsymbol{R}_i = \delta_{ij}$, and the eigenvalues, ω_i, are the excitation energies of the system.

Typically, the Jacobian matrix will be far too large for a complete diagonalization and the eigenvalue problem must be solved using a subspace algorithm. Usually, the Davidson algorithm [42] is employed. This algorithm solves the eigenvalues iteratively, starting with extremal values. As a result, it is not straightforward to find eigenvalues corresponding to core excited states because they are embedded in a continuum of valence excited states. The CVS approximation [114, 116] overcomes this limitation by utilizing the fact that there is almost no coupling between valence- and core excited states. Vector elements corresponding to such couplings can then be set to zero without loss of accuracy and the algorithm will not collapse to valence excited states.

In order to calculate ionization energies, an extremely diffuse orbital can be added to the basis set [116, 128]. Integrals involving this orbital will then be zero and ionization energies can be approximated as excitations to this orbital. The (relative) strength of the corresponding ionization can be approximately estimated by the norm of the so-called Dyson orbitals [276], which can be obtained from the transition density matrix elements for excitations to and from the diffuse orbital [277].

E.3 Results

E.3.1 Geometries

In order to compare our results to experiment, accurate geometries are required. Miller and Clary [266] identified five important conformers of glycine summarized in Figure E.1. Conformer I indicates the numbering scheme adopted in this work where atoms are ordered after their 1s orbital energy in descending order.

The geometries were optimized at the CCSD(T) level using CFOUR [122, 123] and the aug-cc-pVDZ [60] basis set of Dunning and coworkers. Frequencies were also calculated for the different geometries and zero point energies of about 2.15 eV were obtained for all conformers, consistent with earlier studies [266]. Geometries, ground state energies and zero point energies are collected in the supplementary material. The optimized geometries were not strictly symmetric, but most states could nonetheless be clearly assigned as either A' (in the plane), or A'' (out of the plane) for conformer I, II and III. States of A' symmetry are of σ character, and those of A'' symmetry are of π character.

Table E.1: Conformer I valence excitations and ionizations (eV)

	CCSD				CC3				expt.	
	aCDZ		aCTZ		aCDZ		aCTZ		(±0.1) [271]	Asgn
	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE		ΔE		ΔE	
S_1	5.90 (A'')	0.01	5.89 (A'')	0.02	5.88 (A'')		5.83 (A'')		–	
S_2	6.45 (A')	1.62	6.62 (A')	1.57	6.39 (A')		6.51 (A')		–	
S_3	7.27 (A')	7.72	–	–	7.16 (A')		–		–	
S_4	7.37 (A'')	0.06	–	–	7.17 (A'')		–		–	
S_5	7.57 (A')	1.01	–	–	7.49 (A')		–		–	
I_1	10.01 (A')	-	10.24 (A')	-	9.97 (A')		10.13 (A')		10.0	n_N
I_2	10.90 (A')	-	-	-	10.88 (A')		-		11.2	n_O
I_3	12.06 (A'')	-	-	-	12.01 (A'')		-		12.2	π_{OO}
I_4	13.53 (A'')	-	-	-	13.41 (A'')		-		13.7	
I_5	14.37 (A')	-	-	-	14.27 (A')		-		14.4	

E.3.2 Excitation and binding energies of conformer I

In this section we present CCSD and CC3 excitation and ionization energies calculated for the lowest energy conformer, using several different basis sets. Transition strengths for the excitations calculated with CCSD are also given, and they were computed in the length gauge.

In our experience, diffuse basis functions [60] are important for excited state properties, so they have been included in all basis sets. Furthermore, core-valence polarization functions [59] are required for accurate core excitations and are included. For the sake of brevity, we adopt a shorthand notation for the basis sets. For example, aug-cc-pCVDZ will be refered to as aCDZ. We also use mixed basis sets for core excitations and ionizations, and these are indicated with parentheses, i.e., a(CT)Z means aug-cc-pCVTZ on the atom being excited and auc-cc-pVDZ on the rest.

Table E.1 collects the results for the valence excitations and ionizations investigated. Only small changes are observed for the excitation energies, both with respect to basis sets and wavefunction models. Osted *et al.* [278] have also investigated the UV excitations of (among others) glycine at the CCSD level of theory. Using the aD(T)Z basis set, they reported 5.88 eV ($f = 0.00053$), 6.32 eV ($f = 0.01247$), 7.24 eV ($f = 0.04608$), 7.48 eV ($f = 0.01535$) and 7.61 eV ($f = 0.03162$) for their minimum energy conformer. Unfortunately, no experimental gas-phase UV results seem to be available in the literature.

For the ionizations, the effect of changing the basis set is somewhat larger, with the smaller basis set yielding values slightly closer to the experimental one. The deviations from the experimental values for both CC methods is modest, and it does not exceed 0.3 eV.

We now turn our attention to the X-ray spectra. Core excitation energies calculated with CCSD for all types of edges together with their relative intensities are presented in Table E.2, whereas CC3 and experimental excitation energies are given in Table E.3. Changes of about 1 eV are observed when comparing CCSD and CC3 excitation energies, and the A″ states were generally lowered more than the A′. For C1 and O5, the lowest core excitation changed from being an A′ state to an A″ state when going from CCSD to CC3.

Increasing the basis from aCDZ to a(CT)Z lowers the excitation energy by another eV, indicating that triple zeta basis (as well as high level theory) are required to accurately predict core excitation energies. However, only small differences are observed between a(CT)Z and the larger sets, indicating that it is sufficient to use triple zeta only on the atom being excited.

Comparing the results to experiment, too high excitation energies are observed for carbon and too low energies for oxygen, but all CC3 results are within 0.4 eV, with the sole exception of the C1. Relativistic effects would increase all excitation

Table E.2: Conformer I. CCSD core excitation energies (in eV) and corresponding intensities.

	aCDZ		a(CT)Z		aCTZ		a(CQ)Z	
	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
C1	290.79 (A′)	0.96	289.85 (A′)	1.04	289.87 (A′)	1.04	289.79 (A′)	1.07
	291.03 (A″)	0.58	290.07 (A″)	0.62				
	291.96 (A′)	0.34	291.01 (A′)	0.38				
	291.98 (A″)	1.19	291.04 (A″)	1.24				
	292.41 (A′)	0.65	291.45 (A′)	0.64				
C2	290.53 (A″)	7.20	289.48 (A″)	7.55	289.43 (A″)	7.62	289.35 (A″)	7.64
	294.35 (A′)	0.10	293.40 (A′)	0.10				
	295.23 (A′)	0.14	294.28 (A′)	0.15				
	295.63 (A″)	0.03	294.66 (A″)	0.02				
	295.84 (A′)	0.02	294.89 (A′)	0.02				
N3	404.30 (A′)	0.60	402.90 (A′)	0.61	402.90 (A′)	0.61	402.82 (A′)	0.63
	405.41 (A″)	1.89	403.99 (A″)	1.95				
	405.94 (A′)	0.10	404.52 (A′)	0.11				
	406.49 (A″)	0.32	405.08 (A″)	0.33				
	406.57 (A′)	0.59	405.13 (A′)	0.62				
O4	535.87 (A″)	3.75	533.96 (A″)	3.78	533.90 (A″)	3.82	533.83 (A″)	3.79
	539.09 (A′)	0.08	537.14 (A′)	0.08				
	540.01 (A′)	0.44	538.06 (A′)	0.43				
	540.39 (A″)	0.13	538.42 (A″)	0.14				
	540.70 (A′)	0.19	538.74 (A′)	0.20				
O5	539.19 (A′)	1.56	537.34 (A′)	1.55	537.37 (A′)	1.56	537.23 (A′)	1.58
	539.75 (A″)	1.46	537.89 (A″)	1.45				
	541.37 (A′)	0.76	539.49 (A′)	0.75				
	541.74 (A″)	0.37	539.85 (A″)	0.40				
	542.05 (A′)	0.36	540.15 (A′)	0.36				

Table E.3: Conformer I. CC3 core excitation energies and experimental results from Plekan *et al.* [124] (eV). No symmetry could be determined for the degenerate states of C1 with aCDZ

	aCDZ	a(CT)Z	aCTZ	a(CQ)Z	exp.
	289.77 (A*)	288.71 (A″)	288.59 (A″)	288.57 (A″)	287.7
	289.77 (A*)	288.73 (A′)			
C1	290.74 (A′)	289.70 (A′)			289.4
	290.83 (A″)	289.79 (A″)			
	291.06 (A′)	290.02 (A′)			
	290.15 (A″)	288.97 (A″)	288.94 (A″)	288.78 (A″)	288.4
	293.09 (A′)	292.03 (A′)			
C2	293.86 (A′)	292.81 (A′)			
	294.35 (A″)	293.27 (A″)			
	294.47 (A′)	293.41 (A′)			
	402.76 (A′)	401.34 (A′)	401.34 (A′)	401.20 (A′)	401.2
	403.76 (A″)	402.36 (A″)			
N3	403.94 (A′)	402.54 (A′)			402.4
	403.94 (A″)	402.55 (A″)			
	404.55 (A′)	403.15 (A′)			403.8
	533.94 (A″)	532.10 (A″)	532.01 (A″)	531.90 (A″)	532.2
	535.63 (A′)	533.93 (A′)			
O4	536.43 (A′)	534.73 (A′)			
	538.83 (A′)	535.13 (A″)			
	537.01 (A′)	535.31 (A′)			
	536.86 (A″)	535.16 (A″)	535.01 (A″)	534.99 (A″)	535.4
	536.92 (A′)	535.22 (A′)			
O5	538.30 (A′)	536.65 (A′)			537.7
	538.74 (A″)	537.08 (A″)			
	538.90 (A′)	537.24 (A′)			

Table E.4: Conformer I. CCSD and CC3 core ionization energies and comparison with the experimental results from Plekan *et al.* [271] (eV)

		aCDZ	a(CT)Z	aCTZ	a(CQ)Z	exp. (±0.1)
C1	CCSD	294.68	293.75	293.79	293.70	292.3
	CC3	293.37	292.35	292.40	292.23	
C2	CCSD	297.92	296.98	297.04	296.93	295.2
	CC3	296.57	295.52	295.58	295.39	
N3	CCSD	408.56	407.14	407.17	407.08	405.4
	CC3	406.48	405.09	405.11	404.95	
O4	CCSD	542.71	540.75	540.81	540.66	538.4
	CC3	538.91	537.24	537.29	537.09	
O5	CCSD	544.22	542.33	542.38	542.24	540.2
	CC3	540.83	539.21	539.24	539.05	

energies, but the increase would be about 0.1 eV for carbon, 0.2 eV for nitrogen and 0.3 eV for oxygen. Vibrational effects may also influence the position of the peaks. For nitrogen, N_2, this effect decreased the excitation energy slightly [279]. We note that the peak corresponding to C1 is close to the intense peak from C2 and difficult to resolve in experimental spectra. Our results indicate that this peak corresponds to two nearly degenerate states close to a symmetry allowed intersection.

Core ionization energies are compared with experiment in Table E.4. CCSD overestimates the ionization potentials by 1.5–2.5 eV, depending on the edge. The CC3 carbon binding energies are close to the experimental values, whereas nitrogen and especially oxygen ionization energies are too low. The error for oxygen is more than 1 eV and will only partially be offset by including relativistic effects.

E.3.3 Effect of conformers

In Table E.5 we present core excitation energies and transition moments of O5 for the different conformers, calculated using the aCDZ basis set. Only small differences in the excitation energies are observed between the different conformers but II shows considerable differences in the transition moments. In addition, the order of the two lowest states is the opposite of the two other conformers (I and III) with a symmetry plane. However, these changes would be difficult to detect experimentally and might change with a larger basis set [124].

Within CC3, the peaks in the spectrum move closer together, see Table E.6, and the A″ state is the lowest for all three conformers. Conformer II is 0.4 eV lower than the others, but again, this would be hard to resolve in experiment. For the other atoms, smaller differences are observed between the conformers.

Table E.5: Glycine. CCSD core excitations (eV) and intensities for O5. Basis set aCDZ

	I		II		III		IV		V	
	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$	ΔE	$f \times 100$
S_1	539.19	1.56 (A′)	539.17	1.90 (A″)	539.23	1.51 (A′)	539.23	1.58	539.21	1.53
S_2	539.75	1.46 (A″)	539.63	0.47 (A′)	539.62	1.66 (A″)	539.73	1.49	539.68	1.50
S_3	541.37	0.76 (A′)	541.04	0.36 (A′)	541.25	1.04 (A′)	541.44	0.77	541.31	0.89
S_4	541.74	0.37 (A″)	541.22	0.15 (A″)	541.80	0.28 (A″)	541.86	0.41	541.86	0.29
S_5	542.05	0.36 (A′)	541.31	0.27 (A′)	542.03	0.38 (A′)	542.21	0.30	542.20	0.41

Table E.6: Glycine CC3 core excitations O5 (eV), aCDZ

	I	II	III	IV	V
S_1	536.86 (A″)	536.44 (A″)	536.84 (A″)	536.86	536.78
S_2	536.92 (A′)	536.65 (A′)	536.92 (A′)	536.99	536.97
S_3	538.30 (A′)	537.72 (A′)	538.35 (A′)	538.41	538.38
S_4	538.74 (A″)	537.83 (A″)	538.75 (A″)	538.91	538.76
S_5	538.90 (A′)	537.85 (A′)	538.82 (A′)	539.01	538.99

E.4 Conclusions

In this paper we have demonstrated the accuracy of CC3 combined with the CVS approximation to predict core-level spectra. Triple zeta basis sets are required for convergence, but only on the atom being excited. Discrepancies between calculations and experiment can partially be ascribed to relativistic effects. Vibrational effect may also be important, but neither relativistic not vibrational effects were included in this study.

Only small differences are predicted between the conformers of glycine. Conformer II shows some variation in the predicted spectra of O5 and C1. However, the changes are quite modest and would be difficult to observe in experiment because the excited states are close together.

Acknowledgements

H.K. and R.H.M. thank T. J. Martínez for hosting part of this project at Stanford University. H.K. acknowledges financial support from the FP7-PEOPLE-2013-IOF funding scheme (Project No. 625321). S.C. acknowledges financial support from the AIAS-COFUND program (Grant Agreement No. 609033). The COST Actions No. CM1002 “COnvergent Distributed Environment for Computational Spectroscopy (CODECS)”, MP1306 “Modern Tools for Spectroscopy on Advanced Materials (EUSPEC)” and CM1204 “XUV/X-ray light and fast ions for ultrafast chemistry (XLIC)” are also acknowledged.

E.5 Supporting information

Table E.7: Geometry of conformer I, Å

	X	Y	Z
H	2.381241	−0.100621	0.001448
O	1.617858	−0.703469	0.001228
O	0.551882	1.305495	−0.000600
C	0.502925	0.087807	−0.000246
C	−0.772090	−0.754539	−0.002222
H	−0.733956	−1.419463	0.878899
H	−0.734996	−1.412502	−0.888662
N	−2.020891	0.004291	0.001327
H	−2.032275	0.624191	0.815143
H	−2.031512	0.632846	−0.805841

Table E.8: Geometry of conformer II, Å

	X	Y	Z
C	0.610875	0.046401	−0.000142
O	1.698633	0.590019	0.001762
O	0.453312	−1.297663	−0.001159
H	−0.522747	−1.429620	−0.002539
C	−0.710897	0.837585	−0.002688
H	−0.695843	1.486334	−0.894226
H	−0.693871	1.497027	0.880945
N	−1.879677	−0.071719	0.001882
H	−2.472399	0.077665	−0.813273
H	−2.460244	0.070428	0.827065

Table E.9: Geometry of conformer III, Å

	X	Y	Z
C	0.585191	−0.082107	−0.000627
O	1.654781	−0.666393	0.005885
O	0.487664	1.282748	−0.003308
H	1.404587	1.609031	0.000306
C	−0.775551	−0.784298	−0.009842
H	−0.777321	−1.462667	0.860907
H	−0.780645	−1.431194	−0.904475
N	−1.987348	0.036596	0.006411
H	−1.984369	0.671730	−0.794309
H	−1.984917	0.638753	0.832252

Table E.10: Geometry of conformer IV, Å

	X	Y	Z
C	0.492420	−0.081816	−0.023844
O	0.533458	−1.299377	−0.048611
O	1.604534	0.701664	0.089639
H	2.362413	0.092543	0.121347
C	−0.767735	0.754284	−0.165811
H	−0.686416	1.627128	0.503673
H	−0.756737	1.138273	−1.208780
N	−1.939509	−0.054878	0.188806
H	−2.784280	0.365814	−0.200593
H	−1.840115	−0.982089	−0.231934

Table E.11: Geometry of conformer V, Å

	X	Y	Z
C	0.565636	−0.095325	−0.035361
O	1.580608	−0.729527	0.189795
O	0.542934	1.271148	−0.099941
H	1.451914	1.564118	0.084241
C	−0.799001	−0.693338	−0.340586
H	−0.790017	−1.722166	0.050139
H	−0.872613	−0.747170	−1.448469
N	−1.874684	0.073307	0.312821
H	−2.776123	−0.345093	0.074994
H	−1.889152	1.026319	−0.057060

Table E.12: Relative ground state and zero point energies energies, CCSD(T), aug-cc-pVDZ, kJ/mol

	ΔE_0	V_0	total
I	0.0000	208.228	0.0000
II	3.2901	209.190	4.2520
III	6.8944	208.329	6.9957
IV	5.4620	208.200	5.4338
V	10.2096	208.512	10.4934

Paper F

A theoretical and experimental benchmark study of core excited states in N_2

Manuscript

A theoretical and experimental benchmark study of core excited states in N_2

Rolf H. Myhre[1,2], Thomas J. A. Wolf[3], Sonia Coriani[4,5], Markus Gühr[3,6*] and Henrik Koch[1,2*]

[1]*Department of Chemistry, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway*

[2]*Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California 94305, USA*

[3]*PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, US*

[4]*Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Trieste, I-34127, Italy*

[5]*Aarhus Institute of Avanced Studies, University of Aarhus, DK-8000 Århus C, Denmark*

[6]*Department of Physics, University of Potsdam, Potsdam, Germany*

Abstract

The high resolution near edge x-ray absorption fine structure spectrum of nitrogen displays the vibrational structure of the core excited states.This makes nitrogen well suited for assessing the accuracy of different electronic structure methods for core excitations. We report high resolution experimental measurements performed at the SOLEIL synchrotron facility. These are compared with theoretical spectra calculated using coupled cluster theory and algebraic diagrammatic construction theory. The CC3 model is shown to be in quantitative agreement with experiment.

F.1 Introduction

X-ray absorption spectroscopy (XAS) provides a unique ability to study local properties of molecules [107–113]. The research area is in rapid development and, currently, several facilities are under construction, such as the European XFEL in Hamburg and the LCLS II at SLAC. In order to fully understand the insight provided by XAS, accurate theoretical methods are required to interpret the experiments.

Simulating high energy processes such as near edge x-ray absorption fine structure (NEXAFS) spectra involves a number of challenges not encountered in low energy processes such as UV/Vis spectroscopy. Exciting an electron from a core orbital involves a significant reduction in the screening leading to large relaxation effects. The most important are contraction of the valence electron density and repulsion from the electron transferred from the core. Accounting for these effects theoretically is challenging. For example, results from density functional theory often have to be shifted 10 eV or more to agree with experimental results [280, 281], and most methods require shifts of more than 1 eV [45, 124, 282]. More recently, the extended algebraic diagrammatic construction (ADC-x) method has produced results within a few tenths of an eV from experimental values, however, the accuracy appears to partially rely on error cancellation from the basis set [252].

Another challenge when computing core exciting states is that they are embedded in a continuum of Rydberg states. Most electronic structure methods involve solving an eigenvalue problem in order to determine excited electronic states. Usually, some version of the Davidson algorithm [42] is used for this. It solves the eigenproblem iteratively starting with extremal eigenvalues, normally the lowest one, but this is not a good approach for core excitations because there will be many lower energy excited states. Several methods have been proposed to overcome this problem. In the ΔSCF method [243], the energy of the excited state is calculated by restricting the occupation in the core orbitals and the excitation energy is calculated as the difference from the ground state. Another approach is to solve for the eigenvalues using the Lanczos algorithm [44–46]. With this algorithm, the eigenvalues are solved for the whole spectrum simultaneously. However, the algorithm may become numerically unstable and it is necessary to store a large number of vectors on disk. This limits the size of systems where the method can be applied.

Arguably the most successful approach to determining the core excited states is the core-valence separation (CVS) approach [114]. The energy differences between the core and valence orbitals are typically several hundred eVs. Consequently, their overlap integrals become very small and the coupling between them can be neglected. The CVS approach is utilized in ADC [115] and has also been implemented with coupled cluster (CC) [116]. For CC, the differences between CVS and full space results are typically less than 50 meV.

Recently, we reported a new implementation of coupled cluster singles and doubles

with perturbative triples (CC3) [24, 265]. This implementation has now been expanded with the CVS approximation and in this paper we will use an experimental spectrum of nitrogen to assess the accuracy of CC3 for core excited states. The vibrational structure in the spectrum makes it possible to evaluate the shape of the potential energy surface and determine the effect of vibrations on the excitation energy. Because N_2 only has 14 electrons, it is possible to use very large basis sets and minimize the basis set error which can be several eVs for core excitations.

In Section F.2 we will briefly summarize CC theory, particularly CC3, and the CVS approximation. In Section F.3 we describe our computational and experimental approach and in Section F.4 we present our results. Section F.5 contains our concluding remarks and future perspective.

F.2 Theory

In CC theory, the wave function is written as the exponential of the cluster operator, $X = \sum_\mu x_\mu \tau_\mu$, acting on the Hartree-Fock (HF) reference state,

$$|\mathrm{CC}\rangle = \exp(X)\,|\mathrm{HF}\rangle\,. \tag{F.1}$$

The excitation operators, τ_μ, take the reference state to an excited state in the Fock space, $|\mu\rangle = \tau_\mu\,|\mathrm{HF}\rangle$, and x_μ is the corresponding amplitude. In exact theory, the CC formulation is equivalent to full configuration interaction (FCI) up to a normalisation factor, but in practice the cluster operator is truncated at some excitation level and the amplitude equations solved with projection

$$\begin{aligned} E &= \langle \mathrm{HF}|\exp(-X)H\exp(X)\,|\mathrm{HF}\rangle \\ \Omega_\mu &= \langle \mu|\exp(-X)H\exp(X)\,|\mathrm{CC}\rangle = 0 \end{aligned} \tag{F.2}$$

When calculating time dependent properties such as excitation energies and transition moments, the standard methods are CC linear response [31, 32] and equation of motion CC (EOM-CC) [34]. Both methods require the eigenvalues of the non-symmetric Jacobian matrix $\boldsymbol{A}$, whose elements are defined as

$$A_{\mu\nu} = \frac{\partial \Omega_\mu}{\partial x_\nu} = \langle \mu|\exp(-X)[H,\tau_\nu]\exp(X)\,|\mathrm{CC}\rangle\,. \tag{F.3}$$

The eigenvalues of the Jacobian correspond to the vertical excitation energies of the system and it is sufficient to only solve from one side if other properties are not required. If transition moments and other excited state properties are desired, it is necessary to solve the eigenvalue problem from both the left and the right side. As mentioned in the introduction, the eigenproblems are typically solved using an iterative procedure like the Davidson method [42].

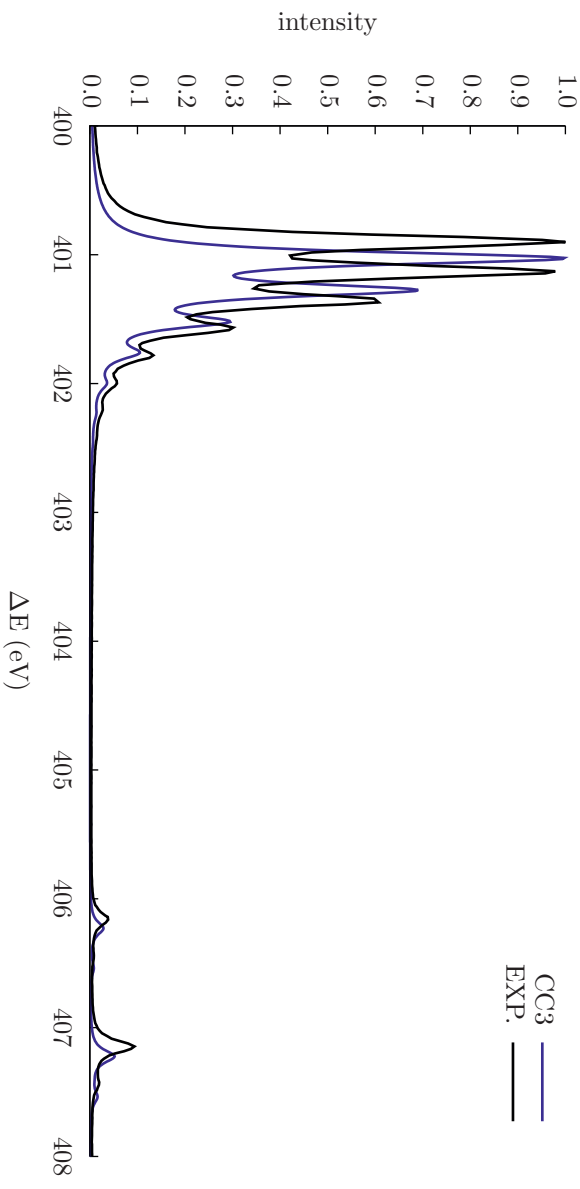


Figure F.1: Comparison of CC3 and experiment for the four lowest excited states. The intensity of the first peak is normalized to 1 and empirical line broadening added to the theoretical spectrum.

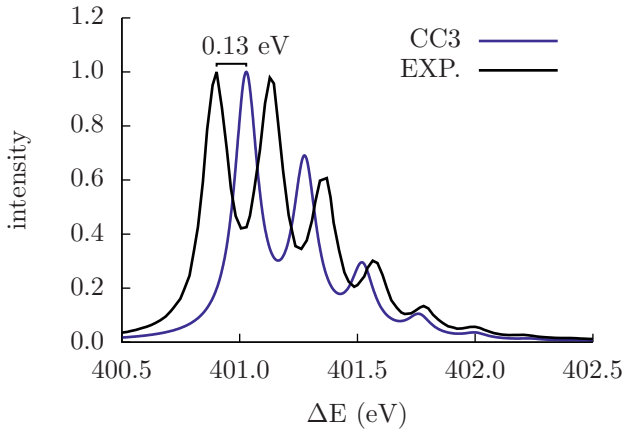


Figure F.2: Comparison of the first peak in the experimental and CC3 spectra. CC3 is shifted 0.13 eV up and the distances between the vibational peaks are a bit too large.

In order to obtain core excited states, the CVS approximation is necessary. In this approximation, elements coupling the core and valence excited states are projected out. For each iteration of the Davidson algorithm, elements of the eigenvectors not corresponding to the core orbitals are set to zero. In this way, the core excited and valence excited states are completely decoupled and the algorithm will (given the appropriate choice of start vectors) converge to the lowest core excited states.

F.3 Experimental and computational details

Calculations were performed at the Stallo computer cluster in Tromsø, Norway [118]. Dalton [99] was used for the CC calculations. Calculations with ADC(2)-x were performed with Q-Chem [153]. Vibrational analysis was performed using the VIBROT [132, 133] program by Sundholm. The CVS approximation [115, 116] was used for core excited state calculations and CC3 was calculated using a new implementation [265] in Dalton. All calculations used the d-aug-cc-pCVQZ [59, 60] basis set. Ionization energies were obtained by including an ultra diffuse orbital and discarding elements in the excitation vectors not referring to this orbital, similar to the CVS approximation. Transition moments were calculated in the length gauge. Theoretical spectra are plotted with an empirical Lorentzian line broadening of 0.06 eV half width at half maximum.

The experimental spectrum was recorded at room temperature at the PLÉIADES soft x-ray beamline, SOLEIL synchrotron, France[134] by measuring the total X-ray induced electron yield. The photon energy resolution was 50 meV. The spectrum was offset-corrected according to Ref. [130].

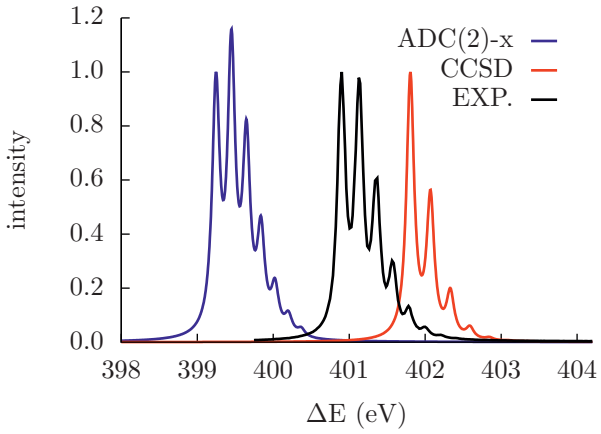


Figure F.3: Comparison of ADC(2)-x, CCSD and experiment. CCSD is too high and ADC is too low.

F.4 Results

In Figure F.1, we compare the first experimental peaks with the theoretical spectrum. Energies of the excited states were computed using CC3 while the transition moments were determined at the CCSD level.

Six electronic states give rise to the peaks in the spectrum. The peaks between 400 and 403 eV are by far the most intense and correspond to excitations from the σ_u core orbital to two sets of equivalent π_g^* orbitals. At about 406.2 eV, the first Rydberg state appears. This is a $^1\Sigma_u^+$ state where an electron is excited from the core σ_u to σ_g^*. The last peak is a combination of two $^1\Pi_u$ states, from core σ_g to π_u^*, and a $^1\Sigma_u^+$, from core σ_g to σ_u^*. Vibrational structure is observed for all these states except the last $^1\Sigma_u^+$ state which is too weak and convoluted with the much stronger $^1\Pi_u$ states. These assignments are consistent with those of Chen *et al.* [130]

Figure F.2 shows the vibrational structure of the first state in closer detail, with the calculated positions and relative intensities of the peaks presented in Table F.1. The absolute error of the $0 \rightarrow 0$ transition is 0.13 eV in CC3 while the distance between the vibrational levels is 0.02 eV too high. Figure F.3 compares the spectra of ADC(2)-x and CCSD to experiment. For ADC(2)-x, the position of the first transition is 1.65 eV too low and the vibrational separation is 0.03 eV too low. Corresponding numbers for CCSD are 0.91 and 0.03 eV, both too high. Equivalent calculations with aug-cc-pCVQZ and aug-cc-pCV5Z showed a decrease in the first excitation energy of 0.04 eV when increasing the basis set. The excitation energy was also calculated using the Lanczos algorithm with and without CVS in the aug-cc-pCVTZ basis. The excitation energy is decreased by 0.02 eV and the intensity

Table F.1: Positions and intenseties of the vibrational structure of the first excited state. Higher vibrational transitions indicated relative to the previous transition.

	ADC(2)-x		CCSD		CC3		Experiment	
transition	ΔE	f	ΔE	f	ΔE	f	ΔE	f
$0 \rightarrow 0$	399.24	1.00	401.80	1.00	401.03	1.00	400.90	1.00
$0 \rightarrow 1$	+0.20	1.14	+0.27	0.52	+0.25	0.65	+0.23	0.95
$0 \rightarrow 2$	+0.20	0.76	+0.26	0.16	+0.25	0.25	+0.23	0.56
$0 \rightarrow 3$	+0.19	0.40	+0.26	0.04	+0.24	0.08	+0.22	0.24
$0 \rightarrow 4$	+0.18	0.18	+0.25	0.01	+0.24	0.02	+0.21	0.09
$0 \rightarrow 5$	+0.18	0.08	+0.25	0.00	+0.23	0.01	+0.21	0.04
$0 \rightarrow 6$	+0.17	0.03	+0.25	0.00	+0.23	0.00	+0.22	0.02

is reduced by 6% when using CVS.

The intensities in the CC3 spectrum fall off too quickly compared to the experiment. They are calculated at the CCSD level and it is possible that CC3 intensities would be closer to experiment. However, the oscillator strength is almost constant across the relevant bond lengths and the structure mostly depend on the potential energy surface. Compared to ADC and CCSD in Table F.1, the ADC potential is much shallower than the CC potentials, resulting in the vibrational levels being closer together. The CCSD potential is deeper than the CC3 potential, resulting in larger energy gaps. The minimum energy of the first core excited CC3 potential is at 1.158 Å, compared to an experimental value of 1.164 Å [130] and 1.187 Å with ADC(2)-x.

Positions and relative intensities of the Rydberg states are presented in Table F.2. The energy gap between the first core excited state and the first Rydberg state is calculated to be 5.20 eV with CC3, 0.06 eV lower than the experimental value. For comparison, the gap is 5.68 eV with ADC(2)-x and 5.57 eV with CCSD. For the gap between the two states, the error is less than 0.01 eV. We note that changing the basis from aug-cc-pCVQZ to d-aug-cc-pCVQZ reduces the gap from the first core excited state by about 0.4 eV for all models.

The energy gaps between the vibrational states are both about 0.03 eV too high for the Rydberg states, indicating that the calculated potential energy surfaces are too deep. Intensities are also too weak for both the $0 \rightarrow 0$ transitions, especially for the $^1\Pi_u$ state. The $0 \rightarrow 1$ transitions, on the other hand, are too strong. In this case it is possible that the error occurs due to the CCSD intensities being too low relative to the first excited state. Furthermore, the $0 \rightarrow 1$ transitions are quite weak and the peaks are convoluted with other states, so it is challenging to obtain accurate intensities, especially for the weak $^1\Sigma_u^+$ state.

We have plotted the CC3 potential energy surfaces in Figure F.4. We have also indicated the relevant vibrational levels with horizontal lines. For the first core excited state, the equilibrium bond length is stretched compared to the ground

Table F.2: Positions and relative intenseties of the vibrational structure of the Rydberg states. Higher vibrational transitions indicated relative to the previous transition.

	CC3		Experiment	
transition	ΔE	$f \times 10$	ΔE	$f \times 10$
$^1\Sigma_u^+$ $0 \rightarrow 0$	406.23	0.29	406.16	0.36
$^1\Sigma_u^+$ $0 \rightarrow 1$	+0.31	0.07	+0.29	0.03
$^1\Pi_u$ $0 \rightarrow 0$	407.22	0.53	407.15	0.91
$^1\Sigma_u^+$ $0 \rightarrow 0$	407.41	0.03	407.36	0.02
$^1\Pi_u$ $0 \rightarrow 1$	+0.32	0.14	+0.28	0.13

state, while for the Rydberg states, it is slightly compressed. When the bond is stretched, the higher $^1\Sigma_u^+$ state goes through a symmetry allowed intersection with the $^1\Pi_u$ state and an avoided crossing with the other $^1\Sigma_u^+$ state. There is also a large number of dark states whose transitions are forbidden by symmetry not shown in the figure.

In Figure F.5 the change in the electron density between the ground state and the first $^1\Pi_u$ state is plotted using Molden [215]. The density is plotted in a plane containing the N_2 molecule and in 3D using isosurfaces. A large density reduction is observed in the cores and a corresponding increase occurs with π symmetry. We note that a superposition of the two core holes is formed because the two atoms are equivalent.

Finally, the first ionization potentials was calculated to be 409.80 and 409.91 eV. This compares to an experimental value of 409.93 eV [283]. We note that CC3 predicts at least one extra vibrational X-ray photoionization spectrum (XPS) peak, however, the resolution of the experiment was not high enough to observe this.

F.5 Conclusion

Calculating core excited states is challenging, not just because of the difficulties of finding the corresponding eigenvalues, but also because of the large relaxation effects that occur. In general, CCSD performs well for valence excited states with typical errors of 0.2 eV compared to CC3 [157]. For core excited states, errors in CCSD of more than 1 eV are common. Furthermore, core excitation processes require much larger basis sets for convergence than valence excitations in our experience. In this paper, we have demonstrated that CC3 can predict spectra with sufficient accuracy to assign peaks to states, but it requires a large basis set. However, basis set requirements can be relaxed by utilizing the fact that the excitation is very local and the large basis set is only needed on the atom being excited. Furthermore, multilevel CC3 can reduce the computational cost by two orders of magnitude [265].

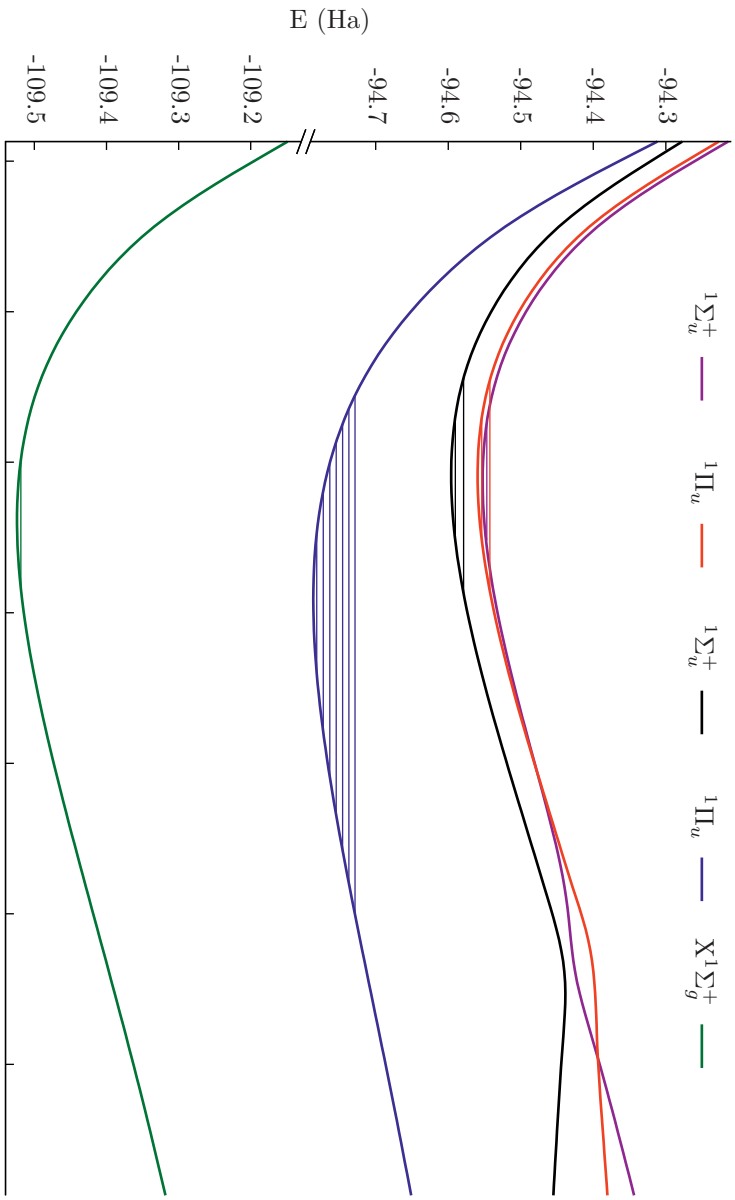


Figure F.4: Potential energy surfaces of excited states and vibrational levels calculated with CC3.

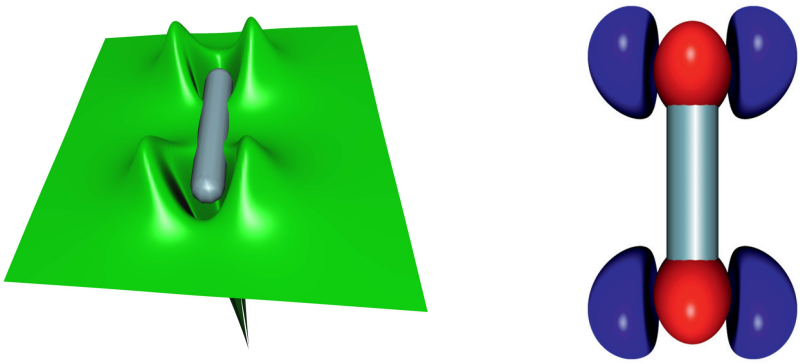


Figure F.5: Electron density change of first $^1\Pi_u$ in the plane and as isosurfaces. Red indicates reduced density and blue increased. Isovalue: 0.01

Vibrational effects can be important in NEXAFS spectra and are required when describing the spectrum of N_2. Both directly in order to describe the vibrational structure, but also due to shifts induced by the zero point energy. In most cases, it is not possible to resolve any vibrational structure in X-ray spectra in detail.

We have not included relativistic effects in the calculations. Using the Douglas-Kroll correction [139, 140], all excitation energies get an increase of 0.2 eV due to contraction of the core orbitals. However, for light elements like Nitrogen, there are important relativistic relaxation effects that will counteract the energy increase [141]. How these effects interact in NEXAFS spectra requires further study.

When calculating NEXAFS, CC3 tends to predict too high excitation energies. This is not the case when simulating XPS where the binding energies tend to be too low. This may be because two indices are fixed, effectively leaving only singles and approximate doubles to describe the relaxation. A potential solution is to calculate the ground state at the CC3 level, but the ionization energies at the full CCSDT level.

Acknowledgments

We thank T. J. Martínez for hosting part of this project at Stanford University and acknowledge computer time from NOTUR through project nn2962k. H.K. acknowledges financial support from the FP7-PEOPLE-2013-IOF funding scheme (Project No. 625321). R.H.M. acknowledges financial support from the COST action "XUV/X-ray light and fast ions for ultrafast chemistry (XLIC)". S.C. acknowledges financial support from the AIAS-COFUND program (Grant agreement No. 609033).

Paper G

Probing ultrafast $\pi\pi^*/n\pi^*$ internal conversion in organic chromophores via K-edge resonant absorption

Manuscript

Probing ultrafast $\pi\pi^*/n\pi^*$ internal conversion in organic chromophores via K-edge resonant absorption

Thomas J. A. Wolf[1], Rolf H. Myhre[2,3], James P. Cryan[1], Sonia Coriani[4,5], Richard Squibb[6], The Thymine Collaboration [1], Henrik Koch[2,3*] and Markus Gühr[1,7*]

[1] *PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, US*

[2] *Department of Chemistry, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway*

[3] *Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California 94305, USA*

[4] *Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Trieste,I-34127, Italy*

[5] *Aarhus Institute of Avanced Studies, University of Aarhus, DK-8000 Århus C, Denmark*

[6] *Department of Physics, University of Gothenburg, Gothenburg, Sweden*

[7] *Department of Physics, University of Potsdam, Potsdam, Germany*

Abstract

Organic chromophores with heteroatoms possess an important excited state relaxation channel from an optically allowed $\pi\pi^*$ to a dark $n\pi^*$ state. We are exploiting the element and site specificity of soft x-ray absorption spectroscopy to selectively follow the electronic change during the $\pi\pi^*/n\pi^*$ internal conversion. As a hole forms in the n orbital during $\pi\pi^*/n\pi^*$ internal conversion, the near edge x-ray absorption fine structure (NEXAFS) spectrum at the heteroatom K-edge exhibits an additional resonance. We demonstrate the concept with the nucleobase thymine, a prototypical heteroatomic chromophore. With the help of time resolved NEXAFS spectroscopy at the oxygen K-edge, we unambiguously show that $\pi\pi^*/n\pi^*$ internal conversion takes place within (60 ± 30) fs. High-level coupled cluster calculations on the isolated molecules used in the experiment confirm the superb electronic structure sensitivity of this new method for excited state investigations.

G.1 Article

The efficient conversion of light into other forms of energy plays a key role in many processes such as photosynthesis or human vision [284, 285]. It is well established that the efficiency of these processes is guaranteed by coupled ultrafast electronic and nuclear dynamics that cannot be described using the Born-Oppenheimer approximation (BOA). The breakdown of the BOA implies that the fundamental details of such mechanisms are notoriously difficult to understand: they proceed on an ultrafast timescale and occur mostly at positions where potential energy surfaces come close or even intersect. From an experimental point of view, it is highly desirable to access the nuclear geometry as well as the electronic degrees of freedom separately, in order to compare them to quantum simulations. The transient nuclear geometry can be best studied by time resolved diffraction techniques using short X-ray [286] or electron pulses [287]. X-ray spectroscopy methods provide a high selectivity on the electronic state, as shown prominently for metal containing molecules at their K- and L-edges [288–290]. We demonstrate the ability of time-resolved soft X-ray spectroscopy to differentiate between distinct electronic excited states of organic molecules, closing an important gap in the classes of molecules to be investigated.

In this article, we concentrate on the internal conversion between excited states of different electronic character, which is a crucial path photoenergy conversion in organic molecules. Organic chromophores exhibit strongly absorbing $\pi\pi^*$ excited states, which can be described in a single electron Hartree-Fock (HF) picture as an electron-hole pair in a formerly occupied and an unoccupied molecular orbital (MO), both with π symmetry. Many of these chromophores, such as azo-switches [286, 291, 292], nucleobases [145, 293–295], and amino acids [296], also contains heteroatoms with electron lone pairs. They therefore exhibit $n\pi^*$ excited states, with a hole in a heteroatom-centered lone pair (n) orbital and an electron in a π^* orbital. The $\pi\pi^*/n\pi^*$ internal conversion provides photochemical pathways to reactions like cis-trans isomerizations and intersystem crossing to the triplet manifold of electronic states governed by the El Sayed selection rules [297]. Unlike $\pi\pi^*$ excited states, $n\pi^*$ states are usually not directly accessible due to low absorption cross-sections from the ground state. The $\pi\pi^*/n\pi^*$ internal conversion through conical intersections is therefore a crucial gateway process for photochemistry.

The preferential localization of the n orbital at the heteroatom has wide-reaching implications for resonant core level spectroscopy using x-rays. In general, near edge x-ray absorption fine structure (NEXAFS) spectra show isolated features due to resonant states below the core ionization edge of an element. Those features are due to transitions from this element's core orbital to an unoccupied valence orbital, for instance a π^* orbital. The core orbital is confined to the immediate vicinity of one particular atom (see Fig. G.1A). The core-to-valence absorption cross-section is strongly dependent on the spatial overlap between the core and the empty valence orbital [298, 299]. Since core ionization potentials of carbon, nitrogen and oxygen are more than 100 eV apart, element- and site-specific probing of the

local electronic structure in an organic molecule is possible with soft x-rays. In the case of excited states, the electron hole in a formerly occupied orbital enables an additional NEXAFS resonance. The spatial overlap makes a 1s-n transition from the strongly localized heteroatom 1s core level to the electron hole of an $n\pi^*$ state more intense than the 1s-π transition to the delocalized π hole of a $\pi\pi^*$ state. Therefore, we expect a negligibly weak time-resolved (TR) NEXAFS signature from the photoexcited $\pi\pi^*$ state to transform into a strong $n\pi^*$ state signature as the molecule undergoes $\pi\pi^*/n\pi^*$ internal conversion, largely independent of geometry changes during the dynamics. For the isolated molecules used in this study, computationally demanding high level coupled cluster (CC) simulations [116, 160, 265] are now feasible and confirm the spectroscopic attribution based on orbital localization.

We test and exemplify the sensitivity of TR-NEXAFS spectroscopy on the $\pi\pi^*/n\pi^*$ transition in the benchmark chromophore thymine, since there is a rich literature on its excited states (see Refs. [145, 293, 294, 300]) and citations therein). Thymine exhibits two high-lying occupied MOs, an oxygen localized n-orbital and a delocalized π-orbital (see Fig. G.1A). Its lowest unoccupied MO (π^*) is similar to the π-orbital in delocalization. The molecule can be excited at 267 nm to a $\pi\pi^*$ state; the lower-lying $n\pi^*$ state is optically dark.

Figure G.1B shows a reduced potential energy sketch from the CC simulations along the two nuclear coordinates, which are expected to be most relevant for the molecular dynamics. In contrast to earlier theoretical studies [143, 144] the $\pi\pi^*/n\pi^*$ conical intersection seam is not isolated by a barrier from the Franck-Condon (FC) point, but directly accessible. After photoexcitation, the nuclear wavepacket is driven out of the FC region by a gradient along the C(5)-C(6) bond elongation towards a saddle point. On its way, it encounters the $\pi\pi^*/n\pi^*$ coonical intersection seam, which indicates a rapid (on a 100 fs scale) internal conversion. In the $n\pi^*$ excited state, a local minimum can be reached from the $\pi\pi^*/n\pi^*$ conical intersection by O(8)-C(4) bond elongation [301].

The $\pi\pi^*$ relaxation in thymine has been experimentally investigated using many methods available in ultrafast technology [145, 294] including our own DUXAP study, where we investigated excited state nonresonant Auger spectra at the oxygen K-edge [110]. It is challenging to attribute signals in ultrafast photoelectron, photoion, absorption, or DUXAP spectroscopy directly to a particular process, like internal conversion, since changes in both, the electronic structure as well as the nuclear geometry, influence the observables. In our own study we were e.g. preferentially sensitive to local C-O bond length changes during relaxation of the $\pi\pi^*$ state. We demonstrate in the following, that our new TR-NEXAFS technique is strongly and selectively sensitive to the ultrafast $\pi\pi^*/n\pi^*$ internal conversion.

The experimental ground state NEXAFS spectrum of thymine is shown in black in Fig. G.2A. It exhibits a double peak π^* resonance. Based on our CC calculations, and in agreement with earlier studies [112], we assign the lower energy peak at 531.4 eV to a linear combination of HF single electron excitations from the O(8)

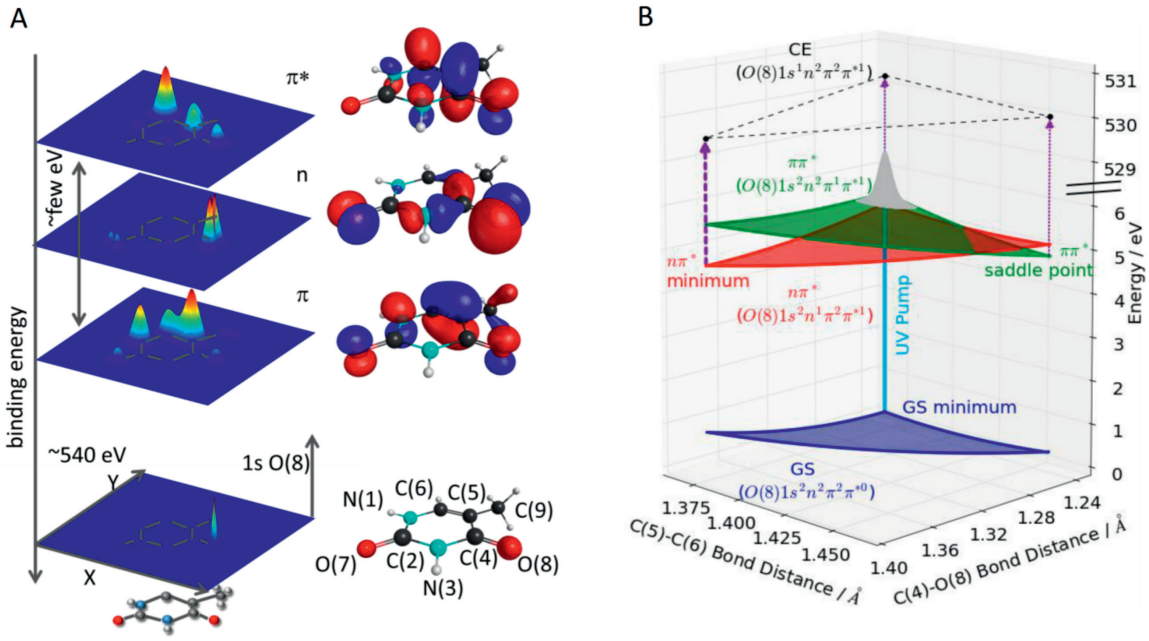


Figure G.1: (A) Isosurface representations (right) and electron density projections onto the molecular plane for the three valence orbitals (Hartree-Fock/6-311G) involved in the characters of the two lowest lying excited states of thymine and a core orbital localized at oxygen(8). The electron density at the position of the core orbital differs strongly for the different valence orbitals. (B) Results from our coupled cluster investigation of the excited state topology along the two most relevant degrees of freedom for relaxation into the $n\pi^*$ minimum. All states are labeled with their electron configuration. Ultraviolet (UV) excitation of the ground state (GS) places a nuclear wavepacket (gray) on the $\pi\pi^*$ excited state. It relaxes through a conical intersection to a minimum in the $n\pi^*$ excited state. According to our calculations, only one core excited state (CE) characterized by an excitation at O(8) is relevant for interpretation of our near-edge x-ray absorption fine structure (NEXAFS) results.

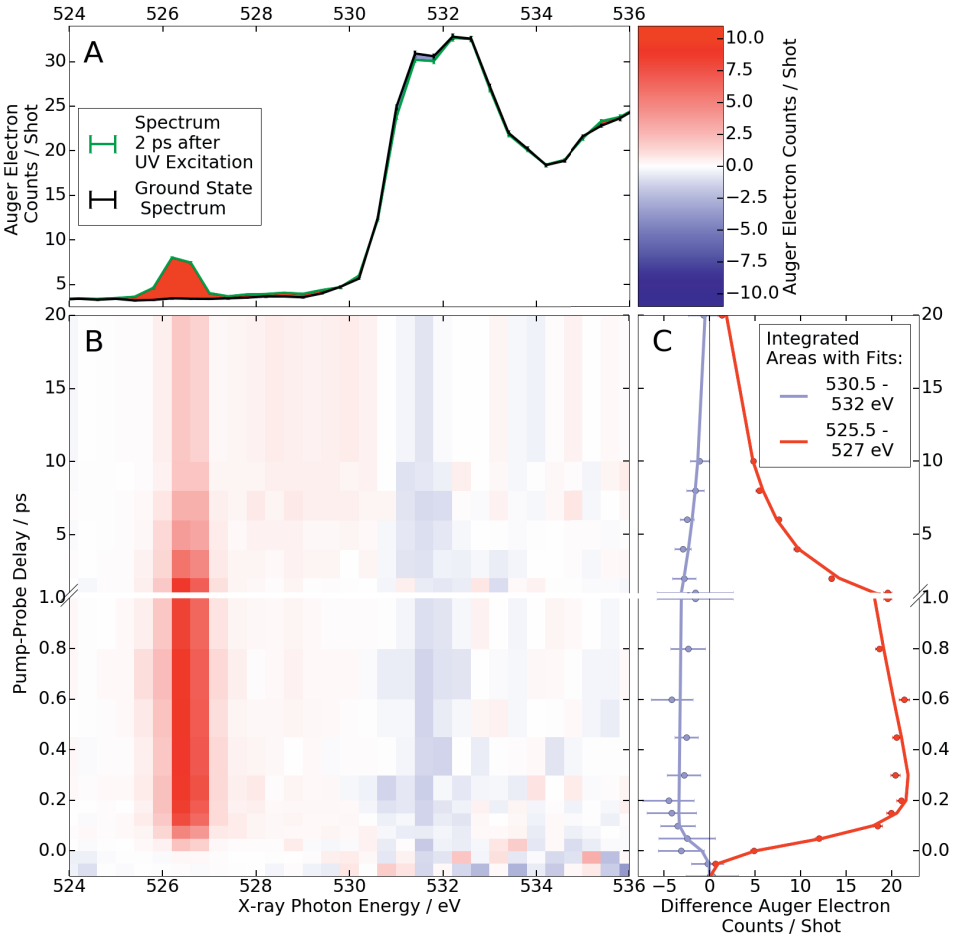


Figure G.2: (A) Representative NEXAFS spectra 2ps after UV excitation and without UV excitation. UV-induced increase in intensity is marked red, UV-induced decrease is light blue. UV excitation leads to the appearance of a new spectral feature around 526.4 eV and a bleach of the ground state π^* resonance at 531.4 eV. (B) False-color plot of time-dependent NEXAFS difference spectra (see color bar in the upper right corner). The UV-induced features at 526.4 eV and 531.4 eV are clearly visible throughout the positive pump-probe delays. (C) Time-dependence of the UV-induced features with fits based on a rate equation model.

1s orbital to several unoccupied π^* orbitals with significant contributions from the aforementioned π^* MO. The linear combination is such that the core excited state possesses a high degree of electron localization at O(8), thus the strong absorption cross section in the Mbarn regime (see a discussion in Section G.2.6). The higher energy peak at 532.2 eV corresponds to an excitation from the O(7) 1s orbital to a different linear i combinationof π^* MOs [112]. The increase in intensity at photon energies beyond the π^* resonances is predominantly due to a smooth feature of K-edge ionization at 537 eV and additional weak resonant transitions [112].

The NEXAFS spectrum taken 2 ps after UV excitation is shown in green in Fig. G.2A. It is a superposition of the excited state spectrum and the ground state spectrum, which is weakened by transfer of an estimated 13 % of the population (see Section G.2.5) to the excited state. The excited state spectrum is redshifted to lower photon energies with respect to the ground state, which is obvious by the background-free signature at 526.4 eV. This signature must be a new core excitation channel either to the n or π electron hole. An additional signature of UV excitation, the intensity reduction in the area of the π^* resonances, is the result of a bleach of the ground state spectrum almost entirely compensated by the redshift of the smooth K-edge ionization feature in the excited state spectrum. The effect is therefore only visible where the ground state exhibits the strongest intensity modulations, i.e. the π^* resonance. It is therefore a direct signature of the ground state depopulation and largely independent of any following excited state dynamics.

The time-dependence of the difference signal (x-ray absorption with UV minus x-ray absorption without UV) is shown in Fig. G.2B. The spectrally integrated time trends at ground state bleach and excited state feature are shown in Fig. G.2C. The temporal onset of the excited state feature exhibits a delay ((60±30) fs according to a rate equation fit, see Section G.2.5) with respect to the temporal overlap between UV and x-ray pulses, which is marked by the bleach onset. The intensity of the 526.4 eV feature is only due to 13 % of the population in the ground state NEXAFS spectrum. Its absorption cross-section is similar to the π^* resonance. Therefore, it must be likewise due to a localized transition, which is the signature of the $n\pi^*$ state, not the $\pi\pi^*$ state. Accordingly, the delay of the $n\pi^*$ signature of (60 ± 30) fs directly reflects the nuclear wavepacket dynamics to access the $\pi\pi^*/n\pi^*$ conical intersection seam, in agreement with our earlier DUXAP study.

Our intuitive interpretation is supported by CC NEXAFS spectra simulations of the ground state and of the excited states at the minimum and saddle point geometries identified in Fig. G.1B. We compare calculated to experimental spectra of the ground state and 2 ps after UV excitation in Fig. G.3. All three simulated excited state spectra exhibit their lowest energy resonance around 526.4 eV. In all cases, the final state is the same O(8)-centered core-excited state (CE). As expected, the oscillator strength in the $n\pi^*$ state beats the $\pi\pi^*$ state by a factor of 40, almost independent of the molecular geometry (see Section G.2.5). We scaled the simulated excited state spectra to the estimated ratio of 13 % excited molecules. Only the simulated $n\pi^*$ state spectrum shows a comparable intensity at the 526.4

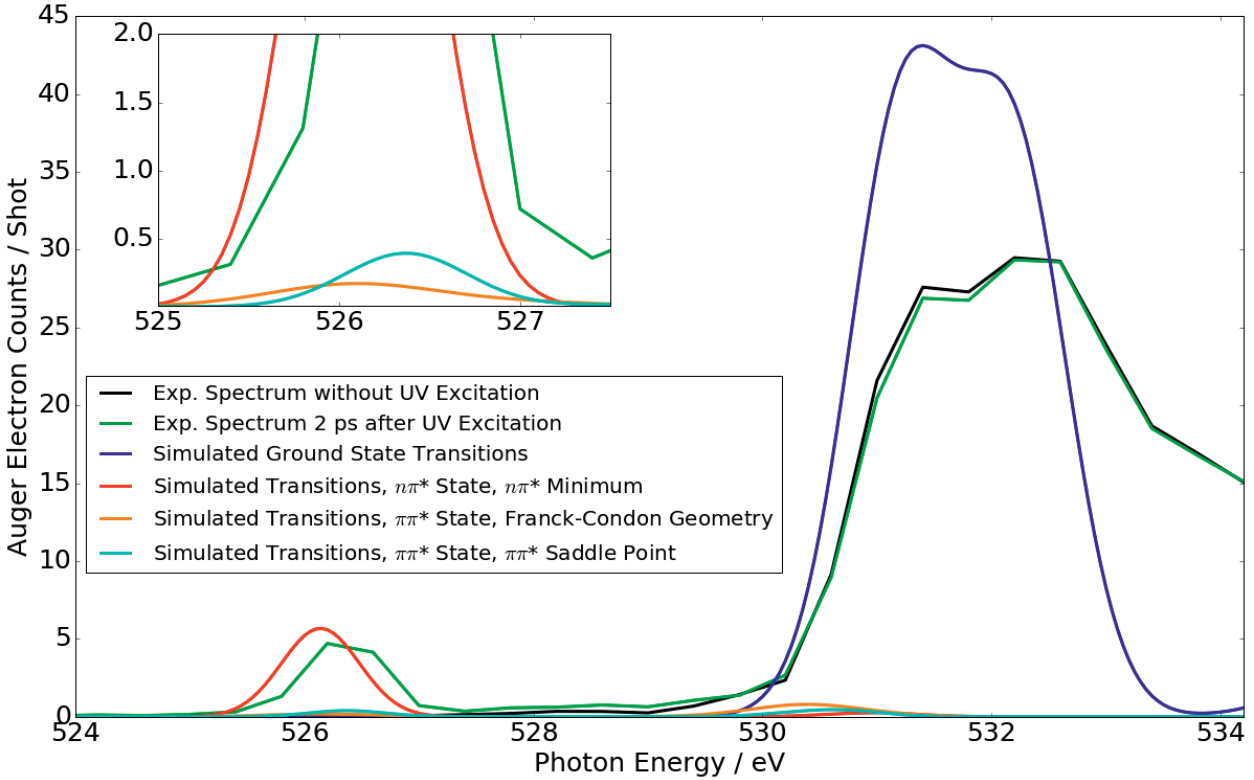


Figure G.3: Simulated spectra of the ground state, the $\pi\pi^*$ state in the Franck-Condon region and at the saddle point, and the nπ^* state at its minimum.Simulated excited state spectra are scaled assuming 13 % excitation. For comarison, the experimental ground state spectrum and the experimental spectrum 2 ps after UV excitation are shown. Relative intensities of simulated and experimental ground state spectra are adjusted to an equal peak area of the π^* resonances. The inset shows a detailed view of the intensity relations at the excited state feature. Contributions from K-edge ionization were not included in the simulations.

eV position.

Thus, with the TR-NEXAFS method presented here, we can confirm population of the $n\pi^*$ state through a directly accessible conical intersection within 60 fs. Comparison of experimentally observed and calculated NEXAFS absorption intensities suggests that internal conversion into the $n\pi^*$ state is a major channel in the relaxation dynamics of thymine. The $n\pi^*$ signature shows a biexponential decay with time constants of (1.9 ± 0.1) ps and (10.5 ± 0.2) ps. This supports a consecutive relaxation process via a level with $n\pi^*$ character to a final level of non-$n\pi^*$ character, to which our method is insensitive. The $n\pi^*$ level is most probably another minimum in the singlet $n\pi^*$ state, since intersystem crossing to a triplet $n\pi^*$ state is forbidden by the El Sayed selection rule. The transition to the final state can be either intersystem crossing to a nearby triplet $\pi\pi^*$ state with strong spin orbit coupling [302] or internal conversion to the ground state [303], which is supported by the recovery of the ground state bleach in our data.

In conclusion, we demonstrated with this work a novel method to selectively investigate ultrafast $\pi\pi^*/n\pi^*$ internal conversion. We use the dominating absorption strength of the oxygen 1s-n resonance to directly monitor this nonadiabatic process. An alternative method, however less selectively sensitive to electronic structure changes, is time-resolved photoelectron spectroscopy using extreme ultraviolet pulses. Here one relies on spectrally resolving ionic continua that single out certain valence states [304], thus exhibiting nonadiabatic transitions by kinetic energy or angular distribution changes [305, 306]. An assignment is, however, only possible by high-level simulations of excited state ionization cross-sections, whereas we confirm the validity of our intuitive orbital-based assignment of core excitations by quantitative CC simulations. Our results prove that the method already works reliably under the current, still challenging conditions of x-ray free-electron laser experiments with low repetition rates and high temporal and spectral jitter. The method has the potential to become a standard tool for ultrafast investigations at the upcoming second generation of ultrafast x-ray sources with up to MHz repetition rates and improved shot-to-shot stability.

Acknowledgments

This work was supported by the AMOS program within the Chemical Sciences, Geosciences, and Biosciences Division of the Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy. Parts of this research were carried out at the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory. LCLS is an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. We thank NOTUR for computer time through Project nn2962k. MG acknowledges funding via the Office of Science Early Career Research Program through the Office of Basic Energy Sciences, U.S. Department of Energy and NB under grant No. DE-SC0012376. MG

is now funded by a Lichtenberg Professorship from the Volkswagen foundation. TJAW thanks the German National Academy of Sciences Leopoldina for a fellowship (Grant No. LPDS2013-14). HK acknowledges support from FP7-PEOPLE-2013-IOF (Project No. 625321). JG acknowledges the European Research Agency via the FP-7 PEOPLE Program (Marie Curie Action 298210). SC acknowledges support from the AIAS-COFUND program (Grant Agreement No. 609033). RF would like to acknowledge financial support from the Knut and Alice Wallenberg Foundation, Sweden, and the Swedish Research Council.

G.2 Supporting material

G.2.1 Experimental Methods

The experiment was performed at the LCLS free electron laser facility, SLAC National Accelerator Laboratory, at the soft x-ray (SXR) instrument [307, 308]. A schematic representation of the experimental setup is shown in Fig. G.4. Thymine was purchased from Sigma Aldrich and evaporated by an effusive oven into an ultra high vacuum chamber at a temperature of 160°C leading to a sample density of 1012 cm^{-3} in the overlap region of optical and x-ray laser [110, 309]. Molecules were excited by 267 nm pulses with 70 fs duration and a focus diameter of 100 μm FWHM. Soft x-ray pulses with 70 fs duration and a focus diameter of 70 μm FWHM were used to probe the sample in the oxygen K-edge spectral region from 520 to 550 eV by simultaneously tuning the FEL and the monochromator of the SXR instrument with an energy resolution of < 0.5 eV [310]. The intensity of the essentially background-free transient feature at 526.4 eV was measured for a wide range of UV pump intensities, to make sure the experiment took place in the linear absorption regime well below saturation. Temporal and spatial overlap of UV and SXR pulses was optimized to a bleach in the Auger spectra of thymine induced by photofragmentation at high UV intensities. Oxygen 1s Auger spectra were recorded with the 2m long LCLS-FELCO magnetic bottle spectrometer [311]. The photon energy dependent absorption cross-section of the sample is proportional to the integrated Auger electron yield. SXR pulses were delayed with respect to UV pulses between -200 fs and 20 ps. To achieve NEXAFS difference spectra, UV laser pulses were blocked on a shot-by-shot basis. LCLS pulses are strongly fluctuating in intensity and relative arrival time. Therefore, both parameters were recorded on a shot-by-shot basis by an optical x-ray cross-correlator [312] and a gas detector after the monochromator, respectively. The dataset was resorted into ≥ 50 fs delay bins and several x-ray intensity bins. Difference spectra from different x-ray intensity bins were averaged

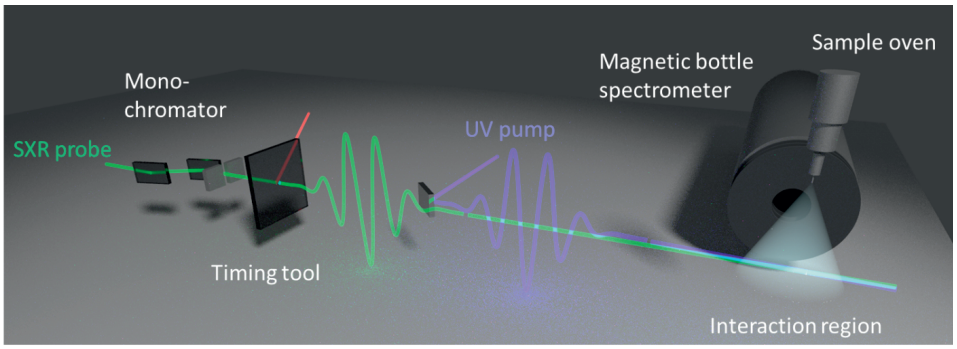


Figure G.4: Schematic representation of the experimental setup. Broad-bandwidth ultrashort soft X-ray (SXR) pulses (green) are monochromatized and focused into the interaction region of the experimental chamber. There, they are quasi-collinearly overlapped with ultrafast UV pulses with a wavelength of 267 nm (violet). The sample thymine is evaporated into the interaction region by an in-vacuum oven. The relative timing between UV and soft X-ray pulses is measured on a shot-by-shot basis using the SXR timing tool to compensate for the 200 fs timing jitter. Auger electron spectra from core-excited thymine molecules are detected in a magnetic bottle photoelectron spectrometer. NEXAFS spectra can be generated by measuring the photon energy dependent integrated Auger electron yield.

G.2.2 Theoretical Methods

The thymine ground state geometry (Table G.1) was optimized with CCSD(T)/aug-cc-pVDZ using CFOUR [122]. Excited state geometries (Table G.2 and G.3) were optimized at the EOM-CCSD/aug-cc-pVDZ level employing Q-Chem [153]. No symmetry restrictions were applied for geometry optimizations. Valence excitation energies were obtained with CC3 using the aug-cc-pCVDZ basis on the oxygens and the aug-cc-pVDZ basis on the other atoms. We employed a newly developed implementation in Dalton [24, 87, 99, 265]. Oxygen 1s to valence excitation energies and oscillator strengths were computed at the CCSD level of theory with the same basis as for the valence excitations using a newly developed linear response code employing core-valence separation and implemented in Dalton [31, 32, 34, 116]. The core to valence excitation energies were offset-corrected by benchmark calculations of the lowest core to valence excitation energies at the CC3/aug-cc-pCVTZ/aug-cc-pVDZ level. With this correction we achieve quantitative agreement with the NEXAFS transition energies within the experimental error bars on a purely *ab initio* basis. The theoretical core excitation energies are not corrected for relativistic effects and we estimate the effect to increase excitation energies by 0.1-0.3 eV.

Thymine exhibits C_s symmetry in the ground state. The two lowest-lying excited states have different representations, $A''(n\pi^*)$ and $A'(\pi\pi^*)$, and the $\pi\pi^*/n\pi^*$ conical intersection is symmetry allowed. We note that no complex eigenvalues of the Jacobian matrix were encountered in the vicinity of the conical intersection seam

[36, 37]. In contrast to earlier studies [143, 144, 313–315], we could not identify a minimum in the $\pi\pi^*$ state. Instead we found a saddle point geometry with C_s symmetry, which is directly accessible from the Franck-Condon region. The energy lowering degrees of freedom of the saddle point are out of plane bending, as confirmed by frequency calculations, which were performed for all observed stationary points [316, 317], (Table G.4). The $n\pi^*$ minimum geometry is distorted from C_s symmetry with O(8) out of the plane. We encountered the $\pi\pi^*/n\pi^*$ conical intersection seam in between the Franck-Condon point and the $\pi\pi^*$ saddle point in close proximity to the latter (energy difference < 0.03 eV). Based on calculated core excitation energies and oscillator strengths, NEXAFS spectra were simulated by convoluting theoretical stick spectra with Gaussians to account for peak broadening and the experimental energy resolution.

G.2.3 Localization of excitations and the single electron picture

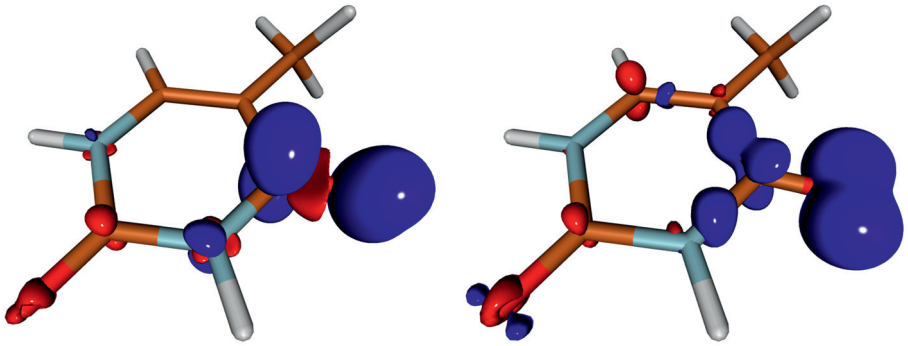
For our intuitive prediction of the expected intensities of excited state NEXAFS features we rely on the Hartree-Fock (HF) molecular orbital (MO) based one-electron picture. In this framework, core-excited as well as valence-excited states are described as a single electron excitation from an occupied to an unoccupied HF-MO. Despite its success in the case of relative intensities for core-excitations from the $\pi\pi^*$and $n\pi^*$ states, it fails in predicting relative intensities for core-excitations from the $n\pi^*$ state and the ground state. The ground state core-excitation leads to the π^* MO in the one-electron picture, which is highly delocalized and exhibits only weak density at the oxygens. Nevertheless, both coupled cluster simulations and experimental results predict the transition intensity to be comparable to the $n\pi^*$ state core-excitation, which involves the strongly localized n MO.

The reason for the failure of the MO picture lies in the HF formalism, which only optimizes occupied MOs. The employed coupled cluster methods calculate valence and core excited states by optimizing a series expansion of electron excitations into the virtual MOs inherently leaving the single electron MO picture. The most appropriate, but less intuitive way to inspect these excitations is therefore to look at electron density changes instead of orbitals (see Fig. G.5). The density changes, however, qualitatively agree with the MO picture in cases of transitions between MOs which are occupied in the HF reference wavefunction.

G.2.4 Rate equation model

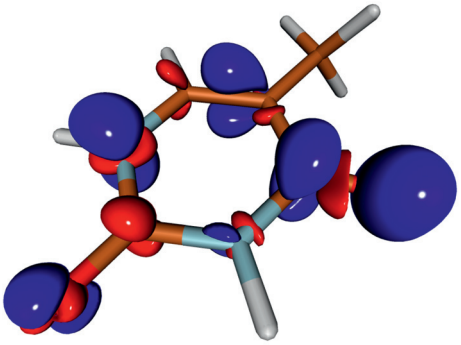
To analyze the transient 526.4 eV feature in the TR-NEXAFS spectra, a rate equation model was employed, which assumes the following chain of subsequent excited state single exponential population transfers with time constants τ_1 to τ_3:

$$\pi\pi^* \text{ state} \rightarrow n\pi^* \text{ state} \rightarrow (n\pi^*)' \text{ state} \rightarrow \text{final state}$$



GS $\rightarrow$ CE

nπ^* $\rightarrow$ CE



$\pi\pi^*$ $\rightarrow$ CE

Figure G.5: Visualization of electron density changes for core excitations from the ground state, the nπ^* state and the $\pi\pi^*$ state to the lowest O(8) 1s core-excited state. In the case of core excitation from the nπ^* state and the $\pi\pi^*$ state, the electron density changes agree with the predictions from the Hartree-Fock molecular orbital based single electron picture. The density change for the nπ^* state core excitation is strongly localized at O(8), since two MOs with strong localization at this oxygen are involved. The density change for the $\pi\pi^*$ state core excitation is delocalized, since it involves delocalized π^* MO apart from the localized O(8) 1s MO. The electron density change for the core excitation from the ground state disagrees with the predictions from the HF MO picture. The latter involves the localized O(8) 1s MO and a delocalized π^* MO. The electron density change, however, is strongly localized due to linear combination of single electron transitions to several π^* MOs.

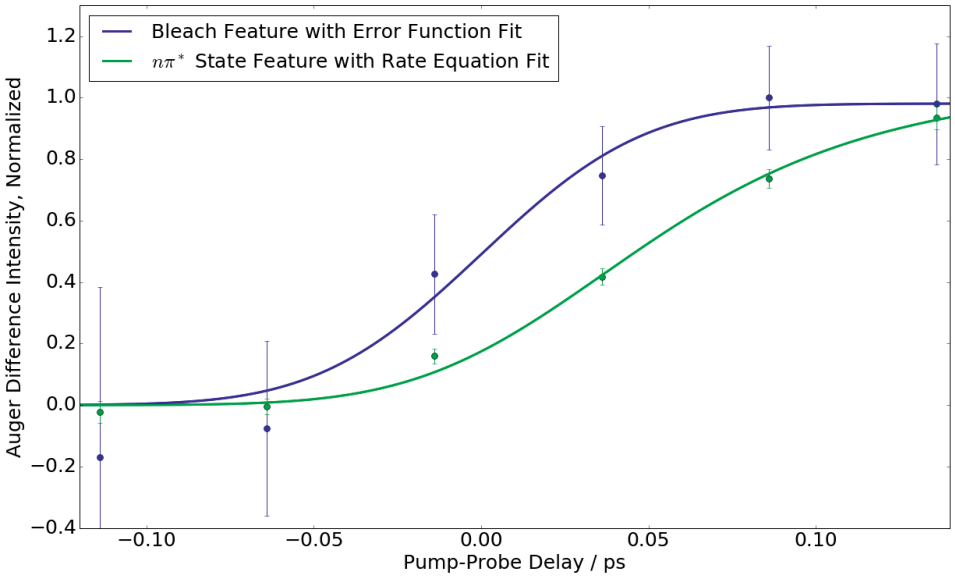


Figure G.6: Representation of the fit of the bleach signal (blue) used to determine the experimental response function and time zero. The experimental data are extracted from a region of interest in the Auger spectra from the photon energy region, where the bleach is observable in the NEXAFS spectra. For better comparison with the analogue dataset from the region of the nπ^* feature (green), the bleach dataset is inverted. Both datasets are normalized to the maximum modulation. The time-dependence of nπ^* feature is fitted with the rate equation model. The delay between the bleach and the nπ^* feature is clearly visible.

Since the 526.4 eV signature decays in a bi-exponential fashion, the step labeled as (nπ^*) had to be included. The character of that step cannot be completely determined by the present results. Its transition moment is only 42% of the nπ^* state according to our fit of the experimental data. It is nevertheless very likely also of nπ^* character, since the calculated oscillator strength for core excitation from the $\pi\pi^*$ state is only 2.5% of the oscillator strength from the nπ^* state, one order of magnitude lower. One possible explanation is that thymine relaxes thermally out of the nπ^* minimum identified in our calculations into a lower lying minimum, from which it then relaxes to the final state, which is either the ground state or a triplet state.

Since the bleach signature only decays slowly within the investigated delay-time window, it is ideal to exctract the exact time zero and the instrument response function (90 fs) using an error function fit. To decrease the noise level, a region of interest with the strongest UV-induced modulations was identified in the resonant Auger spectra from the ground state π^* resonances. The plotted intensities in Fig. G.6 refer to the integrated UV-induced changes in this region of interest. For comparison, Fig. G.6 also shows the analogue signal for the 526.4 eV, which was

fitted (see Fig. G.2 and Fig. G.6) with a weighted sum of the time-dependent populations of the $n\pi^*$ and $(n\pi^*)'$ steps, convoluted with the instrument response function ($g(t)$).

$$I(\Delta t) = g(t) \otimes \text{heavyside} \times \left(\mathrm{I}_1 \frac{\tau_2}{\tau_2 - \tau_1} \left(\mathrm{e}^{-\frac{\Delta t}{\tau_2}} - \mathrm{e}^{-\frac{\Delta t}{\tau_1}} \right) + I_2 \frac{\tau_3}{\tau_2 - \tau_1} \left(\frac{\tau_2}{\tau_3 - \tau_2} \left(e^{-\frac{\Delta t}{\tau_3}} - e^{-\frac{\Delta t}{\tau_2}} \right) - \frac{\tau_1}{\tau_3 - \tau_1} \left(e^{-\frac{\Delta t}{\tau_3}} - e^{-\frac{\Delta t}{\tau_1}} \right) \right) \right) \tag{G.1}$$

The delay between the onsets of the bleach and the 526.4 eV feature is clearly visible in Fig. G.6.

G.2.5 Excited state population analysis

Calculated and experimental intensities of the $n\pi^*$ feature and the ground state π^* resonance have to be compared, to estimate, which fraction of the excited state population is observable in the $n\pi^*$ feature. For this, we chose the same NEXAFS spectra of the ground state and 2 ps after UV excitation as in Fig. G.2A. The ratio of integrated peak areas between the $n\pi^*$ (I2) and ground state π^* resonance features (I1) is 0.053. These can now be compared to the ratio of calculated transition moments for the $n\pi^*$ feature ($\sigma 2$) and the ground state π^* resonance ($\sigma 1$) 0.65. Assuming 100% population transfer from the $\pi\pi^*$ state, the rate equation model predicts the $n\pi^*$ level to contain 36% and the $(n\pi^*)'$ level 57% of the overall relative excited state population (P0). According to the rate equation fit, the transition moment of $(n\pi^*)'$ level $\sigma 2'$ is 42 % of the $n\pi^*$ transition moment $\sigma 2$. This leads to the relation for I2

$$I_2 = f \times P_0 \times \sigma_2 \times (0.36 + 0.57 \times 0.42) \tag{G.2}$$

where f is a conversion factor between experimental intensities and calculated transition moments. It can be evaluated by

$$f = \frac{I_1}{\sigma_1} \tag{G.3}$$

Evaluating both equations gives a value of 13 % for the overall excited state population P_0 relative to the ground state. This moderate excitation ratio fits well together with our expectations of the excitation ratio based on our scan of the UV intensity dependence of the $n\pi^*$ feature intensity. Comparison with the intensity dependence furthermore supports the initial assumption that the relaxation to the $n\pi^*$ state observed in the present experiment is a major channel for the $\pi\pi^*$ population.

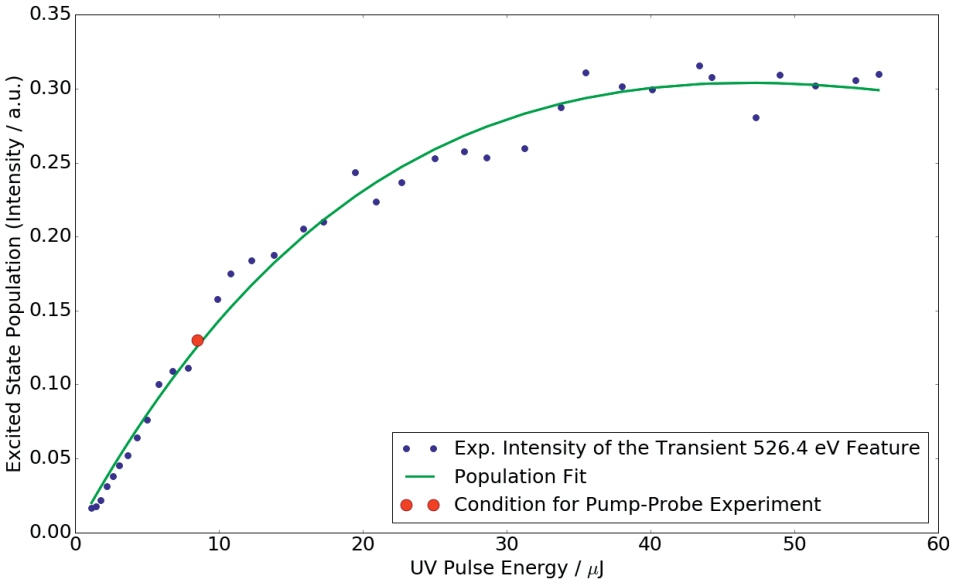


Figure G.7: Intensity of the transient 526.4 eV feature for different UV intensities together with a rate equation fit. The ordinate is rescaled to match the estimated 13 % excited state population for the UV intensity used in the pump–probe experiment.

G.2.6 Intensity scan

To confirm that UV excitation takes place in the linear regime, we investigated the dependence of the nπ^* feature intensity on the UV intensity (see Fig. G.7). At low UV intensities, the nπ^* feature intensity has approximately a linear response. At high intensities, saturation is observable. The most relevant processes contributing to the excited state population are single photon excitation from the ground state and further single photon excitation from the excited state i.e. sequential or resonance-enhanced two photon excitation. The dependence of the relative population P_0 in the excited state on the photon flux F is therefore

$$P_0 = \frac{\sigma_1}{\sigma_1 - \sigma_2}\left(e^{-\sigma_2 \times F} - e^{-\sigma_1 \times F}\right) \tag{G.4}$$

where σ_1 and σ_2 are the ground state and excited state absorption cross-sections. The value of σ_1 is approximately 30 MBarn, the value of σ_2 is unknown. Since absorption of two photons brings the molecule very close to the ionization threshold where the density of states is particularly high, the value of σ_2 can be expected to be higher than σ_1. The absolute upper limit for relative excited state population can be estimated by neglecting sequential two photon excitations i.e. by setting σ_2 to zero. In this case, the saturation value refers to 100% population in the excited state. The UV intensity for the pump-probe experiments leads to 24% of this saturation value. Assuming any value higher than 0 for σ_2 reduces the population

in the excited state at saturation. Assuming 13% excitation at the intensity of the pump probe experiments based on the comparison between experimental and calculated intensities yields a value for σ_2, which is 1.5 times σ_1.

G.2.7 Theoretical results

Table G.1: Ground state minimum geometry (Å).

atom	X	Y	Z
C	−1.6332	0.0426	−0.0001
C	0.7570	0.8384	−0.0001
C	1.1979	−0.5713	0.0000
C	0.2400	−1.5432	0.0000
C	2.6817	−0.8510	0.0002
N	−0.6415	1.0215	−0.0007
N	−1.1219	−1.2509	−0.0002
O	−2.8353	0.2919	0.0005
O	1.5118	1.8105	0.0003
H	−0.9734	1.9835	−0.0002
H	0.4869	−2.6087	0.0000
H	−1.8153	−1.9901	0.0005
H	3.1589	−0.4028	−0.8884
H	3.1587	−0.4025	0.8887
H	2.8721	−1.9376	0.0005

Table G.2: $n\pi^*$ state minimum geometry (Å).

atom	X	Y	Z
C	1.6756	0.0108	0.0230
C	−0.6393	0.8270	−0.0428
C	−1.1709	−0.4819	−0.0168
C	−0.2637	−1.5185	0.0106
C	−2.6680	−0.6951	−0.0181
N	0.7453	1.0515	0.1200
N	1.1227	−1.2492	0.0249
O	2.8776	0.2283	−0.0350
O	−1.3720	1.9491	0.1211
H	1.1236	1.9641	−0.1066
H	−0.5521	−2.5678	−0.0154
H	1.7941	−2.0025	−0.0250
H	−3.1289	−0.2542	0.8817
H	−3.1341	−0.2281	−0.9015
H	−2.8981	−1.7710	−0.0329

Table G.3: $\pi\pi^*$ state saddle point geometry (Å).

atom	X	Y	Z
C	1.6623	0.0922	0.0003
C	−0.6981	0.8798	−0.0002
C	−1.1498	−0.4710	−0.0003
C	−0.1879	−1.5766	−0.0009
C	−2.6197	−0.7577	0.0004
N	0.7372	1.0841	−0.0013
N	1.1229	−1.2303	0.0007
O	2.8833	0.2096	0.0008
O	−1.4102	1.9113	0.0002
H	1.0655	2.0428	−0.0010
H	−0.4451	−2.6345	−0.0012
H	1.8537	−1.9394	0.0014
H	−3.1038	−0.3058	0.8849
H	−3.1047	−0.3049	−0.8832
H	−2.8159	−1.8424	−0.0001

Table G.4: Vibrational frequencies, ν (cm^{-1}), and IR intensities, I (km/mol)

GS		$n\pi^*$		$\pi\pi^*$	
ν	I	ν	I	ν	I
100	0.00	82	2.64	292i	31.79
134	0.72	100	1.89	126i	0.25
143	0.01	135	0.51	94i	13.22
269	2.41	181	19.25	129	0.54
279	0.12	247	4.70	153	0.30
376	22.94	262	2.22	271	7.96
381	21.25	337	4.95	300	5.65
451	18.44	406	30.45	372	18.71
525	52.61	460	21.14	440	35.66
536	6.68	468	91.00	441	76.76
594	1.37	510	15.10	496	11.09
665	88.91	521	47.83	578	0.36
730	4.42	576	6.25	610	0.16
735	25.94	637	47.10	662	4.14
754	1.88	741	33.82	726	12.06
797	3.42	753	3.86	754	65.88
875	19.64	794	7.54	768	9.75
960	11.63	942	26.29	925	17.76
1011	1.94	1012	1.58	968	4.02
1052	0.12	1060	0.32	996	0.39
1154	9.39	1146	62.73	1150	59.95
1199	154.13	1211	44.59	1185	30.92
1249	3.83	1243	13.73	1254	12.70
1373	12.10	1283	5.37	1319	32.57
1386	0.06	1390	84.14	1351	11.99
1403	6.08	1419	0.70	1387	90.04
1424	66.38	1432	13.49	1412	41.87
1453	6.06	1476	22.63	1443	7.86
1477	1.66	1476	6.40	1466	22.10
1499	111.43	1492	0.86	1480	36.99
1699	0.03	1516	27.52	1524	1.79
1739	555.62	1627	45.54	1620	263.15
1778	817.99	1800	742.28	1762	320.77
3034	23.89	3048	24.42	3022	14.56
3109	10.51	3115	13.00	3080	11.34
3123	14.47	3144	13.12	3121	14.83
3212	4.81	3265	1.64	3275	5.94
3594	60.42	3637	70.62	3588	68.66
3643	96.73	3680	76.92	3650	99.19

Table G.5: Thymine CCSD excitation energies, (ΔE), eV, and intensisties ($I, \times 10$) for the three different geometries, aug-cc-pCVDZ/aug-cc-pVDZ.

	GS		$n\pi^*$		$\pi\pi^*$	
	ΔE	I	ΔE	I	ΔE	I
S_1	5.10 ($n\pi^*$)	0.00	3.84 ($n\pi^*$)	0.00	4.48 ($\pi\pi^*$)	2.19
S_2	5.37 ($\pi\pi^*$)	1.95	4.78 ($\pi\pi^*$)	2.41	4.74 ($n\pi^*$)	0.00
S_3	5.77 (πn^*)	0.00	5.51 ($\pi\pi^*$)	0.17	5.37 (πn^*)	0.01
S_4	6.47 ($n\pi^*$)	0.00	5.83 (πn^*)	0.00	6.15 (πn^*)	0.00

Table G.6: Thymine CC3 valence excitation energies, eV, for the three different geometries, aug-cc-pCVDZ/aug-cc-pVDZ. Excited state character in parenthesis.

	GS	$n\pi^*$	$\pi\pi^*$
S_1	4.90 ($n\pi^*$)	3.69 ($n\pi^*$)	4.29 ($\pi\pi^*$)
S_2	5.13 ($\pi\pi^*$)	4.61 ($\pi\pi^*$)	4.40 ($n\pi^*$)
S_3	5.65 (πn^*)	5.12 ($\pi\pi^*$)	5.26 (πn^*)
S_4	6.17 ($\pi\pi^*$)	5.70 (πn^*)	5.60 ($\pi\pi^*$)

Table G.7: Thymine CC3 core excitation energies, eV, for the three different geometries, aug-cc-pCVTZ/aug-cc-pVDZ.

	GS	$n\pi^*$	$\pi\pi^*$
O(8)	531.19	529.74	530.62
O(7)	532.23	531.82	531.84

Table G.8: Thymine CCSD core excitation energies, (ΔE), eV, and intensisties ($I, \times 100$) for the three different geometries from ground state (top), $n\pi^*$ state (middle) and $\pi\pi^*$ (bottom), O(8), aug-cc-pCVDZ/aug-cc-pVDZ

		GS		$n\pi^*$		$\pi\pi^*$	
		ΔE	I	ΔE	I	ΔE	I
GS	S_1	535.46	3.41	534.49	3.24	535.28	3.25
	S_2	538.95	0.04	538.53	0.05	538.82	0.03
	S_3	539.86	0.15	539.31	0.21	539.64	0.32
	S_4	539.95	0.23	539.38	0.45	539.68	0.13
	S_5	539.97	0.08	539.45	0.18	539.77	0.15
$n\pi^*$	S_1	530.36	3.68	530.65	4.32	530.52	3.87
	S_2	533.84	0.00	534.69	0.01	534.06	0.00
	S_3	534.76	0.00	535.47	0.06	534.88	0.28
	S_4	534.84	0.12	535.54	0.00	534.92	0.00
	S_5	534.86	0.00	535.61	0.07	535.01	0.00
$\pi\pi^*$	S_1	530.09	0.17	529.70	0.77	530.80	0.30
	S_2	533.57	0.01	533.74	0.02	534.34	0.01
	S_3	534.49	0.00	534.53	0.14	535.16	0.26
	S_4	534.57	0.12	534.60	0.04	535.19	0.00
	S_5	534.59	0.01	534.66	0.17	535.28	0.01

Table G.9: Thymine CCSD core excitation energies, (ΔE), eV, and intensisties ($I, \times 100$) for the three different geometries from ground state (top), nπ^* state (middle) and $\pi\pi^*$ (bottom), O(7), aug-cc-pCVDZ/aug-cc-pVDZ

		GS		nπ^*		$\pi\pi^*$	
		ΔE	I	ΔE	I	ΔE	I
GS	S_1	536.45	3.17	536.39	3.16	536.42	3.10
	S_2	538.75	0.01	538.71	0.02	538.72	0.01
	S_3	539.55	0.24	539.05	0.05	539.20	0.05
	S_4	539.70	0.04	539.55	0.22	539.50	0.26
	S_4	540.10	0.15	540.06	0.14	540.06	0.17
nπ^*	S_1	531.35	0.00	532.55	0.00	531.66	0.01
	S_2	533.65	0.00	534.87	0.00	533.96	0.00
	S_3	534.45	0.12	535.21	0.08	534.44	0.15
	S_4	534.59	0.00	535.71	0.00	534.75	0.00
	S_5	535.00	0.00	536.22	0.00	535.30	0.00
$\pi\pi^*$	S_1	531.08	0.03	531.61	0.03	531.93	0.04
	S_2	533.38	0.01	533.93	0.01	534.23	0.01
	S_3	534.18	0.00	534.27	0.35	534.72	0.48
	S_4	534.32	0.65	534.77	0.00	535.02	0.00
	S_5	534.73	0.02	535.28	0.00	535.58	0.00

Bibliography

[1] Andrea Battistoni[1], Nora Berrah[2], Christoph Bostedt[3,4], Phil Bucksbaum[1,5], Ryan N. Coffee[3], Giacomo Coslovich[3], Raimund Feifel[6], Kelly J. Gaffney[1], Jakob Grilj[7], Todd J. Martínez[1,7], Shungo Miyabe[1,7], Stefan P. Moeller[3], Melanie Mucke[8], Adi Natan[1], Razib Obaid[2], Timur Osipov[3], Oksana Plekan[9], and Song Wang[1]. The Thymine Collaboration. [1]*Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA* [2]*Department of Physics, University of Connecticut, Storrs, Connecticut 06269, USA* [3]*Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, California 94720, USA* [4]*Advanced Photon Source, Argonne National Laboratory, Lemont, Illinois 60439, USA* [5]*Department of Physics, Stanford University, Stanford, CA 94305, USA* [6]*Department of Physics, University of Gothenburg, Gothenburg, Sweden* [7]*Laboratory of Ultrafast Spectroscopy, Ecole Polytechnique Federal de Lausanne, CH 1015 Switzerland* [7]*Department of Chemistry, Stanford University, Stanford, California 94305, USA* [8]*Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden* [9]*Elettra-Sincrotrone Trieste, I-34149 Basovizza, Trieste, Italy.*

[2] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28:1049–1070, Dec 1926.

[3] P. A. M. Dirac. The quantum theory of the electron. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 117(778):610–624, 1928.

[4] W. H. Eugen Schwarz. An introduction to relativistic quantum chemistry. In Maria Barysz and Yasuyuki Ishikawa, editors, *Relativistic Methods for Chemists*, pages 1–62. Springer Netherlands, Dordrecht, 2010.

[5] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*, chapter Calibration of the Electronic-Structure Models, pages 849–853. John Wiley & Sons, LTD, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England, May 2004.

[6] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[7] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34:1293–1322, Nov 1929.

[8] Josef Paldus and Bogumil Jeziorski. Clifford algebra and unitary group formulations of the many-electron problem. *Theoretica chimica acta*, 73(2):81–103, 1988.

[9] Andrew Baker. *Matrix Groups - An Introduction to Lie Group Theory*. Springer-Verlag London, Philadelphia, PA, 1st edition, 2002.

[10] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:89–110, 1 1928.

[11] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61(1):126–148, 1930.

[12] David Bohm and David Pines. A Collective Description of Electron Interactions. I. Magnetic Interactions. *Phys. Rev.*, 82:625–634, Jun 1951.

[13] F. Coester. Bound states of a many-particle system. *Nucl. Phys.*, 7:421 – 424, 1958.

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

[15] Jiří Čížek. On the correlation problem in atomic and molecular systems. calculation of wavefunction components in ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys.*, 45(11):4256–4266, 1966.

[16] J. Paldus, J. Čížek, and I. Shavitt. Correlation problems in atomic and molecular systems. IV. extended coupled-pair many-electron theory and its application to the BH_3 molecule. *Phys. Rev. A*, 5:50–67, Jan 1972.

[17] K. A. Brueckner. Nuclear saturation and two-body forces. ii. tensor forces. *Phys. Rev.*, 96:508–516, Oct 1954.

[18] Richard A. Chiles and Clifford E. Dykstra. An electron pair operator approach to coupled cluster wave functions. application to he2, be2, and mg2 and comparison with cepa methods. *J. Chem. Phys*, 74(8):4544–4556, 1981.

[19] Simen Kvaal. Variational formulations of the coupled-cluster method in quantum chemistry. *Mol. Phys.*, 111(9-11):1100–1108, 2013.

[20] James B. Robinson and Peter J. Knowles. Approximate variational coupled cluster theory. *J. Chem. Phys.*, 135(4), 2011.

[21] George D. Purvis and Rodney J. Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.*, 76(4):1910–1918, 1982.

[22] Jozef Noga and Rodney J. Bartlett. The full ccsdt model for molecular electronic structure. *J. Chem. Phys.*, 86(12):7041–7050, 1987.

[23] 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.

[24] Henrik Koch, Ove Christiansen, Poul Jørgensen, Alfredo M. Sanchez de Merás, and Trygve Helgaker. The cc3 model: An iterative coupled cluster approach including connected triples. *J. Chem. Phys.*, 106(5):1808–1818, 1997.

[25] Edward G. Hohenstein, Sara I. L. Kokkila, Robert M. Parrish, and Todd J. Martínez. Quartic scaling second-order approximate coupled cluster singles and doubles via tensor hypercontraction: Thc-cc2. *J. Chem. Phys.*, 138(12):–, 2013.

[26] Krishnan Raghavachari, Gary W. Trucks, John A. Pople, and Martin Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.*, 157(6):479 – 483, 1989.

[27] John F. Stanton. Why ccsd(t) works: a different perspective. *Chemical Physics Letters*, 281(1–3):130 – 134, 1997.

[28] Hendrik J. Monkhorst. Calculation of properties with the coupled-cluster method. *International Journal of Quantum Chemistry*, 12(S11):421–432, 1977.

[29] Esper Dalgaard and Hendrik J. Monkhorst. Some aspects of the time-dependent coupled-cluster approach to dynamic response functions. *Phys. Rev. A*, 28:1217–1222, Sep 1983.

[30] Jeppe Olsen and Poul Jørgensen. Linear and nonlinear response functions for an exact state and for an mcscf state. *The Journal of Chemical Physics*, 82(7):3235–3264, 1985.

[31] H. Koch and P. Jørgensen. Coupled cluster response functions. *J. Chem. Phys.*, 93:3333–3344, 1990.

[32] Thomas Bondo Pedersen and Henrik Koch. Coupled cluster response functions revisited. *J. Chem. Phys.*, 106(19):8059–8072, 1997.

[33] D. J. Rowe. Equations-of-motion method and the extended shell model. *Rev. Mod. Phys.*, 40:153–166, Jan 1968.

[34] 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. *J. Chem. Phys.*, 98(9):7029–7039, 1993.

[35] John F. Stanton. Coupled-cluster theory, pseudo-jahn–teller effects and conical intersections. *J. Chem. Phys.*, 115(22):10382–10393, 2001.

[36] Christof Hättig. Structure optimizations for excited states with correlated second-order methods: CC2 and ADC(2). In H. J. Å. Jensen, editor, *Response Theory and Molecular Properties (A Tribute to Jan Linderberg and Poul Jørgensen)*, volume 50 of *Advances in Quantum Chemistry*, pages 37 – 60. Academic Press, 2005.

[37] Andreas Köhn and Attila Tajti. Can coupled-cluster theory treat conical intersections? *J. Chem. Phys.*, 127(4), 2007.

[38] Kotoku Sasagane, Fumihiko Aiga, and Reikichi Itoh. Higher-order response theory based on the quasienergy derivatives: The derivation of the frequency-dependent polarizabilities and hyperpolarizabilities. *J. Chem. Phys.*, 99(5):3738–3778, 1993.

[39] Fumihiko Aiga, Kotoku Sasagane, and Reikichi Itoh. Frequency-dependent hyperpolarizabilities in the møller–plesset perturbation theory. *J. Chem. Phys.*, 99(5):3779–3789, 1993.

[40] Hans Hellmann. *Einführung in die quantenchemie*. F. Deuticke, 1937.

[41] R. P. Feynman. Forces in molecules. *Phys. Rev.*, 56:340–343, Aug 1939.

[42] E. J. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comp. Phys.*, 17:87–94, 1975.

[43] M. Crouzeix, B. Philippe, and M Sadkane. The Davidson method. *SIAM J. Sci. Comput.*, 15:62–76, 1994.

[44] Cornelius Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Stand.*, 45:225–282, 1950.

[45] Sonia Coriani, Thomas Fransson, Ove Christiansen, and Patrick Norman. Asymmetric-lanczos-chain-driven implementation of electronic resonance convergent coupled-cluster linear response theory. *J Chem. Theory Comput.*, 8(5):1616–1628, 2012.

[46] Sonia Coriani, Ove Christiansen, Thomas Fransson, and Patrick Norman. Coupled-cluster response theory for near-edge x-ray-absorption fine structure of atoms and molecules. *Phys. Rev. A*, 85:022507, Feb 2012.

[47] Henrik Koch, Rika Kobayashi, Alfredo Sánchez de Merás, and Poul Jørgensen. Calculation of size-intensive transition moments from the coupled cluster singles and doubles linear response function. *J. Chem. Phys.*, 100(6):4393–4400, 1994.

[48] Danny L. Yeager and Poul Jørgensen. A multiconfigurational time-dependent Hartree-Fock approach. *Chem. Phys. Lett.*, 65(1):77 – 80, 1979.

[49] Björn O. Roos, Peter R. Taylor, and Per E.M. Siegbahn. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chem. Phys.*, 48(2):157 – 173, 1980.

[50] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, Oct 1934.

[51] Rodney J. Bartlett and David M. Silver. Pair-correlation energies in sodium hydride with many-body perturbation theory. *Phys. Rev. A*, 10:1927–1931, Dec 1974.

[52] Jochen Schirmer. Beyond the random-phase approximation: A new approximation scheme for the polarization propagator. *Phys. Rev. A*, 26:2395–2416, Nov 1982.

[53] J. Schirmer and A. B. Trofimov. Intermediate state representation approach to physical properties of electronically excited molecules. *J. Chem. Phys.*, 120(24):11449–11464, 2004.

[54] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*, chapter Perturbation Theory, pages 769–783. John Wiley & Sons, LTD, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England, May 2004.

[55] Philipp H. P. Harbach, Michael Wormit, and Andreas Dreuw. The third-order algebraic diagrammatic construction method (adc(3)) for the polarization propagator for closed-shell molecules: Efficient implementation and benchmarkinga). *The Journal of Chemical Physics*, 141(6), 2014.

[56] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self—consistent molecular orbital methods. xii. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56(5):2257–2261, 1972.

[57] J. Stephen Binkley, John A. Pople, and Warren J. Hehre. Self-consistent molecular orbital methods. 21. small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.*, 102(3):939–947, 1980.

[58] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.

[59] David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. v. core-valence basis sets for boron through neon. *J. Chem. Phys.*, 103(11):4572–4585, 1995.

[60] Rick A. Kendall, Thom H. Dunning, and Robert J. Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *J. Chem. Phys.*, 96(9):6796–6806, 1992.

[61] J. L. Whitten. Coulombic potential energy integrals and approximations. *J. Chem. Phys.*, 58(10):4496–4501, 1973.

[62] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin. On some approximations in applications of xα theory. *J. Chem. Phys.*, 71(8):3396–3402, 1979.

[63] O. Vahtras, J. Almlöf, and M.W. Feyereisen. Integral approximations for lcao-scf calculations. *Chem. Phys. Lett.*, 213(5–6):514 – 518, 1993.

[64] Rick A. Kendall and Herbert A. Früchtl. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.*, 97(1-4):158–163, 1997.

[65] Florian Weigend. A fully direct ri-hf algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.*, 4:4285–4291, 2002.

[66] Nelson H. F. Beebe and Jan Linderberg. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. of Quant. Chem.*, 12(4):683–705, 1977.

[67] I. Røeggen and E. Wisløff-Nilssen. On the beebe-linderberg two-electron integral approximation. *Chemical Physics Letters*, 132(2):154 – 160, 1986.

[68] Henrik Koch, Alfredo Sánchez de Merás, and Thomas Bondo Pedersen. Reduced scaling in electronic structure calculations using cholesky decompositions. *J. Chem. Phys.*, 118(21):9481–9484, 2003.

[69] Francesco Aquilante, Linus Boman, Jonas Boström, Henrik Koch, Roland Lindh, Alfredo Sánchez Merás, and Thomas Bondo Pedersen. Cholesky decomposition techniques in electronic structure theory. In Robert Zalesny, G. Manthos Papadopoulos, G. Paul Mezey, and Jerzy Leszczynski, editors, *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*, pages 301–343. Springer Netherlands, Dordrecht, 2011.

[70] Cdt. Benoît. Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés a un système d'équations linéaires en nombre inférieur a celui des inconnues. — application de la méthode a la résolution d'un système defini d'équations linéaires. *Bulletin géodésique*, 2:67, 1924.

[71] Peter Pulay. Localizability of dynamic electron correlation. *Chem. Phys. Lett.*, 100(2):151 – 154, 1983.

[72] Svein Saebø and Peter Pulay. Local treatment of electron correlation. *Annu. Rev. Phys. Chem.*, 44(1):213–236, 1993.

[73] N. Flocke and Rodney J. Bartlett. A natural linear scaling coupled-cluster method. *J. Chem. Phys.*, 121(22):10935–10944, 2004.

[74] Masato Kobayashi and Hiromi Nakai. Extension of linear-scaling divide-and-conquer-based correlation method to coupled cluster theory with singles and doubles excitations. *J. Chem. Phys.*, 129:044103, 2008.

[75] Janus J. Eriksen, Pablo Baudin, Patrick Ettenhuber, Kasper Kristensen, Thomas Kjærgaard, and Poul Jørgensen. Linear-scaling coupled cluster with perturbative triple excitations: The divide–expand–consolidate ccsd(t) model. *J. Chem. Theory Comput.*, 11(7):2984–2993, 2015.

[76] Frank Neese, Andreas Hansen, and Dimitrios G. Liakos. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys.*, 131:064103, 2009.

[77] Christoph Riplinger, Barbara Sandhoefer, Andreas Hansen, and Frank Neese. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.*, 139:134101, 2013.

[78] Jun Yang, Garnet Kin-Lic Chan, Frederick R. Manby, Martin Schütz, and Hans-Joachim Werner. The orbital-specific-virtual local coupled cluster singles and doubles method. *J. Chem. Phys.*, 136:144105, 2012.

[79] Martin Schütz, Jun Yang, Garnet Kin-Lic Chan, Frederick R. Manby, and Hans-Joachim Werner. The orbital-specific virtual local triples correction: Osv-l(t). *J. Chem. Phys.*, 138:054109, 2013.

[80] Feliu Maseras and Keiji Morokuma. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, 16(9):1170–1179, 1995.

[81] Mats Svensson, Stéphane Humbel, Robert D. J. Froese, Toshiaki Matsubara, Stefan Sieber, and Keiji Morokuma. ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and pt(p(*t*-bu)$_3$)$_2$ + h$_2$ oxidative addition. *J. Phys. Chem.*, 100(50):19357–19363, 1996.

[82] Thomas M. Henderson. Embedding wave function theory in density functional theory. *J. Chem. Phys.*, 125(1), 2006.

[83] Tomasz A. Wesolowski, Sapana Shedge, and Xiuwen Zhou. Frozen-density embedding strategy for multilevel simulations of electronic structure. *Chem. Rev.*, 115(12):5891–5928, 2015. PMID: 25923542.

[84] Sebastian Höfener. Coupled-cluster frozen-density embedding using resolution of identity methods. *J. Comput. Chem.*, 35:1716–1724, 2014.

[85] Andreas Köhn and Jeppe Olsen. Coupled-cluster with active space selected higher amplitudes: Performance of seminatural orbitals for ground and excited state calculations. *J. Chem. Phys.*, 125:174110, 2006.

[86] Piotr Piecuch. Active-space coupled-cluster methods. *Mol. Phys.*, 108:2987–3015, 2010.

[87] Ove Christiansen, Henrik Koch, and Poul Jørgensen. Response functions in the CC3 iterative triple excitation model. *J. Chem. Phys.*, 103(17):7429–7441, 1995.

[88] 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.

[89] S. F. Boys. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.*, 32:296–299, Apr 1960.

[90] Clyde Edmiston and Klaus Ruedenberg. Localized atomic and molecular orbitals. *Rev. Mod. Phys.*, 35:457–464, Jul 1963.

[91] János Pipek and Paul G. Mezey. A fast intrinsic localization procedure applicable for abinitio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.*, 90(9):4916–4926, 1989.

[92] Ida-Marie Høyvik, Branislav Jansik, and Poul Jørgensen. Trust region minimization of orbital localization functions. *J. Chem. Theory Comput.*, 8(9):3137–3146, 2012. PMID: 26605725.

[93] Ida-Marie Høyvik and Poul Jørgensen. Characterization and generation of local occupied and virtual hartree–fock orbitals. *Chem. Rev.*, 116(5):3306–3327, 2016. PMID: 26855066.

[94] Linda Stakvik. Using Localized Molecular Orbitals in the Multi-level Coupled Cluster Approach. Master's thesis, NTNU, 2016.

[95] Ida Marie Høyvik, Rolf Heilemann Myhre, and Henrik Koch. Automated orbital selection for multilevel coupled cluster models. in preparation.

[96] A. V. Luzanov, A. A. Sukhorukov, and V. É. Umanskii. Application of transition density matrix for analysis of excited states. *Theoretical and Experimental Chemistry*, 10(4):354–361, 1976.

[97] Richard L. Martin. Natural transition orbitals. *J. Chem. Phys.*, 118(11):4775–4777, 2003.

[98] Pablo Baudin and Kasper Kristensen. Lofex — a local framework for calculating excitation energies: Illustrations using ri-cc2 linear response theory. *J. Chem. Phys.*, 144(22), 2016.

[99] Kestutis Aidas, Celestino Angeli, Keld L. Bak, Vebjørn Bakken, Radovan Bast, Linus Boman, Ove Christiansen, Renzo Cimiraglia, Sonia Coriani, Pål Dahle, Erik K. Dalskov, Ulf Ekström, Thomas Enevoldsen, Janus J. Eriksen, Patrick Ettenhuber, Berta Fernández, Lara Ferrighi, Heike Fliegl, Luca Frediani, Kasper Hald, Asger Halkier, Christof Hättig, Hanne Heiberg, Trygve Helgaker, Alf Christian Hennum, Hinne Hettema, Eirik Hjertenæs, Stinne Høst, Ida-Marie Høyvik, Maria Francesca Iozzi, Branislav Jansík, Hans Jørgen Aa. Jensen, Dan Jonsson, Poul Jørgensen, Joanna Kauczor, Sheela Kirpekar, Thomas Kjærgaard, Wim Klopper, Stefan Knecht, Rika Kobayashi, Henrik Koch, Jacob Kongsted, Andreas Krapp, Kasper Kristensen, Andrea Ligabue, Ola B. Lutnæs, Juan I. Melo, Kurt V. Mikkelsen, Rolf H. Myhre, Christian Neiss, Christian B. Nielsen, Patrick Norman, Jeppe Olsen, Jógvan Magnus H. Olsen, Anders Osted, Martin J. Packer, Filip Pawlowski, Thomas B. Pedersen, Patricio F. Provasi, Simen Reine, Zilvinas Rinkevicius, Torgeir A. Ruden, Kenneth Ruud, Vladimir V. Rybkin, Pawel Sałek, Claire C. M. Samson, Alfredo Sánchez de Merás, Trond Saue, Stephan P. A. Sauer, Bernd Schimmelpfennig, Kristian Sneskov, Arnfinn H. Steindal, Kristian O. Sylvester-Hvid, Peter R. Taylor, Andrew M. Teale, Erik I. Tellgren, David P.

Tew, Andreas J. Thorvaldsen, Lea Thøgersen, Olav Vahtras, Mark A. Watson, David J. D. Wilson, Marcin Ziolkowski, and Hans Ågren. The Dalton quantum chemistry program system. *WIREs Comput. Mol. Sci.*, 4(3):269–284, 2014.

[100] Dalton, a molecular electronic structure program, release dalton2016, 2015.

[101] Kristin Marie Skjelbred. Calculation of Transition Moments Using the Extended Coupled Cluster Model ECC2. Master's thesis, NTNU, 2014.

[102] Alistair P. Rendell, Timothy J. Lee, and Andrew Komornicki. A parallel vectorized implementation of triple excitations in CCSD(T): application to the binding energies of the AlH_3, AlH_2F, $AlHF_2$ and AlF_3 dimers. *Chem. Phys. Lett.*, 178(5–6):462 – 470, 1991.

[103] OpenMP Architecture Review Board. OpenMP application program interface version 4.0, July 2013. http://www.openmp.org/mp-documents/OpenMP4.0.0.pdf.

[104] MPI Forum. Message Passing Interface (MPI) Forum Home Page, August 2016. http://www.mpi-forum.org/docs/docs.html.

[105] C.J. Milne, T.J. Penfold, and M. Chergui. Recent experimental and theoretical developments in time-resolved x-ray spectroscopies. *Coordination Chemistry Reviews*, 277–278:44 – 68, 2014. Following Chemical Structures using Synchrotron Radiation.

[106] Thomas Fransson, Yoshihisa Harada, Nobuhiro Kosugi, Nicholas A. Besley, Bernd Winter, John J. Rehr, Lars G. M. Pettersson, and Anders Nilsson. X-ray and electron spectroscopy of water. *Chem. Rev.*, 116(13):7551–7569, 2016. PMID: 27244473.

[107] Florian Holch, Dominique Hübner, Rainer Fink, Achim Schöll, and Eberhard Umbach. New set-up for high-quality soft-x-ray absorption spectroscopy of large organic molecules in the gas phase. *J. Electron Spectrosc.*, 184(8–10):452 – 456, 2011.

[108] Claudio Garino, Elisa Borfecchia, Roberto Gobetto, Jeroen A. van Bokhoven, and Carlo Lamberti. Determination of the electronic and structural configuration of coordination compounds by synchrotron-radiation techniques. *Coordin. Chem. Rev.*, 277:130 – 186, 2014. Following Chemical Structures using Synchrotron Radiation.

[109] Vladimir S. Petrović, Marco Siano, James L. White, Nora Berrah, Christoph Bostedt, John D. Bozek, Douglas Broege, Max Chalfin, Ryan N. Coffee, James Cryan, Li Fang, Joseph P. Farrell, Leszek J Frasinski, James M. Glownia, Markus Gühr, Matthias Hoener, David M. P. Holland, Jaehee Kim, Jonathan P. Marangos, Todd Martinez, Brian K. McFarland, Russell S. Minns, Shungo Miyabe, Sebastian Schorb, Roseanne J. Sension, Limor S. Spector, Richard Squibb, Hongli Tao, Jonathan G. Underwood, and Philip H. Bucksbaum. Transient x-ray fragmentation: Probing a prototypical photoinduced ring opening. *Phys. Rev. Lett.*, 108:253006, Jun 2012.

[110] B. K. McFarland, J. P. Farrell, S. Miyabe, F. Tarantelli, A. Aguilar, N. Berrah, C. Bostedt, J. D. Bozek, P. H. Bucksbaum, J. C. Castagna, R. N. Coffee, J. P. Cryan, L. Fang, R. Feifel, K. J. Gaffney, J. M. Glownia, T. J. Martinez, M. Mucke, B. Murphy, A. Natan, T. Osipov, V. S. Petrović, S. Schorb, Th. Schultz, L. S. Spector, M. Swiggers, I. Tenney, S. Wang, J. L. White, W. White, and M. Gühr. Ultrafast X-ray auger probing of photoexcited molecular dynamics. *Nat. Commun.*, 5:4235, Jun 2014.

[111] Chelsea E. Liekhus-Schmaltz, Ian Tenney, Timur Osipov, Alvaro Sanchez-Gonzalez, Nora Berrah, Rebecca Boll, Cedric Bomme, Christoph Bostedt, John D. Bozek, Sebastian Carron, Ryan Coffee, Julien Devin, Benjamin Erk, Ken R. Ferguson, Robert W. Field, Lutz Foucar, Leszek J. Frasinski, James M. Glownia, Markus Gühr, Andrei Kamalov, Jacek Krzywinski, Heng Li, Jonathan P. Marangos, Todd J. Martinez, Brian K. McFarland, Shungo Miyabe, Brendan Murphy, Adi Natan, Daniel Rolles, Artem Rudenko, Marco Siano, Emma R. Simpson, Limor Spector, Michele Swiggers, Daniel Walke, Song Wang, Thorsten Weber, Philip H. Bucksbaum, and Vladimir S. Petrović. Ultrafast isomerization initiated by X-ray core ionization. *Nat. Commun.*, 6:8199, Oct 2015.

[112] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. Prince, A. Trofimov, E. Gromov, I. Zaytseva, and J. Schirmer. A theoretical and experimental study of the near edge x-ray absorption fine structure (NEXAFS) and X-ray photoelectron spectra (XPS) of nucleobases: Thymine and adenine. *Chem. Phys.*, 347:360–375, 2008.

[113] Vitaliy Feyer, Oksana Plekan, Robert Richter, Marcello Coreno, Monica de Simone, Kevin C. Prince, Alexander B. Trofimov, Irina L. Zaytseva, and Jochen Schirmer. Tautomerism in cytosine and uracil: A theoretical and experimental x-ray absorption and resonant auger study. *J. Phys. Chem. A*, 114:10270–10276, 2010.

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

[115] Michael Wormit, Dirk R. Rehn, Philipp H.P. Harbach, Jan Wenzel, Caroline M. Krauter, Evgeny Epifanovsky, and Andreas Dreuw. Investigating excited electronic states using the algebraic diagrammatic construction (adc) approach of the polarisation propagator. *Mol. Phys.*, 112(5-6):774–784, 2014.

[116] Sonia Coriani and Henrik Koch. Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework. *J. Chem. Phys.*, 143(18):181103, 2015.

[117] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25(13):1605–1612, 2004. https://www.cgl.ucsf.edu/chimera/.

[118] About Stallo. `http://hpc.uit.no/en/latest/general/stallo.html`, Dec 2015.

[119] Johanna Carbone, Rolf H. Myhre, Henrik Koch, and Sonia Coriani. An analysis of the performance of core-valence separated coupled cluster methods for core excitations and core ionizations using standard basis sets. in preparation.

[120] J. Schirmer, A. B. Trofimov, K. J. Randall, J. Feldhaus, A. M. Bradshaw, Y. Ma, C. T. Chen, and F. Sette. K-shell excitation of the water, ammonia, and methane molecules using high-resolution photoabsorption spectroscopy. *Phys. Rev. A*, 47:1136–1147, Feb 1993.

[121] James D. Dill and John A. Pople. Self-consistent molecular orbital methods. xv. extended gaussian-type basis sets for lithium, beryllium, and boron. *J. Chem. Phys.*, 62(7):2921–2923, 1975.

[122] J.F. Stanton, J. Gauss, M.E. Harding, P.G. Szalay, A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, L. Cheng, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W.J. Lauderdale, F. Lipparini, D.A. Matthews, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts, J. Almlöf, P.R. Taylor, P.R. Taylor, T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, J. Olsen, A. V. Mitin, and C. van Wüllen. CFOUR, coupled-cluster techniques for computational chemistry http://www.cfour.de, May 2015.

[123] Michael E. Harding, Thorsten Metzroth, , Jürgen Gauss, and Alexander A. Auer*. Parallel calculation of ccsd and ccsd(t) analytic first and second derivatives. *J. Chem. Theory Comput.*, 4(1):64–74, 2008. PMID: 26619980.

[124] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K.C. Prince, and V. Carravetta. An x-ray absorption study of glycine, methionine and proline. *J. Electron Spectrosc.*, 155(1–3):47 – 53, 2007. Scattering, Coincidence and Absorption Studies of MoleculesS-CASM.

[125] Hans Ågren, Vincenzo Carravetta, Olav Vahtras, and M. Lars G. Pettersson. Direct scf direct static-exchange calculations of electronic spectra. *Theor. Chem. Acc.*, 97(1):14–40, 1997.

[126] Kevin Prince. personal communication.

[127] Vincenzo Barone, Malgorzata Biczysko, Julien Bloino, and Cristina Puzzarini. Glycine conformers: a never-ending story? *Phys. Chem. Chem. Phys.*, 15:1358–1363, 2013.

[128] John F. Stanton and Jürgen Gauss. A simple scheme for the direct calculation of ionization potentials with coupled-cluster theory that exploits established excitation energy methods. *The Journal of Chemical Physics*, 111(19):8785–8788, 1999.

[129] Alberto Baiardi, Marco Mendolicchio, Vincenzo Barone, Giovanna Fronzoni, Gustavo Adolfo Cardenas Jimenez, Mauro Stener, Cesare Grazioli, Monica de Simone, and Marcello Coreno. Vibrationally resolved nexafs at c and n k-edges of pyridine, 2-fluoropyridine and 2,6-difluoropyridine: A combined experimental and theoretical assessment. *J. Chem. Phys.*, 143(20), 2015.

[130] C. T. Chen, Y. Ma, and F. Sette. K -shell photoabsorption of the n_2 molecule. *Phys. Rev. A*, 40:6737–6740, Dec 1989.

[131] Philip M. Morse. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Phys. Rev.*, 34:57–64, Jul 1929.

[132] Dage Sundholm. VIBROT. `http://www.chem.helsinki.fi/~sundholm/software/GPL/`.

[133] Dage Sundholm, Jürgen Gauss, and Ansgar Schäfer. Rovibrationally averaged nuclear magnetic shielding tensors calculated at the coupled-cluster level. *J. Chem. Phys.*, 105(24):11051–11059, 1996.

[134] Catalin Miron, Christophe Nicolas, Oksana Travnikova, Paul Morin, Yuping Sun, Faris Gel'mukhanov, Nobuhiro Kosugi, and Victor Kimberg. Imaging molecular potentials using ultrahigh-resolution resonant photoemission. *Nat. Phys.*, 8:1745–2473, Feb 2012.

[135] KP Huber and G Herzberg. Constants of diatomic molecules. In *Molecular Spectra and Molecular Structure*, pages 8–689. Springer, 1979.

[136] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*, chapter Calibration of the Electronic-Structure Models, pages 821–836. John Wiley & Sons, LTD, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England, May 2004.

[137] Ephraim Eliav and Uzi Kaldor. Four-component electronic structure methods. In *Relativistic Methods for Chemists*, pages 1–62. Springer Netherlands, Dordrecht, 2010.

[138] C. G. Darwin. The wave equations of the electron. *Proc. R. Soc. London A*, 118(780):654–680, 1928.

[139] Marvin Douglas and Norman M Kroll. Quantum electrodynamical corrections to the fine structure of helium. *Annals of Physics*, 82(1):89 – 155, 1974.

[140] Bernd A. Hess. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Phys. Rev. A*, 33:3742–3748, Jun 1986.

[141] J. Niskanen, P. Norman, H. Aksela, and H. Ågren. Relativistic contributions to single and double core electron ionization energies of noble gases. *J. Chem. Phys.*, 135(5), 2011.

[142] Wolfgang J. Schreier, Tobias E. Schrader, Florian O. Koller, Peter Gilch, Carlos E. Crespo-Hernández, Vijay N. Swaminathan, Thomas Carell, Wolfgang Zinth, and Bern Kohler. Thymine dimerization in dna is an ultrafast photoreaction. *Science*, 315(5812):625–629, 2007.

[143] Hanneli R. Hudock, Benjamin G. Levine, Alexis L. Thompson, Helmut Satzger, D. Townsend, N. Gador, S. Ullrich, Albert Stolow, , and Todd J. Martínez. Ab initio molecular dynamics and time-resolved photoelectron spectroscopy of electronically excited uracil and thymine. *J. Phys. Chem. A*, 111(34):8500–8508, 2007. PMID: 17685594.

[144] Jaroslaw J. Szymczak, Mario Barbatti, Jason T. Soo Hoo, Jaclyn A. Adkins, Theresa L. Windus, Dana Nachtigallová, and Hans Lischka. Photodynamics simulations of thymine: Relaxation into the first excited singlet state. *J. Phys. Chem. A*, 113(45):12686–12693, 2009. PMID: 19691341.

[145] Chris T. Middleton, Kimberly de La Harpe, Charlene Su, Yu Kay Law, Carlos E. Crespo-Hernández, and Bern Kohler. Dna excited-state dynamics: From single bases to the double helix. *Annu. Rev. Phys. Chem.*, 60(1):217–239, 2009. PMID: 19012538.

[146] A B Trofimov, J Schirmer, V B Kobychev, A W Potts, D M P Holland, and L Karlsson. Photoelectron spectra of the nucleobases cytosine, thymine and adenine. *J. Phys. B: At., Mol. Opt.*, 39(2):305, 2006.

[147] Hui Yu, Jose A. Sanchez-Rodriguez, Marvin Pollum, Carlos E. Crespo-Hernandez, Sebastian Mai, Philipp Marquetand, Leticia Gonzalez, and Susanne Ullrich. Internal conversion and intersystem crossing pathways in uv excited, isolated uracils and their implications in prebiotic chemistry. *Phys. Chem. Chem. Phys.*, 18:20168–20176, 2016.

[148] T Koopmans. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica*, 1(1):104 – 113, 1934.

[149] Markus Drescher, Michael Hentschel, R Kienberger, Matthias Uiberacker, Vladislav Yakovlev, Armin Scrinzi, Th Westerwalbesloh, U Kleineberg, Ulrich Heinzmann, and Ferenc Krausz. Time-resolved atomic inner-shell spectroscopy. *Nature*, 419(6909):803–807, 2002.

[150] P. Auger. Sur les rayons β; secondaires produits dans un gaz par des rayons x. *C.R.A.S.*, 177:169–171, 1923.

[151] L. Meitner. Über die Entstehung der β-Strahl-Spektren radioaktiver Substanzen. *Zeitschrift fur Physik*, 9:131–144, December 1922.

[152] D. CHATTARJI. Chapter eight - auger electron spectroscopy and its application to surface science. In D. CHATTARJI, editor, *The Theory of Auger Transitions*, pages 180 – 250. Academic Press, 1976.

[153] Yihan Shao, Zhengting Gan, Evgeny Epifanovsky, Andrew T.B. Gilbert, Michael Wormit, Joerg Kussmann, Adrian W. Lange, Andrew Behn, Jia Deng, Xintian Feng, Debashree Ghosh, Matthew Goldey, Paul R. Horn, Leif D. Jacobson, Ilya Kaliman, Rustam Z. Khaliullin, Tomasz Kuś, Arie Landau, Jie Liu, Emil I. Proynov, Young Min Rhee, Ryan M. Richard, Mary A. Rohrdanz, Ryan P. Steele, Eric J. Sundstrom, H. Lee Woodcock III, Paul M. Zimmerman, Dmitry Zuev, Ben Albrecht, Ethan Alguire, Brian Austin, Gregory J. O. Beran, Yves A. Bernard, Eric Berquist, Kai Brandhorst, Ksenia B. Bravaya, Shawn T. Brown, David Casanova, Chun-Min Chang, Yunqing Chen, Siu Hung Chien, Kristina D. Closser, Deborah L. Crittenden, Michael Diedenhofen, Robert A. DiStasio Jr., Hainam Do, Anthony D. Dutoi, Richard G. Edgar, Shervin Fatehi, Laszlo Fusti-Molnar, An Ghysels, Anna Golubeva-Zadorozhnaya, Joseph Gomes, Magnus W.D. Hanson-Heine, Philipp H.P. Harbach, Andreas W. Hauser, Edward G. Hohenstein, Zachary C. Holden, Thomas-C. Jagau, Hyunjun Ji, Benjamin Kaduk, Kirill Khistyaev, Jaehoon Kim, Jihan Kim, Rollin A. King, Phil Klunzinger, Dmytro Kosenkov, Tim Kowalczyk, Caroline M. Krauter, Ka Un Lao, Adèle D. Laurent, Keith V. Lawler, Sergey V. Levchenko, Ching Yeh Lin, Fenglai Liu, Ester Livshits, Rohini C. Lochan, Arne Luenser, Prashant Manohar, Samuel F. Manzer, Shan-Ping Mao, Narbe Mardirossian, Aleksandr V. Marenich, Simon A. Maurer, Nicholas J. Mayhall, Eric Neuscamman, C. Melania Oana, Roberto Olivares-Amaya, Darragh P. O'Neill, John A. Parkhill, Trilisa M. Perrine, Roberto Peverati, Alexander Prociuk, Dirk R. Rehn, Edina Rosta, Nicholas J. Russ, Shaama M. Sharada, Sandeep Sharma, David W. Small, Alexander Sodt, Tamar Stein, David Stück, Yu-Chuan Su, Alex J.W. Thom, Takashi Tsuchimochi, Vitalii Vanovschi, Leslie Vogt, Oleg Vydrov, Tao Wang, Mark A. Watson, Jan Wenzel, Alec White, Christopher F. Williams, Jun Yang, Sina Yeganeh, Shane R. Yost, Zhi-Qiang You, Igor Ying Zhang, Xing Zhang, Yan Zhao, Bernard R. Brooks, Garnet K.L. Chan, Daniel M. Chipman, Christopher J. Cramer, William A. Goddard III, Mark S. Gordon, Warren J. Hehre, Andreas Klamt, Henry F. Schaefer III,

Michael W. Schmidt, C. David Sherrill, Donald G. Truhlar, Arieh Warshel, Xin Xu, Alán Aspuru-Guzik, Roi Baer, Alexis T. Bell, Nicholas A. Besley, Jeng-Da Chai, Andreas Dreuw, Barry D. Dunietz, Thomas R. Furlani, Steven R. Gwaltney, Chao-Ping Hsu, Yousung Jung, Jing Kong, Daniel S. Lambrecht, WanZhen Liang, Christian Ochsenfeld, Vitaly A. Rassolov, Lyudmila V. Slipchenko, Joseph E. Subotnik, Troy Van Voorhis, John M. Herbert, Anna I. Krylov, Peter M.W. Gill, and Martin Head-Gordon. Advances in molecular quantum chemistry contained in the q-chem 4 program package. *Mol. Phys.*, 113(2):184–215, 2015.

[154] J. Franck and E. G. Dymond. Elementary processes of photochemical reactions. *Trans. Faraday Soc.*, 21:536–542, 1926.

[155] Edward Condon. A theory of intensity distribution in band systems. *Phys. Rev.*, 28:1182–1201, Dec 1926.

[156] Kasper Hald, Poul Jørgensen, Ove Christiansen, and Henrik Koch. Implementation of electronic ground states and singlet and triplet excitation energies in coupled cluster theory with approximate triples corrections. *J. Chem. Phys.*, 116(14):5963–5970, 2002.

[157] Dániel Kánnár and Péter G. Szalay. Benchmarking coupled cluster methods on valence singlet excited states. *J. Chem. Theory Comput.*, 10:3757–3765, 2014.

[158] Henrik Koch, Hans Jørgen Aa. Jensen, Poul Jørgensen, and Trygve Helgaker. Excitation energies from the coupled cluster singles and doubles linear response function (CCSDLR). Applications to Be, CH^+, CO, and H_2O. *J. Chem. Phys.*, 93:3345, 1990.

[159] Ove Christiansen, Poul Jørgensen, and Christof Hättig. Response functions from fourier component variational perturbation theory applied to a time-averaged quasienergy. *Int. J. Quantum Chem.*, 68:1–52, 1997.

[160] Rodney J. Bartlett and Monika Musial. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79:291–352, 2007.

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

[162] D. H. Friese, N. O. C. Winter, P. Balzerowski, R. Schwan, and C. Hättig. Large scale polarizability calculations using the approximate coupled cluster model cc2 and mp2 combined with the resolution-of-the-identity approximation. *J. Chem. Phys.*, 136:174106, 2012.

[163] Tatiana Korona. XCC2 - a new coupled cluster model for the second-order polarization propagator. *Phys. Chem. Chem. Phys.*, 12:14977–14984, 2010.

[164] I. G. Cuesta, J. S. Marin, A. Sánchez de Merás, F. Pawlowski, and P. Lazzeretti. Assessment of the ctocd-dz method in a hierarchy of coupled cluster methods. *Phys. Chem. Chem. Phys.*, 12:6163–6170, 2010.

[165] Thomas Bondo Pedersen, Henrik Koch, Linus Boman, and Alfredo M.J Sánchez de Merás. Origin invariant calculation of optical rotation without recourse to london orbitals. *Chem. Phys. Lett.*, 393(4–6):319 – 326, 2004.

[166] Jeppe Olsen, Poul Jørgensen, Henrik Koch, Anna Balkova, and Rodney 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(20):8007–8015, 1996.

[167] Björn O. Roos, Kerstin Andersson, Markus P. Fülscher, Per âke Malmqvist, Luis Serrano-Andrés, Kristin Pierloot, and Manuela Merchán. *Advances in Chemical Physics: New Methods in Computational Quantum Mechanics*, volume 93, pages 219–331. John Wiley & Sons, Inc., 1996.

[168] Henrik Koch, Alfredo Sánchez de Merás, Trygve Helgaker, and Ove Christiansen. The integral-direct coupled cluster singles and doubles model. *J. Chem. Phys.*, 104:4157, 1996.

[169] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, LTD, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England, May 2004.

[170] Keiji Morokuma. Oniom and its applications to material chemistry and catalyses. *Bull. Korean Chem. Soc.*, 24:797–801, 2003.

[171] Francesco Aquilante, Thomas Bondo Pedersen, Alfredo Sánchez de Merás, and Henrik Koch. Fast noniterative orbital localization for large molecules. *J. Chem. Phys.*, 125:174101, 2006.

[172] William M. Haynes, editor. *CRC Handbook of Chemistry and Physics*. CRC Press, Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742, 93 edition, June 2012.

[173] P. G. Szalay, A. Karpfen, and H. Lischka. SCF and electron correlation studies on structures and harmonic in-plane force fields of ethylene, trans 1,3-butadiene, and all-trans 1,3,5-hexatriene. *J. Chem. Phys.*, 87:3530, 1987.

[174] Rolf H. Myhre, Alfredo M. J. Sánchez de Merás, and Henrik Koch. Multi-level coupled cluster theory. *J. Chem. Phys.*, 141(22):–, 2014.

[175] Peter Pulay. Localizability of dynamic electron correlation. *Chem. Phys. Lett.*, 100:151–154, 1983.

[176] Svein Sæbø and Peter Pulay. Local configuration interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.*, 113:13–18, 1985.

[177] Peter Pulay and Svein Sæbø. Orbital-invariant formulation and second-order gradient evaluation in møller-plesset perturbation theory. *Theor. Chim. Acta*, 69:357–368, 1986.

[178] Svein Sæbø and Peter Pulay. Fourth-order møller–plessett perturbation theory in the local correlation treatment. i. method. *J. Chem. Phys.*, 86:914, 1987.

[179] Svein Sæbø and Peter Pulay. The local correlation treatment. ii. implementation and tests. *J. Chem. Phys.*, 88:1884, 1988.

[180] Claudia Hampel and Hans-Joachim Werner. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.*, 104:6286, 1996.

[181] Martin Schütz, Georg Hetzer, and Hans-Joachim Werner. Low-order scaling local electron correlation methods. i. linear scaling local mp2. *J. Chem. Phys.*, 111:5691, 1999.

[182] Georg Hetzer, Martin Schütz, Hermann Stoll, and Hans-Joachim Werner. Low-order scaling local correlation methods ii: Splitting the coulomb operator in linear scaling local second-order møller-plesset perturbation theory. *J. Chem. Phys.*, 113:9443, 2000.

[183] Martin Schütz. Low-order scaling local electron correlation methods. iii. linear scaling local perturbative triples correction (t). *J. Chem. Phys.*, 113:9986, 2000.

[184] Martin Schütz and Hans-Joachim Werner. Low-order scaling local electron correlation methods. iv. linear scaling local coupled-cluster (lccsd). *J. Chem. Phys.*, 114:661, 2001.

[185] Marcin Ziółkowski, Branislav Jansik, Thomas Kjærgaard, and Poul Jørgensen. Linear scaling coupled cluster method with correlation energy based error control. *J. Chem. Phys.*, 133:014107, 2010.

[186] Kasper Kristensen, Marcin Ziółkowski, Branislav Jansík, Thomas Kjærgaard, and Poul Jørgensen. A locality analysis of the divide–expand–consolidate coupled cluster amplitude equations. *J. Chem. Theory Comput.*, 7:1677, 2011.

[187] Ida-Marie Høyvik, Kasper Kristensen, Branislav Jansik, and Poul Jørgensen. The divide-expand-consolidate family of coupled cluster methods: Numerical illustrations using second order møller-plesset perturbation theory. *J. Chem. Phys.*, 136:014105, 2012.

[188] Sebastian Höfener, André Severo, Pereira Gomes, and Lucas Visscher. Molecular properties via a subsystem density functional theory formulation: a common framework for electronic embedding. *J. Chem. Phys.*, 136:044104, 2012.

[189] Sebastian Höfener and Lucas Visscher. Calculation of electronic excitations using wave-function in wave-function frozen-density embedding. *J. Chem. Phys.*, 137:204120, 2012.

[190] Christoph Riplinger and Frank Neese. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.*, 138:034106, 2013.

[191] Jun Yang, Yuki Kurashige, Frederick R. Manby, and Garnet K. L. Chan. Tensor factorizations of local second-order møller–plesset theory. *J. Chem. Phys.*, 134:044123, 2011.

[192] Benjamin Helmich and Christof Hättig. A pair natural orbital implementation of the coupled cluster model cc2 for excitation energies. *J. Chem. Phys.*, 139:084114, 2013.

[193] Edward W. Stuebing, John H. Weare, and Robert G. Parr. Discontinuous approximate molecular electronic wave-functions. *Int. J. of Quant. Chem.*, 11(1):81–102, 1977.

[194] Alistair P. Rendell and Timothy J. Lee. Coupled-cluster theory employing approximate integrals: An approach to avoid the input/output and storage bottlenecks. *J. Chem. Phys.*, 101(1):400–408, 1994.

[195] Francesco Aquilante, Thomas Bondo Pedersen, and Roland Lindh. Low-cost evaluation of the exchange fock matrix from cholesky and density fitting representations of the electron repulsion integrals. *J. Chem. Phys.*, 126(19):–, 2007.

[196] Francesco Aquilante, Laura Gagliardi, Thomas Bondo Pedersen, and Roland Lindh. Atomic cholesky decompositions: A route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency. *J. Chem. Phys.*, 130(15):–, 2009.

[197] Richard A. Friesner. Solution of self-consistent field electronic structure equations by a pseudospectral method. *Chem. Phys. Lett.*, 116(1):39 – 43, 1985.

[198] Richard A. Friesner. Solution of the hartree–fock equations by a pseudospectral method: Application to diatomic molecules. *J. Chem. Phys.*, 85(3):1462–1468, 1986.

[199] Murco N. Ringnalda, Youngdo Won, and Richard A. Friesner. Pseudospectral hartree–fock calculations on glycine. *J. Chem. Phys.*, 92(2):1163–1173, 1990.

[200] Todd J. Martífnez, Aseem Mehta, and Emily A. Carter. Pseudospectral full configuration interaction. *J. Chem. Phys.*, 97(3):1876–1880, 1992.

[201] Andrew Komornicki and George Fitzgerald. Molecular gradients and hessians implemented in density functional theory. *J. Chem. Phys.*, 98(2):1398–1421, 1993.

[202] Todd J. Martínez and Emily A. Carter. Pseudospectral møller–plesset perturbation theory through third order. *J. Chem. Phys.*, 100(5):3631–3638, 1994.

[203] Richard A. Friesner, Robert B. Murphy, Michael D. Beachy, Murco N. Ringnalda, W. Thomas Pollard, Barry D. Dunietz, and Yixiang Cao. Correlated ab initio electronic structure calculations for large molecules. *J. Phys. Chem. A*, 103(13):1913–1928, 1999.

[204] Chaehyuk Ko, David K. Malick, Dale A. Braden, Richard A. Friesner, and Todd J. Martínez. Pseudospectral time-dependent density functional theory. *J. Chem. Phys.*, 128(10):–, 2008.

[205] Frank Neese, Frank Wennmohs, Andreas Hansen, and Ute Becker. Efficient, approximate and parallel hartree–fock and hybrid {DFT} calculations. a 'chain-of-spheres' algorithm for the hartree–fock exchange. *Chem. Phys.*, 356(1–3):98 – 109, 2009.

[206] Edward G. Hohenstein, Robert M. Parrish, and Todd J. Martínez. Tensor hypercontraction density fitting. i. quartic scaling second- and third-order møller-plesset perturbation theory. *J. Chem. Phys.*, 137(4):–, 2012.

[207] Robert M. Parrish, Edward G. Hohenstein, Todd J. Martínez, and C. David Sherrill. Tensor hypercontraction. ii. least-squares renormalization. *J. Chem. Phys.*, 137(22):–, 2012.

[208] Edward G. Hohenstein, Robert M. Parrish, C. David Sherrill, and Todd J. Martínez. Communication: Tensor hypercontraction. III. least-squares tensor hypercontraction for the determination of correlated wavefunctions. *J. Chem. Phys.*, 137(22):–, 2012.

[209] Rolf Heilemann Myhre, Alfredo M. J. Sánchez de Merás, and Henrik Koch. The extended CC2 model ECC2. *Mol. Phys.*, 111:1109–1118, 2013.

[210] Jeppe Olsen. The initial implementation and applications of a general active space coupled cluster method. *J. Chem. Phys.*, 113:7140, 2000.

[211] Stefan Dapprich, István Komáromi, K.Suzie Byun, Keiji Morokuma, and Michael J Frisch. A new ONIOM implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM*, 461–462(0):1 – 21, 1999.

[212] Per-Olov Löwdin. Some properties of inner projections. *Int. J. of Quant. Chem.*, S4:231–237, 1970.

[213] Nelson H. F. Beebe and Jan Linderberg. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. of Quant. Chem.*, 12:683–705, 1977.

[214] Jonas Boström, Valera Veryazov, Francesco Aquilante, Thomas Bondo Pedersen, and Roland Lindh. Analytical gradients of the second-order møller–plesset energy using cholesky decompositions. *International Journal of Quantum Chemistry*, 114(5):321–327, 2014.

[215] G.Schaftenaar and J.H. Noordik. Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput.-Aided Mol. Design*, 14:123–134, 2000.

[216] Heike Fliegl, Wim Klopper, and Christof Hättig. Coupled-cluster theory with simplified linear-r12 corrections: The ccsd(r12) model. *J. Chem. Phys.*, 122:084107, 2005.

[217] Fumihiko Aiga, Kotoku Sasagane, and Reikichi Itoh. Frequency-dependent hyperpolarizabilities in the brueckner coupled-cluster theory. *Int. J. Quantum Chem.*, 51(2):87–97, 1994.

[218] Christof Hättig and Florian Weigend. Cc2 excitation energy calculations on large molecules using the resolution of the identity approximation. *J. Chem. Phys.*, 113(13):5154–5161, 2000.

[219] Fu-Ming Tao and Yuh-Kang Pan. Ab initio potential energy curves and binding energies of ar2 and mg2. *Mol. Phys.*, 81(3):507–518, 1994.

[220] Henrik Koch, Berta Fernáández, and Ove Christiansen. The benzene–argon complex: A ground and excited state ab initio study. *J. Chem. Phys.*, 108(7):2784–2790, 1998.

[221] National Center for Biotechnology Information. Pubchem Substance Database. `http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?sid=85230839`, accessed July 2013. SID=85230839.

[222] Rolf Heilemann Myhre, Alfredo M. J. Sánchez de Merás, and Henrik Koch. Multi-level coupled cluster theory, supplementary material, 2014. Supplementary Material Document No.

[223] National Center for Biotechnology Information. Pubchem Compound Database. `http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=524240`, accessed July 2013. CID = 524240.

[224] A. D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648–52, 1993.

[225] M. J. Frisch *et al.* Gaussian09 Revision C.01, 2010. Gaussian Inc. Wallingford CT 2009.

[226] Ove Christiansen, Pekka Manninen, Poul Jørgensen, and Jeppe Olsen. Coupled-cluster theory in a projected atomic orbital basis. *J. Chem. Phys.*, 124:084103, 2006.

[227] Mark S. Gordon, Dmitri G. Fedorov, Spencer R. Pruitt, and Lyudmila V. Slipchenko. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.*, 112(1):632–672, 2012. PMID: 21866983.

[228] Wei Li, Zhigang Ni, and Shuhua Li. Cluster-in-molecule local correlation method for post-hartree-fock calculations of large systems. *Mol. Phys.*, 114(9):1447–1460, 2016.

[229] Evgeny Epifanovsky, Dmitry Zuev, Xintian Feng, Kirill Khistyaev, Yihan Shao, and Anna I. Krylov. General implementation of the resolution-of-the-identity and cholesky representations of electron repulsion integrals within coupled-cluster and equation-of-motion methods: Theory and benchmarks. *J. Chem. Phys.*, 139(13), 2013.

[230] Andreas Köhn and Jeppe Olsen. Coupled-cluster with active space selected higher amplitudes: Performance of seminatural orbitals for ground and excited state calculations. *J. Chem. Phys.*, 125:174110, 2006.

[231] Tomasz Adam Wesolowski and Arieh Warshel. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.*, 97(30):8050–8053, 1993.

[232] National Center for Biotechnology Information. Pubchem Compound Database. `http://pubchem.ncbi.nlm.nih.gov/compound/12298`, accessed January 2015. CID = 12298.

[233] National Center for Biotechnology Information. Pubchem Compound Database. `https://pubchem.ncbi.nlm.nih.gov/compound/638071`, accessed Nov. 12 2015. CID = 638071.

[234] John C. Slater. Statistical exchange-correlation in the self-consistent field. In Per-Olov Löwdin, editor, *Advances in Quantum Chemistry*, volume 6 of *Adv. Quant. Chem.*, pages 1 – 92. Academic Press, 1972.

[235] J. C. Slater and K. H. Johnson. Self-consistent-field $x\alpha$ cluster method for polyatomic molecules and solids. *Phys. Rev. B*, 5:844–853, Feb 1972.

[236] M. Stener, A. Lisini, and P. Decleva. Density functional calculations of excitation energies and oscillator strengths for C1s $\rightarrow \pi^*$ and O1s $\rightarrow \pi^*$ excitations and ionization potentials in carbonyl containing molecules. *Chem. Phys.*, 191(1–3):141 – 154, 1995.

[237] L. Triguero, L. G. M. Pettersson, and H. Ågren. Calculations of near-edge x-ray-absorption spectra of gas-phase and chemisorbed molecules by means of density-functional and transition-potential theory. *Phys. Rev. B*, 58:8097–8110, Sep 1998.

[238] C. R. Natoli, D. K. Misemer, S. Doniach, and F. W. Kutzler. First-principles calculation of x-ray absorption-edge structure in molecular clusters. *Phys. Rev. A*, 22:1104–1108, Sep 1980.

[239] L Fonda. Multiple-scattering theory of x-ray absorption: a review. *J. Phys.: Condens. Mat.*, 4(43):8269, 1992.

[240] A. L. Ankudinov, B. Ravel, J. J. Rehr, and S. D. Conradson. Real-space multiple-scattering calculation and interpretation of x-ray-absorption near-edge structure. *Phys. Rev. B*, 58:7565–7576, Sep 1998.

[241] Peter Schwerdtfeger. The pseudopotential approximation in electronic structure theory. *ChemPhysChem*, 12(17):3143–3155, 2011.

[242] Mathieu Taillefumier, Delphine Cabaret, Anne-Marie Flank, and Francesco Mauri. X-ray absorption near-edge structure calculations with the pseudopotentials: Application to the K edge in diamond and α-quartz. *Phys. Rev. B*, 66:195107, Nov 2002.

[243] P. S. Bagus. Self-consistent-field wave functions for hole states of some ne-like and ar-like ions. *Phys. Rev.*, 139:A619–A634, Aug 1965.

[244] T E Meehan, K Hermann, and F P Larkins. Ab initio calculations of resonant x-ray emission spectra for the n_2o molecule. i. oxygen k spectra. *J. Phys. B: At. Mol. Opt.*, 28(3):357, 1995.

[245] Ulf Ekström, Patrick Norman, Vincenzo Carravetta, and Hans Ågren. Polarization propagator for x-ray spectra. *Phys. Rev. Lett.*, 97:143001, Oct 2006.

[246] Ulf Ekström and Patrick Norman. X-ray absorption spectra from the resonant-convergent first-order polarization propagator approach. *Phys. Rev. A*, 74:042722, Oct 2006.

[247] Nicholas A. Besley and Frans A. Asmuruf. Time-dependent density functional theory calculations of the spectroscopy of core electrons. *Phys. Chem. Chem. Phys.*, 12:12024–12039, 2010.

[248] M Stener, G Fronzoni, and M de Simone. Time dependent density functional theory of core electrons excitations. *Chem. Phys. Lett.*, 373(1–2):115 – 123, 2003.

[249] Yu Zhang, Jason D. Biggs, Daniel Healion, Niranjan Govind, and Shaul Mukamel. Core and valence excitations in resonant x-ray spectroscopy using restricted excitation window time-dependent density functional theory. *J. Chem. Phys.*, 137(19), 2012.

[250] Jan Wenzel, Michael Wormit, and Andreas Dreuw. Calculating core-level excitations and x-ray absorption spectra of medium-sized closed-shell molecules with the algebraic-diagrammatic construction scheme for the polarization propagator. *J. Comput. Chem.*, 35(26):1900–1915, 2014.

[251] Jan Wenzel, Andre Holzer, Michael Wormit, and Andreas Dreuw. Analysis and comparison of cvs-adc approaches up to third order for the calculation of core-excited states. *J. Chem. Phys.*, 142(21), 2015.

[252] Jan Wenzel, Michael Wormit, and Andreas Dreuw. Calculating x-ray absorption spectra of open-shell molecules with the unrestricted algebraic-diagrammatic construction scheme for the polarization propagator. *J. Chem. Theory Comput.*, 10:4583–4598, 2014.

[253] A.B. Trofimov, I.L. Krivdina, J. Weller, and J. Schirmer. Algebraic-diagrammatic construction propagator approach to molecular response properties. *Chem. Phys.*, 329(1–3):1 – 10, 2006.

[254] Marcel Nooijen and Rodney J. Bartlett. Description of core-excitation spectra by the open-shell electron-attachment equation-of-motion coupled cluster method. *J. Chem. Phys.*, 102(17):6735–6756, 1995.

[255] Thomas Fransson, Sonia Coriani, Ove Christiansen, and Patrick Norman. Carbon x-ray absorption spectra of fluoroethenes and acetone: A study at the coupled cluster, density functional, and static-exchange levels of theory. *J. Chem. Phys.*, 138(12), 2013.

[256] Nanna Holmgaard List, Sonia Coriani, Jacob Kongsted, and Ove Christiansen. Lanczos-driven coupled–cluster damped linear response theory for molecules in polarizable environments. *J. Chem. Phys.*, 141(24), 2014.

[257] Joanna Kauczor, Patrick Norman, Ove Christiansen, and Sonia Coriani. Communication: A reduced-space algorithm for the solution of the complex linear response equations used in coupled cluster damped response theory. *J. Chem. Phys.*, 139(21):211102, 2013.

[258] Bo Peng, Patrick J. Lestrange, Joshua J. Goings, Marco Caricato, and Xiaosong Li. Energy-specific equation-of-motion coupled-cluster methods for high-energy excited states: Application to k-edge x-ray absorption spectroscopy. *J. Chem. Theory Comp.*, 11(9):4146–4153, 2015. PMID: 26575909.

[259] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Philadelphia, PA, 3rd edition, 1996.

[260] Jan Wenzel and Andreas Dreuw. Physical properties, exciton analysis, and visualization of core-excited states: An intermediate state representation approach. *J. Chem. Theory Comput.*, 12(3):1314–1330, 2016. PMID: 26845396.

[261] H.-D. Meyer and S. Pal. A band-lanczos method for computing matrix elements of a resolvent. *J. Chem. Phys.*, 91(10):6195–6204, 1989.

[262] Kristin Marie Skjelbred, Rolf Heilemann Myhre, and Henrik Koch. MLCCSD response functions. in preparation.

[263] National Center for Biotechnology Information. Pubchem Compound Database. `https://pubchem.ncbi.nlm.nih.gov/compound/261`, accessed January 2016. CID = 261.

[264] K. C. Prince, R. Richter, M. De Simone, and M. Coreno. X-ray absorption spectra of some small polyatomic molecules. *Surf. Rev. Lett.*, 09(01):159–164, 2002.

[265] Rolf H. Myhre and Henrik Koch. The multilevel CC3 coupled cluster model. *J. Chem. Phys.*, 145(4), 2016.

[266] Thomas F. Miller, III and David C. Clary. Quantum free energies of the conformers of glycine on an ab initio potential energy surface. *Phys. Chem. Chem. Phys.*, 6:2563–2571, 2004.

[267] L. E. Snyder, F. J. Lovas, J. M. Hollis, D. N. Friedel, P. R. Jewell, A. Remijan, V. V. Ilyushin, E. A. Alekseev, and S. F. Dyubko. A rigorous attempt to verify interstellar glycine. *Astrophys. J.*, 619(2):914, 2005.

[268] Vincenzo Barone, Malgorzata Biczysko, Julien Bloino, and Cristina Puzzarini. Characterization of the elusive conformers of glycine from state-of-the-art structural, thermodynamic, and spectroscopic computations: Theory complements experiment. *J. Chem. Theory Comput.*, 9(3):1533–1547, 2013. PMID: 26587615.

[269] Vincenzo Barone, Malgorzata Biczysko, Julien Bloino, and Cristina Puzzarini. The performance of composite schemes and hybrid cc/dft model in predicting structure, thermodynamic and spectroscopic parameters: the challenge of the conformational equilibrium in glycine. *Phys. Chem. Chem. Phys.*, 15(25):10094–10111, 2013.

[270] Michelle L. Gordon, Glyn Cooper, Cynthia Morin, Tohru Araki, Cássia C. Turci, Konstantin Kaznatcheev, , and Adam P. Hitchcock. Inner-shell excitation spectroscopy of the peptide bond: Comparison of the c 1s, n 1s, and o 1s spectra of glycine, glycyl-glycine, and glycyl-glycyl-glycine. *J. Phys. Chem. A*, 107(32):6144–6159, 2003.

[271] Oksana Plekan, Vitaliy Feyer, Robert Richter, Marcello Coreno, Monica de Simone, Kevin C. Prince, , and Vincenzo Carravetta. Investigation of the amino acids glycine, proline, and methionine by photoemission spectroscopy. *J. Phys. Chem. A*, 111(43):10998–11005, 2007. PMID: 17918919.

[272] B. M. Messer, C. D. Cappa, J. D. Smith, K. R. Wilson, M. K. Gilles, R. C. Cohen, and R. J. Saykally. ph dependence of the electronic structure of glycine. *J. Phys. Chem. B*, 109(11):5375–5382, 2005. PMID: 16863204.

[273] K. Kaznacheyev, A. Osanna, , C. Jacobsen, O. Plashkevych, O. Vahtras, , H.Ågren, V. Carravetta, and A. P. Hitchcock. Innershell absorption spectroscopy of amino acids. *J. Phys. Chem. A*, 106(13):3153–3168, 2002.

[274] Y. Zubavichus, M. Zharnikov, A. Schaporenko, and M. Grunze. NEXAFS study of glycine and glycine-based oligopeptides. *J. Electron Spectrosc.*, 134(1):25 – 33, 2004.

[275] Andreas Dreuw and Michael Wormit. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *WIREs Comput Mol Sci*, 5:82–95, 2015.

[276] Reidar Arneberg, Jiri Müller, and Rolf Manne. Configuration interaction calculations of satellite structure in photoelectron spectra of h2o. *Chem. Phys.*, 64(2):249 – 258, 1982.

[277] C. Melania Oana and Anna I. Krylov. Dyson orbitals for ionization from the ground and electronically excited states within equation-of-motion coupled-cluster formalism: Theory, implementation, and examples. *J. Chem. Phys.*, 127(23), 2007.

[278] Anders Osted, Jacob Kongsted, and Ove Christiansen. Theoretical Study of the Electronic Gas-Phase Spectrum of Glycine, Alanine, and Related Amines and Carboxylic Acids. *J. Phys. Chem. A*, 109:1430–1440, 2005.

[279] Rolf H. Myhre, Thomas J. A. Wolf, Sonia Coriani, Markus Gühr, and Henrik Koch. A theoretical and experimental benchmark study of core excited states in N_2. in preparation.

[280] Jan Wenzel, Michael Wormit, and Andreas Dreuw. Calculating core-level excitations and x-ray absorption spectra of medium-sized closed-shell molecules with the algebraic-diagrammatic construction scheme for the polarization propagator. *J. Comput. Chem.*, 35(26):1900–1915, 2014.

[281] Shervin Fatehi, Craig P. Schwartz, Richard J. Saykally, and David Prendergast. Nuclear quantum effects in the structure and lineshapes of the n2 near-edge x-ray absorption fine structure spectrum. *J. Chem. Phys.*, 132(9), 2010.

[282] I. Bâldea, B. Schimmelpfennig, M. Plaschke, J. Rothe, J. Schirmer, A.B. Trofimov, and Th. Fanghänel. C 1s near edge x-ray absorption fine structure (NEXAFS) of substituted benzoic acids - a theoretical and experimental study. *J. Electron Spectrosc.*, 154(3):109 – 118, 2007.

[283] U. Gelius, L. Asplund, E. Basilier, S. Hedman, K. Helenelund, and K. Siegbahn. A high resolution multipurpose esca instrument with x-ray monochromator. *J. Electron. Spectrosc. Relat. Phenom.*, 2:405 – 434, 1973.

[284] Yuan-Chung Cheng and Graham R. Fleming. Dynamics of light harvesting in photosynthesis. *Annu. Rev. Phys. Chem.*, 60(1):241–262, 2009. PMID: 18999996.

[285] Dario Polli, Piero Altoe, Oliver Weingart, Katelyn M. Spillane, Cristian Manzoni, Daniele Brida, Gaia Tomasello, Giorgio Orlandi, Philipp Kukura, Richard A. Mathies, Marco Garavelli, and Giulio Cerullo. Conical intersection dynamics of the primary photoisomerization event in vision. *Nature*, 467:440–443, 2010.

[286] M. P. Minitti, J. M. Budarz, A. Kirrander, J. S. Robinson, D. Ratner, T. J. Lane, D. Zhu, J. M. Glownia, M. Kozina, H. T. Lemke, M. Sikorski, Y. Feng, S. Nelson, K. Saita, B. Stankus, T. Northey, J. B. Hastings, and P. M. Weber. Imaging molecular motion: Femtosecond x-ray scattering of an electrocyclic chemical reaction. *Phys. Rev. Lett.*, 114:255501, Jun 2015.

[287] Jie Yang, Markus Guehr, Theodore Vecchione, Matthew S. Robinson, Renkai Li, Nick Hartmann, Xiaozhe Shen, Ryan Coffee, Jeff Corbett, Alan Fry, Kelly Gaffney, Tais Gorkhover, Carsten Hast, Keith Jobe, Igor Makasyuk, Alexander Reid, Joseph Robinson, Sharon Vetter, Fenglin Wang, Stephen Weathersby, Charles Yoneda, Martin Centurion, and Xijie Wang. Diffractive imaging of a rotational wavepacket in nitrogen molecules with femtosecond megaelectronvolt electron pulses. *Nat. Commun.*, 7, 2016.

[288] Wenkai Zhang, Roberto Alonso-Mori, Uwe Bergmann, Christian Bressler, Matthieu Chollet, Andreas Galler, Wojciech Gawelda, Ryan G. Hadt, Robert W. Hartsock, Thomas Kroll, Kasper S. Kjær, Katharina Kubiček, Henrik T. Lemke, Huiyang W. Liang, Drew A. Meyer, Martin M. Nielsen, Carola Purser, Joseph S. Robinson, Edward I. Solomon, Zheng Sun, Dimosthenis Sokaras, Tim B. van Driel, György Vankó, Tsu-Chien Weng, Diling Zhu, and Kelly J. Gaffney. Tracking excited-state charge and spin dynamics in iron coordination complexes. *Nature*, 509, 2014.

[289] Christian Bressler and Majed Chergui. Molecular structural dynamics probed by ultrafast x-ray absorption spectroscopy. *Annu. Rev. Phys. Chem.*, 61(1):263–282, 2010. PMID: 20055677.

[290] Ph. Wernet, K. Kunnus, I. Josefsson, I. Rajkovic, W. Quevedo, M. Beye, S. Schreck, S. Grübel, M. Scholz, D. Nordlund, W. Zhang, R. W. Hartsock, W. F. Schlotter, J. J. Turner, B. Kennedy, F. Hennies, F. M. F. de Groot, K. J. Gaffney, S. Techert, M. Odelius, and A. Föhlisch. Orbital-specific mapping of the ligand exchange dynamics of fe(co)5 in solution. *Nature*, 520:78–81, 2015.

[291] Matthew J. Comstock, Niv Levy, Armen Kirakosian, Jongweon Cho, Frank Lauterwasser, Jessica H. Harvey, David A. Strubbe, Jean M. J. Fréchet, Dirk Trauner, Steven G. Louie, and Michael F. Crommie. Reversible photomechanical switching of individual engineered molecules at a metallic surface. *Phys. Rev. Lett.*, 99:038301, Jul 2007.

[292] Nobuhiko Hosono, Takashi Kajitani, Takanori Fukushima, Kazuki Ito, Sono Sasaki, Masaki Takata, and Takuzo Aida. Large-area three-dimensional molecular ordering of a polymer brush by one-step processing. *Science*, 330(6005):808–811, 2010.

[293] Roberto Improta, Fabrizio Santoro, and Lluís Blancafort. Quantum mechanical studies on the photophysics and the photochemistry of nucleic acids and nucleobases. *Chem. Rev.*, 116(6):3540–3593, 2016. PMID: 26928320.

[294] Carlos E. Crespo-Hernández, Boiko Cohen, Patrick M. Hare, and Bern Kohler. Ultrafast excited-state dynamics in nucleic acids. *Chem. Rev.*, 104(4):1977–2020, 2004. PMID: 15080719.

[295] Susanne Ullrich, Thomas Schultz, Marek Z. Zgierski, and Albert Stolow. Electronic relaxation dynamics in dna and rna bases studied by time-resolved photoelectron spectroscopy. *Phys. Chem. Chem. Phys.*, 6:2796–2801, 2004.

[296] Martin Klessinger and Josef Michl. *Excited States in Photochemistry of Organic Molecules*. Wiley, New York, 1st edition, 1995.

[297] MA El-Sayed. Spin-orbit coupling and the radiationless processes in nitrogen heterocyclics. *J. Chem. Phys.*, 38(12):2834–2838, 1963.

[298] Joachim Stöhr. *NEXAFS Spectroscopy*. Springer-Verlag Berlin Heidelberg, New York, 1st edition, 1992.

[299] Rolf Manne. Molecular Orbital Interpretation of X-Ray Emission Spectra: Simple Hydrocarbons and Carbon Oxides. *J. Chem. Phys.*, 52(11):5733–5739, 1970.

[300] Karl Kleinermanns, Dana Nachtigallová, and Mattanjah S. de Vries. Excited state dynamics of dna bases. *Int. Rev. Phys. Chem.*, 32(2):308–342, 2013.

[301] Shohei Yamazaki and Tetsuya Taketsugu. Nonradiative deactivation mechanisms of uracil, thymine, and 5-fluorouracil: A comparative ab initio study. *J. Phys. Chem. A*, 116(1):491–503, 2012. PMID: 22171528.

[302] Juan José Serrano-Pérez, Remedios González-Luque, Manuela Merchán, and Luis Serrano-Andrés. On the intrinsic population of the lowest triplet state of thymine. *J. Phys. Chem. B*, 111(41):11880–11883, 2007. PMID: 17927271.

[303] Gunther Zechmann and Mario Barbatti. Photophysics and deactivation pathways of thymine. *J. Phys. Chem. A*, 112(36):8273–8279, 2008. PMID: 18707067.

[304] Valérie Blanchet, Marek Z Zgierski, Tamar Seideman, and Albert Stolow. Discerning vibronic molecular dynamics using time-resolved photoelectron spectroscopy. *Nature*, 401(6748):52–54, 1999.

[305] Paul Hockett, Christer Z Bisgaard, Owen J Clarkin, and Albert Stolow. Time-resolved imaging of purely valence-electron dynamics during a chemical reaction. *Nature Physics*, 7(8):612–615, 2011.

[306] Takuya Horio, Takao Fuji, Yoshi-Ichi Suzuki, and Toshinori Suzuki. Probing ultrafast internal conversion through conical intersection via time-energy map of photoelectron angular anisotropy. *J. Am. Chem. Soc.*, 131(30):10392–10393, 2009. PMID: 19591468.

[307] Georgi L. Dakovski, Philip Heimann, Michael Holmes, Oleg Krupin, Michael P. Minitti, Ankush Mitra, Stefan Moeller, Michael Rowen, William F. Schlotter, and Joshua J. Turner. The Soft X-ray Research instrument at the Linac Coherent Light Source. *J. Synchrotron Radiat.*, 22(3):498–502, May 2015.

[308] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, Ph. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H. D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky, and J Galayda. First lasing and operation of an ångstrom-wavelength free-electron laser. *Nat. Photon.*, 4(9):641–647, 2010.

[309] B. K. McFarland, N. Berrah, C. Bostedt, J. Bozek, P. H. Bucksbaum, J. C. Castagna, R. N. Coffee, J. P. Cryan, L. Fang, J. P. Farrell, R. Feifel, K. J. Gaffney, J. M. Glownia, T. J. Martínez, S. Miyabe, M. Mucke, B. Murphy, A. Natan, T. Osipov, V. S. Petrovic, S. Schorb, Th. Schultz, L. S. Spector, M. Swiggers, F. Tarantelli, I. Tenney, S. Wang, J. L. White, W. White, and M. Gühr. Experimental strategies for optical pump – soft x-ray probe experiments at the lcls. *J. Phys.: Conf. Ser.*, 488(1):012015, 2014.

[310] Philip Heimann, Oleg Krupin, William F. Schlotter, Joshua Turner, Jacek Krzywinski, Florian Sorgenfrei, Marc Messerschmidt, David Bernstein, Jaromir Chalupský, Vera Hájková, Stefan Hau-Riege, Michael Holmes, Libor Juha, Nicholas Kelez, Jan Lüning, Dennis Nordlund, Monica Fernandez Perea, Andreas Scherz, Regina Soufli, Wilfried Wurth, and Michael Rowen. Linac coherent light source soft x-ray materials science instrument optical design and monochromator commissioning. *Rev. Sci. Instrum.*, 82(9), 2011.

[311] L. J. Frasinski, V. Zhaunerchyk, M. Mucke, R. J. Squibb, M. Siano, J. H. D. Eland, P. Linusson, P. v.d. Meulen, P. Salén, R. D. Thomas, M. Larsson, L. Foucar, J. Ullrich, K. Motomura, S. Mondal, K. Ueda, T. Osipov, L. Fang, B. F. Murphy, N. Berrah, C. Bostedt, J. D. Bozek, S. Schorb, M. Messerschmidt, J. M. Glownia, J. P. Cryan, R. N. Coffee, O. Takahashi, S. Wada, M. N. Piancastelli, R. Richter, K. C. Prince, and R. Feifel. Dynamics of hollow atom formation in intense x-ray pulses probed by partial covariance mapping. *Phys. Rev. Lett.*, 111:073002, Aug 2013.

[312] M. Beye, O. Krupin, G. Hays, A. H. Reid, D. Rupp, S. de Jong, S. Lee, W.-S. Lee, Y.-D. Chuang, R. Coffee, J. P. Cryan, J. M. Glownia, A. Föhlisch, M. R. Holmes, A. R. Fry, W. E. White, C. Bostedt, A. O. Scherz, H. A. Durr, and W. F. Schlotter. X-ray pulse preserving single-shot optical cross-correlation method for improved experimental temporal resolution. *Appl. Phys. Lett.*, 100(12), 2012.

[313] Mario Barbatti, Adélia J. A. Aquino, Jaroslaw J. Szymczak, Dana Nachtigallová, Pavel Hobza, Hans Lischka, and Josef Michl. Relaxation mechanisms of uv-photoexcited dna and rna nucleobases. *Proc. Natl. Acad. Sci. U. A.*, 107(50):21453–21458, 2010.

[314] David Asturiol, Benjamin Lasorne, Michael A. Robb, and Lluís Blancafort. Photophysics of the π, π^* and n, π^* states of thymine: MS-CASPT2 minimum-energy paths and CASSCF on-the-fly dynamics. *J. Phys. Chem. A*, 113(38):10211–10218, 2009. PMID: 19722485.

[315] Serhiy Perun, Andrzej L. Sobolewski, and Wolfgang Domcke. Conical intersections in thymine. *J. Phys. Chem. A*, 110(49):13238–13244, 2006. PMID: 17149840.

[316] Jürgen Gauss and John F. Stanton. Analytic CCSD(T) second derivatives. *Chem. Phys. Lett.*, 276(1):70 – 77, 1997.

[317] John F. Stanton and Jürgen Gauss. Analytic energy gradients for the equation-of-motion coupled-cluster method: Implementation and application to the hcn/hnc system. *J. Chem. Phys.*, 100(6):4695–4698, 1994.