HIGH-LEVEL COUPLED-CLUSTER ENERGETICS USING SELECTED-CONFIGURATION-INTERACTION-DRIVEN AND ADAPTIVE MOMENT EXPANSIONS

By

Karthik Gururangan

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry—Doctor of Philosophy

2025

ABSTRACT

It is well established that size-extensive approaches based on the exponential wave function ansatz of coupled-cluster (CC) theory and their extensions to open-shell, multiconfigurational, and excited states are among the best ways of treating many-electron correlation effects. This is especially true in applications involving structural and spectroscopic properties of molecular systems, chemical reaction pathways, noncovalent interactions, and photochemistry. In all these and similar cases, the CC hierarchy, including CCSD, CCSDT, CCSDTQ, and so on, and its equation-of-motion (EOM) and linear-response extensions rapidly converge to the exact, full configuration interaction (CI) limit. Unfortunately, the CCSDT, CCSDTQ, and similar methods, needed to achieve a quantitative description, incur a steep increase in computational costs, rendering their application to larger many-electron systems prohibitively expensive. To address this challenge, this dissertation explores the development of computationally practical alternative approaches based on the CC(P;Q) and externally corrected CC (ec-CC) frameworks, which allow us to study challenging chemical problems, such as molecular bond breaking, biradicals, transition states, and excited states dominated by two- and other many-electron excitations, while avoiding the well-known failures of perturbative CC approximations in the presence of electronic quasi-degeneracies.

The first part of this dissertation introduces two novel CC(P;Q) approaches capable of rapidly converging high-level CC/EOMCC energetics of the CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and similar types, at tiny fractions of the computational costs in an automated fashion, even in cases of stronger correlations. The first methodology combines the CC(P;Q) moment expansions with the information provided by selected CI wave functions obtained using the algorithm abbreviated as CIPSI to systematically recover the results corresponding to any desired level of CC/EOMCC theory. The second approach, called adaptive CC(P;Q), eliminates the reliance on active orbitals or external non-CC information adopted in previous formulations of CC(P;Q) by executing a sequence of CC(P;Q) calculations aimed at converging high-level CC/EOMCC energetics guided solely by the

mathematical structure of the moment expansions. We demonstrate the effectiveness of both the CIPSI-driven and adaptive CC(P;Q) methodologies through a number of molecular applications aimed at recovering the full CCSDT/EOMCCSDT energetics when the noniterative triples corrections to CCSD/EOMCCSD struggle or fail. For the CIPSI-driven approach, we examine the dissociation of F_2, the automerization of cyclobutadiene, and the vertical excitation spectrum of CH^+. The adaptive CC(P;Q) approach is tested on the stretched F_2 and F_2^+ molecules, the automerization of cyclobutadiene, singlet–triplet gaps in organic biradicals, the degenerate Cope rearrangement of bullvalene, and the ground- and excited-state potential energy surfaces of water along the O–H bond-breaking coordinate. To illustrate the computational advantages of the adaptive CC(P;Q) approach, as implemented in the open-source CCpy software package, we also discuss the CPU timings characterizing our calculations for cyclobutadiene and C_nH_{2n+2} linear alkanes with $n = 1$–8.

In the final part of this dissertation, we implement and test a new family of ec-CC approaches designed to recover the exact, full CI, energetics. These methods leverage information about higher-order correlations provided by selected CI wave functions and adopt moment expansions, similar to those used in CC(P;Q), to account for missing higher-order correlation effects. In this new class of ec-CC approaches, termed ec-CC-II, one solves CCSD-like equations for the one- and two-body clusters in the presence of their three-body (T_3) and four-body (T_4) counterparts extracted from the underlying CI wave function, after discarding T_3 and T_4 amplitudes corresponding to CI coefficients that are zero. In this dissertation, we focus on the ec-CC-II approach using T_3 and T_4 clusters extracted from CIPSI calculations, along with its ec-CC-II$_3$ extension, which corrects the ec-CC-II energetics for missing T_3 correlation effects using the appropriately defined CC(P;Q)-like moment expansions. To assess the performance of the CIPSI-based ec-CC-II and ec-CC-II$_3$ methodologies, we apply them to the challenging symmetric dissociation in water, where even high-level CC methods, such as CCSDTQ, struggle to achieve a full-CI-level description.

This thesis is dedicated to my grandparents, Gururangan Gopalaswamy, Vasantha Gururangan, Savitiri Subramanian, and C.N. Subramanian, whose unwavering love and support have been a constant presence throughout my life.

ACKNOWLEDGMENTS

To begin with, I would like to thank my doctoral advisor, Professor Piotr Piecuch. Since my undergraduate days, I have been drawn to the idea of working as a theorist, and I am deeply grateful to Professor Piecuch for giving me the opportunity to become one by introducing me to many-body theory and providing challenging problems for me to work on, while offering constant guidance and intellectual inspiration along the way. Above all, I am thankful for his emphasis on precision and attention to detail in science, as well as for demonstrating the work ethic and discipline required to excel at the highest level. Under his mentorship, I was able to explore a variety of topics in coupled-cluster theory, which allowed me to grow significantly as a scientist and gain confidence in tackling complex problems.

I would also like to thank the remaining members of my guidance committee, namely, Professor Katharine C. Hunt, Professor Angela K. Wilson, and Professor Warren F. Beck, for their support, advice, and patience over the course of my doctoral studies.

I would like to express my sincere gratitude to Professor Elad Harel for introducing me to physical chemistry and giving me the opportunity to work on spectroscopic simulations as an M.S. student in his laboratory. Professor Harel helped foster my interest in quantum mechanics and played a pivotal role in shaping the beginning of my graduate journey. In particular, he instilled in me an active approach to learning, where understanding comes through implementation, a principle that proved invaluable throughout my doctoral studies.

I would like to thank our collaborator, Professor Achintya Kumar Dutta from IIT Mumbai, for many helpful discussions about cost-reduced and relativistic coupled-cluster methods. I also want to thank Dr. Anthony Scemama, and the rest of the team at the CNRS in Université Paul Sabatier (Toulouse III), for developing the CIPSI code in Quantum Package, which was used extensively throughout my research. I am especially grateful to Dr. Scemama for several helpful discussions regarding the design of efficient sparse coupled-cluster codes.

I would also like to extend my thanks to both former and current members of the Piecuch group, with whom I have had the pleasure of working, including Dr. Jun Shen, Dr. J. Emiliano

Deustua, Dr. Ilias Magoulas, Dr. Stephen Yuwono, Dr. Arnab Chakraborty, Ms. Swati S. Priyadarsini, Mr. Tiange Deng, and Mr. Agnibha Hanra. I feel fortunate that we have remained good friends over the years. My sincere gratitude goes to Dr. J. Emiliano Deustua, my first mentor, for helping me navigate the beginning of graduate school and introducing me to scientific programming. I shudder to think how differently my doctoral studies may have turned out if I did not pick the office in Room 87, and I am grateful that we have remained friends and collaborators to this day. I am also deeply thankful to Dr. Ilias Magoulas for his invaluable collaboration during my first project on externally corrected coupled-cluster methods and for always providing kind and thoughtful answers to all of my questions. His enthusiasm for spin-adaptation and diatomic spectroscopy is truly contagious. Finally, I would like to express my profound gratitude to Dr. Stephen Yuwono for his friendship and for engaging in countless discussions, ranging from excited states and Wigner–Wittmer rules to the occasional brain teaser about many-body theory. I have learned a lot from him.

My experience in graduate school would not have been the same without the support of my friends. Among others, I would like to thank Samanta Dalchand, Elizabeth Dalchand, and her husband, James Tavornwattana, Siddharth Moghe, Ashvin Nair, Saavan Patel, Aditya Pradhan, Siddharth Ramakrishnan, Jackson Rigley, Mustafa Sohail, and Shawn Zhao (né Irgen-Gioro) for enriching my life with many good memories, experiences, and laughs. I would also like to thank my friends at Michigan State University, especially Nathan Jansen and Mejdi Mogannam, for their support through the thick and thin of graduate school.

Above all, I am most grateful to my family. To my parents, Raghu Gururangan and Gayatri Gururangan, grandparents, Gururangan Gopalaswamy, Vasantha Gururangan, Savitiri Subramanian, and C.N. Subramanian (*deceased*), brother, Kapil Gururangan, and his wife, Charisma Hooda, thank you all for your unending love and support from near and far while I completed my doctoral studies. To Naomi, words cannot do justice to how much you have done for me over the last six years. Thank you for giving me the strength to keep going, especially during the hardest times.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The primary objective of quantum chemistry is to solve the many-electron Schrödinger equation for a molecular system. The time-independent Schrödinger equation, $H|\Psi_\mu\rangle = E_\mu|\Psi_\mu\rangle$, represents an eigenvalue problem that determines the electronic wave functions $|\Psi_\mu\rangle$ and their corresponding energies E_μ, where $\mu = 0$ denotes the ground state and $\mu > 0$ represents excited states. The nonrelativistic electronic Hamiltonian, expressed in atomic units,

$$H = \sum_{i=1}^{N} \left(-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} \right) + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{1.1}$$

describes a system consisting of N electrons (with spatial coordinates $\mathbf{r}_i$) and M nuclei (with fixed coordinates $\mathbf{R}_A$ and charges Z_A) interacting electrostatically within the Born–Oppenheimer approximation. Extensions of Eq. (1.1), which include the scalar-relativistic and spin-orbit coupling effects, can be derived starting from the one-electron Dirac equation (see Ref. [1] and references therein for further details), but in this dissertation, we focus on a nonrelativistic description of chemical systems. From a many-body perspective, the electronic Hamiltonian consists of one-body terms, which include the electronic kinetic energy and electron-nuclear attraction, and two-body terms arising from electron-electron repulsion. Although these interactions appear simple, the presence of two-body terms in Eq. (1.1) couples the coordinates of all electrons, rendering the many-electron problem analytically intractable. As a result, numerical approaches have long been recognized as the most effective means of solving the many-electron Schrödinger equation.

One of the earliest breakthroughs in solving the many-electron problem came in 1927, just one year after the Schrödinger equation was discovered, when Heitler and London explained covalent bonding in H_2 using a valence-bond ansatz [2]. Shortly thereafter, Slater introduced a determinantal representation for many-fermion states [3] using an antisymmetrized product of single-particle functions, which are called molecular spinorbitals. Building on this idea, Fock applied the variational theorem in conjunction with a Slater determinant ansatz for the

many-electron wave function, leading to the formulation of the Hartree–Fock (HF) method [4]. In the HF approach, the spinorbitals are determined as eigenfunctions of an effective one-electron operator called the Fock operator, which incorporates electron-electron interactions in an approximate fashion. Initially, the HF equations were numerically intractable for all but the simplest systems. However, in 1951, Hall [5] and Roothaan [6] introduced a linear combination of atomic orbitals ansatz for the spinorbitals in the HF method, recasting the complex HF equations into a simple matrix eigenvalue problem within a basis of atomic orbitals. In the Roothaan–Hall formulation, the HF equations are solved by diagonalizing the Fock matrix in this basis, yielding the expansion coefficients and orbital energies that define the resulting spinorbitals, which are iteratively determined in a self-consistent field (SCF) procedure. While the HF method typically recovers most of the total electronic energy, it is well known [7] that the *many-electron correlation effects* missing from the HF mean-field solution are essential for accurately describing chemical transformations and properties. Consequently, a central challenge in quantum chemistry is the development of more advanced methodologies that systematically recover these crucial post-HF correlation effects.

Today, electronic structure calculations of this type are pivotal in investigations across chemistry, physics, and materials science, with a number of widely used computational codes making these techniques readily accessible to both theoreticians and experimentalists. However, despite nearly a century of advances, many contemporary quantum chemical methods struggle to accurately describe quasi-degenerate electronic states in instances involving multiple bond breaking, biradical and polyradical species, and excited states dominated by two- and other many-electron transitions. A defining feature of these challenging systems is the presence of significant multireference (MR), or nondynamical, correlation effects, where multiple quasi-degenerate Slater determinants contribute substantially to the electronic wave function. Such effects commonly arise in bond-breaking processes, open-shell singlet states, and various other low-spin configurations. In addition to capturing MR correlations, it is also essential to account for dynamical correlation effects, which are associated with the

motion of electrons avoiding one another due to Coulombic repulsion, and typically manifest themselves in numerous Slater determinants with small weight in the electronic wave function. In practice, achieving highly accurate descriptions of molecular potential energy surfaces (PESs), properties, and electronic excitation spectra requires a balanced and accurate treatment of both types of correlation.

In principle, all of the aforementioned challenges are addressed by resorting to the full configuration interaction (FCI) approach, which provides exact solutions to the electronic Schrödinger equation, described using a finite basis of spinorbitals, by diagonalizing the Hamiltonian within the full many-electron Hilbert space. The numerically exact energies E_μ and wave functions $|\Psi_\mu\rangle$ can then be obtained by performing FCI calculations using a single-particle basis set that is large enough to converge the complete basis set (CBS) limit, or by extrapolating the CBS limit from results obtained in smaller basis sets. Unfortunately, the dimensionality of the many-electron Hilbert space, which for an N-electron system characterized by K orbitals and total spin S is given by the Weyl–Paldus formula [8, 9], $\frac{2S+1}{K+1}\binom{K+1}{N/2-S}\binom{K+1}{N/2+S+1}$, grows rapidly with the number of electrons and orbitals, rendering the FCI approach inapplicable to systems with more than a few correlated electrons, even when smaller basis sets are employed. For example, in the case of a benzene molecule described using the cc-pVDZ basis set [10], which contains 30 correlated electrons and 108 orbitals (using the frozen-core approximation), the singlet FCI space contains $\sim 10^{34}$ configurations. To put this number in perspective, the largest FCI calculation to date was performed for the propane molecule in a minimum STO-3G basis set [11], which involved only 1.31×10^{12} determinants [12]. Needless to say, the routine use of FCI in realistic chemical problems is impossible given our current computational capabilities. Thus, the central activity in quantum chemistry, which is also a focus of this dissertation, is the development of practical approximations to FCI that provide an accurate description of a wide range of many-electron correlation effects without incurring prohibitive computational costs.

Many of the most successful quantum chemical methodologies in this category are based

on the exponential wave function ansatz [13, 14] of the coupled-cluster (CC) theory [15–20], which offers excellent balance between accuracy and computational cost along with several other desirable mathematical properties. The CC theory initially emerged as a generalization of the many-body perturbation theory (MBPT) expansion, where linked wave function and connected energy diagrams are summed to infinite order, in accordance with the linked [13, 14, 21, 22] and connected [13, 14] cluster theorems. These two theorems ensure that the resulting CC methods satisfy several important theoretical conditions. First, by summing only connected energy diagrams, CC approaches remain size extensive at all levels of approximation, ensuring that accuracy is preserved as the system size increases. Furthermore, due to the exponential form of the wave operator, the CC wave function is size consistent (or separable) in the noninteracting limit if the underlying reference state also separates, which is useful when describing fragmentation phenomena. Most importantly, the CC wave function captures higher-order correlations via products of lower-rank excitation operators. As long as the number of strongly correlated electrons is not too large, this structure ensures rapid convergence toward the exact, FCI, limit while keeping computational costs manageable.

Within the CC framework, there exists two different formalisms depending on the dimensionality of the model space used to provide a zeroth-order description of the many-electron wave function. If the model space is one-dimensional, spanned by a single Slater determinant or configuration state function serving as the Fermi vacuum, the resulting CC methodology is classified as single-reference (SR). In the SRCC theory, which is historically the oldest type of CC formalism, the remaining determinants or configurations entering the many-electron wave function are generated via particle–hole excitations out of the Fermi vacuum. In practice, the Fermi vacuum in SRCC considerations can be obtained using a number of black-box independent particle models (IPMs), including the HF wave function in the restricted (RHF), unrestricted (UHF), or restricted open-shell (ROHF) forms, Brückner theory [23–26], or even Kohn–Sham density functional theory (DFT) [27, 28]. This allows the SRCC calculations to be performed with minimal user-specified input. In con-

trast, a MRCC theory adopts a multidimensional model space spanned by several Slater determinants or configuration state functions, which are chosen to capture the dominant multiconfigurational content of the target wave functions. The remaining, mostly dynamical, correlations are described using particle–hole excitations out of the individual Fermi vacua, as in the Hilbert-space MRCC formalism. Alternatively, one can turn to the Fock space MRCC theory and construct the relevant quasi-degenerate wave functions by starting from a closed-shell vacuum state and adding (removing) electrons to (from) it. Given our interest in describing quasi-degenerate electronic states, which often arise when studying bond dissociations, open shells, and electronic excited states, it may seem more appropriate to adopt an MRCC formalism. Unfortunately, there exists no unique parameterization for MRCC wave functions or the wave operators that generate them, and as a result, several MRCC and MRCC-like ideas exist (see Refs. [29–34] for reviews), which broadly belong to either the state-universal, valence-universal, or state-selective category. The various MRCC schemes also require the user-defined active spaces, which must be relatively small in order to keep the computational cost of these methods manageable, but poor choices may still result in loss of accuracy. Furthermore, due to mathematical complexities associated with their numerical stability and convergence [35–41], the genuine MRCC-based methodologies are currently unable to compete with the ease of use and wider applicability characterizing their SRCC counterparts. Thus, many highly successful strategies for designing CC approaches for MR problems adopt the SRCC perspective, recovering nondynamical correlations dynamically via inclusion of higher-rank particle–hole excitations out of the Fermi vacuum, possibly augmented by particle-nonconserving operators (see, for example, Refs. [42–72]).

In the SRCC formalism, the exact N-electron ground-state wave function is expressed as $|\Psi_0\rangle = e^T|\Phi\rangle$, where $|\Phi\rangle$ is a reference determinant serving as a Fermi vacuum, which is typically the HF state, and $T = \sum_{n=1}^{N} T_n$ is the cluster operator. The many-body components of the cluster operator, T_n, act on $|\Phi\rangle$ to generate the connected contributions to the ground-state wave function corresponding to n-tuply excited determinants, while products of the T_n

clusters resulting from the exponential e^T produce the remaining disconnected but linked contributions to $|\Psi_0\rangle$. The CC ansatz is formally exact (*i.e.*, equivalent to FCI) when all many-body components of the cluster operator are included in the definition of T, however, in practice, the cluster operator is truncated at a lower many-body rank to generate the standard hierarchy of CC approximations. The most basic and practical level of CC theory, obtained when T is truncated at T_2, is the CC method with singles and doubles (CCSD) [73–76], which is characterized by computational steps that scale as $n_o^2 n_u^4$, where n_o (n_u) is the number of correlated occupied (unoccupied) orbitals in $|\Phi\rangle$, or as $\mathscr{N}^6$ with the system size $\mathscr{N}$. The next two levels, namely the CC approach with singles, doubles, and triples, abbreviated as CCSDT [77–80], obtained when T is truncated at T_3, and the CC method with singles, doubles, triples, and quadruples, abbreviated as CCSDTQ [81–84], in which T is truncated at T_4, involve the $n_o^3 n_u^5$ ($\mathscr{N}^8$) and $n_o^4 n_u^6$ ($\mathscr{N}^{10}$) steps, respectively. Higher-level CC approaches that include T_n clusters with $n > 4$ can be analogously defined and come with increasingly larger computational costs scaling as $n_o^n n_u^{n+2}$ ($\mathscr{N}^{2n+2}$).

The ground-state CC formulation can also be extended to describe excited electronic states using the equation-of-motion (EOM) CC formalism [85–89] or its linear-response (LR) CC [90–98] and symmetry-adapted cluster CI (SAC-CI) [99] analogs. In the EOMCC framework, which is the approach adopted in this dissertation, the exact excited states of the N-electron system are defined as $|\Psi_\mu\rangle = R_\mu e^T |\Phi\rangle$, where $R_\mu = r_{\mu,0}\mathbf{1} + \sum_{n=1}^{N} R_{\mu,n}$ ($\mu > 0$) is a linear excitation operator ($\mathbf{1}$ is an identity operator). The EOMCC framework provides a transparent route for computing increasingly accurate and systematically improvable excited-state energetics and properties. The standard series of EOMCC approximations obtained by truncating R_μ and T at a given many-body rank include the basic EOMCCSD method [87–89], defined when R_μ (T) is truncated at $R_{\mu,2}$ (T_2), and its higher-level EOMCCSDT [100–104] and EOMCCSDTQ [102, 103, 105, 106] counterparts that include the $R_{\mu,3}$ (T_3) and $R_{\mu,4}$ (T_4) components of R_μ (T), respectively. The aforementioned EOMCC approximations involve the computational steps that scale the same as their ground-state counterparts. In

analogy to the ground-state case, the approximate EOMCC wave functions retain key theoretical features, including size-extensive total energies and size-intensive vertical transition energies, provided that only one fragment is excited in a noninteracting limit.

It is well established that in the majority of chemical applications, including molecules near equilibrium geometries, chemical bond rearrangements involving smaller numbers of strongly correlated electrons, noncovalent interactions, and photochemistry, the conventional CCSD/EOMCCSD, CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, etc. hierarchy rapidly converges the exact, FCI, limit such that when one reaches the CCSDT/EOMCCSDT or CCSDTQ/EOMCCSDTQ levels, the relevant many-electron correlation effects are typically well described [30, 32]. While the CCSD/EOMCCSD methods are computationally practical approaches, they tend to provide only a qualitative level of accuracy for weakly correlated systems, and often struggle to produce meaningful results in the presence of stronger correlations, such as those arising from multiconfigurational ground states or excited states dominated by two-electron transitions. In order to accurately treat MR problems using the SRCC framework, one needs to include the higher–than–two-body components of the T and R_μ operators, as in the CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and higher-order approaches. Sadly, the routine usage of the more powerful SRCC techniques, such as CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ, is hindered by the respective $\mathscr{N}^8$ and $\mathscr{N}^{10}$ computational steps, which limits their applicability to few-atom systems and smaller basis sets. Thus, a key effort in the field of SRCC method development focuses on designing accurate and reliable approximations to CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ that are capable of reducing the above costs, while being more robust than perturbative approaches of the CCSD(T) [107, 108], CCSDT-n [109–111], and CC3 [112] types and their extensions to higher orders [113–116] and excited electronic states [117–121], which are all known to fail in more MR situations [30, 32, 100, 122–125].

In order to accurately approximate high-level CC/EOMCC energetics with reduced computational costs, even in the presence of substantial electronic quasi-degeneracies and MR

correlation effects, we can turn to the CC(P;Q) framework [126–138]. The CC(P;Q) formalism is based on correcting lower-level CC/EOMCC energetics for missing higher-order correlation effects in a general fashion using the noniterative and nonperturbative energy expansions derived using the method of moments of CC equations (MMCC) [122, 123, 126, 139–150]. In order to do this, each CC(P;Q) calculation begins by defining one subspace of the many-electron Hilbert space called the P space, which contains the excited Slater determinants that, in addition to the reference $|\Phi\rangle$, dominate the ground or excited state of interest. The complementary Q space is spanned by a subset of excited determinants absent from the P space, which are used to correct the CC/EOMCC calculations in the P space [henceforth abbreviated as CC(P) for the ground state and EOMCC(P) for excited states] for the missing correlation effects of interest. By varying the definitions of the P and Q spaces entering the CC(P;Q) framework, we are able to devise a variety of noniterative corrections to both conventional and unconventional CC/EOMCC calculations. To date, work from our group has established five different ways of choosing the underlying P and Q spaces, each one resulting in a different type of CC(P;Q) methodology.

In the first, conventional, form of CC(P;Q), the determinants entering the P and Q spaces are selected purely on the basis of many-body rank. The resulting CC(P;Q) methodologies are equivalent to the left-eigenstate completely renormalized (CR) CC and EOMCC approaches [146–149, 151–153], which correct the energies obtained in standard CC/EOMCC calculations (*e.g.*, CCSD/EOMCCSD, CCSDT/EOMCCSDT, etc.) for the missing correlation effects associated with selected T_n and $R_{\mu,n}$ components of T and R_μ. For example, using the CR-CC(2,3) [146–149], CR-CC(2,3)+Q [153, 154], or CR-CC(2,4) [146, 147, 153, 155] corrections to CCSD, one can approximately account for the many-electron correlation effects arising from connected triples or connected triples and quadruples absent in the initial CCSD computation, respectively. In the case of excited states, one can use the EOM extensions of the CR-CC formalism, including the CR-EOMCC(2,3) [148, 151] triples correction to EOM-CCSD or its rigorously size-intensive δ-CR-EOMCC(2,3) modification [152], to account for

missing higher-order correlation effects resulting from the neglect of connected triples in the preceding EOMCCSD calculation. The CR CC/EOMCC triples corrections can be accurate, even when perturbative CCSD(T)-like approaches fail, such as in problems featuring single bond breaking [146, 147, 149, 156–158] or excited states dominated by two-electron transitions [159–161], and they can be performed at the similar noniterative $n_o^3 n_u^4$, or $\mathscr{N}^7$, cost characterizing the CCSD(T) method. That being said, the typical CR CC/EOMCC triples and quadruples corrections to CCSD/EOMCCSD are not entirely robust in the presence of stronger nondynamical correlation effects that may arise when studying more challenging bond rearrangements [126, 129, 134], some biradical or polyradical species [127, 128, 134–137], and certain excited-state potentials along bond-breaking coordinates [133, 138, 161]. Generally speaking, the aforementioned cases cause problems for any methodology based on noniteratively correcting the CCSD/EOMCCSD energetics for missing higher-order correlation effects due to the presence of substantial coupling between the lower-order cluster and excitations components, such as T_1 and T_2 or $R_{\mu,1}$ and $R_{\mu,2}$, and their T_n and $R_{\mu,n}$ counterparts with $n > 2$ [126–138].

One can address the concern of coupling the lower- and higher-order components of the T and R_μ operators by incorporating the leading subsets of higher–than–doubly excited determinants relevant to the problem into the P space and correcting the resulting CC/EOMCC energetics for the remaining correlations of interest using the complementary Q space. By doing this, we allow the lower-rank components of the cluster and excitation operators, such as T_1, T_2, $R_{\mu,1}$, and $R_{\mu,2}$, to relax in the presence of the leading contributions to their higher–than–two-body T_n and $R_{\mu,n}$ counterparts in the CC(P)/EOMCC(P) calculations prior to determining the CC(P;Q) corrections.

The second variant of the CC(P;Q) theory employs active orbitals to select chemically important subsets of higher–than–doubly excited determinants that are included, in addition to all singly and doubly excited determinants, in the P space, resulting in truncated T and R_μ operators that are equivalent to those adopted in active-space CC/EOMCC method-

ologies [46, 66, 84, 100, 101, 124, 125, 162–169]. The hierarchy of active-orbital-based CC(P;Q) schemes includes approaches like CC(t;3), CC(t,q;3), CC(t,q;3,4), and CC(q;4) [126–129, 131, 132, 138], where the active-space CCSDt/EOMCCSDt [84, 100, 101, 125, 162–164, 166–168] [in the case of CC(t;3)], CCSDtq/EOMCCSDtq [84, 164, 167, 169] [in the cases of CC(t,q;3) or CC(t,q;3,4)], or CCSDTq/EOMCCSDTq [131, 132] [in the case of CC(q;4)] equations are first solved to produce the energies that account for the bulk of the nondynamical and some dynamical correlations, which are then corrected for the remaining dynamical correlation effects due to missing triples [in the cases of CC(t;3) and CC(t,q;3)], quadruples [in the case of CC(q;4)], or triples and quadruples [in the case of CC(t,q;3,4)]. As demonstrated in the past [126–129, 131, 132, 138], the active-orbital-based CC(P;Q) approaches can recover the energetics of high-level CC/EOMCC methods, such as CCSDT/EOMCCSDT in the case of CC(t;3), or CCSDTQ/EOMCCSDTQ in the case of CC(t,q;3), CC(t,q;3,4), or CC(q;4), to within sub-millihartree accuracy using fractions of the computational effort associated with the parent CC/EOMCC calculations. Although the active-orbital-based CC(P;Q) undoubtedly provides a robust strategy for obtaining high-level CC/EOMCC energetics, it relies on user-selected active orbitals to define the underlying P spaces, which means that the corresponding CC(P;Q) calculations cannot be performed in a black-box fashion. Therefore, one may wonder if there is a more easy-to-use form of the CC(P;Q) theory that preserves the effectiveness of the CC(t;3), CC(t,q;3), CC(t,q;3,4), CC(q,4), etc. hierarchy and the improvements it offers over the CR CC/EOMCC corrections, while avoiding the use of user- and system-dependent active orbitals to capture the coupling between the T_1, T_2, $R_{\mu,1}$, and $R_{\mu,2}$ components and their T_3, T_4, $R_{\mu,3}$, $R_{\mu,4}$, etc. counterparts.

To address this challenge, a third variant of CC(P;Q) was introduced within the last few years, which automatically determines the P and Q spaces using information extracted from stochastic Quantum Monte Carlo (QMC) wave function propagations of the CI [170–174] or CC [175–178] types in the many-electron Hilbert space. In the resulting hybrid CC(P;Q) approaches, broadly classified as semi-stochastic CC(P;Q) [130, 133, 134, 136], the P space

associated with a given CIQMC or CC Monte Carlo (CCMC) wave function is defined as all singly and doubly excited determinants and a list of stochastically determined higher–than–doubly excited determinants present in the CIQMC or CCMC wave function, thereby removing the need to manually determine the appropriate P space using active orbitals. As in the fully deterministic active-orbital-based CC(P;Q) approaches, such as CC(t;3), CC(t,q;3), CC(t,q;3,4), and CC(q;4), the companion Q space defining the moment correction spans the remaining higher–than–doubly excited determinants absent in the underlying CIQMC or CCMC wave function such that the resulting semi-stochastic CC(P;Q) approach converges to a high-level target CC method. For example, if the P space is set up to contain all singly and doubly excited determinants and the list of triply excited determinants extracted from the stochastic CIQMC or CCMC wave function, and the Q space spans the remaining triply excited determinants absent from the underlying CIQMC or CCMC wave function at a given propagation time, then the resulting semi-stochastic CC(P;Q) approach converges the full CCSDT energetics, much like the fully deterministic CC(t;3) method. Similarly, if the P space is augmented to include all triply and quadruply excited determinants present in the underlying CIQMC or CCMC wave function and the Q space spans the remaining triply and quadruply excited determinants, the semi-stochastic CC(P;Q) computations converge the CCSDTQ energetics, analogous to the CC(t,q;3,4) approach. Previous work has demonstrated that the semi-stochastic CC(P;Q) approach is capable of converging high-level energetics of the CCSDT [130, 133, 134, 136], CCSDTQ [134], and EOMCCSDT [133] types out of the early stages of CIQMC or CCMC wave function propagations, with minimal reliance on user- and system-dependent inputs. More importantly in the context of this dissertation, the semi-stochastic CC(P;Q) represents a major intellectual advancement by introducing the idea of sampling determinants in the many-electron Hilbert space to identify the important components of the CC/EOMCC ansatz. Naturally, this provokes one to ask whether there exists alternative, potentially better, methods for probing the Hilbert space to obtain the pertinent information about leading higher–than–pair correlations.

Answering this question leads us to the first main topic of this dissertation. In Chapter 3, we introduce a fourth variety of CC(P;Q) capable of converging any high-level CC/EOMCC method of interest, called CIPSI-driven CC(P;Q) [135]. The CIPSI-driven CC(P;Q) methodology replaces the QMC framework originally used to define the P and Q spaces in the semi-stochastic CC(P;Q) computations with the selected CI [179–182] algorithm known as the CI method using perturbative selections made iteratively, or CIPSI [181, 183, 184]. The selected CI approaches, which originated from the pioneering efforts of the late 1960s and early 1970s, have enjoyed a significant renewal of interest in recent years, as their modern implementations have demonstrated the ability to capture the bulk of many-electron correlation effects in a computationally efficient and conceptually straightforward manner [183–195]. For our purposes, selected CI represents an appealing alternative to the QMC framework as a provider of the lists of higher–than–doubly excited determinants within CC(P;Q) considerations due to its ability to construct a CI wave function using a well-defined, systematic sequence of Hamiltonian diagonalizations, resulting in the associated and similarly systematic sequence of P and Q spaces that are subject to little or no stochastic noise. Using the examples of bond breaking in the fluorine molecule, automerization of cyclobutadiene, and the vertical excitation spectrum of CH^+, we demonstrate that the CIPSI-driven CC(P;Q) approach is capable of rapidly converging the high-level CCSDT and EOMCCSDT energetics in a completely black-box fashion using small fractions of the computational costs associated with the parent CCSDT/EOMCCSDT calculations, while eliminating the noise due to random sampling associated with CIQMC or CCMC propagations. In fact, the results collected to date indicate that the CIPSI-driven CC(P;Q) method can be even more efficient that its semi-stochastic predecessor in recovering the near-CCSDT/EOMCCSDT data, employing P spaces derived from CI wave functions that are smaller and more compact than their stochastically determined counterparts thanks, in part, to the more robust perturbative selection mechanism adopted in CIPSI. Just as how the semi-stochastic CC(P;Q) methodology led us to consider alternative approaches for sampling the Hilbert space, the CIPSI-driven

CC(P;Q) approach invites us to scrutinize the way in which we take advantage of sparsity in the many-electron wave function. While CIPSI and other selected CI methods have long exploited the sparsity of the FCI wave function using perturbative arguments to identify the subset of determinants important to the CI expansion, it is tempting to consider whether the underlying CC/EOMCC framework adopted in CC(P;Q) can be used to do the same, providing us with an intrinsic mechanism for selecting the dominant determinants entering our P and Q spaces without turning to non-CC considerations.

In exploring this idea, we arrive at the final form of the CC(P;Q) methodology, which we currently believe to be the most promising and powerful version, called adaptive CC(P;Q) [137, 138]. The adaptive CC(P;Q) approach takes us to an entirely new level by freeing us from the user-defined active orbitals and non-CC (CIQMC, CIPSI) or stochastic (CIQMC, CCMC) concepts adopted in the previous CC(P;Q) work. In Chapter 4 of this dissertation, we introduce the adaptive CC(P;Q) framework capable of converging high-level CC/EOMCC energetics of the CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and similar types, even in cases of stronger correlations, such as those characterizing significant bond rearrangements, biradical species, and larger regions of PESs, where higher–than–two-body components of the cluster operator T and excitation operators R_μ become large, in a fully automated, black-box, manner and at small fractions of the computational costs. The key idea driving this new approach is an adaptive selection of the leading determinants or excitation amplitude types needed to define the T_n and, in the case of excited states, $R_{\mu,n}$, components with $n > 2$, which takes advantage of the intrinsic structure of the CC(P;Q) moment expansions and the associated *a posteriori* energy corrections to capture the remaining correlations of interest, without any reference to the user- and system-dependent or non-CC concepts. After presenting the general theoretical and algorithmic details behind the adaptive CC(P;Q) framework in Chapter 4, we provide a comprehensive set of numerical tests designed to demonstrate its efficiency in converging the high-level CCSDT and EOMCCSDT energetics using tiny fractions of the computational effort. The molecular

examples included in our studies are (i) the significantly stretched F_2 and F_2^+ molecules, (ii) the automerization of cyclobutadiene, (iii) the singlet–triplet gaps in the cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane biradicals, (iv) the degenerate Cope rearrangement in bullvalene, and (v) the ground- and excited-state PESs of the water molecule along its O–H bond-breaking coordinate. All of these examples are characterized by large and highly nonperturbative T_3 and $R_{\mu,3}$ effects, which cannot always be handled using non-iterative triples corrections to the CCSD/EOMCCSD energetics. Finally, to highlight the computational benefits offered by the adaptive CC(P;Q) approach, we conclude Chapter 4 with a discussion of the CPU timings characterizing calculations for cyclobutadiene and C_nH_{2n+2} linear alkanes with $n = 1$—8.

While the CC(P;Q) formalism provides a flexible and efficient framework for devising robust schemes aimed at converging any well-defined level of CC/EOMCC theory, there exists another intriguing route to the study of MR problems within the SRCC framework based on approaching the FCI-level ground-state energetics directly using the externally corrected (ec) CC approaches [195–214]. The ec-CC formalism is based on the observation that, for Hamiltonians containing up to two-body interactions, the CC correlation energy is only dependent on the T_1 and T_2 components of the cluster operator, and the CC amplitude equations projected onto singly and doubly excited determinants do not engage higher-rank components of the cluster operator T_n with $n > 4$. Thus, by solving the CC amplitude equations projected on the singly and doubly excited determinants for the T_1 and T_2 clusters in the presence of their exact three- and four-body counterparts extracted from FCI, with the help of the well-known cluster-analysis relations [215], one obtains the exact T_1 and T_2 clusters and, as a consequence, the exact, FCI energy. This suggests that using external non-CC wave functions capable of generating an accurate representation of T_3 and T_4 clusters, and subsequently solving for T_1 and T_2 in their presence, should not only result in correlation energies that improve upon those obtained with CCSD, where T_3 and T_4 are zero, but also substantially improve the results of calculations used to provide the T_3 and T_4 values.

In reality, and as revealed by our recent mathematical analysis of the ec-CC framework in Ref. [214], this is not always the case. In fact, one must be careful in choosing the external non-CC source entering the ec-CC considerations in order to obtain a useful methodology. The external sources of T_3 and T_4 clusters adopted in the ec-CC methods developed prior to our recent study in Ref. [214] include projected unrestricted HF (PUHF) wave functions used to describe strongly correlated systems involved in modeling the metal-insulator transition [196, 197, 202, 210], and wave functions designed to capture nondynamical correlation effects in *ab initio* calculations, including those resulting from valence-bond [198–200], complete-active-space self-consistent-field (CASSCF) [201, 204, 205, 207], MRCI [203, 206, 207, 209], perturbatively selected CI [208], FCIQMC [195, 211], adaptive CI [212], density-matrix renormalization group (DMRG) [213], and heat-bath CI [213] computations. While the majority of these non-CC sources lead to useful ec-CC methodologies, there are instances [212] where the ec-CC computations do not improve the underlying CI energetics at all. This paradox was resolved in Ref. [214], which showed that when using truncated CI wave functions as sources of T_3 and T_4 clusters in ec-CC considerations, one must adopt the so-called ec-CC-II framework, in which the T_3 and T_4 amplitudes associated with triply and quadruply excited determinants absent in the underlying CI state are discarded *a posteriori*.

In Chapter 5, we report a new type of ec-CC-II approach based on incorporating the information about T_3 and T_4 clusters obtained from CIPSI calculations into the ec-CC equations [214]. In addition to testing the CIPSI-driven ec-CC-II approach, we also introduce its ec-CC-II$_3$ extension that noniteratively corrects the ec-CC-II calculations for the effects of missing T_3 correlations using CC(P;Q)-like moment expansions adapted to the ec-CC context. Using the challenging symmetric dissociation of water as an example, we demonstrate that the new generation of CIPSI-based ec-CC-II approaches, especially ec-CC-II$_3$, is competitive with high-level CC and CI methods, such as CCSDTQ and CISDTQPH, in providing near-FCI energetics, even improving upon CCSDTQ when electronic degeneracies become very strong, using computational steps that are more similar to CCSD and CR-CC(2,3).

CHAPTER 2

THEORETICAL BACKGROUND

2.1 Single-Reference Coupled-Cluster Theory and Its Equation-of-Motion Extension to Excited Electronic States

As mentioned in the Introduction, the exact ground state of an N-electron system within the SRCC theory is expressed as

$$|\Psi_0\rangle = e^T|\Phi\rangle. \tag{2.1}$$

Here, $|\Phi\rangle$ is the reference determinant defining the Fermi vacuum and

$$T = \sum_{n=1}^{N} T_n \tag{2.2}$$

is the cluster operator, with

$$T_n = \sum_{\substack{i_1<\ldots<i_n \\ a_1<\ldots<a_n}} t_{a_1\cdots a_n}^{i_1\cdots i_n} E_{i_1\ldots i_n}^{a_1\ldots a_n} \tag{2.3}$$

designating its n-body (n-particle–n-hole, or np-nh) component, in which $t_{a_1\ldots a_n}^{i_1\ldots i_n}$ are the corresponding cluster amplitudes and $E_{i_1\ldots i_n}^{a_1\ldots a_n} = a^{a_1}\ldots a^{a_n}a_{i_n}\ldots a_{i_1}$ are the elementary np-nh excitation operators that generate the excited determinants $|\Phi_{i_1\ldots i_n}^{a_1\ldots a_n}\rangle$ when acting on $|\Phi\rangle$, with a^p (a_p) representing the usual fermionic creation (annihilation) operator associated with the spinorbital $|p\rangle$. After inserting the CC ansatz, Eq. (2.1), into the Schrödinger equation and multiplying on the left by e^{-T}, we obtain the connected cluster form of the Schrödinger equation,

$$\overline{H}|\Phi\rangle = E_0|\Phi\rangle, \tag{2.4}$$

in which the exact ground-state energy E_0 and reference determinant $|\Phi\rangle$ form an eigenpair of the similarity-transformed Hamiltonian

$$\overline{H} = e^{-T}He^T = (He^T)_C, \tag{2.5}$$

where subscript C denotes the connected operator product (in the sense of diagrammatic MBPT). From Eq. (2.4), it follows that the exact, full CC, ground state satisfies

$$\langle \Phi^{a_1 \ldots a_n}_{i_1 \ldots i_n} | \overline{H} | \Phi \rangle = 0,$$
$$i_1 < \ldots < i_n, \;\; a_1 < \ldots < a_n, \;\; n = 1, \ldots, N, \tag{2.6}$$

with corresponding energy obtained by projecting Eq. (2.4) onto the Fermi vacuum

$$E_0 = \langle \Phi | \overline{H} | \Phi \rangle = \langle \Phi | H | \Phi \rangle + \langle \Phi | [H(T_1 + T_2 + \tfrac{1}{2} T_1^2)]_C | \Phi \rangle. \tag{2.7}$$

If we do not apply any truncations to the CC ansatz, then Eqs. (2.4)–(2.7) represent a projective formulation of the FCI eigenvalue problem.

In the case of approximate CC wave functions, we follow the prescription put forward by Čížek [17] and carefully adapt the formulas given in Eqs. (2.5)–(2.7) to obtain a solvable system of equations applicable to truncated forms of T. In particular, if we define the cluster operator truncated at many-body rank m_A as

$$T^{(A)} = \sum_{n=1}^{m_A} T_n, \tag{2.8}$$

then the system of equations used to determine the cluster amplitudes $t^{i_1 \ldots i_n}_{a_1 \ldots a_n}$, $n \leq m_A$, characterizing $T^{(A)}$ is given by the subset of projections considered in Eq. (2.6) corresponding to the content of the truncated cluster operator,

$$\langle \Phi^{a_1 \ldots a_n}_{i_1 \ldots i_n} | \overline{H}^{(A)} | \Phi \rangle = 0,$$
$$i_1 < \ldots < i_n, \;\; a_1 < \ldots < a_n, \;\; n = 1, \ldots, m_A, \tag{2.9}$$

where

$$\overline{H}^{(A)} = e^{-T^{(A)}} H e^{T^{(A)}} = (H e^{T^{(A)}})_C \tag{2.10}$$

is the associated Hamiltonian transformed with $T^{(A)}$. Once $t^{i_1 \ldots i_n}_{a_1 \ldots a_n}$ are obtained by solving the corresponding CC amplitude equations, Eq. (2.9), the ground-state CC energy is determined *a posteriori* as

$$E_0^{(A)} = \langle \Phi | \overline{H}^{(A)} | \Phi \rangle. \tag{2.11}$$

With the projective formulation summarized in Eqs. (2.8)–(2.11), we can carry out any standard CC approximation based on truncating T at a given many-body rank, including CCSD ($m_A = 2$), CCSDT ($m_A = 3$), CCSDTQ ($m_A = 4$), and so on. The working expressions for the amplitude [Eq. (2.9)] and energy [Eq. (2.11)] equations can be derived in a straightforward manner using the generalized Wick's Theorem or, more elegantly, using the techniques of time-independent many-body diagrammatics [17, 18, 216, 217]. In recent decades, the use of automated coding strategies [218–223] for deriving and implementing the CC equations has also become increasingly popular.

Equation (2.9) represents a system of coupled nonlinear equations involving a large number of variables, namely, the cluster amplitudes. In the first applications of the CC theory, the amplitude equations were solved using a quadratically convergent Newton–Raphson optimization algorithm [17]. In later years, the reduced linear equation [224] method became the preferred approach, with quasi-linearization of nonlinear terms containing T_2 used to improve convergence in highly degenerate problems [225]. In many modern CC codes, including those relevant to this dissertation incorporated in GAMESS [226, 227] and CCpy [228] (see Appendix A of this dissertation for information about the CCpy software), Eq. (2.9) is solved using the Jacobi method (an inexact Newton approach) augmented by convergence accelerators, such as the direct inversion of the iterative subspace (DIIS) algorithm [229–231], to help extrapolate the solution, and energy denominator shifts for obtaining convergence in quasi-degenerate situations. Alternative techniques for solving Eq. (2.9) include the CROP algorithm [232], quasi-Newton [233] and Newton–Krylov [234] methods, and matrix eigenvalue formulations of the CC amplitude system based on dressed CI [235, 236]. Before moving on, it is worth emphasizing that for an approximate CC state, *the eigenvalue relationship in Eq. (2.4) no longer holds.* While the projections of $\overline{H}^{(A)}|\Phi\rangle$ onto excited Slater determinants corresponding to the content of $T^{(A)}$ are equal to zero, as implied by Eq. (2.9), the projections onto excitations absent in $T^{(A)}$ are nonzero, and as we will see in Section 2.2, these remaining projections ultimately help define the noniterative corrections

formulated within the CC(P;Q) framework.

As alluded to in the Introduction, we can also study excited electronic states within the SRCC theory using the EOMCC formalism, or alternatively, its LR CC and SAC-CI analogs. In the EOMCC theory, the excited states of an N-electron system are represented as

$$|\Psi_\mu^{(A)}\rangle = R_\mu^{(A)} e^{T^{(A)}} |\Phi\rangle, \tag{2.12}$$

where the linear excitation operator truncated at many-body rank m_A is

$$R_\mu^{(A)} = r_{\mu,0}\mathbf{1} + \sum_{n=1}^{m_A} R_{\mu,n}, \tag{2.13}$$

with constituent n-body components

$$R_{\mu,n} = \sum_{\substack{i_1<\ldots<i_n \\ a_1<\ldots<a_n}} r_{\mu,a_1\cdots a_n}^{\;\;i_1\cdots i_n} E_{i_1\ldots i_n}^{a_1\ldots a_n}. \tag{2.14}$$

To ensure size-intensivity of the resulting vertical excitation energies, the truncation m_A in $R_\mu^{(A)}$ is the same as that used for $T^{(A)}$. The scalar $r_{\mu,0}$ represents the contribution of the reference state $|\Phi\rangle$ to the μth electronic state $|\Psi_\mu^{(A)}\rangle$. Given that $|\Phi\rangle$ is typically optimized to describe the ground state, its weight in $|\Psi_\mu^{(A)}\rangle$ should be small, if $|\Psi_0^{(A)}\rangle$ and $|\Psi_\mu^{(A)}\rangle$ belong to the same symmetry, or strictly zero, if they belong to different symmetries. For convenience, we define $R_0^{(A)} = \mathbf{1}$ so that Eq. (2.12) can be used to express the CC ansatz for the ground state ($\mu = 0$) and its EOMCC extension for excited electronic states ($\mu > 0$) in a unified manner. As in the ground-state case, the basic EOMCC hierarchy is obtained by varying the truncation in T and R_μ according to m_A, resulting in EOMCCSD ($m_A = 2$), EOMCCSDT ($m_A = 3$), EOMCCSDTQ ($m_A = 4$), and so on. Given the linear nature of the EOMCC ansatz for excited states, the equations used to determine the amplitudes $r_{\mu,a_1\cdots a_n}^{\;\;i_1\cdots i_n}$ and energies $E_\mu^{(A)}$ are obtained in a straightforward fashion by inserting Eq. (2.12) into the Schrödinger equation. After some slight algebraic manipulation, we obtain the matrix eigenvalue equation

$$(\overline{H}_{\rm open}^{(A)} R_{\mu,\rm open}^{(A)})_C |\Phi\rangle = \omega_\mu^{(A)} R_{\mu,\rm open}^{(A)} |\Phi\rangle, \tag{2.15}$$

where $\omega_\mu^{(A)} = E_\mu^{(A)} - E_0^{(A)}$ is the vertical excitation energy characterizing the μth excited state and $\overline{H}^{(A)}_{\rm open} \equiv \overline{H}^{(A)} - E_0^{(A)}\mathbf{1}$ and $R^{(A)}_{\mu,\rm open} \equiv R^{(A)}_\mu - r_{\mu,0}\mathbf{1}$ refer to the diagrams in $\overline{H}^{(A)}$ and $R^{(A)}_\mu$ that contain external fermion lines, respectively. The EOMCC eigenvalue problem given in Eq. (2.15) amounts to diagonalizing of the similarity-transformed Hamiltonian resulting from the ground-state CC computation in the subspace of the Hilbert space corresponding to the content of $R^{(A)}_{\mu,\rm open}$. In practice, the solutions of Eq. (2.15) are obtained using the Hirao–Nakatsuji generalization [237] of the Davidson diagonalization algorithm [238] to non-Hermitian Hamiltonians. The key quantities that must be computed when solving the EOMCC eigenvalue problem using iterative diagonalization routines are the projections of Eq. (2.15) onto the relevant subsets of excited determinants,

$$\sigma_{\mu,a_1\ldots a_n}^{i_1\ldots i_n} = \langle \Phi^{a_1\ldots a_n}_{i_1\ldots i_n} | (\overline{H}^{(A)}_{\rm open} R^{(A)}_{\mu,\rm open})_C | \Phi \rangle,$$
$$i_1 < \ldots < i_n, \quad a_1 < \ldots < a_n, \quad n = 1, \ldots, m_A, \tag{2.16}$$

which are analogs of the σ-vectors used in CI computations. As in the case of the ground-state CC amplitude equations, the working expressions for Eq. (2.16) can be derived using standard algebraic or diagrammatic many-body techniques. After solving Eq. (2.15) to determine the excitation amplitudes $r_{\mu,a_1\cdots a_n}^{i_1\cdots i_n}$ and vertical excitation energies $\omega_\mu^{(A)}$, the scalar part of $R^{(A)}_\mu$ is obtained *a posteriori* as

$$r_{\mu,0} = \langle \Phi | (\overline{H}^{(A)}_{\rm open} R^{(A)}_{\mu,\rm open})_C | \Phi \rangle / \omega_\mu^{(A)}. \tag{2.17}$$

It is important to point out that the similarity-transformed Hamiltonian Eq. (2.10) is not Hermitian [*i.e.*, $(\overline{H}^{(A)})^\dagger \neq \overline{H}^{(A)}$], which means that the bra $\langle \tilde{\Psi}^{(A)}_\mu|$ and its associated ket state $|\Psi^{(A)}_\mu\rangle$ are not trivially related to each other by taking the adjoint. In other words, $\langle \tilde{\Psi}^{(A)}_\mu| \neq (|\Psi^{(A)}_\mu\rangle)^\dagger$. Instead, the correct parameterization of the CC/EOMCC bra state is

$$\langle \tilde{\Psi}^{(A)}_\mu| = \langle \Phi | L^{(A)}_\mu e^{-T^{(A)}}, \tag{2.18}$$

which satisfies the biorthonormality condition

$$\langle \tilde{\Psi}^{(A)}_\mu | \Psi^{(A)}_\nu \rangle = \langle \Phi | L^{(A)}_\mu R^{(A)}_\nu | \Phi \rangle = \delta_{\mu,\nu}. \tag{2.19}$$

The hole–particle deexcitation operator

$$L_\mu^{(A)} = \delta_{\mu,0}\mathbf{1} + \sum_{n=1}^{m_A} L_{\mu,n}, \tag{2.20}$$

with many-body components

$$L_{\mu,n} = \sum_{\substack{i_1<\ldots<i_n \\ a_1<\ldots<a_n}} l_{\mu,i_1\cdots i_n}^{a_1\cdots a_n} (E_{i_1\ldots i_n}^{a_1\ldots a_n})^\dagger, \tag{2.21}$$

is determined by solving the companion left-eigenstate CC/EOMCC equations

$$\langle\Phi|\, L_\mu^{(A)} \overline{H}^{(A)} = E_\mu^{(A)} \langle\Phi| L_\mu^{(A)} \tag{2.22}$$

corresponding to the linear system

$$\begin{aligned} \langle\Phi|\, L_\mu^{(A)} \overline{H}^{(A)} \left|\Phi_{i_1\ldots i_n}^{a_1\ldots a_n}\right\rangle &= E_\mu^{(A)} l_{\mu,i_1\ldots i_n}^{a_1\ldots a_n}, \\ i_1 < \ldots < i_n, \quad a_1 < \ldots < a_n, \quad n &= 1, \ldots, m_A. \end{aligned} \tag{2.23}$$

For the ground state ($\mu = 0$), Eq. (2.23) is solved similarly to the CC amplitude equations using the Jacobi algorithm with DIIS acceleration, since the energy $E_0^{(A)}$ is already known. For excited states ($\mu > 0$), it is still possible to solve Eq. (2.23) using the Jacobi algorithm for a given $E_\mu^{(A)}$ and imposing the normalization given by Eq. (2.19) every iteration, but this approach often leads to poor convergence. Instead, it is easier to solve Eq. (2.22) for $\mu > 0$ in the same way as Eq. (2.15) using the Hirao–Nakatsuji diagonalization algorithm, normalizing the resulting $L_\mu^{(A)}$ operator according to Eq. (2.19) at the end. Alternatively, Eqs. (2.15) and (2.22) can be solved simultaneously to obtain $R_\mu^{(A)}$ and $L_\mu^{(A)}$ together.

As discussed in the Introduction, the higher-level CC/EOMCC methodologies, such as CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ, are capable of providing accurate approximations to the exact, FCI, solution in the majority of chemical problems of interest. We illustrate this point in Figures 2.1 and 2.2 for single bond breaking in hydrogen fluoride and double bond breaking in water, respectively. In both cases, CCSD fails to provide a quantitative description in either the equilibrium or dissociation regions of the PESs. In contrast,

for single bond breaking in hydrogen fluoride (Figure 2.1), the RHF-based CCSDT approach is extremely accurate, while its even higher-level CCSDTQ counterpart is virtually exact. Even in the notoriously challenging case of symmetric O–H dissociation in water (Figure 2.2), the RHF-based CCSDT method remains highly accurate for O–H bond distances of up to and including $2R_e$. As the O–H bonds fully dissociate, the CCSDT calculations using the RHF determinant begin to struggle, but RHF-based CCSDTQ continues to deliver a very accurate description of the entire dissociation process. This highlights the importance of including higher-order excitations to capture stronger nondynamical correlation effects.

Given the limitations of CCSD and the breakdown of CCSDT in the asymptotic region for double bond dissociation, one might consider adopting the symmetry-broken UHF reference to avoid the failures of RHF at larger internuclear distances. However, this is not a robust solution. For single bond breaking in hydrogen fluoride, replacing the RHF determinant with UHF helps reduce the errors obtained using CCSD in the asymptotic region due to mixing the singlet state with solutions of higher multiplicity, but the errors characterizing intermediate H–F distances increase as a consequence, leaving the overall CCSD surface unreliable. When we employ higher-level approximations like CCSDT or CCSDTQ, the choice of reference becomes less relevant. Indeed, this might be anticipated when one is already so close to FCI. For symmetric dissociation in water, while the UHF reference improves the description of the asymptotic region for all CC methods, it again compromises the accuracy in the intermediate region around $2R_e$ due to significant spin contamination, which propagates into the correlated CC state. It is important to recall that symmetric O–H bond dissociation cleaves water into O(^{3}P) and two H(^{2}S) fragments, introducing severe nondynamical correlations associated with the recoupling of four unpaired electrons into a singlet state. From this perspective, it is remarkable that CCSDT remains reliable in this crucial recoupling region and that CCSDTQ can describe the entire bond-breaking potential. This reinforces the idea that higher-level SRCC methods are capable of handling strong nondynamical correlations in chemical problems, making them competitive with more sophisticated MR treatments.

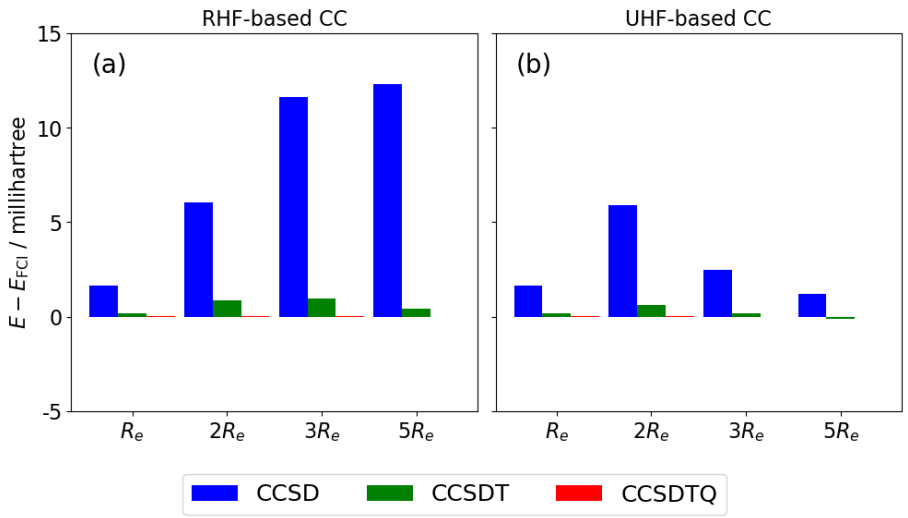

Figure 2.1 Errors, in millihartree, characterizing the RHF-based [(a)] and UHF-based [(b)] all-electron CC calculations relative to FCI (taken from Ref. [140]) for bond breaking in the HF molecule, as described using the DZ basis set [239]. The H–F bond length $R_{\mathrm{H-F}}$ is increased from equilibrium, $R_{\mathrm{H-F}} = R_e$, where $R_e = 1.7832$ bohr, to $2R_e$, $3R_e$, and $5R_e$.

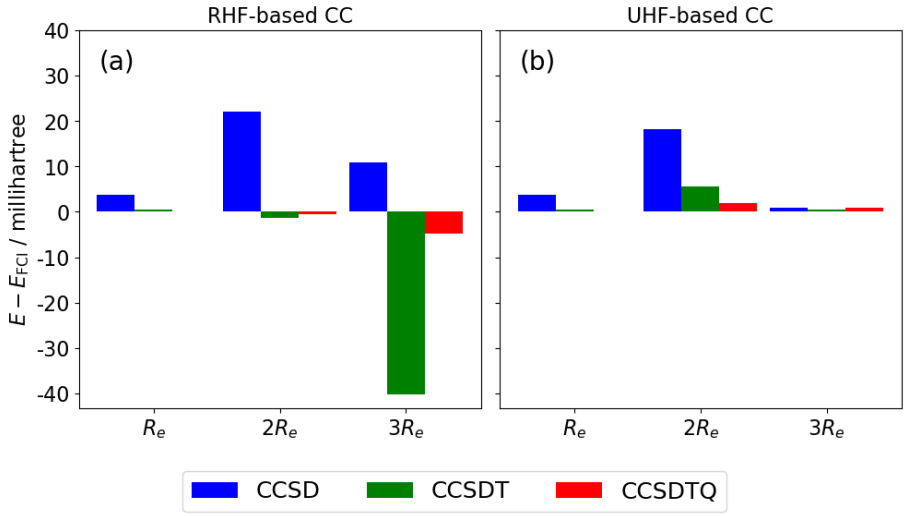

Figure 2.2 Errors, in millihartree, characterizing the RHF-based [(a)] and UHF-based [(b)] all-electron CC calculations relative to FCI (taken from Ref. [240]) for the symmetric dissociation of water, as described using the cc-pVDZ basis set [10]. Following Ref. [240], the H–O–H bond angle is fixed at 110.6 degree and the O–H bonds are simultaneously stretched by factors of 2 and 3 relative to the equilibrium O–H bond length of $R_e = 1.8435$ bohr.

Unfortunately, the computational cost associated with solving the CC/EOMCC equations [Eqs. (2.9) and (2.15)] using high-level truncations in T and R_μ is prohibitively expensive for most practical applications (*e.g.*, $\mathscr{N}^8$ for CCSDT/EOMCCSDT, $\mathscr{N}^{10}$ for CCSDTQ/EOMCCSDTQ, and so on). In order to reduce these costs, standard methodologies like CCSD(T), CCSDT-n, CC3, and their higher-order [113–116] and excited-state [117–121] extensions *directly approximate the parent CC/EOMCC equations using perturbative arguments*. Such strategies may be justified for nondegenerate ground states or when describing excited states dominated by one-electron transitions, but they will generally fail in situations characterized by MR correlations and electronic quasi-degeneracies, which lead to nonperturbative T_n and $R_{\mu,n}$ effects, even when the high-level CC and EOMCC theories remain well-behaved.

This breakdown of perturbative CC/EOMCC approximations is illustrated in Figure 2.3, which compares the ground-state PEC of the F_2 molecule, obtained using CCSD(T), and the excited-state PECs of the lowest-lying $^1\Delta$ state of CH^+, resulting from EOMCCSD(T)(a)* and CC3 calculations, with the corresponding parent CCSDT/EOMCCSDT surfaces. For both of these examples, the PESs obtained using perturbative CC/EOMCC approaches exhibit inaccurate, or even unphysical, behavior, especially at larger internuclear separations, while the full CCSDT/EOMCCSDT calculations remain highly accurate (cf. Ref. [241] and Ref. [101] for numerical results demonstrating the accuracy of CCSDT and EOMCCSDT for the ground state of F_2 and lowest $^1\Delta$ state of CH^+, respectively). Notably, however, the nonperturbative CR-CC(2,3) [Figure 2.3(a)] and CR-EOMCC(2,3) [Figure 2.3(b)] approaches faithfully track their respective parent bond-breaking potentials obtained with CCSDT and EOMCCSDT. In effect, introducing perturbative approximations can undermine the accuracy of high-level CC/EOMCC methodologies, and due to the abundance of such perturbative CC/EOMCC approaches developed in the literature and available in popular codes, this has led some practitioners of quantum chemistry to incorrectly believe that SR CC/EOMCC approaches are inherently unsuitable for describing multiconfigurational states. Given our desire to accurately describe MR correlations effects, we forego perturbative approaches and

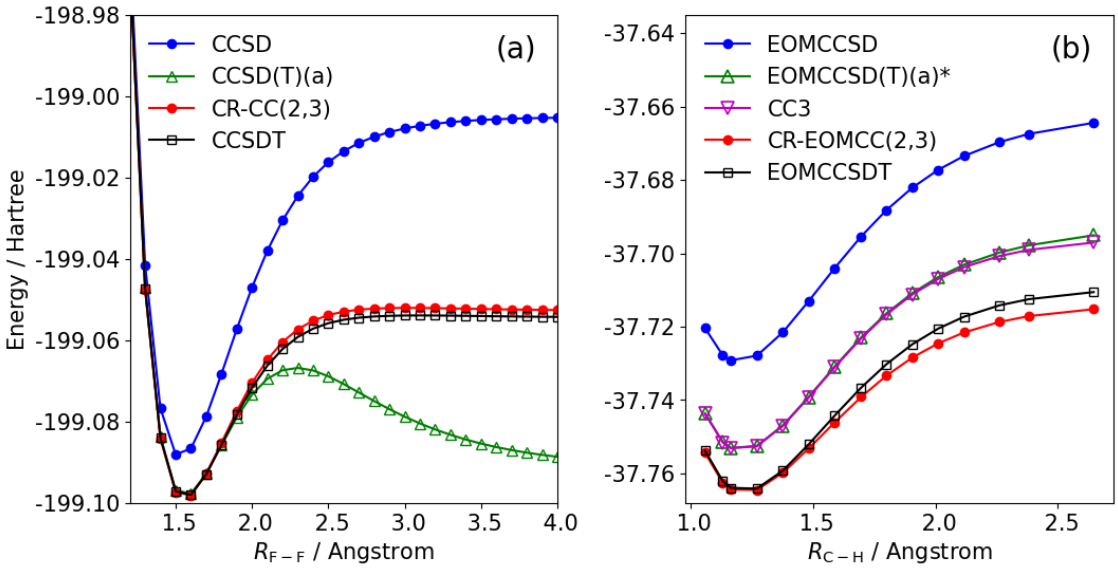

Figure 2.3 A comparison of perturbative and nonperturbative CC/EOMCC approximations for the PESs describing F–F bond breaking in the $X\ ^1\Sigma_g^+$ state of F_2 [(a)] and C–H bond breaking in the lowest-lying $^1\Delta$ state of CH^+ [(b)]. Panel (a) shows the CCSD (blue), CCSD(T) (green), CR-CC(2,3) (red), and parent CCSDT (black) calculations. Panel (b) presents their excited-state counterparts, including EOMCCSD (blue), CR-EOMCC(2,3) (red), and EOMCCSDT (black), alongside the perturbative EOMCCSD(T)(a)* (green) and CC3 (magenta) approaches. The cc-pVDZ basis set [10] was employed for F_2 and the [5s3p1d/3s1p] basis set of Ref. [242] was used for CH^+. The lowest-energy orbitals correlating with 1s shells of F atoms were frozen in all post-RHF steps.

instead turn to the CC(P;Q) formalism as a robust and practical alternative for accurately approximating high-level CC/EOMCC methodologies.

2.2 CC(P;Q) Moment Expansions

The CC(P;Q) formalism takes a different approach compared to standard perturbative methodologies. Instead of approximating the CC/EOMCC equations en masse to reduce the computational costs, *we solve the parent CC/EOMCC equations in a smaller subspace of the Hilbert space and correct the resulting energetics for the missing correlation effects in a nonperturbative fashion*. Thus, our motivation for turning to CC(P;Q) is not only to avoid the failures resulting from perturbation theory, but also to reduce the costs of obtaining high-level CC/EOMCC energetics by only solving a small fraction of the parent CC/EOMCC equations in the iterative steps and using the appropriately designed corrections to handle the remaining correlations. In order to partition the CC/EOMCC equations in a general way

and obtain the compatible noniterative energy corrections, we must adopt the mathematical formulation of the CC(P;Q) theory, which was originally reported in Refs. [126, 127].

Each CC(P;Q) calculation begins by identifying two disjoint subspaces of the N-electron Hilbert space, referred to as the P and Q spaces. The former space, designated as $\mathscr{H}^{(P)}$, is spanned by the excited determinants $|\Phi_K\rangle = E_K|\Phi\rangle$, where E_K is the elementary particle–hole excitation operator generating $|\Phi_K\rangle$ from $|\Phi\rangle$, which, together with $|\Phi\rangle$, dominate the ground state $|\Psi_0\rangle$ or the ground ($\mu = 0$) and excited ($\mu > 0$) states $|\Psi_\mu\rangle$ of the N-electron system of interest. The latter space, designated as $\mathscr{H}^{(Q)}$ [$\mathscr{H}^{(Q)} \subseteq (\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^\perp$, where $\mathscr{H}^{(0)}$ is a one-dimensional subspace spanned by the reference determinant $|\Phi\rangle$], is used to construct the noniterative corrections $\delta_\mu(P;Q)$ to the energies $E_\mu^{(P)}$ obtained in the CC(P)/EOMCC(P) calculations due to the missing higher-order correlation effects.

Once the P and Q spaces are defined, the first step in the CC(P;Q) procedure is to perform the iterative CC(P) and, if excited states are also desired, EOMCC(P) calculations. In a nutshell, the CC(P)/EOMCC(P) equations are almost identical to their counterparts presented in the previous section for CC/EOMCC approximations truncated at many-body rank m_A, except that the T, R_μ, and L_μ operators in CC(P)/EOMCC(P) adopt a more general truncation based on an arbitrary subset of excited determinants spanning the P space. With this in mind, we begin by solving the CC(P) equations,

$$\langle\Phi_K|\overline{H}^{(P)}|\Phi\rangle = 0, \;\; |\Phi_K\rangle \in \mathscr{H}^{(P)}, \tag{2.24}$$

with $\overline{H}^{(P)} = e^{-T^{(P)}} H e^{T^{(P)}}$, to determine the truncated form of the cluster operator T corresponding to the content of the P space,

$$T^{(P)} = \sum_{|\Phi_K\rangle \in \mathscr{H}^{(P)}} t_K E_K, \tag{2.25}$$

where t_Ks are the relevant cluster amplitudes, and the CC(P) ground-state energy

$$E_0^{(P)} = \langle\Phi|\overline{H}^{(P)}|\Phi\rangle. \tag{2.26}$$

We also determine the companion hole–particle deexcitation operator

$$L_0^{(P)} = \mathbf{1} + \sum_{|\Phi_K\rangle \in \mathscr{H}^{(P)}} l_{0,K} (E_K)^\dagger, \tag{2.27}$$

which defines the bra ground state $\langle \tilde{\Psi}_0^{(P)}| = \langle \Phi | L_0^{(P)} e^{-T^{(P)}}$ associated with the CC(P) ket state $|\Psi_0^{(P)}\rangle = e^{T^{(P)}} |\Phi\rangle$ satisfying the normalization condition $\langle \tilde{\Psi}_0^{(P)} | \Psi_0^{(P)} \rangle = 1$, by solving the linear system

$$\langle \Phi | \, L_0^{(P)} \overline{H}^{(P)} \, |\Phi_K\rangle = E_0^{(P)} l_{0,K}, \;\; |\Phi_K\rangle \in \mathscr{H}^{(P)}, \tag{2.28}$$

where $E_0^{(P)}$ is the CC(P) ground-state energy obtained with Eq. (2.26). If there is interest in the ground- as well as excited-state energetics, we follow the CC(P) calculations by the diagonalization of the similarity-transformed Hamiltonian $\overline{H}^{(P)}$ in the P space to obtain the EOMCC(P) energies $E_\mu^{(P)}$ and the corresponding linear excitation and deexcitation operators,

$$R_\mu^{(P)} = r_{\mu,0} \mathbf{1} + \sum_{|\Phi_K\rangle \in \mathscr{H}^{(P)}} r_{\mu,K} E_K \tag{2.29}$$

and

$$L_\mu^{(P)} = \delta_{\mu,0} \mathbf{1} + \sum_{|\Phi_K\rangle \in \mathscr{H}^{(P)}} l_{\mu,K} (E_K)^\dagger, \tag{2.30}$$

respectively, where the amplitudes $r_{\mu,K}$, along with $r_{\mu,0}$, define the EOMCC(P) ket states $|\Psi_\mu^{(P)}\rangle = R_\mu^{(P)} e^{T^{(P)}} |\Phi\rangle$ and their left-eigenstate $l_{\mu,K}$ counterparts define the EOMCC(P) bra states $\langle \tilde{\Psi}_\mu^{(P)}| = \langle \Phi | L_\mu^{(P)} e^{-T^{(P)}}$, which as before, satisfy the biorthonormality condition

$$\langle \tilde{\Psi}_\mu^{(P)} | \Psi_\nu^{(P)} \rangle = \langle \Phi | L_\mu^{(P)} R_\nu^{(P)} | \Phi \rangle = \delta_{\mu,\nu}. \tag{2.31}$$

Once $T^{(P)}$, $L_0^{(P)}$, and $E_0^{(P)}$ and, in the case of excited states, $R_\mu^{(P)}$, $L_\mu^{(P)}$, and $E_\mu^{(P)}$ are obtained, we proceed to the final step, which is the determination of the CC(P;Q) energies

$$E_\mu^{(P+Q)} = E_\mu^{(P)} + \delta_\mu(P;Q), \tag{2.32}$$

where corrections $\delta_\mu(P;Q)$ are given by

$$\delta_\mu(P;Q) = \sum_{|\Phi_K\rangle \in \mathscr{H}^{(Q)}} \ell_{\mu,K}(P)\, \mathfrak{M}_{\mu,K}(P). \tag{2.33}$$

The quantities $\mathfrak{M}_{\mu,K}(P)$ entering Eq. (2.33) represent the generalized moments of the P-space CC/EOMCC equations, which are

$$\mathfrak{M}_{0,K}(P) = \langle \Phi_K | \overline{H}^{(P)} | \Phi \rangle \tag{2.34}$$

in the case of the ground state ($\mu = 0$) [139, 140, 243] and

$$\mathfrak{M}_{\mu,K}(P) = \langle \Phi_K | \overline{H}^{(P)} R_\mu^{(P)} | \Phi \rangle \tag{2.35}$$

for excited states ($\mu > 0$) [142, 143, 244]. Equations (2.34) and (2.35) respectively correspond to projections of the CC(P) and EOMCC(P) equations on the Q-space determinants $|\Phi_K\rangle \in \mathscr{H}^{(Q)}$. The coefficients $\ell_{\mu,K}(P)$ that multiply moments $\mathfrak{M}_{\mu,K}(P)$ in Eq. (2.33) are obtained using the Epstein–Nesbet-like formula

$$\ell_{\mu,K}(P) = \langle \Phi | L_\mu^{(P)} \overline{H}^{(P)} | \Phi_K \rangle / D_{\mu,K}^{(P)}, \tag{2.36}$$

where

$$D_{\mu,K}^{(P)} = E_\mu^{(P)} - \langle \Phi_K | \overline{H}^{(P)} | \Phi_K \rangle. \tag{2.37}$$

One can replace the denominators $D_{\mu,K}^{(P)}$ in Eq. (2.36) by their Møller–Plesset-like analogs, but, as shown in Refs. [128–130, 134, 146, 147, 149, 151], the Epstein–Nesbet-like form of $D_{\mu,K}^{(P)}$ results in a more accurate description. The formulas for the CC(P;Q) moment expansions summarized in Eqs. (2.32)–(2.37) based on the left-eigenstate of the CC(P)/EOMCC(P) problem can be obtained using the general recipes provided in Refs. [126, 146, 147]. A derivation of Eqs. (2.32)–(2.37) along these lines is provided in Appendix B of this thesis.

The CC(P;Q) formalism can be viewed as a generalization of the left-eigenstate CR CC/EOMCC moment expansions and, as such, includes the aforementioned CR-CC(2,3)

approach and its higher-order [129, 131, 153–155] and excited-state [148, 151, 152] extensions, but its key advantage, which the previous moment expansions did not have, is the possibility of making unconventional choices of the P and Q spaces and relaxing the lower-order T_n and $R_{\mu,n}$ components with $n \leq 2$ in the presence of their higher–than–two-body counterparts, such as T_3, T_4, $R_{\mu,3}$, and $R_{\mu,4}$, which the CCSD(T), CR-CC(2,3), and other triples or higher-order corrections to CCSD or EOMCCSD cannot do. In this regard, several creative strategies for selecting the higher–than–doubly excited determinants entering the P and Q spaces in CC(P;Q) considerations have been employed, including (i) using active orbitals to select subsets of higher–than–doubly excited determinants, as in the CC(t;3), CC(t,q;3), CC(t,q;3,4), CC(q;4), etc. hierarchy, [126–129, 131, 132], (ii) relying on CIQMC/CCMC wave function propagations in the many-electron Hilbert space [170–178], adopted in the semi-stochastic CC(P;Q) theories [130, 133, 134, 136], (iii) extracting information from the more deterministic sequence of CIPSI Hamiltonian diagonalizations [181, 183, 184], as in the CIPSI-driven CC(P;Q) framework [135], or (iv) automatically constructing the P and Q spaces based on the information contained within the CC(P;Q) moment expansions, Eq. (2.33), using the adaptive CC(P;Q) approach [137, 138].

All of these different variants of CC(P;Q) offer unique advantages in converging the high-level CC/EOMCC energetics using the same general recipe defined by Eqs. (2.24)–(2.37). The only difference between them is in the method used to partition the manifold of higher–than–doubles excitations between the iterative and noniterative steps of the CC(P;Q) calculation. Each form of CC(P;Q) also offers major savings the computational effort compared to the parent full CC/EOMCC calculations, which were discussed, along with illustrative timings, in Refs. [131, 137] [see Section 2.4 for further illustration of the computational benefits offered by CC(P;Q) methods]. With an exception of the standard CR CC/EOMCC methods and the active-orbital-based CC(P;Q) approaches, the various black-box CC(P;Q) methods, including the semi-stochastic, CIPSI-driven, and adaptive varieties, require that one develops a new strategy for solving the CC(P)/EOMCC(P) equations and computing

the noniterative CC(P;Q) corrections that is capable of handling the potentially spotty subsets of higher–than–doubly excited determinants in the P and Q spaces, which may not form continuous manifolds labeled by occupied and unoccupied orbitals from the respective ranges of indices. A detailed discussion of our novel algorithm for solving the CC(P)/EOMCC(P) equations along with more precise comments about the resulting computational costs is provided in Appendix C of this thesis. In the present context, the CC(P;Q) methodologies that interest us most are the CIPSI-driven and adaptive varieties, which are the subjects of Chapter 3 and 4, respectively. In the next section, we briefly review some of the key details about selected CI methodologies, and in particular, the CIPSI algorithm adopted in the hybrid CC(P;Q) and ec-CC approaches examined in this dissertation.

2.3 Overview of Selected Configuration Interaction

The selected CI approaches, which date back to the late 1960s and early 1970s [179–182], seek to exploit sparsity in the exact, FCI, wave function. As pointed out long ago, a significant fraction of the determinants entering the FCI state have negligible impact on the corresponding energy, and thus, one can construct approximate solutions with near-FCI energies by correctly identifying which configurations to include in the diagonalization and which ones to exclude. This basic observation from the early days of quantum chemistry not only motivated the pioneering selected CI approaches, including the work of Whitten and Hackmeyer [179], Bender and Davidson [180], the MR doubles CI (MRDCI) scheme of Buenker and Peyerimhoff [182], and the CIPSI algorithm introduced by Malrieu and coworkers [181], but it has also inspired a renewed interest in selected CI, with an ever-growing list of methods based on this simple premise being developed in recent years, including the modern reformulation of CIPSI [183, 184], heat-bath CI [191–193, 245], adaptive CI [185, 186], adaptive sampling CI [187, 188], iterative CI [189, 190], Monte Carlo CI [246], coordinate descent CI [247], and even machine-learning-driven CI techniques [248, 249]. All of these selected CI schemes carry out a sequence Hamiltonian diagonalizations in increasingly large, iteratively defined, subspaces of the many-electron Hilbert space, differing primarily in the way that

determinants are selected for inclusion in the variational CI calculation. The majority of selected CI algorithms, including the CIPSI approach of interest to this dissertation, rely on perturbative arguments to identify determinants spanning the diagonalization spaces.

In the CIPSI approach, we seek to obtain an approximation to the FCI wave function by constructing a sequence of subspaces $\mathscr{V}_{\rm int}^{(k)}$, where $k = 0, 1, 2, \ldots$ enumerates the consecutive CIPSI iterations, in which we diagonalize the Hamiltonian to determine the corresponding wave functions $|\Psi_k^{\rm (CIPSI)}\rangle = \sum_{|\Phi_I\rangle \in \mathscr{V}_{\rm int}^{(k)}} c_I^{(k)} |\Phi_I\rangle$ and energies $E_{{\rm var},k}$. The initial subspace $\mathscr{V}_{\rm int}^{(0)}$ can be one-dimensional, if the CIPSI calculations are started from a single determinant, such as the RHF wave function, which is often good enough for ground-state CIPSI calculations, or a relevant singly or doubly excited determinant, if one is interested in targeting specific excited states. Alternatively, a multidimensional $\mathscr{V}_{\rm int}^{(0)}$ space can be used, which may be constructed with the help of a multideterminantal state generated in some preliminary truncated CI computation. Once the initial subspace $\mathscr{V}_{\rm int}^{(0)}$ is defined, the remaining subspaces are constructed via a recursive process in which $\mathscr{V}_{\rm int}^{(k+1)}$ is obtained by augmenting $\mathscr{V}_{\rm int}^{(k)}$ with a subset of the leading singly and doubly excited determinants out of $\mathscr{V}_{\rm int}^{(k)}$ identified with the help of MBPT expansions. Thus, if $\mathscr{V}_{\rm ext}^{(k)}$ is the space of all singly and doubly excited determinants out of the state $|\Psi_k^{\rm (CIPSI)}\rangle$ obtained in the corresponding $\mathscr{V}_{\rm int}^{(k)}$, then for each determinant $|\Phi_\alpha\rangle \in \mathscr{V}_{\rm ext}^{(k)}$, we evaluate the second-order MBPT correction

$$e_{\alpha,k}^{(2)} = |\langle \Phi_\alpha | H | \Psi_k^{\rm (CIPSI)} \rangle|^2 / (E_{{\rm var},k} - \langle \Phi_\alpha | H | \Phi_\alpha \rangle) \tag{2.38}$$

and use the resulting $e_{\alpha,k}^{(2)}$ values to decide which determinants from $\mathscr{V}_{\rm ext}^{(k)}$ should be added to the determinants $|\Phi_I\rangle$ already in $\mathscr{V}_{\rm int}^{(k)}$ to construct the next diagonalization space $\mathscr{V}_{\rm int}^{(k+1)}$. We can also use the $e_{\alpha,k}^{(2)}$ values to calculate the perturbatively corrected CIPSI energies $E_{{\rm var},k} + \Delta E_k^{(2)}$, where

$$\Delta E_k^{(2)} = \sum_{|\Phi_\alpha\rangle \in \mathscr{V}_{\rm ext}^{(k)}} e_{\alpha,k}^{(2)}, \tag{2.39}$$

and, after further manipulations, their $E_{{\rm var},k} + \Delta E_{{\rm r},k}^{(2)}$ counterparts, in which $\Delta E_k^{(2)}$ is replaced by its renormalized $\Delta E_{{\rm r},k}^{(2)}$ form introduced in Ref. [184].

In the modern implementation of CIPSI, formulated in Refs. [183, 184] and available in the Quantum Package software [184], which was used for all CIPSI calculations presented in this dissertation, the process of enlarging $\mathscr{V}_{\rm int}^{(k)}$ to generate $\mathscr{V}_{\rm int}^{(k+1)}$ is executed in the following manner. First, prior to examining the $e_{\alpha,k}^{(2)}$ corrections, one stochastically samples the most important singly and doubly excited determinants out of $|\Psi_k^{\rm (CIPSI)}\rangle$, so that not all singles and doubles from $\mathscr{V}_{\rm int}^{(k)}$ are included in the accompanying $\mathscr{V}_{\rm ext}^{(k)}$ space, only the sampled ones. Next, one arranges the sampled determinants $|\Phi_\alpha\rangle \in \mathscr{V}_{\rm ext}^{(k)}$ in descending order according to their $|e_{\alpha,k}^{(2)}|$ values calculated using Eq. (2.38). The process of enlarging the current subspace $\mathscr{V}_{\rm int}^{(k)}$ to construct the $\mathscr{V}_{\rm int}^{(k+1)}$ space for the subsequent Hamiltonian diagonalization, which starts from the determinants $|\Phi_\alpha\rangle$ characterized by the largest $|e_{\alpha,k}^{(2)}|$ contributions, moving toward those that have smaller $|e_{\alpha,k}^{(2)}|$ values, continues until the number of determinants in $\mathscr{V}_{\rm int}^{(k+1)}$ exceeds the dimension of $\mathscr{V}_{\rm int}^{(k)}$ multiplied by a user-defined factor $f > 1$. In all of the CIPSI calculations considered in this dissertation, we used $f = 2$, which is the default in Quantum Package. In practice, a typical dimension of $\mathscr{V}_{\rm int}^{(k+1)}$ is slightly larger than f times the dimension of $\mathscr{V}_{\rm int}^{(k)}$, since the CIPSI algorithm adds extra determinants to $\mathscr{V}_{\rm int}^{(k+1)}$ to ensure that the resulting $|\Psi_{k+1}^{\rm (CIPSI)}\rangle$ wave function is an eigenstate of the total spin S^2 and S_z operators [250].

The final wave function $|\Psi^{\rm (CIPSI)}\rangle$ of a given CIPSI run and the associated variational ($E_{\rm var}$) and perturbatively corrected [$E_{\rm var} + \Delta E^{(2)}$ and $E_{\rm var} + \Delta E_{\rm r}^{(2)}$] energies are obtained by terminating the above procedure in one of the following two ways: (i) stopping at the first iteration k for which the second-order MBPT correction $|\Delta E_k^{(2)}|$ falls below a user-defined threshold η, indicating that the CIPSI wave function is within a tolerable distance from FCI, or (ii) stopping at the first iteration k for which the number of determinants in the corresponding $\mathscr{V}_{\rm int}^{(k)}$ space exceeds a user-defined input parameter $N_{\rm det(in)}$. Within the context of this dissertation, our primary interest in using CIPSI is for sampling the many-electron Hilbert space to provide useful information for our CIPSI-based CC(P;Q) or ec-CC considerations rather than converging the FCI energetics, so we elect to terminate our

CIPSI calculations using the latter size-based termination option [to ensure that termination according to (i) is not possible, or at least extremely unlikely, we always set $\eta = 10^{-6}$ hartree]. As a result of setting the input parameter f at 2, the sizes of the final wave functions $|\Psi^{(\mathrm{CIPSI})}\rangle$ produced by our CIPSI runs, denoted as $N_{\mathrm{det(out)}}$, were always between $N_{\mathrm{det(in)}}$ and $2N_{\mathrm{det(in)}}$.

To give a feel for how the CIPSI calculations are performed in this dissertation, we provide an example CIPSI run for the ground state of the water molecule in Figure 2.4. In particular, we varied the input parameter $N_{\mathrm{det(in)}}$ in a roughly semi-logarithmic manner, starting at 1 (which returns the RHF state in this example) and ending at a larger value, in this case 100,000. Based on Figure 2.4, we can make a few interesting observations about the behavior of the CIPSI algorithm. First, as expected, both the variational and perturbatively corrected CIPSI energies converge toward the FCI limit as $N_{\mathrm{det(in)}}$ and $N_{\mathrm{det(out)}}$ increase. In fact, we can truly appreciate the sparsity in the FCI wave function when we compare the range of $N_{\mathrm{det(out)}}$ values shown in Figure 2.4, which is minuscule in comparison to the 451,681,246 $S_z = 0$ determinants of the $\mathrm{A}_1(C_{2v})$ symmetry entering the FCI solution. Furthermore, we can see how the perturbative selection scheme adopted in CIPSI influences which types determinants are included in the CI expansions. Based on simple MBPT arguments, we expect that doubly excited determinants form the leading contributions to the FCI expansion when based on an RHF reference, and thus, it is not surprising that CIPSI accumulates almost all doubly excited determinants in the problem when it reaches near-FCI accuracy. After doubly excited determinants, singly, triply, and quadruply excited determinants are picked up next, consistent with their corresponding 4th-order contributions to the energy in perturbation theory. We should keep in mind that our perturbative assessment of the results in Figure 2.4 are only valid when the reference state $|\Phi\rangle$ represents the leading approximation to the FCI solution. The general strength of CIPSI lies in its ability to explore the many-electron Hilbert space in a well-defined and objective manner, adapting the content of the CI wave function to capture the dominant correlations entering the problem. We adopted CIPSI throughout this dissertation, as opposed to alternative selected CI schemes, because

the CIPSI algorithm and its implementation in the Quantum Package code, consistently proves to be among the most accurate ones [194]. Although we employ CIPSI in the context of both CC(P;Q) and ec-CC considerations, the latter serves as our strongest motivation for choosing the most accurate selected CI method. Indeed, in the ec-CC approaches, we rely heavily on the quality of the CI wave function, as we will discuss in the next section.

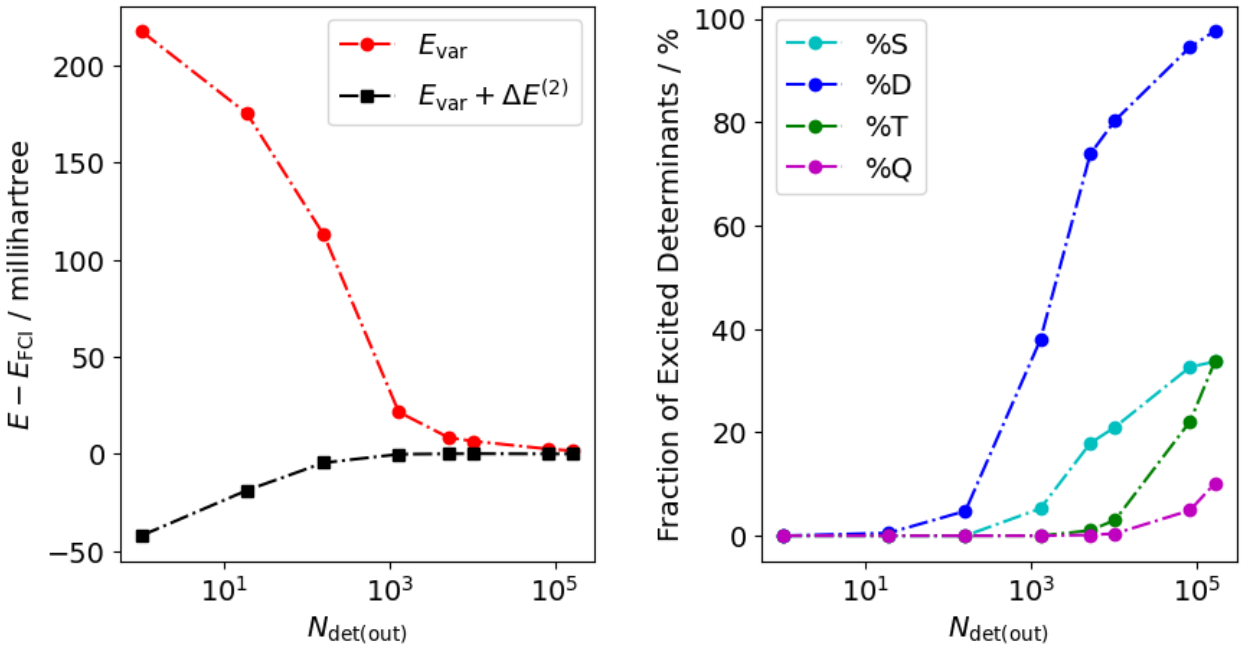

Figure 2.4 An example CIPSI run performed for the ground state of the water molecule, as described using the cc-pVDZ basis set, at the equilibrium C_{2v}-symmetric geometry with H–O–H bond angle of 110.6 degrees and O–H bond length of 1.84345 bohr, also considered in Ref. [240]. The CIPSI calculations were carried out using the protocol described in Section 2.3 in which the input parameter $N_{\text{det(in)}}$ was varied in a roughly semi-logarithmic manner, taking on the values 1, 10, 100, 1,000, 5,000, 10,000, 50,000, and 100,000. The left panel shows the convergence of the variational (E_{var}, in red) and perturbatively corrected ($E_{\text{var}} + \Delta E^{(2)}$, in black) CIPSI energies energies toward the FCI limit as a function of the number of determinants, $N_{\text{det(out)}}$, included in the resulting CIPSI wave function. The right panel shows the corresponding evolution in the fractions of singly (%S, in cyan), doubly (%D, in blue), triply (%T, in green), and quadruply (%Q, in magenta) excited determinants captured by the CIPSI algorithm.

2.4 Externally Corrected Coupled-Cluster Methodology

As discussed in the Introduction, one of the intriguing ways of improving the results of SRCC calculations in MR and strongly correlated situations, which is based on combining

the CC and non-CC (e.g., CI) concepts, is the ec-CC framework [195–214] (see Ref. [210] for a review). The ec-CC approaches are based on the observation that for electronic Hamiltonians containing up to two-body interactions, the CC amplitude equations projected on the singly and doubly excited determinants in the absence of any approximations in T,

$$\langle \Phi_i^a | [H_N \left(1 + T_1 + T_2 + \tfrac{1}{2}T_1^2 + T_3 + T_1T_2 + \tfrac{1}{6}T_1^3\right)]_C |\Phi\rangle = 0 \tag{2.40}$$

and

$$\langle \Phi_{ij}^{ab} | [H_N \left(1 + T_1 + T_2 + \tfrac{1}{2}T_1^2 + T_3 + T_1T_2 + \tfrac{1}{6}T_1^3 + T_4 + \tfrac{1}{2}T_2^2 + \tfrac{1}{2}T_1^2T_2 + T_1T_3 + \tfrac{1}{24}T_1^4\right)]_C |\Phi\rangle = 0, \tag{2.41}$$

respectively, where $H_N = H - \langle \Phi | H | \Phi \rangle$ denotes H in the normal-ordered form with respect to $|\Phi\rangle$, do not engage the higher-rank T_n components of the cluster operator T with $n > 4$. Thus, by solving Eqs. (2.40) and (2.41) for T_1 and T_2 in the presence of the exact T_3 and T_4 amplitudes extracted from FCI using the cluster-analysis relations,

$$\begin{aligned}
T_1^{(\text{ext})} &= C_1 \\
T_2^{(\text{ext})} &= C_2 - \frac{1}{2}C_1^2 \\
T_3^{(\text{ext})} &= C_3 - C_1C_2 + \frac{1}{3}C_1^3 \\
T_4^{(\text{ext})} &= C_4 - C_1C_3 - \frac{1}{2}C_2^2 + C_1^2C_2 - \frac{1}{4}C_1^4,
\end{aligned} \tag{2.42}$$

where C_n represents the n-body component of the CI wave function expansion and $T_n^{(\text{ext})}$ represents the T_n component of T extracted from the CI wave function, produces the exact T_1 and T_2, and as a consequence, the exact, full CI energy. An explicit form of Eq. (2.42) relating the cluster amplitudes t_a^i, t_{ab}^{ij}, t_{abc}^{ijk}, and t_{abcd}^{ijkl} to their c_a^i, c_{ab}^{ij}, c_{abc}^{ijk}, and c_{abcd}^{ijkl} counterparts, which respectively characterize the C_1, C_2, C_3, and C_4 components extracted from CI, is given in Table 2.1. To date, the ec-CC framework has been combined with non-CC wave functions corresponding to PUHF [196, 197, 202, 210], valence-bond [198–200], CASSCF and CASCI [201, 204, 205, 207], MRCI [203, 206, 207, 209], perturbatively selected CI [208], FCIQMC [195, 211], adaptive CI [212], DMRG [213], heat-bath CI [213], and CIPSI [214] calculations,

all of which are designed to provide T_3 and T_4 clusters that accurately capture the strong nondynamical correlations that are difficult to describe using low-level CC calculations alone. In general, we expect that the energies resulting from ec-CC calculations performed using well-behaved approximate T_3 and T_4 will improve upon CCSD, in which T_3 and T_4 are zero, and the non-CC source from which T_3 and T_4 are obtained. While the successful applications of the ec-CC framework do validate this premise, there are also cases where the subsequent ec-CC computations offer no improvements to the underlying non-CC computations [212].

Table 2.1 The programmable expressions for the amplitudes in $T_1^{(\mathrm{ext})}$, $T_2^{(\mathrm{ext})}$, $T_3^{(\mathrm{ext})}$, and $T_4^{(\mathrm{ext})}$ resulting from cluster analysis [Eq. (2.42)] of the underlying ground-state CI wave function $|\Psi^{(\mathrm{CI})}\rangle$ characterized by excitation components C_n.

$T_n^{(\mathrm{ext})}$ Amplitude	Expression[a]
t_a^i	c_a^i
t_{ab}^{ij}	$c_{ab}^{ij} - \mathscr{A}^{ij} c_a^i c_b^j$
t_{abc}^{ijk}	$c_{abc}^{ijk} - \mathscr{A}^{i/jk}\mathscr{A}_{a/bc} c_a^i c_{bc}^{jk} + 2\mathscr{A}^{ijk} c_a^i c_b^j c_c^k$
t_{abcd}^{ijkl}	$c_{abcd}^{ijkl} - \mathscr{A}^{i/jkl}\mathscr{A}_{a/bcd} c_a^i c_{bcd}^{jkl} - \mathscr{A}^{ij/kl} c_{ab}^{ij} c_{cd}^{kl} + 2\mathscr{A}^{ij}\mathscr{A}^{ij/kl}\mathscr{A}_{ab/cd} c_a^i c_b^j c_{cd}^{kl} - 6\mathscr{A}^{ijkl} c_a^i c_b^j c_c^k c_d^l$

[a]The contravariant antisymmetrizers in the expressions are defined as $\mathscr{A}^{pq} = 1 - (pq)$, $\mathscr{A}^{pqr} = \mathscr{A}^{p/qr}\mathscr{A}^{qr}$, and $\mathscr{A}^{pqrs} = \mathscr{A}^{pq/rs}\mathscr{A}^{pq}\mathscr{A}^{rs}$, with $\mathscr{A}^{p/qr} = 1 - (pq) - (pr)$, $\mathscr{A}^{p/qrs} = 1 - (pq) - (pr) - (ps)$, and $\mathscr{A}^{pq/rs} = 1 - (pr) - (ps) - (qr) - (qs) + (pr)(qs)$. The covariant antisymmetrizers $\mathscr{A}_{pq}$, $\mathscr{A}_{pqr}$, $\mathscr{A}_{pqrs}$, $\mathscr{A}_{p/qr}$, $\mathscr{A}_{p/qrs}$, and $\mathscr{A}_{pq/rs}$ are identical to their contravariant counterparts.

In order to gain a deeper understanding of this paradoxical phenomena, we have carried out a detailed theoretical investigation of the ec-CC equations in Ref. [214] and identified two families of ec-CC approaches within the ec-CC framework, called ec-CC-I and ec-CC-II, which differ in the way the cluster-analysis relations, Eq. (2.42), are carried out. In the ec-CC-I approach, one obtains $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ by application of Eq. (2.42) directly, allowing for the potential introduction of completely disconnected components into $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ in the form of $T_3^{(\mathrm{ext})} = -C_1C_2 + \frac{1}{3}C_1^3$ and $T_4^{(\mathrm{ext})} = -C_1C_3 - \frac{1}{2}C_2^2 + C_1^2C_2 - \frac{1}{4}C_1^4$, respectively. Solving for T_1 and T_2 in the presence of these disconnected terms not only results in strongly non-extensive energies, but may also produce poorer energetics that do not even improve upon those obtained using the underlying CI wave function (see the results reported in Refs. [212] and [214] for a demonstration). One can largely remove disconnected terms in the

ec-CC equations by following the ec-CC-II protocol, in which T_1 and T_2 clusters are solved in the presence of $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ extracted from CI using Eq. (2.42) after eliminating $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ components that do not have the companion C_3 and C_4 amplitudes. As shown in Ref. [214], the ec-CC-II method generally results in energetics that are more accurate than those obtained using ec-CC-I.

Depending on whether we adopt the ec-CC-I or ec-CC-II framework, we can identify the different classes of CC and non-CC wave functions that are solutions to Eqs. (2.40) and (2.41), when T_3 and T_4 are obtained from Eq. (2.42), in the form of two theorems. These two theorems, which are proven both algebraically and with a diagrammatic analysis in Appendices A and B of Ref. [214], help one predict whether a given ec-CC approach will improve upon the underlying non-CC energetics. They are given as follows:

1. The ec-CC-I calculation in which T_3 and T_4 are obtained by cluster analysis of the CI wave function that describes singles and doubles fully, as in the CISD, CISDT, CISDTQ, etc., or any other CI method that provides a complete treatment of the single and double excitation manifolds, returns back the underlying CI energy.

2. The ec-CC-II calculation in which T_3 and T_4 are obtained by cluster analysis of the CI wave function that captures all singles, doubles, triples, and quadruples, as in CISDTQ, CISDTQP, CISDTQPH, etc., or any other CI method that provides a complete treatment of the single, double, triple, and quadruple excitation manifolds, returns back the underlying CI energy.

When designing ec-CC methods, the above two theorems are of crucial importance. First, it makes clear that the main shortcoming of the ec-CC-I approach is that it cannot benefit from modern selected-CI-based approaches, which are easily capable of providing a numerically complete treatment of all singly and doubly excited determinants in the problem. Indeed, in CIPSI calculations, the singly (for non-HF references) and doubly excited determinants are, by design, the first excitation manifolds that are explored before attempting to

examine higher-order correlation effects (see Figure 2.2). In contrast, the ec-CC-II method appears to be much more promising for all selected-CI-based ec-CC considerations because it is rather uncommon to use CI wave functions that provide a complete treatment of all excitation manifolds through quadruples for subsequent ec-CC computations. Based on the above theorems, we have concluded that the truncated CI wave functions that are best suited for the ec-CC calculations are those that efficiently sample the many-electron Hilbert space, without saturating the lower-rank excitation manifolds, especially the excitations through quadruples, too rapidly, while adjusting the singly through quadruply excited CI amplitudes to the dominant higher–than–quadruply excited contributions. With these theorems in mind, we have identified the FCIQMC [170–174] and modern CIPSI [183, 184, 194] schemes, which are characterized by tempered growth of the wave function, as optimal non-CC sources of T_3 and T_4 clusters for subsequent ec-CC-II considerations. In this dissertation, we focus on the CIPSI-based ec-CC-II approach, which is developed and tested in Chapter 5.

Inspired by the RMR-CCSD(T) [209] and ACI-CCSD(T) [212] approaches, and drawing upon our previous experience in devising CC(P;Q) moment expansions, we have also introduced a new type of triples correction to the ec-CC-II energetics, abbreviated as ec-CC-II$_3$, which captures the missing, generally higher-order T_3 correlation effects not captured by the CI wave function $|\Psi^{(\mathrm{CI})}\rangle$ using the CC(P;Q)-like moment expansions. In formulating the ec-CC-II$_3$ triples correction to the ec-CC-II energies, we have adapted the basic equations defining the noniterative CC(P;Q) corrections, namely Eqs. (2.32)–(2.37). To be more specific, in the ec-CC-II$_3$ approach, we determine the total energies

$$E_{\mathrm{II}}^{(3)} = E_{\mathrm{II}} + \delta_3, \tag{2.43}$$

where E_{II} is the energy resulting from the underlying ec-CC-II calculation, and the correction δ_3 due to the three-body clusters missing from $|\Psi^{(\mathrm{CI})}\rangle$

$$\delta_3 = \sum_{|\Phi_{ijk}^{abc}\rangle \notin |\Psi^{(\mathrm{CI})}\rangle} \ell_{ijk}^{abc}(2)\, \mathfrak{M}_{abc}^{ijk}(2), \tag{2.44}$$

in which the generalized moments of the ec-CC-II equations are given by

$$\mathfrak{M}_{abc}^{ijk}(2) = \langle \Phi_{ijk}^{abc} | \overline{H}(2) | \Phi \rangle, \tag{2.45}$$

with

$$\overline{H}(2) = e^{-T_1 - T_2} H e^{T_1 + T_2} = (H e^{T_1 + T_2})_C, \tag{2.46}$$

and the coefficients $\ell_{ijk}^{abc}(2)$ multiplying moments $\mathfrak{M}_{abc}^{ijk}(2)$ in Eq. (2.44) are

$$\ell_{ijk}^{abc}(2) = \langle \Phi | (\Lambda_1 + \Lambda_2) \overline{H}(2) | \Phi_{ijk}^{abc} \rangle / (E_{\rm II} - \langle \Phi_{ijk}^{abc} | \overline{H} | \Phi_{ijk}^{abc} \rangle). \tag{2.47}$$

The amplitudes λ_i^a and λ_{ij}^{ab} characterizing the one- and two-body deexcitation operators Λ_1 and Λ_2, respectively, are obtained by solving the corresponding left eigenstate system

$$\begin{aligned} \langle \Phi | (1 + \Lambda_1 + \Lambda_2) \overline{H}(2) | \Phi_i^a \rangle &= E_{\rm II} \lambda_i^a \\ \langle \Phi | (1 + \Lambda_1 + \Lambda_2) \overline{H}(2) | \Phi_{ij}^{ab} \rangle &= E_{\rm II} \lambda_{ij}^{ab}. \end{aligned} \tag{2.48}$$

One particularly appealing aspect of the ec-CC-II and ec-CC-II$_3$ approaches is that they only require solving for the T_1 and T_2 clusters in the iterative steps of the ec-CC calculations, with their $T_3^{(\rm ext)}$ and $T_4^{(\rm ext)}$ counterparts frozen at their values resulting from cluster analysis of the underlying CI wave function. This makes the ec-CC equations very similar to CCSD and easily applicable to larger calculations. In fact, the ec-CC system given in Eqs. (2.40) and (2.41) only differs from its CCSD counterpart by the presence of $\langle \Phi_i^a | (H_N T_3^{(\rm ext)})_C | \Phi \rangle$ in the projection onto singly excited determinants and $\langle \Phi_{ij}^{ab} | (H_N T_3^{(\rm ext)})_C | \Phi \rangle$, $\langle \Phi_{ij}^{ab} | (H_N T_1 T_3^{(\rm ext)})_C | \Phi \rangle$, and $\langle \Phi_{ij}^{ab} | (H_N T_4^{(\rm ext)})_C | \Phi \rangle$ in the projection on doubly excited determinants. With the exception of $\langle \Phi_{ij}^{ab} | (H_N T_1 T_3^{(\rm ext)})_C | \Phi \rangle$, the other three additional terms depend only on the values of $T_3^{(\rm ext)}$ and $T_4^{(\rm ext)}$, which means that they can be precomputed prior to performing the ec-CC iterations. We will now briefly discuss the computational implementation of these terms in the ec-CC-II system.

In order to evaluate $\langle \Phi_i^a | (H_N T_3^{(\rm ext)})_C | \Phi \rangle$, $\langle \Phi_{ij}^{ab} | (H_N T_3^{(\rm ext)})_C | \Phi \rangle$, and $\langle \Phi_{ij}^{ab} | (H_N T_4^{(\rm ext)})_C | \Phi \rangle$, we first note that the disconnected contribution corresponding to each of these terms is

zero. Thus, we may formally remove the connected operator product constraint and insert a resolution of identity within the many-electron Hilbert space between H_N and $T_3^{(\mathrm{ext})}$ or $T_4^{(\mathrm{ext})}$ to obtain

$$\langle \Phi_i^a | H_N T_3^{(\mathrm{ext})} | \Phi \rangle = \sum_{|\Phi_{klm}^{cde}\rangle \in |\Psi^{(\mathrm{CI})}\rangle} \langle \Phi_i^a | H_N | \Phi_{klm}^{cde} \rangle \langle \Phi_{klm}^{cde} | T_3^{(\mathrm{ext})} | \Phi \rangle, \tag{2.49}$$

$$\langle \Phi_{ij}^{ab} | H_N T_3^{(\mathrm{ext})} | \Phi \rangle = \sum_{|\Phi_{klm}^{cde}\rangle \in |\Psi^{(\mathrm{CI})}\rangle} \langle \Phi_{ij}^{ab} | H_N | \Phi_{klm}^{cde} \rangle \langle \Phi_{klm}^{cde} | T_3^{(\mathrm{ext})} | \Phi \rangle, \tag{2.50}$$

and

$$\langle \Phi_{ij}^{ab} | H_N T_4^{(\mathrm{ext})} | \Phi \rangle = \sum_{|\Phi_{klmn}^{cdef}\rangle \in |\Psi^{(\mathrm{CI})}\rangle} \langle \Phi_{ij}^{ab} | H_N | \Phi_{klmn}^{cdef} \rangle \langle \Phi_{klmn}^{cdef} | T_4^{(\mathrm{ext})} | \Phi \rangle, \tag{2.51}$$

where in accordance with the ec-CC-II approach, we have enforced that $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ are zero for all triply and quadruply excited determinants for which there does not exist an associated C_3 or C_4 amplitude, respectively. Since the $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ operators resulting from cluster analysis of $|\Psi^{(\mathrm{CI})}\rangle$ are only as large as their corresponding C_3 or C_4 counterparts, we can assume that these objects can fit into memory as a single vector of amplitudes alongside an associated list of triply or quadruply excited determinants. If the latter assumption is not true, this would imply that we must be saturating the triples and quadruples manifolds in our CI calculation, which would render the ec-CC-II calculations useless anyway. Therefore, our computation of Eqs. (2.49) and (2.50) can be easily executed by looping over all triply excited determinants ($|\Phi_{klm}^{cde}\rangle$) in $|\Psi^{(\mathrm{CI})}\rangle$, and for each one, looping over all singly ($|\Phi_i^a\rangle$) and doubly ($|\Phi_{ij}^{ab}\rangle$) excited determinants and evaluating the matrix elements $\langle \Phi_i^a | H_N | \Phi_{jkl}^{bcd} \rangle$ and $\langle \Phi_{ij}^{ab} | H_N | \Phi_{klm}^{cde} \rangle$, which are then multiplied by the cluster amplitude t_{cde}^{klm} entering $T_3^{(\mathrm{ext})}$ and accumulated into the corresponding projection onto singly or doubly excited determinants. Similarly, Eq. (2.51) is evaluated by looping over all quadruply excited determinants ($|\Phi_{klmn}^{cdef}\rangle$) captured by CI, and for each one, looping over all doubly excited determinants ($|\Phi_{ij}^{ab}\rangle$) and multiplying the matrix elements $\langle \Phi_{ij}^{ab} | H_N | \Phi_{klmn}^{cdef} \rangle$ with the corresponding t_{cdef}^{klmn} amplitude in $T_4^{(\mathrm{ext})}$ and adding the result to the projection onto doubly excited determinants. The relevant matrix elements of H_N entering Eqs. (2.49)–(2.51) are straightforward

to evaluate using Wick's Theorem or many-body diagrams. Unfortunately, the additional $\langle\Phi_{ij}^{ab}|(H_N T_1 T_3^{(\mathrm{ext})})_C|\Phi\rangle$ term in Eq. (2.41) cannot be handled using similar one-time precomputation. Due to the presence of T_1 clusters, this terms must be recomputed each iteration, making the computational costs of our ec-CC-II iterations asymptotically $(d/D)n_o^3 n_u^4$, where d is the number of triply excited determinants captured by $|\Psi^{(\mathrm{CI})}\rangle$ out of the total number of determinants D spanning the triples manifold. Fortunately, for smaller values of (d/D), the iterative ec-CC-II steps have costs that are approximately the same as that of CCSD. For higher fractions of triply excited determinants in the underlying CI state, the costs may scale as large as $\mathscr{N}^7$ with the system size $\mathscr{N}$. It is also worth pointing out that earlier ec-CC studies [212] demonstrated that one can approximate the quantity $\langle\Phi_{ij}^{ab}|(H_N T_1 T_3^{(\mathrm{ext})})_C|\Phi\rangle$ entering the ec-CC-II equations by its $\langle\Phi_{ij}^{ab}|(H_N T_1^{(\mathrm{ext})} T_3^{(\mathrm{ext})})_C|\Phi\rangle$ counterpart obtained from the cluster analysis of $|\Psi^{(\mathrm{CI})}\rangle$ without significant loss of accuracy. Although we have not pursued such approximations, they would undoubtedly be helpful in improving the computational efficiency if our existing ec-CC-II codes. Finally, the triples correction δ_3, Eq. (2.44), can be determined using simple CR-CC(2,3)-like [or CCSD(T)-like] expressions involving T_1, T_2, Λ_1, and Λ_2, which involve noniterative computational steps that scale as $n_o^3 n_u^4$. Therefore, under most circumstances, the ec-CC steps of the ec-CC-II and ec-CC-II$_3$ calculations should be comparable to those characterizing the CCSD and CR-CC(2,3) methods, respectively.

CHAPTER 3

MERGING THE CC(P;Q) FORMALISM WITH SELECTED CONFIGURATION INTERACTION

In the previous chapter, we have introduced the framework of CC(P;Q) moment expansions and how its flexibility allows us to devise computationally tractable schemes for accurately capturing higher-order correlation effects in the CC/EOMCC calculations associated with the T_n and $R_{\mu,n}$ components with $n > 2$, even in cases where these higher–than–two-body components of T and R_μ become large, nonperturbative, and strongly coupled to their lower-rank T_1, T_2, $R_{\mu,1}$, and $R_{\mu,2}$ counterparts and conventional noniterative corrections of the CR-CC or CR-EOMCC types fail. The original active-orbital-based CC(P;Q) approaches, including the CC(t;3), CC(t,q;3,4), CC(q;4), etc. hierarchy [126–129, 131, 132, 138], are capable of recovering high-level CC/EOMCC energies of the CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and so on, types by including a subset of leading higher–than–doubly excited determinants in the underlying P spaces selected with the help of active orbitals and correcting the resulting CCSDt, CCSDtq, CCSDTq, etc. calculations for missing correlations using the suitably defined CC(P;Q) moment expansions. However, as mentioned in the Introduction, such schemes rely on the user- and system-dependent active orbitals, rendering these methods no longer computational black boxes. Subsequent work explored the combination of the deterministic CC(P;Q) formalism with stochastic wave function sampling procedures in which the higher–than–doubly excited determinants included in the P spaces defining the CC(P;Q) calculations are automatically identified using the information extracted CIQMC or CCMC calculations. The resulting semi-stochastic CC(P;Q) methodology [130, 133, 134] allows us to recover high-level CC/EOMCC energetics in a black-fox fashion at small fractions of the computational costs with the help of relatively early-stage CIQMC/CCMC calculations.

Inspired by the success of the semi-stochastic CC(P;Q) approach, in this chapter, we present CIPSI-driven CC(P;Q) approach, which was originally introduced in Ref. [135], and

represents a more deterministic version of the semi-stochastic CC(P;Q) scheme. Instead of relying on stochastic wave function propagations of the CIQMC or CCMC types, in the CIPSI-driven CC(P;Q) approach, the subsets of leading higher–than–doubly excited determinants entering the P spaces are identified using the information extracted from the sequences of Hamiltonian diagonalization defining the selected CI algorithm abbreviated as CIPSI [181, 183, 184]. The primary advantages of using CIPSI instead of CIQMC/CCMC as a provider of the lists of higher–than–doubly excited determinants are (i) the stochastic noise associated with random sampling of the many-electron Hilbert space is largely or entirely eliminated, (ii) the perturbative selection scheme driving the CIPSI algorithm is typically more reliable than CIQMC/CCMC in identifying the leading higher–than–doubly excited determinants in the problem, resulting in smaller-sized P spaces used in the CC(P;Q) computations, and (iii) the CIPSI Hamiltonian diagonalizations provide easy and natural access to excited-state wave functions, which are significantly more challenging to describe using CIQMC/CCMC calculations. After summarizing the key algorithmic details used in our CIPSI-driven CC(P;Q) calculations aimed at converging high-level CC/EOMCC energetics for ground and excited electronic states, we demonstrate the usefulness of the CIPSI-driven CC(P;Q) approach on a few molecular examples, including the dissociation of F_2 and the automerization of cyclobutadiene as well as the vertical excitation spectrum of the CH^+ molecule, where we recover the electronic energies corresponding to the full CCSDT and EOMCCSDT calculations based on the information extracted from compact CI wave functions originating from relatively inexpensive Hamiltonian diagonalizations.

3.1 Theoretical and Algorithmic Details

Having already discussed the key ingredients of the CC(P;Q) and CIPSI formalisms relevant to this work in Chapter 2, we can describe the CIPSI-driven CC(P;Q) algorithm aimed at converging the high-level CC and EOMCC energetics, in which the higher–than–doubly excited determinants entering the P spaces used in the CC(P;Q) calculations are generated in an automated fashion with the help of CIPSI Hamiltonian diagonalizations. In

setting up and executing the CIPSI-driven CC(P;Q) calculations, we employ the following procedure:

1. Given a reference state $|\Phi\rangle$, which in all of the CIPSI-driven CC(P;Q) calculations reported in Figure 3.1 and Tables 3.1–3.5 was the RHF determinant (for open-shell states, we would use the high-spin ROHF reference), choose an input parameter $N_{\text{det(in)}}$, used to terminate the CIPSI wave function growth, and execute a CIPSI run to generate the ground state or, if desired, the ground- and excited-state wave functions $|\Psi^{(\text{CIPSI})}\rangle$ spanned by $N_{\text{det(out)}}$ determinants. If the goal of the CIPSI-driven CC(P;Q) calculation is to determine the ground-state energy or the energies of several electronic states belonging to the same irreducible representation (irrep) as the ground state, one can initiate the corresponding CIPSI run from the reference determinant $|\Phi\rangle$. If, in addition to the ground state, one also wishes to determine excited states belonging to different irreps as the ground state, one can generate the initial CIPSI subspace $\mathscr{V}_{\text{int}}^{(0)}$ with the help of some preliminary (e.g., CIS) calculation.

2. After the CIPSI run defined by a given $N_{\text{det(in)}}$ value is completed, extract a list or, if multiple electronic states belonging to multiple irreps are targeted, lists of higher–than–doubly excited determinants relevant to the target CC/EOMCC theory level from the ground-state or ground- and excited-state wave functions $|\Psi^{(\text{CIPSI})}\rangle$ to define the P space(s) needed to set up the ground-state CC(P) or ground- and excited-state CC(P)/EOMCC(P) calculations. If the goal is to converge the CCSDT/EOMCCSDT-level energetics, the P space for the CC(P) calculations and the EOMCC(P) calculations for excited states belonging to the same irrep as the ground state can be defined as all singly and doubly excited determinants plus the triply excited determinants contained in the corresponding CIPSI ground state $|\Psi^{(\text{CIPSI})}\rangle$. For excited states belonging to a different irrep than the ground state, the P space is defined to contain all singly and doubly excited determinants in addition to the triply excited determinants con-

tained in the lowest-energy CIPSI wave function of the target irrep. In particular, in our current implementation, a distinct P space is constructed for each irrep associated with the point group of the molecule.

3. Solve the CC(P) and, if excited states are targeted, EOMCC(P) equations in the P spaces or P spaces obtained in the previous step. If we are targeting the CCSDT- and EOMCCSDT-level energetics we define $T^{(P)} = T_1 + T_2 + T_3^{(\mathrm{CIPSI})}$, $R_\mu^{(P)} = r_{\mu,0}\mathbf{1} + R_{\mu,1} + R_{\mu,2} + R_{\mu,3}^{(\mathrm{CIPSI})}$, and $L_\mu^{(P)} = \delta_{\mu,0}\mathbf{1} + L_{\mu,1} + L_{\mu,2} + L_{\mu,3}^{(\mathrm{CIPSI})}$, where the list, or lists, of triples in $T_3^{(\mathrm{CIPSI})}$, $R_{\mu,3}^{(\mathrm{CIPSI})}$, and $L_{\mu,3}^{(\mathrm{CIPSI})}$ is/are extracted from the ground-state or ground- and excited-state wave function(s) $|\Psi^{(\mathrm{CIPSI})}\rangle$, as summarized in point 2. For the excited states belonging to irreps other than that of the ground state, we construct the similarity-transformed Hamiltonian $\overline{H}^{(P)}$, to be diagonalized in the EOMCC(P) calculations, in the same way as in the ground-state computations, but then use the appropriate irrep-specific list of triply excited determinants to define $R_{\mu,3}^{(\mathrm{CIPSI})}$ and $L_{\mu,3}^{(\mathrm{CIPSI})}$. We follow a similar procedure when targeting the CCSDTQ/EOMCCSDTQ-level energetics, in which case $T^{(P)} = T_1 + T_2 + T_3^{(\mathrm{CIPSI})} + T_4^{(\mathrm{CIPSI})}$, $R_\mu^{(P)} = r_{\mu,0}\mathbf{1} + R_{\mu,1} + R_{\mu,2} + R_{\mu,3}^{(\mathrm{CIPSI})} + R_{\mu,4}^{(\mathrm{CIPSI})}$, and $L_\mu^{(P)} = \delta_{\mu,0}\mathbf{1} + L_{\mu,1} + L_{\mu,2} + L_{\mu,3}^{(\mathrm{CIPSI})} + L_{\mu,4}^{(\mathrm{CIPSI})}$.

4. Use the information obtained in Step 3 to determine corrections $\delta_\mu(P;Q)$, Eq. (2.33), which describe the remaining correlations of interest that were not captured by the CIPSI-based CC(P) and EOMCC(P) calculations. If the goal is to converge the CCSDT/EOMCCSDT energetics, define the Q space(s) needed to calculate corrections $\delta_\mu(P;Q)$ as the remaining triply excited determinants that are not contained in corresponding P space(s) considered in points 2 and 3. If the target approach is CCSDTQ/EOMCCSDTQ, define the relevant Q space(s) as the triply and quadruply excited determinants absent in corresponding P space(s). Add the resulting corrections $\delta_\mu(P;Q)$ to $E_\mu^{(P)}$ to obtain the CC(P;Q) energies $E_\mu^{(P+Q)}$, Eq. (2.32).

5. To check convergence, repeat Steps 1–4 for a larger value of $N_{\mathrm{det(in)}}$. The CIPSI-

driven CC(P;Q) calculations can be regarded as converged if the difference between consecutive $E_\mu^{(P+Q)}$ energies falls below a user-defined threshold. In analogy to the semi-stochastic CIQMC/CCMC-based semi-stochastic CC(P;Q) framework of Refs. [130, 133, 134], one can also stop if the fraction(s) of higher–than–doubly excited determinants contained in the final CIPSI state(s) $|\Psi^{(\mathrm{CIPSI})}\rangle$ is (are) sufficiently large to produce the desired accuracy level.

The above algorithm can be carried out to perform the CIPSI-driven CC(P;Q) computations aimed at converging the results corresponding to any high-level CC/EOMCC theory, such as CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and so on. In our initial studies to date exploring the CIPSI-driven CC(P;Q) methodology, we have focused our efforts on the CIPSI-based CC(P;Q) approach aimed at converging the CCSDT energetics for the ground state and EOMCCSDT energetics for excited states. The resulting codes, which have been incorporated in the open-source CCpy software package available on GitHub [228] and interfaced with the RHF/ROHF and integral transformation routines in GAMESS [227, 251], are capable of parsing the lists of triply excited determinants extracted from the ground- or ground- and excited-state CIPSI wave functions $|\Psi^{(\mathrm{CIPSI})}\rangle$, generated with the Quantum Package 2.0 code [184], to set up the relevant P space(s) as described in Step 2 of the above CIPSI-driven CC(P;Q) algorithm. The corresponding Q space(s) is/are automatically defined as the remaining triples absent in the $|\Psi^{(\mathrm{CIPSI})}\rangle$ wave function(s).

By design, as the input parameter $N_{\mathrm{det(in)}}$ used to terminate CIPSI runs increases, producing longer and longer CI expansions to represent wave functions $|\Psi^{(\mathrm{CIPSI})}\rangle$, the CC($P$;$Q$) energies $E_\mu^{(P+Q)}$ approach their CCSDT/EOMCCSDT parents. The underlying CC(P) and EOMCC(P) calculations converge the CCSDT/EOMCCSDT energetics too, but, as further elaborated on in our numerical examples, by ignoring the triples that were not captured by CIPSI, they do it at a much slower rate. In examining the convergence of the CIPSI-driven CC(P)/EOMCC(P) and CC(P;Q) energies toward CCSDT/EOMCCSDT, in Tables 3.1–3.5 and Fig. 3.1, we sampled the $N_{\mathrm{det(in)}}$ values in a roughly semi-logarithmic manner,

starting from $N_{\rm det(in)} = 1$. Since all of the calculations reported in this work adopted RHF determinants as reference functions, the $|\Psi^{(\rm CIPSI)}\rangle$ state becomes the RHF determinant and the resulting CC(P)/EOMCC(P) and CC(P;Q) energies become identical to those obtained in the RHF-based CCSD/EOMCCSD and CR-CC(2,3)/CR-EOMCC(2,3) calculations, respectively, when $N_{\rm det(in)} = 1$. Thus, we can regard the $N_{\rm det(in)}$ input variable defining CIPSI computations as the parameter connecting CCSD/EOMCCSD [in the CC(P)/EOMCC(P) case] or CR-CC(2,3)/CR-EOMCC(2,3) [in the case of CC(P;Q) runs] with CCSDT/EOMCCSDT. In the case of the CIPSI-driven CC(P) calculations discussed in this chapter, $T^{(P)} = T_1 + T_2 + T_3^{(\rm CIPSI)}$, where $T_3^{(\rm CIPSI)} = \sum_{|\Phi_{ijk}^{abc}\rangle \in |\Psi^{(\rm CIPSI)}\rangle} t_{abc}^{ijk} E_{ijk}^{abc}$ denotes the T_3 operator defined using the lists of triply excited determinants $|\Phi_{ijk}^{abc}\rangle$ contained in the final ground-state CIPSI wave function $|\Psi^{(\rm CIPSI)}\rangle$ (we use the usual notation in which i, j, k and a, b, c designate the occupied and unoccupied spin-orbitals in $|\Phi\rangle$, respectively, and E_{ijk}^{abc} is the elementary triple excitation operator generating $|\Phi_{ijk}^{abc}\rangle$ from $|\Phi\rangle$). For the excited-state applications employing the CIPSI-based EOMCC(P) approach, we use $R_\mu^{(P)} = r_{\mu,0}\mathbf{1} + R_{\mu,1} + R_{\mu,2} + R_{\mu,3}^{(\rm CIPSI)}$ and $L_\mu^{(P)} = \delta_{\mu,0}\mathbf{1} + L_{\mu,1} + L_{\mu,2} + L_{\mu,3}^{(\rm CIPSI)}$, where $R_{\mu,3}^{(\rm CIPSI)}$ and $L_{\mu,3}^{(\rm CIPSI)}$ are defined analogously to $T_3^{(\rm CIPSI)}$ using the appropriate list of triply excited determinants. The noniterative corrections $\delta_\mu(P;Q)$, Eq. (2.33), which in the case of the CIPSI-driven CC(P;Q) approach developed in this work captures the T_3 and $R_{\mu,3}$ effects not described by $T_3^{(\rm CIPSI)}$ and $R_{\mu,3}^{(\rm CIPSI)}$ and which involves the summation over the remaining triply excited determinants that are not included in $|\Psi^{(\rm CIPSI)}\rangle$, is determined using the appropriate form of the CC(P;Q) moment expansions, *i.e.*, $\delta_\mu(P;Q) = \sum_{|\Phi_{ijk}^{abc}\rangle \notin |\Psi^{(\rm CIPSI)}\rangle} \ell_{ijk}^{abc}(\mu)\, \mathfrak{M}_{abc}^{ijk}(\mu)$.

In addition to rapidly converging the parent CCSDT and EOMCCSDT energetics, as demonstrated in Section 3.2, and in analogy to the active-orbital-based [126–129] and semi-stochastic [130, 133, 134] CC(P;Q) approaches, the CIPSI-driven CC(P;Q) methodology examined in this work offers significant savings in the computational effort compared to full CCSDT and EOMCCSDT. This is largely related to the fact that, as shown in our calculations for F_2, cyclobutadiene, and CH^+ reported in Section 3.2, the convergence of the

CIPSI-driven CC(P;Q) energies toward their respective CCSDT or EOMCCSDT parents with the wave function termination parameter $N_{\rm det(in)}$, with the number of determinants used to generate the final CIPSI state $N_{\rm det(out)}$, and with the fractions of triples in the P space captured by the CIPSI algorithm is very fast. Indeed, the CPU times associated with the CIPSI runs using smaller $N_{\rm det(in)}$ values, resulting in smaller diagonalization spaces, are relatively short compared to the converged CIPSI computations. Next, as explained in detail in Refs. [130, 133, 134], the CC(P)/EOMCC(P) calculations using small fractions of triples in the P space, which is all one needs to converge the CCSDT- or EOMCCSDT-level energetics in the CIPSI-driven CC(P;Q) runs, are much faster than the corresponding CCSDT or EOMCCSDT computations. Furthermore, as also explained in Refs. [130, 133, 134], the computational cost of determining the CC(P;Q) correction $\delta_\mu(P;Q)$ is less than the cost of a single iteration of CCSDT/EOMCCSDT.

In examining the CIPSI-driven CC(P)/EOMCC(P) and CC(P;Q) energies shown in Tables 3.1–3.5 and Fig. 3.1, we are primarily interested in how fast they converge toward their parent CCSDT/EOMCCSDT values as $N_{\rm det(in)}$ and the fraction of triples in the P space increase. In the case of our calculations for F_2 and the automerization of cyclobutadiene, it is also helpful to examine the convergence of the $E_{\rm var}$, $E_{\rm var}+\Delta E^{(2)}$, and $E_{\rm var}+\Delta E^{(2)}_{\rm r}$ energies characterizing the ground-state CIPSI wave function $|\Psi^{\rm (CIPSI)}\rangle$. In reporting the variational and perturbatively corrected ground-state energies for F_2 (Tables 3.1 and 3.2 and Figure 3.1) and cyclobutadiene (Table 3.3) obtained in the corresponding CIPSI computations, we do what is often done in CIPSI calculations (see, e.g., Refs. [184, 194]) and compare them to their counterparts obtained by extrapolating the data obtained in the CIPSI runs defined by the largest $N_{\rm det(in)}$ values to the FCI limit. Specifically, following the procedure used in Ref. [194], we performed a linear fit of the last four $E_{{\rm var},k} + \Delta E^{(2)}_{{\rm r},k}$ energies leading to the final $|\Psi^{\rm (CIPSI)}\rangle$ state obtained for the largest value of $N_{\rm det(in)}$ in a given CIPSI sequence, plotted against the corresponding $\Delta E^{(2)}_{{\rm r},k}$ corrections, and extrapolated the resulting line to the $\Delta E^{(2)}_{{\rm r},k} = 0$ limit.

3.2 Numerical Examples

3.2.1 Dissociation of F_2

Our first example is the frequently studied dissociation of the fluorine molecule, as described by the cc-pVDZ basis set [10]. We chose this example, since it is well established that the CCSDT approach provides an accurate description of bond breaking in F_2 (see, *e.g.*, Refs. [146, 147, 241, 252]) and since we previously used it to benchmark the CC(P;Q)-based CC(t;3) approach and the semi-stochastic CC(P;Q) methods driven by the CIQMC and CCMC propagations. The results of our calculations for the F_2/cc-pVDZ system, in which the F–F bond length R was stretched from its equilibrium, $R_e = 2.66816$ bohr, value, where electron correlation effects are largely dynamical in nature, to $1.5R_e$, $2R_e$, and $5R_e$, where they gain an increasingly nondynamical character, are summarized in Table 3.1 and Fig. 3.1. The complexity of electron correlations in F_2 manifests itself in the rapidly growing magnitude of T_3 contributions as the F–F distance increases, as exemplified by the unsigned differences between the CCSDT and CCSD energies, which are 9.485 millihartree at $R = R_e$, 32.424 millihartree at $R = 1.5R_e$, 45.638 millihartree at $R = 2R_e$, and 49.816 millihartree at $R = 5R_e$, when the cc-pVDZ basis set is employed. The T_3 correlations grow with R so fast that in the $R = 2R_e$–$5R_e$ region, they become larger than the depth of the CCSDT potential (estimated at $\sim$44 millihartree when the CCSDT energy at $R = R_e$ is subtracted from its $R = 5R_e$ counterpart) and highly nonperturbative. The latter feature of T_3 contributions in the stretched F_2 molecule can be seen by examining the errors relative to CCSDT obtained in the CCSD(T) calculations at $R = 1.5R_e$, $2R_e$, and $5R_e$, which are -5.711, -23.596, and -39.348 millihartree, respectively, when the cc-pVDZ basis set is used. As shown in Table 3.1 [see the $N_{\mathrm{det(in)}} = 1$ CC(P;Q) energies], the CR-CC(2,3) triples correction to CCSD helps, reducing the large errors characterizing CCSD(T) to 1.735 millihartree at $R = 1.5R_e$, 1.862 millihartree at $R = 2R_e$, and 1.613 millihartree at $R = 5R_e$, which are much more acceptable, but, as demonstrated in our earlier active-orbital-based and semi-stochastic CC(P;Q) studies [126, 130, 134], further error reduction requires the relaxation of T_1 and T_2 clusters

in the presence of the dominant T_3 contributions. This is precisely what the CIPSI-driven CC(P;Q) methodology, where we use CIPSI runs to identify the leading triple excitations for the inclusion in the P space, allows us to do.

Indeed, with as little as 1,006–1,442 $S_z = 0$ determinants of the $A_g(D_{2h})$ symmetry in the final Hamiltonian diagonalization spaces (we used D_{2h} group, which is the largest Abelian subgroup of the $D_{\infty h}$ symmetry group of F_2, in our calculations), generated by the inexpensive $N_{\text{det(in)}} = 1,000$ CIPSI runs at $R = 1.5R_e$, $2R_e$, and $5R_e$, which capture very small fractions, on the order of 0.1–0.2 %, of all triples, the errors in the resulting CC(P;Q) energies relative to CCSDT are 0.202 millihartree at $R = 1.5R_e$, 0.132 millihartree at $R = 2R_e$, and 0.144 millihartree at $R = 5R_e$. This is an approximately tenfold error reduction compared to the CR-CC(2,3) calculations, in which T_1 and T_2 clusters, obtained with CCSD, are decoupled from T_3, and an improvement of the faulty CCSD(T) energetics by orders of magnitude. As explained in detail in Refs. [130, 133, 134] within the context of the CIQMC/CCMC-driven CC(P;Q) approaches, with the fractions of triples in the relevant P spaces being so small, the post-CIPSI steps of the CC(P;Q) calculations are not much more expensive than the CCSD-based CR-CC(2,3) computations and a lot faster than the corresponding CCSDT computations. The CC(P;Q) calculations using $N_{\text{det(in)}} = 1,000$ do not offer any improvements over CR-CC(2,3) at the equilibrium geometry, since the final diagonalization space of the underlying CIPSI run does not yet contain any triply excited determinants, and the CR-CC(2,3) energy at $R = R_e$ is already very accurate anyway, but with the relatively small additional effort corresponding to $N_{\text{det(in)}} = 10,000$, which results in 10,150 $S_z = 0$ determinants of the $A_g(D_{2h})$ symmetry in the final CIPSI diagonalization space and only 1.2 % of all triples in the P space, the unsigned error in the CC(P;Q) energy relative to its CCSDT parent substantially decreases, from 0.240 millihartree, when $N_{\text{det(in)}} \leq 1,000$, to 67 microhartree, when $N_{\text{det(in)}}$ is set at 10,000. The use of $N_{\text{det(in)}} = 10,000$ for the remaining three geometries considered in Table 3.1 and Fig. 3.1 produces similarly compact $|\Psi^{(\text{CIPSI})}\rangle$ wave functions, spanned by 11,578–19,957 determinants, similarly small fractions

of triples in the corresponding P spaces, ranging from 1.5 % at $R = 1.5R_e$ to 2.2 % at $R = 5R_e$, and even smaller errors in the CC(P;Q) energies relative to CCSDT.

It is clear from Table 3.1 and Fig. 3.1 that the convergence of the CIPSI-driven CC(P;Q) energies toward CCSDT with the wave function termination parameter $N_{\mathrm{det(in)}}$, with the number of determinants used to generate the final CIPSI state $|\Psi^{(\mathrm{CIPSI})}\rangle$ [$N_{\mathrm{det(out)}}$], and with the fraction of triples in the P space captured by the CIPSI procedure is very fast. The uncorrected CC(P) energies converge to CCSDT too, but they do it at a considerably slower rate than their CC(P;Q) counterparts. For example, the CIPSI-driven CC(P) calculations reduce the 9.485, 32.424, 45.638, and 49.816 millihartree errors relative to CCSDT obtained with CCSD to 1.419, 0.991, 0.922, and 0.764 millihartree, respectively, when $N_{\mathrm{det(in)}} = 50,000$, which translates in the $N_{\mathrm{det(out)}}$ values ranging between 65,172 and 92,682 and about 5–9 % of all triples included in the underlying P spaces, but the errors characterizing the corresponding CC(P;Q) energies are already at the level of 20–30 microhartree at this stage, which is obviously a substantial improvement over the CC(P) results. It is also worth noticing that the convergence of the CIPSI-driven CC(P) and CC(P;Q) energies toward their CCSDT parents with $N_{\mathrm{det(in)}}$ [or $N_{\mathrm{det(out)}}$] is considerably faster than the convergence of the corresponding variational and perturbatively corrected CIPSI energies toward the extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ values. This is in line with the above observations that indicate that the CIPSI-driven CC(P;Q) calculations are capable of recovering the parent CCSDT energetics, even when electronic quasi-degeneracies and T_3 clusters become significant, out of the unconverged CIPSI runs that rely on relatively small diagonalization spaces. We observed similar patterns when comparing the semi-stochastic, CIQMC- and CCMC-driven, CC(P)/EOMCC(P) and CC(P;Q) calculations with the underlying CIQMC/CCMC propagations [130, 133, 134, 253].

In analogy to the previously considered deterministic, active-orbital-based [126, 127, 129, 131, 132] and semi-stochastic, CIQMC/CCMC-based [130, 134] CC(P;Q) studies, the convergence of the CIPSI-driven CC(P;Q) computations toward the parent CCSDT energetics

remains equally rapid when we use basis sets larger than cc-pVDZ. This is illustrated in Table 3.2, where we show the results of the CIPSI-driven CC(P;Q) calculations for the F_2 molecule at $R = 2R_e$ using the cc-pVTZ basis set [10]. As pointed out in Ref. [134], and in analogy to the cc-pVDZ basis, the T_3 contribution characterizing the stretched F_2/cc-pVTZ system in which the internuclear separation is set at twice the equilibrium bond length, estimated by forming the difference between the CCSDT and CCSD energies at -62.819 millihartree, is not only very large, but also larger, in absolute value, than the corresponding CCSDT dissociation energy, which is about 57 millihartree, when the CCSDT energy at R_e is subtracted from its $5R_e$ counterpart. It is also highly nonperturbative at the same time, as demonstrated by the -26.354 millihartree error relative to CCSDT obtained with CCSD(T). Again, the CR-CC(2,3) triples correction to CCSD, equivalent to the $N_{\mathrm{det(in)}} = 1$ CC(P;Q) calculation in Table 3.2, works a lot better than CCSD(T), but the 4.254 millihartree error relative to CCSDT remains. With as little as 5,118 $S_z = 0$ determinants of the $A_g(D_{2h})$ symmetry in the final diagonalization space obtained by the nearly effortless $N_{\mathrm{det(in)}} = 5,000$ CIPSI run, which captures 0.03 % of all triples, the difference between the CC(P;Q) and CCSDT energies decreases to 0.345 millihartree, and with the help of the $N_{\mathrm{det(in)}} = 50,000$ CIPSI calculation, which is still relatively inexpensive, resulting in 82,001 $S_z = 0$ $A_g(D_{2h})$-symmetric determinants in the final diagonalization space and less than 1 % of the triples in the P space, the error in the CC(P;Q) energy relative to its CCSDT parent reduces to less than 0.1 millihartree. Similarly to the cc-pVDZ basis, the convergence of the CIPSI-driven CC(P;Q) energies toward CCSDT with $N_{\mathrm{det(in)}}$, $N_{\mathrm{det(out)}}$, and the fraction of triples in the P space captured by the CIPSI algorithm is not only fast, when the larger cc-pVTZ basis set is employed, but also much faster than in the case of the uncorrected CC(P) calculations. Once again, as $N_{\mathrm{det(in)}}$ increases, the rate of convergence of the CIPSI-driven CC(P) and CC(P;Q) energies toward their CCSDT parent is higher than those characterizing the corresponding variational and perturbatively corrected CIPSI energies in their attempt to recover the extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energy.

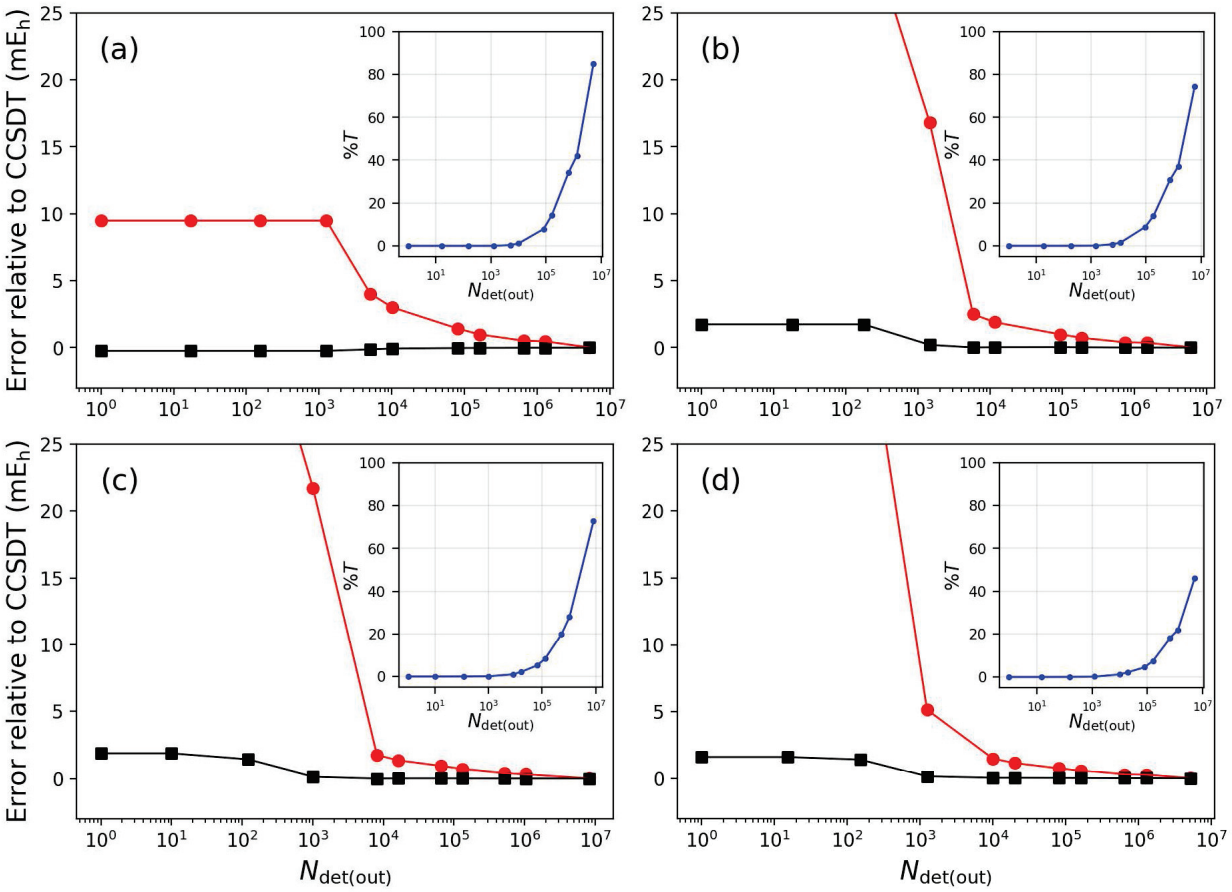

Figure 3.1 Convergence of the CIPSI-based CC(P) (red lines and circles) and CC(P;Q) (black lines and squares) energies toward their CCSDT parents as functions of the actual numbers of determinants, $N_{\mathrm{det(out)}}$, defining the sizes of the final wave functions $|\Psi^{(\mathrm{CIPSI})}\rangle$ generated in the underlying CIPSI runs, for the F_2/cc-pVDZ molecule in which the F–F bond length R was set at (a) R_e, (b) $1.5R_e$, (c) $2R_e$, and (d) $5R_e$, where $R_e = 2.66816$ bohr is the equilibrium geometry. The P spaces used in the CC(P) calculations consisted of all singles and doubles and subsets of triples contained in the final $|\Psi^{(\mathrm{CIPSI})}\rangle$ states of the underlying CIPSI runs, whereas the Q spaces needed to compute the corresponding CC(P;Q) corrections $\delta(P;Q)$ were defined as the remaining triples absent in $|\Psi^{(\mathrm{CIPSI})}\rangle$. The insets show the percentages of triples captured by the CIPSI runs as functions of $N_{\mathrm{det(out)}}$. Adapted from Ref. [135].

Table 3.1 Convergence of the CIPSI-based CC(P) and CC(P;Q) energies toward CCSDT, alongside the variational and perturbatively corrected CIPSI energies, for the F_2/cc-pVDZ molecule in which the F–F bond length R was set at R_e, $1.5R_e$, $2R_e$, and $5R_e$, where $R_e = 2.66816$ bohr is the equilibrium geometry. In all post-RHF calculations, the two lowest-lying core orbitals were frozen and the Cartesian components of d orbitals were employed throughout. Each CIPSI run was initiated from the RHF reference determinant and the MBPT-based stopping parameter η and selection factor f were set at 10^{-6} hartree and 2, respectively. Adapted from Ref. [135].

R/R_e	$N_{\mathrm{det(in)}}$ / $N_{\mathrm{det(out)}}$	% T	E_{var}[a]	$E_{\mathrm{var}} + \Delta E^{(2)}$[a]	$E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$[a]	CC(P)[b]	CC($P;Q$)[b]
1.0	1 / 1	0	418.057[c]	−94.150[d]	−12.651	9.485[e]	−0.240[f]
	1,000 / 1,266	0	65.926	1.480	2.079	9.485	−0.240
	5,000 / 5,072	0.4	23.596	−0.133(0)	−0.069(0)	4.031	−0.129
	10,000 / 10,150	1.2	19.197	0.045(2)	0.084(2)	3.010	−0.067
	50,000 / 81,288	7.9	11.282	0.133(1)	0.145(1)	1.419	−0.031
	100,000 / 162,430	14.5	9.222	0.138(1)	0.146(1)	0.983	−0.020
	500,000 / 649,849	34.3	5.630	0.092(1)	0.095(1)	0.519	−0.009
	1,000,000 / 1,300,305	42.2	4.816	0.072(0)	0.074(0)	0.464	−0.008
	5,000,000 / 5,187,150	85.1	1.161	0.015(2)	0.016(2)	0.023	−0.001
1.5	1 / 1	0	541.109[c]	−130.718[d]	137.819	32.424[e]	1.735[f]
	1,000 / 1,442	0.1	77.306	5.218	5.948	16.835	0.202
	5,000 / 5,773	0.7	21.091	0.811(2)	0.856(2)	2.490	0.009
	10,000 / 11,578	1.5	17.333	0.811(2)	0.839(2)	1.892	0.028
	50,000 / 92,682	8.8	10.879	0.762(1)	0.771(1)	0.991	0.033
	100,000 / 185,350	13.9	9.243	0.632(1)	0.639(1)	0.727	0.023
	500,000 / 742,754	30.8	5.586	0.391(1)	0.393(1)	0.390	0.005
	1,000,000 / 1,484,218	37.1	4.795	0.330(0)	0.332(0)	0.362	0.004
	5,000,000 / 5,907,228	74.3	1.165	0.079(2)	0.079(2)	0.028	−0.000
2.0	1 / 1	0	640.056[c]	−159.482[d]	289.080	45.638[e]	1.862[f]
	1,000 / 1,006	0.1	105.265	5.589	7.036	21.727	0.132
	5,000 / 8,118	1.1	17.355	0.787(1)	0.815(1)	1.725	−0.003
	10,000 / 16,291	2.1	14.555	0.860(1)	0.878(1)	1.338	0.012
	50,000 / 65,172	5.2	11.064	0.800(1)	0.810(1)	0.922	0.015
	100,000 / 130,448	8.4	9.410	0.655(1)	0.662(1)	0.695	0.009
	500,000 / 521,578	19.8	5.929	0.375(1)	0.378(1)	0.400	0.005
	1,000,000 / 1,043,539	28.0	4.820	0.306(0)	0.308(0)	0.314	0.002
	5,000,000 / 8,190,854	72.8	0.764	0.047(1)	0.047(1)	0.009	−0.000
5.0	1 / 1	0	730.244[c]	−183.276[d]	430.051	49.816[e]	1.613[f]
	1,000 / 1,241	0.2	70.879	6.966	7.491	5.154	0.144
	5,000 / 9,977	1.2	14.531	1.033(0)	1.050(0)	1.489	0.029
	10,000 / 19,957	2.2	12.550	1.039(0)	1.050(0)	1.156	0.029
	50,000 / 79,866	4.6	9.025	0.764(1)	0.770(1)	0.764	0.022
	100,000 / 159,668	7.6	7.495	0.580(1)	0.584(1)	0.584	0.013
	500,000 / 639,593	18.0	4.391	0.276(0)	0.277(0)	0.294	0.003
	1,000,000 / 1,278,976	22.0	3.682	0.238(0)	0.239(0)	0.259	0.003
	5,000,000 / 5,099,863	46.1	0.675	0.041(1)	0.041(1)	0.009	−0.000

[a]Energies at each internuclear separation are reported as errors, in millihartree, relative to the extrapolated $E_{\mathrm{var}}+\Delta E_{\mathrm{r}}^{(2)}$ energy found using a linear fit based on the last four $E_{\mathrm{var},k}+\Delta E_{\mathrm{r},k}^{(2)}$ values leading to the largest CIPSI wave function obtained with $N_{\mathrm{det(in)}} = 5,000,000$ following the procedure used in Ref. [194]. These extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energies at $R = R_e$, $1.5R_e$, $2R_e$, and $5R_e$ are −199.104422(6), −199.069043(1), −199.060152(8), and −199.059647(11) hartree, respectively.

[b]Energies reported as errors relative to CCSDT, in millihartree. The total CCSDT energies at $R = R_e$, $1.5R_e$, $2R_e$, and $5R_e$ are −199.102796, −199.065882, −199.058201, and −199.058586 hartree, respectively.

[c]Equivalent to RHF.

[d]Equivalent to the second-order MBPT energy using the Epstein–Nesbet denominator.

[e]Equivalent to CCSD.

[f]Equivalent to CR-CC(2,3).

Table 3.2 Convergence of the CIPSI-based CC(P) and CC(P;Q) energies toward CCSDT, alongside the variational and perturbatively corrected CIPSI energies, for the F_2/cc-pVTZ molecule in which the F–F bond length R was fixed at $2R_e$, where $R_e = 2.66816$ bohr is the equilibrium geometry. In all post-RHF calculations, the two lowest-lying core orbitals were frozen and the spherical components of d and f orbitals were employed throughout. Each CIPSI run was initiated from the RHF reference determinant and the MBPT-based stopping parameter η and selection factor f were set at 10^{-6} hartree and 2, respectively. Adapted from Ref. [135].

$N_{\mathrm{det(in)}}$ / $N_{\mathrm{det(out)}}$	% T	E_{var}[a]	$E_{\mathrm{var}} + \Delta E^{(2)}$[a]	$E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$[a]	CC(P)[b]	CC(P;Q)[b]
1 / 1	0	758.849[c]	−165.740[d]	340.460	62.819[e]	4.254[f]
10 / 18	0	441.567	−0.554	31.337	62.819	4.254
100 / 156	0.00	393.749	6.420	28.790	58.891	3.683
1,000 / 1,277	0.01	253.172	13.595(0)	20.323	42.564	1.579
5,000 / 5,118	0.03	123.591	10.874(1)	12.149(1)	18.036	0.345
10,000 / 10,239	0.06	73.122	7.202(5)	7.636(5)	11.439	0.198
50,000 / 82,001	0.84	29.674	3.371(2)	3.428(2)	4.898	0.061
100,000 / 163,866	1.58	27.002	3.068(2)	3.113(2)	4.157	0.049
500,000 / 655,859	3.75	22.301	2.517(1)	2.547(1)	3.111	0.014
1,000,000 / 1,311,633	5.58	20.244	2.292(1)	2.316(1)	2.739	0.009
5,000,000 / 5,253,775	13.3	14.499	1.645(1)	1.657(1)	1.866	−0.015

[a]Energies are reported as errors, in millihartree, relative to the extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energy found using a linear fit based on the last four $E_{\mathrm{var},k} + \Delta E_{\mathrm{r},k}^{(2)}$ values leading to the largest CIPSI wave function obtained with $N_{\mathrm{det(in)}} = 5,000,000$ following the procedure used in Ref. [194]. The extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energy is −199.242119(59) hartree.

[b]Energies reported as errors relative to CCSDT, in millihartree. The total CCSDT energy is −199.238344 hartree..

[c]Equivalent to RHF.

[d]Equivalent to the second-order MBPT energy using the Epstein–Nesbet denominator.

[e]Equivalent to CCSD.

[f]Equivalent to CR-CC(2,3).

3.2.2 Automerization of Cyclobutadiene

A more demanding test, shown in Table 3.3, is the frequently examined [127, 130, 134, 212, 254–268] automerization of cyclobutadiene. In this case, one of the key challenges is an accurate determination of the activation barrier, which requires a well-balanced description of the nondegenerate, rectangle-shaped, closed-shell reactant (or the equivalent product) species, in which electron correlation effects are largely dynamical in nature, and the quasi-degenerate, square-shaped, transition state characterized by substantial nondynamical correlations associated with its strongly diradical character. Experimental estimates of the activation barrier for the automerization of cyclobutadiene, which range from 1.6 to 10 kcal/mol [256, 257], are not very precise, but the most accurate single- and multireference *ab initio* computations, compiled, for example, in Refs. [127, 255, 268], place it in the 6–10 kcal/mol range. This, in particular, applies to the CCSDT approach [127, 254], which is of the primary interest in the present study. Indeed, if we, for example, use the reactant and transition-state geometries obtained with the multireference average-quadratic CC (MR-AQCC) approach [269, 270] in Ref. [262] and the cc-pVDZ basis set, the CCSDT value of the activation energy characterizing the automerization of cyclobutadiene becomes 7.627 kcal/mol [127], in good agreement with the best *ab initio* calculations carried out to date. By adopting the same geometries and basis set in this initial benchmark study of the CIPSI-driven CC(P;Q) methodology, we can examine if the CC(P;Q) calculations using the P spaces constructed with the help of CIPSI runs are capable of converging this result. The main challenge here is that the T_3 effects, estimated as the difference between the CCSDT and CCSD energies, are not only very large, but also hard to balance. When the cc-pVDZ basis set used in this study is employed, they are -26.827 millihartree for the reactant and -47.979 millihartree for the transition state. Furthermore, in the case of the transition state, the coupling of the lower-rank T_1 and T_2 clusters with their higher-rank T_3 counterpart is so large that none of the noniterative triples corrections to CCSD provide a reasonable description of the activation barrier [127, 254, 255]. This, in particular, applies to the CR-CC(2,3) approach, equivalent

to the $N_{\mathrm{det(in)}} = 1$ CC(P;Q) calculation, which produces an activation barrier exceeding 16 kcal/mol, when the cc-pVDZ basis set is employed, instead of less than 8 kcal/mol obtained with CCSDT (see Table 3.3). The failure of CR-CC(2,3) to provide an accurate description of the activation energy is a consequence of its inability to accurately describe the transition state. Indeed, the difference between the CR-CC(2,3) and CCSDT energies at the transition-state geometry is 14.636 millihartree, when the cc-pVDZ basis set is employed, as opposed to only 0.848 millihartree obtained for the reactant. As discussed in detail in Refs. [127, 255], other triples corrections to CCSD, including the widely used CCSD(T) approach, face similar problems. It was previously demonstrated in Refs. [127, 130, 134] that the deterministic CC(P;Q)-based CC(t;3) approach and the semi-stochastic CC(P;Q) calculations using CIQMC and CCMC are capable of accurately approximating the CCSDT energies of the reactant and transition-state species and the CCSDT activation barrier, so it is interesting to explore if the CIPSI-driven CC(P;Q) methodology can do the same.

As shown in Table 3.3, the CC(P;Q) calculations using CIPSI to identify the dominant triply excited determinants for the inclusion in the P space are very efficient in converging the CCSDT energetics. With the final diagonalization spaces spanned by a little over 110,000 $S_z = 0$ determinants of the $\mathrm{A}_g(D_{2h})$ symmetry (we used D_{2h} group for both the D_{2h}-symmetric reactant and the D_{4h}-symmetric transition state in our calculations), generated in the relatively inexpensive CIPSI runs defined by $N_{\mathrm{det(in)}} = 100,000$ that capture 0.1 % of all triples, the 0.848 millihartree, 14.636 millihartree, and 8.653 kcal/mol errors in the reactant, transition-state, and activation energies relative to CCSDT obtained with CR-CC(2,3) are reduced by factors of 2–4, to 0.382 millihartree, 3.507 millihartree, and 1.961 kcal/mol, respectively, when the CC(P;Q) approach is employed. When $N_{\mathrm{det(in)}}$ is increased to 500,000, resulting in about 890,000–900,000 $S_z = 0$ determinants of the $\mathrm{A}_g(D_{2h})$ symmetry in the final diagonalization spaces used by CIPSI and 1.0–1.2 % of the triples in the resulting P spaces, the errors in the CC(P;Q) reactant, transition-state, and activation energies relative to CCSDT become 0.267 millihartree, 0.432 millihartree, and 0.104 kcal/mol. Clearly, these are

great improvements compared to the initial $N_{\mathrm{det(in)}} = 1$, i.e., CR-CC(2,3), values, especially if we realize that with the fractions of triples being so small, the post-CIPSI steps of the CC(P;Q) computations are not only a lot faster than the parent CCSDT runs, but also not much more expensive than the corresponding CR-CC(2,3) calculations, as elaborated on in Refs. [130, 133, 134].

In analogy to the previously discussed case of bond breaking in F_2, the convergence of the CIPSI-driven CC(P;Q) energies toward CCSDT for the reactant and transition-state species defining the automerization of cyclobutadiene with $N_{\mathrm{det(in)}}$, $N_{\mathrm{det(out)}}$, and the fractions of triples in the relevant P spaces captured by the underlying CIPSI runs is not only very fast, but also significantly faster than that characterizing the uncorrected CC(P) calculations. For each of the two species, the CC(P) energies converge toward their CCSDT parent in a steady fashion, but, as shown in Table 3.3, their convergence is rather slow, emphasizing the importance of correcting the results of the CC(P) calculations for the missing triple excitations not captured by the CIPSI runs using smaller diagonalization spaces. Similarly to the previously examined active-orbital-based [126, 127, 129, 131, 132] and CIQMC/CCMC-based [130, 134] CC(P;Q) approaches, the moment correction $\delta_0(P;Q)$, defined by Eq. (2.33), is very effective in this regard. Another interesting observation, which can be made based on the results presented in Table 3.3, is that while the CC(P) energies for the individual reactant and transition-state species converge toward their CCSDT parent values in a steady fashion, the corresponding activation barrier values behave in a less systematic manner, oscillating between about -1 and 1 kcal/mol when $N_{\mathrm{det(in)}} = 500,000$–$15,000,000$. One might try to eliminate this behavior, which is a consequence of a different character of the many-electron correlation effects in the reactant and transition-state species, by merging the P spaces used to perform the CC(P) calculations for the two structures, but, as shown in Table 3.3, the CC(P;Q) correction $\delta_0(P;Q)$, which is highly effective in capturing the missing T_3 correlations, takes care of this problem too. As $N_{\mathrm{det(in)}}$, $N_{\mathrm{det(out)}}$, and the fractions of triples in the P spaces used in the CC(P) calculations for the reactant and transition state increase,

the CC(P;Q) values of the activation barrier converge toward its CCSDT parent rapidly and in a smooth manner, eliminating, at least to a large extent, the need to equalize the P spaces used in the underlying CC(P) steps. As in the case of bond breaking in the fluorine molecule, the convergence of the CIPSI-driven CC(P) and CC(P;Q) energies toward their CCSDT parents with $N_{\mathrm{det(in)}}/N_{\mathrm{det(out)}}$ is considerably faster than the convergence of the variational and perturbatively corrected CIPSI energies toward the extrapolated $E_{\mathrm{var}}+\Delta E_{\mathrm{r}}^{(2)}$ values, but we must keep in mind that the calculated CCSDT and extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energies, while representing the respective parent limits for the CC(P;Q) and CIPSI calculations, are fundamentally different quantities, especially when higher–than–triply excited cluster components, which are not considered in the present calculations, become significant. As one might anticipate, the $N_{\mathrm{det(in)}}$ values needed to accurately represent the CCSDT energies of the reactant and transition-state species of cyclobutadiene by the CIPSI-driven CC(P;Q) approach are considerably larger than those used in the analogous CC(P;Q) calculations for the smaller F_2 system, but they are orders of magnitude smaller than the values of $N_{\mathrm{det(in)}}$ required to obtain the similarly well converged $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energetics in the underlying CIPSI runs.

Table 3.3 Convergence of the CIPSI-based CC(P) and CC(P;Q) energies toward CCSDT, alongside the variational and perturbatively corrected CIPSI energies, for the reactant (R) and transition-state (TS) species involved in the automerization of cyclobutadiene, as described by the spherical cc-pVDZ basis set, and for the corresponding barrier height. In all post-RHF calculations, the four lowest-lying core orbitals were frozen. Each CIPSI run was initiated from the RHF reference determinant and the MBPT-based stopping parameter η and selection factor f were set at 10^{-6} hartree and 2, respectively. Adapted from Ref. [135].

Species	$N_{\mathrm{det(in)}}$ / $N_{\mathrm{det(out)}}$	% T	E_{var}[a]	$E_{\mathrm{var}} + \Delta E^{(2)}$[a]	$E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$[a]	CC(P)[b]	CC(P;Q)[b]
R	1 / 1	0	598.120[c]	−83.736[d]	120.809	26.827[e]	0.848[f]
	50,000 / 55,653	0.0	121.880	26.065(182)	28.096(178)	25.468	0.678
	100,000 / 111,321	0.1	109.688	23.819(163)	25.397(160)	22.132	0.382
	500,000 / 890,582	1.2	93.413	19.049(141)	20.167(139)	16.260	0.267
	1,000,000 / 1,781,910	2.0	89.989	18.322(137)	19.348(135)	15.359	0.251
	5,000,000 / 7,125,208	7.9	78.122	16.311(123)	17.045(122)	10.794	0.150
	10,000,000 / 14,253,131	11.8	73.250	15.514(115)	16.146(114)	9.632	0.127
	15,000,000 / 28,493,873	25.8	60.872	12.842(96)	13.260(95)	4.817	0.046
TS	1 / 1	0	632.707[c]	−102.816[d]	282.246	47.979[e]	14.636[f]
	50,000 / 56,225	0.0	146.895	45.357(180)	47.696(176)	42.132	9.563
	100,000 / 112,481	0.1	130.832	36.716(183)	38.673(179)	31.723	3.507
	500,000 / 899,770	1.0	93.288	18.106(139)	19.251(137)	14.742	0.432
	1,000,000 / 1,800,183	1.6	89.049	17.458(142)	18.482(140)	13.645	0.412
	5,000,000 / 7,195,780	5.4	78.472	15.587(124)	16.346(123)	10.720	0.260
	10,000,000 / 14,400,744	9.7	71.784	14.397(114)	15.016(113)	8.358	0.155
	15,000,000 / 28,793,512	15.2	63.375	12.587(102)	13.058(101)	7.080	0.108
Barrier	1 / 1 ; 1	0 ; 0	21.703[c]	−11.973[d]	101.303	13.274[e]	8.653[f]
	50,000 / 55,653 ; 56,225	0.0 ; 0.0	15.697	12.106(161)	12.299(157)	10.457	5.576
	100,000 / 111,321 ; 112,481	0.1 ; 0.1	13.268	8.093(154)	8.331(151)	6.018	1.961
	500,000 / 890,582 ; 899,770	1.2 ; 1.0	−0.079	−0.592(124)	−0.574(122)	−0.953	0.104
	1,000,000 / 1,781,910 ; 1,800,183	2.0 ; 1.6	−0.590	−0.542(124)	−0.544(122)	−1.075	0.101
	5,000,000 / 7,125,208 ; 7,195,780	7.9 ; 5.4	0.220	−0.454(110)	−0.439(109)	−0.047	0.069
	10,000,000 / 14,253,131 ; 14,400,744	11.8 ; 9.7	−0.920	−0.701(102)	−0.710(100)	−0.800	0.017
	15,000,000 / 28,493,873 ; 28,793,512	25.8 ; 15.2	1.571	−0.159(88)	−0.127(87)	1.420	0.039

[a]Energies reported as errors, in millihartree, relative to the extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energy found using a linear fit based on the last four $E_{\mathrm{var},k} + \Delta E_{\mathrm{r},k}^{(2)}$ values leading to the largest CIPSI wave function obtained with $N_{\mathrm{det(in)}} = 15,000,000$ following the procedure used in Ref. [194]. These extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energies for the R and TS species are −154.249292(314) and −154.235342(321) hartree, respectively. The barrier heights are reported as errors, in kcal/mol, relative to the reference value of 8.753(0) kcal/mol obtained using the extrapolated $E_{\mathrm{var}} + \Delta E_{\mathrm{r}}^{(2)}$ energies of the R and TS species.

[b]Energies reported as errors relative to CCSDT, in millihartree. The total CCSDT energies of the R and TS species are −154.244157 and −154.232002 hartree, respectively. Barrier heights are reported as errors in kcal/mol relative to the CCSDT value of 7.627 kcal/mol.

[c]Equivalent to RHF.

[d]Equivalent to the second-order MBPT energy using the Epstein–Nesbet denominator.

[e]Equivalent to CCSD.

[f]Equivalent to CR-CC(2,3).

3.2.3 Electronic Excitation Spectrum of CH^+

The previous examples of bond breaking in F_2 and automerization of cyclobutadiene convincingly demonstrate that the CIPSI-driven CC(P;Q) approach is capable of rapidly and efficiently recovering the CCSDT energetics using small fractions of triply excited determinants included in the underlying P spaces, which are identified with the help of relatively inexpensive CIPSI wave calculations characterized by small $N_{\rm det(in)}$, $N_{\rm det(out)}$, and % T values. In this section, we will examine if the same remains true in an excited-state application, where the goal is to converge the EOMCCSDT energetics for the electronic excitation spectrum of CH^+ using the CIPSI-driven CC(P;Q) methodology.

Following the analogous study of the CH^+ molecule carried out in Refs. [133, 253] using the semi-stochastic CC(P;Q) approach, we employed the [5s3p1d/3s1p] basis set of Ref. [242] and performed all-electron calculations for the ground and three lowest-energy singlet excited states of the $\Sigma^+(C_{\infty v})$ symmetry, denoted as $n\,^1\Sigma^+$, n = 1–4, where $n = 1$ corresponds to the ground state and n = 2–4 represent excited states, the two lowest singlet $\Pi(C_{\infty v})$ states, labeled $1\,^1\Pi$ and $2\,^1\Pi$, and the two lowest singlet $\Delta(C_{\infty v})$ states, denoted as $1\,^1\Delta$ and $2\,^1\Delta$. As discussed in points 1 and 2 of the algorithm outlined in Section 3.1, our current implementation of the CIPSI-driven CC(P;Q) approach constructs separate P spaces for each irrep of the molecular point group by extracting the list of triply excited determinants contained in the corresponding CIPSI wave function for the lowest-energy state of a given spatial symmetry. Accordingly, the P (and Q) spaces defining the CC(P;Q) calculations for the $^1\Sigma^+$, $^1\Pi$, and $^1\Delta$ states of CH^+ were constructed using the information obtained from the CIPSI computations describing the $1\,\Sigma^+$, $1\,^1\Pi$, and $1\,^1\Delta$ states of CH^+, respectively. Our calculations were performed using the C_{2v} subgroup of the full non-Abelian $C_{\infty v}$ symmetry describing the CH^+ molecule, and as a result, the underlying CIPSI runs for the $1\,^1\Sigma^+$ ground state were initiated in a usual fashion using the RHF determinant $|\Phi\rangle$, whereas the CIPSI calculations for the $1\,^1\Pi$ and $1\,^1\Delta$ states of CH^+ were carried out by defining the initial $\mathscr{V}_{\rm int}^{(0)}$ subspace containing a singly excited determinant of the $B_1(C_{2v})$ symmetry describing

the $3\sigma \rightarrow 1\pi$ excitation and a doubly excited determinant of the $A_2(C_{2v})$ symmetry characterizing the $3\sigma^2 \rightarrow 1\pi^2$ transition, respectively. The results of our CIPSI-driven CC(P;Q) calculations for the CH^+ molecule in the ground-state equilibrium geometry, in which the C–H bond length was set to $R = R_e = 2.13713$ bohr, as well as the stretched geometry with internuclear separation of $R = 2R_e = 4.27426$ bohr are reported in Tables 3.4 and 3.5, respectively.

We begin by discussing the CIPSI-driven CC(P;Q) calculations for the ground and low-lying excited states of the CH^+ molecule at the equilibrium structure, shown in Table 3.4. When the C–H bond length is set to $R = R_e$, the $1\,^1\Sigma^+$ ground state is nondegenerate and characterized by predominantly dynamical correlations. Similarly, most of the excited states are dominated by one-electron excitations, except for the $2\,^1\Sigma^+$, $1\,^1\Delta$, and $2\,^1\Delta$ states of CH^+, which contain significant contributions from two-electron transitions corresponding to $3\sigma^2 \rightarrow 1\pi^2$ in the case of $2^1\Sigma^+$ and $1^1\Delta$ and $2\sigma\,3\sigma \rightarrow 1\pi^2$ in the case of $2^1\Delta$, and which require an accurate treatment of the T_3 and $R_{\mu,3}$ correlation effects. As indicated by the $N_{\rm det(in)} = 1$ data in Table 3.4, the CR-CC(2,3) (for the $1^1\Sigma^+$ ground state) and CR-EOMCC(2,3) (excited states) calculations are very accurate in describing the ground and excited states of CH^+ at its equilibrium structure, and with the exception of the $2\,^1\Sigma^+$ and $2\,^1\Pi$ states, result in energies that are within 1 millihartree of the parent CCSDT and EOMCCSDT data. Even for the $2\,^1\Sigma^+$ and $2\,^1\Pi$ states of CH^+, the CR-EOMCC(2,3) calculations are characterized by rather modest errors of 1.373 and 2.805 millihartree relative to EOMCCSDT, respectively. In analogy to the previously examples studying the F_2 molecule and cyclobutadiene, the CC(P;Q) calculations based on the tiny $N_{\rm det(in)} = 1,000$ CIPSI runs, which capture 1.4 % of $S_z = 0$ triply excited determinants of the $A_1(C_{2v})$ symmetry for the $^1\Sigma^+$ states and 3.9 % of triply excited determinants of the $S_z = 0$ $B_1(C_{2v})$ symmetry for the $^1\Pi$ states, substantially improve the CR-EOMCC(2,3) results, reducing the 1.373 and 2.805 millihartree errors relative to EOMCCSDT obtained with CR-EOMCC(2,3) for the $2\,^1\Sigma^+$ and $2\,^1\Pi$ states of CH^+ to 0.465 and 0.478 millihartree, respectively. The convergence of the CIPSI-driven

CC(P;Q) calculations toward CCSDT/EOMCCSDT is so rapid that when the wave function input parameter $N_{\mathrm{det(in)}}$ is set to 5,000, which is minuscule when compared to the 31,912, 27,180, or 22,012, $S_z = 0$ triply excited determinants of the $\Sigma^+(C_{\infty v})$, $\Pi(C_{\infty v})$, and $\Delta(C_{\infty v})$ symmetries included in the full CCSDT/EOMCCSDT calculations, respectively, the errors in the resulting CC(P;Q) energetics relative to CCSDT/EOMCCSDT range between 0.003 and 0.128 millihartree. As in the ground-state case, it is worth noting that the CIPSI-based CC(P;Q) calculations provide the CCSDT- and EOMCCSDT-level energetics using small fractions of triply excited determinants in the underlying P spaces identified with the help of extremely small CIPSI computations, which result in massive reductions in computational effort compared to the parent CCSDT/EOMCCSDT methodologies.

Similar remarks about the performance of the CIPSI-driven CC(P;Q) calculations in converging the CCSDT and EOMCCSDT energetics apply when examining the results shown in Table 3.5 for the CH^+ molecule in the significantly stretched geometry. In analogy to the case of bond breaking in F_2, as the C–H bond length increases from $R = R_e$ to $2R_e$, the strength of the T_3 and $R_{\mu,3}$ correlations increases substantially. In particular, all of the excited states of CH^+ considered in this work acquire significant contributions from two-electron excitations, which cannot be described in the basic EOMCCSD calculations. Indeed, as seen in the $N_{\mathrm{det(in)}} = 1$ data reported in Table 3.5, the errors relative to EOMCCSDT obtained with EOMCCSD for the $2^1\Sigma^+$, $3^1\Sigma^+$, $4^1\Sigma^+$, $1^1\Pi$, $2^1\Pi$, $1^1\Delta$, and $2^1\Delta$ states of CH^+ are 17.140, 19.930, 32.639, 13.522, 21.200, 44.495, and 144.414 millihartree, respectively. Evidently, when the C–H bond is stretched, none of the low-lying singlet excited states of CH^+ considered in this study can be quantitatively studied using the basic EOMCCSD method, and for the $^1\Delta$ states, EOMCCSD cannot provide a qualitative description. In order to cope with the substantial contributions from two-electron excitations in these excited states, the T_3 and $R_{\mu,3}$ effects must be incorporated. As expected, the noniterative CR-EOMCC(2,3) corrections help improve the poor performance of EOMCCSD, reducing the 17.140, 19.930, 32.639, 13.522, 21.200, 44.495, and 144.414 millihartree errors relative to

EOMCCSDT obtained using EOMCCSD for the $2\ ^1\Sigma^+$, $3\ ^1\Sigma^+$, $4\ ^1\Sigma^+$, $1\ ^1\Pi$, $2\ ^1\Pi$, $1\ ^1\Delta$, and $2\ ^1\Delta$ states of CH^+ to 1.646, -2.869, 12.657, 2.304, -1.428, -4.524, and -63.405 millihartree, respectively. Although the noniterative CR-EOMCC(2,3) triples correction provides a significant improvement over EOMCCSD, it is clear that ignoring the coupling between the T_3 and $R_{\mu,3}$ components and their lower-rank T_1, T_2, $R_{\mu,1}$, and $R_{\mu,2}$ counterparts in this case results in substantial errors when describing the $4\ ^1\Sigma^+$ and $2\ ^1\Delta$ states of CH^+ in addition to non-negligible errors for the remaining states.

When the C–H bond length is stretched, the CIPSI-driven CC(P;Q) approach becomes an indispensable and extremely efficient scheme for providing a highly accurate description of the vertical excitation spectrum of CH^+. As in the case of $R = R_e$, one does not have to perform much additional work to improve the CR-CC(2,3) and CR-EOMCC(2,3) results using the CIPSI-based CC(P;Q) calculations. Again, when the CIPSI input parameter $N_{\text{det(in)}}$ is set to 1,000, the large 12.657 and -63.405 millihartree errors relative to EOMCCSDT obtained CR-EOMCC(2,3) for the $4\ ^1\Sigma^+$ and $2\ ^1\Delta$ states of CH^+ are reduced by nearly a hundred-fold to 0.999 and 0.855 millihartree, respectively. At the same time, the 1.646, -2.869, 2.304, -1.428, and -4.524 millihartree differences between the CR-EOMCC(2,3) and EOMCCSDT energetics for the $2\ ^1\Sigma^+$, $3\ ^1\Sigma^+$, $1\ ^1\Pi$, $2\ ^1\Pi$, and $1\ ^1\Delta$ states of CH^+ are reduced to 1.095, -1.282, 0.047, 0.095, and 0.047 millihartree, respectively, when the CIPSI-driven CC(P;Q) calculations based on the $N_{\text{det(in)}} = 1,000$ CIPSI wave functions are employed. We note, once again, that the CIPSI wave functions computed using $N_{\text{det(in)}} = 1,000$ are tiny, containing just 1.8 % of the $S_z = 0$ $A_1(C_{2v})$-symmetric triply excited determinants, 3.2 % of the $S_z = 0$ $B_1(C_{2v})$-symmetric triply excited determinants, and 3.7 % of the $S_z = 0$ $A_2(C_{2v})$-symmetric triply excited determinants, which means that the CC(P;Q) steps are not much more expensive than their CR-CC(2,3)/CR-EOMCC(2,3) counterparts and the CIPSI diagonalizations, involving so few determinants, add negligible additional computational cost to the calculations. Similar to the $R = R_e$ case, increasing $N_{\text{det(in)}}$ from 1,000 to 5,000 provides a near-perfect description of the parent CCSDT/EOMCCSDT energetics.

Table 3.4 Convergence of the CIPSI-based CC(P)/EOMCC(P) and CC($P;Q$) energies toward CCSDT/EOMCCSDT for CH^+, calculated using the [5s3p1d/3s1p] basis set of Ref. [242], at the C–H internuclear distance $R = R_e = 2.13713$ bohr. The P spaces used in the CC(P) and EOMCC(P) calculations were defined as all singles, all doubles, and subsets of triples extracted from CIPSI wave functions for the lowest states of the relevant symmetries. Each CIPSI run was initiated using the $\mathscr{V}_{\rm int}^{(0)}$ space spanned by the appropriate determinant [the RHF state for the $^1\Sigma^+$ states, the $3\sigma \rightarrow 1\pi$ singly excited determinant of the $B_1(C_{2v})$ symmetry for the $^1\Pi$ states, and the $3\sigma^2 \rightarrow 1\pi^2$ doubly excited determinant of the $A_2(C_{2v})$ symmetry for the $^1\Delta$ states], and the MBPT-based stopping parameter η and selection factor f were set at 10^{-6} hartree and 2, respectively. The Q spaces used in constructing the CC($P;Q$) corrections consisted of the triples not captured by CIPSI.

$N_{\rm det(in)}$	$1\,^1\Sigma^+$			$2\,^1\Sigma^+$		$3\,^1\Sigma^+$		$4\,^1\Sigma^+$		$1\,^1\Pi$			$2\,^1\Pi$		$1\,^1\Delta$			$2\,^1\Delta$	
	P[a]	$(P;Q)$[b]	%T[c]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	%T[c]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	%T[c]	P[a]	$(P;Q)$[b]
1[d]	1.845	0.063	0	19.694	1.373	3.856	0.787	5.537	0.954	3.080	0.792	0	11.656	2.805	34.304	−0.499	0	34.685	0.350
1,000	0.619	0.017	1.4	9.744	0.465	2.275	0.192	1.793	0.118	0.636	0.123	3.9	3.699	0.478	2.040	0.160	2.6	10.027	0.425
5,000	0.100	0.003	8.5	1.351	0.086	0.200	0.060	0.310	0.012	0.128	0.033	13.9	2.043	0.157	0.077	0.015	17.0	1.012	0.056
10,000	0.034	0.001	16.3	0.404	0.020	0.031	0.019	0.149	0.008	0.035	0.012	23.3	1.188	0.040	0.022	0.005	26.9	0.312	0.013
50,000	0.002	0.000	47.3	0.052	0.001	−0.078	0.001	−0.005	0.001	0.002	0.000	51.1	0.063	0.002	0.001	0.000	56.2	0.130	0.003
100,000	0.001	0.000	68.9	0.012	0.000	−0.107	−0.002	−0.007	0.001	0.000	0.000	70.9	0.019	0.001	0.000	0.000	73.3	0.011	0.000

[a]Errors, in millihartree, characterizing the CC(P) (the $1\,^1\Sigma^+$ ground state) and EOMCC(P) (excited states) energies relative to the corresponding CCSDT and EOMCCSDT data, which are −38.019516, −37.702621, −37.522457, −37.386872, −37.900921, −37.498143, −37.762113, and −37.402308 hartree for the $1\,^1\Sigma^+$, $2\,^1\Sigma^+$, $3\,^1\Sigma^+$, $4\,^1\Sigma^+$, $1\,^1\Pi$, $2\,^1\Pi$, $1\,^1\Delta$, and $2\,^1\Delta$ states of CH^+, respectively.

[b]Errors, in millihartree, in the CC($P;Q$) energies relative to the corresponding CCSDT and EOMCCSDT data provided in footnote (a).

[b]The %T values are the percentages of triply excited determinants contained in the CIPSI wave function for the lowest state of a given symmetry [the $1\,^1\Sigma^+ = 1\,^1A_1(C_{2v})$ ground state for the $^1\Sigma^+$ states, the $^1B_1(C_{2v})$ component of the $1\,^1\Pi$ state for the $^1\Pi$ states, and the $^1A_2(C_{2v})$ component of the $1\,^1\Delta$ state for the $^1\Delta$ states].

[c]The CC(P) and EOMCC(P) energies at $N_{\rm det(in)} = 1$ are identical to the energies obtained in the CCSD and EOMCCSD calculations. The $N_{\rm det(in)} = 1$ CC($P;Q$) energies are equivalent to the CR-CC(2,3) (the ground state) and the CR-EOMCC(2,3) (excited states) energies.

Table 3.5 Same as Table 3.4 for the stretched C–H internuclear distance $R = 2R_e = 4.27426$ bohr.

$N_{det(in)}$	$1\,^1\Sigma^+$			$2\,^1\Sigma^+$		$3\,^1\Sigma^+$		$4\,^1\Sigma^+$		$1\,^1\Pi$			$2\,^1\Pi$		$1\,^1\Delta$			$2\,^1\Delta$	
	P[a]	$(P;Q)$[b]	%T[c]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	%T[c]	P[a]	$(P;Q)$[b]	P[a]	$(P;Q)$[b]	%T[c]	P[a]	$(P;Q)$[b]
1[d]	5.002	0.012	0	17.140	1.646	19.930	−2.869	32.639	12.657	13.552	2.304	0	21.200	−1.428	44.495	−4.524	0	144.414	−63.405
1,000	0.434	0.006	1.8	7.456	1.095	10.848	−1.282	9.609	0.999	0.452	0.047	3.2	1.758	0.095	0.430	0.047	3.7	8.160	0.855
5,000	0.060	0.002	7.7	0.852	0.033	0.890	0.011	1.860	0.109	0.052	0.006	11.0	0.159	0.012	0.047	0.009	11.0	0.401	0.010
10,000	0.015	0.001	15.6	0.333	0.010	0.332	0.012	0.741	0.054	0.013	0.002	19.2	0.043	0.003	0.013	0.004	18.9	0.123	0.002
50,000	0.001	0.000	40.4	0.016	0.000	0.018	0.001	0.046	0.001	0.001	0.000	44.7	0.004	0.001	0.001	0.001	41.5	0.006	0.000
100,000	0.000	0.000	57.3	0.004	0.000	0.005	0.001	0.018	0.000	0.000	0.000	61.2	0.002	0.000	0.001	0.001	58.6	0.002	0.000

[a]Errors, in millihartree, characterizing the CC(P) (the $1\,^1\Sigma^+$ ground state) and EOMCC(P) (excited states) energies relative to the corresponding CCSDT and EOMCCSDT data, which are −37.900394, −37.704834, −37.650242, −37.495275, −37.879532, −37.702345, −37.714180, and −37.494031 hartree for the $1\,^1\Sigma^+$, $2\,^1\Sigma^+$, $3\,^1\Sigma^+$, $4\,^1\Sigma^+$, $1\,^1\Pi$, $2\,^1\Pi$, $1\,^1\Delta$, and $2\,^1\Delta$ states of CH^+, respectively.

[b]Errors, in millihartree, in the CC($P;Q$) energies relative to the corresponding CCSDT and EOMCCSDT data provided in footnote (a).

[b]The %T values are the percentages of triply excited determinants contained in the CIPSI wave function for the lowest state of a given symmetry [the $1\,^1\Sigma^+ = 1\,^1A_1(C_{2v})$ ground state for the $^1\Sigma^+$ states, the $^1B_1(C_{2v})$ component of the $1\,^1\Pi$ state for the $^1\Pi$ states, and the $^1A_2(C_{2v})$ component of the $1\,^1\Delta$ state for the $^1\Delta$ states].

[c]The CC(P) and EOMCC(P) energies at $N_{det(in)} = 1$ are identical to the energies obtained in the CCSD and EOMCCSD calculations. The $N_{det(in)} = 1$ CC($P;Q$) energies are equivalent to the CR-CC(2,3) (the ground state) and the CR-EOMCC(2,3) (excited states) energies.

CHAPTER 4

THE ADAPTIVE CC(P;Q) FRAMEWORK

In Chapter 3, we introduced a new formulation of CC(P;Q) in which the content of the P and Q spaces is automatically generated with the help of the information provided by external CIPSI calculations. Similar to its semi-stochastic CC(P;Q) counterpart, which exploited the CIQMC or CCMC wave function propagations in order to identify the higher–than–doubly excited determinants spanning the P and Q spaces, the CIPSI-driven CC(P;Q) approach successfully replaces the user- and system-dependent selection of active orbitals used in the original active-space CC(P;Q) hierarchy with more black-box non-CC considerations.

In general, however, the reliance on non-CC sources of information in the CC(P;Q) framework is not entirely ideal for several reasons. First of all, from a purely practical point of view, it less convenient to have our CC(P;Q) calculations based on interfacing multiple independent computational modules, as this hinders the accessibility of these methods for non-expert users. On a more fundamental level, the linear wave function ansatz adopted in the CIPSI and CIQMC approaches does not ensure size-extensivity by construction, unlike the CC/EOMCC calculations. Thus, one may wonder how reliable CI-based calculations are in providing information about the importance of the *connected* contributions to higher-rank excitations when studying larger many-electron systems (especially when attempting to recover the thermodynamic limit, where all truncated CI approximations fail). Finally, for most problems of chemical interest, any Hilbert space sampling procedure starting from a one-dimensional initial subspace will likely spend significant time during its early stages sampling the singly (for non-HF references) and doubly excited determinants in the many-electron Hilbert space. This is a waste of time from the point of view of our CC(P;Q) calculations, which often include all singly and doubly excited determinants in the P spaces *a priori*. Needless to say, it would be better to directly explore the higher-rank excitation manifolds of interest using the information contained within the lower-level CC/EOMCC wave functions present from the very beginning of the CC(P;Q) calculations.

In this chapter, we aim to solve the above-mentioned challenges by formulating a novel adaptive CC(P;Q) approach capable of rapidly converging high-level CC/EOMCC energetics of the CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and similar types in a fully automated fashion, even in cases of stronger correlations, which is free from the use of active orbitals and information provided by the non-CC or stochastic sources exploited in the previous variants of CC(P;Q). The key idea, which is discussed in Section 4.1, is an adaptive selection of the excitation manifolds defining higher–than–two-body components of the cluster and excitation operators driven by the CC(P;Q) moment expansions. The benefits of the resulting adaptive CC(P;Q) methodology are illustrated in Section 4.2 using a number of molecular examples, including the significantly stretched F_2 and F_2^+ molecules and automerization of cyclobutadiene (Section 4.2.1), the singlet–triplet gaps in organic biradicals (Section 4.2.2), the degenerate Cope rearrangement in bullvalene (Section 4.2.3), and the ground- and excited-state PESs of the water molecule along its O–H bond-breaking coordinate (Section 4.2.4), where the goal is to recover the full CCSDT and EOMCCSDT energetics when the noniterative triples corrections to CCSD/EOMCCSD struggle to provide a highly accurate description, or fail entirely. Along with demonstrating the rapid convergence toward the parent CCSDT/EOMCCSDT results, we discuss the CPU timings in Section 4.3 to highlight vast reductions in the computational effort relative to the high-level CC/EOMCC calculations offered by the adaptive CC(P;Q) calculations.

4.1 Theory and Algorithmic Details

4.1.1 Adaptive CC($\boldsymbol{P}$;$\boldsymbol{Q}$): General considerations

We now proceed to describing the adaptive CC(P;Q) methodology applicable to converging any high-level CC or EOMCC theory of interest in a fully automated fashion. In developing the adaptive CC(P;Q) approach, we have been inspired by the CIPSI algorithm [181, 183, 184], and especially its modern implementation in Refs. [183, 184]. As a result, many of the elements entering our adaptive CC(P;Q) approach have direct analogs in the CIPSI framework, and because of this strong similarity, we will first briefly recapitulate the

key elements in CIPSI, which were discussed in Section 2.3.

We recall that the main idea of CIPSI is a series of Hamiltonian diagonalizations in increasingly large, iteratively defined, subspaces of the many-electron Hilbert space, which are followed by correcting the resulting energies using expressions originating from the second-order many-body perturbation theory to estimate the remaining correlations. Adopting the notation used in Section 2.3, if $\mathscr{V}_{\rm int}^{(k)}$, where $k = 0, 1, 2, \ldots$ enumerates the consecutive CIPSI iterations, designates the current diagonalization subspace, the $\mathscr{V}_{\rm int}^{(k+1)}$ space for the subsequent Hamiltonian diagonalization is constructed by arranging the candidate determinants $|\Phi_\alpha\rangle$ from outside $\mathscr{V}_{\rm int}^{(k)}$ for a potential inclusion in $\mathscr{V}_{\rm int}^{(k+1)}$ in descending order according to the absolute values of the perturbative corrections $e_{\alpha,k}^{(2)}$ associated with them, starting from the $|\Phi_\alpha\rangle$s characterized by the largest $|e_{\alpha,k}^{(2)}|$ contributions, moving toward those that have smaller $|e_{\alpha,k}^{(2)}|$ values, and continuing until the number of determinants in $\mathscr{V}_{\rm int}^{(k+1)}$ reaches a desired dimension (in the CIPSI algorithm of Refs. [183, 184], until the dimension of $\mathscr{V}_{\rm int}^{(k+1)}$ exceeds that of $\mathscr{V}_{\rm int}^{(k)}$ by the user-defined factor $f > 1$). We can adopt a similar strategy in designing P spaces for the CC(P;Q) computations.

Indeed, within the CC(P;Q) framework, we can interpret the moment expansions Eq. (2.33) as a sum of contributions $\delta_{\mu,K}(P;Q)$ due to the individual determinants from the Q space, $|\Phi_K\rangle \in \mathscr{H}^{(Q)}$, evaluated as

$$\delta_{\mu,K}(P;Q) = \ell_{\mu,K}(P)\,\mathfrak{M}_{\mu,K}(P), \tag{4.1}$$

which, in analogy to the perturbative $e_{\alpha,k}^{(2)}$ corrections that measure the significance of the candidate determinants $|\Phi_\alpha\rangle$ in CIPSI, determine the importance of the Q-space determinants $|\Phi_K\rangle$. One can, therefore, propose an adaptive, self-improving, CC(P;Q) scheme, in which we construct an approximation to the high-level CC/EOMCC approach of interest (in the numerical examples in Section 4.2, CCSDT and EOMCCSDT) by a series of CC(P;Q) calculations using increasingly large, iteratively defined, P spaces $\mathscr{H}^{(P)}(k)$, where $k = 0, 1, 2, \ldots$ enumerates the consecutive CC(P;Q) computations, with the corresponding Q subspaces $\mathscr{H}^{(Q)}(k)$ being defined as complementary excitation spaces, such that $\mathscr{H}^{(P)}(k) \oplus \mathscr{H}^{(Q)}(k)$ is

always equivalent to the entire excitation manifold appropriate for the CC/EOMCC method we are targeting, independent of k (when targeting CCSDT/EOMCCSDT, all singly, doubly, and triply excited determinants, when targeting CCSDTQ/EOMCCSDT, all singly, doubly, triply, and quadruply excited determinants, etc.). With this iterative scheme in mind, Eq. (2.33) adapted to the kth P space $\mathscr{H}^{(P)}(k)$ and its complementary Q space $\mathscr{H}^{(Q)}(k)$ containing the excited determinants of interest not included in $\mathscr{H}^{(P)}(k)$ can be written as

$$\delta_{\mu}(P^{(k)};Q^{(k)}) = \sum_{|\Phi_K\rangle \in \mathscr{H}^{(Q)}(k)} \delta_{\mu,K}(P^{(k)},Q^{(k)}), \tag{4.2}$$

where

$$\delta_{\mu,K}(P^{(k)},Q^{(k)}) = \ell_{\mu,K}(P^{(k)})\,\mathfrak{M}_{\mu,K}(P^{(k)}) \tag{4.3}$$

is the contribution to $\delta_{\mu}(P^{(k)};Q^{(k)})$ that corresponds to a given Q-space determinant $|\Phi_K\rangle \in \mathscr{H}^{(Q)}(k)$. For clarity of this description, the P and Q symbols seen in Eqs. (4.2) and (4.3), which represent the $\mathscr{H}^{(P)}(k)$ and $\mathscr{H}^{(Q)}(k)$ spaces used in the kth iteration of the adaptive CC(P;Q) procedure, are labeled with the additional superscript (k).

The initial P space $\mathscr{H}^{(P)}(0)$ can be a conveniently chosen zeroth-order excitation manifold, such as the space of singly and doubly excited determinants, $|\Phi_i^a\rangle$ and $|\Phi_{ij}^{ab}\rangle$, respectively, and the remaining subspaces are constructed via a recursive process analogous to that used in CIPSI, where the P space $\mathscr{H}^{(P)}(k+1)$ is obtained by augmenting its $\mathscr{H}^{(P)}(k)$ predecessor with a subset of the leading Q-space determinants $|\Phi_K\rangle \in \mathscr{H}^{(Q)}(k)$ [when targeting CCSDT/EOMCCSDT, the leading triply excited determinants $|\Phi_{ijk}^{abc}\rangle$ outside $\mathscr{H}^{(P)}(k)$, when targeting CCSDTQ/EOMCCSDT, the leading triply and quadruply excited determinants outside $\mathscr{H}^{(P)}(k)$, etc.] identified with the help of corrections $\delta_{\mu}(P^{(k)};Q^{(k)})$, Eq. (4.2). In analogy to CIPSI, one can enlarge the current subspace $\mathscr{H}^{(P)}(k)$ to construct the $\mathscr{H}^{(P)}(k+1)$ space for the subsequent CC(P;Q) computation by arranging the candidate determinants $|\Phi_K\rangle \in \mathscr{H}^{(Q)}(k)$ in descending order according to the $\delta_{\mu,K}(P^{(k)};Q^{(k)})$ corrections associated with them, starting from the $|\Phi_K\rangle$s characterized by the largest $|\delta_{\mu,K}(P^{(k)};Q^{(k)})|$

contributions, moving toward those that have smaller $|\delta_{\mu,K}(P^{(k)};Q^{(k)})|$ values, and continuing until the total number of the Q-space determinants in $\mathscr{H}^{(P)}(k+1)$ reaches a certain fraction of all higher-rank determinants relevant to the CC/EOMCC approach of interest (e.g., triples when targeting CCSDT/EOMCCSDT or triples and quadruples when targeting CCSDTQ/EOMCCSDT). The adaptive CC(P;Q) algorithm is also shown graphically in Figure 4.1 below.

Clearly, the adaptive CC(P;Q) procedure, as described above, guarantees convergence toward the high-level CC/EOMCC theory, but, following the computational cost analysis of the CC(P;Q) methods [130, 133, 134] (cf., also, Refs. [126–129, 131, 132, 135, 136]), in order to be an attractive approach, it has to be capable of recovering the target CC/EOMCC energetics to a very good accuracy with small fractions of higher-rank determinants relevant to the CC/EOMCC method of interest in the underlying P spaces. In order to test the validity and efficiency of this novel, self-driven CC(P;Q) model, we focused on implementing the adaptive CC(P;Q) approach that aims at converging the CC/EOMCC calculations with a full treatment of triples.

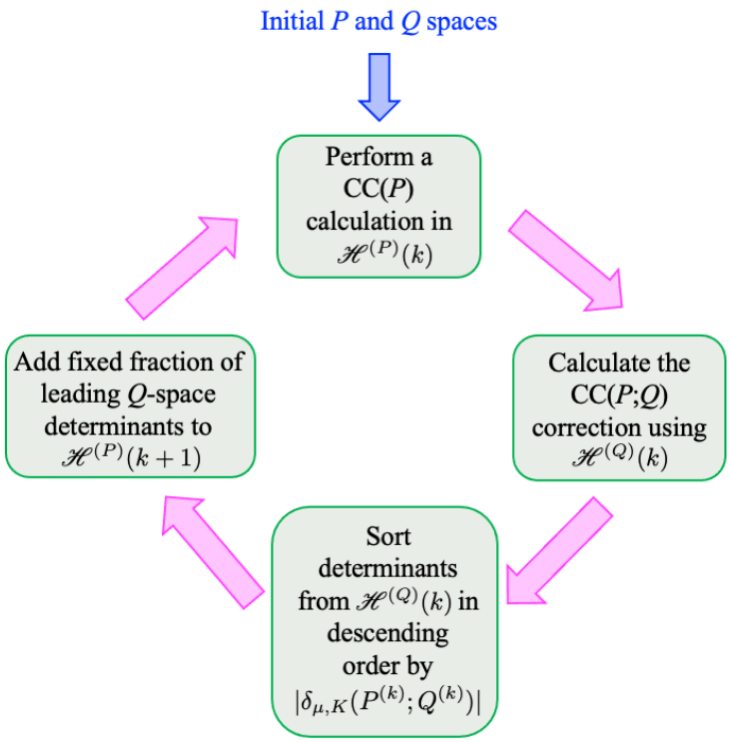

Figure 4.1 Schematic depiction of the adaptive CC(P;Q) algorithm described in Section 4.2.1, which is generally applicable to converging the energetics corresponding to any high-level CC/EOMCC calculation of interest. The recursive procedure begins with an initial definition of the P space, which we always take to be spanned by all singly and doubly excited determinants, and the corresponding Q space appropriately defined for the target level of CC/EOMCC theory (*e.g.,* triply excited determinants if the goal is to converge CCSDT/EOMCCSDT or triply and quadruply excited determinants if one is interested in converging CCSDT/EOMCCSDTQ, etc.), where the sum of P and Q spaces remains fixed for all calculations executed during iterative adaptive CC(P;Q) algorithm. Once the initial P and Q spaces are defined, the adaptive CC(P;Q) calculation is carried out following the cyclic sequence of steps shown in this figure.

4.1.2 Adaptive CC(P;Q) approaches aimed at CCSDT/EOMCCSDT and their computational implementation

Our adaptive CC(P;Q) methodology aimed at converging CCSDT/EOMCCSDT applies Eqs. (4.2) and (4.3) to a situation where the kth P space $\mathscr{H}^{(P)}(k)$ is spanned by all singly and doubly excited determinants and a subset of triply excited determinants identified by the adaptive CC(P;Q) algorithm and the associated Q space $\mathscr{H}^{(Q)}(k)$ consists of the remaining triples not included in $\mathscr{H}^{(P)}(k)$, *i.e.*,

$$\delta_\mu(P^{(k)};Q^{(k)}) = \sum_{|\Phi^{abc}_{ijk}\rangle \in \mathscr{H}^{(Q)}(k)} \delta^{(k)}_{ijk,abc}(\mu), \tag{4.4}$$

where

$$\delta^{(k)}_{ijk,abc}(\mu) = \ell^{abc}_{\mu,ijk}(P^{(k)})\, \mathfrak{M}^{ijk}_{\mu,abc}(P^{(k)}) \tag{4.5}$$

is the individual energy correction corresponding to a given triply excited determinant $|\Phi^{abc}_{ijk}\rangle \in \mathscr{H}^{(Q)}(k)$. The $\ell^{abc}_{\mu,ijk}(P^{(k)})$ and $\mathfrak{M}^{ijk}_{\mu,abc}(P^{(k)})$ quantities in Eq. (4.5) are the coefficients $\ell_{\mu,K}(P)$ and moments $\mathfrak{M}_{\mu,K}(P)$ entering the CC(P;Q) correction $\delta_\mu(P;Q)$, Eq. (2.33), adapted to the above definitions of the $\mathscr{H}^{(P)}(k)$ and $\mathscr{H}^{(Q)}(k)$ spaces.

In the adaptive CC(P;Q) code that we incorporated in the CCpy package available on GitHub [228], the $k = 0$ P space $\mathscr{H}^{(P)}(0)$, used to initiate the calculations, is defined as all singly and doubly excited determinants and the corresponding Q space $\mathscr{H}^{(Q)}(0)$ consists of all triples. This means that the $k = 0$ CC(P)/EOMCC(P) and CC(P;Q) energies are respectively identical to those obtained with CCSD and CR-CC(2,3) for the ground state and EOMCCSD and CR-EOMCC(2,3) for excited states. We then follow the recursive procedure described in Section 4.1 by moving more and more triply excited determinants from the Q to P spaces. We enlarge the kth subspace $\mathscr{H}^{(P)}(k)$, which consists of all singles and doubles and, when $k > 0$, the subset of triples identified in the previous iteration, to construct the $\mathscr{H}^{(P)}(k+1)$ space for the subsequent CC(P;Q) computation by arranging the candidate triply excited determinants $|\Phi^{abc}_{ijk}\rangle \in \mathscr{H}^{(Q)}(k)$ in descending order according to the $\delta^{(k)}_{ijk,abc}(\mu)$ corrections associated with them, starting from the triples characterized

by the largest $|\delta^{(k)}_{ijk,abc}(\mu)|$ contributions, moving toward those that have smaller $|\delta^{(k)}_{ijk,abc}(\mu)|$ values, and continuing until the total number of triply excited determinants in $\mathscr{H}^{(P)}(k+1)$ represents an increase of the number of triples in $\mathscr{H}^{(P)}(k)$ by $f_{\mathrm{Ad}}\%$, where f_{Ad} is a user-defined growth rate (analogous to the selection factor f used in CIPSI). When enlarging the $\mathscr{H}^{(P)}(k)$, $k = 0, 1, 2, \ldots$, spaces in our numerical tests to date, we have found that setting $0 < f_{\mathrm{Ad}} \leq 1$ results in optimal performance, allowing us to use as small fractions of triples in the consecutive CC(P)/EOMCC(P) runs as possible, so that the iterative CC(P)/EOMCC(P) steps preceding the determination of the $\delta_\mu(P^{(k)};Q^{(k)})$ corrections are much less expensive than those of the CCSDT or EOMCCSDT targets, while still providing rapid convergence within respect to k. In medium-sized systems containing ∼20–30 correlated electrons and ∼100 orbitals, an upper bound of $f_{\mathrm{Ad}} = 1$ is reasonable, since the numbers of triply excited determinants used to enlarge the P spaces are not too large. When we examine even larger systems, such as the bullvalene molecule studied in Section 4.2.3, which contains 50 correlated electrons and 440 orbitals, 1% of the triples manifold corresponds to more than 10^9 triply excited determinants. In these cases, we have also found success by setting the growth rate in an automated fashion according to $f_{\mathrm{Ad}} = (N_S + N_D)/N_T$, where N_S, N_D, and N_T represent the numbers of singly, doubly, and triply excited determinants relevant to the electronic state of interest, respectively. This choice of f_{Ad} grows the P spaces using the size of the CCSD problem as an increment, which ensures that the resulting CC(P)/EOMCC(P) and CC(P;Q) calculations do not become overwhelmed by the number of triply excited determinants.

In the adaptive CC(P;Q) computations reported in this work, we distinguish between the relaxed and unrelaxed schemes. In the relaxed variant of the adaptive CC(P;Q), we solve the CC(P)/EOMCC(P) equations for the singly, doubly, and triply excited cluster amplitudes corresponding to the content of each $\mathscr{H}^{(P)}(k)$ space and recompute the corresponding triples corrections $\delta^{(k)}_{ijk,abc}(\mu)$ accordingly, increasing the number of triples, when going from $\mathscr{H}^{(P)}(k)$ to $\mathscr{H}^{(P)}(k+1)$, by $f_{\mathrm{Ad}}\%$. In the unrelaxed CC(P;Q) approach, we simply pick a particular fraction of triples for inclusion in the P space (say, $3f_{\mathrm{Ad}}\%$ or $5f_{\mathrm{Ad}}\%$) that have the largest

initial $|\delta^{(0)}_{ijk,abc}(\mu)|$ values determined using the T_1, T_2, $R_{\mu,1}$, and $R_{\mu,2}$ amplitudes obtained with CCSD/EOMCCSD [i.e., the largest absolute values of the $\delta^{(0)}_{ijk,abc}(\mu)$ contributions to the triples correction of CR-CC(2,3) or CR-EOMCC(2,3)] and solve the CC(P)/EOMCC(P) equations for the singly, doubly, and triply excited cluster amplitudes with this particular fraction of triples, adding the correction due to the missing T_3 and $R_{\mu,3}$ correlations using Eq. (4.2). The relaxed variant of the adaptive CC($P;Q$) approach has an advantage of approaching the target CCSDT/EOMCCSDT energetics as $k \to \infty$, but the unrelaxed scheme, which does not require an iterative construction of multiple $\mathscr{H}^{(P)}(k)$ spaces, is less expensive.

It is also worth mentioning that when implementing the adaptive CC($P;Q$) approach aimed at converging the ground-state CCSDT energetics, we have the option of invoking the so-called two-body approximation, which was successfully used in some of the earlier CC($P;Q$)-related work [126–129, 214]. In the two-body approximation, the ground-state moments $\mathfrak{M}^{ijk}_{0,abc}(P^{(k)})$ and coefficients $\ell^{abc}_{0,ijk}(P^{(k)})$ entering Eq. (4.5) are replaced by $\langle\Phi^{abc}_{ijk}|\overline{H}^{(P)}(2)|\Phi\rangle$ and $\langle\Phi|(\mathbf{1} + L_{0,1} + L_{0,2})\,\overline{H}^{(P)}(2)|\Phi^{abc}_{ijk}\rangle/(E^{(P)}_0 - \langle\Phi^{abc}_{ijk}|\overline{H}^{(P)}(2)|\Phi^{abc}_{ijk}\rangle)$, respectively, where $\overline{H}^{(P)}(2) = e^{-T_1-T_2}He^{T_1+T_2}$, with T_1 and T_2 designating the one- and two-body components of the cluster operator $T^{(P)}$ obtained in the CC(P) calculations in $\mathscr{H}^{(P)}(k)$, is an approximation to the true similarity-transformed Hamiltonian $\overline{H}^{(P)} = e^{-T_1-T_2-T^{(P)}_3}He^{T_1+T_2+T^{(P)}_3}$. The one- and two-body deexcitation operators $L_{0,1}$ and $L_{0,2}$, which enter the formula for the $\ell^{abc}_{0,ijk}(P^{(k)})$ coefficients, are obtained by solving the left eigenvalue problem involving $\overline{H}^{(P)}(2)$ in the space spanned by singly and doubly excited determinants [this step in which the ground-state left CC(P) eigenvalue problem is replaced by its simpler CCSD-like counterpart is not possible in the EOMCC(P) case since the right and left EOMCC eigenvectors must be obtained within the same subspace of the many-electron Hilbert space]. Thus, in analogy to $T^{(P)}$, the three-body component $L^{(P)}_{0,3}$ of the deexcitation operator $L^{(P)}_0$ corresponding to the P space $\mathscr{H}^{(P)}(k)$ is neglected. The two-body approximation avoids the more complex computational steps associated with the use of the complete

form of $\overline{H}^{(P)}$ in determining the $\delta^{(k)}_{ijk,abc}(0)$ corrections, while preserving the philosophy of the CC(P;Q) algorithm and accounting for the relaxation of T_1, T_2, $L_{0,1}$, and $L_{0,2}$ in the presence of the three-body components of $T^{(P)}$ obtained in the CC(P) calculations in $\mathscr{H}^{(P)}(k)$.

4.2 Numerical Examples

4.2.1 Bond Breaking in F_2 and F_2^+ and Automerization of Cyclobutadiene

In order to illustrate the benefits offered by the adaptive CC(P;Q) methodology, we begin by discussing our initial test of the relaxed and unrelaxed adaptive CC(P;Q) approaches based on Eqs. (4.2) and (4.5), especially their ability to recover the parent CCSDT energetics when the noniterative triples corrections to CCSD struggle, using the significantly stretched F_2 and F_2^+ molecules and the reactant (R) and transition-state (TS) species involved in the automerization of cyclobutadiene, along with the corresponding barrier height, as examples. With the exception of F_2^+, these systems were examined earlier in Chapter 3 within the context of the CIPSI-driven CC(P;Q) calculations. In analogy to Sections 3.2.1 and 3.2.2, we employed the cc-pVTZ basis set [10] to describe the fluorine molecule and its cation (Table 4.1), in which the respective F–F bond lengths r were stretched to $2r_e$, where r_e represents the equilibrium geometry (2.66816 bohr for F_2 and 2.49822 bohr for F_2^+), and the cc-pVDZ basis [10] for cyclobutadiene (Table 4.2), where the geometries of the D_{2h}-symmetric R and D_{4h}-symmetric TS species correspond to the structures taken from Ref. [262] optimized using the MR-AQCC approach.

We chose the stretched F_2 and F_2^+ molecules and the automerization of cyclobutadiene as initial examples because (i) all of these systems feature large, and in the case of F_2^+ and the TS species of cyclobutadiene, nonperturbative, T_3 correlations and represent meaningful tests for our CC(P;Q)-based approximations to CCSDT (see the discussions in Refs. [126, 127, 134, 135, 137] for more precise remarks), and (ii) it allows us to maintain consistency with previous studies on the active-orbital-based CC(t;3) [126, 127], CIQMC-/CCMC-based semi-stochastic CC(P;Q) [130, 134], and the CIPSI-driven CC(P;Q) approach [135] (cf. Sections 3.2.1 and 3.2.2), which also considered these examples. Here, we show that the

fully automated, "black-box", adaptive CC(P;Q) methodology, as implemented in our CCpy package [228], is equally, and oftentimes more, efficient than its active-orbital-based, semi-stochastic, and CIPSI-based CC(P;Q) predecessors in improving the CR-CC(2,3) results and generating the CCSDT-quality data at small fractions of the computational effort associated with the conventional CCSDT calculations. In all of our adaptive CC(P;Q) calculations for F_2, F_2^+, and cyclobutadiene, the two-body approximation was invoked in constructing the CC(P;Q) moment expansions and the core orbitals correlating with the 1s shells of fluorine and carbon were kept frozen in the post-SCF steps.

Indeed, as shown in Tables 4.1 and 4.2, the convergence of the adaptive CC(P;Q) calculations toward the parent CCSDT energetics is very fast. This includes the more challenging multireference situations created by the stretched F_2 and F_2^+ molecules and the TS structure of cyclobutadiene, where, T_3 correlations are large, nonperturbative, and difficult to capture, resulting in failures of methods like CCSD(T) and, in the case of the latter two systems, of CR-CC(2,3), as well as the weakly correlated cyclobutadiene R species, which the CCSD(T) and CR-CC(2,3) methods can handle (see Refs. [127, 137]), although not perfectly. The relaxed variant of the adaptive CC(P;Q) algorithm is generally most accurate. For the most demanding cases of the stretched F_2^+ and TS species of cyclobutadiene, where the coupling of the lower-rank T_1 and T_2 clusters with their higher-rank T_3 counterpart is the largest, it reduces the 10.971 and 14.636 millihartree errors relative to CCSDT obtained with CR-CC(2,3) and the similarly large errors obtained with CCSD(T) to a 0.1 millihartree level using as little as 2–3% of all triply excited determinants in the underlying P spaces. With only 2% of all triples in the P space, the difference between the activation energies characterizing the automerization of cyclobutadiene obtained with the relaxed variant of the adaptive CC(P;Q) approach and full CCSDT is less than 0.1 kcal/mol, as opposed to the orders of magnitude larger, 8.653 kcal/mol, errors relative to CCSDT resulting from the CR-CC(2,3) calculations. The rate with which the energies resulting from the adaptive CC(P;Q) calculations based on the relaxed algorithm approach the parent CCSDT energetics, observed in

Tables 4.1 and 4.2, is certainly most encouraging.

As one might anticipate, and as confirmed in Tables 4.1 and 4.2, the unrelaxed variant of the adaptive CC(P;Q) methodology, in which one picks a particular fraction of triples for inclusion in the P space based on their contributions to the CR-CC(2,3) correction to CCSD, is less accurate for the stretched F_2^+ molecule, the TS structure of cyclobutadiene, and the associated barrier height than its relaxed counterpart, which reaches a desired fraction of triply excited determinants in the P space through a sequence of recursively generated subspaces, but the results of the unrelaxed CC(P;Q) calculations, especially given their simplicity and lower computational costs, are excellent too. As shown in Tables 4.1 and 4.2, with only 2–3% of all triply excited determinants in the underlying P spaces, the adaptive CC(P;Q) computations based on the unrelaxed algorithm reduce the 10.971 millihartree, 14.636 millihartree, and 8.653 kcal/mol unsigned errors relative to CCSDT characterizing the CR-CC(2,3) calculations for the $r = 2r_e$ F_2^+ system, the TS species involved in the automerization of cyclobutadiene, and the corresponding barrier height, respectively, to a chemical accuracy (1 millihartree or 1 kcal/mol) level. Compared to the adaptive CC(P;Q) calculations using the relaxed scheme, the convergence rate toward the parent CCSDT energetics characterizing the unrelaxed approach is slower, but the fact that one can obtain such high accuracies with just a few percent of all triples in the underlying P spaces, when the CR-CC(2,3) corrections to CCSD fail or struggle, is encouraging. It is worth noticing that with only 1% of all triply excited determinants in the relevant P spaces, which is the smallest fraction of triples considered in this work, the adaptive CC(P;Q) computations reduce the 10.971 millihartree, 14.636 millihartree, and 8.653 kcal/mol errors obtained with CR-CC(2,3) for the challenging $r = 2r_e$ F_2^+ system, the TS structure of cyclobutadiene, and the activation energy characterizing the automerization of cyclobutadiene relative to CCSDT to 2.173 millihartree, 0.601 millihartree, and 0.412 kcal/mol, respectively. Given the fact that 1% is also the incremental fraction of triples used to enlarge the $\mathscr{H}^{(P)}(k)$ spaces in the relaxed CC(P;Q) calculations reported in this study, the relaxed and unrelaxed CC(P;Q) compu-

tations using the leading 1% of all triply excited determinants identified by the adaptive CC(P;Q) algorithm are equivalent.

While the P spaces containing only 1% of all triply excited determinants may not be rich enough to bring the errors characterizing the adaptive CC(P;Q) calculations for the stretched F_2^+ ion and the TS structure of cyclobutadiene relative to CCSDT to a 0.1 millihartree level, they are sufficient for achieving high accuracies of this type in the adaptive CC(P;Q) computations for the stretched fluorine molecule and the cyclobutadiene R species. As shown in Tables 4.1 and 4.2, the adaptive CC(P;Q) calculations for the $r = 2r_e$ F_2 system and the R structure of cyclobutadiene using only 1% of all triply excited determinants in the underlying P spaces reduce the 4.254 and 0.848 millihartree differences between the CR-CC(2,3) and CCSDT energies to about 60 microhartree. Clearly, these are dramatic improvements, especially given the small computational effort involved. Furthermore, unlike in the $r = 2r_e$ F_2^+ and cyclobutadiene TS systems, the results of the adaptive CC(P;Q) computations for the stretched fluorine molecule and the R species of cyclobutadiene using larger fractions of triples in the corresponding P spaces do not change much when the relaxed algorithm is replaced by its simpler unrelaxed counterpart. We can rationalize these observations as follows. In the case of the stretched F_2 molecule, T_3 correlations are large and nonperturbative, so that one is much better off by using the CR-CC(2,3) triples correction to CCSD instead of perturbative approaches like CCSD(T), but the coupling of T_1 and T_2 clusters with T_3 is not as well pronounced as in the stretched F_2^+ and TS species of cyclobutadiene and can, therefore, be captured by injecting a tiny fraction of the leading triply excited determinants into the CC(P) calculations preceding the determination of the noniterative $\delta_0(P;Q)$ correction. As a result, the incorporation of a larger fraction of triples in the underlying P space and the iterative construction of the P space via the relaxed algorithm are not necessary to obtain high accuracies in the adaptive CC(P;Q) computations for the stretched F_2. In the case of the R species of cyclobutadiene, T_3 correlations are relatively small, perturbative, and largely decoupled from those captured by T_1 and T_2 clusters and

their powers entering the CC wave function ansatz, making the use of larger fractions of triples in the P space and the iteratively constructed $\mathscr{H}^{(P)}(k)$ spaces less necessary.

Table 4.1 Convergence of the energies resulting from the relaxed (Rel.) and unrelaxed (Unrel.) variants of the adaptive CC(P;Q) approach and the underlying CC(P) computations toward CCSDT for the F_2 and F_2^+ molecules described by the cc-pVTZ basis set, in which the F–F bond lengths r were fixed at $2r_e$, with r_e representing the relevant equilibrium geometries (2.66816 bohr for F_2 and 2.49822 bohr for F_2^+). The %T values are the percentages of the triply excited determinants of the $S_z = 0$ $A_{1g}(D_{2h})$, for F_2, and $S_z = 1/2$ $B_{3g}(D_{2h})$, for F_2^+, symmetries identified by the adaptive CC(P;Q) algorithm and included, alongside all singles and doubles, in the respective P spaces. The Q spaces used in computing the CC(P;Q) corrections were defined as the remaining triples not included in the associated P spaces. In increasing the numbers of triply excited determinants in the P spaces employed in the relaxed calculations, a 1% growth rate was assumed throughout. In all post-RHF (F_2) and post-ROHF (F_2^+) calculations, the two lowest core orbitals were kept frozen. Adapted from Ref. [137].

	F_2				F_2^+			
% T	CC(P) Unrel.[a]	CC(P;Q) Unrel.[a]	CC(P) Rel.[a]	CC(P;Q) Rel.[a]	CC(P) Unrel.[a]	CC(P;Q) Unrel.[a]	CC(P) Rel.[a]	CC(P;Q) Rel.[a]
0	62.819[b]	4.254[c]	62.819[b]	4.254[c]	76.291[b]	10.971[c]	76.291[b]	10.971[c]
1[d]	3.076	0.063	3.076	0.063	6.071	2.173	6.071	2.173
2	2.103	0.089	2.052	0.057	4.061	1.560	2.599	0.191
3	1.586	0.104	1.539	0.070	2.970	1.146	1.707	−0.026
4	1.243	0.098	1.212	0.071	2.100	0.684	1.330	−0.007
5	1.009	0.105	0.985	0.080	1.680	0.549	1.077	0.010
100	−199.238344[e]				−198.606409[e]			

[a]The CC(P) and CC(P;Q) energies are reported as errors relative to CCSDT in millihartree.
[b]Equivalent to CCSD.
[c]Equivalent to CR-CC(2,3).
[d]For %T = 1, the CC(P) and CC(P;Q) energies obtained in the relaxed and unrelaxed calculations are identical.
[e]Total CCSDT energy in hartree.

Table 4.2 Convergence of the energies resulting from the relaxed (Rel.) and unrelaxed (Unrel.) variants of the adaptive CC($P;Q$) approach and the underlying CC(P) computations toward CCSDT for the reactant (R) and transition-state (TS) structures involved in the automerization of cyclobutadiene, as described by the cc-pVDZ basis set, optimized in the MR-AQCC calculations reported in Ref. [262], along with the corresponding barrier heights. The %T values are the percentages of the $S_z = 0$ triply excited determinants of the $A_{1g}(D_{2h})$ symmetry identified by the adaptive CC($P;Q$) algorithm and included, alongside all singles and doubles, in the respective P spaces. In analogy to F_2 and F_2^+, the Q spaces adopted in computing the CC($P;Q$) corrections consisted of the triply excited determinants not included in the associated P spaces and in increasing the numbers of triples in the P spaces used in the relaxed calculations, a 1% growth rate was assumed throughout. In all post-RHF calculations, the four lowest core orbitals were kept frozen. Adapted from Ref. [137].

	R				TS				Barrier Height	
%T	CC(P)	CC($P;Q$)	CC(P)	CC($P;Q$)	CC(P)	CC($P;Q$)	CC(P)	CC($P;Q$)	CC($P;Q$)	CC($P;Q$)
	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[b]	Rel.[b]
0	26.827[c]	0.848[d]	26.827[c]	0.848[d]	47.979[c]	14.636[d]	47.979[c]	14.636[d]	8.653[d]	8.653[d]
1[e]	12.622	−0.055	12.622	−0.055	12.622	0.601	12.622	0.601	0.412	0.412
2	10.143	−0.013	10.121	−0.030	10.180	0.598	9.250	−0.169	0.384	−0.087
3	8.610	0.016	8.588	−0.002	8.643	0.561	7.821	−0.111	0.343	−0.068
4	7.501	0.037	7.482	0.022	7.571	0.568	6.827	−0.040	0.334	−0.038
5	6.637	0.050	6.620	0.035	6.722	0.559	6.052	0.010	0.320	−0.016
100	−154.244157[f]				−154.232002[f]				7.627[g]	

[a]The CC(P) and CC($P;Q$) energies of the R and TS species are reported as errors relative to CCSDT in millihartree.
[b]The CC($P;Q$) values of the barrier height are reported as errors relative to CCSDT in kcal/mol.
[c]Equivalent to CCSD.
[d]Equivalent to CR-CC(2,3).
[e]For %T = 1, the CC(P) and CC($P;Q$) energies obtained in the relaxed and unrelaxed calculations are identical.
[f]Total CCSDT energy in hartree.
[g]The CCSDT barrier height in kcal/mol.

4.2.2 Singlet–Triplet Gaps in Organic Biradicals

Our next application of the adaptive CC(P;Q) approach demonstrates its usefulness in converging the CCSDT data characterizing the total energies of the lowest-lying singlet and triplet states, along with the corresponding singlet–triplet gaps, in three organic biradicals, namely, cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane. The nature of the lowest-lying singlet and triplet states in biradicals, along with the associated $\Delta E_{\rm S-T}$ values (throughout this work, we define $\Delta E_{\rm S-T}$ as $E_{\rm S} - E_{\rm T}$, where $E_{\rm S}$ and $E_{\rm T}$ are the electronic energies of the relevant singlet and triplet states), is a fascinating problem from both an experimental and theoretical perspective, as biradicals often play key roles as reaction intermediates in thermal and photochemical pathways [271–277] as well as functional materials used in molecular magnets [278–280], battery electrodes [281], and organic photovoltaics [282–285]. The accurate computational determination of the singlet–triplet gaps in biradicals remains, however, a difficult task because it requires balancing the strong nondynamical many-electron correlation effects characterizing the low-spin singlet states with the predominantly dynamical correlations associated with their high-spin triplet counterparts [128, 136, 149, 265, 286–298].

In this section, we explore the performance of both the relaxed and unrelaxed adaptive CC(P;Q) approaches in recovering the CCSDT-level energetics of the lowest singlet and triplet states of cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane, as well as the corresponding singlet–triplet gaps. Our choice of molecules was motivated by the recent semi-stochastic CC(P;Q) study in Ref. [136], where it was demonstrated that the CIQMC-based CC(P;Q) approach was rather effective in converging the results for these three systems obtained with CCSDT calculations using fractions of the computational effort. Therefore, it is interesting to examine whether the newly developed adaptive CC(P;Q) approach is capable of doing the same, but in a fully automated fashion that is free from the active-orbital-based or external QMC and selected CI considerations adopted in the earlier CC(P;Q) work. Following Ref. [136], our CC calculations for the lowest-energy singlet

states of cyclobutadiene, cyclopentadienyl cation, and trimethylenemethan were carried out using the respective $S_z = 0$ RHF determinants as the Fermi vacuum, whereas their lowest triplet counterparts were determined using CC computations employing the corresponding high-spin $S_z = 1$ ROHF determinants as the reference function. As in the previous study discussed in Section 4.2.1, we are especially interested in how effective the adaptive CC(P;Q) calculations are in improving the CR-CC(2,3) energetics and converging the parent CCSDT data as a function of the numbers of triply excited determinants included in the underlying P spaces. The results of our adaptive CC(P;Q) computations for the lowest-lying singlet and triplet states of cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane, along with the corresponding vertical (cyclobutadiene and cyclopentadienyl cation) or adiabatic (trimethylenemethane) singlet–triplet gap values are reported in Tables 4.3–4.5. As in our adaptive CC(P;Q) calculations for F_2, F_2^+, and cyclobutadiene discussed in the previous section, we employed the two-body approximation to evaluate the triples correction $\delta_0(P;Q)$ and the lowest-energy orbitals correlating with the 1s shells of carbon atoms were frozen in all post-RHF/ROHF steps.

We begin our discussion by pointing out that the CR-CC(2,3) approach, represented by the % T = 1 data in Tables 4.3–4.4, is generally unable to provide accurate $\Delta E_{\mathrm{S-T}}$ values for the three biradicals considered in this study. As shown in Tables 4.3–4.4, the errors in the singlet–triplet gaps for cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane relative to full CCSDT obtained using CR-CC(2,3) are 9.222, 3.765, and 8.128 kcal/mol, respectively. The error is most pronounced in the case of cyclobutadiene, where the CR-CC(2,3) calculations are unable to correctly identify its lowest-energy singlet state as the ground state (the determination of the ground state of cyclobutadiene has a long history in quantum chemistry; see, *e.g.*, Refs. [299–303] for some of the earliest *ab initio* calculations predicting the $X\ ^1B_{1g}$ ground state, which were confirmed in many subsequent theoretical studies, such as those reported in Refs. [136, 254, 262, 288, 293–295, 297, 298, 304–306]). For all three biradicals considered here, we should note that the failure in CR-CC(2,3) is al-

most entirely due to its inability to accurately describe the relevant open-shell singlet state. Indeed, when we compare the errors relative to CCSDT obtained using CR-CC(2,3) for the lowest-lying singlet states of cyclobutadiene ($X\,^1\mathrm{B}_{1g}$), cyclopentadienyl cation ($A\,^1\mathrm{E}'_2$), and trimethylenemethane ($B\,^1\mathrm{A}_1$), which are 14.636, 6.245, and 13.371 millihartree, respectively, to the much smaller -0.060, 0.245, and 0.418 millihartree errors characterizing their $A\,^3\mathrm{A}_{2g}$, $X\,^3\mathrm{A}'_2$, and $X\,^3\mathrm{A}'_2$ counterparts, it becomes clear that the large errors in the singlet–triplet gap values obtained using CR-CC(2,3) arise from the inability of CR-CC(2,3) to properly balance the description of the lowest high-spin triplet states, which are of largely SR character and dominated by dynamical correlations, with the stronger MR correlations characterizing their two-reference open-shell singlet counterparts. Failure of CR-CC(2,3) [126, 127, 136] and, as demonstrated, for example, in Refs. [127, 255], of other noniterative triples corrections to CCSD, including CCSD(T) [127, 255], to accurately describe the lowest-energy singlet states of biradicals is, in significant part, a consequence of the inability of all such methods to capture the coupling of the lower-rank T_1 and T_2 clusters with T_3, which can become large, nonperturbative, and significant enough to substantially alter T_1 and T_2 amplitudes compared to their CCSD values. This means that to improve the description of the lowest-energy singlet state in cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane, and obtain accurate $\Delta E_{\mathrm{S-T}}$ values, one should relax the T_1 and T_2 clusters, adjusting them to the presence of T_3 correlations, prior to determining noniterative triples corrections. The adaptive CC(P;Q) methodology allows us to do this in a computationally efficient manner, avoiding expensive CCSDT iterations, by incorporating the leading triply excited determinants identified with the help of the $\delta_0(P;Q)$ corrections, Eqs. (4.2) and (4.5), into the underlying P spaces and correcting the resulting CC(P) energies for the remaining T_3 effects using the CC(P;Q) moment expansions.

When we move on to examining the adaptive CC(P;Q) calculations that include the leading 1 % of triply excited determinants in the underlying P spaces, reported in Table 4.3–4.5, which for the sake of brevity in this discussion is abbreviated as CC(P;Q)[%T =

1], we immediately see the benefits of relaxing T_1 and T_2 clusters in the presence of leading T_3 correlations. The large 14.636, 6.245, and 13.371 millihartree differences between the CR-CC(2,3) and CCSDT energetics for the $X\ ^1\mathrm{B}_{1g}$ state of cyclobutadiene, the $A\ ^1\mathrm{E}'_2$ state of cyclopentadienyl cation, and the $B\ ^1\mathrm{A}_1$ state of trimethylenemethane are respectively reduced to 0.601, 1.322, and 0.171 millihartree when the CR-CC(2,3) calculations are replaced by their adaptive CC(P;Q)[%T = 1] counterparts. At the same time, the errors relative to CCSDT characterizing the adaptive CC(P;Q)[%T = 1] calculations for the corresponding lowest-energy triplet states, which are −0.187 millihartree for the $A\ ^3\mathrm{A}_{2g}$ state of cyclobutadiene, 0.031 millihartree for the $X\ ^3\mathrm{A}'_2$ state of cyclopentadienyl cation, and 0.253 millihartree for the $X\ ^3\mathrm{A}'_2$ state of trimethylenemethane, are roughly the same (cyclobutadiene) or improved (cyclopentadienyl cation and trimethylenemethane) compared to the already very accurate results obtained using CR-CC(2,3). As a result of relaxing the T_1 and T_2 clusters in the presence of higher-order correlation effects prior to determining the CC(P;Q) triples corrections in the calculations for the lowest-energy singlet states of cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane, the corresponding $\Delta E_{\mathrm{S-T}}$ values obtained with the adaptive CC(P;Q)[%T = 1] approach are substantially more accurate than their counterparts obtained in the CR-CC(2,3) calculations. In particular, the adaptive CC(P;Q)[%T = 1] calculations reduce the 9.222, 3.765, and 8.128 kcal/mol errors relative to full CCSDT characterizing the singlet–triplet gaps of cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane computed with CR-CC(2,3) to 0.495, 0.810, and −0.052 kcal/mol, respectively. In a similar fashion to our study testing the adaptive CC(P;Q) approach on bond breaking in F_2 and F_2^+ and automerization of cyclobutadiene, we observe that the least expensive CC(P;Q) calculations considered in our study, which use just 1% of triply excited determinants in the underlying P spaces, are capable of reducing the errors obtained using CR-CC(2,3) and providing a chemically accurate description (within 1 millihartree or 1 kcal/mol) relative to the parent CCSDT approach.

Naturally, we can study the convergence of the adaptive CC(P;Q) calculations toward

the full CCSDT limit as the fractions of triply excited determinants entering the corresponding P spaces are increased. As shown in Tables 4.3–4.5, the relaxed variant of the adaptive CC(P;Q) calculations using 2–5% of triply excited determinants in the underlying P spaces (similarly abbreviated as CC(P;Q)[%T = n], for n = 2–5) demonstrably converge the parent CCSDT results. Focusing on the description of the $\Delta E_{\mathrm{S-T}}$ values, the 0.495 kcal/mol error in the singlet–triplet gap of cyclobutadiene relative to CCSDT obtained with the adaptive CC(P;Q)[%T = 1] calculations are reduced to −0.046, −0.040, −0.015, and 0.003 kcal/mol when the fractions of triply excited determinants entering the relaxed adaptive CC(P;Q) are increased to 2% , 3%, 4%, and 5%, respectively. In a similar fashion, the 0.810 kcal/mol error in the $\Delta E_{\mathrm{S-T}}$ value relative to CCSDT obtained with the adaptive CC(P;Q)[%T = 1] calculations are reduced to −0.076, −0.128, −0.111, and 0.096 kcal/mol in the relaxed adaptive CC(P;Q) computations employing 2% , 3%, 4%, and 5% of triply excited determinants, respectively. In the case of trimethylenemethane, the singlet–triplet gap obtained using the initial adaptive CC(P;Q)[%T = 1] calculations was already very accurate, but it is reassuring to observe that the −0.052 kcal/mol error relative to CCSDT obtained using the CC(P;Q) calculations with 1% of triply excited determinants remains to within (−0.133)–(−0.067) kcal/mol of the CCSDT data when the fractions of triply excited determinants are increased from 1% to 2–5%.

It is interesting to note, however, that the relaxed adaptive CC(P;Q)[%T = n] calculations with $n > 1$ are more accurate than their unrelaxed counterparts when describing the lowest-lying singlet states of cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane. Indeed, when we compare the differences between the total energies of the lowest-lying singlet states of cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane obtained in the adaptive CC(P;Q)[%T = 1] calculations and CCSDT with their counterparts corresponding to the unrelaxed adaptive CC(P;Q)[%T = 5] approach, we see that the unrelaxed selection scheme is not quite as effective as the relaxed variant in presence of stronger MR correlations. For example, the 0.601 millihartree error relative to CCSDT ob-

tained in the adaptive CC(P;Q)[%T = 1] calculations for the $X\,^1\mathrm{B}_{1g}$ state of cyclobutadiene becomes 0.559 millihartree when the unrelaxed CC(P;Q)[%T = 5] approach is employed. Similarly minor changes are observed for the $A\,^1\mathrm{E}'_2$ state cyclopentadienyl cation, in which the 1.322 millihartree differences between the energy obtained in the adaptive CC(P;Q)[%T = 1] and CCSDT become 0.860 millihartree when 5% of triply excited determinants are used in the underlying P spaces defining the unrelaxed CC(P;Q) methodology. In the case of trimethylenemethane, the 0.171 millihartree error relative to CCSDT resulting from the initial adaptive CC(P;Q)[%T = 1] computations characterizing the $B\,^1\mathrm{A}_1$ state actually increases slightly to 0.245 millihartree when the CC(P;Q)[%T = 1] approach is replaced with the unrelaxed CC(P;Q)[%T = 5] method. Thus, we can conclude that the relaxed adaptive CC(P;Q) model is more effective than its unrelaxed counterpart in converging the full CCSDT description of the lowest-energy singlet and triplet state, along with the gap between them, for the cyclobutadiene, cyclopentadienyl cation, and trimethylenemethane biradicals, especially if a precision beyond standard chemical accuracy is desired. Although the better performance of the relaxed adaptive CC(P;Q) algorithm is to be expected (and also demonstrated in our calculations discussed in Section 4.2.1), given that the relaxed adaptive CC(P;Q) calculations *de facto* determine the lists of leading higher–than–doubly excited determinants entering the CC(P;Q) steps using a more accurate description of many-electron correlation effects than their unrelaxed counterparts, one should keep in mind that this improvement in accuracy comes with higher computational costs, since, as mentioned in Section 4.1, the relaxed adaptive CC(P;Q) model requires an iterative construction of multiple $\mathscr{H}^{(P)}(k)$ spaces. Therefore, while both the unrelaxed and relaxed adaptive CC(P;Q) models are highly efficient and viable tools for studying the electronic structure of organic biradicals, the specific choice of algorithm should be carefully selected by considering the balance of computational costs with the desired accuracy levels for the application of interest.

Table 4.3 Convergence of the energies resulting from the relaxed (Rel.) and unrelaxed (Unrel.) variants of the adaptive CC(P;Q) approach and the underlying CC(P) computations for the $X\,^1\mathrm{B}_{1g}$ and $A\,^3\mathrm{A}_{2g}$ states of cyclobutadiene, as described by the cc-pVDZ basis set, and of the corresponding singlet-triplet gap, toward their parent CCSDT values. All calculations were performed at the D_{4h}-symmetric transition-state geometry on the $X\,^1\mathrm{B}_{1g}$ potential optimized in the MR-AQCC calculations reported in Ref. [262]. The %T values are the percentages of the triply excited determinants of the $S_z = 0$ $\mathrm{A}_{1g}(D_{2h})$ symmetry, for the $X\,^1\mathrm{B}_{1g}$ state, and the $S_z = 1$ $\mathrm{B}_{1g}(D_{2h})$, for the $A\,^3\mathrm{A}_{2g}$ state, identified by the adaptive CC(P;Q) algorithm and included, alongside all singles and doubles, in the respective P spaces. In computing the CC(P;Q) corrections , the Q space consisted of the triply excited determinants not included in the associated P spaces and in increasing the numbers of triples in the P spaces used in the relaxed calculations, a 1% growth rate was assumed throughout. In all post-RHF/ROHF calculations, the four lowest core orbitals were kept frozen.

	$X\,^1\mathrm{B}_{1g}$				$A\,^3\mathrm{A}_{2g}$				$X\,^1\mathrm{B}_{1g} - A\,^3\mathrm{A}_{2g}$	
%T	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P;Q)	CC(P;Q)
	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[b]	Rel.[b]
0	47.979[c]	14.636[d]	47.979[c]	14.636[d]	23.884[c]	−0.060[d]	23.884[c]	−0.060[d]	9.222[d]	9.222[d]
1	12.622	0.601	12.622	0.601	11.565	−0.187	11.565	−0.187	0.495	0.495
2	10.180	0.598	0.250	−0.169	9.290	−0.097	9.292	−0.095	0.436	−0.046
3	8.643	0.561	7.821	−0.111	7.865	−0.048	7.867	−0.046	0.383	−0.040
4	7.571	0.568	6.827	−0.040	6.833	−0.018	6.835	−0.016	0.368	−0.015
5	6.722	0.559	6.052	0.010	6.029	0.003	6.031	0.005	0.349	0.003
100	−154.232002[f]				−154.224380[f]				−4.783[g]	

[a]The CC(P) and CC(P;Q) energies of the $X\,^1\mathrm{B}_{1g}$ and $A\,^3\mathrm{A}_{2g}$ states of cyclobutadiene are reported as errors relative to CCSDT in millihartree.

[b]The CC(P;Q) values of the singlet–triplet gap are reported as errors relative to CCSDT in kcal/mol.

[c]Equivalent to CCSD.

[d]Equivalent to CR-CC(2,3).

[e]For %T = 1, the CC(P) and CC(P;Q) energies obtained in the relaxed and unrelaxed calculations are identical.

[f]Total CCSDT energy in hartree.

[g]The CCSDT singlet–triplet gap in kcal/mol.

Table 4.4 Convergence of the energies resulting from the relaxed (Rel.) and unrelaxed (Unrel.) variants of the adaptive CC(P;Q) approach and the underlying CC(P) computations for the $X\,^3\mathrm{A}_2'$ and $A\,^1\mathrm{E}_2'$ states of cyclopentadienyl cation, as described by the cc-pVDZ basis set, and of the corresponding singlet–triplet gap toward their parent CCSDT values. All calculations were performed at the D_{5h}-symmetric geometry of the $X\,^3\mathrm{A}_2'$ state optimized using the unrestricted CCSD/cc-pVDZ approach in Ref. [288]. The %T values are the percentages of the triply excited determinants of the $S_z = 1$ $\mathrm{B}_1(C_{2v})$ symmetry, for the $X\,^3\mathrm{A}_2'$ state, and the $S_z = 0$ $\mathrm{A}_1(C_{2v})$, for the $A\,^1\mathrm{E}_2'$ state, identified by the adaptive CC(P;Q) algorithm and included, alongside all singles and doubles, in the respective P spaces. In computing the CC(P;Q) corrections consisted of the triply excited determinants not included in the associated P spaces and in increasing the numbers of triples in the P spaces used in the relaxed calculations, a 1% growth rate was assumed throughout. In all post-RHF/ROHF calculations, the five lowest core orbitals were kept frozen.

	$X\,^3\mathrm{A}_2'$				$A\,^1\mathrm{E}_2'$				$A\,^1\mathrm{E}_2' - X\,^3\mathrm{A}_2'$	
%T	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P;Q)	CC(P;Q)
	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[b]	Rel.[b]
0	28.840[c]	0.245[d]	28.840[c]	0.245[d]	38.572[c]	6.245[d]	38.572[c]	6.245[d]	3.765[d]	3.765[d]
1[e]	12.614	0.031	12.614	0.031	14.613	1.322	14.613	1.322	0.810	0.810
2	10.026	0.084	10.026	0.086	11.630	1.085	10.418	−0.035	0.628	−0.076
3	8.436	0.107	8.437	0.110	9.856	0.984	8.704	−0.094	0.550	−0.128
4	7.296	0.119	7.298	0.123	8.597	0.918	7.562	−0.054	0.501	−0.111
5	7.461	0.145	6.420	0.129	7.619	0.860	6.681	−0.023	0.448	−0.096
100	−192.615924[f]				−192.589235[f]				16.747[g]	

[a]The CC(P) and CC(P;Q) energies of the $X\,^3\mathrm{A}_2'$ and $A\,^1\mathrm{E}_2'$ states of cyclopentadienyl cation are reported as errors relative to CCSDT in millihartree.

[b]The CC(P;Q) values of the singlet–triplet gap are reported as errors relative to CCSDT in kcal/mol.

[c]Equivalent to CCSD.

[d]Equivalent to CR-CC(2,3).

[e]For %T = 1, the CC(P) and CC(P;Q) energies obtained in the relaxed and unrelaxed calculations are identical.

[f]Total CCSDT energy in hartree.

[g]The CCSDT singlet–triplet gap in kcal/mol.

Table 4.5 Convergence of the energies resulting from the relaxed (Rel.) and unrelaxed (Unrel.) variants of the adaptive CC(P;Q) approach and the underlying CC(P) computations for the $X\,^3\mathrm{A}_2'$ and $B\,^1\mathrm{A}_1$ states of trimethylenemethane, as described by the cc-pVDZ basis set, and of the corresponding adiabatic singlet–triplet gap toward their parent CCSDT values. The D_{3h}- and C_{2v}-symmetric geometries of the $X\,^3\mathrm{A}_2'$ and $B\,^1\mathrm{A}_1$ states, respectively, optimized in the SF-DFT/6-31G(d) calculations, were taken from Ref. [286]. The %T values are the percentages of the triply excited determinants of the $S_z = 1$ $\mathrm{B}_1(C_{2v})$ symmetry, for the $X\,^3\mathrm{A}_2'$ state, and the $S_z = 0$ $\mathrm{A}_1(C_{2v})$, for the $A\,^1\mathrm{E}_2'$ state, identified by the adaptive CC(P;Q) algorithm and included, alongside all singles and doubles, in the respective P spaces. In computing the CC(P;Q) corrections consisted of the triply excited determinants not included in the associated P spaces and in increasing the numbers of triples in the P spaces used in the relaxed calculations, a 1% growth rate was assumed throughout. In all post-RHF/ROHF calculations, the four lowest core orbitals were kept frozen.

	$X\,^3\mathrm{A}_2'$				$B\,^1\mathrm{A}_1$				$B\,^1\mathrm{A}_1 - X\,^3\mathrm{A}_2'$	
%T	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P;Q)	CC(P;Q)
	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[a]	Unrel.[a]	Rel.[a]	Rel.[a]	Unrel.[b]	Rel.[b]
0	19.202[c]	0.418[d]	19.202[c]	0.418[d]	58.052[c]	13.371[d]	58.052[c]	13.371[d]	8.128[d]	8.128[d]
1[e]	9.630	0.253	9.630	0.253	9.487	0.171	9.487	0.171	−0.052	−0.052
2	7.729	0.241	7.729	0.242	7.506	0.207	7.255	0.031	−0.021	−0.133
3	6.536	0.227	6.536	0.229	6.309	0.229	6.080	0.064	0.001	−0.104
4	5.672	0.215	5.673	0.218	5.456	0.240	5.250	0.085	0.015	−0.083
5	5.000	0.204	5.001	0.206	4.801	0.245	4.616	0.099	0.026	−0.067
100	−155.466242[f]				−155.431596[f]				21.740[g]	

[a]The CC(P) and CC(P;Q) energies of the $X\,^3\mathrm{A}_2'$ and $B\,^1\mathrm{A}_1$ states of trimethylenemethane are reported as errors relative to CCSDT in millihartree.

[b]The CC(P;Q) values of the singlet–triplet gap are reported as errors relative to CCSDT in kcal/mol.

[c]Equivalent to CCSD.

[d]Equivalent to CR-CC(2,3).

[e]For %T = 1, the CC(P) and CC(P;Q) energies obtained in the relaxed and unrelaxed calculations are identical.

[f]Total CCSDT energy in hartree.

[g]The CCSDT singlet–triplet gap in kcal/mol.

4.2.3 The Cope Rearrangement of Bullvalene

As a final example to demonstrate the capabilities of our adaptive CC(P;Q) approach in converging the CCSDT energetics, we applied it to study the degenerate Cope rearrangement in bullvalene ($C_{10}H_{10}$), depicted in Figure 4.2. First predicted by Doering and Roth

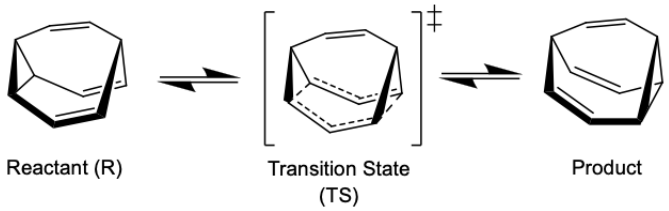

Figure 4.2 Schematic of the Cope rearrangement reaction in bullvalene.

[307] and synthesized by Schröder in 1963 [308], bullvalene is a well-known fluxional molecule in which rapid carbon-carbon bond rearrangements occur on the timescale of typical NMR measurements [309, 310]. The Cope rearrangement of bullvalene constantly shuffles nuclei between equivalent reactant (R) and product conformers by passing through a typical aromatic pericylic transition-state species (TS) [311]. As one might expect, the Cope rearrangement in bullvalene has been the subject of numerous experimental [312–319] as well as theoretical [320–325] studies. In particular, gas-phase NMR measurements of bullvalene estimate the activation barrier for isomerization to be 13.8 ± 0.2 kcal/mol [312], while standard DFT calculations [320] and thermochemical protocols like CBS-QB3 [323] predict barrier heights of 11.3 and 12.5 kcal/mol, respectively. As we can expect after examining the automerization of cyclobutadiene, lower-level quantum chemical methods that cannot capture and balance the correlation effects characterizing the R and TS species are likely to struggle.

In this regard, we highlight two recent theoretical studies, which served as the inspiration for the present work, examining the R and TS structures of bullvalene with the goal of treating many-electron correlations at a higher level compared to previous attempts. In the most recent study, Lesiuk performed SVD-CCSDT+ calculations augmented with perturbative quadruples corrections to obtain a barrier height of 14.6 kcal/mol [325], which is in line with an earlier prediction of 14.9 kcal/mol made by Karton [324] using a combination of

CCSD(T) calculations with post-CCSD(T) correlation effects estimated using a composite scheme. While the barrier heights resulting from these two studies differ from the gas-phase NMR measurements by ∼1 kcal/mol, both Lesiuk and Karton demonstrated that the vast majority of the correlation effects are due to the T_3 clusters (cf., also, the results of the CCSD calculations provided in Table 4.6). Given that both the SVD-CCSDT method [326] and the CCSD(T) approach rely on perturbative treatments of T_3 [while this is obvious in the case of CCSD(T), one should not forget that the SVD-CCSDT approach describes the three-body clusters within a truncated basis resulting from singular value decomposition of a low-order CC3-like approximation to T_3], it is interesting to examine whether nonperturbative CC(P;Q) calculations can help resolve this discrepancy.

Our adaptive CC(P;Q) calculations for the R and TS structures of bullvalene, along with the corresponding barrier heights, are presented in Table 4.6. The C_1-symmetric geometries characterizing the R and TS species were taken from Ref. [325], where they were optimized using the B3LYP-D3/pc-2 approach. All calculations for bullvalene employed the cc-pVTZ [10] basis set and the lowest-energy orbitals correlating with the 1s shells of carbon atoms were frozen in all post-RHF steps. Using the cc-pVTZ basis, the bullvalene molecule contains 50 correlated electrons and 440 orbitals and represents the largest system studied using the adaptive CC(P;Q) method to date. In order to efficiently handle a single-particle basis containing over 400 orbitals, we employed the Cholesky decompositions (CD) of the electron repulsion integrals, which allows us to factorize the two-electron integrals as a product of three-center integrals in an automated fashion. A key feature of the CD approach is that it approximates the two-electron integrals with arbitrary precision according to a user-specified energy tolerance parameter. In particular, the two-electron integrals resulting from the CD are identical to their canonical (*i.e.*, non-CD-based) counterparts up to n digits when a tolerance of 10^{-n} hartree is used. Our CD-based CC(P;Q) calculations used a tolerance of 10^{-7} hartree in the CD step in order to ensure that the energies resulting from our calculations remain accurate relative to their canonical counterparts through at least seven

digits. Another modification that we had to make due to the larger size of this system was in the choice of the growth rate used to increase the numbers of triply excited determinants in the P spaces entering our adaptive CC(P;Q) calculations. In our previous studies, we adopted a default growth rate of 1%, but in the bullvalene molecule using the cc-pVTZ basis set, a 1% increment results in the inclusion of roughly 5×10^9 triply excited determinants in each iteration, which leads to excessive computation and storage costs. To mitigate this issue, we used an alternative strategy, in which we set the growth increment for %T equal to the ratio between the total size of the CCSD problem (*e.g.*, the total number of all singly and doubly excited determinants) and the total size of the triples manifold. For the bullvalene molecule in the cc-pVTZ basis set, there are 20,250 singly excited determinants, 151,601,625 doubly excited determinants, and 547,548,876,000 triply excited determinants of the $S_z = 0$ symmetry, resulting in our chosen growth increment of 0.03%.

Before discussing the results of our adaptive CC(P;Q) calculations for bullvalene summarized in Table 4.6, we mention that, unlike in previous examples, we could not perform the parent CCSDT calculations. Therefore, in evaluating the results of our CC(P;Q) computations for the isomerization barrier height, we treat the experimentally determined value of 13.8 $\pm$ 0.2 kcal/mol as a reliable benchmark result. We begin our discussion by pointing out that the CCSD calculations shown in Table 4.6 are entirely inadequate for predicting the barrier for Cope rearrangement in bullvalene. As a result of neglecting T_3 correlations, the 18.4 kcal/mol barrier predicted with CCSD lies far from the experimental range of 13.6–14.0 kcal/mol. On the other hand, the CR-CC(2,3) calculations predict a much improved barrier height of 14.5 kcal/mol, which still differs from experiment by 0.5–0.9 kcal/mol. Interestingly, the 14.5 kcal/mol barrier height predicted by the CR-CC(2,3) approach is very similar to the corresponding 14.6 and 14.9 kcal/mol reaction barriers obtained using the SVD-CCSDT approach with perturbative quadruples corrections and composite CCSD(T)/CCSDT(Q) calculations of Refs. [325] and [324], respectively. However, as emphasized in the previous examples, the CR-CC(2,3) method may result in errors due to ignoring the coupling be-

tween the T_1, T_2, and T_3 clusters. Indeed, when the CR-CC(2,3) calculations are replaced by CC(P;Q)[%T = 0.03], we reduce the 14.6 kcal/mol barrier predicted using CR-CC(2,3) by 0.8 kcal/mol to obtain 13.8 kcal/mol with CC(P;Q)[%T = 0.03], which matches perfectly with the experimentally determined mean barrier height. The agreement between the barrier height resulting from the CC(P;Q)[%T = 0.03] calculations and experiment is certainly impressive, especially considering that inclusion of just 0.03% of triply excited determinants in the underlying P spaces makes our CC(P;Q) calculations computationally practical (far more practical than full CCSDT, which could not be carried out). By examining the total energies reported in Table 4.6, we find that the TS structure is more sensitive to the inclusion of coupling between the T_1, T_2, and T_3 clusters compared to the R species. When we move from CR-CC(2,3) to CC(P;Q)[%T = 0.03], the total energy of the TS changes by 2.917 millihartree compared to 1.778 millihartree in the R species. Furthermore, as we increase the fractions of triply excited determinants entering the P spaces used in our adaptive CC(P;Q) calculations from 0.03% to 0.09%, we find that the resulting barrier heights remain at 13.8 kcal/mol. To lend further confidence to the reliability of our results, we also see that the differences in total energies for the R and TS species obtained in the CC(P;Q)[%T = 0.06] and CC(P;Q)[%T = 0.09] calculations are just 0.186 millihartree, in the case of the R, and 0.154 millihartree, in the case of the TS. Thus, we have reason to believe that we have converged the parent CCSDT limit to within a fraction of a millihartree.

Although the nearly perfect agreement between our adaptive CC(P;Q) calculations and the gas-phase NMR data is most encouraging, we cannot claim that our calculations are of strictly higher quality than those reported in Refs. [324, 325]. While we have treated the T_3 effects more completely at the CCSDT level, which is an improvement over the previous two studies, we have not made any attempt to describe the T_4 correlations, which were approximately included in Refs. [324, 325]. Future work aimed at treating T_4 clusters within the adaptive CC(P;Q) calculations, either fully as in CCSDTQ or approximately as in CR-CC(3,4), will be required to verify the accuracy of the present results.

Table 4.6 Total energies, in hartree, of the R and TS structures involved in the Cope rearrangement of bullvalene obtained using the adaptive CC(P;Q) calculations, alongside the corresponding barrier heights, in kcal/mol. The geometries of the C_1-symmetric R and TS structures of bullvalene were taken from Ref. [325], where they were optimized using the B3LYP-D3/pc-2 approach. The lowest orbitals correlating with the 1s shells of carbon atoms were frozen in all post-RHF steps and the cc-pVTZ basis set was employed throughout. In carrying out the CD of the two-electron integrals, we used a tolerance parameter 10^{-7} hartree to ensure that the energies resulting from our CC(P;Q) calculations are identical to their canonical counterparts through at least seven digits.

	R		TS		Barrier Height[a,b]	
%T[c]	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)
0	−386.182025[d]	−386.262955[e]	−386.151090[d]	−386.238226[e]	18.4[d]	14.5[e]
0.03	−386.201340	−386.264733	−386.184958	−386.241143	9.3	13.8
0.06	−386.206525	−386.265043	−386.191629	−386.241416	8.3	13.8
0.09	−386.210004	−386.265229	−386.195949	−386.241570	7.8	13.8

[a]The barrier heights, in kcal/mol, corresponding to the difference between the energies of the TS and R structures of bullvalene. In reporting barrier heights, we have included the zero-point vibrational energy correction of −1.0 kcal/mol obtained in Ref. [325] (cf., also, Ref. [324]).

[b]The experimentally determined barrier height for the Cope rearrangement of bullvalene in the gas phase is 13.8 ± 0.2 kcal/mol [312].

[c]The %T values reflect the numbers of $S_z = 0$ triply excited determinants identified by the relaxed variant of the adaptive CC(P;Q) algorithm that are included, in addition to all singly and doubly excited determinants, in the underlying P spaces. The increment of 0.03%, used to increase the numbers of triply excited determinants entering the P spaces, corresponds to the ratio between the sum of the total number of $S_z = 0$ singly and doubly excited determinants to the total number of $S_z = 0$ triply excited determinants.

[d]Equivalent to CCSD.

[e]Equivalent to CR-CC(2,3).

4.2.4 Ground- and Excited-State Potential Surfaces for O–H Bond-Breaking in Water

In the previous three sections, we have carried out studies designed to test the effectiveness of the adaptive CC(P;Q) algorithm in converging the ground-state CCSDT energetics for chemically relevant examples involving bond breaking, isomerization reactions, and single–triplet gaps in biradicals. Our last example, which is inspired by Refs. [161, 327], focuses on converging the ground- and excited-state PESs of the water molecule along the O–H bond-breaking coordinate corresponding to the $H_2O \rightarrow H + OH$ dissociation obtained with CCSDT/EOMCCSDT. As we will show below, the CCSDT and EOMCCSDT water potentials are nearly exact, closely matching the FCI data reported in Refs. [161, 327], but they cannot be accurately approximated using any of the more recently developed triples corrections to EOMCCSD [142, 148, 151, 152, 159, 161, 328, 329], including the left-eigenstate CR-EOMCC(2,3) approach [126, 148, 151] or its rigorously size-intensive δ-CR-EOMCC(2,3) modification [152, 159] that rely on the moment energy expansions [126, 139, 140, 142, 143, 146–148, 151], especially when examining stretched regions of certain excited-state water potentials when T_3 and $R_{\mu,3}$ correlations become larger, nonperturbative, and strongly coupled to their lower-rank T_1, T_2, $R_{\mu,1}$, and $R_{\mu,2}$ counterparts [126, 127, 133].

In this section, we compare the ground- and excited-state PESs of the water molecule along the O–H bond-breaking coordinate obtained using three different forms of CC(P;Q), including (i) the conventional CR-CC(2,3)/CR-EOMCC(2,3) triples corrections to CCSD/EOMCCSD, (ii) the active-orbital-based CC(t;3) approach based on correcting the underlying CCSDt/EOMCCSDt calculations for missing T_3 and $R_{\mu,3}$ effects, and (iii) the adaptive CC(P;Q) approach employing the relaxed algorithm, relative to the parent CCSDT/EOMCCSDT surfaces. We show that both the CC(t;3) and adaptive CC(P;Q) calculations provide results that closely approximate the full CCSDT/EOMCCSDT data for the PESs of water using significantly reduced computational costs while improving the CR-CC(2,3) and CR-EOMCC(2,3) energetics in stretched regions of the O–H bond-breaking potentials. In

addition, we show that the adaptive CC(P;Q) calculations produce PESs of water that are nearly identical to their active-orbital-based CC(t;3) counterparts, while using much smaller fractions of triply excited determinants in the underlying P spaces.

The nuclear geometries needed to construct the ground- and excited-state PESs of the water molecule along the $H_2O \rightarrow H + OH$ dissociation path, obtained by considering 11 values of the O–H bond separation R_{OH} ranging from 1.3 to 4.4 bohr, with the remaining O–H bond length and H–O–H bond angle optimized using the CCSD/cc-pVTZ method, were taken from Ref. [327]. As pointed out in Refs. [161, 327], the resulting dissociation pathway on the ground-state PES has several desirable characteristics. For example, the equilibrium values of R_{OH} and H–O–H bond angle resulting from the CCSD/cc-pVTZ geometry optimization, of 1.809 bohr and 103.9 degree, respectively, are in good agreement with experiment and, as R_{OH} increases, the second O–H bond length approaches its equilibrium value in the ground-state OH radical and the angular potential flattens. For each water structure corresponding to a particular value of R_{OH}, we carried out the CR-CC(2,3)/CR-EOMCC(2,3), active-orbital-based CC(P;Q) [i.e., CC(t;3)], adaptive CC(P;Q), and parent CCSDT/EOMCCSDT calculations for the four lowest $A'(C_s)$-symmetric singlet states, which include the ground state $X\,^1A'$ and $n\,^1A'$ excited states with $n = 1$–3, three lowest $A'(C_s)$-symmetric triplet states (denoted as $n\,^3A'$, $n = 1$–3), two lowest $A''(C_s)$-symmetric singlet states (denoted as $n\,^1A''$, $n = 1, 2$), and three lowest triplet states of the $A''(C_s)$ symmetry (denoted as $n\,^3A''$, $n = 1$–3). To enable comparisons of our CCSDT/EOMCCSDT data with the FCI and MRCC energetics obtained in Refs. [327] (FCI and MRCC) and [161] (FCI), all calculations reported in this work were performed with the TZ basis set of Ref. [327], used in Ref. [161] as well. The RHF determinant, used in our CC/EOMCC calculations as a reference, consisted of four $a'(C_s)$-symmetric and one $a''(C_s)$-symmetric orbitals. The lowest-energy orbital correlating with 1s shell of oxygen was frozen in post-RHF steps. To define the CC(t;3) and underlying CCSDt/EOMCCSDt calculations, the three highest occupied and two lowest unoccupied orbitals in the RHF reference correlating with 2p shell of oxygen

and 1s shells of hydrogens were treated as active. Two of the three active occupied and both active unoccupied orbitals were of the $a'(C_s)$ symmetry. The third active occupied orbital was $a''(C_s)$-symmetric. The adaptive CC(P;Q) calculations employed the relaxed algorithm, discussed in Section 4.1, in which we assumed a 1% growth rate in the numbers of triply excited determinants entering the underlying P spaces. The percentages of triply excited determinants characterizing our adaptive CC(P), EOMCC(P), and CC(P;Q) calculations were defined as fractions of the $S_z = 0$ triples of the $A'(C_s)$ ($^1A'$ and $^3A'$ states) or $A''(C_s)$ ($^1A''$ and $^3A''$ states) symmetry identified with the adaptive CC(P;Q) algorithm. The CCSD, EOMCCSD, CR-CC(2,3), CR-EOMCC(2,3), CCSDt, EOMCCSDt, CC(t;3), CCSDT, and EOMCCSDT computations were performed using our in-house CC/EOMCC codes interfaced with the RHF and integral transformation routines in GAMESS [227]. The adaptive CC(P;Q) calculations were executed using our recently developed CCpy package available on GitHub [228] that can efficiently handle the potentially spotty subsets of triply excited determinants entering the underlying CC(P) and EOMCC(P) computations, which may not form continuous manifolds labeled by occupied and unoccupied orbitals from the respective ranges of indices, to achieve the desired speedups compared to CCSDT/EOMCCSDT (see Section 4.3 and Appendix C for further information).

The results of our CC/EOMCC calculations are shown in Tables 4.7–4.9, Fig. 4.3, and Tables D.1–D.13 in Appendix D. Table 4.7 summarizes the mean unsigned error (MUE) and nonparallelity error (NPE) values characterizing the CCSDT and EOMCCSDT potential cuts of water computed in this work relative to their FCI counterparts obtained in Refs. [161, 327]. The MUE and NPE values characterizing the CCSD/EOMCCSD, CR-CC(2,3)/CR-EOMCC(2,3), adaptive CC(P)/EOMCC(P) and CC(P;Q), CCSDt/EOMCCSDt, and CC(t;3) PESs relative to their CCSDT/EOMCCSDT parents are reported in Tables 4.8 and 4.9, respectively. The ground- and excited-state PES cuts of water corresponding to the $H_2O \rightarrow H + OH$ dissociation channels that correlate with the $X\ ^2\Pi$ ground state and the lowest-energy $^2\Sigma^+$ and $^2\Sigma^-$ states of OH obtained with CCSD/EOMCCSD,

CR-CC(2,3)/CR-EOMCC(2,3), adaptive CC(P)/EOMCC(P) and CC(P;Q), CCSDt/EOM-CCSDt, CC(t;3), and CCSDT/EOMCCSDT are shown in Fig. 4.3. Tables D.2–D.13 in Appendix D provide total electronic energies of all the calculated states.

Accuracy of the CCSDT and EOMCCSDT Approaches

Given that all three variants of the CC(P;Q) methodology tested in this work aim at converging or accurately approximating the CCSDT/EOMCCSDT energetics, it is important to assess how well the CCSDT and EOMCCSDT approaches perform in describing the ground- and excited-state PESs of water along the O–H bond-breaking coordinate relative to their exact, FCI, counterparts. As shown in Table 4.7, the CCSDT method provides a highly accurate description of the $X\,^1\mathrm{A}'$ state, with the MUE and NPE values relative to the ground-state FCI potential being as small as 0.63 and 1.14 millihartree, respectively. EOMCCSDT is similarly accurate in describing the excited-state potentials, with only three – out of 11 examined in this study – characterized by the MUE and NPE values exceeding 1 and 3 millihartree, respectively, but even in this case, which includes the $2\,^1\mathrm{A}'$, $3\,^1\mathrm{A}'$, and $3\,^3\mathrm{A}'$ states, the MUE and NPE values, of 1.28 and 3.19 millihartree for the $2\,^1\mathrm{A}'$ PES, 1.53 and 4.48 millihartree for the $3\,^1\mathrm{A}'$ PES, and 1.41 and 5.50 millihartree for the $3\,^3\mathrm{A}'$ PES, remain rather small, especially when we realize that these three states are located hundreds of millihartrees above the ground state (see Tables D.1 in Appendix D). Most importantly, the EOMCCSDT approach accurately approximates the FCI water potentials in the highly stretched $R_{\mathrm{OH}} \geq 2.4$ bohr region, where all excited states of interest in this study acquire a substantial MR character that EOMCCSD cannot capture. When compared with the MRCC results reported in Ref. [327], we observe that the SR CCSDT and EOMCCSDT methods can also rival, or even outperform, the state-of-the-art MR treatments. For example, for the $1\,^3\mathrm{A}'$, $1\,^1\mathrm{A}'$, $2\,^3\mathrm{A}''$, $2\,^1\mathrm{A}''$, $3\,^1\mathrm{A}'$, and $3\,^3\mathrm{A}'$ states, the EOMCCSDT NPE values relative to FCI are lower – sometimes substantially – than those resulting from the nR-GMS-SU-CCSD and (N,M)-CCSD calculations. For the $X^1\mathrm{A}'$, $1^1\mathrm{A}''$, $1^3\mathrm{A}''$, $2^3\mathrm{A}'$, and $2^1\mathrm{A}'$ potentials, the CCSDT and EOMCCSDT NPE values relative to FCI are within $\sim$1 millihartree from the best MRCC

results reported in Ref. [327]. It is clear from these comparisons that we can treat the ground- and excited-state potentials of water along the O–H bond-breaking coordinate obtained with CCSDT and EOMCCSDT as highly accurate benchmarks for evaluating performance of the different CC(P;Q) approaches explored in this work.

Table 4.7 The MUE and NPE values, in millihartree, relative to full CI characterizing the ground-state CCSDT and excited-state EOMCCSDT potentials of the water molecule, as described by the TZ basis set of Ref. [327], along the O–H bond-breaking coordinate corresponding to the $H_2O \rightarrow H + OH$ dissociation. Adapted from Ref. [138].

	$X\,^1A'$	$1\,^1A''$	$1\,^3A'$	$1\,^3A''$	$1\,^1A'$	$2\,^3A'$	$2\,^3A''$	$2\,^1A'$	$2\,^1A''$	$3\,^3A''$	$3\,^1A'$	$3\,^3A'$
MUE	0.63	0.78	0.28	0.90	0.92	0.93	0.43	1.28	0.59	0.63	1.53	1.41
NPE	1.14	1.85	0.83	2.14	2.13	2.33	1.35	3.19	0.57	2.67	4.48	5.50

Performance of Different CC(P;Q) Approaches Relative to CCSDT and EOM-CCSDT

We begin by discussing the first variant of the CC(P;Q) methodology of interest in the present study corresponding to the CR-CC(2,3) and CR-EOMCC(2,3) triples corrections to CCSD and EOMCCSD (also examined in Ref. [161]). In the vicinity of the equilibrium geometry on the ground-state PES, all CC/EOMCC approaches, including CR-CC(2,3)/CR-EOMCC(2,3), and even CCSD/EOMCCSD, perform well. Indeed, the errors relative to CCSDT in the $X\,^1A'$ potential computed with CCSD in the R_{OH} = 1.3–2.0 bohr region range between 2.771 and 3.562 millihartree. CR-CC(2,3) reduces them to ∼0.2–0.3 millihartree. For the excited-state potentials in the R_{OH} = 1.3–2.0 bohr region, the largest error obtained with EOMCCSD relative to EOMCCSDT, which occurs when the high-lying $3\,^1A'$ state is considered, is 3.392 millihartree. The largest error characterizing CR-EOMCC(2,3) in the same region, encountered when the $3\,^3A''$ PES is examined, is 1.660 millihartree. For nearly all excited states considered in this work, the differences between the CR-EOMCC(2,3) and EOMCCSDT potentials in the R_{OH} = 1.3–2.0 bohr region are about 1 millihartree, making them virtually identical around the equilibrium geometry. The CR-EOMCC(2,3) PESs are also more parallel to their EOMCCSDT counterparts than the corresponding EOMCCSD surfaces.

The situation changes when we move toward larger $R_{\rm OH}$ values, where all electronic states of water considered in this study develop a strong MR character, resulting in failures of the CCSD and EOMCCSD methods, especially when the $X\,^1\mathrm{A}'$, $1\,^1\mathrm{A}''$, $1\,^3\mathrm{A}''$, $1\,^1\mathrm{A}'$, $2\,^3\mathrm{A}'$, $2\,^3\mathrm{A}''$, $2\,^1\mathrm{A}'$, $3\,^3\mathrm{A}''$, $3\,^1\mathrm{A}'$, and $3\,^3\mathrm{A}'$ potentials are examined. These failures become particularly dramatic for the $2\,^3\mathrm{A}''$, $2\,^1\mathrm{A}'$, $3\,^3\mathrm{A}''$, and $3\,^1\mathrm{A}'$ states, where errors relative to EOMCCSDT resulting from the EOMCCSD calculations in the $R_{\rm OH} = 2.8$–4.4 bohr region become as large as 28.470, 32.540, 65.442, and 32.035 millihartree, respectively, although no state considered in our computations is accurately described when T_3 and $R_{\mu,3}$ correlations are neglected and $R_{\rm OH} > 2.4$ bohr. The CR-CC(2,3) and CR-EOMCC(2,3) triples corrections reduce the excessive errors characterizing the $X\,^1\mathrm{A}'$, $1\,^1\mathrm{A}''$, $1\,^3\mathrm{A}''$, $1\,^1\mathrm{A}'$, $2\,^3\mathrm{A}'$, $2\,^3\mathrm{A}''$, $2\,^1\mathrm{A}'$, $3\,^3\mathrm{A}''$, $3\,^1\mathrm{A}'$, and $3\,^3\mathrm{A}'$ potentials obtained with CCSD/EOMCCSD at larger values of $R_{\rm OH}$ and the associated MUEs and NPEs shown in Tables 4.8 and 4.9, which are 6.154–28.262 and 8.453–64.663 millihartree, respectively, in the CCSD/EOMCCSD case and 0.543–6.075 and 0.692–42.170 millihartree when the CR-CC(2,3)/CR-EOMCC(2,3) data are examined, but several problems remain, especially when dealing with the $2\,^3\mathrm{A}''$ and $3\,^3\mathrm{A}''$ states. In the former case, the CR-EOMCC(2,3) approach produces large, 11.891–13.976 millihartree, errors relative to EOMCCSDT in the $R_{\rm OH} = 4.0$–4.4 bohr region, resulting in the qualitatively incorrect asymptotic behavior of the CR-EOMCC(2,3) $2\,^3\mathrm{A}''$ potential and the NPE of 13.043 millihartree. The CR-EOMCC(2,3) PES for the $3\,^3\mathrm{A}''$ state presents an even more distressing situation, with a massive, 37.018 millihartree, error relative to EOMCCSDT at $R_{\rm OH} = 2.8$ bohr associated with a bump in the $3\,^3\mathrm{A}''$ potential seen in Fig. 4.3(b) and the even larger NPE value of 42.170 millihartree. These failures of CR-EOMCC(2,3) are consistent with our earlier observations [126, 127, 130, 133, 135, 137] that none of the triples corrections to CCSD/EOMCCSD can provide accurate results when T_3 and $R_{\mu,3}$ correlations become substantial and strongly coupled to their lower-rank counterparts, as is the case when examining the $2\,^3\mathrm{A}''$ and $3\,^3\mathrm{A}''$ potentials at larger $R_{\rm OH}$ values.

The above challenges encountered in the CR-EOMCC(2,3) calculations can be addressed

by turning to the CC(t;3) method. As explained in the Introduction and in Chapter 2, CC(t;3) is a CC(P;Q) approach in which energies obtained in the CCSDt/EOMCCSDt calculations that incorporate the leading triply excited determinants identified with the help of active orbitals in the P spaces used in the iterative CC(P)/EOMCC(P) steps are corrected for the missing, mostly dynamical, T_3 and $R_{\mu,3}$ correlations using Eq. (2.33). The numerical results in Tables 4.8 and 4.9, with further details provided by Tables D.2–D.13 in Appendix D, and the potential curves shown in Fig. 4.3(h) demonstrate that the CC(t;3) method readily addresses the shortcomings of the CR-EOMCC(2,3) approach and its ground-state CR-CC(2,3) counterpart. This becomes particularly clear when examining the $2\,^3\mathrm{A}''$ and $3\,^3\mathrm{A}''$ potentials that are poorly described by CR-EOMCC(2,3). For example, the CC(t;3) calculations reduce the large, 11.891–13.976 millihartree, errors relative to EOMCCSDT obtained with CR-EOMCC(2,3) for the $2^3\mathrm{A}''$ PES in the $R_{\mathrm{OH}} = 4.0$–4.4 bohr region to 0.481–0.625 millihartree. The massive error of 37.018 millihartree produced by CR-EOMCC(2,3) for the $3\,^3\mathrm{A}''$ state at $R_{\mathrm{OH}} = 2.8$ bohr is reduced in the CC(t;3) calculations by more than two orders of magnitude, to 0.133 millihartree, eliminating the unphysical bump in the CR-EOMCC(2,3) potential for this state in the $R_{\mathrm{OH}} = 2.4$–3.2 bohr region altogether. The major improvements offered by CC(t;3) in describing the $2\,^3\mathrm{A}''$ and $3\,^3\mathrm{A}''$ PESs are also reflected in the corresponding NPE values relative to EOMCCSDT, which decrease from 13.043 and 42.170 millihartree obtained with CR-EOMCC(2,3) to the minuscule 0.255 and 0.608 millihartree, respectively. In general, the CC(P;Q)-based CC(t;3) method provides a highly accurate description of the ground and excited states of water along the O–H bond-breaking coordinate, greatly improving the CR-CC(2,3)/CR-EOMCC(2,3) energetics and closely matching the nearly exact CCSDT/EOMCCSDT potentials at small fractions of the computational costs associated with CCSDT/EOMCCSDT. This is manifested by the tiny MUE and NPE values relative to CCSDT/EOMCCSDT characterizing the CC(t;3) calculations for the 12 electronic states of water reported in this work, which range from 0.166 to 0.951 millihartree for MUEs and 0.110 to 0.608 millihartree for NPEs (see Tables

4.8 and 4.9). By capturing the missing T_3 and $R_{\mu,3}$ correlations, the CC(t;3) corrections are also very effective in improving the underlying CCSDt/EOMCCSDt potentials, especially in reducing the MUE values relative to CCSDT/EOMCCSDT characterizing the CCSDt and EOMCCSDt calculations, from 1.477–2.420 millihartree in CCSDt/EOMCCSDt to fractions of a millihartree in CC(t;3).

We conclude by commenting on the results of our adaptive CC(P;Q) calculations, which are based on the same basic principles as those employed in the CC(t;3) considerations, but do not rely on active orbitals to obtain accurate results, allowing us to converge the high-level CCSDT and EOMCCSDT energetics in an entirely black-box fashion. As shown in Tables 4.8 and 4.9, panels (d) and (f) of Fig. 4.3, and Tables D.2–D.13 in Appendix D, the adaptive CC(P;Q) approach is remarkably effective in accurately approximating the CCSDT/EOMCCSDT water potentials. Already with a tiny 1% of triply excited determinants in the underlying P spaces, the adaptive CC(P;Q) calculations offer major improvements in the CR-CC(2,3)/CR-EOMCC(2,3) data. They reduce the MUE and NPE values of 0.543 and 0.692 millihartree, respectively, relative to CCSDT, obtained with CR-CC(2,3) for the ground-state potential, to less than 0.3 millihartree. The improvements in the description of the 11 excited states of water considered in this work offered by the adaptive CC(P;Q) approach using the leading 1% of triply excited determinants in the underlying P spaces, which, as in Section 4.2.2, will be abbreviated as CC(P;Q)[%T = 1], are similarly impressive, especially when we realize that in the case of the C_s-symmetric water structures and the TZ basis set used in our calculations, 1% of triply excited determinants amounts to only about 300 T_3 and $R_{\mu,3}$ amplitudes, as opposed to 31,832 $A'(C_s)$-symmetric and 32,232 $A''(C_s)$-symmetric $S_z = 0$ triples used by full CCSDT/EOMCCSDT. These improvements can be best seen when comparing the $2\,^3A''$ and $3\,^3A''$ potentials obtained with CR-EOMCC(2,3) with their CC(P;Q)[%T = 1] counterparts. In the case of the $2\,^3A''$ PES, the MUE and NPE values of 5.033 and 13.043 millihartree relative to EOMCCSDT resulting from the CR-EOMCC(2,3) computations are reduced to 1.177 and 1.081 millihartree, respectively, when

the CC(P;Q)[%T = 1] approach is employed, helping to alleviate the inaccurate behavior of CR-EOMCC(2,3) in the asymptotic part of the $2\,^3\mathrm{A}''$ potential. The poor description of the $3\,^3\mathrm{A}''$ PES by the CR-EOMCC(2,3) method becomes much more reasonable in the CC(P;Q)[%T = 1] calculations as well. The 37.018 millihartree error relative to EOM-CCSDT obtained with CR-EOMCC(2,3) at R_{OH} = 2.8 bohr reduces to 6.740 millihartree when the adaptive CC(P;Q)[%T = 1] approach is employed. The MUE and NPE values characterizing the CR-EOMCC(2,3) $3\,^3\mathrm{A}''$ potential, of 6.075 and 42.170 millihartree, respectively, decrease in the CC(P;Q)[%T = 1] calculations to 1.949 and 5.933 millihartree and, as shown in Fig. 4.3 (d), the unphysical bump in the CR-EOMCC(2,3) PES for this state in the R_{OH} = 2.4–3.2 bohr region disappears. The generally small MUE and NPE values characterizing the CC(P;Q)[%T = 1] calculations for the 12 PESs of water considered in this study, relative to their CCSDT/EOMCCSDT parents, which range from 0.275 to 1.949 millihartree for MUEs and 0.216 to 5.933 millihartree for NPEs, are certainly encouraging. As shown in Tables 4.8 and 4.9, Fig. 4.3(f), and Tables D.2–D.13 in Appendix D, the situation gets even better when the fraction of triply excited determinants included in the P spaces defining the adaptive CC(P;Q) calculations increases to 2%. The MUEs and NPEs relative to CCSDT/EOMCCSDT characterizing the resulting CC(P;Q)[%T = 2] computations for the 12 potential cuts of water examined in this work reduce to 0.197–0.993 and 0.133–1.572 millihartree, respectively. In the case of the $2\,^3\mathrm{A}''$ and $3\,^3\mathrm{A}''$ potentials that cause major troubles to the CR-EOMCC(2,3) method, the already small MUEs obtained in the CC(P;Q)[%T = 1] calculations, of 1.177 and 1.949 millihartree, decrease to 0.737 and 0.993 millihartree, respectively, when the fraction of triply excited determinants incorporated in the underlying P spaces grows from 1% to 2%. The corresponding NPE values decrease from 1.081 and 5.933 millihartree in the CC(P;Q)[%T = 1] case to 0.237 and 1.572 millihartree, when the CC(P;Q)[%T = 2] approach is employed. For some electronic potentials of water, the NPEs characterizing the adaptive CC(P;Q)[%T = 2] computations are slightly larger than those obtained with the CC(t;3) approach, but the overall performance of the adaptive

CC(P;Q)[%T = 2] and CC(t;3) methods is very similar. This is promising for the future applications of the adaptive CC(P;Q) framework since typical CC(t;3) computations, in addition to requiring the user to select active orbitals, employ considerably larger fractions of triply excited determinants in the iterative steps of the underlying CC(P;Q) algorithm than their black-box adaptive CC(P;Q) counterparts. For example, the CC(t;3) calculations reported in this work used about 38% of all triples in the iterative CCSDt/EOMCCSDt steps, as opposed to only 2% used in the adaptive CC(P;Q) computations.

Table 4.8 The MUE values, in millihartree, relative to CCSDT/EOMCCSDT characterizing the ground- and excited-state potential cuts of the water molecule, as described by the TZ basis set of Ref. [327], along the O–H bond-breaking coordinate corresponding to the $H_2O \rightarrow H + OH$ dissociation obtained with the different CC(P)/EOMCC(P) and CC(P;Q) approaches. Adapted from Ref. [138].

					%T = 1[e]		%T = 2[f]	
State	CCSD[a]	CR(2,3)[b]	CCSDt[c]	CC(t;3)[d]	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)
$X\,^1\mathrm{A}'$	6.154	0.543	2.047	0.166	2.176	0.275	1.616	0.197
$1\,^1\mathrm{A}''$	7.666	1.517	1.683	0.599	3.159	0.755	2.271	0.622
$1\,^3\mathrm{A}'$	2.819	1.137	1.741	0.951	2.061	0.793	1.532	0.616
$1\,^3\mathrm{A}''$	7.693	1.080	1.658	0.627	3.237	0.854	2.339	0.665
$1\,^1\mathrm{A}'$	10.103	1.544	1.602	0.589	3.578	0.590	2.420	0.494
$2\,^3\mathrm{A}'$	8.846	1.225	1.682	0.646	3.379	0.665	2.322	0.530
$2\,^3\mathrm{A}''$	9.258	5.033	1.565	0.590	5.384	1.177	3.532	0.737
$2\,^1\mathrm{A}'$	14.460	2.988	2.276	0.718	3.230	0.782	2.359	0.618
$2\,^1\mathrm{A}''$	2.337	1.750	1.477	0.553	3.461	0.870	2.535	0.579
$3\,^3\mathrm{A}''$	28.262	6.075	2.420	0.542	8.111	1.949	4.973	0.993
$3\,^1\mathrm{A}'$	13.547	2.410	1.688	0.804	5.151	1.152	3.120	0.643
$3\,^3\mathrm{A}'$	8.305	2.481	1.656	0.854	3.794	1.048	2.671	0.676

[a] CCSD for the ground state and EOMCCSD for excited states.
[b] CR-CC(2,3) for the ground state and CR-EOMCC(2,3) for excited states.
[c] CCSDt/EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.
[d] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.
[e] CC(P)/EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.
[f] CC(P)/EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

Table 4.9 The NPE values, in millihartree, relative to CCSDT/EOMCCSDT characterizing the ground- and excited-state potential cuts of the water molecule, as described by the TZ basis set of Ref. [327], along the O–H bond-breaking coordinate corresponding to the $H_2O \rightarrow H + OH$ dissociation obtained with the different CC(P)/EOMCC(P) and CC(P;Q) approaches. Adapted from Ref. [138].

					%T = 1[e]		%T = 2[f]	
State	CCSD[a]	CR(2,3)[b]	CCSDt[c]	CC(t;3)[d]	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)
$X\,^1\mathrm{A}'$	8.453	0.692	0.687	0.110	1.072	0.216	0.801	0.159
$1\,^1\mathrm{A}''$	16.093	4.592	0.446	0.275	1.569	0.433	1.304	0.198
$1\,^3\mathrm{A}'$	5.004	0.442	0.452	0.492	0.801	0.593	1.005	0.281
$1\,^3\mathrm{A}''$	17.445	3.303	0.447	0.240	2.070	0.430	1.360	0.133
$1\,^1\mathrm{A}'$	21.043	4.637	0.387	0.244	2.692	0.313	1.380	0.168
$2\,^3\mathrm{A}'$	20.127	3.712	0.826	0.257	2.264	0.246	1.233	0.218
$2\,^3\mathrm{A}''$	29.017	13.043	0.572	0.255	6.852	1.081	2.761	0.237
$2\,^1\mathrm{A}'$	33.260	8.330	2.231	0.271	2.216	0.761	1.434	0.508
$2\,^1\mathrm{A}''$	7.971	3.983	0.601	0.263	3.129	0.611	0.837	0.636
$3\,^3\mathrm{A}''$	64.663	42.170	3.601	0.608	23.376	5.933	10.679	1.572
$3\,^1\mathrm{A}'$	30.846	4.757	0.463	0.603	6.019	2.207	2.862	0.993
$3\,^3\mathrm{A}'$	21.562	5.670	0.712	0.392	3.149	1.574	1.413	0.582

[a] CCSD for the ground state and EOMCCSD for excited states.
[b] CR-CC(2,3) for the ground state and CR-EOMCC(2,3) for excited states.
[c] CCSDt/EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.
[d] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.
[e] CC(P)/EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.
[f] CC(P)/EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

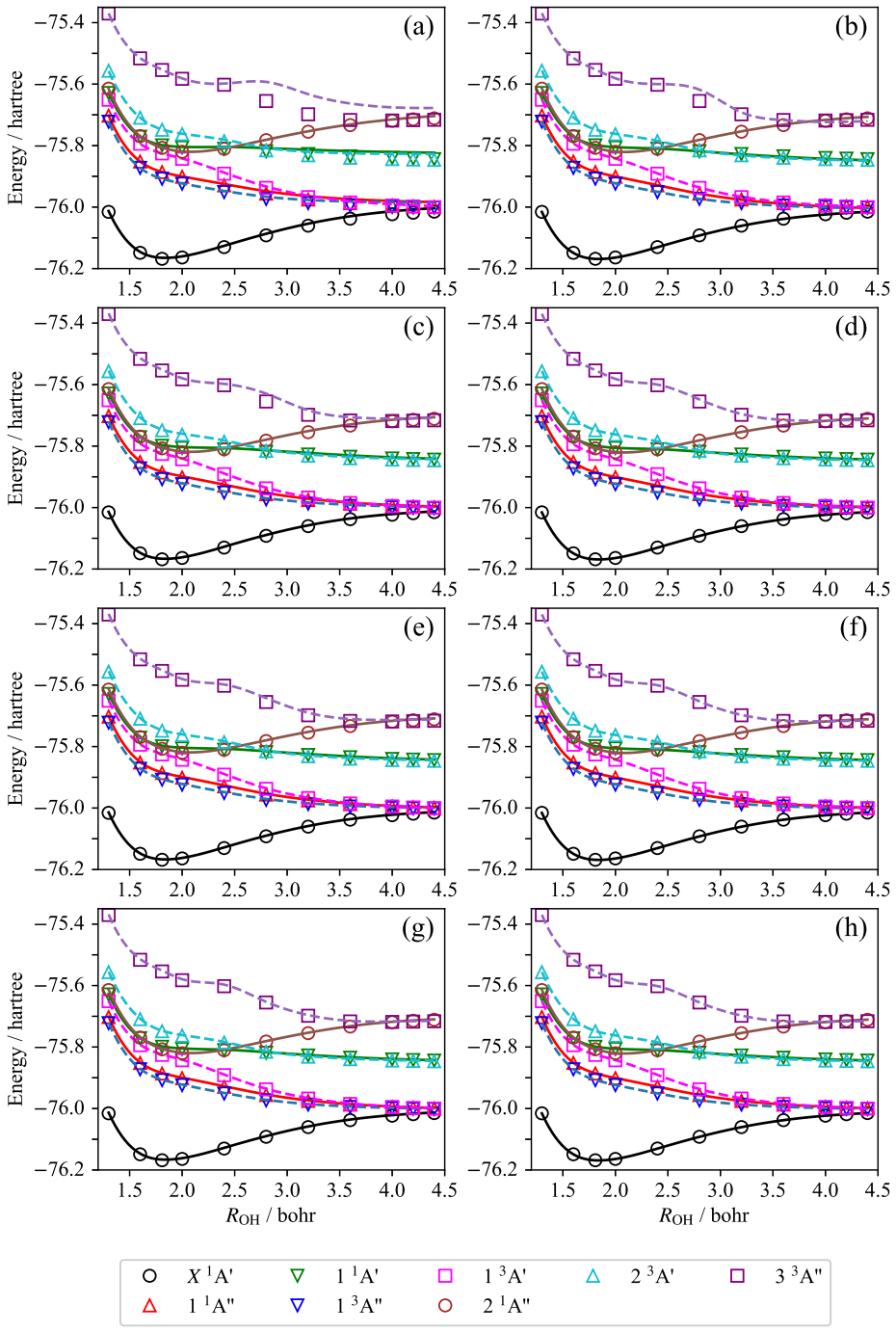

Figure 4.3 A comparison of the potential cuts of water, plotted as functions of R_{OH} and corresponding to the $H_2O \rightarrow H + OH$ dissociation channels that correlate with the $X\ ^2\Pi$ ground state and the lowest-energy $^2\Sigma^+$ and $^2\Sigma^-$ states of OH, obtained with (a) CCSD/EOMCCSD, (b) CR-CC(2,3)/CR-EOMCC(2,3), (c) CC(P)/EOMCC(P)[%T = 1], (d) CC(P;Q)[%T = 1], (e) CC(P)/EOMCC(P)[%T = 2], (f) CC(P;Q)[%T = 2], (g) CCSDt/EOMCCSDt, and (h) CC(t;3) with their full CCSDT/EOMCCSDT counterparts. The splined CCSD/EOMCCSD, CR-CC(2,3)/CR-EOMCC(2,3), adaptive CC(P)/EOMCC(P) and CC(P;Q), CCSDt/EOMCCSDt, and CC(t;3) data are represented by the solid and dashed lines, whereas the open circles, squares, triangles, and inverted triangles correspond to the parent CCSDT/EOMCCSDT energetics.

4.3 Computational Performance of Adaptive CC($P;Q$)

In presenting our numerical results of the adaptive CC($P;Q$) methodology in Section 4.2, we have emphasized that the CC($P;Q$) calculations rapidly converge their CCSDT/EOMCCSDT parents, such that only a tiny fraction of triply excited determinants are needed in the underlying P spaces in order to obtain energetics that are within a fraction of a millihartree of the full CCSDT/EOMCCSDT data. As one might anticipate, the use of small subsets of leading higher–than–doubly excited determinants in the iterative CC(P)/EOMCC(P) steps and CC($P;Q$) corrections provides enormous computational savings compared to the full CCSDT and EOMCCSDT approaches. In this section, we elaborate on these computational benefits offered by the adaptive CC($P;Q$) approach by discussing the single-core CPU timings characterizing our adaptive CC($P;Q$) calculations for the TS structure involved in the automerization of cyclobutadiene, previously examined in Section 4.2.1, and for the series of C_nH_{2n+2} linear alkanes with $n = 1$–8. As in the case of all of the adaptive CC($P;Q$) calculations performed in this work, we used the implementation included in the open-source CCpy package available on GitHub [228], which is based on the Numpy library and augmented by hand-coded Fortran routines in the computationally intensive parts [in particular, we use Fortran for the iterative CC(P)/EOMCC(P) algorithm described in Appendix C]. All reported timings presented in this section correspond to single-core runs on the Precision 7920 workstation from Dell equipped with Intel Xeon Silver 4114 2.2 GHz processor boards. In presenting the CPU timings, we have ignored the computational times associated with the execution of the integral, SCF, and integral transformation routines used to generate the one- and two-electron molecular integrals in the RHF basis, which was carried out using GAMESS, and the integral sorting operations preceding the CC steps performed using CCpy.

4.3.1 Timings for Calculations of Cyclobutadiene

We will begin by examining the computational timings characterizing our adaptive CC($P;Q$) calculations for the TS structure of cyclobutadiene, as described by the cc-pVDZ basis set, examined in Section 4.2.1. To facilitate our discussion, we focus on the unrelaxed CC($P;Q$)

calculations using 1%, 3%, and 5% of the $S_z = 0$ triply excited determinants of the $A_{1g}(D_{2h})$ symmetry, which, as shown in Table 4.2, reduce the 14.636 millihartree errors relative to CCSDT obtained with CR-CC(2,3) to as little as 0.601, 0.561, and 0.559 millihartree, respectively, when the unrelaxed CC($P;Q$) approach is employed. To get useful insights, in addition to the total CPU times, we report the timings associated with the three key stages of the adaptive CC($P;Q$) calculations. In the case of the unrelaxed CC($P;Q$) algorithm considered in Table 4.10, these three key stages include (i) the P space determination, which consists of the initial CR-CC(2,3) run followed by the analysis of the $\delta^{(0)}_{ijk,abc}(0)$ contributions to the resulting triples correction to CCSD, needed to identify a desired fraction of triply excited determinants for inclusion in the subsequent CC(P) computation, (ii) the iterative CC(P) calculation using the P space $\mathscr{H}^{(P)}(1)$ consisting of all singly and doubly excited determinants and a subset of triply excited determinants identified in stage (i), and (iii) the determination of the noniterative $\delta_0(P;Q)$ correction to the CC(P) energy obtained in stage (ii) to capture the remaining T_3 correlations with the help of the complementary Q space $\mathscr{H}^{(Q)}(1)$ using Eq. (4.2) in which we set k at 1.

The illustrative CPU timings included in Table 4.10, combined with the previously discussed energetics compared to CCSDT/EOMCCSDT, clearly demonstrate the enormous benefit offered by the adaptive CC($P;Q$) methodology proposed in this study. Indeed, the adaptive CC($P;Q$) runs using 1–5% of triples in the underlying P spaces accelerate the parent CCSDT computations for the cyclobutadiene/cc-pVDZ system in its TS geometry with only minimal loss of accuracy by factors on the order of 67–81, reducing the 14.636 millihartree errors relative to CCSDT obtained with the CR-CC(2,3) triples corrections to CCSD to small fractions of a millihartree.

Table 4.10 Computational timings characterizing the various CC calculations for the cyclobutadiene/cc-pVDZ system in the transition-state (TS) geometry optimized with MR-AQCC in Ref. [262], including CCSD, CR-CC(2,3), and CCSDT and the unrelaxed variants of the adaptive CC(P;Q) approach, abbreviated as CC(P;Q)[%T = x], where $x = 1, 3$, and 5, which used the leading x percent of triply excited determinants, identified on the basis of the largest $\delta^{(0)}_{ijk,abc}(0)$ contributions to the CR-CC(2,3) triples correction, in addition to all singles and doubles, in constructing the respective P spaces. The Q spaces used to determine the CC(P;Q)[%T = x] corrections to the CC(P)[%T = x] energies consisted of the remaining $(100 - x)$ percent of triples not included in the corresponding P spaces. In all post-RHF calculations, the four lowest core orbitals were kept frozen. Adapted from Ref. [137].

Method	CPU Time[a]			
	P Space Determination[b]	Iterative Steps[c]	Noniterative Steps[d]	Total
CCSD	–	0.5	–	0.5
CR-CC(2,3)	–	0.8	0.3	1.1
CC(P;Q)[%T = 1]	1.9	3.4	0.3	5.6
CC(P;Q)[%T = 3]	1.9	4.0	0.3	6.2
CC(P;Q)[%T = 5]	2.0	4.5	0.3	6.8
CCSDT	–	456	–	456

[a] All reported timings, in CPU minutes, correspond to single-core runs on the Precision 7920 workstation from Dell equipped with Intel Xeon Silver 4114 2.2 GHz processor boards. No advantage of the D_{4h} symmetry of the TS structure of cyclobutadiene or its D_{2h} Abelian subgroup was taken in the post-RHF steps. The computational times associated with the execution of the integral, RHF, and integral transformation and sorting routines preceding the CC steps are ignored.

[b] The timings associated with determining the P spaces for the unrelaxed CC(P;Q)[%T = x] computations, where $x = 1, 3$, and 5, include the times required to execute the corresponding initial CR-CC(2,3) runs plus the times spent on analyzing the $\delta^{(0)}_{ijk,abc}(0)$ contributions to the resulting triples corrections to CCSD needed to identify the top x percent of triply excited determinants. The timings associated with constructing the P spaces for the iterative steps of the CCSD, CR-CC(2,3), and CCSDT calculations are not reported, since in these four cases these spaces are *a priori* known (consisting of all singly and doubly excited determinants in the case of the CCSD and CR-CC(2,3) methods, and of all singly, doubly, and triply excited determinants in the CCSDT case).

[c] In executing the iterative steps of each CC calculation, a convergence threshold of 10^{-7} hartree was assumed. The timings corresponding to the iterative steps include the times required to construct and solve the relevant CC amplitude equations and, in the case of CR-CC(2,3) and CC(P;Q), the times needed to construct and solve the companion left eigenstate problems involving the respective similarity-transformed Hamiltonians [using, in the case of CC(P;Q), the two-body approximation discussed in the text].

[d] In the language of Q spaces adopted by the CC(P;Q) formalism, the computational times required to determine the noniterative triples corrections of CR-CC(2,3) and CC(P;Q) correspond to all triples in the case of CR-CC(2,3) and the remaining $(100 - x)$ percent of triples not included in the relevant P spaces in the case of CC(P;Q)[%T = x] ($x = 1, 3$, and 5).

The computational timings collected in Table 4.10 allow us to draw a few other conclusions. While we will continue working on improving our codes, the results in Table 4.10 already show a reasonably high efficiency of our current implementation of the adaptive CC(P;Q) approach in the CCpy package. This is particularly true in the case of the most challenging CC(P) routines which, in analogy to the previously formulated semi-stochastic [130, 133, 134, 136] and CIPSI-driven [135] CC(P;Q) approaches, are used to solve the CC equations in the P spaces containing spotty subsets of triply excited determinants that do not form continuous manifolds labeled by occupied and unoccupied orbitals from the respective ranges of indices assumed in conventional CC programming. General remarks about the key algorithmic ingredients that must be considered when coding the CC(P;Q) approaches, in order to benefit from the potentially enormous speedups offered by methods in this category when the excitation manifolds used in the iterative CC(P) steps [and their excited-state EOMCC(P) counterparts] are small and spotty, were provided Refs. [130, 133, 134] (cf., also, Ref. [253]), while Appendix C of this document covers the specific strategies adopted in our current implementation of the adaptive CC(P;Q) method, as coded in the CCpy package, in greater detail. One of the most useful consequences of our current implementation of the adaptive CC(P;Q) approach aimed at accurately approximating the full CCSDT energetics, in addition to the major savings in the computational effort compared to CCSDT, is the observation that the CPU timings characterizing the CC(P) calculations preceding the determination of the noniterative $\delta_0(P,Q)$ corrections grow slowly with the fraction of triply excited determinants included in the underlying P spaces (see Table 4.10). This is consistent with the computational cost analysis of the CC(P;Q) methods using small fractions of higher–than–doubly excited cluster and excitation amplitudes in the iterative CC(P) and EOMCC(P) steps in Refs. [130, 133, 134, 253].

Last, but not least, it is interesting to observe in Table 4.10 that the CPU timings characterizing the adaptive CC(P;Q) computations capable of providing the near-CCSDT energetics, while larger than those required by its CR-CC(2,3) counterpart, are not very far

removed from the costs of the CR-CC(2,3) calculations, which are generally less robust and much less accurate, especially when T_3 correlations are larger and nonperturbative and the coupling between the lower-rank T_1 and T_2 and higher-rank T_3 clusters becomes significant. This is mainly because the CPU times needed to complete the iterative CC(P) steps, when the fraction of triples incorporated in the P space is as tiny as 1–3%, which is, as shown in Table 4.2, enough to recover the CCSDT energetics to within fractions of a millihartree, even when CR-CC(2,3) fails or struggles, are only a few times longer than those characterizing CCSD. This is not too surprising. When the fraction of triples in the P space is tiny, the noniterative triples corrections of the adaptive CC($P;Q$) and CR-CC(2,3) calculations have virtually identical costs, since the corresponding Q spaces, which are spanned by either all [CR-CC(2,3)] or nearly all [CC($P;Q$)] triply excited determinants, are very similar. One can, therefore, anticipate that much of the difference between the CPU timings characterizing the CR-CC(2,3) and adaptive CC($P;Q$) runs, especially if we put aside the P space determination steps, is driven by the time spent on the CC(P) iterations, and this is exactly what we see in Table 4.10. The storage requirements characterizing the adaptive CC($P;Q$) calculations are not far removed from those required by their CR-CC(2,3) counterpart either, if the fractions of triples in the underlying P spaces are tiny, since the vectors representing the $T^{(P)} = T_1 + T_2 + T_3^{(P)}$ operators are not much longer than those used to store $T_1 + T_2$. In summary, it is encouraging that with the introduction of the adaptive CC($P;Q$) methodology, which is only a few times more expensive than the CR-CC(2,3) approach, we can accelerate the full CCSDT computations by orders of magnitude with only minimal loss of accuracy, even when T_3 effects are highly nonperturbative, causing CCSD(T) to fail, and even when the coupling among T_1, T_2, and T_3 components of the cluster operator becomes so large that the CR-CC(2,3) correction to CCSD is no longer reliable.

4.3.2 Computational Scalability of Adaptive CC(P;Q) for the Series C_nH_{2n+2} Linear Alkanes

The previous example showcasing the computational timings of our adaptive CC(P;Q) calculations for cyclobutadiene help us appreciate how much we can reduce the CPU effort of the parent CCSDT calculations without sacrificing accuracy. Indeed, the fact that we are able to reduce the 6–7 hours of CPU time required for CCSDT to just 5–7 minutes is most encouraging. In this next example, we examine the computational performance of the adaptive CC(P;Q) approach using the series of C_nH_{2n+2} linear alkanes, with n = 1–8, as described using the cc-pVDZ basis set [10]. For the larger alkanes with $n > 5$, the CCSDT calculations are no longer affordable on our well above-average computational resources. The geometries for the C_nH_{2n+2}, with n = 1–8, molecules were taken from Ref. [326], where they were optimized using B3LYP/cc-pVTZ method. For each molecule C_nH_{2n+2}, the CC calculations involve $6n + 2$ correlated electrons and $23n + 10$ correlated orbitals. In all of our calculations, we have frozen the lowest-lying orbitals correlating with the 1s shells of carbon atoms and the two-body approximation for the ground-state $\delta_0(P;Q)$ correction was employed throughout. We present our timings for the CC(P;Q)[%T = 1] calculations (using an growth increment of 1%, there is no distinction between the relaxed and unrelaxed calculations in this case), which are presented in Figure 4.4 using the same decomposition adopted in Table 4.10 in terms of the preparatory P space determination phase, the iterative CC(P) and left-CCSD-like calculations, and the noniterative CC(P;Q) correction steps

Upon examining Figure 4.4, the benefits offered by the adaptive CC(P;Q)[%T = 1] calculations is evident. In particular, the CPU timings required for the CCSDT calculations increase dramatically as the length of the alkane chain increases. While the CPU timings of 0.3, 20.7, and 133.7 minutes taken by the CCSDT calculations for methane, ethane, and propane are not excessively large, they are still between one and two orders of magnitude larger than the 0.04, 0.5, and 3.77 minutes required by their CC(P;Q)[%T = 1] counterparts. When we examine the butane molecule, the CCSDT calculations consume 1,252.1 minutes

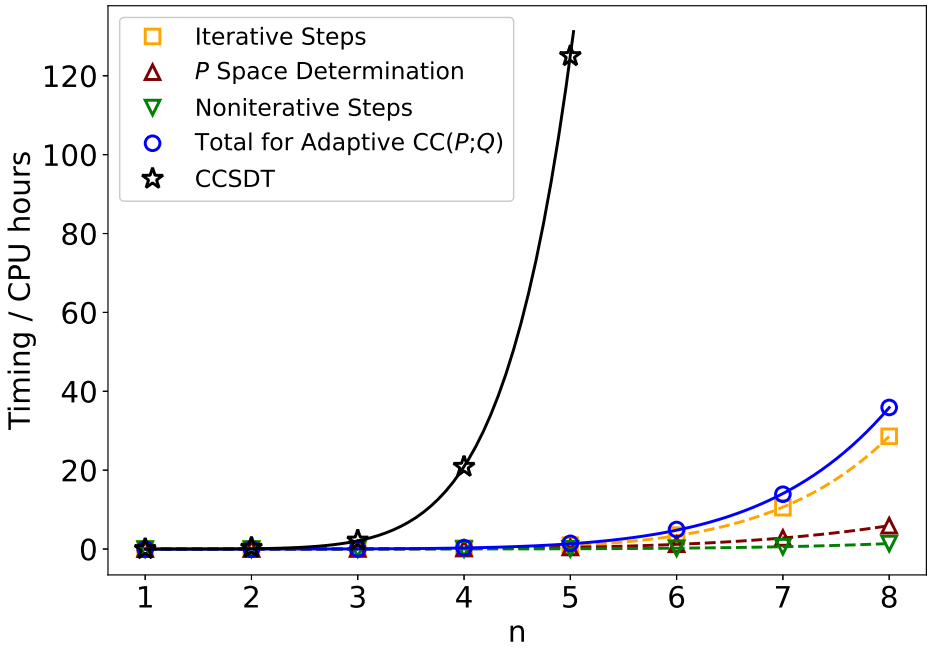

Figure 4.4 Comparison of the computational timings (in CPU hours) characterizing the post-RHF steps of the parent CCSDT and adaptive CC(P;Q)[%T = 1] calculations, decomposed into the timings for P space determination, iterative steps, and noniterative steps, for the series of C_nH_{2n+2} linear alkanes, with n = 1–8, described by the cc-pVDZ basis, using the geometries optimized with B3LYP/cc-pVTZ in Ref. [326]. The timings presented in this figure correspond to single-core runs on the Precision 7920 workstation from Dell equipped with Intel Xeon Silver 4114 2.2 GHz processor boards. In reporting timings, we have ignored the computational times associated with the execution of the integral, SCF, and integral transformation routines used to generate the one- and two-electron molecular integrals in the RHF basis, and the integral sorting operations.

of CPU time, which is roughly 21 hours. In comparison, the CC(P;Q)[%T = 1] calculations for butane complete in just 21.0 minutes. In fact, within one day, which is approximately the time required to perform the CCSDT calculation for butane, the adaptive CC(P;Q) method can be used to obtain results for systems as large as heptane, which consumes 833.1 minutes of CPU time. Within two days of CPU time (specifically, 2153.9 CPU minutes), the CC(P;Q)[%T = 1] calculations can be completed for a system as large as octane. Therefore, this computational scalability test demonstrates that the adaptive CC(P;Q) method can be applied to systems far larger compared to those that can be studied using the high-level CCSDT approach.

It is also interesting to note that our larger-scale computational scalability test confirms

many of the conclusions originally drawn from the calculations on the TS structure of cyclobutadiene, examined in Section 4.3.1. In particular, we observe that the total CPU time of the CC(P:Q)[%T = 1] calculations is dominated by the time required to execute the underlying iterative CC(P) step. For the series of eight alkanes studies here, the fraction of CPU time spent on the iterative steps grows from 50–60% of the total computational timing for n = 1–4 to 68–80% when n = 5–8, with the remaining time mostly taken up by the P space determination step. The increase in the fraction of time spent in the iterative portion of the calculation with n reflects not only the fact that the computational steps involved in the CC(P) calculation grow faster with the system size than either the P space determination or noniterative CC(P;Q) steps, but also the iterative nature of the CC(P) computation. For larger systems, we can expect that the total timing for our adaptive CC(P;Q) calculations will be almost entirely determined by the time required for the iterative CC(P) step, and it will become increasingly more important that we continue improving our CC(P) and EOMCC(P) algorithms in order to take better advantage of modern computational accelerators.

CHAPTER 5

EXTERNALLY CORRECTED COUPLED-CLUSTER METHODS USING CIPSI AND MOMENT EXPANSIONS

In Chapters 3 and 4, we presented two novel strategies for obtaining high-level CC/EOMCC energetics by exploiting the flexibility of the CC(P;Q) moment expansions to judiciously include the most important higher–than–doubly excited determinants in the P spaces used in the CC(P)/EOMCC(P) calculations prior to determining corrections for the missing higher-order correlation effects. In this way, we are able to accurately recover the energetics corresponding to any level of CC/EOMCC theory of interest, including, for example, CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and so on, at small fractions of the computational costs, which, as shown in the last section, are on the order of the basic CCSD and CR-CC(2,3) calculations. The final topic pursued in this thesis seeks to tackle an even loftier goal: converging the exact, FCI, energetics directly at a cost comparable to that of CCSD. In our attempt to address this challenge, we rely on the ec-CC framework, introduced in Section 2.4, and develop a new family of ec-CC approaches in which we solve for the T_1 and T_2 clusters in the presence of their T_3 and T_4 counterparts extracted from CIPSI wave functions and correct the resulting energetics for missing higher-order correlation effects using CC(P;Q)-like moment expansions.

5.1 Theory and Algorithmic Details

We begin by recalling the basic premise underlying all ec-CC schemes, namely that for Hamiltonians containing up to two-body interactions, the CC amplitude equations [Eq. (2.9)] projected onto singly and doubly excited determinants in the absence of any approximations, *i.e.*, $\langle \Phi_i^a | \overline{H} | \Phi \rangle = 0$ and $\langle \Phi_{ij}^{ab} | \overline{H} | \Phi \rangle = 0$, respectively, where $\overline{H} = e^{-T} H e^{T}$, do not engage any terms involving the T_n components of the cluster operator with $n > 4$. As alluded to in Section 2.4, solving Eqs. (2.40) and (2.41) for T_1 and T_2 in the presence of the exact T_3 and T_4 amplitudes extracted from full CI using the cluster-analysis relations, Eq. (2.42), produces the exact T_1 and T_2, as well as the resulting full CI energy.

In this chapter, we focus on the ec-CC approaches based on CIPSI wave functions [214]. As discussed in Section 2.4, when combining the ec-CC framework with modern selected CI wave functions, such as CIPSI, one should adopt the ec-CC-II methodology, which removes the explicitly disconnected diagrams from the cluster analysis of the CI state, as opposed to its ec-CC-I counterpart. Doing so allows us to benefit from the efficient exploration of the many-electron Hilbert space performed by algorithms like CIPSI and gives us the opportunity to improve upon the CIPSI calculations as long as the CIPSI wave function does not saturate the singles through quadruples excitation manifolds. As we will see in our numerical results discussed in the next section, we can make excellent use of this window of opportunity and help accelerate the convergence of the CIPSI wave functions toward the FCI limit in a computationally efficient manner. In performing the CIPSI-driven ec-CC calculations reported in this work, we carried out the following protocol:

1. Given a reference state $|\Phi\rangle$, which in our calculations reported in Table 5.1 was the RHF determinant, choose an input parameter $N_{\mathrm{det(in)}}$, used to terminate the CIPSI wave function growth, and execute a CIPSI run to generate the ground state wave function $|\Psi^{(\mathrm{CIPSI})}\rangle$ spanned by $N_{\mathrm{det(out)}}$ determinants.

2. After the CIPSI run defined by a given $N_{\mathrm{det(in)}}$ value is completed, extract the C_1, C_2, C_3, and C_4 components of the CIPSI wave function $|\Psi^{(\mathrm{CIPSI})}\rangle$.

3. Perform the cluster analysis of $|\Psi^{(\mathrm{CIPSI})}\rangle$ according to Eq. (2.42) to obtain the corresponding $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$. In the case of the CIPSI-driven ec-CC calculations performed in this study, we adhere to the ec-CC-II protocol, in which we remove all $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ amplitudes resulting from Eq. (2.42) that do not have the corresponding C_3 and C_4 coefficient.

4. Solve the ec-CC equations, Eq. (2.40) and Eq. (2.41), to obtain the T_1 and T_2 operators relaxed in the presence of $T_3^{(\mathrm{ext})}$ and $T_4^{(\mathrm{ext})}$ extracted from CIPSI in the preceding step.

After the ec-CC equations are solved to obtain T_1 and T_2, solve the companion left-CCSD-like system Eq. (2.47) to determine the deexcitation operators Λ_1 and Λ_2.

5. Using T_1, T_2, Λ_1, and Λ_2, determine correction δ_3 to the ec-CC-II energy due to T_3 correlations missing from the CIPSI wave function according to Eq. (2.44).

6. To check convergence, repeat Steps 1–5 for a larger value of $N_{\mathrm{det(in)}}$. The CIPSI-driven ec-CC calculations can be regarded as converged if the difference between consecutive ec-CC-II$_3$ energies falls below a user-defined threshold.

In the next section, we perform a test to measure how effective the newly developed CIPSI-driven ec-CC-II and ec-CC-II$_3$ approaches are in recovering the exact, FCI, energetics for the challenging problem of symmetric dissociation in water.

5.2 Symmetric Dissociation in the Water Molecule

The results of our CIPSI-based all-electron ec-CC-II and ec-CC-II$_3$ calculations for the C_{2v}-symmetric double bond dissociation of the water molecule, as described by the cc-pVDZ basis set [10], are summarized in Table 5.1. The molecular geometry, taken from Ref. [240], simultaneously stretches both O–H bonds without changing the $\angle$H–O–H angle. In addition to the equilibrium geometry, designated as $R = R_e$, we also considered two stretches of the O–H bonds by factors of 2 and 3, designated as $R = 2R_e$ and $R = 3R_e$, respectively. To maintain consistency with the study in Ref. [240], we correlated all electrons in the problem and used spherical components of d orbitals throughout. In all CIPSI and ec-CC calculations, the RHF determinant was used as a reference.

We use the double bond dissociation of water described within the cc-pVDZ basis set to showcase the performance of the CIPSI-driven ec-CC-II and ec-CC-II$_3$ approaches since it is small enough to perform full CI, while still posing significant challenges for many quantum chemical approaches at stretched geometries due to substantial non-dynamical correlation effects, resulting in large triply and quadruply excited CI and CC amplitudes. In particular, examining the results in Ref. [214] indicates that CCSD produces significant errors relative to

full CI of 3.744, 22.034, and 10.849 millihartree at $R = R_e$, $2R_e$, and $3R_e$, respectively. The errors at equilibrium arise from the neglect of dynamical correlations described by T_3 and T_4, and as expected, CCSDT and CCSTDQ produce nearly exact results, with errors of 0.493 and 0.019 millihartree relative to full CI, respectively. Once the two O–H bonds are stretched to $2R_e$, CCSD is failing, and CCSDT, lacking T_4 correlations, overshoots full CI by 1.403 millihartree, while CCSDTQ remains practically exact. The most challenging MR geometry occurs at the large $3R_e$ stretch, where CCSDT completely fails, overshooting full CI by a massive 40.126 millihartree, and even CCSDTQ produces a sizeable -4.733 millihartree error relative to full CI. Given the failure of even the powerful CCSDTQ method, it is interesting to see if the ec-CC-II and ec-CC-II$_3$ methodologies, which are formally much less expensive than CCSDTQ, can outperform it and reach the full-CI–level energetics even in these challenging MR geometries.

The results in Table 5.1 demonstrate that both the CIPSI-driven ec-CC-II and ec-CC-II$_3$, particularly the latter method, offer significant improvements over the underlying variational CIPSI computations and achieve a fast convergence toward full CI, even in the challenging $3R_e$ geometry. As shown in Table 5.1, using only 5,000 – 6,000 $S_z = 0$ determinants of the $A_1(C_{2v})$ symmetry in the CIPSI diagonalization spaces, corresponding to the small $N_{\mathrm{det(in)}} = 5,000$ CIPSI run that captures approximately 40–50% of singles, 20–70% of doubles, 1–2% of triples, and 0.2% of quadruples, the ec-CC-II approach reduces the 8.445, 34.996, and 28.755 millihartree errors relative to full CI obtained using the CIPSI method at $R = R_e$, $2R_e$, and $3R_e$, respectively, to 2.445, 4.102, and 4.161 millihartree. This also represents an improvement over the 3.744, 22.034, and 10.849 millihartree errors obtained using the CCSD method, where T_3 and T_4 correlations are neglected. Meanwhile, the triples-corrected ec-CC-II$_3$ approach greatly reduces the errors obtained using ec-CC-II to 0.012, 0.226, and 1.507 millihartree at each geometry, respectively. As a matter of fact, compared to the truncated CI and CC energies found in Ref. [214], the ec-CC-II$_3$ energies are substantially more accurate than the CCSDT, CISDTQ, and even CISDTQP calculations, each of which

are considerably more expensive to perform. For the largest stretch of $3R_e$, the ec-CC-II$_3$ method is actually more accurate than even CCSDTQ, which is remarkable, considering that the ec-CC-II and ec-CC-II$_3$ methods have the storage requirements and computational steps that are similar to CCSD and CR-CC(2,3), respectively. It is also worthwhile to note that the ec-CC-I approach, which retains the problematic fully disconnected terms in $T_3^{(\mathrm{CIPSI})}$ and $T_4^{(\mathrm{CIPSI})}$, produces errors relative to full CI of 8.123, 12.707, and 9.546 millihartree at each geometry. These energies are very similar to those produced using CIPSI alone, which is a consequence of CIPSI saturating the single and double excitation manifolds, resulting in ec-CC-I returning energies very close to CIPSI, in accordance with Theorem 1. Increasing the value of $N_{\mathrm{det(in)}}$ to 50,000, produces larger CIPSI wave functions, containing around 80,000–90,000 determinants of the $S_z = 0$ $A_1(C_{2v})$ symmetry. Using this CIPSI wave function in the ec-CC-I approach results in essentially no change to the CIPSI energetics, indicating that the single and double excitation manifolds are numerically complete. On the other hand, only about 20% of triples and 4% of quadruples are captured using these wave functions, hence the ec-CC-II method is still able to improve upon CIPSI, reducing its 2.612, 2.436, and 0.906 millihartree errors relative to full CI at $R = R_e$, $2R_e$, and $3R_e$, respectively, to 0.626, 0.788, and 0.341 millihartree, respectively. The ec-CC-II$_3$ method, which corrects for missing T_3 correlations, reduces the errors obtained using ec-CC-II at R_e and $2R_e$ to 0.168 and 0.515 millihartree, respectively, while unable to further correct the ec-CC-II results at $3R_e$, producing a 0.358 millihartree error relative to full CI. For values of $N_{\mathrm{det(in)}}$ equal to 50,000 and larger, the ec-CC-II$_3$ appears ineffective at accelerating the convergence of the ec-CC-II energetics toward full CI, which is likely due to the fact that it does not take into account the missing T_4 correlations, which are substantial, particularly in the highly stretched $3R_e$ geometry.

Looking at the overall picture emerging from the results in Table 5.1, it is quite clear that the ec-CC-II and ec-CC-II$_3$, particularly the latter ones, offer a rapid convergence to within a millihartree of full CI using relatively small CIPSI diagonalization spaces. They also converge

to full CI much faster than the variational CIPSI energies. The efficacy of CIPSI-driven ec-CC-II and ec-CC-II$_3$ reinforces the view that the selected CI approaches like CIPSI, which are capable of efficiently sampling the many-electron Hilbert space through a tempered evolution of the wave function, without saturating the lower-rank excitation manifolds, are of the most use in ec-CC considerations. We also mention that, while the convergence of the ec-CC-II and ec-CC-II$_3$ approaches is impressive, it is at times outperformed by the perturbatively corrected CIPSI energies, indicating that at least for the small problems studied so far, the CIPSI and other selected CI models are powerful methods in their own right. To gain further insight into the utility of CIPSI-driven ec-CC-II and ec-CC-II$_3$ approaches, further testing is required, particularly on larger systems described with bigger basis sets, where the perturbatively corrected CIPSI approach may begin to struggle in converging the exact, FCI, limit.

Table 5.1 Convergence of the energies resulting from the all-electron CIPSI calculations initiated from the RHF wave function and the corresponding CIPSI-based ec-CC energies toward FCI for the H_2O molecule, as described by the cc-pVDZ basis set, at the equilbrium and two displaced geometries in which both O–H bonds are stretched by factors of 2 and 3. [a] Adapted from Ref. [214].

$N_{\mathrm{det(in)}}$ / $N_{\mathrm{det(out)}}$	%S[b]	%D[b]	%T[b]	%Q[b]	CIPSI[c]			ec-CC[c]		
					E_{var}	$E_{\mathrm{var}}+\Delta E^{(2)}$	$E_{\mathrm{var}}+\Delta E_r^{(2)}$	I	II	II_3
					$R=R_e$					
1 / 1	0	0	0	0	217.822[d]	−42.098	−21.684	3.744[e]	3.744[e]	0.344[f]
1,000 / 1,299	15.2	37.9	0	0.0	21.589	−0.109	−0.024	11.019	3.637	0.236
5,000 / 5,216	51.5	73.8	1.0	0.2	8.445	0.098	0.108	8.123	2.455	0.012
10,000 / 10,448	60.6	80.4	3.1	0.4	6.587	0.184	0.190	6.441	1.887	0.063
50,000 / 83,762	93.9	94.7	22.1	4.9	2.612	0.089	0.090	2.610	0.626	0.168
100,000 / 167,425	93.9	97.8	34.5	10.0	1.743	0.064	0.064	1.742	0.410	0.190
500,000 / 665,840	100	99.7	69.0	31.9	0.435	0.008	0.008	0.435	0.120	0.105
1,000,000 / 1,308,003	100	99.9	82.5	45.6	0.229	0.004	0.004	0.229	0.069	0.066
					$R=2R_e$					
1 / 1	0	0	0	0	363.956[d]	−180.621	105.169	22.034[e]	22.034[e]	−0.548[f]
1,000 / 1,399	36.4	18.1	0.3	0.0	34.996	1.884	2.139	26.505	9.639	−0.454
5,000 / 5,664	54.5	43.1	1.8	0.2	13.817	0.605	0.638	12.707	4.102	0.226
10,000 / 11,350	57.6	55.6	3.6	0.4	9.011	0.375	0.388	8.653	2.677	0.321
50,000 / 90,880	93.9	85.3	21.6	3.7	2.436	0.084	0.085	2.429	0.788	0.515
100,000 / 181,579	100	91.6	31.7	6.8	1.418	0.046	0.046	1.417	0.467	0.356
500,000 / 718,316	100	97.7	59.6	21.4	0.273	0.009	0.009	0.273	0.147	0.138
1,000,000 / 1,390,678	100	99.3	73.2	31.9	0.137	0.003	0.003	0.137	0.074	0.072
					$R=3R_e$					
1 / 1	0	0	0	0	567.554[d]	−227.583	420.893	10.849[e]	10.849[e]	−40.556[f]
1,000 / 1,437	30.3	8.8	0.3	0.1	28.755	1.010	1.183	22.107	8.467	−1.934
5,000 / 5,793	39.4	21.4	1.8	0.2	9.919	0.337	0.354	9.546	4.161	1.507
10,000 / 11,603	54.5	28.6	3.5	0.4	5.258	0.157	0.161	5.093	1.687	0.699
50,000 / 92,707	87.9	70.8	14.9	2.5	0.906	0.031	0.031	0.902	0.341	0.358
100,000 / 184,903	90.9	78.5	22.7	4.4	0.483	0.011	0.011	0.483	0.177	0.228
500,000 / 703,445	97.0	89.3	46.0	13.5	0.071	−0.001	−0.001	0.071	0.049	0.050
1,000,000 / 1,215,321	97.0	92.8	56.3	19.5	0.029	−0.001	−0.001	0.029	0.017	0.017

[a]The equilibrium geometry, $R = R_e$, and the geometries that represent a simultaneous stretching of both O–H bonds by factors of 2 and 3 without changing the ∠(H–O–H) angle were taken from Ref. [240].

[b]%S, %D, %T, and %Q are, respectively, the percentages of the singly, doubly, triply, and quadruply excited $S_z = 0$ determinants of A_1 symmetry captured during the CIPSI computations.

[c]Errors relative to FCI in millihartree (see Table I of Ref. [214] for the FCI energies).

[d]Equivalent to RHF.

[e]Equivalent to CCSD.

[e]Equivalent to CR-CC(2,3).

CHAPTER 6

CONCLUDING REMARKS AND FUTURE OUTLOOK

In this dissertation, we explored recent advances in the development of new SR CC and EOMCC methodologies aimed at addressing multiconfigurational problems in chemistry. The primary goal was to introduce black-box approaches capable of determining molecular PESs involving significant bond rearrangements, biradical transition states, singlet–triplet gaps, and excited states dominated by both one- and many-electron transitions with lower computational costs compared to other methods that aim at similar precision. Specifically, we introduced, developed, and tested several novel and practical *ab initio* quantum chemistry methods: the CIPSI-driven CC(P;Q), the adaptive CC(P;Q), and the CIPSI-based ec-CC-II and ec-CC-II$_3$ schemes. The first two methodologies exploit the flexibility of the CC(P;Q) framework to automatically converge high-level CC/EOMCC energetics, such as CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, etc., at small fractions of the computational costs, while the latter ec-CC-based approaches offer an intriguing route to converging exact, FCI, energetics directly, using computational steps akin to CCSD or CR-CC(2,3).

In the first two chapters of this dissertation, we introduced the many-electron Schrödinger equation and discussed the computational challenges associated with obtaining numerically exact solutions using the FCI method. We highlighted the practical necessity of resorting to approximate methods and emphasized the preeminent role of SRCC theory in *ab initio* quantum chemistry, as well as its EOM extension to excited electronic states, which provide a hierarchy of increasingly accurate approximations that allows us to converge FCI or near-FCI descriptions for most chemical applications using polynomial computational steps. After presenting the necessary theoretical background on the SR CC and EOMCC formalisms, we demonstrated that higher-level CC/EOMCC approximations, such as CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ, remain robust in the presence of electronic quasi-degeneracies, closely approximating the exact FCI energetics. This was illustrated using bond-breaking in HF and the challenging symmetric dissociation of water. We then

argued that, while high-level methods like CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ can accurately handle significant MR correlations, traditional perturbative approximations, such as CCSD(T), rob us of this benefit and are inadequate for studying chemical bond rearrangements, open-shell systems like biradicals, and excited states dominated by multi-electron transitions. To better describe MR problems within the SRCC framework, we moved away from perturbation theory and adopted the more powerful CC(P;Q) framework. This formalism allows us to partition the manifold of higher-than-doubly excited determinants in high-level CC/EOMCC theories, selecting a small subset of important determinants for inclusion in the P space, which is used to define the cluster and excitation operators and perform the iterative CC(P)/EOMCC(P) step, and correcting the resulting energetics for the remaining correlations of interest captured with the help of the complementary Q space using the nonperturbative CC(P;Q) moment energy expansions.

In Chapters 3 and 4 of this thesis, we introduced two novel CC(P;Q) methodologies based on this idea, which are capable of converging high-level CC/EOMCC energetics in an automated fashion, even in situations where higher–than–pair correlations are large, nonperturbative, and difficult to capture with noniterative corrections to lower-level CC/EOMCC calculations. The first approach combined CC(P;Q) moment expansions with the information provided by CIPSI wave functions, which were used to identify the subsets of higher-than-doubly excited determinants entering the underlying P and Q spaces. The resulting CIPSI-driven CC(P;Q) approach was implemented to converge the CCSDT and EOMCCSDT energetics and tested on several molecular examples, including dissociation of F_2, the automerization of cyclobutadiene, and the vertical excitation spectrum of CH^+. In all cases, we demonstrated that the CIPSI-driven CC(P;Q) method successfully recovered the parent CCSDT/EOMCCSDT results with tiny fractions of the computational costs using P spaces containing a relatively small number of triply excited determinants identified with the help of compact CI wave function expansions. Inspired by the success of the quasi-perturbative CIPSI algorithm, the second CC(P;Q) approach introduced in this dissertation aimed at

creating a CC-based analog of CIPSI, which we called adaptive CC(P;Q). In this approach, the excitation manifolds defining the higher–than–two-body components of the T and R_μ operators are adaptively selected based on the information contained within the CC(P;Q) moment expansions themselves. This novel methodology represents a practical and powerful approach for converging high-level CC/EOMCC energetics, such as CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, and similar types, freeing us from the reliance on active orbitals and external information from non-CC or stochastic sources used in previous formulations of CC(P;Q). The effectiveness of the adaptive CC(P;Q) methodology was demonstrated in numerous applications, where the goal was to recover the CCSDT/EOMCCSDT energetics when noniterative triples corrections to CCSD/EOMCCSD struggle or fail. These tests included the significantly stretched F_2 and F_2^+ molecules, the automerization of cyclobutadiene, singlet–triplet gaps in several organic biradicals, the degenerate Cope rearrangement of bullvalene, and the ground- and excited-state PESs of the water molecule along its O–H bond-breaking coordinate. In addition to demonstrating rapid convergence toward the parent CC results, we discussed the CPU timings for the TS of cyclobutadiene and for a series of C_nH_{2n+2} linear alkanes for n = 1–8 to highlight the significant reductions in computational effort relative to CCSDT achieved by the adaptive CC(P;Q) calculations.

In the final section of this dissertation, we introduced a new generation of ec-CC approaches designed to recover the near-FCI energetics, even in the presence of stronger MR correlation effects, by leveraging information about T_3 and T_4 clusters extracted from CIPSI, combined with CC(P;Q)-like moment corrections. The resulting ec-CC-II methodology, along with its ec-CC-II$_3$ extension that accounts for missing T_3 correlations, was tested on the symmetric double dissociation of water, where we demonstrated that the CIPSI-driven ec-CC calculations, particularly ec-CC-II$_3$, can achieve accuracies on the order of 0.1 millihartree relative to FCI, using relatively inexpensive CIPSI runs with small diagonalization spaces. These results successfully compete with high-level CC and CI methods, such as CCSDTQ and CISDTQPH, at a fraction of the computational cost, and they improve upon

CCSDTQ in the dissociation regime when electronic degeneracies become particularly strong.

The ideas and methodologies developed in this dissertation open several avenues for future work. Regarding the CIPSI-based CC(P;Q) methodologies, one potential direction is to replace the CIPSI algorithm with alternative selected CI schemes, such as heat-bath CI. While heat-bath CI calculations are generally less accurate than CIPSI (as seen by comparing the heat-bath CI results in Ref. [195] with their counterparts obtained using CIPSI in Ref. [194]), they offer the advantage of being computationally less expensive due to using a simplified selection scheme. Since our focus on selected CI within the context of CC(P;Q) is primarily to identify the leading higher–than–doubly excited determinants for setting up CC/EOMCC calculations, rather than converging full FCI energetics, we may benefit from accelerating the preliminary CI steps, even if the resulting wave CI functions are less accurate. Another promising area for future work is the incorporation of state-specific P spaces in selected-CI-driven CC(P;Q) calculations. This would allow us to tailor the P spaces to better capture the many-electron correlation effects specific to each state of interest. Since selected CI diagonalizations provide both ground- and excited-state wave functions, which can be parsed to identify leading higher–than–doubly excited determinants, constructing the necessary state-specific P spaces and integrating them with our existing CIPSI-driven CC(P;Q) codes is straightforward. This approach has the potential to further accelerate convergence of high-level CC/EOMCC energetics. Lastly, we aim to extend the capabilities of our CIPSI-driven CC(P;Q) codes to converge even higher-level CC/EOMCC energetics, such as CCSDTQ/EOMCCSDTQ. Work in the latter direction is already underway.

In the same vein, we anticipate significant future work focused on extending the capabilities of our adaptive CC(P;Q) codes to converge higher levels of CC/EOMCC theory, including CCSDTQ/EOMCCSDTQ, and eventually, possibly, the exact FCI limit. Achieving the latter would involve performing adaptive CC(P;Q) calculations without placing restrictions on the determinants entering our moment energy expansions. This untruncated adaptive CC(P;Q) approach would take advantage of the original purpose of MMCC expan-

sions as corrections toward FCI, ultimately transforming adaptive CC(P;Q) into a form of selected CC. The primary obstacle to realizing this selected CC approach is developing a CC(P)/EOMCC(P) code capable of handling any list of higher–than–doubly excited determinants in an efficient manner. Most codes that are used to perform high-level CC/EOMCC calculations introduce computational intermediates in order to recast the equations in a form that is quasi-linear in the higher–than–two-body components of the cluster operator. Optimally factorized CC/EOMCC equations of this kind are straightforward to obtain if one can limit the maximum excitation rank in the T and R_μ operators *a priori*, such as when targeting CCSDT/EOMCCSDT or CCSDTQ/EOMCCSDTQ. However, if the structure of the CC/EOMCC equations is left unspecified until the list of higher–than–doubly excited determinants is provided, it becomes challenging to design a code that leverages intermediates without incurring excessive CPU or memory costs. To our knowledge, the only robust method for performing general CC calculations of this type rely on less efficient string-based CI-like approaches [102, 330–332]. One potential brute-force solution could involve constructing a library of factorized CC/EOMCC equations up to a high excitation rank and dynamically assembling the appropriate CC(P)/EOMCC(P) code on the fly. However, this approach requires further investigation to fully understand its feasibility.

As we move toward higher levels of CC/EOMCC theory, it is critical that we continue improving our adaptive CC(P;Q) codes from a technical perspective. In particular, we believe that developing new algorithms to reduce the computational cost of filtering and selecting subsets of higher–than–doubly excited determinants entering the P spaces during the evaluation of CC(P;Q) corrections will become increasingly important. One technical optimization of this type involves adopting stochastic selection algorithms, which have proven to be very successful in accelerating CIPSI computations (see Ref. [183] for details). To make progress, one may draw inspiration from the recently developed stochastic closed-shell CCSD(T) method [333], in which the triples correction $\delta_{(T)}$ is decomposed as $\delta_{(T)} = \sum_{a<b<c} \delta_{abc}$, where $\delta_{abc} = \sum_{i<j<k} \ell_{0,ijk}^{abc} \, \mathfrak{M}_{0,abc}^{ijk}$, with $\mathfrak{M}_{0,abc}^{ijk} = \langle \Phi_{ijk}^{abc} | (HT_2)_C | \Phi \rangle$

and $\ell_{0,ijk}^{abc} = (\mathfrak{M}_{0,abc}^{ijk})^{\dagger}/D_{abc}^{ijk}$. A Monte Carlo procedure based on sampling triplets of unoccupied spinorbitals (a,b,c), with $a < b < c$, is employed to estimate $\delta_{(T)}$ as $\delta_{(T)}^{(\mathrm{MC})} = \frac{1}{N_{\mathrm{trials}}} \sum_{(a,b,c)} n_{(a,b,c)} \delta_{abc}$, where N_{trials} is the number of Monte Carlo samples and $n_{(a,b,c)}$ is the number of times that the triplet (a,b,c) is sampled. For each triplet (a,b,c), the corresponding correction δ_{abc} is computed and stored, allowing reuse of previously computed δ_{abc} values in subsequent trials, and in the limit that $N_{\mathrm{trials}} \to \infty$, the Monte Carlo estimate $\delta_{(T)}^{(\mathrm{MC})}$ converges to the deterministic value $\delta_{(T)}$. To accelerate the convergence, the authors of Ref. [333] sample the triplets (a,b,c) according to the probability distribution $P \propto \max(\epsilon, \Delta_{(a,b,c)})$, with $\Delta_{(a,b,c)} \propto (f_a^a + f_b^b + f_c^c)^{-1}$, where f_p^p is the energy of spinorbital $|p\rangle$, which places more weight on corrections δ_{abc} associated with virtual orbitals near the Fermi level. The arbitrary constant ϵ (set to 0.2 hartree in Ref. [333]) is introduced to ensure that P does not diverge. In a similar fashion, one could extend this stochastic algorithm to the CC(P;Q) framework by decomposing the moment energy expansion, Eq. (2.33), into a sum over corrections due to higher–than–doubly excited determinants in the Q space, $\delta_{\mu}(P;Q) = \delta_{\mu,3} + \delta_{\mu,4} + \ldots$, where $\delta_{\mu,n} = \sum_{|\Phi_{i_1\ldots i_n}^{a_1\ldots a_n}\rangle \in \mathscr{H}^{(Q)}} \delta_{i_1\ldots i_n, a_1\ldots a_n}(\mu)$ is the correction corresponding to n-tuply excited determinants in the Q space, with $n > 2$. The corrections $\delta_{\mu,n}$ are approximated using the Monte Carlo estimator $\delta_{\mu,n}^{(\mathrm{MC})} = \frac{1}{N_{\mathrm{trials}}} \sum_{(a_1,\ldots,a_n)} n_{(a_1,\ldots,a_n)} \delta_{a_1\ldots a_n}(\mu)$, where $\delta_{a_1\ldots a_n}(\mu) = \sum_{|\Phi_{i_1\ldots i_n}^{a_1\ldots a_n}\rangle \in \mathscr{H}^{(Q)}} \ell_{\mu,i_1\ldots i_n}^{a_1\ldots a_n}(P)\, \mathfrak{M}_{\mu,a_1\ldots a_n}^{i_1\ldots i_n}(P)$. Following Ref. [333], one could sample tuples $(a_1,\ldots,a_n)$ associated with determinants $|\Phi_{i_1\ldots i_n}^{a_1\ldots a_n}\rangle \in \mathscr{H}^{(Q)}$ according to the distribution $P \propto \max(\epsilon, \Delta_{(a_1,\ldots,a_n)})$, with $\Delta_{(a_1,\ldots,a_n)} \propto (\sum_{i=1}^{n} f_{a_i}^{a_i})^{-1}$, which corresponds to the Møller–Plesset energy denominator, or one could use the more accurate Epstein–Nesbet form. This approach would not only reduce the computational cost of evaluating the corrections associated with higher–than–doubly excited determinants in the Q space, but also stochastically filter the subsets of determinants considered for potential inclusion in the P spaces, thereby improving the efficiency of the adaptive CC(P;Q) method.

In the longer term, we also propose combining the adaptive CC(P;Q) methodology with the cluster-in-molecule (CIM) local correlation framework [334–339], including its single-

environment [340] and multilevel [341] extensions, which has previously been used to develop local CCD, CCSD, CCSD(T), and CR-CC(2,3) approaches. These CIM-based CC methods achieve computational costs that scale linearly with the system size, while maintaining sub-kcal/mol accuracy relative to canonical calculations, successfully extending the applicability of CC approaches to much larger molecules containing hundreds of atoms. Undoubtedly, obtaining high-level CC energetics corresponding to CCSDT, CCSDTQ, and similar methods for large chemical systems remains a key goal for future SRCC development work. We believe that the adaptive CC(P;Q) approach introduced in this dissertation is particularly well-suited for this purpose, as it is fully automated, does not require the user-defined selection of active orbitals, and operates independently of information from external calculations.

Finally, regarding ec-CC approaches, we envision that future work should begin with extensive testing of the existing CIPSI-driven ec-CC-II and ec-CC-II$_3$ methods to evaluate their ability to converge FCI in larger many-electron systems described using more realistic basis sets. This will provide invaluable insight into how CIPSI-based ec-CC approaches perform in more practical scenarios, where obtaining FCI is unfeasible. In particular, this will help us understand how to efficiently extract high-quality T_3 and T_4 clusters from CIPSI without being hindered by the growing cost of Hamiltonian diagonalizations. In this context, detailed studies varying the selection factor f (as initially explored in Ref. [214]) and investigating well-chosen CI expansions for initiating CIPSI runs could be highly beneficial. Methodologically, one might consider correcting the ec-CC-II energetics for the missing T_4 correlation effects to accelerate convergence toward FCI. In fact, we have already developed the ec-CC-II$_{3,4}$ methodology, in which the energies from ec-CC-II calculations are corrected for the missing T_3 and T_4 effects using the CC(P;Q)-like moment expansions discussed in Section 2.4. Good initial tests for the new ec-CC-II$_{3,4}$ method include the symmetric dissociation of water in a larger basis set, where well-converged and extrapolated CIPSI energetics can serve as near-FCI benchmarks, and the ground state of benzene, for which highly accurate FCI estimates were obtained in Ref. [195].

REFERENCES

[1] K. Dyall and K. Faegri, *Introduction to Relativistic Quantum Chemistry* (Oxford University Press, USA, 2007).

[2] W. Heitler and F. London, Z. Phys. **44**, 455 (1927).

[3] J. C. Slater, Phys. Rev. **34**, 1293 (1929).

[4] V. Fock, Z. Physik **61**, 126 (1930).

[5] G. G. Hall, Proc. R. Soc. A **205**, 541 (1951).

[6] C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[7] P.-O. Löwdin, "Correlation problem in many-electron quantum mechanics i. review of different approaches and discussion of some current ideas," in *Adv. Chem. Phys.* (John Wiley & Sons, Ltd, 1958) pp. 207–322.

[8] H. Weyl, *The Classical Groups: Their Invariants and Representations*, Princeton Landmarks in Mathematics and Physics No. no. 1, pt. 1 (Princeton University Press, 1946).

[9] J. Paldus, J. Chem. Phys. **61**, 5321 (1974).

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

[11] W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).

[12] H. Gao, S. Imamura, A. Kasagi, and E. Yoshida, J. Chem. Theory Comput. **20**, 1185 (2024).

[13] J. Hubbard, Proc. R. Soc. Lond., Ser. A **240**, 539 (1957).

[14] N. M. Hugenholtz, Physica **23**, 481 (1957).

[15] F. Coester, Nucl. Phys. **7**, 421 (1958).

[16] F. Coester and H. Kümmel, Nucl. Phys. **17**, 477 (1960).

[17] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).

[18] J. Čížek, Adv. Chem. Phys. **14**, 35 (1969).

[19] J. Čížek and J. Paldus, Int. J. Quantum Chem. **5**, 359 (1971).

[20] J. Paldus, J. Čížek, and I. Shavitt, Phys. Rev. A **5**, 50 (1972).

[21] K. A. Brueckner, Phys. Rev. **100**, 36 (1955).

[22] J. Goldstone, Proc. R. Soc. London A Math. Phys. Eng. Sci. **239**, 267 (1957).

[23] R. K. Nesbet, Phys. Rev. **109**, 1632 (1958).

[24] R. A. Chiles and C. E. Dykstra, J. Chem. Phys. **74**, 4544 (1981).

[25] L. Z. Stolarczyk and H. J. Monkhorst, Int. J. Quantum Chem., Symp. **26**, 267 (1984).

[26] N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari, and G. W. Trucks, Chem. Phys. Lett. **164**, 185 (1989).

[27] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).

[28] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[29] I. Lindgren and D. Mukherjee, Phys. Rep. **151**, 93 (1987).

[30] J. Paldus and X. Li, Adv. Chem. Phys. **110**, 1 (1999).

[31] P. Piecuch and K. Kowalski, Int. J. Mol. Sci. **3**, 676 (2002).

[32] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).

[33] D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, Chem. Rev. **112**, 182 (2012).

[34] F. A. Evangelista, J. Chem. Phys. **149**, 030901 (2018).

[35] K. Kowalski and P. Piecuch, Phys. Rev. A **61**, 052506 (2000).

[36] J. Paldus, P. Piecuch, L. Pylypow, and B. Jeziorski, Phys. Rev. A **47**, 2738 (1993).

[37] P. Piecuch, R. Toboła, and J. Paldus, Chem. Phys. Lett. **210**, 243 (1993).

[38] P. Piecuch and J. Paldus, Phys. Rev. A **49**, 3479 (1994).

[39] K. Kowalski and P. Piecuch, Int. J. Quantum Chem. **80**, 757 (2000).

[40] K. Jankowski, J. Paldus, I. Grabowski, and K. Kowalski, J. Chem. Phys. **97**, 7600 (1992), 101, 1759 (1994) [Erratum].

[41] K. Jankowski, J. Paldus, I. Grabowski, and K. Kowalski, J. Chem. Phys. **101**, 3085

(1994).

[42] M. Nooijen and R. J. Bartlett, J. Chem. Phys. **102**, 3629 (1995).

[43] M. Nooijen and R. J. Bartlett, J. Chem. Phys. **102**, 6735 (1995).

[44] S. Hirata, M. Nooijen, and R. J. Bartlett, Chem. Phys. Lett. **328**, 459 (2000).

[45] M. Musiał and R. J. Bartlett, J. Chem. Phys. **119**, 1901 (2003).

[46] J. R. Gour, P. Piecuch, and M. Włoch, J. Chem. Phys. **123**, 134113 (2005).

[47] J. R. Gour, P. Piecuch, and M. Włoch, Int. J. Quantum Chem. **106**, 2854 (2006).

[48] J. R. Gour and P. Piecuch, J. Chem. Phys. **125**, 234107 (2006).

[49] M. Nooijen and J. G. Snijders, Int. J. Quantum Chem. Symp. **26**, 55 (1992).

[50] M. Nooijen and J. G. Snijders, Int. J. Quantum Chem. **48**, 15 (1993).

[51] J. F. Stanton and J. Gauss, J. Chem. Phys. **101**, 8938 (1994).

[52] R. J. Bartlett and J. F. Stanton, in *Reviews in Computational Chemistry*, Vol. 5, edited by K. B. Lipkowitz and D. B. Boyd (VCH Publishers, New York, 1994) pp. 65–169.

[53] M. Musiał, S. A. Kucharski, and R. J. Bartlett, J. Chem. Phys. **118**, 1128 (2003).

[54] M. Musiał and R. J. Bartlett, Chem. Phys. Lett. **384**, 210 (2004).

[55] Y. J. Bomble, J. C. Saeh, J. F. Stanton, P. G. Szalay, M. Kállay, and J. Gauss, J. Chem. Phys. **122**, 154107 (2005).

[56] M. Kamiya and S. Hirata, J. Chem. Phys. **125**, 074111 (2006).

[57] M. Nooijen and R. J. Bartlett, J. Chem. Phys. **106**, 6441 (1997).

[58] M. Nooijen and R. J. Bartlett, J. Chem. Phys. **107**, 6812 (1997).

[59] M. Wladyslawski and M. Nooijen, in *Low-Lying Potential Energy Surfaces*, ACS Symposium Series, Vol. 828, edited by M. R. Hoffmann and K. G. Dyall (American Chemical Society, Washington, D.C., 2002) pp. 65–92.

[60] M. Nooijen, Int. J. Mol. Sci. **3**, 656 (2002).

[61] K. W. Sattelmeyer, H. F. Schaefer, III, and J. F. Stanton, Chem. Phys. Lett. **378**, 42

(2003).

[62] M. Musiał, A. Perera, and R. J. Bartlett, J. Chem. Phys. **134**, 114108 (2011).

[63] M. Musiał, S. A. Kucharski, and R. J. Bartlett, J. Chem. Theory Comput. **7**, 3088 (2011).

[64] T. Kuś and A. I. Krylov, J. Chem. Phys. **135**, 084109 (2011).

[65] T. Kuś and A. I. Krylov, J. Chem. Phys. **136**, 244109 (2012).

[66] J. Shen and P. Piecuch, J. Chem. Phys. **138**, 194102 (2013).

[67] J. Shen and P. Piecuch, Mol. Phys. **112**, 868 (2014).

[68] A. O. Ajala, J. Shen, and P. Piecuch, J. Phys. Chem. A **121**, 3469 (2017).

[69] J. Shen and P. Piecuch, Mol. Phys. **119**, e1966534 (2021).

[70] S. Gulania, E. F. Kjønstad, J. F. Stanton, H. Koch, and A. I. Krylov, J. Chem. Phys. **154**, 114115 (2021).

[71] M. Musiał, M. Olszówka, D. I. Lyakh, and R. J. Bartlett, J. Chem. Phys. **137**, 174102 (2012).

[72] K. Gururangan, A. K. Dutta, and P. Piecuch, J. Chem. Phys. **162**, 061101 (2025).

[73] G. D. Purvis, III and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).

[74] J. M. Cullen and M. C. Zerner, J. Chem. Phys. **77**, 4088 (1982).

[75] G. E. Scuseria, A. C. Scheiner, T. J. Lee, J. E. Rice, and H. F. Schaefer, III, J. Chem. Phys. **86**, 2881 (1987).

[76] P. Piecuch and J. Paldus, Int. J. Quantum Chem. **36**, 429 (1989).

[77] M. R. Hoffmann and H. F. Schaefer, III, Adv. Quantum Chem. **18**, 207 (1986).

[78] J. Noga and R. J. Bartlett, J. Chem. Phys. **86**, 7041 (1987), 89, 3401 (1988) [Erratum].

[79] G. E. Scuseria and H. F. Schaefer, III, Chem. Phys. Lett. **152**, 382 (1988).

[80] J. D. Watts and R. J. Bartlett, J. Chem. Phys. **93**, 6104 (1990).

[81] N. Oliphant and L. Adamowicz, J. Chem. Phys. **95**, 6645 (1991).

[82] S. A. Kucharski and R. J. Bartlett, Theor. Chim. Acta **80**, 387 (1991).

[83] S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **97**, 4282 (1992).

[84] P. Piecuch and L. Adamowicz, J. Chem. Phys. **100**, 5792 (1994).

[85] K. Emrich, Nucl. Phys. A **351**, 379 (1981).

[86] K. Emrich, Nucl. Phys. A **351**, 397 (1981).

[87] J. Geertsen, M. Rittby, and R. J. Bartlett, Chem. Phys. Lett. **164**, 57 (1989).

[88] D. C. Comeau and R. J. Bartlett, Chem. Phys. Lett. **207**, 414 (1993).

[89] J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).

[90] H. J. Monkhorst, Int. J. Quant. Chem., Symp. **11**, 421 (1977).

[91] D. Mukherjee and P. K. Mukherjee, Chem. Phys. **39**, 325 (1979).

[92] E. Dalgaard and H. J. Monkhorst, Phys. Rev. A **28**, 1217 (1983).

[93] H. Sekino and R. J. Bartlett, Int. J. Quantum Chem. **26**, 255 (1984).

[94] M. Takahashi and J. Paldus, J. Chem. Phys. **85**, 1486 (1986).

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

[96] H. Koch, H. J. A. Jensen, P. Jørgensen, and T. Helgaker, J. Chem. Phys. **93**, 3345 (1990).

[97] A. E. Kondo, P. Piecuch, and J. Paldus, J. Chem. Phys. **102**, 6511 (1995).

[98] A. E. Kondo, P. Piecuch, and J. Paldus, J. Chem. Phys. **104**, 8566 (1995).

[99] H. Nakatsuji, Chem. Phys. Lett. **59**, 362 (1978).

[100] K. Kowalski and P. Piecuch, J. Chem. Phys. **115**, 643 (2001).

[101] K. Kowalski and P. Piecuch, Chem. Phys. Lett. **347**, 237 (2001).

[102] M. Kállay and P. R. Surjan, J. Chem. Phys. **113**, 1359 (2000).

[103] S. Hirata, M. Nooijen, and R. J. Bartlett, Chem. Phys. Lett. **326**, 255 (2000).

[104] S. A. Kucharski, M. Włoch, M. Musiał, and R. J. Bartlett, J. Chem. Phys. **115**, 8263 (2001).

[105] S. Hirata, J. Chem. Phys. **121**, 51 (2004).

[106] M. Kállay and J. Gauss, J. Chem. Phys. **121**, 9257 (2004).

[107] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[108] J. D. Watts, J. Gauss, and R. J. Bartlett, J. Chem. Phys. **98**, 8718 (1993).

[109] Y. S. Lee, S. A. Kucharski, and R. J. Bartlett, J. Chem. Phys. **81**, 5906 (1984), 82, 5761 (1985) [Erratum].

[110] M. Urban, J. Noga, S. J. Cole, and R. J. Bartlett, J. Chem. Phys. **83**, 4041 (1985), 85, 5383 (1986) [Erratum].

[111] J. Noga, R. J. Bartlett, and M. Urban, Chem. Phys. Lett. **134**, 126 (1987).

[112] H. Koch, O. Christiansen, P. Jørgensen, A. M. Sánchez de Merás, and T. Helgaker, J. Chem. Phys. **106**, 1808 (1997).

[113] S. Kucharski and R. Bartlett, J. Chem. Phys. **108**, 9221 (1998).

[114] S. A. Kucharski and R. J. Bartlett, Chem. Phys. Lett. **158**, 550 (1989).

[115] Y. J. Bomble, J. F. Stanton, , M. Kállay, and J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

[116] M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

[117] J. D. Watts and R. J. Bartlett, Chem. Phys. Lett. **233**, 81 (1995).

[118] O. Christiansen, H. Koch, and P. Jørgensen, J. Chem. Phys. **105**, 1451 (1996).

[119] J. D. Watts and R. J. Bartlett, Chem. Phys. Lett. **258**, 581 (1996).

[120] O. Christiansen, H. Koch, and P. Jørgensen, J. Chem. Phys. **103**, 7429 (1995).

[121] D. A. Matthews and J. F. Stanton, J. Chem. Phys. **145**, 124102 (2016).

[122] P. Piecuch, K. Kowalski, I. S. O. Pimienta, and M. J. McGuire, Int. Rev. Phys. Chem. **21**, 527 (2002).

[123] P. Piecuch, K. Kowalski, I. S. O. Pimienta, P.-D. Fan, M. Lodriguito, M. J. McGuire, S. A. Kucharski, T. Kuś, and M. Musiał, Theor. Chem. Acc. **112**, 349 (2004).

[124] P. Piecuch, Mol. Phys. **108**, 2987 (2010).

[125] K. Kowalski and P. Piecuch, J. Chem. Phys. **113**, 8490 (2000).

[126] J. Shen and P. Piecuch, Chem. Phys. **401**, 180 (2012).

[127] J. Shen and P. Piecuch, J. Chem. Phys. **136**, 144104 (2012).

[128] J. Shen and P. Piecuch, J. Chem. Theory Comput. **8**, 4968 (2012).

[129] N. P. Bauman, J. Shen, and P. Piecuch, Mol. Phys. **115**, 2860 (2017).

[130] J. E. Deustua, J. Shen, and P. Piecuch, Phys. Rev. Lett. **119**, 223003 (2017).

[131] I. Magoulas, N. P. Bauman, J. Shen, and P. Piecuch, J. Phys. Chem. A **122**, 1350 (2018).

[132] S. H. Yuwono, I. Magoulas, J. Shen, and P. Piecuch, Mol. Phys. **117**, 1486 (2019).

[133] S. H. Yuwono, A. Chakraborty, J. E. Deustua, J. Shen, and P. Piecuch, Mol. Phys. **118**, e1817592 (2020).

[134] J. E. Deustua, J. Shen, and P. Piecuch, J. Chem. Phys. **154**, 124103 (2021).

[135] K. Gururangan, J. E. Deustua, J. Shen, and P. Piecuch, J. Chem. Phys. **155**, 174114 (2021).

[136] A. Chakraborty, S. H. Yuwono, J. E. Deustua, J. Shen, and P. Piecuch, J. Chem. Phys. **157**, 134101 (2022).

[137] K. Gururangan and P. Piecuch, J. Chem. Phys. **159**, 084108 (2023).

[138] K. Gururangan, J. Shen, and P. Piecuch, Chem. Phys. Lett. **862**, 141840 (2025).

[139] P. Piecuch and K. Kowalski, in *Computational Chemistry: Reviews of Current Trends*, Vol. 5, edited by J. Leszczyński (World Scientific, Singapore, 2000) pp. 1–104.

[140] K. Kowalski and P. Piecuch, J. Chem. Phys. **113**, 18 (2000).

[141] K. Kowalski and P. Piecuch, J. Chem. Phys. **113**, 5644 (2000).

[142] K. Kowalski and P. Piecuch, J. Chem. Phys. **120**, 1715 (2004).

[143] K. Kowalski and P. Piecuch, J. Chem. Phys. **115**, 2966 (2001).

[144] K. Kowalski and P. Piecuch, J. Chem. Phys. **122**, 074107 (2005).

[145] P.-D. Fan, K. Kowalski, and P. Piecuch, Mol. Phys. **103**, 2191 (2005).

[146] P. Piecuch and M. Włoch, J. Chem. Phys. **123**, 224105 (2005).

[147] P. Piecuch, M. Włoch, J. R. Gour, and A. Kinal, Chem. Phys. Lett. **418**, 467 (2006).

[148] M. Włoch, M. D. Lodriguito, P. Piecuch, and J. R. Gour, Mol. Phys. **104**, 2149 (2006), 104, 2991 (2006) [Erratum].

[149] M. Włoch, J. R. Gour, and P. Piecuch, J. Phys. Chem. A **111**, 11359 (2007).

[150] P. Piecuch, M. Włoch, M. Lodriguito, and J. R. Gour, in *Recent Advances in the Theory of Chemical and Physical Systems*, Progress in Theoretical Chemistry and Physics, Vol. 15, edited by S. Wilson, J.-P. Julien, J. Maruani, E. Brändas, and G. Delgado-Barrio (Springer, Dordrecht, 2006) pp. 45–106.

[151] P. Piecuch, J. R. Gour, and M. Włoch, Int. J. Quantum Chem. **109**, 3268 (2009).

[152] G. Fradelos, J. J. Lutz, T. A. Wesołowski, P. Piecuch, and M. Włoch, J. Chem. Theory Comput. **7**, 1647 (2011).

[153] P. Piecuch, M. Włoch, and A. J. C. Varandas, in *Topics in the Theory of Chemical and Physical Systems*, Progress in Theoretical Chemistry and Physics, Vol. 16, edited by S. Lahmar, J. Maruani, S. Wilson, and G. Delgado-Barrio (Springer, Dordrecht, 2007) pp. 63–121.

[154] P. Piecuch, M. Włoch, and A. J. C. Varandas, Theor. Chem. Acc. **120**, 59 (2008).

[155] M. Horoi, J. R. Gour, M. Włoch, M. D. Lodriguito, B. A. Brown, and P. Piecuch, Phys. Rev. Lett. **98**, 112501 (2007).

[156] Y. Ge, M. S. Gordon, and P. Piecuch, J. Chem. Phys. **127**, 174106 (2007).

[157] Y. Ge, M. S. Gordon, P. Piecuch, M. Włoch, and J. R. Gour, J. Phys. Chem. A **112**, 11873 (2008).

[158] A. G. Taube, Mol. Phys. **108**, 2951 (2010).

[159] P. Piecuch, J. A. Hansen, and A. O. Ajala, Mol. Phys. **113**, 3085 (2015).

[160] P. Piecuch, J. A. Hansen, D. Staedter, S. Faure, and V. Blanchet, J. Chem. Phys. **138**,

201102 (2013).

[161] J. J. Lutz and P. Piecuch, Comput. Theor. Chem. **1040–1041**, 20 (2014).

[162] N. Oliphant and L. Adamowicz, J. Chem. Phys. **94**, 1229 (1991).

[163] N. Oliphant and L. Adamowicz, J. Chem. Phys. **96**, 3739 (1992).

[164] P. Piecuch, N. Oliphant, and L. Adamowicz, J. Chem. Phys. **99**, 1875 (1993).

[165] L. Adamowicz, P. Piecuch, and K. B. Ghose, Mol. Phys. **94**, 225 (1998).

[166] K. B. Ghose, P. Piecuch, and L. Adamowicz, J. Chem. Phys. **103**, 9331 (1995).

[167] P. Piecuch, S. A. Kucharski, and R. J. Bartlett, J. Chem. Phys. **110**, 6103 (1999).

[168] P. Piecuch and L. Adamowicz, J. Chem. Phys. **102**, 898 (1995).

[169] P.-D. Fan and S. Hirata, J. Chem. Phys. **124**, 104108 (2006).

[170] G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).

[171] D. Cleland, G. H. Booth, and A. Alavi, J. Chem. Phys. **132**, 041103 (2010).

[172] W. Dobrautz, S. D. Smart, and A. Alavi, J. Chem. Phys. **151**, 094104 (2019).

[173] K. Ghanem, A. Y. Lozovoi, and A. Alavi, J. Chem. Phys. **151**, 224108 (2019).

[174] K. Ghanem, K. Guther, and A. Alavi, J. Chem. Phys. **153**, 224115 (2020).

[175] A. J. W. Thom, Phys. Rev. Lett. **105**, 263004 (2010).

[176] R. S. T. Franklin, J. S. Spencer, A. Zoccante, and A. J. W. Thom, J. Chem. Phys. **144**, 044111 (2016).

[177] J. S. Spencer and A. J. W. Thom, J. Chem. Phys. **144**, 084108 (2016).

[178] C. J. C. Scott and A. J. W. Thom, J. Chem. Phys. **147**, 124105 (2017).

[179] J. L. Whitten and M. Hackmeyer, J. Chem. Phys. **51**, 5584 (1969).

[180] C. F. Bender and E. R. Davidson, Phys. Rev. **183**, 23 (1969).

[181] B. Huron, J. P. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).

[182] R. J. Buenker and S. D. Peyerimhoff, Theor. Chim. Acta. **35**, 33 (1974).

[183] Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, J. Chem. Phys. **147**, 034101 (2017).

[184] Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Fertié, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J.-P. Malrieu, M. Véril, M. Caffarel, P.-F. Loos, E. Giner, and A. Scemama, J. Chem. Theory Comput. **15**, 3591 (2019).

[185] J. B. Schriber and F. A. Evangelista, J. Chem. Phys. **144**, 161106 (2016).

[186] J. B. Schriber and F. A. Evangelista, J. Chem. Theory Comput. **13**, 5354 (2017).

[187] N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, J. Chem. Phys. **145**, 044112 (2016).

[188] N. M. Tubman, C. D. Freeman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, J. Chem. Theory Comput. **16**, 2139 (2020).

[189] W. Liu and M. R. Hoffmann, J. Chem. Theory Comput. **12**, 1169 (2016), 12, 3000 (2016) [Erratum].

[190] N. Zhang, W. Liu, and M. R. Hoffmann, J. Chem. Theory Comput. **16**, 2296 (2020).

[191] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, J. Chem. Theory Comput. **12**, 3674 (2016).

[192] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, J. Chem. Theory Comput. **13**, 1595 (2017).

[193] J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, J. Chem. Phys. **149**, 214110 (2018).

[194] P.-F. Loos, Y. Damour, and A. Scemama, J. Chem. Phys. **153**, 176101 (2020).

[195] J. J. Eriksen, T. A. Anderson, J. E. Deustua, K. Ghanem, D. Hait, M. R. Hoffmann, S. Lee, D. S. Levine, I. Magoulas, J. Shen, N. M. Tubman, K. B. Whaley, E. Xu, Y. Yao, N. Zhang, A. Alavi, G. K.-L. Chan, M. Head-Gordon, W. Liu, P. Piecuch, S. Sharma, S. L. Ten-no, C. J. Umrigar, and J. Gauss, J. Phys. Chem. Lett. **11**, 8922 (2020).

[196] J. Paldus, J. Čížek, and M. Takahashi, Phys. Rev. A **30**, 2193 (1984).

[197] P. Piecuch and J. Paldus, Int. J. Quantum Chem., Symp. **40**, 9 (1991).

[198] J. Paldus and J. Planelles, Theor. Chim. Acta **89**, 13 (1994).

[199] J. Planelles, J. Paldus, and X. Li, Theor. Chim. Acta **89**, 33 (1994).

[200] J. Planelles, J. Paldus, and X. Li, Theor. Chim. Acta **89**, 59 (1994).

[201] L. Z. Stolarczyk, Chem. Phys. Lett. **217**, 1 (1994).

[202] P. Piecuch, R. Toboła, and J. Paldus, Phys. Rev. A **54**, 1210 (1996).

[203] X. Li and J. Paldus, J. Chem. Phys. **107**, 6257 (1997).

[204] X. Li, G. Peris, J. Planelles, F. Rajadall, and J. Paldus, J. Chem. Phys. **107**, 90 (1997).

[205] G. Peris, J. Planelles, and J. Paldus, Int. J. Quantum Chem. **62**, 137 (1997).

[206] X. Li and J. Paldus, J. Chem. Phys. **108**, 637 (1998).

[207] G. Peris, F. Rajadell, X. Li, J. Planelles, and J. Paldus, Mol. Phys. **94**, 235 (1998).

[208] G. Peris, J. Planelles, J.-P. Malrieu, and J. Paldus, J. Chem. Phys. **110**, 11708 (1999).

[209] X. Li and J. Paldus, J. Chem. Phys. **124**, 174101 (2006).

[210] J. Paldus, J. Math Chem. **55**, 477 (2017).

[211] J. E. Deustua, I. Magoulas, J. Shen, and P. Piecuch, J. Chem. Phys. **149**, 151101 (2018).

[212] G. J. R. Aroeira, M. M. Davis, J. M. Turney, and H. F. Schaefer, J. Chem. Theory Comput. **17**, 182 (2021).

[213] S. Lee, H. Zhai, S. Sharma, C. Umrigar, and G. Chan, J. Chem. Theory Comput. **17**, 3414 (2021).

[214] I. Magoulas, K. Gururangan, P. Piecuch, J. E. Deustua, and J. Shen, J. Chem. Theory Comput. **17**, 4006 (2021).

[215] J. Čížek, J. Paldus, and L. Šroubková, Int. J. Quantum Chem. **3**, 227 (1969).

[216] J. Paldus and J. Čížek (Academic Press, 1975) pp. 105–197.

[217] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge Molecular Science (Cambridge University Press, 2009).

[218] X. Li and J. Paldus, J. Chem. Phys. **101**, 8812 (1994).

[219] S. Hirata, J. Phys. Chem. A **107**, 9887 (2003).

[220] M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945 (2001).

[221] D. I. Lyakh, V. V. Ivanov, and L. Adamowicz, J. Chem. Phys. **122**, 024108 (2005).

[222] N. C. Rubin and A. E. D. III, Mol. Phys. **119**, e1954709 (2021).

[223] F. A. Evangelista, J. Chem. Phys. **157**, 064111 (2022).

[224] I. Purvis, George D. and R. J. Bartlett, J. Chem. Phys. **75**, 1284 (1981).

[225] P. Piecuch and L. Adamowicz, J. Chem. Phys. **100**, 5857 (1994).

[226] P. Piecuch, S. A. Kucharski, K. Kowalski, and M. Musiał, Comp. Phys. Commun. **149**, 71 (2002).

[227] G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irle, J. Ivanic, K. Kowalski, S. S. Leang, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. G. Vallejo, B. Westheimer, M. Włoch, P. Xu, F. Zahariev, and M. S. Gordon, J. Chem. Phys. **152**, 154102 (2020).

[228] K. Gururangan and P. Piecuch, "CCpy: A Coupled-Cluster Package Written in Python," see https://github.com/piecuch-group/ccpy.

[229] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).

[230] P. Pulay, J. Comput. Chem. **3**, 556 (1982).

[231] T. P. Hamilton and P. Pulay, J. Chem. Phys. **84**, 5728 (1986).

[232] P. Ettenhuber and P. Jørgensen, J. Chem. Theory Comput. **11**, 1518 (2015).

[233] E. F. Kjønstad, S. D. Folkestad, and H. Koch, J. Chem. Phys. **153**, 014104 (2020).

[234] C. Yang, J. Brabec, L. Veis, D. B. Williams-Young, and K. Kowalski, Front. Chem. **8**,

1 (2020).

[235] L. Meissner, Mol. Phys. **115**, 2629 (2017).

[236] J. Meller, J. P. Malrieu, and R. Caballol, J. Chem. Phys. **104**, 4068 (1996).

[237] K. Hirao and H. Nakatsuji, J. Comput. Phys. **45**, 246 (1982).

[238] E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).

[239] T. H. Dunning, Jr., J. Chem. Phys. **53**, 2823 (1970).

[240] J. Olsen, F. Jorgensen, H. Koch, A. Balková, and R. J. Bartlett, J. Chem. Phys. **104**, 8007 (1996).

[241] M. Musiał and R. J. Bartlett, J. Chem. Phys. **122**, 224102 (2005).

[242] J. Olsen, A. M. Sánchez de Merás, H. Jensen, and P. Jørgensen, Chem. Phys. Lett. **154**, 380 (1989).

[243] K. Jankowski, J. Paldus, and P. Piecuch, Theor. Chim. Acta **80**, 223 (1991).

[244] K. Kowalski and P. Piecuch, J. Chem. Phys. **116**, 7411 (2002).

[245] A. D. Chien, A. A. Holmes, M. Otten, C. J. Umrigar, S. Sharma, and P. M. Zimmerman, J. Phys. Chem. A **122**, 2714 (2018).

[246] J. Coe, A. M. Carrascosa, M. Simmermacher, A. Kirrander, and M. J. Paterson, J. Chem. Theory Comput. **18**, 6690 (2022).

[247] Z. Wang, Y. Li, and J. Lu, J. Chem. Theory Comput. **15**, 3558 (2019).

[248] J. Coe, J. Chem. Theory Comput. **14**, 5739 (2018).

[249] J. J. Goings, H. Hu, C. Yang, and X. Li, J. Chem. Theory Comput. **17**, 5482 (2021).

[250] V. G. Chilkuri, T. Applencourt, K. Gasperich, P.-F. Loos, and A. Scemama, in *Advances in Quantum Chemistry*, Advances in Quantum Chemistry (Academic Press, 2021).

[251] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[252] K. Kowalski and P. Piecuch, Chem. Phys. Lett. **344**, 165 (2001).

[253] J. E. Deustua, S. H. Yuwono, J. Shen, and P. Piecuch, J. Chem. Phys. **150**, 111101 (2019).

[254] A. Balková and R. J. Bartlett, J. Chem. Phys. **101**, 8972 (1994).

[255] D. I. Lyakh, V. F. Lotrich, and R. J. Bartlett, Chem. Phys. Lett. **501**, 166 (2011).

[256] D. W. Whitman and B. K. Carpenter, J. Am. Chem. Soc. **104**, 6473 (1982).

[257] B. K. Carpenter, J. Am. Chem. Soc. **105**, 1700 (1983).

[258] B. A. Hess, P. Čarsky, and L. J. Schaad, J. Am. Chem. Soc. **105**, 695 (1983).

[259] A. F. Voter and W. A. Goddard III, J. Am. Chem. Soc. **108**, 2830 (1986).

[260] P. Čarsky, R. J. Bartlett, G. Fitzgerald, J. Nova, and V. Špirko, J. Chem. Phys. **89**, 3008 (1988).

[261] O. Demel and J. Pittner, J. Chem. Phys. **124**, 144112 (2006).

[262] M. Eckert-Maksić, M. Vazdar, M. Barbatti, H. Lischka, and Z. B. Maksić, J. Chem. Phys. **125**, 064310 (2006).

[263] K. Bhaskaran-Nair, O. Demel, and J. Pittner, J. Chem. Phys. **129**, 184105 (2008).

[264] P. B. Karadakov, J. Phys. Chem. A **112**, 7303 (2008).

[265] O. Demel, K. R. Shamasundar, L. Kong, and M. Nooijen, J. Phys. Chem. A **112**, 11895 (2008).

[266] J. Shen, T. Fang, S. Li, and Y. Jiang, J. Phys. Chem. A **112**, 12518 (2008).

[267] X. Li and J. Paldus, J. Chem. Phys. **131**, 114103 (2009).

[268] T. Zhang, C. Li, and F. A. Evangelista, J. Chem. Theory Comput. **15**, 4399 (2019).

[269] P. G. Szalay and R. J. Bartlett, Chem. Phys. Lett. **214**, 481 (1993).

[270] P. G. Szalay and R. J. Bartlett, J. Chem. Phys. **103**, 3600 (1995).

[271] A. G. Myers, P. S. Dragovich, and E. Y. Kuo, J. Am. Chem. Soc. **114**, 9369 (1992).

[272] S. Pedersen, J. L. Herek, and A. H. Zewail, Science **266**, 1359 (1994).

[273] J. A. Berson, Acc. Chem. Res. **11**, 446 (1978).

[274] M. Z. Zgierski, S. Patchkovskii, and E. C. Lim, J. Chem. Phys. **123**, 081101 (2005).

[275] M. Z. Zgierski, S. Patchkovskii, T. Fujiwara, and E. C. Lim, J. Phys. Chem. A **109**, 9384 (2005).

[276] W. Park, J. Shen, S. Lee, P. Piecuch, M. Filatov, and C. H. Choi, J. Phys. Chem. Lett. **12**, 9720 (2021).

[277] M. Abe, J. Ye, and M. Mishima, Chem. Soc. Rev. **41**, 3808 (2012).

[278] D. A. Dougherty, Acc. Chem. Res. **24**, 88 (1991).

[279] J. P. Malrieu, R. Caballol, C. J. Caldazo, C. de Graaf, and N. Guihéry, Chem. Rev. **114**, 429 (2014).

[280] D. Cho, K. C. Ko, and J. Y. Lee, Int. J. Quant. Chem. **116**, 578 (2016).

[281] Z. Sun, Z. Zeng, and J. Wu, Acc. Chem. Res. **47**, 2582 (2014).

[282] T. Minami and M. Nakano, J. Phys. Chem Lett. **3**, 145 (2012).

[283] G. J. Hedley, A. Ruseckas, and I. D. W. Samuel, Chem. Rev. **117**, 796 (2017).

[284] J. Niklas and O. G. Poluektov, Adv. Energy Mater. **7**, 1602226 (2017).

[285] M. B. Smith and J. Michl, Chem. Rev. **110**, 6891 (2010).

[286] L. V. Slipchenko and A. I. Krylov, J. Chem. Phys. **117**, 4694 (2002).

[287] X. Li and J. Paldus, J. Chem. Phys. **129**, 174101 (2008).

[288] T. Saito, S. Nishihara, S. Yamanaka, Y. Kitagawa, T. Kawakami, S. Yamada, H. Isobe, M. Okumura, and K. Yamaguchi, Theor. Chem. Acc. **130**, 749 (2011).

[289] D. H. Ess, E. R. Johnson, X. Hu, and W. Yang, J. Phys. Chem. A **115**, 76 (2011).

[290] M. Abe, Chem. Rev. **113**, 7011 (2013).

[291] A. J. Garza, C. A. Jiménez-Hoyos, and G. E. Scuseria, J. Chem. Phys. **140**, 244102 (2014).

[292] C. U. Ibeji and D. Ghosh, Phys. Chem. Chem. Phys. **17**, 9849 (2015).

[293] A. O. Ajala, J. Shen, and P. Piecuch, J. Phys. Chem. A **121**, 3469 (2017).

[294] S. J. Stoneburner, J. Shen, A. O. Ajala, P. Piecuch, D. G. Truhlar, and L. Gagliardi, J. Chem. Phys. **147**, 164120 (2017).

[295] P. M. Zimmerman, J. Phys. Chem. A **121**, 4712 (2017).

[296] J. Shen and P. Piecuch, Mol. Phys. **119**, e1966534 (2021).

[297] S. Gulania, E. F. Kjønstad, J. F. Stanton, H. Koch, and A. I. Krylov, J. Chem. Phys. **154**, 114115 (2021).

[298] J.-N. Boyn and D. A. Mazziotti, J. Chem. Phys. **154**, 134103 (2021).

[299] D. P. Craig, Proc. R. Soc. London, Ser. A **202**, 498 (1950).

[300] R. J. Buenker and S. D. Peyerimhoff, J. Chem. Phys. **48**, 354 (1968).

[301] W. T. Borden, J. Am. Chem. Soc. **97**, 5968 (1975).

[302] H. Kollmar and V. Staemmler, J. Am. Chem. Soc. **99**, 3583 (1977).

[303] W. T. Borden, E. R. Davidson, and P. Hart, J. Am. Chem. Soc. **100**, 388 (1978).

[304] K. Nakamura, Y. Osamura, and S. Iwata, Chem. Phys. **136**, 67 (1989).

[305] S. V. Levchenko and A. I. Krylov, J. Chem. Phys. **120**, 175 (2004).

[306] E. Monino, M. Boggio-Pasqua, A. Scemama, D. Jacquemin, and P.-F. Loos, J. Phys. Chem. A **126**, 4664 (2022).

[307] W. E. Doering and W. R. Roth, Tetrahedron **19**, 715 (1963).

[308] G. Schröder, Angew. Chem., Int. Ed. Engl. **2**, 481 (1963).

[309] W. E. Doering and W. R. Roth, Angew. Chem. Internat. Edit. **2**, 115 (1963).

[310] A. Ault, J. Chem. Educ. **78**, 924 (2001).

[311] M. J. S. Dewar and W. W. Schoeller, J. Am. Chem. Soc. **6**, 1481 (1971).

[312] P. O. Moreno, C. Suarez, M. Tafazzoli, and N. S. True, J. Phys. Chem. **96**, 10206 (1992).

[313] M. Saunders, Tetrahedron Lett. **4**, 1699 (1963).

[314] A. Allerhand and H. S. Gutowsky, J. Am. Chem. Soc. **87**, 4092 (1965).

[315] G. Schröder and J. F. M. Oth, Angew. Chem., Int. Ed. Engl. **6**, 414 (1967).

[316] J. F. M. Oth, K. Müllen, J.-M. Gilles, and G. Schröder, Helv. Chim. Acta **57**, 1415 (1974).

[317] H. Günther and J. Ulmen, Tetrahedron **30**, 3781 (1974).

[318] H. Nakanishi and O. Yamamoto, Tetrahedron Lett. **15**, 1803 (1974).

[319] R. Poupko, H. Zimmermann, and Z. Luz, J. Am. Chem. Soc. **106**, 5391 (1984).

[320] D. A. Hrovat, E. C. Brown, R. V. Williams, H. Quast, and W. T. Borden, J. Org. Chem. **70**, 2627 (2005).

[321] O. Yahiaoui, L. F. Pašteka, B. Judeel, and T. Fallon, Angew. Chem. Internat. Edit. **57**, 2570 (2018).

[322] Y.-Y. Ma, M. Yan, H.-R. Li, Y.-B. Wu, X.-X. Tian, H.-G. Lu, and S.-D. Li, Sci. Rep. **9** (2019), 10.1038/s41598-019-53488-5.

[323] M. Khojandi, A. Seif, E. Zahedi, L. R. Domingo, and M. Karimkhani, New J. Chem. **44**, 6543 (2020).

[324] A. Karton, Chem. Phys. Lett. **759**, 138018 (2020).

[325] M. Lesiuk, J. Chem. Theory Comput. **18**, 6537 (2022).

[326] M. Lesiuk, J. Chem. Theory Comput. **16**, 453 (2020).

[327] X. Li and J. Paldus, J. Chem. Phys. **133**, 024102 (2010).

[328] S. Hirata, M. Nooijen, I. Grabowski, and R. J. Bartlett, J. Chem. Phys. **114**, 3919 (2001), 115, 3967 (2001) [Erratum].

[329] T. Shiozaki, K. Hirao, and S. Hirata, J. Chem. Phys. **126**, 244106 (2007).

[330] J. Olsen, J. Chem. Phys. **113**, 7140 (2000).

[331] S. Hirata and R. J. Bartlett, Chem. Phys. Lett. **321**, 216 (2000).

[332] M. L. Abrams and C. D. Sherrill, Chem. Phys. Lett. **404**, 284 (2005).

[333] Y. Damour, A. Gallo, and A. Scemama, "Stochastically accelerated perturbative triples

correction in coupled cluster calculations," (2024), arXiv:2405.16962 .

[334] S. Li, J. Ma, and Y. Jian, J. Comput. Chem. **23**, 237 (2002).

[335] S. Li, J. Shen, W. Li, and Y. Jiang, J. Chem. Phys. **125**, 074109 (2006).

[336] W. Li, P. Piecuch, and J. R. Gour, in *Theory and Applications of Computational Chemistry - 2008*, AIP Conference Proceedings, Vol. 1102, edited by D.-Q. Wei and X.-J. Wang (American Institute of Physics, Melville, New York, 2009) pp. 68–113.

[337] W. Li, P. Piecuch, and J. R. Gour, in *Advances in the Theory of Atomic and Molecular Systems: Conceptual and Computational Advances in Quantum Chemistry*, Progress in Theoretical Chemistry and Physics, Vol. 19, edited by P. Piecuch, J. Maruani, G. Delgado-Barrio, and S. Wilson (Springer, Dordrecht, 2009) pp. 131–195.

[338] W. Li, J. R. Gour, P. Piecuch, and S. Li, J. Chem. Phys. **131**, 114109 (2009).

[339] Z. N. Wei Li and S. Li, Mol. Phys. **114**, 1447 (2016).

[340] W. Li and P. Piecuch, J. Phys. Chem. A **114**, 8644 (2010).

[341] W. Li and P. Piecuch, J. Phys. Chem. A **114**, 6721 (2010).

[342] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, WIREs Comput. Mol. Sci. **8**, e1340 (2018).

[343] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, J. Chem. Phys. **153**, 024109 (2020).

[344] A. I. Krylov, Chem. Phys. Lett. **338**, 375 (2001).

[345] L. V. Slipchenko and A. I. Krylov, J. Chem. Phys. **123**, 084107 (2005).

[346] C. E. Smith, R. A. King, and T. D. Crawford, J. Chem. Phys. **122**, 054110 (2005).

[347] P. Neogrády and M. Urban, Int. J. Quantum Chem. **55**, 187 (1994).

[348] R. A. Chiles and C. E. Dykstra, Chem. Phys. Lett. **80**, 69 (1981).

[349] S. M. Bachrach, R. A. Chiles, and C. E. Dykstra, J. Chem. Phys. **75**, 2270 (1981).

[350] J. Paldus, M. Takahashi, and R. W. H. Cho, Phys. Rev. B **30**, 4267 (1984).

[351] J. Paldus, J. Čížek, and M. Takahashi, Phys. Rev. A **30**, 2193 (1984).

[352] P. Piecuch and J. Paldus, Theoret. Chim. Acta **78**, 65 (1990).

[353] P. Piecuch and J. Paldus, Int. J. Quantum Chem. **40**, 9 (1991).

[354] J. Paldus and P. Piecuch, Int. J. Quantum Chem. **42**, 135 (1992).

[355] P. Piecuch, R. Toboła, and J. Paldus, Int. J. Quantum Chem. **55**, 133 (1995).

[356] P. Piecuch, R. Toboła, and J. Paldus, Phys. Rev. A **54**, 1210 (1996).

[357] I. Magoulas, J. Shen, and P. Piecuch, Mol. Phys. **120**, e2057365 (2022).

[358] R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6769 (1992).

APPENDIX A

CCPY: AN OPEN-SOURCE COUPLED-CLUSTER PACKAGE IMPLEMENTED IN PYTHON

In this appendix, we provide information about the CCpy package available on GitHub [228], which is a user-friendly and flexible open-source Python-driven software package for performing CC and EOMCC calculations for many-electron systems. The CCpy package was originally developed to supplement the research reported in this dissertation, and in particular, it contains the functionality required to reproduce the results of the calculations reported in chapters 3–5 of this dissertation. In recent months, CCpy has also been incorporated in the PySCF ecosystem as an official extension module. CCpy offers a rich and diverse selection of CC approaches, which includes a wide range of methods belonging to the conventional, LR, spin-flip (SF), active-orbital-based, CR, externally corrected, approximate coupled-pair (ACP), and CC(P;Q) categories as well as their EOM extensions aimed at describing electronically excited, singly and doubly electron attached (EA, DEA), and singly and doubly ionized (IP, DIP) states of many-electron systems.

CCpy is mainly written in Python, to ensure ease of use and extensibility, and uses Fortran and Numpy libraries within the computationally critical routines to maintain good performance relative to established production-level CC/EOMCC codes, such as those distributed in GAMESS [227, 251]. The CCpy software package focuses on executing post-SCF steps of the CC/EOMCC calculations and currently uses interfaces with either PySCF [342, 343] or GAMESS to perform the initial HF computations and obtain the transformed one- and two-electron integrals in the molecular orbital basis that are used to set up the correlated CC calculations. A general interface that can be used to initialize CCpy calculations using RHF/ROHF-like reference state information and one- and two-electron integrals provided by an FCIDUMP file is also included. In the remainder of this appendix, we aim to categorize and provide a brief overview of the different types of CC/EOMCC calculations that can be performed using CCpy.

Ground-State CC Methodologies

CCpy provides a range of conventional CC methods defined by truncating the cluster operator T at a specific many-body rank. Standard CC approaches, including CCD, CCSD, CCSDT, and CCSDTQ, are available for RHF, UHF, and ROHF references. As emphasized throughout this document, the higher-level CCSDT and CCSDTQ methods are included to give the user access to powerful, albeit computationally expensive, approaches for obtaining quantitative descriptions for correlation in many-electron systems. While these are useful for benchmarking purposes, CCpy offers more practical options for obtaining high-level CC energetics using active-space CC methodologies and the more general CC(P;Q) formalism.

A distinctive feature of CCpy is its general CC(P) module, which performs CC calculations in a P space spanned by all singly and doubly excited determinants along with an arbitrary subset of triply excited determinants. Like the other CC options, the CC(P) implementation is compatible with RHF, UHF, and ROHF references. The CC(P) method offers a huge amount of flexibility in defining CC calculations that include T_3 clusters. While the CC(P) module can be used to carry out CCSDT computations (in a fashion that takes advantage of permutational and molecular point group symmetry), its primary strength lies in allowing to user to perform various kinds of nonperturbative approximations to CCSDT. For example, through the general CC(P) code, CCpy supports active-space CCSDt calculations along with its lower-cost CCSDt(II) and CCSDt(III) analogs introduced in Ref. [126]. Auxiliary routines for generating the appropriate lists of triply excited determinants corresponding to the active-space CCSDt-type calculations are included in CCpy as well. Other varieties of CC(P) calculations, including those entering the semi-stochastic, CIPSI-based, and adaptive CC(P;Q) approaches, will be discussed in the later section on CC(P;Q) methods.

EOMCC Approaches for Studying Electronic Excited States

A number of EOMCC approaches for describing electronically excited states are available in CCpy as well. Within the standard EOMCC hierarchy, CCpy includes the basic EOMCCSD approach as well as its higher-level EOMCCSDT counterpart. As discussed earlier

in this dissertation, the EOMCCSDT approach is capable of providing an accurate treatment of many-electron correlation effects needed to describe both singly and doubly excited states accurately, but at significant computational cost. CCpy provides cheaper, yet robust, alternatives to the EOMCCSDT calculations via the general EOMCC(P) module, which solves the EOMCCSDT-like equations in a P space spanned by all singly and doubly excited determinants and a subset of triply excited determinants. With the help of the previously mentioned functions that generate the lists of triply excited determinants corresponding to active-space truncations schemes, CCpy can perform the EOMCCSDt calculations in addition to the analogously defined EOMCCSDt(II) and EOMCCSDt(III) counterparts. As in the case of the ground-state CC methods, all particle-conserving EOMCC options are compatible with RHF, UHF, and ROHF references. CCpy also includes two SF-EOMCC methods, namely SF-EOMCCSD [344] and SF-EOMCC(2,3) [345], which can be used to describe certain low-spin ground and excited states by applying the spin-flipping excitation operator truncated at the two-body (SF-EOMCCSD) or three-body [SF-EOMCC(2,3)] level to a high-spin (*e.g.*, ROHF- or UHF-based) CCSD reference wave function.

When performing EOMCC calculations, the choice of the guess used to initiate the non-Hermitian variant of the Davidson algorithm is often critical, especially when attempting to describe states dominated by two-electron transitions. CCpy follows the philosophy of GAMESS and provides the user with both the EOMCCS and active-space EOMCCSd guess routines. While the former performs a typical CIS-like diagonalization, which is appropriate for describing singly excited states, the latter also includes a subset of valence double excitations needed to locate low-lying doubly excited states. The EOMCCSd guess is also often helpful in improving convergence for certain excited states. Both the EOMCCS and EOMCCSd guess routines can be used to target states according to irrep and spin multiplicity.

EOMCC Approaches for Studying Electron Attachment and Ionization

CCpy includes a rather wide selection of particle-nonconserving EOMCC approaches of the EA/IP- and DEA/DIP-EOMCC types [42–72], which are versatile tools for studying

open-shell systems. In the EA/IP-EOMCC methodology, the ground and excited states of a target $(N+1)$-electron (EA) or $(N-1)$-electron (IP) system are determined by diagonalizing the similarity-transformed CC Hamiltonian of the underlying N-electron closed-shell core in the appropriate subspace of the Fock space spanned by $(N\pm1)$-electron excited Slater determinants. Such approaches can be used to determine EAs and IPs of closed-shell molecules as well as electronic spectra and PESs of radicals. Similarly, the DEA/DIP-EOMCC approaches diagonalize the N-electron core Hamiltonian in the relevant subspace containing $(N+2)$-electron (DEA) or $(N-2)$-electron (DIP) excitations. The DEA/DIP-EOMCC approaches are especially well-suited to describing the MR correlations characterizing biradicals and PESs along single bond-breaking coordinates. Unlike the particle-conserving or SF CC/EOMCC methods, the EA/IP- and DEA/DIP-EOMCC approaches provide a rigorously spin- and symmetry-adapted description of open-shell electronic states. To preserve this key benefit, the EA/IP- and DEA/DIP-EOMCC methods included in CCpy are designed to work with RHF references only. Although the EA/IP/DEA/DIP-EOMCC codes in CCpy can be used with any RHF/ROHF-like set of orbitals, they are currently set up to use the RHF orbitals of the underlying N-electron ground state. Orbitals of the target system can be used, but this currently requires manual modification of the input scripts.

CCpy includes the basic IP-EOMCCSD($2h$-$1p$) [44, 49, 50, 52, 56–58], EA-EOMCCSD($2p$-$1h$) [42–44, 51, 52, 57, 58], DEA-EOMCCSD($3p$-$1h$) [57, 58, 62, 63, 70], and DIP-EOMCCSD($3h$-$1p$) [57–62, 64, 65] approaches, which respectively treat excitations up to the $2p$-$1h$, $2h$-$1p$, $3p$-$1h$, and $3h$-$1p$ level on top of the CCSD description of the underlying N-electron closed-shell core. In order to obtain a highly accurate description of the target attached or ionized system, it is necessary to include the higher–than–two-body correlations corresponding to the $3p$-$2h$ (in the case of EA-EOMCC), $3h$-$2p$ (in the case of IP-EOMCC), $4p$-$2h$ (in the case of DEA-EOMCC), or $4h$-$2p$ (in the case of DIP-EOMCC) excitations on top of CCSD. Therefore, CCpy includes the EA-EOMCCSD($3p$-$2h$) [44, 46–48], IP-EOMCCSD($3h$-$2p$) [44, 46–48], DEA-EOMCCSD($4p$-$2h$) [66–69], and DIP-EOMCCSD($4h$-$2p$) [66, 67] approaches that

allow one to do this as well. As in the previous cases, the higher-level EA/IP/DEA/DIP-EOMCC methodologies come with much higher computational costs in the diagonalization steps, which scale as $\mathcal{N}^7$ (in the case of EA/IP-EOMCC methods) or $\mathcal{N}^8$ (in the case of DEA/DIP-EOMCC methods) with the system size. In order to provide options to help reduce these costs while maintaining high accuracy, CCpy also provides a few routines for performing general particle-nonconserving EOMCC(P) calculations that diagonalize the CCSD similarity-transformed Hamiltonian in a flexibly defined subspace of the Fock space. In particular, the EA-EOMCC(P) option performs diagonalization in the subspace spanned by all $1p$ and $2p$-$1h$ determinants and a subset of $3p$-$2h$ determinants. Similarly, the IP-EOMCC(P) option executes the diagonalization in a subspace containing all $1h$, $2h$-$1p$, and a subset of $3h$-$2p$ excitations. Likewise, the DEA-EOMCC(P) routine diagonalizes the CCSD Hamiltonian in the $(N+2)$-electron subspace spanned by all $2p$ and $3p$-$1h$ excitations and a general selection of $4p$-$2h$ excitations. With the help of these P-space EOMCC routines, CCpy can perform the active-space EA-EOMCCSDt($3p$-$2h$) [46–48], IP-EOMCCSDt($3h$-$2p$) [46–48], and DEA-EOMCCSD($4p$-$2h$)$\{N_u\}$ [66–69] calculations, again, with the help of auxiliary routines used to generate the requisite lists of $(N \pm 1)$- or $(N+2)$-electron excitations [the P-space variant of the DIP-EOMCCSD($4h$-$2p$), which can be used to execute the analogous DIP-EOMCCSD($4h$-$2p$)$\{N_o\}$ [66, 67] calculations, is currently under development].

CCpy also offers a more limited selection of even higher-level particle-nonconserving EOMCC methods that replace the the CCSD similarity-transformed Hamiltonian of the N-electron reference system with its more accurate CCSDT counterpart in the diagonalization step. These approaches, which include EA-EOMCCSDT($3p$-$2h$) [44, 45], IP-EOMCCSDT($3h$-$2p$) [44, 53–56], and DIP-EOMCCSDT($4h$-$2p$) [72], are useful when one needs to balance the descriptions of the target ionized or attached electronic states and the N-electron closed-shell ground state, as in studies where the goal is to accurately determine the IPs/DIPs or EAs/DEAs of a molecular system [72]. If one is only interested in energy differences between the target ground and excited electronic states, which is the case when determining excitation

spectra and PESs of open-shell radicals and biradicals, then replacing the underlying CCSD similarity-transformed Hamiltonian with its CCSDT counterpart is of less importance.

As in the case of the particle-conserving EOMCC options, CCpy offers multiple initial guess routines for performing the EA/IP/DEA/DIP-EOMCC calculations. For the EA/IP-EOMCC options, CCpy can compute the basic $1p/1h$ roots for describing simple Koopmans' states as well as more sophisticated guesses that include the valence $2p$-$1h/2h$-$1p$ excitations, needed to describe the more strongly correlated doublet and quartet states in the radical excitation spectrum. Similarly, DEA/DIP-EOMCC calculations can be initiated using $2p/2h$ guess vectors, which can be augmented with valence $3p$-$1h/3h$-$1p$ excitations to describe more strongly correlated excited states.

LR CC Methods for Ground and Excited States

CCpy includes a limited selection of LR CC methodologies for describing ground and excited electronic states based on the perturbative CCn hierarchy []. In this category, CCpy includes the CC3 approach and its higher-level CC4 counterpart, which solve an approximate form of the CCSDT and CCSDTQ equations, respectively. In order to study excited electronic states, CCpy includes the EOM extension of the CC3 approach, abbreviated as EOMCC3, which is equivalent to the original LR formulation of CC3 for excitation energies (the two will differ for properties other than energy). The CC3 and EOMCC3 calculations are compatible with RHF, UHF, and ROHF references, whereas, the CC4 implementation is currently only available for RHF references. When using noncanonical orbitals, like ROHF, the occupied-virtual block of the Fock matrix is treated as first order. The CCn calculations using noncanonical orbitals are not orbitally invariant unless one semicanonicalizes the occupied and virtual orbitals prior to entering the CC3/EOMCC3 routine [346].

Perturbative CC/EOMCC Methods

Although we have cautioned against the use of perturbative CC/EOMCC methodologies when studying systems featuring MR correlations and electronic quasi-degeneracies, CCpy includes a selection of perturbative CC/EOMCC corrections, which can be useful for more

benign SR systems and benchmarking purposes. In particular, we provide the CCSD(T) correction to the CCSD energetics in addition to EOMCCSD(T)(a)*, IP-EOMCCSD(T)(a)*, and EA-EOMCCSD(T)(a)* methodologies of Ref. [121], which perturbatively correct the EOMCCSD, IP-EOMCCSD(2h-1p), and EA-EOMCCSD(2h-1p) calculations for three-body correlation effects, respectively. Recently, we have also formulated and implemented the DIP-EOMCCSD(T)(a) approximation [72] to the parent DIP-EOMCCSDT(4h-2p) theory, which incorporates the T_3 correlation effects in the DIP-EOMCC calculations using the CCSD(T)(a) framework adopted in Ref. [121]. All perturbative CC/EOMCC methods can be used with RHF, UHF, and ROHF references. As in the previous case regarding noncanonical CCn calculations, the occupied-virtual block of the Fock matrix is treated as first order, and orbital invariance may be lost unless semicanonicalization steps are performed [108, 347].

CR CC/EOMCC Approaches

In the majority of chemical problems, it is better to address higher–than–pair correlations using the CR CC/EOMCC corrections based on the MMCC framework instead of their perturbative counterparts discussed in the preceding section. CCpy offers a number of CR-CC methodologies, including the CR-CC(2,3) triples correction to CCSD along with its CR-CC(2,4) extension that accounts for the T_4 effects missing from CCSD as well. The higher-level CR-CC(3,4) method has also recently been included in CCpy, which corrects the CCSDT energetics for missing four-body clusters, resulting in a very robust approximation to the parent CCSDTQ energetics.

The CR-EOMCC methods available in CCpy include the CR-EOMCC(2,3) triples correction to the EOMCCSD total energies along with the rigorously size-intensive δ-CR-EOMCC(2,3) correction to the EOMCCSD vertical excitation energies. Recently, we have also extended the CR-EOMCC corrections to the EA/IP-EOMCC frameworks, resulting in the EA-CR-EOMCC(2,3) and IP-CR-EOMCC(2,3) approaches, which correct the energetics resulting from the EA-EOMCCSD(2p-1h) and IP-EOMCCSD(2h-1p) calculations for missing 3p-2h and 3h-2p correlations, respectively. With the exception of the EA/IP-based

CR-EOMCC methods, which are designed to work with RHF references to maintain spin symmetry, all CR CC/EOMCC methods can be used with RHF, UHF, and ROHF references.

CC($P;Q$) Methodologies for Converging High-Level CC/EOMCC Energetics

As discussed throughout this dissertation, there are situations when MR correlations become stronger, leading to substantial coupling between higher- and lower-rank cluster and excitation operators that render the CR CC/EOMCC corrections to the lower-level CC/EOMCC energetics ineffective. To address this challenge, CCpy specializes in performing highly efficient and robust CC($P;Q$) computations aimed at converging high-level CC/EOMCC energetics at tiny fractions of the computational effort. The CC($P;Q$) schemes currently available in CCpy aimed at converging the full CCSDT/EOMCCSDT energetics include the active-orbital-based CC(t;3) method, the semi-stochastic CC($P;Q$) and CIPSI-driven CC($P;Q$) approaches, as well as the adaptive CC($P;Q$) methodology. Extensions of CC($P;Q$) of the adaptive and CIPSI-driven varieties aimed at converging the CCSDTQ/EOMCCSDTQ and similar energetics are currently being developed. We have also recently implemented the active-orbital based EA-CC(t;3) and IP-CC(t;3) approaches, which, in analogy to the CC(t;3) method, are capable of recovering the EA-EOMCCSD(3p-2h) and IP-EOMCCSD(3h-2p) and energetics to within a fraction of a millihartree by correcting the results obtained from EA-EOMCCSDt(3p-2h) and IP-EOMCCSDt(3h-2p) calculations for the missing 3p-2h and 3h-2p correlations, respectively. The development of black-box variants of these EA/IP-CC($P;Q$) approaches driven by CIPSI and adaptive moment expansions will be subjects of future studies as well.

ACP Approaches for Describing Strongly Correlated Systems

One of the most interesting ways of describing strong correlation in many-electron systems using CC theory is found by turning to the ACP framework (see, for example, Refs. [210, 348–357]). The existing ACP methods and their various modifications retain all doubly excited cluster amplitudes, while using subsets of nonlinear diagrams of the CCD/CCSD equations. This eliminates failures of conventional CC approaches, including CCSD and

even CCSDT or CCSDTQ, in strongly correlated situations created by the Mott metal-insulator transitions, as modeled by linear chains, rings, or lattices of hydrogen atoms and the π-electron networks described by the Hubbard and Pariser–Parr–Pople Hamiltonians that model one-dimensional metallic systems with periodic boundary conditions. The basic ACP approaches available in CCpy include the ACCD and ACCSD methods. The T_3 correlations missing from the ACCD/ACCSD calculations can be incorporated by turning to higher-level ACCSDT approach or its more practical ACCSDt approximation, which includes the leading contributions to T_3 clusters using an active-space treatment similar to CCSDt. Triples corrections to the ACCSD approach based on the CR-CC framework are available with the ACC(2,3) option, while corrections to the ACCSDt energetics for missing T_3 correlations are obtained via the ACC(t;3) method.

Hybrid ec-CC Methodologies Combining CI and CC Wave Functions

The final class of CC/EOMCC approaches available in CCpy belong to the ec-CC hierarchy, discussed earlier in this dissertation. In particular, CCpy includes the CIPSI-driven ec-CC-II and ec-CC-II$_3$ approaches discussed in Section 2.4 and Chapter 5 of this dissertation as well as its ec-CC-II$_{3,4}$ extension that corrects for T_4 correlations missing from the CIPSI wave function as well. As in the CIPSI-based CC(P;Q) implementation in CCpy, the determinants and associated coefficients characterizing the external CI wave function must be stored in the format adopted by Quantum Package. Once this is done, the ec-CC options in CCpy are capable of extracting the corresponding T_3 and T_4 via cluster analysis and perform the subsequent ec-CC-II and ec-CC-II$_3$ calculations. In fact, other non-CC sources of T_3 and T_4 clusters can also be used in conjunction with the ec-CC-II/ec-CC-II$_3$ routines in CCpy if the requisite CI coefficients and determinants are provided to CCpy in an external file. In this way, the ec-CC module in CCpy is very flexible and allows one to perform a variety of hybrid ec-CC calculations in a straightforward and unified fashion.

APPENDIX B

DERIVATION OF THE CC(P;Q) MOMENT EXPANSIONS

In this appendix, we provide a derivation of the key formulas defining the CC(P;Q) energy corrections, Eqs. (2.33)–(2.37), introduced in Chapter 2. Following Refs. [126, 146, 147], the derivation starts by considering the asymmetric energy expression involving the CC(P)/EOMCC(P) ket state

$$E_\mu^{(\mathrm{FCI})} = \langle \Psi_\mu^{(\mathrm{FCI})} | H R_\mu^{(P)} e^{T^{(P)}} | \Phi \rangle / \langle \Psi_\mu^{(\mathrm{FCI})} | R_\mu^{(P)} e^{T^{(P)}} | \Phi \rangle. \tag{B.1}$$

Equation (B.1) is easily verified by acting on $\langle \Psi_\mu^{(\mathrm{FCI})}|$ with H on the left via the "turnover rule". Without loss of generality, the exact, FCI, bra state can be parameterized as $\langle \Psi_\mu^{(\mathrm{FCI})}| = \langle \Phi | \mathscr{L}_\mu e^{-T^{(P)}}$, where the deexcitation operator

$$\mathscr{L}_\mu = \mathcal{L}_\mu^{(P)} + \delta \mathscr{L}_\mu^{(P)}, \tag{B.2}$$

with

$$\mathscr{L}_\mu^{(P)} = \delta_{\mu,0} \mathbf{1} + \sum_{|\Phi_K\rangle \in \mathscr{H}^{(P)}} \ell_{\mu,K} (E_K)^\dagger \tag{B.3}$$

and

$$\delta \mathscr{L}_\mu^{(P)} = \sum_{|\Phi_K\rangle \in (\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^\perp} \ell_{\mu,K} (E_K)^\dagger, \tag{B.4}$$

is formally obtained by solving the adjoint FCI eigenvalue problem in the entire many-electron Hilbert space

$$\langle \Phi | \mathscr{L}_\mu \overline{H}^{(P)} = E_\mu^{(\mathrm{FCI})} \langle \Phi | \mathscr{L}_\mu. \tag{B.5}$$

Using this form of the FCI state, Eq. (B.1) becomes

$$E_\mu^{(\mathrm{FCI})} = \langle \Phi | \mathscr{L}_\mu \overline{H}^{(P)} R_\mu^{(P)} | \Phi \rangle / \langle \Phi | \mathscr{L}_\mu R_\mu^{(P)} | \Phi \rangle. \tag{B.6}$$

Expanding the quantity $\mathscr{L}_\mu$ into many-body components according to Eq. (B.2), Eq. (B.6) can be written as

$$E_\mu^{(\mathrm{FCI})} = [\langle \Phi | \mathscr{L}_\mu^{(P)} R_\mu^{(P)} | \Phi \rangle]^{-1} [\langle \Phi | \mathscr{L}_\mu^{(P)} \overline{H}^{(P)} R_\mu^{(P)} | \Phi \rangle + \langle \Phi | \delta \mathscr{L}_\mu^{(P)} \overline{H}^{(P)} R_\mu^{(P)} | \Phi \rangle], \tag{B.7}$$

where we have used the fact that $\langle\Phi|\delta\mathscr{L}_\mu^{(P)}R_\mu^{(P)}|\Phi\rangle = 0$. Next, we impose the normalization

$$\langle\Phi|\mathscr{L}_\mu^{(P)}R_\mu^{(P)}|\Phi\rangle = 1, \tag{B.8}$$

which, while not identical to the biorthonormality relationship satisfied by $L_\mu^{(P)}$ and $R_\mu^{(P)}$ [Eq. (2.31)], is nonetheless a reasonable condition given the similarity between $\mathscr{L}_\mu^{(P)}$ and $L_\mu^{(P)}$. With the help of Eq. (B.8), which also allows us to identify $E_\mu^{(P)} = \langle\Phi|\mathscr{L}_\mu^{(P)}\overline{H}^{(P)}R_\mu^{(P)}|\Phi\rangle$, Eq. (B.7) takes on the more revealing form

$$E_\mu^{(\mathrm{FCI})} = E_\mu^{(P)} + \langle\Phi|\delta\mathscr{L}_\mu^{(P)}\overline{H}^{(P)}R_\mu^{(P)}|\Phi\rangle, \tag{B.9}$$

and the corrections to the CC(P)/EOMCC(P) energetics toward FCI can be defined as

$$E_\mu^{(\mathrm{FCI})} - E_\mu^{(P)} = \langle\Phi|\delta\mathscr{L}_\mu^{(P)}\overline{H}^{(P)}R_\mu^{(P)}|\Phi\rangle. \tag{B.10}$$

Inserting a resolution of identity between $\delta\mathscr{L}_\mu^{(P)}$ and $\overline{H}^{(P)}$ in Eq. (B.10) results in an exact form of the moment energy expansions

$$E_\mu^{(\mathrm{FCI})} - E_\mu^{(P)} = \sum_{|\Phi_K\rangle\in(\mathscr{H}^{(0)}\oplus\mathscr{H}^{(P)})^\perp} \ell_{\mu,K}\,\mathfrak{M}_{\mu,K}(P), \tag{B.11}$$

where $\mathfrak{M}_{\mu,K}(P)$ are the generalized moments of CC(P) [Eq. (2.34)] and EOMCC(P) [Eq. (2.35)] equations and $\ell_{\mu,K}$ are the amplitudes entering the operator $\delta\mathscr{L}_\mu$ characterizing the exact, FCI, bra state $\langle\tilde{\Psi}_\mu^{(\mathrm{FCI})}|$.

As it stands, Eq. (B.11) provides the formulas for the corrections to the energies resulting from CC(P)/EOMCC(P) calculations toward the exact, FCI limit. Unfortunately, we have not yet arrived at a computationally feasible recipe for obtaining these corrections, since the amplitudes $\ell_{\mu,K}$ multiplying moments $\mathfrak{M}_{\mu,K}(P)$ in Eq. (B.11) are determined by solving the exact left-eigenstate problem Eq. (B.5). Simply put, there is no free lunch in obtaining FCI. However, the advantage of Eq. (B.11) lies in the fact that approximations to the coefficients $\ell_{\mu,K}$, even relatively low-order ones, can be used to provide useful noniterative corrections. In the rigorously size-extensive biorthogonal CC(P;Q) framework pursued in this work, we adopt the approximation to $\ell_{\mu,K}$ resulting from the quasi-perturbative Löwdin partitioning of

Eq. (B.5). In particular, by projecting Eq. (B.5) onto the set of determinants corresponding to the content of $\delta\mathscr{L}_\mu^{(P)}$, we obtain

$$\langle\Phi|\mathscr{L}_\mu^{(P)}\overline{H}^{(P)}|\Phi_K\rangle + \langle\Phi|\delta\mathscr{L}_\mu^{(P)}\overline{H}^{(P)}|\Phi_K\rangle = E_\mu^{(\mathrm{FCI})}\ell_{\mu,K}, \quad |\Phi_K\rangle \in (\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^\perp, \quad \text{(B.12)}$$

and assuming that $\overline{H}^{(P)}$ is approximately diagonal in the complementary space [meaning, $\langle\Phi_L|\overline{H}^{(P)}|\Phi_K\rangle \approx \delta_{KL}\langle\Phi_K|\overline{H}^{(P)}|\Phi_K\rangle$ for $|\Phi_K\rangle, |\Phi_L\rangle \in (\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^\perp$], we can solve for the amplitudes $\ell_{\mu,K}$ in Eq. (B.12) directly to obtain

$$\ell_{\mu,K} = \langle\Phi|\mathscr{L}_\mu^{(P)}\overline{H}^{(P)}|\Phi_K\rangle/(E_\mu^{(\mathrm{FCI})} - \langle\Phi_K|\overline{H}^{(P)}|\Phi_K\rangle), \quad |\Phi_K\rangle \in (\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^\perp. \quad \text{(B.13)}$$

We can then replace the exact quantities $\mathscr{L}_\mu^{(P)}$ and $E_\mu^{(\mathrm{FCI})}$ in Eq. (B.14) with their respective $L_\mu^{(P)}$ and $E_\mu^{(P)}$ approximations obtained by solving the CC(P)/EOMCC(P) problem to arrive at a computationally tractable form of the coefficients $\ell_{\mu,K}$, namely,

$$\ell_{\mu,K}(P) = \langle\Phi|L_\mu^{(P)}\overline{H}^{(P)}|\Phi_K\rangle/(E_\mu^{(P)} - \langle\Phi_K|\overline{H}^{(P)}|\Phi_K\rangle), \quad |\Phi_K\rangle \in (\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^\perp. \quad \text{(B.14)}$$

Finally, we substitute Eq. (B.14) into Eq. (B.11), limiting ourselves to correcting the CC(P) and EOMCC(P) energetics for correlation effects due to $|\Phi_K\rangle$ in the Q-space $\mathscr{H}^{(Q)}$ rather than in the entire remaining subspace of the Hilbert space $(\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^\perp$. This results in the noniterative CC(P;Q) moment corrections summarized in Eqs. (2.33)–(2.37).

APPENDIX C

EFFICIENT ALGORITHM FOR THE CC(P) AND EOMCC(P) APPROACHES AIMED AT CONVERGING CCSDT AND EOMCCSDT

This appendix contains information about our novel algorithm for solving the CC(P) and EOMCC(P) equations within a P space spanned by all singly and doubly excited determinants and a subset of triply excited determinants, which could be obtained with the help of active orbitals, CIQMC/CCMC wave function propagations, CIPSI calculations, or the adaptive CC(P;Q) moment expansions. Our focus will primarily be on the algorithm used to solve the CC(P) equations aimed at converging CCSDT. Our implementation for the EOMCC(P) approach, which solves a subset of the EOMCCSDT equations, and the companion left-eigenstate CC(P)/EOMCC(P) methods are virtually the same as the CC(P) case after an appropriate definition of one- and two-body intermediates.

C.1 General Considerations

We begin by analyzing the structure of the CC(P)/EOMCC(P) equations, in which the cluster and EOM excitation operators in the P space are $T^{(P)} = T_1 + T_2 + T_3^{(P)}$ and $R_\mu^{(P)} = r_{\mu,0} + R_{\mu,1} + R_{\mu,2} + R_{\mu,3}^{(P)}$, respectively. Let us, for now, assume that the $T_3^{(P)}$ and $R_{\mu,3}^{(P)}$ operators are defined over distinct P spaces, denoted as $\mathscr{H}_\mu^{(P)}$, where $\mu = 0$ corresponds to the P space for the ground state and $\mu > 0$ denote P spaces for excited states. Focusing on the ground-state CC(P) case, if we define the CCSD-like similarity-transformed Hamiltonian $\overline{H}^{(2)} = e^{-T_1-T_2} H e^{T_1+T_2}$ along with the normal-product form of $\overline{H}^{(2)}$ relative to $|\Phi\rangle$, $\overline{H}_N^{(2)} \equiv \overline{H}^{(2)} - E_0^{(P)}\mathbf{1}$, which naturally arise as computational intermediates in the CCSDT system, then Eq. (2.24) becomes

$$\mathfrak{M}_{0,K}(\text{CCSD}) + \langle \Phi_K | [\overline{H}_N^{(2)}, T_3^{(P)}] | \Phi \rangle = 0, \;\; |\Phi_K\rangle \in \mathscr{H}_0^{(P)}, \tag{C.1}$$

where $\mathfrak{M}_{0,K}(\text{CCSD}) = \langle \Phi_K | \overline{H}_N^{(2)} | \Phi \rangle$ are generalized moments of CCSD equations and we have used the fact that the CC(P) amplitude equations corresponding to projections onto singly, doubly, and triply excited determinants eliminate terms that are nonlinear in $T_3^{(P)}$. The programmable expressions for the one- and two-body components of $\overline{H}_N^{(2)}$, denoted as

$\bar{h}_p^q$ and $\bar{h}_{pq}^{rs}$, are given in Table C.1. Similarly, the σ-vectors [Eq. (2.16)] corresponding to projections of the EOMCC(P) eigenvalue problem on the relevant singly, doubly, and subset of triply excited determinants are given by

$$\mathfrak{M}_{\mu,K}(\text{CCSD}) + \langle \Phi_K | [\overline{H}_N^{(2)}, R_{\mu,3}^{(P)}] | \Phi \rangle + \langle \Phi_K | [\overline{X}_\mu^{(2)}, T_3^{(P)}] | \Phi \rangle, \;\; |\Phi_K\rangle \in \mathscr{H}_\mu^{(P)}, \tag{C.2}$$

where $\mathfrak{M}_{\mu,K}(\text{CCSD}) = \langle \Phi_K | \overline{X}_\mu^{(2)} | \Phi \rangle$ are the generalized moments of EOMCCSD equations, with $\overline{X}_\mu^{(2)} = \overline{H}_N^{(2)}(r_{\mu,0} + R_{\mu,1} + R_{\mu,2})$ denoting the intermediates used to factorize the contributions to the EOMCCSDT equations due to three- and four-body components of the CC(P) similarity-transformed Hamiltonian. Thanks to the definition of $\overline{H}_N^{(2)}$ and $\overline{X}_\mu^{(2)}$, the Eqs. (C.1) and (C.2) are linear in $T_3^{(P)}$ and $R_{\mu,3}^{(P)}$. In addition, Eqs. (C.1) and (C.2) have similar structures, and indeed, we will exploit this similarity to implement the EOMCC(P) algorithm by simply recycling code written for the CC(P) case. Before we can discuss these details further, we must first address a concern regarding the potential inclusion of size-intensivity-violating terms in the EOMCC(P) equations.

Equation (C.2) formally contains contributions corresponding to the ground state in the form of $r_{\mu,0}[\mathfrak{M}_{0,K}(\text{CCSD}) + \langle \Phi_K | [\overline{H}_N^{(2)}, T_3^{(P)}] | \Phi \rangle]$ for $|\Phi_K\rangle \in \mathscr{H}_\mu^{(P)}$. In situations where the P spaces for the ground and excited states are identical, meaning that $\mathscr{H}_0^{(P)} = \mathscr{H}_\mu^{(P)} \equiv \mathscr{H}^{(P)}$ for all μ, the ground-state contributions vanish as a result of solving the CC(P) amplitude equations [Eq. (C.1)] and size-intensivity of the resulting excitation energies is automatically retained. This is always the case for conventionally truncated EOMCC approximations as well as active-orbital-based CC(P;Q) methods based on the underlying CCSDt/EOMCCSDt, CCSDtq/EOMCCSDtq, etc. calculations, which *de facto* adopt a single P space for all electronic states. For the semi-stochastic, CIPSI-driven, or adaptive CC(P;Q) approaches, we have the possibility of constructing separate P spaces for each state, and a decision for how to handle the ground-state contributions in the resulting EOMCC(P) equations is required. In the semi-stochastic excited-state CC(P;Q) studies of Refs. [133, 253] and the CIPSI-based CC(P;Q) calculations for CH^+ reported in Section 3.2.3 of this dissertation, the terms proportional to $r_{\mu,0}$ were eliminated in the EOMCC(P) calculations

by adopting a single P space for all electronic states belonging to a particular symmetry. However, in our adaptive CC(P;Q) calculations for water discussed in Section 4.2.4, separate P spaces were constructed for each electronic state, and unless we enforce overlap between the P spaces for the relevant ground and excited states, the EOMCC(P) equations formally retain terms proportional to $r_{\mu,0}$. Our solution for this issue is quite simple: *we drop the size-intensive-violating terms from Eq. (C.2)*, which in practice, is accomplished by simply redefining the computational intermediates $\overline{X}^{(2)}_{\mu} = [\overline{H}^{(2)}_{N}(R_{\mu,1} + R_{\mu,2})]_C$. The one- and two-body components $\bar{x}_{\mu,p}^{\ q}$ and $\overline{x}_{\mu,pq}^{\ rs}$ entering the newly defined quantity $\overline{X}^{(2)}_{\mu}$ are provided in Table C.2. Although our neglect of size-intensive-violating terms may appear ad hoc, it is worth mentioning that this idea is also adopted in other approximations to EOMCCSDT, including EOMCCSDT-n [117, 119] and EOMCCSD(T)(a) [121]. Similar considerations also enter the rigorously size-intensive δ-CR-EOMCC(2,3) [152] triples correction to EOMCCSD as well as its perturbatively defined analogs [119, 328, 329].

With the matter of size-intensivity behind us, we now discuss the similarity between Eqs. (C.1) and (C.2). In particular, the second term in Eq. (C.1) is identical to the second term in Eq. (C.2) if one replaces replaces $R^{(P)}_{\mu,3}$ in the latter by $T^{(P)}_{3}$. The third term in Eq. (C.2) is also the same as the second term in Eq. (C.1) after replacing $\overline{H}^{(2)}_{N}$ by $\overline{X}^{(2)}_{\mu}$. Thus, it should not come as a surprise that we can adopt an essentially identical strategy for solving the EOMCC(P) equations as in the CC(P) case. To be more specific, our EOMCC(P) implementation is obtained by making two copies of the CC(P) source code. In the first copy, the quantity $T^{(P)}_{3}$ is replaced with $R^{(P)}_{\mu,3}$, while in the second copy, the matrix elements of $\overline{H}^{(2)}_{N}$ are replaced by $\overline{X}^{(2)}_{\mu}$. With this in mind, we can restrict our attention for the remainder of this appendix on the CC(P) case. Since we are only considering the ground state, we will drop the subscript μ from the P spaces $\mathscr{H}^{(P)}_{\mu}$ and work with a single P space $\mathscr{H}^{(P)}_{0} \equiv \mathscr{H}^{(P)}$ from here on out.

Returning to our development of the CC(P) algorithm, the next step is to expand the commutator in Eq. (C.1) and insert a resolution of identity between the resulting products

of $\overline{H}_N^{(2)}$ and $T_3^{(P)}$ to obtain the working equation used in our CC(P) algorithm,

$$\underbrace{\mathfrak{M}_{0,K}(\text{CCSD})}_{\text{(I)}} + \underbrace{\sum_{|\Phi_{lmn}^{def}\rangle \in \mathscr{H}^{(P)}} \langle \Phi_K | \overline{H}_N^{(2)} | \Phi_{lmn}^{def} \rangle t_{def}^{lmn}}_{\text{(II)}} = 0, \;\; |\Phi_K\rangle \in \mathscr{H}^{(P)}. \tag{C.3}$$

In what follows, we will describe the strategies adopted for evaluating terms (I) and (II) separately.

Table C.1 The one- and two-body components of the CCSD-like similarity-transformed Hamiltonian $\overline{H}_N^{(2)}$ expressed in terms of the one- and two-body cluster amplitudes, t_a^i and t_{ab}^{ij} respectively, and matrix elements of the Fock and two-electron interaction operators, denoted as $f_p^q \equiv \langle p|f|q\rangle$ and $v_{pq}^{rs} \equiv \langle pq|v|rs\rangle - \langle pq|v|sr\rangle$.

Component of $\overline{H}_N^{(2)}$	Expression[a]
$\bar{h}_m^e$	$f_m^e + v_{mn}^{ef} t_f^n$
$\bar{h}_j^i$	$f_j^i + \bar{h}_j^e t_e^i + v_{jm}^{ie} t_e^m + \frac{1}{2} v_{jn}^{ef} t_{ef}^{in}$
$\bar{h}_a^b$	$f_a^b - \bar{h}_m^b t_a^m + v_{am}^{be} t_e^m - \frac{1}{2} v_{mn}^{bf} t_{af}^{mn}$
$\bar{h}_{mn}^{ef}$	v_{mn}^{ef}
$\bar{h}_{am}^{ef}$	$v_{am}^{ef} - v_{mn}^{fe} t_a^n$
$\bar{h}_{mn}^{ie}$	$v_{mn}^{ie} + v_{mn}^{fe} t_f^i$
$\bar{h}_{ab}^{ef}$	$v_{ab}^{ef} + \frac{1}{2} v_{mn}^{ef} \tau_{ab}^{mn} - \mathscr{A}_{ab} v_{am}^{ef} t_b^m$
$\bar{h}_{mn}^{ij}$	$v_{mn}^{ij} + \frac{1}{2} v_{mn}^{ef} \tau_{ef}^{ij} + \mathscr{A}^{ij} v_{nm}^{je} t_e^i$
$\bar{h}_{am}^{ie}$	$v_{am}^{ie} + v_{am}^{fe} t_f^i - \bar{h}_{nm}^{ie} t_a^n + v_{mn}^{ef} t_{af}^{in}$
$\bar{h}_{am}^{ij}$	$v_{am}^{ij} + \bar{h}_m^e t_{ae}^{ij} - \bar{h}_{nm}^{ij} t_a^n + \frac{1}{2} v_{am}^{ef} t_{ef}^{ij} + \mathscr{A}^{ij} (\bar{h}_{mn}^{jf} t_{af}^{in} + \chi'^{ie}_{\;\;am} t_e^j)$
$\bar{h}_{ab}^{ie}$	$v_{ab}^{ie} - \bar{h}_m^e t_{ab}^{im} + v_{ab}^{ef} t_f^i + \frac{1}{2} \bar{h}_{mn}^{ie} t_{ab}^{mn} - \mathscr{A}_{ab} (\chi_{am}^{ie} t_b^m - v_{bn}^{ef} t_{af}^{in})$
$\chi'^{ie}_{\;\;am}$	$v_{am}^{ie} + \frac{1}{2} v_{am}^{ef} t_e^i$
χ_{am}^{ie}	$\bar{h}_{am}^{ie} + \frac{1}{2} \bar{h}_{nm}^{ie} t_a^n$
τ_{ab}^{ij}	$t_{ab}^{ij} + \mathscr{A}^{ij} t_a^i t_b^j$

[a] In each expression, summation is carried out over repeated upper and lower indices.

Table C.2 The one- and two-body components of the intermediates $\overline{X}^{(2)}$ expressed in terms of the one- and two-body components of the cluster operator, EOM excitation operator, and the CCSD-like similarity-transformed Hamiltonian, entering Eq. (C.2) that are introduced in order to evaluate the contributions to the EOMCC(P) equations due to the three- and four-body components of the CC(P) similarity-transformed Hamiltonian. In this table, we drop the subscript μ labeling the one- and two-body components of R_μ and $\overline{X}^{(2)}_\mu$ so that $r_{\mu,a}^{\ \ i} \equiv r_a^i$, $r_{\mu,ab}^{\ \ ij} \equiv r_{ab}^{ij}$, $\bar{x}_{\mu,p}^{\ \ q} \equiv \bar{x}_p^q$, and $\bar{x}_{\mu,pq}^{\ \ rs} \equiv \bar{x}_{pq}^{rs}$.

Component of $\overline{X}^{(2)}_\mu$	Expression[a]
$\bar{x}_m^e$	$\bar{h}_{mn}^{ef} r_f^n$
$\bar{x}_j^i$	$\bar{h}_j^e r_e^i + \bar{h}_{jn}^{if} r_f^n + \frac{1}{2}\bar{h}_{jn}^{ef} r_{ef}^{in}$
$\bar{x}_a^b$	$-\bar{h}_m^b r_a^m + \bar{h}_{an}^{bf} r_f^n - \frac{1}{2}\bar{h}_{mn}^{bf} r_{af}^{mn}$
$\bar{x}_{ab}^{ef}$	$-\mathscr{A}_{ab}\bar{h}_{am}^{ef} r_b^m + \frac{1}{2}\bar{h}_{mn}^{ef} r_{ab}^{mn}$
$\bar{x}_{mn}^{ij}$	$\mathscr{A}^{ij}\bar{h}_{nm}^{je} r_e^i + \frac{1}{2}\bar{h}_{mn}^{ef} r_{ef}^{ij}$
$\bar{x}_{am}^{ie}$	$\bar{h}_{am}^{fe} r_f^i - \bar{h}_{nm}^{ie} r_a^n + \bar{h}_{mn}^{ef} r_{af}^{in}$
$\bar{x}_{am}^{ij}$	$\mathscr{A}^{ij}[\bar{h}_{mn}^{je} r_{ae}^{in} + \bar{h}_{am}^{ie} r_e^j] - \bar{h}_{nm}^{ij} r_a^n + \frac{1}{2}\bar{h}_{am}^{ef} r_{ef}^{ij}$
$\bar{x}_{ab}^{ie}$	$\bar{h}_{ab}^{fe} r_f^i + \frac{1}{2}\bar{h}_{mn}^{ie} r_{ab}^{mn} - \bar{x}_m^e t_{ab}^{im} + \mathscr{A}^{ab}[\bar{h}_{am}^{ie} r_b^m + \bar{h}_{am}^{fe} r_{fb}^{im}]$

[a] In each expression, summation is carried out over repeated upper and lower indices.

C.2 Evaluation of Term (I)

The first term in Eq. (C.3) is a generalized moment of the CCSD equations. In particular, when $|\Phi_K\rangle \equiv |\Phi_i^a\rangle$ or $|\Phi_K\rangle \equiv |\Phi_{ij}^{ab}\rangle$, the corresponding one- and two-body moments of the CCSD equations, $\mathfrak{M}_{0,a}^{\ i}$(CCSD) and $\mathfrak{M}_{0,ab}^{\ ij}$(CCSD), represent the standard CCSD amplitude equations. These terms are computed in a fully vectorized fashion (*i.e.*, avoiding the use of explicit loops) by taking advantage of efficient matrix transposition and multiplication routines provided by BLAS. Because our CC(P;Q) calculations aimed at converging the full CCSDT energetics include all singly and doubly excited determinants in the P spaces, evaluating $\mathfrak{M}_{0,a}^{\ i}$(CCSD) and $\mathfrak{M}_{0,ab}^{\ ij}$(CCSD) involves the conventional $n_o^2 n_u^4$ computational steps characterizing CCSD. When $|\Phi_K\rangle$ represents a triply excited determinant in the P space, $|\Phi_{ijk}^{abc}\rangle$, the generalized three-body moment of the CCSD equations is evaluated using

$$\mathfrak{M}_{0,abc}^{\ ijk}(\text{CCSD}) = \mathscr{A}^{ijk}\mathscr{A}_{abc}\frac{1}{4}(\sum_e \bar{h}_{ab}^{ie} t_{ec}^{jk} - \sum_m I_{am}^{ij} t_{bc}^{mk}), \tag{C.4}$$

where $\mathscr{A}^{pqr} \equiv \mathscr{A}_{pqr} = 1 - (pq) - (pr) - (qr) + (pqr) + (prq)$ is a three-index antisymmetrizer and $I_{am}^{ij} = \bar{h}_{am}^{ij} - \sum_e \bar{h}_m^e t_{ae}^{ij}$ [cf. Eqs. (59) and (62) of Ref. [151]]. Within every iteration of the CC(P) solution procedure, the $\bar{h}_p^q$, $\bar{h}_{pq}^{rs}$, and I_{am}^{ij} intermediates are precomputed and stored so that Eq. (C.4) can be evaluated for each triply excited determinant $|\Phi_{ijk}^{abc}\rangle \in \mathscr{H}^{(P)}$ via dot products between $\bar{h}_{ab}^{ie}$ and I_{am}^{ij} and the T_2 cluster amplitudes. This allows us to reduce the number of CPU operations associated with determining the contribution of $\mathfrak{M}_{0,abc}^{\ ijk}$(CCSD) to the CC($P$) equations relative to its CCSDT counterpart by a factor of (D/d), where D is the number of all triples and d is the number of triples included in the P space. As we will discuss later on in Section C.3.2.3, Eq. (C.4) is also used to absorb the contributions to the CC(P) equations due to three-body components of $\overline{H}_N^{(2)}$, which formally enter in term (II), with the help of suitable modifications to the quantities $\bar{h}_{ab}^{ie}$ and I_{am}^{ij}.

C.3 Evaluation of Term (II)

We now turn our attention to term (II) in Eq. (C.3), which represents the contributions to the CC(P) equations that are linear in $T_3^{(P)}$. In principle, term (II) is nothing more than a product between the matrix representation of $\overline{H}_N^{(2)}$ in the singles-triples ($\langle\Phi_i^a|\overline{H}_N^{(2)}|\Phi_{lmn}^{def}\rangle$),

doubles-triples ($\langle\Phi_{ij}^{ab}|\overline{H}_N^{(2)}|\Phi_{lmn}^{def}\rangle$), and triples-triples ($\langle\Phi_{ijk}^{abc}|\overline{H}_N^{(2)}|\Phi_{lmn}^{def}\rangle$) sectors of the P space and the vector of three-body cluster amplitudes. Precomputing the matrix $\overline{H}_N^{(2)}$ for use in a matrix-vector product each iteration is not practical for most problems due to the size of the singles-triples, doubles-triples, and triples-triples blocks. The strategy we adopt instead is to compute matrix elements of $\overline{H}_N^{(2)}$ on the fly, but in order to be efficient, this must be carried out such that we skip the evaluation of $\langle\Phi_K|\overline{H}_N^{(2)}|\Phi_{lmn}^{def}\rangle$ terms which are zero.

C.3.1 Projections onto Singly and Doubly Excited Determinants

In the case of the projections onto singly and doubly excited determinants, this can be accomplished using an amplitude-driven strategy, where we iterate over the triply excited determinants in the P space and evaluate all contributions to the CC(P) equations projected on singly and doubly excited determinants coming from a given amplitude t_{abc}^{ijk} corresponding to $|\Phi_{ijk}^{abc}\rangle \in \mathscr{H}^{(P)}$. This strategy is outlined in Algorithm C.1. There are four distinct diagrams in the CC(P) equations corresponding to projections onto singly and doubly excited determinants that involve $T_3^{(P)}$, referred to as Diagrams 1–4. The specific expressions used to evaluate these diagrams in Algorithm C.1 can be derived using standard many-body techniques. The only potentially murky point is the use of an additional summation over index permutations operators in lines 6, 12, and 19 of Algorithm C.1, which, from a diagrammatic perspective, exchanges the labels carried by internal lines connecting the Hamiltonian vertex and $T_3^{(P)}$ with those carried by external lines emanating from the $T_3^{(P)}$ vertex going into the projection. The presence of these additional permutations, which does not normally enter our diagrammatic considerations, is consequence of working with a permutationally unique set of triply excited determinants and cluster amplitudes [*i.e.*, the summations over permutations could removed if one includes the redundant triply excited determinants and corresponding amplitudes in the list used in the CC(P) procedure].

Algorithm C.1 Contributions to term (II) in Eq. (C.3) corresponding to projections onto $|\Phi_i^a\rangle$ and $|\Phi_{ij}^{ab}\rangle$, denoted as D_a^i and D_{ab}^{ij}, respectively.

for $|\Phi_{ijk}^{abc}\rangle \in \mathscr{H}^{(P)}$ **do** ▷ Loop over triply excited determinants in the P space

  Get $t_3 \leftarrow t_{abc}^{ijk}$ ▷ The three-body cluster amplitude associated with $|\Phi_{ijk}^{abc}\rangle$

  `// Permutations corresponding to (k/ij)(c/ab)`

  for $\mathscr{P} \in \{(1), (ik), (jk), (ac), (bc), (ik)(ac), (ik)(bc), (jk)(ac), (jk)(bc)\}$ **do**

    $\mathscr{P} D_a^i \leftarrow \mathscr{P} D_a^i + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}_{ij}^{ab} \times t_3$ ▷ Diagram 1

    $\mathscr{P} D_{ab}^{ij} \leftarrow \mathscr{P} D_{ab}^{ij} + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}_k^c \times t_3$ ▷ Diagram 2

  end for

  `// Permutations corresponding to (j/ik)(c/ab)`

  for $\mathscr{P} \in \{(1), (ij), (jk), (ac), (ab), (ij)(ac), (ij)(ab), (jk)(ac), (jk)(ab)\}$ **do**

    for $m = 1,\ n_o$ **do**

      $\mathscr{P} D_{ab}^{mj} \leftarrow \mathscr{P} D_{ab}^{mj} - (-1)^{\mathscr{P}} \mathscr{P} \bar{h}_{ik}^{mc} \times t_3$ ▷ Diagram 3

    end for

  end for

  `// Permutations corresponding to (k/ij)(b/ac)`

  for $\mathscr{P} \in \{(1), (ik), (jk), (ab), (bc), (ik)(ab), (ik)(bc), (jk)(ab), (jk)(bc)\}$ **do**

    for $e = 1,\ n_u$ **do**

      $\mathscr{P} D_{eb}^{ij} \leftarrow \mathscr{P} D_{eb}^{ij} + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}_{ek}^{ac} \times t_3$ ▷ Diagram 4

    end for

  end for

end for

for $|\Phi_{ij}^{ab}\rangle \in \mathscr{H}^{(P)}$ **do**

  $D_{ab}^{ij} \leftarrow \mathscr{A}^{ij} \mathscr{A}_{ab} D_{ab}^{ij}$ ▷ Antisymmetrize the two-body residual

end for

C.3.2 Projections onto Triply Excited Determinants

In the contributions to term (II) of Eq. (C.3) corresponding to projections onto triply excited determinants in the P space, the nonzero matrix elements of the $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{lmn}^{def} \rangle$, with $|\Phi_{ijk}^{abc}\rangle, |\Phi_{lmn}^{def}\rangle \in \mathscr{H}^{(P)}$, type are most clearly represented using the diagrams shown in Figure C.1, which involve the one- and two-body components of $\overline{H}_N^{(2)}$ [the contributions due to three-body components of $\overline{H}_N^{(2)}$ can be embedded within the computation of term (I) projected onto triply excited determinants; see Section C.3.2.3 for further details]. In Figure

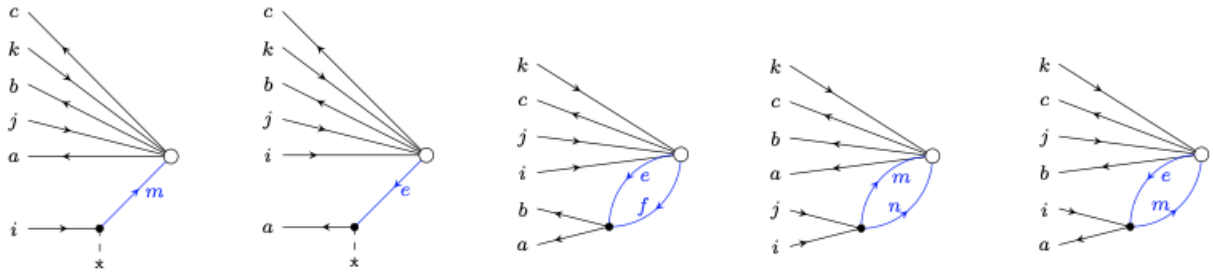

Figure C.1 Hugenholtz diagrams corresponding to the contributions in Eq. (C.3) originating from the one- and two-body components of $\overline{H}_N^{(2)}$. In these diagrams, the open circles are T_3 vertices, and interaction vertices represent the one- and two-body components of the CCSD-like similarity-transformed Hamiltonian $\overline{H}^{(2)}$.

C.1, the six black external lines in each diagram specify the determinant in the bra, $|\Phi_{ijk}^{abc}\rangle$, while the blue internal lines connecting the $\overline{H}_N^{(2)}$ and T_3 vertices, in addition to the black lines emanating from the $T_3^{(P)}$ and going into the projection, determine the corresponding kets $|\Phi_{lmn}^{def}\rangle \in \mathscr{H}^{(P)}$. As a consequence of the ket and bra sharing the indices carried by a subset of the black lines, the resulting matrix elements of $\overline{H}_N^{(2)}$ are guaranteed to be nonzero. To be more specific, the nonzero matrix element types corresponding to the diagram in Figure C.1, from left to right, are $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mjk}^{abc} \rangle$, $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{ijk}^{ebc} \rangle$, $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{ijk}^{efc} \rangle$, $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mnk}^{abc} \rangle$, and $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mjk}^{ebc} \rangle$, respectively. Thus, in order to efficiently construct the CC(P) equations, we must find a way to selectively iterate over pairs of triply excited determinants in the bra and ket that share some or all of their hole and particle indices.

C.3.2.1 Radix Sort Algorithm for the List of Triply Excited Determinants

In order to accomplish this task, we must be able to judiciously sort the list of triply excited determinants entering the CC(P) calculation without encountering exorbitant CPU or memory costs. One particularly clever way of doing this is based on a technique known as the radix sort. Radix sorting the list of triply excited determinants will allow us to organize determinants according to a subset of their hole and particle indices in such a way that we can directly identify connected pairs that share a particular combination of hole and particle indices. This strategy lies at the heart of our novel CC(P) algorithm.

To acquaint ourselves with the radix sorting procedure, suppose we want to arrange the numbers (53, 89, 150, 36, 633, 233) in ascending order. In the radix sort, we recursively sort the sequence of numbers according to a given set of *keys*. Different sorting tasks may require different definitions of the keys, but for the simple case of sorting nonnegative integers, the appropriate keys are the digits of each number, from least significant (rightmost) to most significant (leftmost). In this example, we must sort according to the least significant digit first and move toward the most significant digit in order to obtain a correctly sorted list in the end. Sorting the array according to the leftmost digit, we obtain (150, 53, 633, 233, 36, 89). Note that the individual sorting passes in the radix sort algorithm must be *stable*, meaning that objects with equivalent key values retain their original order. Sorting the list according to the middle digit results in (633, 233, 36, 150, 53, 89), and sorting according to the rightmost digit gives the sorted array (36, 53, 89, 150, 233, 633). It is worth making a few comments about the time complexity of the radix sort algorithm. In practice, each individual sorting pass is accomplished using a simple *counting sort* algorithm. The counting sort algorithm uses $n+k$ steps, where n is the number of objects being sorted according to a given key and k is the number of values that the key can take on. Since the radix sort runs the counting sort algorithm m times, where m is the number of keys, the time complexity of radix sort is $\mathcal{O}(m(n+k))$, assuming that each key has the same range of values (in the case of sorting numbers according to their digits, each key takes on values from 0–9). In

particular, the $\mathcal{O}(m(n+k))$ complexity of radix sort can be substantially less than the $\mathcal{O}(n \log n)$ complexity of more common state-of-the-art sorting algorithms if $m, k \ll n$.

Let us now switch to the context that interests us. In our CC(P) algorithm, we will ultimately want to perform a radix sort on the list of triply excited determinants entering the P space. In what follows, we define Φ_3 as a $d \times 6$ matrix containing the list of triply excited determinants in the P space, in which each row of Φ_3 corresponds to a particular triply excited determinant. Each triply excited determinant $|\Phi_{ijk}^{abc}\rangle \in \mathscr{H}^{(P)}$ is represented as a tuple of the particle indices followed by the hole indices, (a, b, c, i, j, k), where $a < b < c$ and $i < j < k$. For example, a system with 4 electrons ($i, j, k = 1, 2, 3,$ or 4) and 4 spatial orbitals ($a, b, c = 5, 6, 7,$ or 8) would have all triply excited determinants arranged in Φ_3 (in no particular order) as follows

$$\begin{bmatrix}
5 & 6 & 7 & 1 & 2 & 3 \\
5 & 6 & 7 & 1 & 3 & 4 \\
5 & 6 & 7 & 2 & 3 & 4 \\
5 & 6 & 7 & 1 & 2 & 4 \\
5 & 7 & 8 & 1 & 2 & 3 \\
5 & 7 & 8 & 1 & 3 & 4 \\
5 & 7 & 8 & 2 & 3 & 4 \\
5 & 7 & 8 & 1 & 2 & 4 \\
6 & 7 & 8 & 1 & 2 & 3 \\
6 & 7 & 8 & 1 & 3 & 4 \\
6 & 7 & 8 & 2 & 3 & 4 \\
6 & 7 & 8 & 1 & 2 & 4 \\
5 & 6 & 8 & 1 & 2 & 3 \\
5 & 6 & 8 & 1 & 3 & 4 \\
5 & 6 & 8 & 2 & 3 & 4 \\
5 & 6 & 8 & 1 & 2 & 4
\end{bmatrix}$$

With the above concrete representation of Φ_3, we now introduce the critical notion of $uvxw$-major sorting of Φ_3. In particular, the array Φ_3 is said to be sorted in $uvxw$-major order if the rows in Φ_3 characterized by the same values in columns u, v, x, and w appear consecutively one after another. In order to bring Φ_3 into $uvxw$-major order, we can perform a radix sort on Φ_3, where the keys are now the columns u, v, x, and w. Given the above form of Φ_3 for a 4-electron–4-orbital system, let us bring it to a 1235-major order using the radix sort. Unlike in the previous case of sorting nonnegative integers digit-by-digit, the order in which we sort the columns 1, 2, 3, and 5 does not matter. Therefore, we will proceed from left to right, starting by sorting column 1 and finishing with sorting column 5.

$$\begin{bmatrix}
5&6&7&1&2&3\\5&6&7&1&3&4\\5&6&7&2&3&4\\5&6&7&1&2&4\\5&7&8&1&2&3\\5&7&8&1&3&4\\5&7&8&2&3&4\\5&7&8&1&2&4\\6&7&8&1&2&3\\6&7&8&1&3&4\\6&7&8&2&3&4\\6&7&8&1&2&4\\5&6&8&1&2&3\\5&6&8&1&3&4\\5&6&8&2&3&4\\5&6&8&1&2&4
\end{bmatrix}
\rightarrow
\begin{bmatrix}
5&6&7&1&2&3\\5&6&7&1&3&4\\5&6&7&2&3&4\\5&6&7&1&2&4\\5&7&8&1&2&3\\5&7&8&1&3&4\\5&7&8&2&3&4\\5&7&8&1&2&4\\5&6&8&1&2&3\\5&6&8&1&3&4\\5&6&8&2&3&4\\5&6&8&1&2&4\\6&7&8&1&2&3\\6&7&8&1&3&4\\6&7&8&2&3&4\\6&7&8&1&2&4
\end{bmatrix}
\rightarrow
\begin{bmatrix}
5&6&7&1&2&3\\5&6&7&1&3&4\\5&6&7&2&3&4\\5&6&7&1&2&4\\5&6&8&1&2&3\\5&6&8&1&3&4\\5&6&8&2&3&4\\5&6&8&1&2&4\\5&7&8&1&2&3\\5&7&8&1&3&4\\5&7&8&2&3&4\\5&7&8&1&2&4\\6&7&8&1&2&3\\6&7&8&1&3&4\\6&7&8&2&3&4\\6&7&8&1&2&4
\end{bmatrix}
\rightarrow
\begin{bmatrix}
5&6&7&1&2&3\\5&6&7&1&3&4\\5&6&7&2&3&4\\5&6&7&1&2&4\\5&6&8&1&2&3\\5&6&8&1&3&4\\5&6&8&2&3&4\\5&6&8&1&2&4\\5&7&8&1&2&3\\5&7&8&1&3&4\\5&7&8&2&3&4\\5&7&8&1&2&4\\6&7&8&1&2&3\\6&7&8&1&3&4\\6&7&8&2&3&4\\6&7&8&1&2&4
\end{bmatrix}
\rightarrow
\begin{bmatrix}
5&6&7&1&2&3\\5&6&7&1&2&4\\5&6&8&1&2&3\\5&6&8&1&3&4\\5&6&8&2&3&4\\5&6&8&1&2&4\\5&7&8&1&2&3\\5&7&8&1&3&4\\5&7&8&2&3&4\\5&7&8&1&2&4\\6&7&8&1&2&3\\6&7&8&1&3&4\\6&7&8&2&3&4\\6&7&8&1&3&4\\6&7&8&2&3&4\\6&7&8&1&2&4
\end{bmatrix}$$

In each step, we have permuted the rows in Φ_3 such that the indices in the red columns are sorted. By the end of the radix sort, we obtain the matrix Φ_3 in the 1235-major order, in which all the rows with the same values appearing in columns 1, 2, 3, and 5 are placed next to one another. We can perform this $uvxw$-major sorting for any combination of (distinct) keys defined by the integers u, v, x, and w, where u, v, x, w = 1–6. Based on the previous analysis of the radix sort, we expect the time complexity of these sorting steps to scale as $\mathcal{O}(4(d+n_u))$, where n_u is the number of unoccupied orbitals in the problem, which represents the largest range of values that can be taken on by any key in the radix sort. Given that the number of triply excited determinants in the P space, d, is much greater than n_u, the radix sort roughly has complexity $\mathcal{O}(4d)$, which is typically less than $\mathcal{O}(d \log d)$ resulting from conventional sorting algorithms.

In the above example, we also notice that as a result of sorting Φ_3 in the 1235-major order, Φ_3 can be partitioned into blocks, where each block is associated with a specific combination of the values appearing columns 1, 2, 3, and 5. For example, the first two rows of Φ_3 above form a block characterized by the values 5, 6, 7, and 2 in columns 1, 2, 3, and 5, respectively. Similarly, rows 3 and 4 form another block with the values 5, 6, 8 and 2 in columns 1, 2, 3, and 5, respectively. In this way, we can find 8 distinct blocks in the above Φ_3, which are associated with the respective tuples of values (5,6,7,2), (5,6,8,2), (5,7,8,2), (6,7,8,2), (5,6,7,3), (5,6,8,3), (5,7,8,3), and (6,7,8,3) in columns 1, 2, 3, and 5. Note that the number of rows within each block are associated with the different combinations of indices

that appear in the unsorted columns, which are columns 4 and 6 in this case. Given this blocked structure of Φ_3, it is useful to number each block according to a lexical index denoted as y. In the above Φ_3 for example, the block characterized by (5,6,7,2) can be assigned $y = 1$, and the next block, associated with (5,6,8,2), can be $y = 2$, and so on. In the end we will map each block to a particular lexical index $y = 1$–8. In a general $uvxw$-sorting, the lexical indices y range between 1 and the total number of $uvxw$-blocks that can exist within the list Φ_3, which we denote as n_{uvxw}.

It follows from this discussion that if each block in Φ_3 is associated with a particular lexical index y, and if the determinants in a given block share the values in columns u, v, x, and w, which we arbitrarily denote as p, q, r, and s, respectively, then there must exist a one-to-one mapping between the tuple (p, q, r, s) and the lexical index y. In our considerations, it will be helpful to know this mapping in order to move back and forth between tuples (p, q, r, s) and corresponding indices y. We will denote this mapping as $y = f_{uvxw}(p, q, r, s)$. In particular, the function $f_{uvxw}(p, q, r, s)$ returns the lexical index y associated with the block in Φ_3 containing rows in which the column u has value p, column v has value q, column x has value r, and column w as value s. For a general list of triply excited determinants Φ_3, determining a computable formula for the mapping function $f_{uvxw}(p, q, r, s)$ is not straightforward, and may even be impossible depending on the nature of Φ_3 (this amounts to finding a perfect hashing function). However, we can pre-assign the lexical indices corresponding to all possible tuples (p, q, r, s) that can appear in list Φ_3 and store them in $f_{uvxw}(p, q, r, s)$, effectively treating $f_{uvxw}(p, q, r, s)$ as a lookup table that returns the lexical index y for a given combination of p, q, r, and s values. Thus, in our implementation, the function $f_{uvxw}(p, q, r, s)$ is represented as a four-dimensional array, where each dimension of the array is associated with a particular argument. In this way, we can trivially retrieve the lexical index of the block in Φ_3 sorted in $uvxw$-major order associated with a given p, q, r, and s. Adopting the previous example of 1235-major sorting, the array f_{1235} is an $n_u \times n_u \times n_u \times n_o$ array. Based on our assignment of the lexical indices $y = 1$–8, we would

have $f_{1235}(5,6,7,2) = 1$, $f_{1235}(5,6,8,2) = 2$, $f_{1235}(5,7,8,2) = 3$, and so on.

The last piece of information we need to efficiently access the sorted determinants in Φ_3 is a final auxiliary array, which we denote as A_{uvxw}. The array A_{uvxw} has length equal to the number of lexical indices associated with $uvxw$-major sorting, n_{uvxw}. The entry at $A_{uvxw}(y)$ gives the row index in Φ_3 that starts the sorted block associated with lexical index y. The ending row of a given block characterized by y is *de facto* $A_{uvxw}(y+1)-1$, where we enforce that $A_{uvxw}(n_{uvxw}) = d+1$. Again, in our previous example of 1235-major sorting, $A_{1235}(1) = 1$, $A_{1235}(2) = 3$, $A_{1235}(3) = 5$, ..., $A_{1235}(7) = 15$, and $A_{1235}(8) = 17$. Note that the arrays $f_{uvxw}(p,q,r,s)$ and A_{uvxw} are meant to be used in tandem. Thus, for a given (p,q,r,s), one can use $f_{uvxw}(p,q,r,s)$ to retrieve the lexical index y and then identify the rows in Φ_3 lying between $A_{uvxw}(y)$ and $A_{uvxw}(y+1)-1$ (inclusively) to obtain those triply excited determinants that possess indices p, q, r, and s in column positions u, v, x, and w.

C.3.2.2 Contributions Due to One- and Two-Body Components of $\overline{H}_N^{(2)}$

Armed with an understanding of how we can use the radix sort to organize the list of triply excited determinants in any $uvxw$-major order and retrieve any desired block of triply excited determinants in it, we can apply it to form and evaluate matrix elements of the $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mjk}^{abc} \rangle$, $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{ijk}^{ebc} \rangle$, $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{ijk}^{efc} \rangle$, $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mnk}^{abc} \rangle$, and $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mjk}^{ebc} \rangle$ types corresponding to the diagrams in Figure C.1. Let us focus on matrix elements of the $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{ijk}^{efc} \rangle$ type corresponding to the particle-particle ladder diagram in Figure C.1, which represents the most expensive term in CCSDT. For this matrix element, we must sort Φ_3 in a 3456-major order. However, we should be mindful of the fact that we are working with permutationally unique lists of triply excited determinants. Therefore, in addition to 3456-major sorting, we should also consider contributions originating from 1456- and 2456-major sorting configurations as well. Similarly, matrix elements of the $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mnk}^{abc} \rangle$ type will require sorting Φ_3 in 1234-, 1235-, and 1236-major order and matrix elements of the $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{mjk}^{ebc} \rangle$ category involve sorting Φ_3 in nine different ways: 2356-, 2345-, 2346-, 1356-, 1345-, 1346-, 1256-, 1245-, and 1246-major order. The matrix elements of the

$\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}_N|\Phi^{abc}_{mjk}\rangle$ and $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}_N|\Phi^{ebc}_{ijk}\rangle$ types technically involve sorting with respect to five indices. This can be done, but it requires storing 5-dimensional arrays in the lookup table f. To avoid creation of prohibitively large auxiliary arrays, we perform the computation of $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}_N|\Phi^{abc}_{mjk}\rangle$ and $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}_N|\Phi^{ebc}_{ijk}\rangle$ while evaluating $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}_N|\Phi^{abc}_{mnk}\rangle$ and $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}_N|\Phi^{efc}_{ijk}\rangle$, respectively (see Algorithms C.2 and C.3). Our implementation of the five diagrams in Figure C.1 defining the contributions to term (II) in Eq. C.3 corresponding to projections onto triply excited determinants resulting from one- and two-body components of $\overline{H}^{(2)}_N$ are given in Algorithms C.2–C.6. As in the case of the expressions given in Algorithm C.1, the structure of the expressions match what one would obtain using many-body diagrammatics, except that additional loops over permutations of the indices attached to internal lines and their external counterparts emanating from $T_3^{(P)}$ are added to accommodate the use of a permutationally unique set of triply excited determinants in Φ_3.

C.3.2.3 Contributions Due to Three-Body Components of $\overline{H}^{(2)}_N$

Compared to the contributions due to one- and two-body components of $\overline{H}^{(2)}_N$ in term (II) of Eq. (C.3) corresponding to projection onto triply excited determinants, the strategy for handling the three-body components of $\overline{H}^{(2)}_N$ is, fortunately, much simpler. The only three-body components of $\overline{H}^{(2)}_N$ that we must explicitly consider arise from contributions of the $\langle\Phi^{abc}_{ijk}|(H_N T_2 T_3^{(P)})_C|\Phi\rangle$ type. The best approach to handling these terms results from first contracting H_N with $T_3^{(P)}$ to form $\eta^{ie}_{ab} = -\frac{1}{2}\bar{h}^{ef}_{mn}t^{imn}_{abf}$ and $\eta^{ij}_{am} = \frac{1}{2}\bar{h}^{ef}_{mn}t^{ijn}_{aef}$. Then, after defining $I^{ie}_{ab} = \bar{h}^{ie}_{ab} + \eta^{ie}_{ab}$ and $I'^{ij}_{\ \ am} = I^{ij}_{am} + \eta^{ij}_{am}$, where $\bar{h}^{ie}_{ab}$ and I^{ij}_{am} are the intermediates entering Eq. (C.4), we can embed the contributions due to three-body components of $\overline{H}^{(2)}_N$ within the evaluation of Eq. (C.4) by replacing $\bar{h}^{ie}_{ab}$ and I^{ij}_{am} by I^{ie}_{ab} and $I'^{ij}_{\ \ am}$, respectively. All that remains is finding an efficient strategy for computing η^{ie}_{ab} and η^{ij}_{am} given the spotty subset of triply excited determinants defining $T_3^{(P)}$. Fortunately, these terms can be evaluated using the same amplitude-driven method adopted in Algorithm C.1. The routines for evaluating η^{ie}_{ab} and η^{ij}_{am}, which help us embed the contributions of three-body components of $\overline{H}^{(2)}_N$ into the computation of Eq. (C.4), are shown in Algorithm C.7.

Algorithm C.2 Particle–particle ladder diagram contributions to term (II) in Eq. (C.3) corresponding to projections onto $|\Phi_{ijk}^{abc}\rangle \in \mathscr{H}^{(P)}$. The resulting projections are stored in the $d \times 1$ array r_3, where d is the number of triply excited determinants in the P space.

for $I = 1, \ldots, d$ **do** ▷ Loop over projections $\langle\Phi_{ijk}^{abc}|$

  $a = \Phi_3(I,1)$ $b = \Phi_3(I,2)$ $c = \Phi_3(I,3)$
  $i = \Phi_3(I,4)$ $j = \Phi_3(I,5)$ $k = \Phi_3(I,6)$

  `// Sorting order 1456 [permutations correspond to (a/bc)]`
  for $\mathscr{P} \in \{(1), (ab), (ac)\}$ **do**
    $a' = \mathscr{P}a$
    $y = f_{1456}(a', i, j, k)$
    for $J = A_{1456}(y), A_{1456}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi_{ijk}^{a'ef}\rangle$
      $e = \Phi_3(J,2)$
      $f = \Phi_3(J,3)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}_{bc}^{ef} \times t_3(J)$ ▷ Evaluate $\langle\Phi_{ijk}^{abc}|\overline{H}^{(2)}|\Phi_{ijk}^{a'ef}\rangle$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{A}_{bc}\mathscr{A}_{ef}[\mathscr{P}\,(\bar{h}_c^f \delta_{be}) \times t_3(J)]$
    end for
  end for

  `// Sorting order 2456 [permutations correspond to (b/ac)]`
  for $\mathscr{P} \in \{(1), (ab), (bc)\}$ **do**
    $b' = \mathscr{P}b$
    $y = f_{2456}(b', i, j, k)$
    for $J = A_{2456}(y), A_{2456}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi_{ijk}^{eb'f}\rangle$
      $e = \Phi_3(J,1)$
      $f = \Phi_3(J,3)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}_{ac}^{ef} \times t_3(J)$ ▷ Evaluate $\langle\Phi_{ijk}^{abc}|\overline{H}^{(2)}|\Phi_{ijk}^{eb'f}\rangle$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{A}_{ac}\mathscr{A}_{ef}[\mathscr{P}\,(\bar{h}_c^f \delta_{ae}) \times t_3(J)]$
    end for
  end for

  `// Sorting order 3456 [permutations correspond to (c/ab)]`
  for $\mathscr{P} \in \{(1), (ac), (bc)\}$ **do**
    $c' = \mathscr{P}c$
    $y = f_{3456}(c', i, j, k)$
    for $J = A_{3456}(y), A_{3456}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi_{ijk}^{efc'}\rangle$
      $e = \Phi_3(J,1)$
      $f = \Phi_3(J,2)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}_{ab}^{ef} \times t_3(J)$ ▷ Evaluate $\langle\Phi_{ijk}^{abc}|\overline{H}^{(2)}|\Phi_{ijk}^{efc'}\rangle$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{A}_{ab}\mathscr{A}_{ef}[\mathscr{P}\,(\bar{h}_b^f \delta_{ae}) \times t_3(J)]$
    end for
  end for
end for

Algorithm C.3 Hole–hole ladder diagram contributions to term (II) in Eq. (C.3) corresponding to projections onto $|\Phi^{abc}_{ijk}\rangle \in \mathscr{H}^{(P)}$. The resulting projections are stored in the $d \times 1$ array r_3, where d is the number of triply excited determinants in the P space.

for $I = 1, \ldots, d$ **do** ▷ Loop over projections $\langle\Phi^{abc}_{ijk}| \in \mathscr{H}^{(P)}$

  $a = \Phi_3(I,1)$ $b = \Phi_3(I,2)$ $c = \Phi_3(I,3)$
  $i = \Phi_3(I,4)$ $j = \Phi_3(I,5)$ $k = \Phi_3(I,6)$

  `// Sorting order 1234 [permutations correspond to (i/jk)]`
  for $\mathscr{P} \in \{(1), (ij), (ik)\}$ **do**
    $i' = \mathscr{P} i$
    $y = f_{1234}(a, b, c, i')$
    for $J = A_{1234}(y), A_{1234}(y+1) - 1$ **do** ▷ Loop $|\Phi^{abc}_{i'mn}\rangle$
      $m = \Phi_3(J,5)$
      $n = \Phi_3(J,6)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{jk}_{mn} \times t_3(J)$ ▷ Evaluate $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}|\Phi^{abc}_{i'mn}\rangle$
      $r_3(I) \leftarrow r_3(I) - (-1)^{\mathscr{P}} \mathscr{A}^{jk} \mathscr{A}^{mn} [\mathscr{P} (\bar{h}^{j}_{m} \delta_{nk}) \times t_3(J)]$
    end for
  end for

  `// Sorting order 1235 [permutations correspond to (j/ik)]`
  for $\mathscr{P} \in \{(1), (ij), (jk)\}$ **do**
    $j' = \mathscr{P} j$
    $y = f_{1235}(a, b, c, j')$
    for $J = A_{1235}(y), A_{1235}(y+1) - 1$ **do** ▷ Loop $|\Phi^{abc}_{mj'n}\rangle$
      $m = \Phi_3(J,4)$
      $n = \Phi_3(J,6)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ik}_{mn} \times t_3(J)$ ▷ Evaluate $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}|\Phi^{abc}_{mj'n}\rangle$
      $r_3(I) \leftarrow r_3(I) - (-1)^{\mathscr{P}} \mathscr{A}^{ik} \mathscr{A}^{mn} [\mathscr{P} (\bar{h}^{i}_{m} \delta_{nk}) \times t_3(J)]$
    end for
  end for

  `// Sorting order 1236 [permutations correspond to (k/ij)]`
  for $\mathscr{P} \in \{(1), (ik), (jk)\}$ **do**
    $k' = \mathscr{P} k$
    $y = f_{1236}(a, b, c, k')$
    for $J = A_{1236}(y), A_{1236}(y+1) - 1$ **do** ▷ Loop $|\Phi^{abc}_{mnk'}\rangle$
      $m = \Phi_3(J,4)$
      $n = \Phi_3(J,5)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ij}_{mn} \times t_3(J)$ ▷ Evaluate $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}|\Phi^{abc}_{mnk'}\rangle$
      $r_3(I) \leftarrow r_3(I) - (-1)^{\mathscr{P}} \mathscr{A}^{ij} \mathscr{A}^{mn} [\mathscr{P} (\bar{h}^{i}_{m} \delta_{nj}) \times t_3(J)]$
    end for
  end for
end for

Algorithm C.4 Particle–hole ring diagram contributions to term (II) in Eq. (C.3) corresponding to projections onto $|\Phi^{abc}_{ijk}\rangle \in \mathscr{H}^{(P)}$. The resulting projections are stored in the $d \times 1$ array r_3, where d is the number of triply excited determinants in the P space. [Part 1 of 3]

for $I = 1, \ldots, d$ **do** ▷ Loop over projections $\langle\Phi^{abc}_{ijk}| \in \mathscr{H}^{(P)}$

  $a = \Phi_3(I,1)$ $b = \Phi_3(I,2)$ $c = \Phi_3(I,3)$
  $i = \Phi_3(I,4)$ $j = \Phi_3(I,5)$ $k = \Phi_3(I,6)$

  `// Sorting order 2356 [permutations correspond to (a/bc)(i/jk)]`
  for $\mathscr{P} \in \{(1), (ij), (ik), (ab), (ac), (ij)(ab), (ij)(ac), (ik)(ab), (ik)(ac)\}$ **do**
    $j' = \mathscr{P}j$ $k' = \mathscr{P}k$
    $b' = \mathscr{P}b$ $c' = \mathscr{P}c$
    $y = f_{2356}(b', c', j', k')$
    for $J = A_{2356}(y), A_{2356}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{eb'c'}_{mj'k'}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,1)$
      $m = \Phi_3(J,4)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ie}_{am} \times t_3(J)$ ▷ Evaluate $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}|\Phi^{eb'c'}_{mj'k'}\rangle$
    end for
  end for

  `// Sorting order 2346 [permutations correspond to (a/bc)(j/ik)]`
  for $\mathscr{P} \in \{(1), (ij), (jk), (ab), (ac), (ij)(ab), (ij)(ac), (jk)(ab), (jk)(ac)\}$ **do**
    $i' = \mathscr{P}i$ $k' = \mathscr{P}k$
    $b' = \mathscr{P}b$ $c' = \mathscr{P}c$
    $y = f_{2346}(b', c', i', k')$
    for $J = A_{2346}(y), A_{2346}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{eb'c'}_{i'mk'}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,1)$
      $m = \Phi_3(J,5)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{je}_{am} \times t_3(J)$ ▷ Evaluate $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}|\Phi^{eb'c'}_{i'mk'}\rangle$
    end for
  end for

  `// Sorting order 2345 [permutations correspond to (a/bc)(k/ij)]`
  for $\mathscr{P} \in \{(1), (ik), (jk), (ab), (ac), (ik)(ab), (ik)(ac), (jk)(ab), (jk)(ac)\}$ **do**
    $i' = \mathscr{P}i$ $j' = \mathscr{P}j$
    $b' = \mathscr{P}b$ $c' = \mathscr{P}c$
    $y = f_{2345}(b', c', i', j')$
    for $J = A_{2345}(y), A_{2345}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{eb'c'}_{i'j'm}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,1)$
      $m = \Phi_3(J,6)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ke}_{am} \times t_3(J)$ ▷ Evaluate $\langle\Phi^{abc}_{ijk}|\overline{H}^{(2)}|\Phi^{eb'c'}_{i'j'm}\rangle$
    end for
  end for
end for

Algorithm C.5 Same as Algorithm C.4. [Part 2 of 3]

for $I = 1, \ldots, d$ **do** ▷ Loop over projections $\langle \Phi^{abc}_{ijk}| \in \mathscr{H}^{(P)}$

  $a = \Phi_3(I,1)$ $b = \Phi_3(I,2)$ $c = \Phi_3(I,3)$
  $i = \Phi_3(I,4)$ $j = \Phi_3(I,5)$ $k = \Phi_3(I,6)$

  `// Sorting order 1356 [permutations correspond to (b/ac)(i/jk)]`
  for $\mathscr{P} \in \{(1), (ij), (ik), (ab), (bc), (ij)(ab), (ij)(bc), (ik)(ab), (ik)(bc)\}$ **do**
    $j' = \mathscr{P} j$ $k' = \mathscr{P} k$
    $a' = \mathscr{P} a$ $c' = \mathscr{P} c$
    $y = f_{1356}(a', c', j', k')$
    for $J = A_{1356}(y), A_{1356}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{a'ec'}_{mj'k'}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,2)$
      $m = \Phi_3(J,4)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ie}_{bm} \times t_3(J)$ ▷ Evaluate $\langle \Phi^{abc}_{ijk}| \overline{H}^{(2)} |\Phi^{a'ec'}_{mj'k'}\rangle$
    end for
  end for

  `// Sorting order 1346 [permutations correspond to (b/ac)(j/ik)]`
  for $\mathscr{P} \in \{(1), (ij), (jk), (ab), (bc), (ij)(ab), (ij)(bc), (jk)(ab), (jk)(bc)\}$ **do**
    $i' = \mathscr{P} i$ $k' = \mathscr{P} k$
    $a' = \mathscr{P} a$ $c' = \mathscr{P} c$
    $y = f_{1346}(a', c', i', k')$
    for $J = A_{1346}(y), A_{1346}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{a'mc'}_{i'mk'}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,2)$
      $m = \Phi_3(J,5)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{je}_{bm} \times t_3(J)$ ▷ Evaluate $\langle \Phi^{abc}_{ijk}| \overline{H}^{(2)} |\Phi^{a'ec'}_{i'mk'}\rangle$
    end for
  end for

  `// Sorting order 1345 [permutations correspond to (b/ac)(k/ij)]`
  for $\mathscr{P} \in \{(1), (ik), (jk), (ab), (ac), (ik)(ab), (ik)(ac), (jk)(ab), (jk)(ac)\}$ **do**
    $i' = \mathscr{P} i$ $j' = \mathscr{P} j$
    $a' = \mathscr{P} a$ $c' = \mathscr{P} c$
    $y = f_{1345}(a', c', i', j')$
    for $J = A_{1345}(y), A_{1345}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{a'ec'}_{i'j'm}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,2)$
      $m = \Phi_3(J,6)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ke}_{bm} \times t_3(J)$ ▷ Evaluate $\langle \Phi^{abc}_{ijk}| \overline{H}^{(2)} |\Phi^{a'ec'}_{i'j'm}\rangle$
    end for
  end for
end for

Algorithm C.6 Same as Algorithm C.4. [Part 3 of 3]

for $I = 1, \ldots, d$ **do** ▷ Loop over projections $\langle \Phi^{abc}_{ijk}| \in \mathscr{H}^{(P)}$

  $a = \Phi_3(I,1)$ $b = \Phi_3(I,2)$ $c = \Phi_3(I,3)$
  $i = \Phi_3(I,4)$ $j = \Phi_3(I,5)$ $k = \Phi_3(I,6)$

  `// Sorting order 1256 [permutations correspond to (c/ab)(i/jk)]`
  for $\mathscr{P} \in \{(1), (ij), (ik), (ac), (bc), (ij)(ac), (ij)(bc), (ik)(ac), (ik)(bc)\}$ **do**
    $j' = \mathscr{P} j$ $k' = \mathscr{P} k$
    $a' = \mathscr{P} a$ $b' = \mathscr{P} b$
    $y = f_{1256}(a', b', j', k')$
    for $J = A_{1256}(y), A_{1256}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{a'b'e}_{mj'k'}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,3)$
      $m = \Phi_3(J,4)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ie}_{cm} \times t_3(J)$ ▷ Evaluate $\langle \Phi^{abc}_{ijk} | \overline{H}^{(2)} | \Phi^{a'b'e}_{mj'k'} \rangle$
    end for
  end for

  `// Sorting order 1246 [permutations correspond to (c/ab)(j/ik)]`
  for $\mathscr{P} \in \{(1), (ij), (jk), (ac), (bc), (ij)(ac), (ij)(bc), (jk)(ac), (jk)(bc)\}$ **do**
    $i' = \mathscr{P} i$ $k' = \mathscr{P} k$
    $a' = \mathscr{P} a$ $b' = \mathscr{P} b$
    $y = f_{1246}(a', b', i', k')$
    for $J = A_{1246}(y), A_{1246}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{a'b'e}_{i'mk'}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,3)$
      $m = \Phi_3(J,5)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{je}_{cm} \times t_3(J)$ ▷ Evaluate $\langle \Phi^{abc}_{ijk} | \overline{H}^{(2)} | \Phi^{a'b'e}_{i'mk'} \rangle$
    end for
  end for

  `// Sorting order 1245 [permutations correspond to (c/ab)(k/ij)]`
  for $\mathscr{P} \in \{(1), (ik), (jk), (ac), (bc), (ik)(ac), (ik)(bc), (jk)(ac), (jk)(bc)\}$ **do**
    $i' = \mathscr{P} i$ $j' = \mathscr{P} j$
    $a' = \mathscr{P} a$ $b' = \mathscr{P} b$
    $y = f_{1245}(a', b', i', j')$
    for $J = A_{1245}(y), A_{1245}(y+1) - 1$ **do** ▷ Loop over kets $|\Phi^{a'b'e}_{i'j'm}\rangle \in \mathscr{H}^{(P)}$
      $e = \Phi_3(J,3)$
      $m = \Phi_3(J,6)$
      $r_3(I) \leftarrow r_3(I) + (-1)^{\mathscr{P}} \mathscr{P} \bar{h}^{ke}_{cm} \times t_3(J)$ ▷ Evaluate $\langle \Phi^{abc}_{ijk} | \overline{H}^{(2)} | \Phi^{a'b'e}_{i'j'm} \rangle$
    end for
  end for
end for

Algorithm C.7 Evaluation of η_{ab}^{ie} and η_{am}^{ij} intermediates resulting from factorization of three-body components of $\overline{H}_N^{(2)}$

for $|\Phi_{ijk}^{abc}\rangle \in \mathscr{H}^{(P)}$ **do** ▷ Loop over triply excited determinants in the P space

  Get $t_3 \leftarrow t_{abc}^{ijk}$ ▷ The three-body cluster amplitude associated with $|\Phi_{ijk}^{abc}\rangle$

  `// Permutations corresponding to (j/ik)(c/ab)`

  for $\mathscr{P} \in \{(1), (ij), (jk), (ac), (ab), (ij)(ac), (ij)(ab), (jk)(ac), (jk)(ab)\}$ **do**

    for $m = 1,\ n_o$ **do**

      $\mathscr{P}\eta_{am}^{ik} \leftarrow \mathscr{P}\eta_{am}^{ik} + (-1)^{\mathscr{P}}\mathscr{P}\bar{h}_{mj}^{ab} \times t_3$ ▷ Evaluate $\eta_{am}^{ij} = \frac{1}{2}\bar{h}_{mn}^{ef}t_{aef}^{ijn}$

    end for

  end for

  `// Permutations corresponding to (k/ij)(b/ac)`

  for $\mathscr{P} \in \{(1), (ik), (jk), (ab), (bc), (ik)(ab), (ik)(bc), (jk)(ab), (jk)(bc)\}$ **do**

    for $e = 1,\ n_u$ **do**

      $\mathscr{P}\eta_{ca}^{ke} \leftarrow \mathscr{P}\eta_{ca}^{ke} - (-1)^{\mathscr{P}}\mathscr{P}\bar{h}_{ij}^{eb} \times t_3$ ▷ Evaluate $\eta_{ab}^{ie} = -\frac{1}{2}\bar{h}_{mn}^{ef}t_{abf}^{imn}$

    end for

  end for

end for

C.4 Computational Efficiency and Benefits of the Novel CC(P)/EOMCC(P) Algorithms

Finally, we make a few remarks about the computational costs and additional benefits characterizing our novel CC(P)/EOMCC(P) algorithm driven by radix sorting of the list of triply excited determinants. As stated earlier, the cost of the radix sorting operations scale linearly with d, with a small prefactor. For the most expensive particle-particle ladder diagram in CCSDT, the CC(P) algorithm based on computing matrix elements of the $\langle \Phi_{ijk}^{abc} | \overline{H}_N^{(2)} | \Phi_{ijk}^{efc} \rangle$ type will involve $D \times n_u^2$, or $n_o^3 n_u^5$, floating point operations in the limit that all triply excited determinants are included in Φ_3, *i.e.*, $(d/D) \rightarrow 1$. In fact, the $n_o^3 n_u^5$ operation count of the CC(P) method using the complete manifold of triply excited determinants in the P space exactly matches the costs of CCSDT. The same is true for the EOMCC(P) calculations, which aim at converging EOMCCSDT. In practice, the CC(P)/EOMCC(P) computations are carried out with $(d/D) \ll 1$ (*e.g.*, using a few percent of triply excited determinants or less), and in this regime, the cost of the CC(P)/EOMCC(P) calculations is roughly $d \times \left(\frac{d}{D} n_u\right)^2$. Therefore, we can expect reductions in CPU effort compared to the canonical $n_o^3 n_u^5$ cost of CCSDT or EOMCCSDT by a factor on the order of $(d/D)^2$.

There are also a number of additional benefits offered by this general approach to solving the CC(P)/EOMCC(P) equations, which we will briefly comment on as well. First of all, the resulting CC(P)/EOMCC(P) routines are optimal in both CPU operation count and storage, and they can work with any list of triply excited determinants. This not only includes the spotty subsets of triply excited determinants obtained using semi-stochastic, CIPSI-based, or adaptive CC(P;Q) methodologies, but also the more structured lists adopted in active-orbital-based CC(t;3) calculations. Alternative choices of lists can also be used to perform different types of CC(P;Q) calculations not considered in this dissertation, such as core-valence-separated EOMCC schemes, which are useful for studying core excitation and ionization spectra. Another advantage offered by our CC(P)/EOMCC(P) routine is that it works directly with permutationally unique lists of triply excited determinants. This

helps reduce the costs associated with storing the t_{abc}^{ijk} and $r_{\mu,abc}^{ijk}$ arrays by only retaining the unique elements with $i < j < k$ and $a < b < c$ (the use of such triangular arrays would not be compatible with the optimized BLAS routines adopted in most conventional CC/EOMCC codes). Finally, by including only those triply excited determinants of a particular irrep, the CC(P)/EOMCC(P) routine can trivially take advantage of speedups offered by molecular point group symmetry, which is usually more difficult to implement in high-level CC/EOMCC codes.

To illustrate these benefits and highlight the efficiency of our novel CC(P) algorithm, we compare the single-core CPU timings characterizing the active-orbital-based CC(t;3) calculations for the stretched F_2 molecule in the aug-cc-pVQZ [10, 358] basis set obtained using CCpy and GAMESS in Table C.3. Our calculations used an active space consisting of the five highest-energy occupied orbitals, corresponding to the π_u, π_g, and σ_g shells, and the lowest-energy σ_u orbital unoccupied in the RHF reference. To demonstrate the capability of our CC(P) code to take advantage of speedups offered by molecular point-group symmetry, we performed CC(t;3) calculations using CCpy based on two different lists of triply excited determinants. The first list included all triply excited determinants selected with the help of active orbitals, consistent with the truncation of T_3 adopted in CCSDt, while the second list retained the subset of triply excited determinants present in the first list belonging to the $S_z = 0$ $A_g(D_{2h})$ symmetry. The CCSDt and CC(t;3) calculations performed with GAMESS relies on the highly efficient and automatically generated vectorized codes, which cannot take advantage of point-group symmetry at this time.

Upon comparing the timings reported in Table C.3, it is clear that CCpy is competitive with GAMESS in performing the CC(t;3) calculations. Indeed, the CPU timing obtained with GAMESS is within 5% of its counterpart resulting from the CCpy calculations that do not advantage of the D_{2h} Abelian symmetry of F_2. More importantly, the timings per iteration of the CCSDt calculations executed using CCpy, without using symmetry, and GAMESS differ by just 1.4 minutes. This is impressive, especially when we consider that

the CCSDt code in CCpy adopts a devectorized strategy for constructing the amplitude equations corresponding to projections onto triply excited determinants in the P space. In contrast, the CCSDt code in GAMESS is fully vectorized, making use of vendor-optimized BLAS routines for matrix transposition and multiplication that achieve peak performance on modern hardware. Typically, vectorization is always more efficient, and indeed, this is reflected in the slightly better performance of the CCSDt routines present in GAMESS, but the fact that the CC(P) algorithm in CCpy can remain very competitive speaks to its efficiency. In most circumstances, a vectorized implementation will be $\sim$10x faster than a devectorized one, but our CC(P) code is able to close this gap with a combination of efficient looping and removal of the CPU operations associated with permutationally redundant three-body amplitudes, which must be considered in the BLAS operations.

The flexibility of the CC(P) approach also gives us the ability to speed up our calculations by filtering the list of triply excited determinants to only retain those of the ground-state symmetry. The timings for our CC(t;3) calculations that exploit the D_{2h} symmetry group of F_2 reported in Table C.3 demonstrate that our CC(P) code can be even more efficient than the vectorized implementation in GAMESS when taking advantage of point-group symmetry to shorten the list of triply excited determinants. Indeed, with the help of spatial symmetry, the total CPU timings of the calculations performed with CCpy and GAMESS, which are 177.4 and 185.7 minutes, respectively, are reduced to 67.4 minutes. One may wonder why the speedup characterizing the calculation exploiting symmetry is only approximately a factor of three when the number of $S_z = 0$ triply excited determinants entering the CCSDt calculations is reduced by a factor of eight for the $A_g(D_{2h})$-symmetric ground state of F_2 considered here. This sub-optimal speedup is, in part, because a significant fraction of the CPU time corresponding to each iteration of the CCSDt calculation is spent constructing the $\overline{H}_N^{(2)}$ intermediates. Fortunately, the computation of $\overline{H}_N^{(2)}$ is completely vectorized. Future work may consider accelerating this step using high-performance parallel architectures, like GPUs.

Table C.3 Comparison of the computational timings characterizing the CC(t;3) calculations for the stretched F_2 molecule, as described using the aug-cc-pVQZ basis set, obtained using CCpy and GAMESS. The F–F bond length in F_2 was set to $2R_e$, where $R_e = 2.66816$ bohr defines the equilibrium geometry, and the core orbitals correlating with the 1s shells of fluorine were frozen in all post-RHF steps. The active space defining the CC(t;3) and underlying CCSDt calculations consisted of the highest-energy π_u, π_g, and σ_g shells occupied in the RHF reference and the lowest-energy σ_u orbital unoccupied in the reference state.

Software	CPU Time[a]		
	Iterative Steps[b]	Noniterative Steps[c]	Total
CCpy[d]	175.8 (7.4[f])	1.6	177.4
CCpy[e]	65.9 (2.6[f])	1.5	67.4
GAMESS[d]	183.7 (6.0[f])	2.0	185.7

[a] All reported timings, in CPU minutes, correspond to single-core runs on the Precision 7920 workstation from Dell equipped with Intel Xeon Silver 4114 2.2 GHz processor boards. The computational times associated with the execution of the integral, RHF, and integral transformation and sorting routines preceding the CC steps are ignored.

[b] In executing the iterative steps of the CC(t;3) calculations, a convergence threshold of 10^{-7} hartree was assumed. The timings corresponding to the iterative steps include the times required to construct and solve the CCSDt amplitude equations and the companion left-CCSD-like eigenstate involving the respective similarity-transformed Hamiltonians using the two-body approximation.

[c] In the language of Q spaces adopted by the CC(P;Q) formalism, the computational times required to determine the noniterative triples corrections of CC(t;3) correspond to the remaining triply excited determinants missing from the P space characterizing the CCSDt calculations.

[d] No advantage of the $D_{\infty h}$ symmetry of F_2 or its D_{2h} Abelian subgroup was taken in the post-RHF steps.

[e] The P space defining the CCSDt and CC(t;3) calculations was filtered to retain the $S_z = 0$ triply excited determinants of the $A_g(D_{2h})$ symmetry.

[f] The average time, in CPU minutes, corresponding to a single iteration of the CCSDt calculation.

C.5 Prospects for Future Improvements

To conclude this appendix, we offer a few additional remarks about our ideas for improving the current CC(P)/EOMCC(P) routines. By and large, the CC(P)/EOMCC(P) algorithm that we have presented thus far is highly efficient. There are no unnecessary FLOPs used when constructing the CC/EOMCC equations and, as just discussed, the performance of our CC(P;Q) implementation in CCpy is competitive with the vectorized routines available in GAMESS. Key technical improvements to our current implementation will come from fine-tuning the operations executed within the loops over higher–than–doubly excited determinants, or perhaps reorganization of the looping structure.

First of all, the mapping given by $f_{uvxw}(p, q, r, s)$ should be represented using a hash

table or dictionary, rather than a full multidimensional array. The latter is not only unnecessary, since only a subset of (p, q, r, s) tuples enter the problem, but also consumes excessive amounts of memory. A large $f_{uvxw}(p, q, r, s)$ object may even lead to deterioration in performance for larger problems when cache-misses become more frequent while performing lookups for lexical indices y. Indeed, the most natural and efficient data structure for $f_{uvxw}(p, q, r, s)$ is a hash table that maps tuples of integers (keys) to lexical indices y (values). In order to create this hash table, a given tuple (p, q, r, s) must be encoded to a key using a hash function. Although off-the-shelf hashing functions exist, one could avoid hash conflicts altogether by encoding each (p, q, r, s) into the linear index $I = (p-1)n_2n_3n_4+(q-1)n_3n_4+(r-1)n_4+s+1$, where n_i, $i = 1$–4, refer to the number of possible values (either n_u or n_o) that each index can take on. Modified linear indices can be defined to address triangular subarrays corresponding to conditions resulting from permutational symmetry, like $p < q$, $p < q < r$, $p < q$, $r < s$, and $p < q < r < s$. Thus, $f_{uvxw}(p, q, r, s)$ could be represented as a dictionary that uses linear indices as keys and returns lexical indices y as values. Even better, the values in $f_{uvxw}(p, q, r, s)$ may directly correspond to the starting and ending positions of a given $uvxw$-block in Φ_3, removing the need for the addressing array A_{uvxw}. Introducing a dictionary would not only reduce memory requirements and potentially speed up the code, but it would also allow us to extend our sorting schemes to sort more than four indices without introducing large arrays. This dictionary representation would be invaluable when implementing CC(P)/EOMCC(P) methods aimed at converging CCSDTQ/EOMCCSDTQ, where sorting according to six indices is necessary.

Another potentially interesting idea for future optimization involves restructuring the loops in Algorithms C.2–C.6. Currently, we perform an outer loop over higher–than–doubly excited determinants in the P space, and for each one, we locate the connected determinants and loop over them, performing computations for each one. One potential drawback of this structure is that the lexical index y is obtained using $f_{uvxw}(p, q, r, s)$ at least once per iteration of the outer loop. Given that this outer loop is generally quite long (entailing potentially

billions of iterations), performance may suffer if the the operation $y = f_{uvxw}(p, q, r, s)$ is not highly efficient. We already discussed that using a dense array for $f_{uvxw}(p, q, r, s)$ should be replaced with a hash table to reduce expensive cache misses, but even a hash table will lead to slowdowns when billions of lookups are required. Thus, we may benefit from restructuring our current algorithm so that instead of iterating over determinants directly, we loop over the distinct lexical indices (or sorted blocks) in our problem, and for each one, we locate the connected blocks of determinants. This optimization has the advantage of minimizing the number of times the map $f_{uvxw}(p, q, r, s)$ is used, which may be beneficial for larger problems.

Finally, with the previous two optimizations taken into account, the CC(P)/EOMCC(P) code is likely at near-optimal performance on a single core. Thus, it is natural to consider parallelizing routines, such as those in Algorithms C.2–C.6. To accomplish this, we may distribute the work over higher–than–doubly excited determinants to separate processes using message passing interface (MPI) instructions. Within each process, the work can be subsequently subdivided using OpenMP directives. This would represent a natural parallel scheme, similar to the one adopted in the CIPSI code in Quantum Package (see Ref. [184] for further details).

APPENDIX D

CC/EOMCC AND FCI ENERGIES FOR THE GROUND- AND EXCITED-STATE POTENTIAL CUTS OF WATER

This appendix contains the results of the CC and EOMCC calculations for the ground- and excited-state PES cuts of the water molecule, as described by the TZ basis set of Ref. [327], corresponding to the $H_2O \rightarrow H + OH$ dissociation, discussed in Section 4.2.3 of Chapter 4 of this dissertation, along with the associated full CI data taken from Refs. [161, 327]. Table D.1 compares the results of the CCSDT and EOMCCSDT calculations with full CI. The remaining Tables D.2–D.13 compare the energies of the ground and excited states of water considered in this work obtained with the CCSD/EOMCCSD, CR-CC(2,3)/CR-EOMCC(2,3), CCSDt/EOMCCSDt, CC(t;3), and adaptive CC(P;Q) approaches for the selected values of the O–H bond-breaking coordinate defining the $H_2O \rightarrow H + OH$ dissociation pathway with the parent CCSDT/EOMCCSDT data.

Table D.1 Total ground and excited electronic state energies of water, as described by the TZ basis set of Ref. [327], along the O–H bond-breaking coordinate, R_{OH}, in bohr. All energies are reported as $E + 75$ hartree. The values in parentheses are errors in the CCSDT/EOMCCSDT energies relative to FCI, in millihartree. The FCI data were taken from Refs. [327] and [161]. The lowest-energy orbital correlating with the 1s shell of oxygen was kept frozen in post-RHF calculations. Adapted from Ref. [138].

	$X\,^1\mathrm{A}'$		$1\,^1\mathrm{A}''$		$1\,^3\mathrm{A}'$	
R_{OH}	Full CI	CCSDT	Full CI	EOMCCSDT	Full CI	EOMCCSDT
1.3	−1.01567	−1.015602 (0.07)	−0.70187	−0.702504 (−0.63)	−0.65020	−0.650784 (−0.58)
1.6	−1.14894	−1.148832 (0.11)	−0.85127	−0.851860 (−0.59)	−0.79340	−0.793913 (−0.51)
1.809[a]	−1.16847	−1.168316 (0.15)	−0.88566	−0.886181 (−0.52)	−0.82511	−0.825513 (−0.40)
2.0	−1.16417	−1.163964 (0.21)	−0.90098	−0.901368 (−0.39)	−0.84272	−0.842928 (−0.21)
2.4	−1.13058	−1.130217 (0.36)	−0.92945	−0.929326 (0.12)	−0.89158	−0.891420 (0.16)
2.8	−1.09255	−1.091972 (0.58)	−0.95679	−0.956186 (0.60)	−0.93780	−0.937554 (0.25)
3.2	−1.06129	−1.060470 (0.82)	−0.97626	−0.975300 (0.96)	−0.96789	−0.967653 (0.24)
3.6	−1.03889	−1.037848 (1.04)	−0.98862	−0.987453 (1.17)	−0.98540	−0.985183 (0.22)
4.0	−1.02451	−1.023323 (1.19)	−0.99617	−0.994949 (1.22)	−0.99516	−0.994978 (0.18)
4.2	−1.01967	−1.018463 (1.21)	−0.99870	−0.997508 (1.19)	−0.99822	−0.998050 (0.17)
4.4	−1.01603	−1.014832 (1.20)	−1.00065	−0.999498 (1.15)	−1.00046	−1.000308 (0.15)
	$1\,^3\mathrm{A}''$		$1\,^1\mathrm{A}'$		$2\,^3\mathrm{A}'$	
R_{OH}	Full CI	EOMCCSDT	Full CI	EOMCCSDT	Full CI	EOMCCSDT
1.3	−0.72170	−0.722266 (−0.57)	−0.62821	−0.628899 (−0.69)	−0.55577	−0.556467 (−0.70)
1.6	−0.87187	−0.872393 (−0.52)	−0.77055	−0.771190 (−0.64)	−0.70873	−0.709251 (−0.53)
1.809[a]	−0.90726	−0.907714 (−0.45)	−0.79860	−0.799146 (−0.55)	−0.74823	−0.748568 (−0.34)
2.0	−0.92363	−0.923946 (−0.32)	−0.80575	−0.806139 (−0.39)	−0.76263	−0.762831 (−0.20)
2.4	−0.95148	−0.951300 (0.18)	−0.81062	−0.810399 (0.22)	−0.78565	−0.785421 (0.23)
2.8	−0.97420	−0.973528 (0.67)	−0.81932	−0.818515 (0.81)	−0.81625	−0.815525 (0.73)
3.2	−0.98837	−0.987281 (1.09)	−0.82860	−0.827420 (1.18)	−0.83261	−0.831462 (1.15)
3.6	−0.99640	−0.995001 (1.40)	−0.83618	−0.834787 (1.39)	−0.84103	−0.839577 (1.45)
4.0	−1.00091	−0.999363 (1.55)	−0.84172	−0.840275 (1.44)	−0.84553	−0.843920 (1.61)
4.2	−1.00236	−1.000790 (1.57)	−0.84379	−0.842366 (1.42)	−0.84693	−0.845298 (1.63)
4.4	−1.00345	−1.001888 (1.56)	−0.84546	−0.844080 (1.38)	−0.84797	−0.846350 (1.62)
	$2\,^3\mathrm{A}''$		$2\,^1\mathrm{A}'$		$2\,^1\mathrm{A}''$	
R_{OH}	Full CI	EOMCCSDT	Full CI	EOMCCSDT	Full CI	EOMCCSDT
1.3	−0.62356	−0.624198 (−0.64)	−0.54185	−0.542618 (−0.77)	−0.61400	−0.614700 (−0.70)
1.6	−0.78052	−0.781036 (−0.52)	−0.68730	−0.687943 (−0.64)	−0.76861	−0.769204 (−0.59)
1.809[a]	−0.82109	−0.821469 (−0.38)	−0.71786	−0.718367 (−0.51)	−0.80694	−0.807409 (−0.47)
2.0	−0.83571	−0.835975 (−0.26)	−0.72455	−0.724899 (−0.35)	−0.81992	−0.820292 (−0.37)
2.4	−0.82785	−0.828038 (−0.19)	−0.71631	−0.716154 (0.16)	−0.81008	−0.810444 (−0.36)
2.8	−0.80175	−0.801866 (−0.12)	−0.72567	−0.724539 (1.13)	−0.78177	−0.782217 (−0.45)
3.2	−0.77863	−0.778512 (0.12)	−0.73203	−0.730092 (1.94)	−0.75430	−0.754820 (−0.52)
3.6	−0.76634	−0.765773 (0.57)	−0.73289	−0.730473 (2.42)	−0.73276	−0.733362 (−0.60)
4.0	−0.76393	−0.763214 (0.72)	−0.72999	−0.727678 (2.31)	−0.71846	−0.719183 (−0.72)
4.2	−0.76335	−0.762712 (0.64)	−0.72738	−0.725296 (2.08)	−0.71372	−0.714539 (−0.82)
4.4	−0.76232	−0.761785 (0.54)	−0.72415	−0.722336 (1.81)	−0.71037	−0.711300 (−0.93)
	$3\,^3\mathrm{A}''$		$3\,^1\mathrm{A}'$		$3\,^3\mathrm{A}'$	
R_{OH}	Full CI	EOMCCSDT	Full CI	EOMCCSDT	Full CI	EOMCCSDT
1.3	−0.36985	−0.370374 (−0.52)	−0.39097	−0.391520 (−0.55)	−0.44370	−0.444133 (−0.43)
1.6	−0.51578	−0.516222 (−0.45)	−0.57861	−0.578996 (−0.39)	−0.61947	−0.619787 (−0.32)
1.809[a]	−0.55368	−0.554100 (−0.42)	−0.63701	−0.637243 (−0.23)	−0.67629	−0.676486 (−0.20)
2.0	−0.58208	−0.582672 (−0.59)	−0.66800	−0.668055 (−0.05)	−0.71482	−0.714899 (−0.08)
2.4	−0.60132	−0.602051 (−0.73)	−0.69672	−0.696531 (0.19)	−0.73737	−0.737345 (0.03)
2.8	−0.65695	−0.655011 (1.94)	−0.67863	−0.678272 (0.36)	−0.70581	−0.705526 (0.28)
3.2	−0.69947	−0.698187 (1.28)	−0.66287	−0.661466 (1.40)	−0.67660	−0.675863 (0.74)
3.6	−0.71715	−0.716719 (0.43)	−0.66382	−0.660620 (3.20)	−0.65564	−0.654127 (1.51)
4.0	−0.71837	−0.718451 (−0.08)	−0.66878	−0.665985 (2.80)	−0.64297	−0.640046 (2.92)
4.2	−0.71706	−0.717253 (−0.19)	−0.66943	−0.665657 (3.77)	−0.63921	−0.635295 (3.92)
4.4	−0.71581	−0.716073 (−0.26)	−0.66889	−0.664958 (3.93)	−0.63687	−0.631806 (5.06)

[a] The equilibrium value of the O–H bond length obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.2 The total electronic energies, reported as errors relative to CCSDT in millihartree, obtained with the CCSD, CR-CC(2,3), CCSDt, CC(t;3), and adaptive CC(P;Q) approaches for the $X\,^1\mathrm{A}'$ state of the water molecule, as described by the TZ basis set of Ref. [327], along the O–H bond-breaking coordinate, R_{OH}, in bohr. The lowest-energy orbital correlating with the 1s shell of oxygen was kept frozen in post-RHF steps. Adapted from Ref. [138].

					%T = 1[d]		%T = 2[e]	
R_{OH}	CCSD	CR(2,3)[a]	CCSDt[b]	CC(t;3)[c]	CC(P)	CC(P;Q)	CC(P)	CC(P;Q)
1.3	2.771	−0.226	2.068	−0.182	1.696	−0.163	1.374	−0.133
1.6	3.063	−0.269	2.229	−0.205	1.713	−0.171	1.331	−0.124
1.809[f]	3.307	−0.298	2.317	−0.216	1.857	−0.196	1.390	−0.144
2.0	3.562	−0.325	2.361	−0.219	2.141	−0.230	1.652	−0.175
2.4	4.230	−0.398	2.328	−0.206	2.689	−0.335	2.132	−0.258
2.8	5.150	−0.500	2.187	−0.178	2.768	−0.379	2.087	−0.284
3.2	6.389	−0.613	2.023	−0.151	2.575	−0.369	1.898	−0.275
3.6	7.929	−0.724	1.872	−0.130	2.338	−0.350	1.667	−0.231
4.0	9.622	−0.828	1.753	−0.116	2.115	−0.308	1.469	−0.195
4.2	10.448	−0.875	1.709	−0.112	2.043	−0.276	1.405	−0.183
4.4	11.224	−0.918	1.674	−0.110	1.997	−0.247	1.374	−0.162

[a] CR-CC(2,3) calculations.

[b] CCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] CC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] CC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.3 The total electronic energies, reported as errors relative to EOMCCSDT in millihartree, obtained with the EOMCCSD, CR-EOMCC(2,3), EOMCCSDt, CC(t;3), and adaptive CC(P;Q) approaches for the $1\,^1\mathrm{A}''$ state of the water molecule, as described by the TZ basis set of Ref. [327], along the O–H bond-breaking coordinate, R_{OH}, in bohr. The lowest-energy orbital correlating with the 1s shell of oxygen was kept frozen in post-RHF steps. Adapted from Ref. [138].

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.081	0.917	1.601	0.708	2.404	0.615	2.049	0.549
1.6	0.049	1.125	1.661	0.718	2.458	0.612	1.983	0.595
1.809[f]	0.301	1.145	1.710	0.702	2.573	0.656	2.124	0.587
2.0	1.015	1.005	1.794	0.648	2.943	0.700	2.480	0.660
2.4	4.319	0.505	1.972	0.485	3.802	0.620	3.093	0.534
2.8	7.904	−0.067	1.873	0.443	3.944	0.587	2.802	0.527
3.2	10.917	−0.875	1.705	0.480	3.973	0.696	2.569	0.574
3.6	13.263	−1.810	1.594	0.537	3.637	0.869	2.261	0.681
4.0	14.933	−2.700	1.542	0.595	3.130	0.934	1.935	0.686
4.2	15.537	−3.095	1.531	0.623	2.889	1.021	1.897	0.725
4.4	16.012	−3.448	1.527	0.649	2.995	0.999	1.789	0.725

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.4 Same as Table D.3 for the $1\,^3\mathrm{A}'$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.351	0.884	1.550	0.693	2.150	0.495	1.827	0.457
1.6	−0.188	1.024	1.603	0.719	2.191	0.483	1.721	0.458
1.809[f]	0.185	1.055	1.646	0.754	2.192	0.474	1.638	0.465
2.0	1.118	1.019	1.758	0.776	2.435	0.514	1.865	0.482
2.4	3.596	1.099	2.001	0.909	2.461	0.712	2.058	0.610
2.8	4.501	1.163	1.915	1.005	2.330	0.873	1.711	0.705
3.2	4.653	1.176	1.801	1.043	2.061	0.959	1.469	0.738
3.6	4.488	1.208	1.740	1.093	1.801	1.031	1.272	0.727
4.0	4.166	1.263	1.716	1.133	1.710	1.046	1.131	0.708
4.2	3.977	1.294	1.713	1.153	1.681	1.067	1.104	0.712
4.4	3.783	1.326	1.714	1.184	1.660	1.066	1.053	0.708

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.5 Same as Table D.3 for the $1\,^3A''$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.325	0.887	1.545	0.722	2.427	0.679	2.012	0.595
1.6	−0.205	1.088	1.596	0.730	2.175	0.658	1.893	0.624
1.809[f]	0.046	1.110	1.641	0.716	2.537	0.728	2.029	0.636
2.0	0.757	0.999	1.732	0.669	2.650	0.777	2.370	0.724
2.4	3.847	0.663	1.957	0.532	3.621	0.764	3.151	0.696
2.8	7.203	0.326	1.890	0.490	4.063	0.771	3.147	0.666
3.2	10.310	−0.213	1.721	0.517	4.246	0.874	2.851	0.679
3.6	13.094	−0.899	1.598	0.566	3.869	1.088	2.535	0.728
4.0	15.388	−1.594	1.533	0.621	3.407	0.988	2.025	0.653
4.2	16.323	−1.909	1.518	0.648	3.377	1.035	1.923	0.670
4.4	17.120	−2.192	1.510	0.687	3.231	1.033	1.791	0.645

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.6 Same as Table D.3 for the $1\,^1\mathrm{A}'$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.018	0.964	1.466	0.660	2.364	0.472	1.092	0.435
1.6	0.298	1.129	1.495	0.695	2.455	0.475	1.918	0.418
1.809[f]	0.890	1.189	1.530	0.706	2.676	0.520	1.984	0.478
2.0	1.950	1.147	1.599	0.691	3.053	0.547	2.369	0.518
2.4	5.922	0.781	1.829	0.545	4.025	0.434	3.117	0.407
2.8	10.351	0.088	1.853	0.489	4.809	0.490	3.282	0.467
3.2	14.047	−0.790	1.718	0.462	5.056	0.632	3.147	0.504
3.6	17.027	−1.737	1.596	0.501	4.533	0.746	2.656	0.540
4.0	19.348	−2.649	1.527	0.551	3.678	0.695	2.258	0.575
4.2	20.261	−3.063	1.509	0.578	3.414	0.734	2.070	0.561
4.4	21.024	−3.448	1.499	0.600	3.296	0.744	1.915	0.532

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.7 Same as Table D.3 for the $2\,^3\mathrm{A}'$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.930	0.877	1.291	0.705	2.160	0.529	2.026	0.538
1.6	−0.057	0.950	1.545	0.708	2.159	0.523	1.989	0.506
1.809[f]	0.958	1.006	1.781	0.733	2.488	0.593	2.008	0.545
2.0	1.669	1.122	1.913	0.781	2.878	0.742	2.478	0.637
2.4	4.621	0.995	2.117	0.722	3.750	0.729	2.898	0.656
2.8	8.219	0.352	1.961	0.545	4.360	0.585	3.051	0.540
3.2	11.561	−0.328	1.750	0.524	4.423	0.660	2.913	0.533
3.6	14.620	−1.109	1.608	0.551	4.030	0.739	2.389	0.519
4.0	17.204	−1.892	1.531	0.591	3.674	0.741	2.059	0.462
4.2	18.274	−2.253	1.510	0.612	3.555	0.769	1.918	0.450
4.4	19.197	−2.590	1.497	0.632	3.695	0.701	1.818	0.439

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.8 Same as Table D.3 for the $2\,^3\mathrm{A}''$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.547	0.933	1.370	0.736	2.392	0.726	2.314	0.655
1.6	0.286	1.044	1.620	0.711	2.430	0.672	2.165	0.608
1.809[f]	1.181	1.049	1.825	0.679	2.808	0.710	2.323	0.623
2.0	1.877	1.052	1.927	0.647	3.071	0.840	2.540	0.714
2.4	2.210	1.111	1.740	0.596	3.906	1.066	3.334	0.808
2.8	2.703	1.472	1.477	0.543	4.215	1.145	3.384	0.777
3.2	5.070	2.801	1.355	0.505	5.281	1.284	3.706	0.831
3.6	11.868	6.744	1.409	0.484	7.607	1.634	4.470	0.844
4.0	21.864	11.891	1.504	0.481	9.244	1.753	4.926	0.804
4.2	25.767	13.294	1.503	0.483	9.172	1.595	4.889	0.736
4.4	28.470	13.976	1.482	0.625	9.097	1.523	4.806	0.705

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.9 Same as Table D.3 for the $2\,^1\mathrm{A}'$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.720	0.865	1.300	0.679	2.302	0.473	2.069	0.477
1.6	0.014	0.918	1.509	0.663	2.197	0.400	1.830	0.405
1.809[f]	0.920	0.908	1.677	0.643	2.655	0.446	1.908	0.414
2.0	1.886	0.872	1.785	0.627	2.819	0.433	2.219	0.464
2.4	5.132	1.267	1.853	0.772	4.413	0.597	3.264	0.533
2.8	11.838	0.215	1.896	0.892	3.710	0.786	2.598	0.611
3.2	18.528	−2.178	2.218	0.868	3.149	0.999	2.364	0.656
3.6	25.315	−4.968	2.706	0.743	3.299	1.109	2.405	0.753
4.0	30.370	−6.883	3.185	0.743	3.606	1.101	2.504	0.912
4.2	31.801	−7.063	3.377	0.646	3.653	1.161	2.440	0.841
4.4	32.540	−6.732	3.531	0.621	3.722	1.090	2.353	0.731

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.10 Same as Table D.3 for the $2\,^1\mathrm{A}''$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	−0.688	0.912	1.340	0.722	2.153	0.702	2.209	0.624
1.6	−0.078	0.963	1.582	0.684	2.202	0.607	2.049	0.583
1.809[f]	0.649	0.905	1.776	0.642	2.565	0.652	2.171	0.593
2.0	1.252	0.845	1.875	0.604	2.685	0.777	2.465	0.686
2.4	1.371	0.776	1.717	0.562	3.247	0.890	2.886	0.701
2.8	1.147	0.846	1.483	0.523	3.039	0.805	2.577	0.578
3.2	1.309	1.103	1.348	0.493	3.586	0.790	2.588	0.588
3.6	2.177	1.642	1.289	0.473	3.757	0.783	2.493	0.544
4.0	4.156	2.805	1.274	0.461	4.680	1.140	2.823	0.734
4.2	5.598	3.693	1.277	0.459	4.871	1.217	2.820	0.637
4.4	7.284	4.759	1.283	0.460	5.282	1.208	2.800	0.098

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.11 Same as Table D.3 for the $3\,^3\mathrm{A}''$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	0.779	1.263	1.977	0.604	2.175	0.807	2.001	0.699
1.6	1.104	1.518	2.431	0.662	2.645	0.864	2.261	0.735
1.809[f]	1.241	1.660	2.825	0.663	2.800	1.019	2.330	0.823
2.0	1.399	1.643	3.427	0.557	3.421	1.199	2.768	0.962
2.4	3.843	2.230	5.071	0.322	4.367	2.134	3.916	1.551
2.8	62.239	37.018	2.134	0.133	25.551	6.740	12.680	2.195
3.2	65.442	6.569	2.247	0.531	14.778	2.541	8.348	1.144
3.6	53.786	−0.457	1.909	0.511	10.641	1.780	6.420	0.852
4.0	42.844	−4.801	1.606	0.741	7.891	1.486	4.998	0.678
4.2	39.743	−5.151	1.522	0.737	7.713	1.491	4.603	0.665
4.4	38.464	−4.513	1.470	0.496	7.241	1.373	4.380	0.623

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.12 Same as Table D.3 for the $3\,^1A'$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	1.189	1.471	1.698	0.995	2.737	0.526	2.168	0.425
1.6	1.674	1.605	1.820	1.052	2.551	0.534	2.005	0.497
1.809[f]	2.399	1.598	1.870	1.042	2.741	0.624	1.855	0.513
2.0	3.392	1.517	1.874	1.010	3.087	0.770	2.392	0.631
2.4	5.272	0.714	1.659	0.795	4.087	0.584	3.029	0.482
2.8	6.462	0.852	1.586	0.684	3.981	0.571	2.994	0.525
3.2	12.832	2.422	1.712	0.742	5.674	1.479	3.886	0.732
3.6	23.485	5.089	1.782	0.676	7.575	1.980	4.353	0.674
4.0	29.845	5.471	1.661	0.449	8.569	2.733	4.717	1.418
4.2	30.434	3.564	1.493	0.643	7.649	1.735	3.591	0.561
4.4	32.035	2.203	1.411	0.754	8.006	1.133	3.338	0.612

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.

Table D.13 Same as Table D.3 for the $3\,^3A'$ state.

					%T = 1[d]		%T = 2[e]	
R_{OH}	EOMCCSD	CR(2,3)[a]	EOMCCSDt[b]	CC(t;3)[c]	EOMCC(P)	CC(P;Q)	EOMCC(P)	CC(P;Q)
1.3	0.720	1.272	1.929	0.985	2.377	0.498	1.910	0.414
1.6	1.352	1.531	1.980	1.066	2.595	0.625	2.005	0.558
1.809[f]	2.074	1.592	1.991	1.045	2.694	0.734	1.933	0.572
2.0	2.950	1.393	2.060	0.923	3.051	0.732	2.542	0.629
2.4	3.158	0.986	1.760	0.674	3.907	0.704	3.037	0.618
2.8	4.223	1.175	1.541	0.726	3.931	0.702	2.879	0.580
3.2	6.986	1.528	1.452	0.772	4.083	0.799	2.817	0.580
3.6	11.320	2.330	1.412	0.802	4.111	1.203	2.793	0.676
4.0	16.730	3.829	1.379	0.800	4.471	1.611	3.003	0.858
4.2	19.566	5.000	1.363	0.777	4.984	1.847	3.134	0.958
4.4	22.282	6.656	1.348	0.826	5.525	2.073	3.323	0.996

[a] CR-EOMCC(2,3) calculations.

[b] EOMCCSDt calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[c] CC(t;3) calculations using the active space consisting of the three highest occupied and two lowest unoccupied RHF orbitals.

[d] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 1% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[e] EOMCC(P) and CC(P;Q) calculations using P spaces consisting of all singly and doubly excited determinants and 2% of triply excited determinants identified by the adaptive CC(P;Q) algorithm.

[f] The equilibrium value of the O–H bond length in the ground electronic state of water, as obtained in Ref. [327] using the CCSD/cc-pVTZ method.